Theor Appl Genet (2011) 122:705–722
DOI 10.1007/s00122-010-1480-9

ORIGINAL PAPER

# Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme

**J. Corbi · M. Debieu · A. Rousselet · P. Montalent · M. Le Guilloux · D. Manicacci · M. I. Tenaillon**

Received: 6 April 2010 / Accepted: 22 October 2010 / Published online: 9 November 2010


**Abstract** Maize domestication from teosinte (*Zea mays* ssp. *parviglumis*) was accompanied by an increase of kernel size in landraces. Subsequent breeding has led to a diversification of kernel size and starch content among major groups of inbred lines. We aim at investigating the effect of domestication on duplicated genes encoding a key enzyme of the starch pathway, the ADP-glucose pyrophosphorylase (AGPase). Three pairs of paralogs encode the AGPase small (SSU) and large (LSU) subunits mainly expressed in the endosperm, the embryo and the leaf. We first validated the putative sequence of $LSU_{leaf}$ through a comparative expression assay of the six genes. Second, we investigated the patterns of molecular evolution on a 2 kb coding region homologous among the six genes in three panels: teosintes, landraces, and inbred lines. We corrected for demographic effects by relying on empirical distributions built from 580 previously sequenced ESTs. We found contrasted patterns of selection among duplicates: three genes exhibit patterns of directional selection during domestication ($SSU_{end}$, $LSU_{emb}$) or breeding ($LSU_{leaf}$), two exhibit patterns consistent with diversifying ($SSU_{leaf}$) and balancing selection ($SSU_{emb}$) accompanying maize breeding. While patterns of linkage disequilibrium did not reveal sign of coevolution between genes expressed in the same organ, we detected an excess of non-synonymous substitutions in the small subunit functional domains highlighting their role in AGPase evolution. Our results offer a different picture on AGPase evolution than the one depicted at the Angiosperm level and reveal how genetic redundancy can provide flexibility in the response to selection.

Communicated by J. Yu.

D. Manicacci and M. I. Tenaillon contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-010-1480-9) contains supplementary material, which is available to authorized users.

J. Corbi · M. Le Guilloux · M. I. Tenaillon (✉)
CNRS, UMR 0320/UMR 8120 Génétique Végétale,
Ferme du Moulon, 91190 Gif sur Yvette, France
e-mail: tenaillon@moulon.inra.fr

M. Debieu · D. Manicacci
Université Paris Sud, UMR 0320/UMR
8120 Génétique Végétale, Ferme du Moulon,
91190 Gif sur Yvette, France

M. Debieu
Department of Plant Breeding and Genetics,
Max-Planck Institute for Plant Breeding Research,
Carl-Von-Linné-Weg 10, 50829 Cologne, Germany

A. Rousselet · P. Montalent
INRA, UMR 0320/UMR 8120 Génétique Végétale,
Ferme du Moulon, 91190 Gif sur Yvette, France

**Abbreviations**
SSU Small subunit
LSU Large subunit

## Introduction

The origin and adaptation of crop species result from domestication per se and subsequent genetic improvement. Numerous studies have explored the genetic consequences of these processes in agronomically important plants. These studies typically aim to characterize the loss of genetic diversity due to domestication and breeding and to identify the genetic determinants of the cultivated

phenotype. The loss of diversity associated with domestication and improvement bottleneck(s) covers a wide range of variation from around 20% in maize (Tenaillon et al. 2004), to 50% in soybean (Hyten et al. 2006), 55% in sunflower (Liu and Burke 2006), 69% in bread wheat (Haudry et al. 2007), 80 and 90% in rice subspecies *indica* and *japonica*, respectively (Zhu et al. 2007), and 87% in barley (Kilian et al. 2006). In addition, several genes targeted by selection during domestication and post domestication processes have been identified (for a review see Doebley et al. 2006; Glémin and Bataillon 2009). To date, these studies have generated two important observations. First, most genes that contribute to phenotypes associated with domestication are transcriptional regulators, but a broader set of genes—including transcriptional regulators and structural genes—has contributed to varietal differences (Doebley et al. 2006). Second, selection during domestication has often proceeded from standing genetic variation rather than from de novo mutations, as evidenced by the presence of cryptic variability in wild populations for domestication-related traits (Burke et al. 2002; Jaenicke-Despres et al. 2003; Nesbitt and Tanksley 2002; Weber et al. 2007).

Maize has received much attention because of its agronomic importance. Both archeological (Piperno and Flannery 2001) and molecular data (Matsuoka et al. 2002) suggest that modern maize arose from a single domestication event from *Zea mays* ssp. *parviglumis* (thereafter called teosinte) around 9,000 years ago in Mexico. Once domesticated, maize cultivation expanded throughout the Americas and was subsequently introduced in Europe (Rebourg et al. 2003). Overall, up to 4% of the maize genome may have been targeted by selection and thus potentially contributed to the cultivated phenotype (Wright et al. 2005). Moreover, this estimate likely stands as a lower bound given the reduced power of the methods used to detect selection (Tenaillon and Tiffin 2008; Teshima et al. 2006). Top-down and bottom-up approaches have led to the identification of several genes or candidate regions involved in the genetic determination of the cultivated phenotype, notably genes/regions controlling the architecture of the kernel, of the inflorescence and the architecture of the plant (Clark et al. 2004; Doebley et al. 1997; Gallavotti et al. 2004; Hanson et al. 1996; Palaisa et al. 2004, 2003; Tenaillon et al. 2004; Vigouroux et al. 2002; Vollbrecht et al. 2005; Wang et al. 2005; Weber et al. 2007; Wright et al. 2005; Yamasaki et al. 2005; Zhao et al. 2008).

Maize strongly differs from teosinte for kernel content. On average, starch content represents 53% of kernel dry matter in teosinte but 73% in maize (Flint-Garcia et al. 2009). The starch biosynthesis is complex, consisting of over 20 genes (Myers et al. 2000; Pan 2000). The upstream part of this pathway controls starch production, whereas the downstream part governs the ratio of amylose over amylopectin that affects pasting properties and starch gelatinization. This pathway has been scanned for evidence of selection in panels of maize inbred lines in three previous studies (Fan et al. 2009; Whitt et al. 2002; Wilson et al. 2004). Whitt et al. (2002) focused on six genes, *Amylose extender1* (*Ae1*), *Brittle2* (*Bt2*), *Shrunken1* (*Sh1*), *Shrunken2* (*Sh2*), *Sugary* (*Su1*) and *Waxy* (*Wx*). *Bt2* and *Sh2* genes encode the small (SSU) and the large (LSU) subunits of the ADP-glucose pyrophosphorylase (AGPase) expressed in the endosperm, named thereafter $SSU_{end}$ and $LSU_{end}$. Whitt et al. (2002) uncovered evidence of positive selection on three of the six genes (*Ae1*, *Su1* and $SSU_{end}$), suggesting that these genes contribute to agronomic phenotypes. In contrast, $LSU_{end}$ has evolved under strong selective constraint even prior to domestication (Manicacci et al. 2007). Note, however, that the signatures of selection depend on the sample used. For example, in Chinese inbred lines characterized by a glutinous starch (i.e., small amount of amylose), the *waxy1* locus (*wx1*) was targeted by selection (Fan et al. 2009) even though it was not found to be under selection by Whitt et al. (2002). A limitation of these studies however resides in the confusion of two selective steps, domestication and plant breeding. This is critical for at least two reasons. First, we expect different genes to have been targeted during each process. For instance, domestication per se may have affected starch production whereas pasting properties may have been targeted during plant breeding. Second, domestication is generally accompanied by strong directional selection whereas breeding may involve diversifying selection to enhance differences between varieties (Doebley et al. 2006). The confusion between those two types of selection may be misleading our search for the footprints of selection.

In the present study, we focus on genes encoding AGPase, the enzyme that catalyzes the rate-limiting reaction that leads to the production of ADP-glucose from Glucose-1-phosphate and ATP. In maize, six paralogous genes are known to encode AGPase. They all derive from a single ancestral gene that has undergone several rounds of duplication, including a first duplication common to all terrestrial plants (Georgelis et al. 2008; Patron and Keeling 2005) that led to the specialization of a large and a small subunit (LSU and SSU, respectively), and a last duplication specific to maize (Rosti and Denyer 2007). The three pairs of paralogs are expressed in different tissues, one pair being mainly expressed in the endosperm cytosol, $LSU_{end}$ (*Shrunken2* or *Sh2*; Bhave et al. 1990) and $SSU_{end}$ (*Brittle2* or *Bt2*; Bae et al. 1990), one in the embryo plastids, hereafter $LSU_{emb}$ (*Agplemzm*; Giroux et al. 1995) and $SSU_{emb}$ (*AGP2* or *Agpsemzm*; Giroux and Hannah 1994; Hannah et al. 2001), and one in the leaf plastids, hereafter

$LSU_{leaf}$ and $SSU_{leaf}$ (*L2* or *Agpslzm*; Hannah et al. 2001; Prioul et al. 1994).

AGPase catalysis and regulation have been investigated in *E. coli* (Ballicora et al. 2007; Bejar et al. 2006; Lee and Preiss 1986) and potato tubers (Fu et al. 1998; Ballicora et al. 1998; Frueauf et al. 2003; Kavakli et al. 2001; Kim et al. 2007). The studies have identified regions crucial for enzyme activity, including sites of substrate fixation, catalytic sites and allosteric regulation motifs. In maize, mutagenesis experiments have also defined important functional domains, such as the $LSU_{end}$ C terminus region involved in the enzyme allostery (Giroux et al. 1996), the $SSU_{end}$ N terminus involved in enzyme stability (Linebarger et al. 2005), and the between-subunit-interaction region involved in allosteric regulation (Boehlein et al. 2010; Georgelis et al. 2009).

Thus far, only AGPase paralogs that are expressed in the endosperm have been investigated with regard to genetic diversity and selection during maize domestication and breeding. Most of the starch resources used during the heterotrophic stage of the seedling growth is provided by the endosperm, but it may be of interest to study the non-endosperm versions of the genes. For example, starch accumulation in the embryo may also play a role in seedling development and fitness. Moreover, transient starch is produced in leaves during photosynthesis and hydrolyzed at night, providing the plant with sugar supplies throughout the night (Zeeman et al. 2007). Leaf-expressed AGPase is reversibly activated by a redox post-translational reaction (reduction of a disulfide bridge between small subunits), leading to a partial activation during the day and inactivation during the night (Hendriks et al. 2003; Geigenberger et al. 2005).

Because, AGPase activity is essential in all three organs that produce starch and because starch content and quality were likely targeted during maize domestication and breeding, we decided to investigate patterns of selection across the three pairs of duplicated genes before, during, and after domestication. As a first step, we characterize the sequence of the unknown large subunit paralog expressed in leaves and confirm its identification through comparative expression analysis among paralogs. We then sequence a ~2 kb homologous region of the six paralogs in panels of teosinte, landraces and inbred lines. In each pair, we analyzed a common sample of individuals belonging to each of the three panels. This sequenced region represents part of the coding portion of the gene, including some of the sites that were identified as associated with starch viscosity and amylose content in $LSU_{end}$ (Wilson et al. 2004). We search for selection in these data by accounting for demographic effects using a previously published data set containing sequencing information in 580 ESTs (Wright et al. 2005). Finally, we studied coevolution between subunit genes preferentially expressed in the same organ, paying specific attention to a previously characterized motif of 55 amino acids that plays a crucial role in the interaction between AGPase subunits (Cross et al. 2005; Hwang et al. 2005).

## Materials and methods

We evaluated DNA sequence diversity at six paralogous genes encoding the small (SSU) and large (LSU) AGPase subunits expressed in different organs: the endosperm, the embryo and the leaf. For the sake of clarity, paralogous gene names were coded according to the subunit type (small or large) and their major organ of expression.

### $LSU_{leaf}$ identification

Because $LSU_{leaf}$ was not identified prior to our study, we performed a Blastn search using the sequence of the AGPase large subunit gene expressed in *Hordeum vulgare* leaves as a query (acc. U66876). We retrieved a cDNA sequence from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) corresponding to the 3′ end of the gene (acc. AY108622) and a genomic sequence from maizeGDB version 4 (http://www.maizegdb.org/) corresponding to the 5′-end part of the gene (acc. AZM4_51892). We then concatenated both sequences to design specific primers for sequence amplification in all panels. To confirm that this sequence was preferentially expressed in leaves, we performed an RT-PCR expression assay.

### Expression assay

We compared the expression profiles of the six paralogous genes ($SSU_{end}$, $LSU_{end}$, $SSU_{emb}$, $LSU_{emb}$, $SSU_{leaf}$, and $LSU_{leaf}$) in several organs using semi-quantitative reverse transcription (RT)-PCR. This experiment was conducted using the inbred line *MBS847* (Mike Brayton Seeds breeding company). RNA was extracted from organs in which the genes are supposedly preferentially expressed, namely the endosperm (dissected from kernels collected 21 days after pollination), the embryo (dissected from kernels collected 21 days after pollination) and the leaves (including young autotrophic leaves, heterotrophic leaves and the ear leaf) as well as six other organs including spathes, silks, ovules, young kernels (collected 3 days after pollination), pollen grains and roots. Total RNA was extracted using the RNeasy plant mini kit (Qiagen) and DNase-treated according to the manufacturer's instructions (Ambion). First-strand cDNA synthesis was carried out with the RevertAid Moloney murine leukemia virus reverse transcriptase (Fermentas) and random hexamers (Pharmacia) using 5 µg of total RNA. $2.5 \times 10^5$ copies of

GeneAmplimer pAW109 RNA (Applied Biosystems) were added to each reaction as a positive control RT. Constitutive expression of *18S* ribosomal gene was then used to standardize cDNA quantity among organs. Expression of *18S* was tested in a 25 µl reaction mix containing Taq buffer, 3 mM of $MgCl_2$, 200 µM of dNTP, 0.5 µM of each primer, and 1.25 units of Taq polymerase (Qiagen). Thermocycling conditions were 5 min at 94°C, then 10 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 30 s, the final elongation step was 72°C for 5 min (PCR primers forward 5′-STD5-3′ and reverse, 5′-STD6-3′). To control for successful reverse transcription in each sample, the same PCR conditions were applied using GeneAmplimer pAW109 PCR primers (5′-CATGTCAAATTTCACTGCTTCATC-3′ and 5′-TGACCACCCAGCCATCCTT-3′) with 27 cycles.

Specific primers were designed for the six paralogous genes (Online Resource 1) and tested for specificity by sequencing PCR amplification products obtained from genomic DNA before being used in RT-PCR assays. Because alternative splicing of the exon 1 of $SSU_{end}$ leads to a cytosolic form expressed in endosperm and an additional minor plastidial form expressed in leaves (Cossegal et al. 2008; Rosti and Denyer 2007), we designed two pairs of primers that specifically amplify each form (Online Resource 1). Genes were amplified in a 25 µl reaction mix containing Taq buffer, 200 µM of dNTP, 0.4 µM of each primer, 1 unit of Taq polymerase (Qiagen) and additional $MgCl_2$ at various concentrations depending on the gene (200 µM for $SSU_{end}$, 100 µM for $LSU_{end}$, 150 µM for $SSU_{emb}$, 100 µM for $LSU_{leaf}$) as well as 1× Q buffer (Qiagen) for $SSU_{leaf}$ and $LSU_{leaf}$. The number of PCR cycles for each gene (Online Resource 1) was adjusted to be in the linear phase of amplification in the main organ of expression, i.e., the product was visualized on agarose gels stained with ethidium bromide, and the number of cycles was defined in order to obtain a band of comparable intensity as the 20 ng band of the SmartLadder $M_r$ marker (MW-1700; Eurogentec).

## Plant material

Paralogous genes expressed in embryo and leaves ($SSU_{emb}$, $LSU_{emb}$, $SSU_{leaf}$, and $LSU_{leaf}$) were sequenced in a common sample containing 26 inbred lines and 12 landraces (Online Resource 2). Inbred lines and landrace accessions were chosen in a core collection defined by SSR analysis by Camus-Kulandaivelu et al. (2006). Because data from $LSU_{end}$ were already available for an extended sample of 51 inbred lines and seven landraces (Manicacci et al. 2007), we sequenced its partner, $SSU_{end}$, in this same sample, which differs substantially from the one sequenced for the two other pairs of paralogs. Finally, a common sample of 15 teosintes was sequenced across all six paralogs. The sequence data obtained, by using genomic DNA extracted from leaf material, were deposited in the EMBL/GenBank Data Library under accession numbers: DQ019876–DQ019928, HM749333–HM749341 for $LSU_{end}$, HM749342–HM749614 for $SSU_{end}$, $SSU_{emb}$, $LSU_{emb}$, $SSU_{leaf}$, and $LSU_{leaf}$. One individual of *Tripsacum dactyloides* was used as outgroup (CIMMYT accession 4563).

## DNA amplification, cloning, and sequencing

For the six paralogous genes ($SSU_{end}$, $LSU_{end}$, $SSU_{emb}$, $LSU_{emb}$, $SSU_{leaf}$, and $LSU_{leaf}$) we PCR-amplified a homologous genomic region spanning ~2 kb and containing both introns and exons (Fig. 1). PCR conditions and primers are available from Online Resource 3. For inbred lines, direct sequencing of PCR product was performed by Genoscreen (Lille, France) or GENOME EXPRESS, Inc. (Grenoble, France). For landraces, teosintes and *Tripsacum dactyloides*, which are potentially heterozygous, PCR products were cloned either using the pGEM-TA (Promega) or the TOPO cloning (Invitrogen) technology and a single clone was sequenced. We verified singleton polymorphisms either by direct resequencing on genomic DNA, searching for the presence of double versus single peaks, or by reamplifying, recloning, and resequencing several clones per individual when frequent indels prevented us from reading bi-allelic sequences (this last procedure was applied for $LSU_{end}$). Singletons that could not be verified were considered as PCR errors and corrected in the

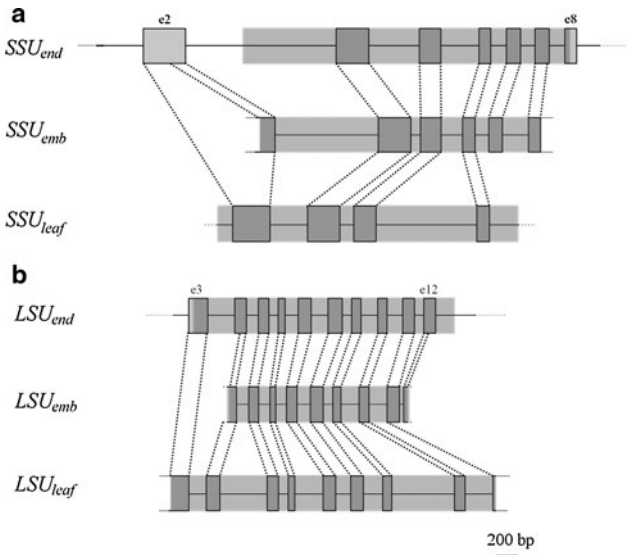


**Fig. 1** Structure of the paralogous genes encoding AGPase small (**a**) and large (**b**) subunits in endosperm, embryo and leaf. *Boxes* symbolize exons, separated by *lines* representing introns. *Vertical dashed lines* connecting genes indicate homology in exonic regions. *Shaded areas* correspond to the regions that were sequenced in this study

alignments. Sequences were assembled and aligned into contigs using Staden v.5.1 (Staden 1996) and manually edited using BioEdit v.4.8.8 (Hall 1999). Coding regions were assigned in all six genes according to previously described ORFs in the $LSU_{end}$ (Shaw and Hannah 1992), $SSU_{end}$ (Bae et al. 1990), and $SSU_{leaf}$ (Prioul et al. 1994). Note that coding regions are well conserved among genes encoding a given subunit, which facilitates their annotation.

## Mapping

In order to validate the mapping position of each amplified fragment, polymorphisms between inbred lines B73 and Mo17, or between FV2 and F252, were used to map the fragments using the mapping populations IBM or LHRF, respectively, available at UMR de Génétique Végétale, Le Moulon (Falque et al. 2005). This allowed for validating the position of the $LSU_{end}$ on chromosome 3 bin 09 (Bhave et al. 1990), $SSU_{end}$ on chromosome 4 bin 05 (Teas and Teas 1953), and $SSU_{leaf}$ on chromosome 3 bin 07 (Causse et al. 1996). The $LSU_{leaf}$ was located on chromosome 1 bin 09–10, as expected from the genomic sequence (AZM4-51892). Genes expressed in the embryo were located on chromosome 6 bin 07 and chromosome 2 bin 06–08 for $LSU_{emb}$ and $SSU_{emb}$, respectively.

## Nucleotide diversity and differentiation

We estimated per site nucleotide diversity $\pi$ (Nei 1987) and Watterson's $\theta$ (1975) at all sites or silent sites ($\pi_{TOT}$, $\pi_{sil}$, $\theta_{TOT}$, $\theta_{sil}$) using DnaSP v5 software (Rozas 2009) on four panels: American lines, European lines, landraces, and teosintes (Online Resource 3). Significant differences in diversity between panels were assessed by a 1,000-replicate permutation procedure. Nucleotide differentiation between panels was measured by *Fst* and *Ks** (Hudson et al. 1992) and their significance was estimated from a 1,000-replicate permutation procedure following Achaz et al. (2004). Owing to the fact that we sequenced a homologous region (Fig. 1) among paralogs that derived by duplication from a single ancestral copy, we expect to control for structural constraints linked to the partition in introns and exons and the presence of particular functional domains.

## Neutrality tests

We tested whether deviation from neutrality was detectable before, during, or after domestication by applying neutrality tests to the different panels that trace back maize evolutionary history: teosintes, landraces, American and European inbred lines. To assess whether selection occurred during domestication or breeding we compared teosintes to landraces, and landraces to inbred lines respectively. To increase statistical power, we also performed neutrality tests after grouping all cultivated maize (landraces and inbred lines) in a single panel.

Neutrality tests included Fu's *Fs* (1997), Tajima's *D* (Tajima 1989), Fu and Li's *F* and *D* (1993) and Fay and Wu's *H* (2000). All tests were computed in DnaSP v5 (Rozas 2009) except Fay and Wu's *H* which was computed online using Guillaume Achaz' website (http://wwwabi.snv.jussieu.fr/achaz/neutrality-test.html). *Tripsacum dactyloides* was used as outgroup when required (Fu and Li's *D* and *F*, Fay and Wu's *H*). As recombination influences the significance of all tests, we estimated the population recombination rate, *C*, using LDhat v2.1 (McVean et al. 2002) for all panels and averaged these likelihood estimates over five simulations. *P* values for the neutrality tests were obtained using the coalescent simulator of DnaSP v5 with recombination.

We performed a multilocus likelihood HKA test (Wright and Charlesworth 2004) which accounts for differences in mutation rate between loci. We compared the likelihood of a model in which our AGPase loci were neutral against a model in which they were under selection. Besides testing for deviation from neutrality at the focus loci, we estimated *k*, the degree to which the diversity level increased ($k > 1$) or decreased ($k < 1$) under selection. We compared AGPase genes to 11 reference loci that had been previously sequenced on a sample of 9 tropical or dent American lines (Tenaillon et al. 2001). Because control loci were sequenced on American lines, we only considered American lines in our sample (see Online Resource 2 for further details). We first applied the HKA test considering 11 control neutral loci and one of the six AGPase loci ($SSU_{end}$, $LSU_{end}$ $SSU_{emb}$, $LSU_{emb}$, $SSU_{leaf}$, $LSU_{leaf}$) independently, and further considered nested models in which several AGPase loci can be simultaneously considered as targeted by selection.

We also conducted the McDonald and Kreitman (1991) test on coding regions to compare variation within a panel (polymorphism) to the divergence from the outgroup at synonymous and non-synonymous sites using DnaSP v5. Four different panels were tested: inbred lines (American + European), landraces, cultivated maize (inbreds + landraces) and teosintes. We also tested for deviation from neutrality in cultivated maize by comparing all pairs of genes within LSUs and within SSUs, leading to three McDonald–Kreitman tests for SSU gene copies and three tests for LSU gene copies. Before executing this latter test, we controlled for non-saturation of the data among paralogous lineages using PAML v4 software (Yang 2007).

Finally, in order to assess deviation from neutrality in functional domains involved in regulation and interaction,

as opposed to the other gene regions, we first aligned the 3 full-length coding regions retrieved from GenBank for paralogs encoding SSUs (1.35 kb) and LSUs (1.30 kb), respectively. We then compared in the two alignments the number of synonymous and non-synonymous substitutions in functional domains versus the rest of the sequence using a $\chi^2$ test with 1 degree of freedom.

### Linkage disequilibrium

Linkage disequilibrium among polymorphic sites excluding non-informative sites (singletons) was investigated using TASSEL v2.1 software (Bradbury et al. 2007), considering inbred lines, landraces, cultivated maize or teosintes. Significance of the square correlation coefficient ($r^2$, Weir 1996) was determined using two-sided Fisher's exact test and 10,000-replicate permutation test (Weir 1996).

### Use of a genome-wide data set as control

In order to account for the demographic effect caused by the domestication bottleneck, we compared our observed values of $\theta_{TOT}$, $\pi_{TOT}$, *Fst* and Tajima's *D* to values previously reported in a genome-wide data set. To do so, we reanalyzed a data set produced by Wright et al. (2005) where 774 ESTs were sequenced over 100 to 900 bp in a sample of 14 American lines and 16 teosintes. From this data set, we isolated a sample of 580 ESTs based on the following criteria: (1) each alignment contained at least 8 sequences of inbred lines and 8 sequences of teosintes and (2) loci with evidence of selection according to Wright et al. (2005) were discarded. Based on the analysis of these 580 ESTs, we obtained distributions for $\theta_{TOT}$, $\pi_{TOT}$ and Tajima's *D* in American lines and teosintes, respectively. Among those 580 loci, we used 502 loci that fulfilled the conditions mentioned above both in American inbreds and teosintes to calculate the distribution of the differentiation between American lines and teosintes as measured by *Fst* (Hudson et al. 1992) using SITES (Hey and Wakeley 1997). Note for medians and 95% confidence intervals (CI) were calculated from those distributions and the CI was used to assess significance of the departure from neutrality. We compared these distributions to the data obtained for the six paralogous genes on the American line and teosinte panels.

We further used Wright et al. (2005)'s data set to quantify the reduction of diversity due to the domestication bottleneck on $\theta_{TOT}$. To do so, we estimated the value of the multilocus maximum likelihood of $\theta_{TOT}$ (ML$\theta$) using the recursion equations of Hudson et al. (1992) implemented by Theta Curve, a perl script kindly provided by Ross-Ibarra (http://www.plantsciences.ucdavis.edu/faculty/rossibarra/code/files/0503c041bb15048e0ec5d31f656b923f-10.html) for American lines and for teosintes, respectively. We slightly modified the perl script to calculate a likelihood value of $\theta_{TOT}$ per base pair at each of the 502 loci and then sum the log-likelihood across loci. The bottleneck effect on $\theta_{TOT}$ was measured as the ratio of the multilocus ML$\theta$ estimated from American lines over the multilocus ML$\theta$ estimated from teosintes. Again, multilocus ML$\theta$ estimated from American lines and teosintes and the reduction of diversity from American and teosintes were compared to observed values for the paralogous AGPase genes.

## Results

### Validation of $LSU_{leaf}$ paralog and expression profiles

We identified through database search a putative large subunit paralog of the ADP-glucose pyrophosphorylase (AGPase) expressed in leaves ($LSU_{leaf}$). Using semi-quantitative RT-PCR, we compared its expression pattern to the expression pattern of the 5 paralogous genes previously identified $SSU_{end}$, $LSU_{end}$, $SSU_{emb}$, $LSU_{emb}$ and $LSU_{leaf}$. The expression pattern of the putative $LSU_{leaf}$ gene was different from the other paralogs and its expression in leaf tissues was substantial. Both observations were consistent with the putative $LSU_{leaf}$ being the gene encoding the large subunit expressed in the leaf. In addition, we observed that the two partners, $LSU_{leaf}$ and $SSU_{leaf}$ exhibited similar expression profiles (Fig. 2) suggesting that both genes participate in the elaboration of a single AGPase. Note that $LSU_{leaf}$ was also expressed in the embryo and overall exhibited a broader expression pattern than the two other large subunit genes, $LSU_{emb}$ and $LSU_{end}$ that were exclusively expressed in 21 DAP embryo and 21 DAP endosperm, respectively (Fig. 2). Finally, the expression of $SSU_{end}$ and $SSU_{emb}$, although predominant in endosperm and embryo, respectively, also extended to other organs (Online Resource 4).

### Sequence polymorphism data

All 6 loci were sampled in ≥26 inbred lines, ≥7 landraces and in ≥15 teosintes but some individuals proved difficult to amplify for some loci, even with several different PCR primers and primer combinations. Because most analyses were carried out within a locus, differences in the sample did not affect our outcome. However, for some tests, comparisons between paralogs expressed in the same organ were useful, and we therefore also defined a common sample for each pair of paralogs encompassing 24, 21 and 25 inbred lines; 12, 11 and 7 landraces; and 12, 10 and 7

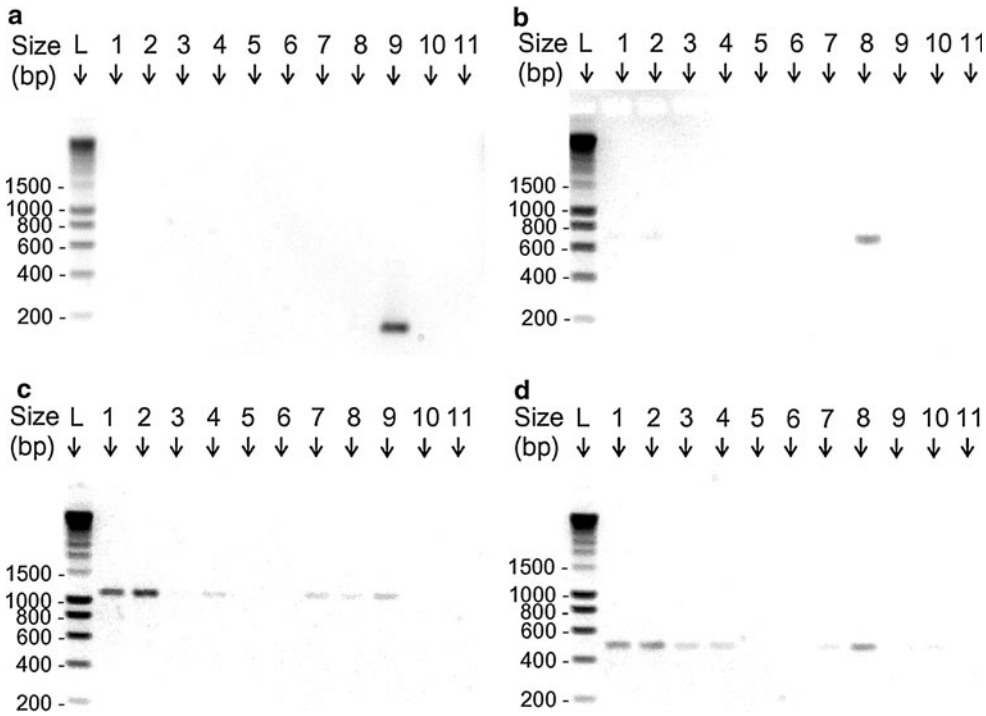


**Fig. 2** Semi-quantitative RT-PCR profiles of $LSU_{end}$ (**a**), $LSU_{emb}$ (**b**), $LSU_{leaf}$ (**c**) and $SSU_{leaf}$ (**d**) paralogous genes in 11 organs. *1* heterotrophic leaf; *2* autotrophic leaf; *3* ear leaf; *4* spathe; *5* silk; *6* ovule; *7* kernel 3 days after pollination (DAP); *8* embryo 21 DAP; *9* endosperm 21 DAP; *10* pollen; *11* root; *L* molecular size ladder

teosintes for the paralogs expressed in the leaf, embryo and endosperm respectively (Table 2). All neutrality tests were performed with both the entire sample of inbreds, landraces, and teosintes for each locus, and the pairwise common sample as defined above (referred to as the common sample throughout the text). We reported results only for the entire sample and reported those from the common sample only when they differed from the former.

One of the goals of our study was to compare levels of genetic diversity among the panels tracing back maize history, i.e., teosintes, landraces, American, and European lines in each paralog. Depleted levels of genetic diversity, for instance, can be indicative of past positive selection. However, because genome-wide demographic effects associated with repeated bottlenecks, accompanying domestication and breeding, generate loss of diversity, it is important to analyze patterns of genetic diversity in a genomic context. Typically, non-selected AGPase genes should have levels of nucleotide diversity consistent with the genome-wide level of diversity, whereas AGPase genes selected before, during or after domestication should exhibit a locally more severe loss of diversity in the corresponding panel. We undertook three approaches to study whether levels of nucleotide diversity may have been shaped by selection during the processes of domestication or breeding.

First, we built empirical distributions of $\theta_{TOT}$ in teosintes and American lines from a genome-wide survey of 580 ESTs (from Wright et al. 2005) and placed our observed values of $\theta_{TOT}$ on these distributions. Values observed for all six paralogous genes in American lines were included in the 95% CI of the empirical $\theta_{TOT}$ distribution. Similarly, $\theta_{TOT}$ values observed in teosintes were included in the 95% CI of the empirical distribution. Thus, this approach provides little indication that the six AGPase genes deviate from background diversity levels.

Second, we estimated from the same genome-wide survey the genomic effect of the combined domestication/breeding bottleneck (from teosintes to American lines) on nucleotide diversity as the ratio of the multilocus likelihood estimate of $\theta_{TOT}$ in American lines (average $\theta_{TOT} = 0.00871$) over the multilocus likelihood estimate of $\theta_{TOT}$ in teosintes (average $\theta_{TOT} = 0.01119$). This ratio (0.78) indicated that 78% of the teosinte variation was captured by American lines. In other words, the genome-wide reduction of diversity in coding regions from the Wright et al. data (2005) was ~22%. On average, observed values for AGPase genes exhibited a higher loss from teosintes to American lines for the small subunit paralogous genes (63.9%) than for the large subunit paralogs (36.1%, Fig. 3). The highest decrease was observed for the $SSU_{end}$ (84.8%).

Third, to evaluate the significance of the diversity loss at successive stages (from teosintes to landraces, from landraces to American inbred lines and from American to European inbred lines), we performed resampling tests that examined each gene individually (see "Materials and methods"). This permutation procedure revealed a significant loss of diversity from teosintes to landraces except for

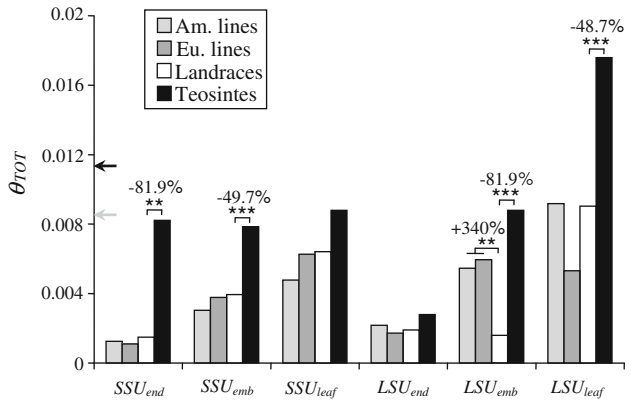


**Fig. 3** Nucleotide diversity ($\theta_{TOT}$) among American lines, European lines, landraces and teosintes at the six AGPase paralogous genes. Differences in nucleotide diversity between teosintes and landraces and between landraces and inbred lines were tested by a permutation procedure (*$P < 5\%$, **$P < 1\%$, ***$P < 1‰$). Only significant results are shown. *Arrows* pointing to the *Y*-axis indicate multilocus likelihood estimates of $\theta_{TOT}$ among American lines (*light gray*) and teosintes (*black*) from 580 neutral loci

$LSU_{end}$ and $SSU_{leaf}$ (Fig. 3). However, the reduction in diversity was not significant from landraces to modern inbred lines or from American to European inbred lines for any of the six genes. Therefore, the only discernible signal of selection by this method is associated with domestication. Note that an unexpected significant *gain* of diversity from landraces to inbred lines was observed in $LSU_{emb}$ (+340%, Fig. 3).

## Differentiation

If selection and/or drift drives the evolution of a locus during domestication or breeding, we expect a higher differentiation between teosintes and landraces, or between landraces and inbred lines, respectively. In other words, two panels may exhibit a similar level of diversity but may be highly differentiated because of contrasted allele frequencies driven by selection and/or drift. In order to investigate differentiation between panels, we followed the same rationale as described above first by performing resampling procedures, and second by comparing our data to a genome-wide survey.

Using a resampling procedure, we found that four paralogs presented a significant differentiation between teosintes and landraces: $SSU_{end}$ ($K_s^* = 2.014$, $P < 10^{-3}$), $LSU_{emb}$ ($Fst = 0.155$, $P = 2.3\ 10^{-2}$; $K_s^* = 1.572$, $P < 10^{-3}$), $SSU_{leaf}$ ($Fst = 0.130$, $P = 1.1\ 10^{-2}$; $K_s^* = 2.336$, $P = 3.0\ 10^{-3}$) and $LSU_{leaf}$ ($Fst = 0.172$, $P = 1.8\ 10^{-2}$; $K_s^* = 2.579$, $P = 9.0\ 10^{-3}$). Only $SSU_{leaf}$ presented a significant differentiation between landraces and inbred lines ($Fst = 0.189$, $P = 3.0\ 10^{-3}$; $K_s^* = 2.063$, $P = 8.0\ 10^{-3}$) and no differentiation was found among inbred lines from the two continents. The genome-wide differentiation between teosintes and American inbred lines was evaluated from 502 ESTs using a one-sided test for *Fst*. The comparison of our observed values at six paralogs to the genome-wide distribution of *Fst* values evidenced an unusual high level of differentiation, above the 95% CI, at $SSU_{end}$ (Fig. 4a).

Overall, our results were in good agreement with a mild effect of the breeding bottleneck. Four paralogs ($SSU_{end}$, $LSU_{emb}$, $SSU_{leaf}$, $LSU_{leaf}$) displayed strong differentiation between teosintes and landraces. $SSU_{end}$ was the most extreme outlier. Because the examination of European and American lines revealed neither a significant loss of diversity nor a significant differentiation, we pooled them in a single inbred line panel for subsequent analyses.

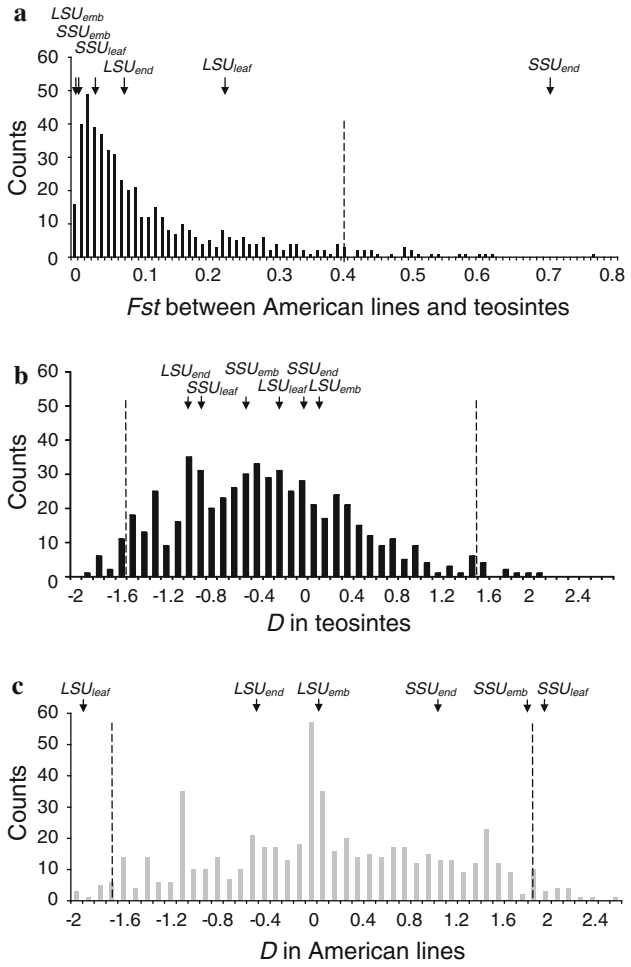


**Fig. 4** Frequency distributions estimated from neutral loci for **a** nucleotide differentiation as measured by *Fst* between American lines and teosintes (502 loci were used), **b** Tajima's *D* in teosintes (580 loci were used) and **c** Tajima's *D* in American lines (580 loci were used). *Dotted vertical lines* indicate 95% distribution limits [one-sided test in (**a**) and two-sided test in (**b**) and (**c**)]. Values for each AGPase paralog are indicated by *vertical arrows*

### Deviation from neutrality: frequency spectrum, haplotype structure and linkage disequilibrium

In addition to the description of nucleotide diversity and differentiation, several tests (reviewed in Nielsen 2005) were applied to evaluate the deviation from neutrality at the six AGPase paralogous genes among teosintes, landraces, and inbred lines (Table 1). Some tests relied on the frequency spectrum such as Tajima's *D*, Fu and Li's *F* and *D*, Fay and Wu's *H* and others relied on the haplotype structuring such as the Fu's *Fs*. Typically, directional selection (positive or negative) increases the proportion of low frequency variants in the region surrounding the causal mutation, producing negative Tajima's *D* and Fu and Li's *F* and *D* values. An opposite pattern is expected under balancing selection. Positive selection also tends to increase the proportion of high frequency derived variants that are hitchhiked with the causal mutation, resulting in negative Fay and Wu's *H* values. Finally, selection increases transiently the correlation among alleles from different loci, thereby creating extended linkage disequilibrium, and a correlative deficit in haplotypes (positive *Fu's Fs* values).

Because most of these tests are not robust to demographic factors such as bottlenecks (Nielsen 2005), we compared the Tajima's *D* values observed for the AGPase paralogs to the distributions of Tajima's *D* for 580 neutral loci in teosintes and American lines. A significant positive value for Tajima's D denotes an excess of balanced variants whereas a significant negative value of Tajima's *D* denotes an excess of rare variants. To detect both kinds of deviations from neutrality, we performed a two-sided test.

Note that Tajima's *D* exhibited a broader distribution in American lines (Fig. 4c), with a positive median compared to teosintes (−0.52; Fig. 4b). It likely resulted from the loss of rare variants through the domestication and/or plant breeding bottlenecks that led both to higher Tajima's *D* values and a wider variance, since drift intensity is expected to vary among genomic regions (Tenaillon et al. 2004; Wright et al. 2005). We present below the results of the neutrality tests performed at each locus for the teosinte, the landrace and the inbred panels, respectively.

Evidence of selection among teosintes with negative Tajima's *D*, Fu & Li's *D* and *F* values was detected for $SSU_{leaf}$ (Table 1). However, the Tajima's *D* value fell within the distribution for neutral loci (Fig. 4b) and there was no evidence of significant LD between sites in the teosinte panel (Fig. 5a) suggesting that the bias in allele frequency may be due to demographic rather than selective effect. The only significant test for $LSU_{end}$ was the negative Fu's *Fs* that indicates an excess of haplotypes. This statistic was slightly higher (−3.581) and not significant when using the common sample of 7 teosintes out of the 12. Although not significant, the tendency shown by the negative values of Tajima's *D* and Fu and Li's *D* and *F* suggested that $LSU_{end}$ may have evolved under purifying selection long before domestication, consistent with what was previously reported by Manicacci et al. (2007).

Among landraces, two genes, $SSU_{end}$ and $LSU_{emb}$, exhibited significant negative Tajima's *D* accompanied by a strong reduction in diversity from teosinte to landraces and a significant negative Fay and Wu's *H* (Table 1). Additionally, $LSU_{leaf}$ presented a strong deficit in haplotypes, as revealed by the significant Fu's *Fs* but no significant LD was detected in the landrace panel (data not shown). Those results remained unchanged using the common sample.

In inbred lines, three AGPase paralogs exhibited consistent signs of departure from neutrality, $LSU_{leaf}$, $SSU_{leaf}$ and $SSU_{emb}$ (Table 1). These results were all confirmed using the common sample of inbred lines. $LSU_{leaf}$ stood as an outlier in Tajima's *D* distribution in American lines (Fig. 4c) and exhibited a consistent pattern of directional selection with negative values of Tajima's *D*, Fu and Li's *F*, Fay and Wu's *H* and a deficit in haplotypes (Table 1). This pattern was also clear when pooling together landraces and inbred lines (Table 1) and is consistent with a substantial increase of LD from teosintes to inbred lines (Fig. 5a, b). The second gene that exhibited a signal of selection among inbred lines was $SSU_{leaf}$, with a highly significant deficit in haplotypes (Table 1) and several instances of highly significant LD between polymorphic sites in the inbred line panel (Fig. 5b) contrasting with a very low level of LD in teosintes (Fig. 5a). Although none of the Tajima's, and Fu and Li's tests was significant (Table 1), when considering demography, $SSU_{leaf}$ Tajima's *D* fell outside the 95% CI (Fig. 4c) suggesting an excess of balanced polymorphisms. Finally, $SSU_{emb}$ exhibited an unusual pattern with a strong deficit in haplotypes (significant positive Fu's *Fs*, Table 1) consistent with its borderline significant Tajima's *D* value (Fig. 4c) and significant LD between polymorphic sites (data not shown). The neighbor-joining tree built for all $SSU_{emb}$ sequences (Fig. 6) showed a remarkable structure, with two main clades, i.e., haplotypic groups, each containing teosintes, landraces, and inbred lines. Those inbred lines were recorded for a number of traits such as photoperiod sensitivity, flowering time, kernel and embryo weight in a 2-year experimental trial (Camus-Kulandaivelu et al. 2006; Manicacci et al. 2009). Interestingly, one clade encompassed inbred lines with a higher ratio of embryo weight relative to the kernel and earlier flowering phenotypes than the other clade.

Overall, both $SSU_{end}$ and $LSU_{emb}$ seem to have been targeted by selection during domestication consistently

**Table 1** Summary statistics of nucleotide diversity and neutrality tests for the six AGPase paralogous genes

| Genes | Panel[a] | $N$[b] | $L$[c] | $S$[d] | Sing[e] | NHap[f] | $C$[g] | $\pi_{sil}$ | $\theta_{sil}$ | Tajima | Fu and Li | | Fay and Wu | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | $D$ | $D$ | $F$ | $H$ | $Fs$ |
| $SSU_{end}$ | Inbred lines | 55 | 2,468 | 13 | 2 | 9 | 0.866 | 0.00181 | 0.00142 | 0.9110 | 0.2828 | 0.2389 | −1.9382 | 1.422 |
| | Maize | 62 | 2,466 | 14 | 3 | 10 | 0.597 | 0.00176 | 0.00150 | 0.6086 | −0.4291 | −0.4324 | −1.8570 | 0.844 |
| | Landraces | 7 | 2,468 | 9 | 9 | 3 | 0.389 | 0.00124 | 0.00177 | **−1.5945**** | 0.5385 | 0.0970 | **−3.5925*** | 1.835 |
| | Teosintes | 11 | 2,451 | 59 | 26 | 11 | 1.091 | 0.01048 | 0.01045 | −0.0049 | −1.0855 | −0.9382 | −0.0565 | −2.260 |
| $LSU_{end}$ | Inbred lines | 30 | 2,363 | 19 | 8 | 8 | 1.065 | 0.00239 | 0.00262 | −0.4555 | 0.2109 | 0.0279 | −1.1157 | 1.242 |
| | Maize | 37 | 2,363 | 22 | 11 | 11 | 0.993 | 0.00233 | 0.00293 | −0.7777 | −0.7176 | −0.7849 | −1.0111 | −0.473 |
| | Landraces | 7 | 2,364 | 11 | 9 | 6 | 3.763 | 0.00210 | 0.00249 | −0.9342 | **−1.7757*** | −1.6146 | 0.8688 | −1.751 |
| | Teosintes | 12 | 2,364 | 20 | 12 | 10 | 0.982 | 0.00269 | 0.00344 | −1.0562 | −1.0639 | −1.0966 | 0.3282 | **−3.619*** |
| $SSU_{emb}$ | Inbred lines | 22 | 2,100 | 26 | 9 | 6 | 0.341 | 0.00663 | 0.00502 | 1.2234 | −0.4014 | 0.1354 | −0.1938 | **6.654*** |
| | Maize | 34 | 2,100 | 28 | 2 | 13 | 5.188 | 0.00669 | 0.00482 | **1.3646*** | **1.6170*** | **1.6902*** | 0.0675 | 1.479 |
| | Landraces | 12 | 2,100 | 25 | 6 | 10 | 18.047 | 0.00681 | 0.00582 | 0.7553 | 0.4573 | 0.5559 | 0.6640 | −1.532 |
| | Teosintes | 11 | 2,090 | 48 | 23 | 11 | 13.385 | 0.01031 | 0.01137 | −0.4780 | −0.4490 | −0.4527 | 0.6422 | −3.023 |
| $LSU_{emb}$ | Inbred lines | 23 | 1,502 | 30 | 1 | 8 | 0.332 | 0.00986 | 0.00838 | 0.6761 | **1.6653*** | **1.4349*** | −0.2192 | **4.054*** |
| | Maize | 34 | 1,501 | 30 | 1 | 9 | 0.497 | 0.00748 | 0.00757 | −0.0405 | **1.6634*** | 1.1637 | −0.9705 | 3.344 |
| | Landraces | 11 | 1,502 | 7 | 7 | 3 | 1.150 | 0.00131 | 0.00246 | **−1.8969***** | 0.2409 | −0.3440 | **−4.8825**** | 1.085 |
| | Teosintes | 13 | 1,502 | 40 | 16 | 12 | 11.872 | 0.01374 | 0.01361 | 0.1264 | −0.1769 | −0.1265 | −0.0929 | −2.581 |
| $SSU_{leaf}$ | Inbred lines | 26 | 1,957 | 38 | 9 | 7 | 0.144 | 0.00994 | 0.00766 | 1.1266 | 0.1446 | 0.5355 | 0.3918 | **8.551**** |
| | Maize | 38 | 1,957 | 40 | 4 | 13 | 1.094 | 0.01072 | 0.00732 | **1.6462*** | 1.1854 | **1.4965*** | 0.9647 | **4.338*** |
| | Landraces | 12 | 1,960 | 38 | 22 | 8 | 4.234 | 0.00855 | 0.00965 | −0.5172 | **−1.5585*** | −1.3368 | 0.1616 | 1.021 |
| | Teosintes | 15 | 1,957 | 56 | 28 | 13 | 33.827 | 0.00997 | 0.01324 | **−1.0667*** | **−1.2138*** | **−1.2335*** | 1.0221 | −2.357 |
| $LSU_{leaf}$ | Inbred lines | 24 | 2,210 | 80 | 46 | 9 | 6.065 | 0.00652 | 0.01246 | **−1.8908**** | −1.2712 | **−1.6114*** | −2.9721 | **3.982*** |
| | Maize | 36 | 2,203 | 104 | 43 | 13 | 9.942 | 0.00736 | 0.01455 | **−1.8575***** | −1.0920 | **−1.5220*** | −2.4040 | **3.267*** |
| | Landraces | 12 | 2,231 | 63 | 19 | 5 | 0.493 | 0.00950 | 0.01195 | −0.9657 | 0.2638 | −0.1300 | −1.6952 | **7.168**** |
| | Teosintes | 12 | 2,221 | 122 | 56 | 11 | 6.417 | 0.02164 | 0.02288 | −0.2969 | −0.6652 | −0.5876 | 0.3746 | 0.249 |

Significant values are indicated in bold (*$P < 5\%$; **$P < 1\%$; ***$P < 1‰$)

[a] The "inbred lines" panel includes both American and European lines and the "Maize" panel includes inbred lines and landraces

[b] Number of sequences excluding the outgroup

[c] Length of sequence alignment excluding gaps (in base pairs)

[d] Number of polymorphic sites

[e] Number of singletons

[f] Number of haplotypes

[g] Average value of likelihood estimates of the population recombination rate ($C$) obtained from five simulations

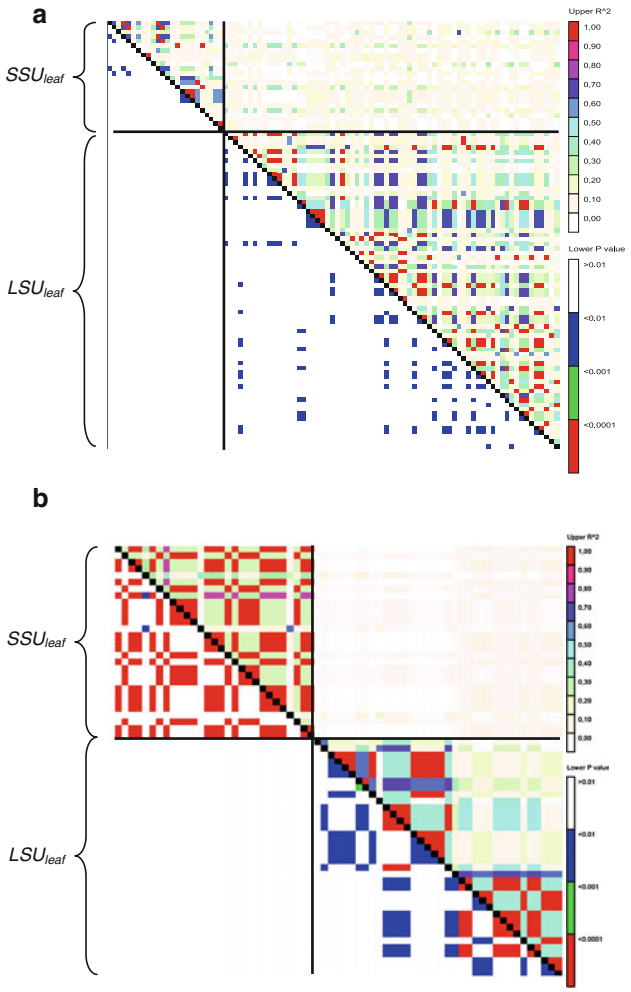


**Fig. 5** Linkage disequilibrium as estimated by the square of the coefficient of correlation ($r^2$) among informative sites and level of significance for $SSU_{leaf}$ and $LSU_{leaf}$ genes considering 12 teosintes (**a**) and 24 maize inbred lines (**b**)

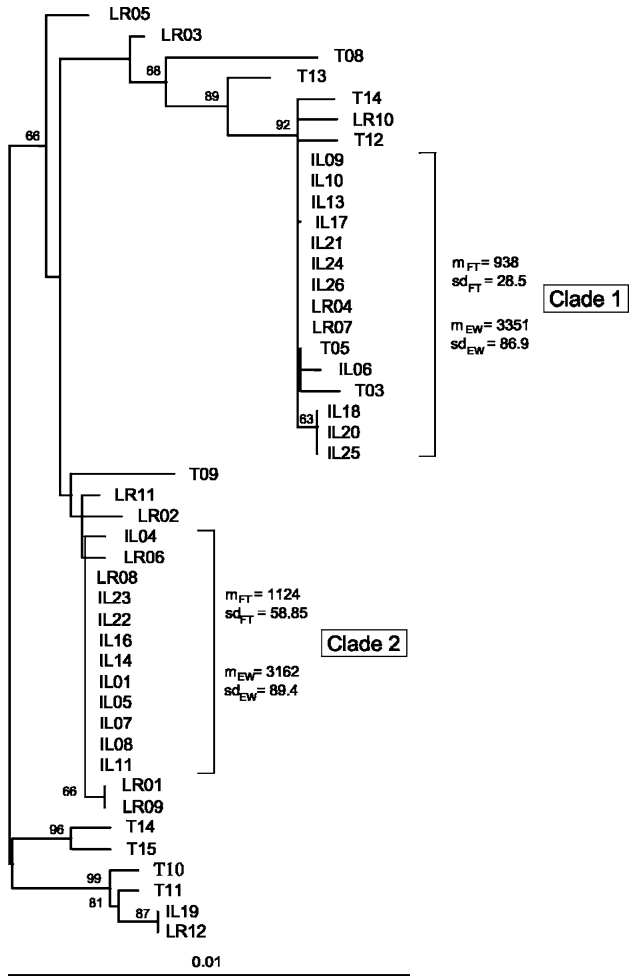


**Fig. 6** Neighbor-joining tree based on the sequence alignment of $SSU_{emb}$ (IL1 to IL13: American lines; IL14 to IL26: European lines; LR01 to LR12: landraces; T03 to T15: teosintes). Bootstrap values higher than 60% are indicated. Mean (m) and standard deviation (sd) for flowering time (FT) measured as days to pollen under long-day conditions expressed in thermal time (Camus-Kulandaivelu et al. 2006) and the ratio of embryo weight relative to the kernel (EW, NIRS estimation; Baye et al. 2006) are indicated for both inbred lines groups (*Clade 1* early flowering and high embryo weight; *Clade 2* late flowering and low embryo weight)

with a significant loss of diversity between teosintes and landraces and a significant differentiation between teosintes and landraces as evaluated by resampling tests. Among inbred lines, contrasted patterns of selection were observed. While $LSU_{leaf}$ seems to have been subjected to positive selection, $SSU_{leaf}$ and $SSU_{emb}$ exhibited patterns consistent with diversifying and balancing selection respectively. The segregation of two major haplotypes in $SSU_{emb}$ correlates with flowering time and the ratio of embryo weight relative to the kernel. Note that using the restricted common sample of inbred lines in $SSU_{end}$, i.e., 29 out of the 55 inbred lines, resulted in statistical significance of two neutrality tests, the Fay and Wu's $H$ (−4.3184) and the Fu's $Fs$ (5.626). These results indicated signs of positive selection in $SSU_{end}$ but because they were sample-dependent, this interpretation had to be taken with caution.

## Comparing level of polymorphism and divergence

We investigated whether the six paralogs have undergone selective pressures leading to an increase ($k > 1$) or decrease ($k < 1$) in nucleotide diversity applying a multilocus likelihood HKA test. By comparing the level of polymorphism over divergence to an outgroup species (*T. dactyloides*), the HKA test allows to control for heterogeneity of the mutation rate among loci. Because we used a maximum likelihood approach, values of the $k$ index may vary according to the simulations. We first considered models including 11 neutral loci plus one of each AGPase locus. Three loci appeared to be under selection: $SSU_{end}$, $SSU_{leaf}$ and $LSU_{leaf}$ ($P = 9.29\ 10^{-6}$, $P = 1.33\ 10^{-2}$ and

$P = 1.86\ 10^{-4}$, respectively). We found that selection acting on $SSU_{end}$ and $LSU_{leaf}$ led to a decrease in nucleotide diversity with $k = 0.07$ and $k = 0.05$, respectively, whereas selection acting on $SSU_{leaf}$ led to an increase in diversity with $k = 4.5$. We then used nested models that included 14 loci in which $SSU_{end}$, $SSU_{leaf}$, and $LSU_{leaf}$ were considered under selection either individually or taken together (Table 2). Considering a single paralog under selection, $SSU_{end}$ exhibited the most significant evidence of selection (model B, Table 2). Building a model that includes two loci under selection, we obtained a better significance by adding $SSU_{end}$ (model E and F) rather that $SSU_{leaf}$ (E vs. B) or $LSU_{leaf}$ (F vs. B) suggesting that $SSU_{end}$ has undergone stronger selection than $SSU_{leaf}$ and $LSU_{leaf}$. Additionally, we obtained a better significance by adding $LSU_{leaf}$ (model F) rather than $SSU_{leaf}$ (model E), suggesting that $LSU_{leaf}$ had undergone stronger selection than $SSU_{leaf}$ (Table 2). Finally, a model considering all three paralogs under selection (model H) performed significantly better than models considering two paralogs under selection. Under this model, both $SSU_{end}$ and $LSU_{leaf}$ loci would have undergone directional selection ($k = 0.07$ and $k = 0.11$, respectively, under model H), whereas $SSU_{leaf}$ would have been subjected to diversifying selection ($k = 3.51$, model H). HKA test was also performed on the common sample to test for selection at $SSU_{end}$, $LSU_{leaf}$ and $SSU_{leaf}$ and similar conclusions were reached. Altogether, HKA results confirmed those obtained with neutrality tests based on the frequency spectrum.

We applied the McDonald-Kreitman's test (1991) in order to compare the ratio of polymorphism over divergence at synonymous and non-synonymous sites considering each locus individually. Selection increases or reduces the number of non-synonymous mutations relative to the number of synonymous mutations, and this effect is stronger in divergence than in polymorphism. We therefore expect deviations of the ratio when the locus is under selection. For a given paralog, we used as outgroup either *Tripsacum dactyloides* or the sequence of one paralog encoding the same type of subunit (either SSU or LSU) to estimate divergence at both synonymous and non-synonymous sites. To test for saturation, we estimated the synonymous substitution rate along branches of a topology including all six paralogs. Values above 2.79 in the three branches leading to the *LSU* paralogs were indicative of saturation at synonymous site. In contrast, no evidence of saturation was detected among *SSU* paralogs ($dS < 0.35$ in all branches). Therefore, we only considered small subunit paralogous genes to compute the McDonald-Kreitman's test. This test was applied on the different panels (inbred lines, landraces, cultivated maize and teosintes) but did not reveal any significant deviation from neutrality (data not shown).

## Selection in functional domains and coevolution

In addition to selection at individual genes, we also considered selection on functionally important motifs within a gene and selection on interacting partners, i.e., paralogs

**Table 2** Multilocus likelihood ratio tests (LRT) for neutrality at $SSU_{end}$, $SSU_{leaf}$ and $LSU_{leaf}$ duplicated genes

| Model | Loci under selection | Comparison | LR stat[a] | *P* value[b] | *k*[c] | | |
|---|---|---|---|---|---|---|---|
| | | | | | $SSU_{end}$ | $SSU_{leaf}$ | $LSU_{leaf}$ |
| A | | | | | 1 | 1 | 1 |
| B | $SSU_{end}$ | B vs. A | 18.87 | $1.4 \times 10^{-5}$*** | 0.07 | 1 | 1 |
| C | $SSU_{leaf}$ | C vs. A | 10.04 | $1.5 \times 10^{-3}$* | 1 | 5.37 | 1 |
| D | $LSU_{leaf}$ | D vs. A | 13.51 | $2.4 \times 10^{-4}$** | 1 | 1 | 0.06 |
| E | $SSU_{end}$ and $SSU_{leaf}$ | E vs. B | 7.93 | $4.8 \times 10^{-3}$ | 0.10 | 3.76 | 1 |
| | | E vs. C | 16.76 | $4.2 \times 10^{-5}$*** | 0.10 | 3.76 | 1 |
| F | $SSU_{end}$ and $LSU_{leaf}$ | F vs. B | 12.44 | $4.2 \times 10^{-4}$** | 0.04 | 1 | 0.05 |
| | | F vs. D | 17.79 | $2.5 \times 10^{-5}$*** | 0.04 | 1 | 0.05 |
| G | $SSU_{leaf}$ and $LSU_{leaf}$ | G vs. C | 10.22 | $1.4 \times 10^{-3}$* | 1 | 3.51 | 0.07 |
| | | G vs. D | 6.75 | $9.4 \times 10^{-3}$ | 1 | 3.51 | 0.07 |
| H | $SSU_{end}$, $SSU_{leaf}$ and $LSU_{leaf}$ | H vs. E | 13.31 | $2.6 \times 10^{-4}$** | 0.07 | 3.51 | 0.11 |
| | | H vs. F | 8.80 | $3.0 \times 10^{-3}$* | 0.07 | 3.51 | 0.11 |
| | | H vs. G | 19.85 | $8.4 \times 10^{-6}$*** | 0.07 | 3.51 | 0.11 |

[a] Likelihood ratio statistics (1 *df*)

[b] *P* value considering that the likelihood ratio statistics follows a $\chi^2$ distribution. Significance after correction for multiple testing is indicated as: * $P < 5\%$, ** $P < 1\%$, *** $P < 1‰$

[c] *k* measures the degree of diversity increase ($k > 1$) or reduction ($k < 1$)

encoding the LSU and the SSU expressed in the same organ, through compensatory mutations. We considered as functionally important motifs those previously described in the literature as involved in enzyme regulation and in the interaction with the partner subunit. We tested whether non-synonymous substitution accumulated differentially in functionally known domains as opposed to interdomain regions. Because of saturation (see above), this test was performed on SSU paralogs only. We detected an excess of non-synonymous mutations in functional domains (16 synonymous vs. 33 non-synonymous mutations) as compared to interdomain regions (123 synonymous vs. 44 non-synonymous mutations; $\chi_1^2 = 27.76$, $P$ value $= 1.4 \times 10^{-7}$). This suggests a crucial role of the functional regions in the evolution of the small subunit.

Finally, we tested for coevolution between subunits expressed in a same organ by assessing the significance of linkage disequilibrium between polymorphic sites belonging to genes encoding for interacting subunits: $LSU_{leaf}$ and $SSU_{leaf}$, $LSU_{end}$ and $SSU_{end}$, and $LSU_{emb}$ and $SSU_{emb}$. We detected only weak level of LD that was never significant and therefore no evidence for coevolution between interacting partners.

## Discussion

In maize, three pairs of paralogous genes encode the large (LSU) and small (SSU) subunits of the most rate-limiting enzyme during starch synthesis, AGPase. These three pairs are preferentially expressed in either endosperm, embryo or leaves. The LSU gene expressed in leaves had not been characterized, and here we have isolated $LSU_{leaf}$ and report comparative patterns of all six AGPase paralogs. We have found that $LSU_{leaf}$ has a very similar expression pattern to $SSU_{leaf}$, suggesting that the sequence we isolated corresponds to the large subunit expressed in leaves. The $LSU_{leaf}$ expression pattern differs strikingly from those of the other LSUs. It exhibits a broader expression than $LSU_{end}$ and $LSU_{emb}$, both of which were expressed exclusively in one tissue under our assay conditions.

In agreement with the results of Rosti and Denyer (2007) and Cossegal et al. (2008), we have found that SSU genes had a broader expression pattern than LSU genes. In some phases of angiosperm evolution, a single small subunit may have interacted with several large subunits because of non concomitant duplications of both subunit lineages (Georgelis et al. 2008). This observation may explain the broader expression of SSU genes and also suggests that AGPase enzyme activity is limited to regions of LSU gene expression. Moreover, this considerations may explain why some sequenced angiosperm genomes have more LSU than SSU paralogous genes—e.g., 4 versus 2 in *O. sativa* (Akihiro et al. 2005), 4 versus 1 in *A. thaliana* (The *Arabidopsis* genome initiative 2000; Villand et al. 1993), and 6 versus 1 in *Populus trichocarpa* (Tuskan et al. 2006), respectively. Indeed, although a same number of SSU and LSU genes has been described in maize, the complete genome sequencing may reveal additional AGPase genes. In fact, using one amino acid reference sequence for each of the six paralogs, we searched the maize cDNA database in maizesequence.org. We retrieved hits above 45% identity over the entire protein length, which corresponds to the minimum identity observed across all pairs of previously described LSUs and SSUs. This revealed four additional positions on chromosomes 1, 3, 7 and 8, which suggests that other AGPase genes persist in this species. LSU-SSU interactions in maize may be therefore more complex than previously thought and comparable at what is reported in other species.

Among angiosperms, duplications in the SSU lineage occurred after the monocot-dicot divergence (Georgelis et al. 2008). Among grasses, $SSU_{end}$ is expressed either in the endosperm cytosol or in the leaf plastid as a result of the alternative splicing of the first exon (exon 1a and 1b; Rosti and Denyer 2007). In maize, an additional duplication (anterior to domestication) gave rise to $SSU_{end}$ and $SSU_{leaf}$. While the alternative splicing of the first exon has been shown to persist in $SSU_{end}$, exon 1a is spliced out in $SSU_{leaf}$ (Rosti and Denyer 2007). As a result, $SSU_{end}$ and $SSU_{leaf}$ duplicates mostly diverged at exons 1a and 1b, exon 1a exhibits 56% similarity and exon 1b exhibits 89% similarity, whereas the rest of the sequence displays on average 94% similarity between the two paralogs (Rosti and Denyer 2007). When using two different pairs of primers, specific of each first exon (1a and 1b) for $SSU_{end}$, we obtained similar patterns of expression than those reported by Cossegal et al. (2008), i.e., a higher expression of $SSU_{end}$ (1a-specific) in endosperm and a higher expression of $SSU_{leaf}$ (1b-specific) in leaves (data not shown).

### Effect of domestication and breeding on duplicates

Selection during domestication and breeding has increased kernel starch content, suggesting that starch content has been under selection (Flint-Garcia et al. 2009). In addition, inbred lines exhibit different starch content and kernel weight according to their origin (Manicacci et al. 2009) as illustrated in Fig. 7. Indeed, European and Northern Flint inbred lines appear to contain lower level of starch than Corn Belt Dent and Stiff Stalk lines. Northern and European Flint lines are adapted to temperate climates and early flowering. As a consequence of a short filling period, they tend to have smaller kernels than Corn Belt Dent and Stiff Stalk inbred lines, that instead have been selected for

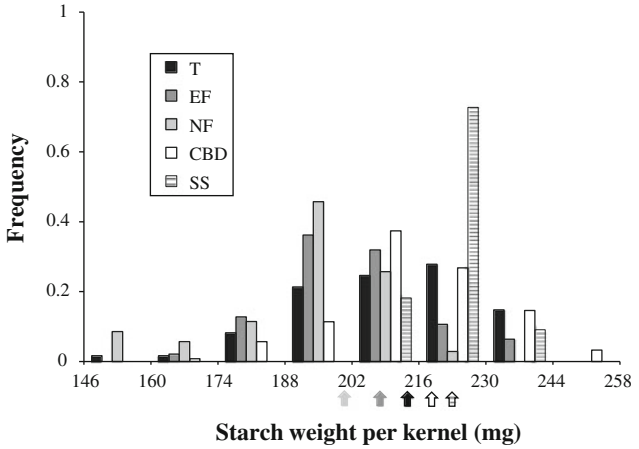


**Fig. 7** Frequency distribution of starch weight per kernel in five major inbred line groups (*T* tropical; *EF* European Flint; *NF* Northern Flint; *CBD* Corn Belt Dent; *SS* Stiff Stalk). Data are taken from a panel of 375 inbred lines (Camus-Kulandaivelu et al. 2006). Only 278 inbred lines with group membership higher than 75% were considered (61, 47, 35, 123 and 11 inbred lines for T, EF, NF, CBD and SS, respectively). Arrows indicate the mean kernel starch weight for each group, calculated by weighting inbred line phenotypic values by group membership

longer life cycles, larger kernels with higher starch content (Duvick and Cassman 1999). Given the variability of starch content among groups of inbred lines, we also expect diversifying selection to have shaped the genetic diversity of some genes in the starch pathway during maize expansion and plant breeding.

Among the six genes encoding AGPase, five exhibited evidence of selection during domestication and/or breeding. Only $LSU_{end}$ did not exhibit any sign of selection in both the landrace and the inbred line panels. This gene appears to have evolved under strong selective constraint long before domestication, as evidenced by an extremely low level of diversity in both wild and cultivated panels confirming previous results reported by Manicacci et al. (2007). In contrast, $SSU_{end}$ exhibited evidence of selection during the domestication process but did not seem to be involved in varietal differentiation, strengthening the results previously obtained by Whitt et al. (2002).

We observed a complex pattern of diversity for $LSU_{emb}$. We reported directional selection among landraces. In addition, while selection and demographic effects are expected to decrease nucleotide diversity within landrace and inbred line panels, nucleotide diversity increased from landraces to inbred lines. In order to determine whether the inbred line diversity emerged recently from the landrace pool, we constructed a neighbor-joining tree based on sequences of all panels (data not shown). Among three main haplotypes, two were shared by inbred lines and teosintes and the third one encompassed landraces and inbred lines. This distance tree therefore did not support the emergence of inbred line diversity from landraces through de novo accumulation of mutations. Furthermore, the multiplication of elite material is highly controlled and gene flow from teosintes to inbred lines has likely been very limited and too small to explain the observed increase in nucleotide diversity among inbred lines. Besides, if some introgression occurred, it should have led to the reduction of genetic differentiation between both panels, which was not observed. An alternative hypothesis that could explain our results is a bias in the landrace sample. However, we have not encountered a similar effect for other AGPase genes sequenced on the same landrace panel. We therefore lack a clear explanation for this pattern; a panel including more landraces should help elucidating this issue.

The pattern of diversity of $SSU_{emb}$ suggests divergent selection during breeding. Interestingly, the structure in haplotype diversity, as reported in a distance tree (Fig. 6), was associated to the ratio of embryo weight relative to kernel among inbred lines (two-sided $t$ test, $t = 2.23$, $df = 23$, $P = 0.038$) but also correlated with flowering time. In maize, Northern/Southern geographic structure is correlated with divergence in flowering time and, in a lower extent, with kernel size and endosperm starch content. Typically, maize adapted to temperate climate, such as Northern Flints, have a shorter period of kernel filling and less favorable conditions for seedling development. This adaptation has resulted in divergent selection between tropical and temperate material for genes involved in kernel resource accumulation and/or seedling development, as observed in the vicinity of the flowering time gene *Dwarf8* (Camus-Kulandaivelu et al. 2006, 2008). A formal association genetics mapping, including a larger panel and correction for genetic structure, would be necessary to validate the role of $SSU_{emb}$ polymorphism on phenotypic variation of the ratio of embryo weight relative to kernel.

$LSU_{leaf}$ gene has responded to selection mostly during breeding. The intense selection on kernel starch content that occurred since domestication may have had an indirect effect on $LSU_{leaf}$, because of the role this gene plays in the diurnal accumulation of transient starch that is hydrolyzed at night, providing sink tissues with sugar supplies (Zeeman et al. 2007). This is particularly important during kernel filling. Furthermore, one of the most important phenotypic changes that accompanied domestication is the reduction in the number of axillary branches (Doebley et al. 1995) and the subsequent reduction in leaf number. This reduction in global leaf area may have led to a selective pressure for either increased gene expression or enzyme efficiency in the remaining leaves. Whether the modification affected AGPase regulation or structure, it may have had consequences on the molecular diversity pattern of the coding sequence because of linkage

disequilibrium (Drummond et al. 2005). Comparative expression of AGPase in leaves between inbred lines, landraces, and teosintes would help distinguishing between regulatory and structural changes.

Finally, $SSU_{leaf}$ also revealed signs of diversifying selection among cultivated maize, a pattern that could not be explained by demographic effects. Among inbred lines, three haplotypes were clearly differentiated and correlated to flowering time, one haplotype including mainly early flowering inbred lines whereas the two others encompassed mainly late-flowering inbred lines (two-sided test, $t = 3.80$, $df = 23$, $P < 10^{-3}$). The haplotype divergence either may result from direct selection on the $SSU_{leaf}$ gene or from linkage disequilibrium between $SSU_{leaf}$ and a gene involved in flowering time subjected to divergent selection. This second hypothesis cannot be ruled out since, in maize, numerous QTL involved in flowering time variation have been found all over the genome (Buckler et al. 2009). Here also, an association genetics approach coupled with a precise study of linkage disequilibrium around $SSU_{leaf}$ would provide more insights into this question.

### Comparing evolution at two evolutionary scales

The history of the AGPase multigenic family has been investigated in Angiosperms in several studies (Georgelis et al. 2008, 2007, 2009). These studies have shown that the *SSU* paralogs have evolved under selective constraint whereas the *LSU* paralogs have undergone accelerated evolutionary changes including positive selection at several sites following duplications. Some of these sites are located in regulatory motifs, such as the 55-amino acid region involved in the between-subunit interaction, the glucose-1-phosphate binding region or the C-terminal regulatory region.

In our intraspecific studies, both SSU and LSU paralogs had similar levels of diversity. However, we found an excess of non-synonymous substitutions in functional regions of SSU genes, suggesting the action of positive selection on SSU functional domains, but positive selection was detected on LSU functional domains only for interspecific contrasts. The differences between intra- and interspecific contrasts may not be too surprising, because selective pressures acting at inter- and intra-specific levels are expected to differ. For example, De Mita et al. (2006) reported that the patterns of diversity of the *NORK* gene does not deviate from neutrality within *Medicago truncatula* ssp. *trunlacuta*, whereas several sites under positive selection are involved in the divergence among species of *Medicago* and other genera. As opposed to the wild *Medicago truncatula*, maize has undergone recent selection through domestication and breeding, which has probably resulted in more pronounced intraspecific patterns of selection on putative target genes, such as AGPase paralogs.

Finally, the interaction between subunits raises the question of coevolution between genes. Coevolution between amino acids is known to generally affect functional regions or interaction motifs (Hakes et al. 2007; Yeang and Haussler 2007) and has been observed for AGPase during angiosperm radiation both within SSU and between interacting subunits (Corbi et al., unpublished). We investigated patterns of substitution in the 55 amino acid motif identified as involved in between-subunit interaction (Cross et al. 2005). This motif did not display divergence between maize and teosinte in any of the paralogs, whereas it strongly differed among paralogs. This suggests that the between-subunit interaction motif has evolved in relation to the specialization of expression, long before domestication. Patterns of linkage disequilibrium did not reveal any sign of coevolution between genes expressed in the same organ. So if coevolution is visible at an interspecific scale, it is not at a more recent intraspecific scale.

## Conclusion

AGPase paralogous genes have undergone contrasting selection pressures during maize domestication and breeding that are neither specific to the large subunit, the small subunit nor the organ of expression. While the small subunit is known as the most constrained during angiosperm AGPase evolution, both small and large subunits appear to have been targets of selection during and after maize domestication. By generating redundancy, duplication is usually considered as a process that increases robustness to mutation, i.e., that makes more mutations to be selectively neutral (without phenotypic effect; Conant and Wagner 2004; Hickman and Rusche 2007; Wagner 2005). There are few cases reported in the literature showing that gene duplication has also been advantageous for the evolution of new phenotypes. A recent example in tomato reports how gene duplication has allowed altered gene expression and the evolution of an elongated fruit, a shape that was likely selected by early domesticators (Xiao et al. 2008). Our results suggest that genetic redundancy may also provide evolutionary potential in the response to selection through higher flexibility leading to contrasted patterns of selection among AGPase paralogs.

**Acknowledgments** We would like to thank Catherine Damerval, Alain Charcosset and Johann Joets for helpful discussions throughout this study. We are very grateful to Brandon Gaut for providing comments on the manuscript and carefully checking the language. Guillaume Achaz and Frantz Depaulis have brought insightful suggestions on the data analyses. We are thankful to Jean-Louis Prioul

for his physiological view of maize AGPase evolution. Pierre Garrabos helped with data collection, Céline Ridel and Matthieu Falque performed the mapping. We are grateful to Létizia Camus-Kulandaivelu and Brigitte Gouesnard for providing accessions. We also would like to thank the two anonymous reviewers who brought insightful comments to the manuscript. This work was partly supported by a grant from the Bureau des Ressources Génétiques (project no. 8 2005–2007) to DM and a grant from the Agence Nationale de la Recherche (ANR-05-JCJC-0067-01) to MIT. JC was financed by a PhD fellowship from Centre National de la Recherche Scientifique.

## References

Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, Wakeley J (2004) A robust measure of HIV-1 population turnover within chronically infected individuals. Mol Biol Evol 21:1902–1912

Akihiro T, Mizuno K, Fujimura T (2005) Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. Plant Cell Physiol 46:937–946

Bae JM, Giroux M, Hannah LC (1990) Cloning and characterization of the brittle-2 gene of maize. Maydica 35:312–322

Ballicora MA, Fu Y, Nesbitt NM, Preiss J (1998) ADP-Glucose pyrophosphorylase from potato tubers. Site-directed mutagenesis studies of the regulatory sites. Plant Physiol 118:265–274

Ballicora MA, Erben ED, Yazaki T, Bertolo AL, Demonte AM, Schmidt JR, Aleanzi M, Bejar CM, Figueroa CM, Fusari CM, Iglesias AA, Preiss J (2007) Identification of regions critically affecting kinetics and allosteric regulation of the *Escherichia coli* ADP-glucose pyrophosphorylase by modeling and pentapeptide-scanning mutagenesis. J Bacteriol 189:5325–5333

Baye T, Pearson T, Settles A (2006) Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. J Cereal Sci 43:236–243

Bejar CM, Jin X, Ballicora MA, Preiss J (2006) Molecular architecture of the glucose 1-phosphate site in ADP-glucose pyrophosphorylases. J Biol Chem 281:40473–40484

Bhave MR, Lawrence S, Barton C, Hannah LC (1990) Identification and molecular characterization of shrunken-2 cDNA clones of maize. Plant Cell 2:581–588

Boehlein SK, Shaw JR, Hannah LC, Stewart JD (2010) Probing allosteric binding sites of the maize endosperm ADP-glucose pyrophosphorylase. Plant Physiol 152:85–95

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267

Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Camus-Kulandaivelu L, Chevin L-M, Tollon C, Charcosset A, Manicacci D, Tenaillon MI (2008) Patterns of variation of the Tb1–D8 region shed light into early maize evolutionary history. Genetics 180:1107–1121

Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, Vienne D (1996) A composite map of expressed sequences in maize. Genome 39:418–432

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene tb1. Proc Natl Acad Sci USA 101:700–707

Conant GC, Wagner A (2004) Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. Proc Biol Sci 271:89–96

Cossegal M, Chambrier P, Mbelo S, Balzergue S, Martin-Magniette ML, Moing A, Deborde C, Guyon V, Perez P, Rogowsky P (2008) Transcriptional and metabolic adjustments in ADP-glucose pyrophosphorylase-deficient bt2 maize kernels. Plant Physiol 146:1553–1570

Cross JM, Clancy M, Shaw JR, Boehlein SK, Greene TW, Schmidt RR, Okita TW, Hannah LC (2005) A polymorphic motif in the small subunit of ADP-glucose pyrophosphorylase modulates interactions between the small and large subunits. Plant J 41:501–511

De Mita S, Santoni S, Hochu I, Ronfort J, Bataillon T (2006) Molecular evolution and positive selection of the symbiotic gene NORK in *Medicago truncatula*. J Mol Evol 62:234–244

Doebley J, Stec A, Gustus C (1995) Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333–346

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc Natl Acad Sci USA 102:14338–14343

Duvick DN, Cassman KG (1999) Post-green revolution trends in yield potential of temperate maize in the North-Central United States. Crop Sci 39:1622–1630

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene Loci. Genetics 170:1957–1966

Fan L, Bao J, Wang Y, Yao J, Gui Y, Hu W, Zhu J, Zeng M, Li Y, Xu Y (2009) Post-domestication selection in the maize starch pathway. PLoS ONE 4:e7612

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Flint-Garcia S, Bodnar A, Scott M (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. TAG Theor Appl Genet 119:1129–1142

Frueauf JB, Ballicora MA, Preiss J (2003) ADP-glucose pyrophosphorylase from potato tuber: site-directed mutagenesis of homologous aspartic acid residues in the small and large subunits. Plant J 33:503–511

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Fu YX, Li WH (1993) Maximum likelihood estimation of population parameters. Genetics 134:1261–1270

Fu Y, Ballicora MA, Leykam JF, Preiss J (1998) Mechanism of reductive activation of potato tuber ADP-glucose pyrophosphorylase. J Biol Chem 273:25045–25052

Gallavotti A, Zhao Q, Kyozuka J, Meeley RB, Ritter MK, Doebley JF, Pe ME, Schmidt RJ (2004) The role of barren stalk1 in the architecture of maize. Nature 432:630–635

Geigenberger P, Kolbe A, Tiessen A (2005) Redox regulation of carbon storage and partitioning in response to light and sugars. J Exp Bot 56:1469–1479

Georgelis N, Braun EL, Shaw JR, Hannah LC (2007) The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. Plant Cell 19:1458–1472

Georgelis N, Braun EL, Hannah LC (2008) Duplications and functional divergence of ADP-glucose pyrophosphorylase genes in plants. BMC Evol Biol 8:232

Georgelis N, Shaw J, Hannah L (2009) Phylogenetic analysis of ADP-glucose pyrophosphorylase subunits reveals a role of subunit interfaces in the allosteric properties of the enzyme. Plant Physiol 151:67

Giroux MJ, Hannah LC (1994) ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. Mol Gen Genet 243:400–408

Giroux M, Smith-White B, Gilmore V, Hannah LC, Preiss J (1995) The large subunit of the embryo isoform of ADP glucose pyrophosphorylase from maize. Plant Physiol 108:1333–1334

Giroux MJ, Shaw J, Barry G, Cobb BG, Greene T, Okita T, Hannah LC (1996) A single mutation that increases maize seed weight. Proc Natl Acad Sci USA 93:5824–5829

Glémin S, Bataillon T (2009) A comparative view of the evolution of grasses under domestication. New Phytol 183:273–290

Hakes L, Lovell SC, Oliver SG, Robertson DL (2007) Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proc Natl Acad Sci USA 104:7999–8004

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hannah LC, Shaw JR, Giroux MJ, Reyss A, Prioul JL, Bae JM, Lee JY (2001) Maize genes encoding the small subunit of ADP-glucose pyrophosphorylase. Plant Physiol 127:173–183

Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF (1996) Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics 143:1395–1407

Haudry A, Cenci A, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Glémin S, David J (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24(7):1506–1517

Hendriks JHM, Kolbe A, Gibon Y, Stitt M, Geigenberger P (2003) ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of *Arabidopsis* and other plant species. Plant Physiol 133:838–849

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. Genetics 145:833–846

Hickman MA, Rusche LN (2007) Substitution as a mechanism for genetic robustness: the duplicated deacetylases Hst1p and Sir2p in *Saccharomyces cerevisiae*. PLoS Genet 3:e126

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151

Hwang S-K, Salamone PR, Okita TW (2005) Allosteric regulation of the higher plant ADP-glucose pyrophosphorylase is a product of synergy between the two subunits. FEBS Lett 579:983–990

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671

Jaenicke-Despres V, Buckler ES, Smith BD, Gilbert MT, Cooper A, Doebley J, Paabo S (2003) Early allelic selection in maize as revealed by ancient DNA. Science 302:1206–1208

Kavakli IH, Park JS, Slattery CJ, Salamone PR, Frohlick J, Okita TW (2001) Analysis of allosteric effector binding sites of potato ADP-glucose pyrophosphorylase through reverse genetics. J Biol Chem 276:40834–40840

Kilian B, Özkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F (2006) Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol Genet Genomics 276:230–241

Kim D, Hwang S-K, Okita TW (2007) Subunit interactions specify the allosteric regulatory properties of the potato tuber ADP-glucose pyrophosphorylase. Biochem Biophys Res Commun 362:301–306

Lee YM, Preiss J (1986) Covalent modification of substrate-binding sites of *Escherichia coli* ADP-glucose synthetase. Isolation and structural characterization of 8-azido-ADP-glucose-incorporated peptides. J Biol Chem 261:1058–1064

Linebarger CRL, Boehlein SK, Sewell AK, Shaw J, Hannah LC (2005) Heat stability of maize endosperm ADP-glucose pyrophosphorylase is enhanced by insertion of a cysteine in the N terminus of the small subunit. Plant Physiol 139:1625–1634

Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321–330

Manicacci D, Falque M, Le Guillou S, PiÈgu B, Henry A-M, Le Guilloux M, Damerval C, De Vienne D (2007) Maize Sh2 gene is constrained by natural selection but escaped domestication. J Evol Biol 20:503–516

Manicacci D, Camus-Kulandaivelu L, Fourmann M, Arar C, Barrault S, Rousselet A, Feminias N, Consoli L, Frances L, Mechin V, Murigneux A, Prioul JL, Charcosset A, Damerval C (2009) Epistatic interactions between Opaque2 transcriptional activator and its target gene CyPPDK1 control kernel trait variation in maize. Plant Physiol 150:506–520

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

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654

McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160:1231–1241

Myers AM, Morell MK, James MG, Ball SG (2000) Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol 122:989–997

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, p 256

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus *Lycopersicon*. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162: 365–379

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197–218

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. Proc Natl Acad Sci USA 101:9885–9890

Pan D (2000) Starch synthesis in maize. In: Gupta AK, Kaur N (eds) Carbohydrate reserves in plants: synthesis and regulation. Elsevier, Amsterdam, pp 125–146

Patron NJ, Keeling PJ (2005) Common evolutionary origin of starch biosynthetic enzymes in green and red algae. J Phycol 41:1131

Piperno DR, Flannery KV (2001) The earliest archaeological maize (Zea mays L.) from highland Mexico: new accelerator mass

spectrometry dates and their implications. Proc Natl Acad Sci USA 98:2101–2103

Prioul JL, Jeannette E, Reyss A, Grégory N, Giroux M, Hannah LC, Causse M (1994) Expression of ADP-glucose pyrophosphorylase in maize (*Zea mays* L.) grain and source leaf during grain filling. Plant Physiol 104:179–187

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction to Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106(5):895–903

Rosti S, Denyer K (2007) Two paralogous genes encoding small subunits of ADP-glucose pyrophosphorylase in maize, Bt2 and L2, replace the single alternatively spliced gene found in other cereal species. J Mol Evol 65:316–327

Rozas J (2009) DNA sequence polymorphism analysis using DnaSP. Methods Mol Biol 537:337–350

Shaw JR, Hannah LC (1992) Genomic Nucleotide Sequence of a Wild-Type Shrunken-2 Allele of *Zea mays*. Plant Physiol 98:1214–1216

Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5:233–241

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

Teas H, Teas A (1953) Heritable characters in maize. Description and linkage of brittle endosperm-2. J Hered 44:156–158

Tenaillon MI, Tiffin PL (2008) The quest for adaptive evolution: a theoretical challenge in a maze of data. Curr Opin Plant Biol 11:110–115

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21:1214–1225

Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? Genome Res 16:702–712

The Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (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

Villand P, Olsen OA, Kleczkowski LA (1993) Molecular characterization of multiple cDNA clones for ADP-glucose pyrophosphorylase from *Arabidopsis thaliana*. Plant Mol Biol 23:1279–1284

Vollbrecht E, Springer PS, Goh L, ESt Buckler, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. Nature 436:1119–1126

Wagner A (2005) Robustness, evolvability, and neutrality. FEBS Lett 579:1772–1778

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

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

Weber A, Clark RM, Vaughn L, Sanchez-Gonzalez Jde J, Yu J, Yandell BS, Bradbury P, Doebley J (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). Genetics 177:2349–2359

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, MA

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci USA 99:12959–12962

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Wright SI, Charlesworth B (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168:1071–1076

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17:2859–2872

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yeang CH, Haussler D (2007) Detecting coevolution in and among protein domains. PLoS Comput Biol 3:e211

Zeeman SC, Smith SM, Smith AM (2007) The diurnal metabolism of leaf starch. Biochem J 401:13–28

Zhao Q, Thuillet AC, Uhlmann NK, Weber A, Rafalski JA, Allen SM, Tingey S, Doebley J (2008) The role of regulatory genes during maize domestication: evidence from nucleotide polymorphism and gene expression. Genetics 178:2133–2143

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24: 875–888