Theor Appl Genet
DOI 10.1007/s00122-009-1188-x

ORIGINAL PAPER

Nucleotide diversity and molecular evolution of the *PSY1* gene in *Zea mays* compared to some other grass species

Zhiyuan Fu · Jianbing Yan · Yanping Zheng · Marilyn L. Warburton · Jonathan H. Crouch · Jian-Sheng Li

Received: 26 April 2009 / Accepted: 9 October 2009

Abstract Phytoene synthase (PSY), which is encoded by the phytoene synthase 1 (*PSY1*) gene, is the first rate-limiting enzyme in the plant carotenoid biosynthetic pathway. In order to examine the genetic diversity and evolution pattern of *PSY1* within the Andropogoneae, sequences of 76 accessions from 5 species (maize, teosinte, tripsacum, coix, and sorghum) of the Andropogoneae were tested, along with 4 accessions of rice (*Oryza sativa* L.) included as outliers. Both the number and the order of exons and introns were relatively conserved across the species tested. Three domains were identified in the coding sequence, including signal peptide (SP), PSY, and highly conserved squalene synthase (SQS) domain. Although no positive selection signal was detected at an overall coding level among all species tested, the SP domain and the region upstream of the SQS–PSY domain appear to have undergone rapid evolution, as evidenced by a high d_N/d_S ratio (>1.0). At the nucleotide level, positive selection and balancing selection were detected only among the yellow maize germplasm and the white maize germplasm, respectively. The phylogenetic tree based on full-length sequences of *PSY1*-like regions supported the monophyletic theory of the Andropogoneae and the closest relationship between Zea and Tripsacum among the Andropogoneae. Coix, which was theorized to have a closer relationship with maize due to similarities in morphology and chromosome number, has been shown in this study to have diverged relatively early from the other Andropogoneae, including maize.

Communicated by J. Yu.

Z. Fu and J. Yan contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1188-x) contains supplementary material, which is available to authorized users.

Z. Fu · J. Yan · Y. Zheng · J.-S. Li (✉)
National Maize Improvement Center of China,
Key Laboratory of Crop Genomics and Genetic Improvement,
China Agricultural University, Yuanmingyuan West Road,
Haidian, 100193 Beijing, China
e-mail: lijiansheng@cau.edu.cn; lijs@163bj.com

J. Yan · J. H. Crouch
Genetic Resources Program,
International Maize and Wheat Improvement Center (CIMMYT),
Apdo. Postal 6-641, 06600 Mexico D.F., Mexico

M. L. Warburton
USDA-ARS Corn Host Plant Resistance Research Unit,
Box 9555, Mississippi State, MS 39762, USA

Introduction

Cereal crops domestication has had a unique impact on the history of human civilization. Among these cereal crops, maize represents a rather unique case for having undergone much more drastic morphological modifications during domestication (Sang 2009). Archaeological and molecular evidence demonstrate that maize (*Zea mays* L. ssp. *mays*) was domesticated from teosinte (Doebley et al. 1984, 1987; Doebley 1990; Piperno and Flannery 2001) and that the *Z. mays* L. ssp. *parviglumis* is likely to be the direct progenitor of maize (Matsuoka et al. 2002). There are about 59,000 genes in the maize genome (Messing et al. 2004), but only 2% of these (~1,200) were involved in the process of domestication of maize from teosinte, or more recent crop improvement (Wright et al. 2005). Identifying these artificially selected genes is the key to understand the mechanism and history of maize domestication and improvement. Recent studies demonstrate that important novel morphological

Published online: 03 November 2009

changes associated with domestication involved only a small number of regulatory genes (Buckler et al. 2006; Doebley and Stec 1993). To date, two key genes associated with domestication have been cloned in maize: *tb1*, controlling plant architecture (Doebley et al. 1997; Doebley 2004); and *tga1*, controlling glume architecture (Wang et al. 2005). A number of genes associated with maize improvement have also been well studied, including the two transcriptional regulators *c1* and *r1*, controlling the colorful hues of maize kernels (Hanson et al. 1996); the *sh2* and *su1* genes, contributing to the sweetness of maize kernels (Whitt et al. 2002); the *vgt1* gene, controlling the flowering time (Ducrocq et al. 2008; Salvi et al. 2007); and the phytoene synthase 1 (*PSY1*) gene, influencing endosperm color and carotenoid accumulation in maize kernels (Palaisa et al. 2003).

New powerful statistical tools and the reduced cost of DNA sequencing make it more feasible to search for or reconfirm the effect of artificially selected genes on past evolution and continued future improvement of modern maize. Studying these genes can help us to understand how plant development was redirected to meet the demands of a hungry world (Doebley et al. 2006). Nucleotide diversity in maize reflects the rich history of human selection and migration, combined with high level of recombination, and out-breeding characteristic of this species (Buckler and Thornsberry 2002). Phylogenetic analysis based on nucleotide diversity of a genome is a promising strategy for furthering the understanding of a species' evolution at the molecular level.

The *PSY1* gene is a representative gene related to maize improvement for controlling an important human nutritional trait, the quantity of provitamin A in maize kernels (Doebley et al. 2006). Among the three copies of *PSY* in maize (*PSY1*, *PSY2* and *PSY3*), only *PSY1*, which encodes the first rate-limiting enzyme in the plant carotenoid biosynthetic pathway, phytoene synthase (PSY) (Bird et al. 1991; Bramley et al. 1992; Fray and Grierson 1993; Giuliano et al. 1993; Kumagai et al. 1995; von Lintig et al. 1997), has an effect on endosperm color (Gallagher et al. 2004; Li et al. 2008b; Palaisa et al. 2003). The yellow endosperm phenotype controlled by *PSY1* has been a strong target of breeding selection since the 1930s, when the nutritional advantage of increased carotenoids in yellow maize (yellow endosperm maize) was recognized (Mangelsdorf and Fraps 1931). This selection on *PSY1* during maize improvement has been verified in previous studies (Palaisa et al. 2003, 2004), in which yellow maize lines showed a significant negative Tajima's *D* and 19-fold lower nucleotide polymorphism compared to white maize lines, and no linkage disequilibrium (LD) decay compared to LD decay within 1 kb in white maize lines (Palaisa et al. 2003). The causative mutation was suggested to be a regulatory change in the promoter of the *PSY1* gene, and selection resulted in the reduction of nucleotide diversity within the yellow maize *PSY1* allele extending 600 kb downstream and 200 kb upstream from *PSY1* (Palaisa et al. 2004). They concluded that the sequence encoding white endosperm was the ancestral state of the *PSY1* gene, and it was the *PSY1* rather than *PSY2* gene, which underwent strong positive selection and influenced carotenoid accumulation in yellow endosperm. In addition, sorghum can also accumulate endosperm carotenoids (FAO 1995; Salas Fernandez et al. 2008), but teosinte, coix, and tripsacum have white endosperm only. Sequence information from the *PSY1* gene in these species can answer questions regarding the structure of the *PSY1* gene in these species; in which species directional selection has occurred; and the phylogenetic relationship between the key species in the Andropogoneae.

In this study, we cloned *PSY1*-like genes in maize, teosinte, tripsacum, coix, sorghum, and rice and analyzed the nucleotide diversity and phylogenetic relationships of this gene among these species with three objectives: (1) to investigate how variable the DNA sequences are in each species and how selection has affected this variation in the *PSY1* gene across Andropogoneae; (2) to detect the variation, selection and LD of the *PSY1* gene in yellow maize using various different germplasm resources; (3) to identify the interspecific relationship among Andropogoneae as revealed by the *PSY1* phylogenetic tree.

Materials and methods

Plant materials and DNA extraction

All maize inbred lines were planted on the Agronomy Farm at China Agricultural University (Beijing, BJ, E 116°46′, N 39°92′) in spring of 2005; genotypes of coix, rice, sorghum and teosinte were grown in the greenhouse. Fresh leaves of 2-week-old plants were harvested and frozen at −70°C. DNA was extracted using the CTAB method (Murray and Thompson 1980). Eight coix accessions were included in the study, five from China and three from exotic germplasm. A set of 56 maize genotypes were tested, including 38 yellow inbred lines (19 normal lines, 15 high-oil lines, and 4 high provitamin A lines) and 18 white lines. None of the maize lines used in this study except for B73 and Mo17 has been previously characterized at the *PSY1* locus. Alleles of *PSY1* were also amplified from teosinte (7 accessions), rice (4 varieties), and sorghum (4 yellow accessions). One *Tripsacum* species was included as the outgroup for the Tajima's *D* and HKA analysis within the genus, *Zea*. Identification and information about each of the accessions used in this study are listed in Supplementary Table S1.

Primer design and PCR

To obtain the *PSY1*-like sequences among the grasses, the highly similar regions among the aligned reference sequences of maize (GeneBank ZMU32636), sorghum (PlantGDB SbGSStuc11-12-04.5154.1) and rice (AP005750) were used to design conserved primers (Supplementary Table S2) using the Primer Premier 5.0 program (Clarke and Gorley 2001) with the following conditions: product size between 400 and 1,300 bases, primer size between 18 and 22 bases, annealing temperature between 55 and 60°C, and ideal GC content between 50 and 60%. The sequenced regions covered open reading frames (ORFs) of the *PSY1* gene. PCR was performed using a PTC-200 thermocycler under the following conditions: denaturation at 94°C for 4 min; amplification for 5 cycles of 95°C for 1 min, 60°C for 1 min (reducing temperature by 1°C each cycle) and 72°C for 1.5 min; followed by 30 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1.5 min, or 35 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1.5 min; and final extension at 72°C for 10 min. The 30 µl PCR mixture consisted of 60 ng DNA, 3 µl 10× PCR buffer (containing Mg^{2+}), 0.6 µl 10 µM of each primer, 2.4 µl 2.5 mM of dNTPs, and 3 units of EasyTaq (Transgen, China).

DNA sequencing and assembly

PCR products were analyzed by agarose gel electrophoresis, and were purified using the Recycle Kit (Biotech Ltd) and sequenced directly or cloned using the Promega pGEM-T Easy Vector according to the manufacturer's protocol. The T7 sequence (5′-GTAATACGACTCACTATAGGGC-3′) and SP6 sequence (5′-ATTTAGGTGACACTATAGAATA-3′) were used to identify and sequence positive clones. No <3 positive clones were sequenced for each sample using an ABI 3730 sequencer. For heterozygous teosinte, coix and tripsacum samples, at least five clones were sequenced to control for PCR errors versus true sequence heterogeneity, and the DNA sequence with the same content in at least three clones was used. Vector NTI Contig Express software (Informax, North Bethesda, MD, USA) was used to delete the vector sequence and assemble the gene sequences. All sequences of the tested species were submitted to GenBank and are identified by the accession number from FJ971174 to FJ971252. Gene structure was predicted separately for each species using the software packages FGENESH (http://linuxl.softberry.com) and NNPP (http://www.fruitfly.org). Alignments were performed among all species with the ClustalX version 8.1 (Thompson et al. 1997) and Muscle (Edgar 2004) software packages and improved manually in the BioEdit software package (http://www.mbio.ncsu.edu/BioEdit/bioedit) using the reference sequences of rice, sorghum and maize as controls.

DNA polymorphism analysis

The software package DNAsp Version 4.20.2 (http://www.ub.es/dnasp) was used to calculate summary data about the sequences, including π (Nei 1987), Watterson's estimator of θ (Nei 1987; Watterson 1975), Tajima's D (Tajima 1989) and the ratio of nonsynonymous substitution to synonymous substitution (d_N/d_S) (Nielsen and Yang 1998). The polymorphic sites and Insertion/Deletion (InDel) changes were extracted and combined into haplotypes using TASSEL Version 2.0 (Bradbury et al. 2007). The r^2 (Hill and Robertson 1968) values were calculated in Tassel and LD, measured as r^2, between pairs of polymorphic sites was plotted against the physical distance among all sites in Microsoft Excel. The measurements were performed by averaging r^2 values over a distance of 200 bp and plotting the values against distance. To detect positive selection footprints at the *PSY1* gene in *Z. mays*, the Hudson–Kreitman–Aguade (HKA) test was performed. The tripsacum sequence was used to calculate divergence, and the *ADH1* gene sequence (forward: 5′-AGATCAATCCTCAGGCTCCC-3′, reverse: 5′-TTCAGGGTCCTTTCGTTCA-3′) and *GLB1* gene sequence (forward: 5′-TACCTCCGCTTTAGTTCTGC-3′, reverse: 5′-CGAAGAAGAACTGCCAAAGG-3′) (Tenaillon et al. 2001; Tiffin and Gaut 2001) were used as neutral control loci. The overall HKA P value was obtained by summing the individual χ^2 values for the two control genes.

Phylogenetic and evolutionary analyses

Phylogenetic analyses were performed using the software packages PAUP version 4.0 beta (Swofford 1998) and Phylip (Felsenstein 1993). Alignment gaps were treated as missing data, and branches with zero length were collapsed. To find all shortest trees and to identify multiple tree islands, the fault maximum-parsimony method was used with tree bisection–reconnection (TBR) branch swapping and random order of taxon addition. Both heuristic search and bootstrap support for nodes were estimated with 100 and 1,000 replicates, respectively. For maximum likelihood (ML) method analyses (Felsenstein 1981), the gamma shape parameter alpha was set to 0.5, and the transition: transversion ratio was set to the default value of 2.0. The neighbor-joining method of Phylip was used with the following conditions: equally weighted characters, and transition: transversion ratio of 2.0. Phylogenetic trees were shown in TreeView (Page 1996) or MEGA 4.0 (Tamura et al. 2007), using rice as the outgroup. The strict consensus tree was evaluated, and bootstrap value of <50% for each branch was set as a cutoff.

Results

Sequence and structure of the *PSY1*-like gene in the different species

The *PSY1*-like genes shared a similar overall structure across the different species tested, with the same number of exons (6) and introns (5) (Table 1). Four of the exons (the second, third, fourth and fifth) were found to be conserved in length and sequence order across all accessions tested. The length of the first exon was longer in rice (436 and 439 bp) and sorghum (433 and 439 bp) than in the other species tested, which ranged between 406 and 415 bp. The sixth exon in coix was very different from the other species in both length and sequence. The mean identity of all sequences is 96.7% at the nucleotide level with a range of 87.2–99.9% (Table 2).

Despite this high nucleotide similarity among *PSY1*-like sequences, remarkable differences at the amino acid level are found in the COOH and NH_2 terminals of the predicted proteins (Supplementary Fig. S1). Sorghum has the greatest level of differentiation in amino acid arrangement at the COOH terminal compared to the other taxa, while coix shows the most differences at the NH_2 terminal. The amino acids at the end of the first exon (residues 100–130) are highly variable among the tested species (Fig. 1, Supplementary Fig. S1), while the sequence between these regions appears to be relatively conserved. Three domains were identified in the amino acid alignments among the tested species, except rice (Fig. 1, Supplementary Fig. S1). The signal peptide (SP) domain (residues 1–19 in the alignment) is relatively well conserved; the only major variation among the tested species is the insertion of two amino acids (alanine and aspartic acid) in the sorghum sequence. The other two domains, squalene synthase (SQS) and PSY, were identified between the 259th and 320th residues in aligned sequences. In the species tested, the SQS signal domain is completely conserved and the PSY signal domain is highly conserved with only one polymorphism identified at the 311th residue in the aligned sequences, with threonine in sorghum and arginine in the other species.

Nucleotide variation of the *PSY1*-like gene across grass species

Detailed comparison of identified variation in the present and a previous study (Palaisa et al. 2003) is listed in supplementary Table S3. There were several major new variants

Table 1 *PSY1*-like gene structure in different species (length of each feature given in bp)

Taxa	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Intron 3	Exon 4	Intron 4	Exon 5	Intron 5	Exon 6	Full length
Rice	436/439[a]	98	51	828	173	597	236	108	193	85	174	2,979/2,982
Coix	406	98	51	612	173	665	236	94	193	77	186	2,791
Sorghum	433/439	98	51	614	173	652	236	103	193	100	174	2,827/2,833
Tripsacum	415	146	51	624	173	582	236	84	193	75	174	2,753
Teosinte	406/409	100	51	612	173	898	236	94	193	77	174	3,014/3,017
Maize	406	98	51	613	173	885	236	94	193	77	174	3,000

[a] Different length of the exon

Table 2 Mean identities and d_N/d_S ratios for *PSY1*-like genes in different species

Species	*N*	Identity (%)	d_N	d_S	d_N/d_S
All	80	96.7 (87.2–99.9)[a]	0.025 (0.000–0.148)	0.104 (0.000–0.546)	0.240 (0.000–1.227)
Maize	56	98.9 (96.5–100)	0.002 (0.000–0.007)	0.008 (0.000–0.019)	0.266 (0.000–5.500)
Maize (yellow)	38	99.7 (98.3–100)	0.001 (0.000–0.002)	0.001 (0.000–0.008)	0.505 (0.000–5.500)
Maize (white)	18	98.3 (96.5–99.9)	0.003 (0.000–0.007)	0.010 (0.001–0.017)	0.258 (0.000–1.692)
Teosinte	7	98.4 (97.2–99.3)	0.002 (0.000–0.004)	0.040 (0.003–0.064)	0.052 (0.000–0.393)
Teosinte (luxuriantes)	4	98.5 (97.7–99.3)	0.001 (0.000–0.001)	0.035 (0.003–0.062)	0.016 (0.000–0.084)
Teosinte (zea)	3	98.7 (98.3–99.1)	0.004 (0.002–0.004)	0.011 (0.009–0.014)	0.325 (0.237–0.404)
Sorghum	4	99.9 (99.8–99.9)	0.001 (0.000–0.003)	0.001 (0.000–0.003)	0.840 (0.000–1.190)
Coix	8	98.6 (97.4–99.8)	0.003 (0.001–0.006)	0.013 (0.003–0.022)	0.226 (0.393–0.767)
Rice	4	96.4 (95.0–97.9)	0.001 (0.000–0.001)	0.005 (0.000–0.007)	0.172 (0.000–0.180)

N Number of materials used

[a] Range of identity

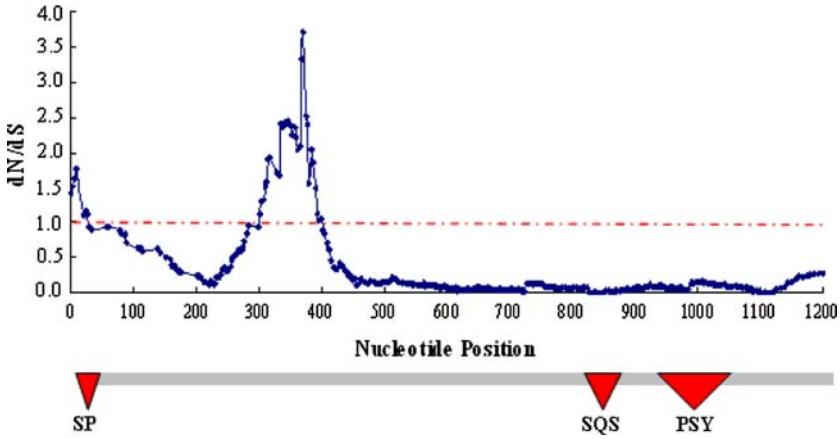

Fig. 1 Sliding-window analysis of the estimated ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) along the *PSY1* coding sequence among the Andropogoneae. The mean d_N/d_S ratio for all pairwise sequence comparisons at each window is depicted as a rhombus on the graph. The d_N/d_S estimate was computed for windows of 60 bp with 3 bp intervals between each pair of windows. The three domains are indicated *below* the graph

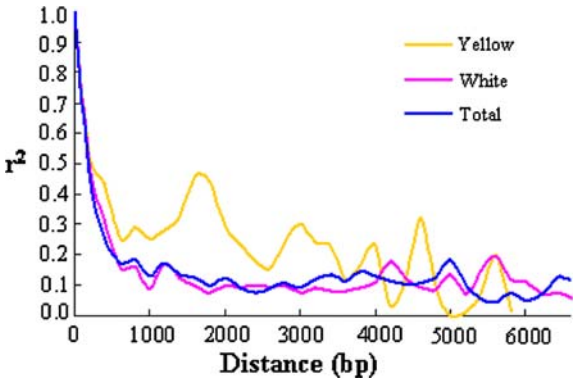

Fig. 2 Linkage disequilibrium decay in the *PSY1* gene sequence in *yellow* and *white* maize lines. Measurement was performed by averaging r^2 values over a distance of 200 bp and plotting the values against distance (bp)

found in this study, including the 644 bp InDel in the 5′UTR, and two silent mutations and one nonsynonymous mutation at positions 2050, 2128 and 2129, respectively. Plots of the pair-wise LD measure r^2 for *PSY1* indicate that LD declines to 0.1 in yellow maize lines at a distance of greater than 4 kb. This is a much slower rate of decay than in white maize lines, which declines to 0.1 in <2 kb (Fig. 2).

Major differences were observed between the UTR sequences of maize compared to the other species tested in this study, which led to only partial UTR sequences being cloned from the other species (Table 3). The three large insertions described above which were identified in yellow maize were not found in the other species investigated in this study. Conversely, we observed unique sequences in the 5′UTR of rice compared to the other species, including several small InDels ranging between 15 and 44 bp before the start codon. The *Zea* teosinte UTR sequences are more closely related to white maize than to the *luxuriantes* section species of teosinte. Teosinte and coix also contain a trinucleotide repeat (CCA) upstream of the start site. Most of these variants are similar to those reported previously in white maize, but one new allele, a $(CCA)_3$ repeat, was observed in teosinte and coix. Further upstream of the CCA repeat, a 484 bp InDel was observed in *Zea luxurians*, which has no sequence similarity to the same alignment position of other teosinte accessions. The tested sorghum accessions have no changes in the UTRs and the intraspecific changes in the 3′UTR of coix and rice are inconspicuous. In coding sequences, the major SNP and InDel differences between the species investigated occur in the first and the last exons. In all species, more SNPs were observed in the introns than the other regions and InDels mainly occurred in untranslated regions (except for sorghum and coix). The set of teosinte accessions investigated displayed more sequence diversity (9.20 polymorphisms/100 bp) than the tested maize accessions (5.66 polymorphisms/100 bp) and this trend is consistent across all regions (Table 3). In all tested species, coix had the highest level of variation with 9.78 polymorphisms/100 bp, and sorghum had the lowest, 0.66 polymorphisms/100 bp. Among the Andropogoneae, π and θ values were very similar between the white maize and *Zea* teosinte (Table 4). Coix had the highest π and θ values for all segregating sites with 24.17×10^3 and 26.40×10^3, respectively, which are 36 and 14 times higher than yellow maize. Coix also displayed a relatively high π value for nonsynonymous sites (9.15×10^3), which is 22-fold more diverse than yellow maize, and threefold more than white maize and *Zea* teosinte (Table 4).

Phylogenetic analyses of the *PSY1*-like gene

The phylogenetic tree of *PSY1* gene (Fig. 3) reveals that the Andropogoneae of the Panicoid subfamily is monophyletic. The Tripsacum taxa can be united with Zea with 100% bootstrap support; and the Sorghum clade is joined to the Tripsacum + Zea clade at a bootstrap value of 59%. All coix studied here were placed in a single clade which can

Table 3 Summary of the natural variation of different regions of the *PSY1* gene in different species

Taxon	*N*	SNP (number/100 bp)			InDel (number/100 bp)			All (number/100 bp)			Total (number/100 bp)
		UTR	Intron	Exon	UTR	Intron	Exon	UTR	Intron	Exon	Full length
Maize (all)	56	1.20	2.07	1.38	0.52	0.49	0.00	1.72	2.56	1.38	5.66
Maize (yellow)	38	0.57	0.79	0.41	0.28	0.22	0.00	0.85	1.01	0.41	2.27
Maize (white)	18	1.11	1.94	1.05	0.46	0.4	0.00	1.57	2.34	1.05	4.96
Teosinte (all)	7	2.45	3.66	1.13	1.13	0.67	0.16	3.58	4.33	1.29	9.20
Teosinte (luxuriantes)	4	1.96	2.61	0.49	0.88	0.5	0.16	2.84	3.11	0.65	6.60
Teosinte (zea)	3	1.81	1.28	0.73	0.39	0.28	0.00	2.2	1.56	0.73	4.49
Sorghum	4	0.09	0.19	0.08	0.09	0.13	0.08	0.18	0.32	0.16	0.66
Coix	8	0.83	4.23	3.45	0.24	0.95	0.08	1.07	5.18	3.53	9.78
Rice	4	0.15	0.68	0.16	0.59	0.11	0.08	0.74	0.79	0.24	1.77

Number/100 bp: the total mutants in one set × 100/sequence length in intraspecific alignment

The minimum extraction frequency for polymorphism = 2/*N*

N Number of materials used

Table 4 Summary of *PSY1* DNA sequence variation in Andropogoneae

Parameter	Coix	Maize			Teosinte		Sorghum
		White	Yellow	All maize	Section zea	Section luxuriantes	
	273/8[a]	155/18	40/38	169/56	75/3	283/4	15/4
Total π ($\times 10^3$)	24.17	10.79	0.67	6.68	9.74	37.47	1.83
Silent π ($\times 10^3$)	27.72	12.88	0.73	7.93	11.27	46.03	1.65
Nonsynonymous π ($\times 10^3$)	9.15	2.58	0.42	1.83	2.89	0.00	0.00
Total θ ($\times 10^3$)	26.4	9.77	1.90	8.16	9.80	37.47	1.49
Silent θ ($\times 10^3$)	0.00	11.71	2.17	9.71	11.34	46.03	0.00
Nonsynonymous θ ($\times 10^3$)	0.00	2.18	0.77	2.12	2.89	0.00	2.09
Tajima's *D*	−0.64	0.43	−2.35**	−0.64	–	–	1.62
HKA (*P* value)		6.65 (0.0099)	38.80 (0.0000)	9.03 (0.0027)	4.52 (0.0335)	0.09 (NS)	–

– Tajima's *D* cannot be detected because of <4 lineages tested among the intraspecies

** Significant at $P = 0.01$ level

[a] Number of segregating sites to number of sequences used

be united with the Zea + Tripsacum + Sorghum clade with 100% bootstrap support, lending weight to the placement of Coix in the Andropogoneae rather than the "Maydeae" and other monoecious genera.

The phylogenetic tree of the Zea species showed that the yellow maize lines and all other white endosperm Zea were placed in one clade with 95% bootstrap support (Fig. 4). Figure 4 reveals the formation of subclades within the entire group as well, with a single major clade for all yellow maize lines, dispersal of the white lines into two groups, and other Zea species joining the group in known order of relationship. Within the group of yellow maize lines, three clades can be easily distinguished based on the different insertion pattern observed in the UTRs. Three lines with both the 390 bp insertion in the 5′UTR and the 345 bp insertion in the 3′UTR group together, and are further combined into a much larger clade with lines that have the 390 bp insertion in the 5′UTR but lack the 3′UTR insertion (with a bootstrap value of 98%). The two yellow lines, HZS and Chang7-2, which lack both insertions, form their own separate clades unrelated to either the other two clades of yellow maize and to each other; at the sequence level, they are differentiated from each other by the 18 bp InDel in the 3′UTR. Within the white *Z. mays* lines, one clade contains only white endosperm maize, and the other contains white endosperm maize plus ssp. *mexicana* and *huehuetenangensis*. The *parviglumis* teosinte is clustered with the yellow + white *Z. mays* clade with 82% bootstrap support, and this large clade is next joined by the Tripsacum clade and one *diploperennis* species. Finally, the two

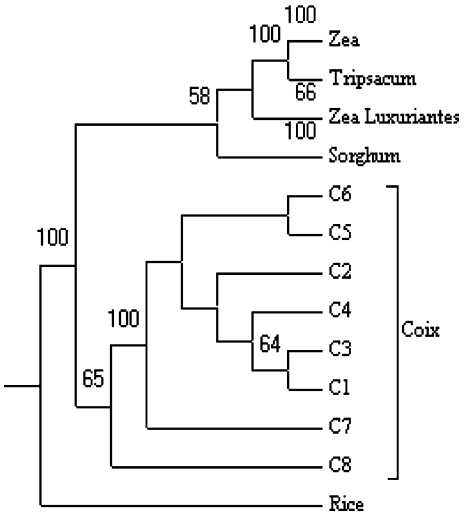

Fig. 3 Neighbor-joining tree in Andropogoneae based on the full-length DNA sequences of the *PSY1* gene. Bootstrap values are shown as percentages over 1,000 replicates

luxuriantes species form sister clades with 100% bootstrap support, joined by one more *diploperennis* species, and this *luxuriantes*/*diploperennis* clade is the last to join the rest of the *Zea* lines in this study.

Testing for positive selection of *PSY1*

The Tajima's *D* test for *PSY1* sequences from teosinte and coix species are all nonsignificant (Table 4). Sorghum lines tested in this study have similar nucleotide diversity as yellow maize lines, but reveal no significant effects of selection (Table 4). The three domains of *PSY1*-like sequences showed different nucleotide substitution patterns, with the highest mean d_N/d_S value (1.479) in the SP domain, and the lowest in the PSY (0.043) and SQS (0.097) domains (Table 2; Fig. 1). The region between residues 100 and 130 had the highest d_N/d_S ratio (1.98) in the entire coding sequence. Mean d_N/d_S ratios for the entire *PSY1* sequence were highest in sorghum (0.840), high for yellow maize (0.505), moderate in white maize, coix and *Zea* teosinte (0.258, 0.266, and 0.325, respectively) and very low in *luxuriantes* teosinte (0.016). Although the d_N/d_S ratios for all species tested are very different, all are <1.0, providing no significant signs of positive selection. The results of HKA test for selection in maize support the conclusions from the Tajima's *D* test, in that yellow maize lines showed significant values for artificial selection ($P = 0.0000$). The 3′UTR showed a positive, significant Tajima's *D* value in white maize lines, implying balancing selection in white germplasm (Table 5). The lowest π and θ values were found in the 3′UTR for white maize and 5′UTR for yellow maize (Table 5), indicating targeted selection in these regions of *PSY1*.

Discussion

Structural domains within the *PSY1*-like gene sequences in the different species

The predicted structure and amino acid content of *PSY1* in rice, sorghum, and maize in this study are in agreement with previously reported results (Gallagher et al. 2004; Li et al. 2008a), indicating that the same sequence was successfully cloned. All species tested here have the same number of exons and introns with similar length and sequence order, implying conservation of this gene sequence during the evolution of the grasses. There are three domains in the aligned amino acid sequences among the species tested. The SP domain, encoding most secretory proteins that work across the endoplasmic reticular membrane, is observed in all tested species except rice, which was also seen in a previous study (Gallagher et al. 2004). Lacking the SP domain entirely, the rice PSY1 protein cannot be located on the endoplasmic reticular membrane, so carotenoids cannot accumulate in rice endosperm (Gallagher et al. 2004). They do accumulate when the *ZmPSY1* gene with the SP domain is inserted into rice via genetic engineering (Paine et al. 2005), confirming the need for this domain in *PSY1* for carotenoid accumulation in the endosperm. The variation in the helix-breaking residues of the SP domain region between sorghum and yellow maize may influence the efficiency of cleavage by the signal peptidase, which may partially explain the differences in functional expression reported in the endosperm of sorghum and yellow maize (Salas Fernandez et al. 2008). The SQS and PSY domains are highly conserved (Fig. 1, Supplementary Fig. S1), which may be due to its necessary interaction with upstream products (Doebley et al. 2006). The region between residues 100 and 130 has a very different evolutionary pattern among the species tested, with the highest d_N/d_S ratio (much greater than 1.0), providing strong evidence for its importance for differential evolution within the different species, rather than for a conserved evolution within all grasses.

Nucleotide diversity in the different species

The level of structural heterogeneity among *PSY1* sequences (Table 3) is different from that seen in the *tb1* gene (Lukens and Doebley 2001), implying that SNPs might also be important for creating novel variation during species evolution. The *PSY1* gene has a higher d_N/d_S ratio in coding regions relative to other plant genes, which is also seen in the directionally selected *tb1* gene (Lukens and Doebley 2001; Purugganan 1998); this implies a rapid evolution of *PSY1* sequences. Among the Andropogoneae, coix has the highest level of variation

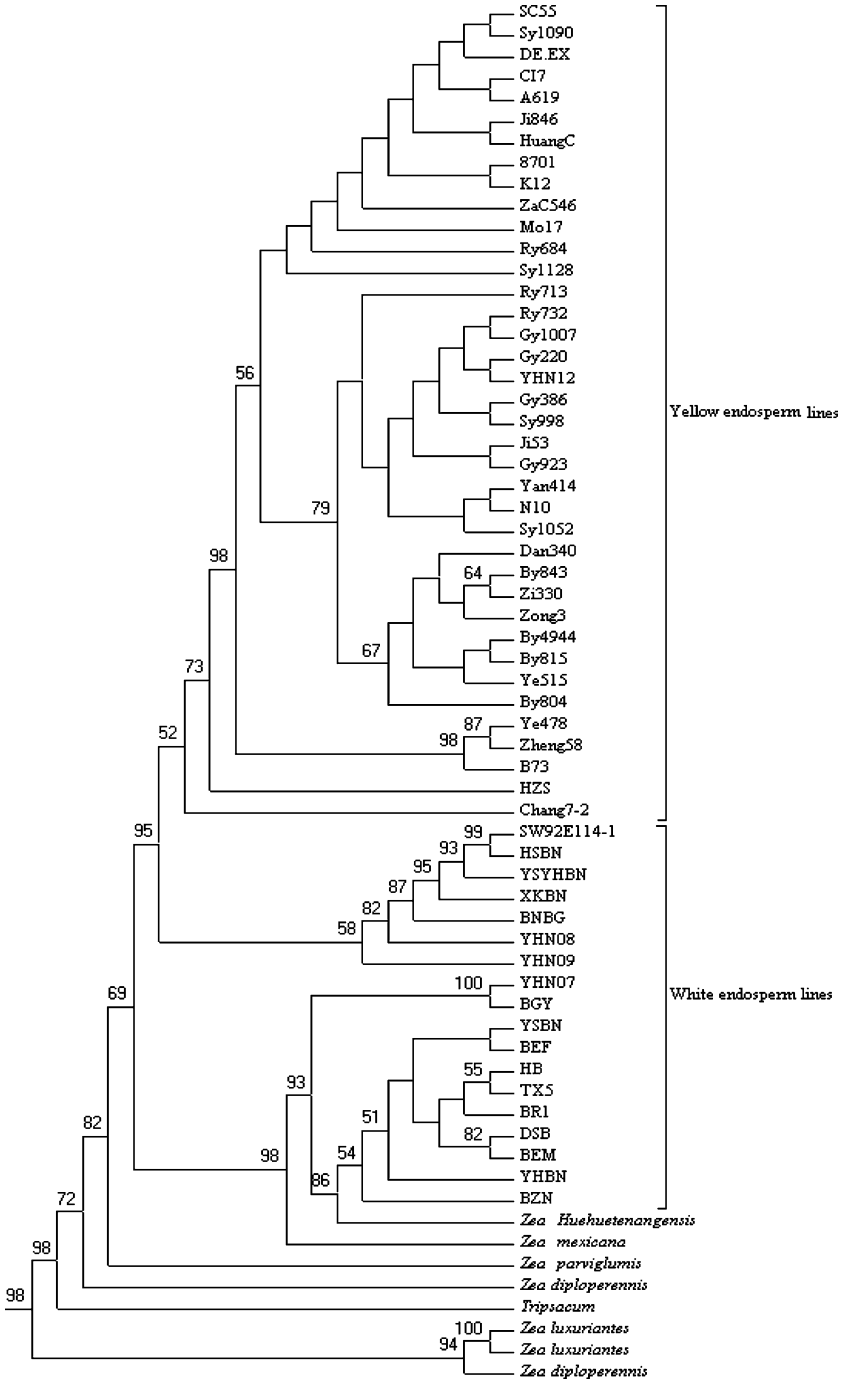

Fig. 4 Maximum-likelihood tree in *Zea* based on full-length DNA sequences of the *PSY1* gene. Cutoff value for consensus tree is 50%. Bootstrap values are shown as percentages over 1,000 replicates

with 9.78 polymorphisms/100 bp and the highest π and θ values, probably due to the lack of artificial selection. The two white wild maize subspecies (*Z. mays* ssp. *parviglumus* and *Z. mays* ssp. *mexicana*) have a similar diversity level, which is higher than the yellow maize subspecies (yellow *Z. mays mays*) along with the lowest diversity level among the *Zea* species. These results imply that white is the ancestral state of *PSY1* gene, and yellow may be a post-domestication mutation, a conclusion similarly reached by Palaisa et al. (2003). In this study, we see that the haplotypes composed of the 378 and 390 bp InDels, and not the two core SNPs defined by Palaisa et al. (2003), completely associate with yellow lines in both data sets. These results allow the possibility to definitively separate the white and yellow classes of maize based on the sequence of the 5′UTR, where the large insertions in the 5′UTR are in LD with most of the SNPs in the exons of all yellow, and no white, maize lines tested.

Previous research has shown that total carotenoid content of diverse maize germplasm varied from 5.5 to

Table 5 Summary of natural variation of different regions of the *PSY1* gene in maize

Region	Parameter	Total	White	Yellow
5′UTR	*N*	53	48	6
	π ($\times 10^3$)	8.69	13.42	0.24
	θ ($\times 10^3$)	8.24	9.81	0.94
	Tajima's *D*	0.19 NS	1.52 NS	−2.01*
CDS	*N*	19	18	5
	π ($\times 10^3$)	3.73	5.81	0.52
	θ ($\times 10^3$)	3.71	4.56	0.97
	Tajima's *D*	0.02 NS	1.08 NS	−1.18 NS
Intron	*N*	55	52	20
	π ($\times 10^3$)	5.78	9.33	0.82
	θ ($\times 10^3$)	8.25	10.43	2.7
	Tajima's *D*	−1.03 NS	−0.44 NS	−2.32**
3′UTR	*N*	17	7	15
	π ($\times 10^3$)	3.23	3.43	1.22
	θ ($\times 10^3$)	3.85	2.07	3.5
	Tajima's *D*	−0.49 NS	2.22*	−2.09*

N Number of segregating sites, *NS* not significant, *CDS* coding sequences

* Significant at $P = 0.05$ level

** Significant at $P = 0.01$ level

66.0 μg/g with an average of 22.8 μg/g (Harjes et al. 2008). A QTL with a large effect on total carotenoid content and its components was mapped to the same genomic region as the *PSY1* gene using a segregating population generated by crossing two yellow maize lines (Chander et al. 2008; Wong et al. 2004). Since only loci with alleles having a different phenotypic effect on a trait can be identified via linkage mapping, this implies that there should be sufficient variation among different *PSY1* alleles in yellow maize germplasm to continue improvement of this trait without the need for introducing crosses with more distant germplasm, or genes from unrelated species via transgenesis. Unlike the results of Palaisa et al. (2003), we detected LD decay in the *PSY1* gene in yellow lines, allowing further analysis of association. The 644 bp insertion at the transcriptional site in the data set presented here may regulate the expression of the *PSY1* gene. Variants found among the yellow maize lines from Chinese temperate germplasm (Supplementary Table S3, S4), combined with the new confirmation that LD does decay (albeit slowly) in yellow maize demonstrate that association analysis in modern yellow maize germplasm is feasible. Comparing allelic variation at the *PSY1* locus with changes in carotenoid accumulation will enable the development of marker-assisted selection tools to improve the efficiency and impact of breeding for enhanced provitamin A content in maize.

The *PSY1*-like gene phylogeny

There are many reports of the evolution and phylogenetic relationships of species within the Andropogoneae based on phenotypic, archaeological, and cytotaxonomic studies. More recently, molecular evidence based on variation in coding sequences of individual gene has been used to validate and refine these conclusions based on chloroplast gene phylogeny (Davis and Soreng 1993), ribosomal gene phylogeny (Buckler and Holtsford 1996), and nuclear gene phylogeny (Bomblies and Doebley 2005; Giussani et al. 2001; Lukens and Doebley 2001; Mason-Gamer et al. 1998; Mathews et al. 2000, 2002; Spangler et al. 1999). Although exons, sometimes with only low statistical support, are variable enough to provide extensive resolution within taxonomic families, diversity within introns and UTRs is particularly useful for differentiating between closely related species (Mason-Gamer et al. 1998).

Our suggestion that the Andropogoneae of the Panicoid subfamily is monophyletic differs from the results from the *tb1* gene (Lukens and Doebley 2001), but is in agreement with reports based on other genes (Clark et al. 1995; Giussani et al. 2001; Mason-Gamer et al. 1998; Mathews et al. 2000, 2002; Spangler et al. 1999). It appears that sorghum is more closely related to Zea than the other species tested within the Andropogoneae, although with weak bootstrap support (58%). This has been reported previously based on sequence analysis of several other nuclear genes (Bomblies and Doebley 2005; Lukens and Doebley 2001; Mathews et al. 2002; Spangler et al. 1999). Results of the current study also confirm Tripsacum as the sister genus of Zea, as reported previously (Bomblies and Doebley 2005; Giussani et al. 2001; Lukens and Doebley 2001; Mason-Gamer et al. 1998; Mathews et al. 2000, 2002; Spangler et al. 1999). The topological structure of the phylogenetic tree within the Andropogoneae described in the current study is the same as that reported for the *GBSSI* gene (Mason-Gamer et al. 1998), ribosomal ITS sequences (Buckler and Holtsfor 1996) and the nuclear gene *tb1* (Lukens and Doebley 2001). All support that coix should be classified to Andropogoneae and that the term "Maydea" should be abandoned in taxonomy. The relationship of Coix clade with the Zea + Tripsacum + Sorghum clade implies that coix may have evolved before the divergence of maize and sorghum, but after the divergence of the Andropogoneae and rice. This hypothesis is supported by the fact that maize, sorghum and coix all share the same chromosome number, and all have been hypothesized to be descended from an ancestral species that had a haploid genome of 5 ($n = 5$) (Paterson et al. 2004). The same ancestral species is thought to have evolved into three new genomes, one with five chromosomes, one following chromosomal duplication to form maize and related species with ten chromosomes, and one

following chromosomal split to form sorghum with ten chromosomes (Paterson et al. 2009; SwigoHová et al. 2004).

In this study, the *luxuriantes* section species of teosinte are the very last to join with the other *Zea* accessions, instead of clustering with the *Zea* section species of teosinte and white *Z. mays mays* as previously reported (Palaisa et al. 2003), providing a fine phylogenetic tree for *Zea* species. Another difference compared to previous studies is that the yellow clade in the present study can be further separated into four subclades according to the haplotypes of the 390 (378/644 bp), 345 and 18 bp InDels in the UTRs. In addition to the common haplotypes identified in both studies, we found an additional rare haplotype in the maize inbred HZS. These haplotypes work well in differentiating the yellow lines for both data sets.

Evidence for positive selection

Amino acid changes among species in the *PSY1* sequence were constrained, although different amino acid positions probably experienced different degrees of selection. Variations of the d_N/d_S ratio (<1.0) and nonsignificant Tajima's *D* indicated there is no evidence for positive selection overall in *PSY1* among all accessions tested here. Variations of the d_N/d_S ratio among the Andropogoneae can be explained by unequal levels of purifying selection among lineages. However, the SP domain and the region upstream of the SQS–PSY domain appear to have undergone rapid evolution because a high d_N/d_S ratio (>1.0) was evidenced. Other methods were also used to test the variation of this region, like sliding window and HKA test (data not show). Analysis of Tajima's *D* test for CDS through Sliding window had negative significant p value in the SP domain (−1.7599) and the region between residues 100 and 130 (−1.6452). When rice was used as an outgroup for HKA test of the SP domain, significant level for sorghum ($*P = 0.0401$), coix ($*P = 0.0179$), teosinte ($*P = 0.0321$) and maize ($*P = 0.0061$) were detected. Within *Z. mays* ssp. *mays*, we could summarize the evolution pattern of *PSY1* as follows: the first exon of *PSY1* (especially the SP domain and the residues 100–130) has primarily been subjected to selection pressure during the evolution of the grasses. This is followed by positive selection on regulatory regions altering the expression levels of *PSY1* in maize after divergence from the other grasses (Doebley et al. 2006; Palaisa et al. 2003). This caused evolutionary differences between species for this gene. On the other hand, resembling selection in this region could be also created by sequencing errors and/or population demographic changes. Therefore, the further validation will be needed for this proposed evolution pattern of *PSY1*. While, sorghum, in which *PSY3* gene is the major gene influencing the carotenoids accumulation in endosperm, may have different evolution pattern from that of maize as evidenced by no selection signals on the *PSY1* gene and no QTLs for carotenoids co-located with *PSY1* (Salas Fernandez et al. 2008). The *PSY1* gene is necessary for yellow endosperm color and carotenoid accumulation in maize (Gallagher et al. 2004; Palaisa et al. 2003). The unusually low diversity level and slow LD decay accompanied by an excess of rare alleles (Tajima's $D = -2.35$), significant HKA value and unusually low π_M/π_T (0.0179) value compared to neutral genes (≈0.75) (Clark et al. 2004) observed in the *PSY1* gene in yellow maize lines suggest its involvement in maize improvement via artificial selection. The relatively faster LD decay and the significant positive Tajima's *D* value in the 3′UTR in white maize lines verify that *PSY1* has high diversity and has experienced balancing selection (Tajima 1989). The selective sweep reducing the sequence diversity in yellow maize only causes yellow maize lines to form a separate clade in phylogenetic analyses using the *PSY1* gene.

Acknowledgments This research was supported by the National Natural Science Foundation of China (30821140352) and the specific project grants from the Harvest Plus Program and targeted funds from the World Bank and European Commission as well as from USAID, UK DFID and Canadian CIDA to International Maize and Wheat Improvement Center. Authors also greatly appreciate both anonymous reviewers for their invaluable comments.

References

Bird CR, Ray JA, Fletcher JD, Boniwell JM, Bird AS, Teulieres C, Blain I, Bramley PM, Schuch W (1991) Using antisense RNA to study gene function: inhibition of carotenoid biosynthesis in transgenic tomatoes. Nat Biotechnol 9:635–639

Bomblies K, Doebley JF (2005) Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). Mol Biol Evol 22:1082–1094

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

Bramley P, Teulieres C, Blain I, Bird C, Schuch W (1992) Biochemical characterization of transgenic tomato plants in which carotenoid synthesis has been inhibited through the expression of antisense RNA to pTOM5. Plant J 2:343–349

Buckler ES, Holtsford TP (1996) Zea systematics: ribosomal its evidence. Mol Biol Evol 13:612–622

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9:172–176

Chander S, Guo YQ, Yang XH, Zhang J, Lu XQ, Yan JB, Song TM, Rocheford TR, Li JS (2008) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theor Appl Genet 16:223–233

Clark LG, Zhang W, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on ndhF sequence data. Syst Bot 20:436–460

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

Clarke KR, Gorley RN (2001) PRIMER v5: user manual/tutorial. PRIMER-E Ltd, Plymouth, p 91

Davis JI, Soreng RJ (1993) Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. Am J Bot 80:1444–1454

Doebley J (1990) Molecular evidence and the evolution of maize. Econ Bot 44:6–27

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

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. Genetics 134:559–570

Doebley JF, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (Gramineae). Syst Bot 9:203–218

Doebley JF, Renfroe W, Blanton A (1987) Restriction site variation in the *Zea* chloroplast genome. Genetics 117:139–147

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

Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433–2437

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Res 32:1792–1797

FAO (1995) Sorghum and millets in human nutrition. Food and nutrition series v.27, chap 2 and 5. Food and Agriculture Organization of the United Nations, Rome

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1993) PHYLIP (Phylogeny inference package) Ver.3.57c. Department of Genetics, University of Washington, Seattle, WA. http://evolution.genetics.washington.edu/phylip.html

Fray RG, Grierson D (1993) Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Mol Biol 22:589–602

Gallagher CE, Matthews PD, Li FQ, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grassed. Plant Physiol 135:1776–1783

Giuliano G, Bartley GE, Scolnik PA (1993) Regulation of carotenoid biosynthesis during tomato development. Plant Cell 5:379–387

Giussani LM, Cota-Sánchez JH, Zuloaga FO, Kellogg EA (2001) A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C4 photosynthesis. Am J Bot 88:1993–2012

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

Harjes CE, Rocheford T, Bai L, Brutnell T, Kandianis CB, Sowinski S, Stapleton A, Vallabhaneni R, Williams M, Wurtzel E, Yan JB, Buckler ES (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319:330–333

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proc Natl Acad Sci USA 92:1679–1683

Li FQ, Vallabhaneni R, Yu J, Wurtzel ET (2008a) *PSY3*, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced root carotenogenesis. Plant Physiol 146:1333–1345

Li FQ, Vallabhaneni R, Yu J, Rocheford T, Wurtzel ET (2008b) The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress-tolerance. Plant Physiol 147:1334–1346

Lukens L, Doebley J (2001) Molecular evolution of the *teosinte branched* gene among maize and related grassed. Mol Biol Evol 18:627–638

Mangelsdorf PC, Fraps GS (1931) A direct quantitative relationship between vitamin A in corn and the number of genes for yellow pigmentation. Science 73:241–242

Mason-Gamer RJ, Weil CE, Kellogg (1998) Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol Bio Evol 15:1658–1673

Mathews S, Tsai RC, Kellogg EA (2000) Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. Am J Bot 87:96–107

Mathews S, Spangler RE, Mason-Gamer RJ, Kellogg EA (2002) Phylogeny of *Andropogoneae* inferred from phytochrome B, GBSSI, and NDHF. Int J Plant Sci 163:441–450

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

Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF, Wing RA (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci USA 101:14349–14354

Murray MG, Thompson WF (1980) Rapid isolation of high-molecular weight plant DNA. Nucleic Acid Res 8:4321

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

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of golden rice through increased pro-vitamin A content. Nat Biotechnol 23:482–487

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

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al (2009) The sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

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

Purugganan M (1998) The molecular evolution of development. Bioessays 20:700–711

Salas Fernandez MG, Hamblin MT, Li L, Rooney WL, Tuinstra MR, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Sci 48:1732–1743

Salvi S, Sponza G, Morgante M, Tomes D, Niu XM, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggenann E, Li BL, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao GH, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic

sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104:11376–11381

Sang T (2009) Genes and mutations underling domestication transitions in grasses. Plant Physiol 149:63–70

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in *Sorghum* (*Poaceae*) using ndhF sequences. Syst Bot 24:267–281

SwigoHová Z, Lai JS, Ma JX, Ramakrishna VL, Liaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Swofford DL (1998) PAUP*, phylogenetic analysis using parsimony (* and other methods) version 4.0 beta. Sinauer Associates, Sunderland, MA

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Res 25:4876–4882

Tiffin P, Gaut BS (2001) Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics 158:401–412

Von Lintig J, Welsch R, Bonk M, Giuliano G, Batschauer A, Kleinig H (1997) Light-dependent regulation of carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. Plant J 12:625–634

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 genetic models without recombination. Theor Popul Biol 7:256–276

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

Wong JC, Lambert RJ, Wurtzel ET (2004) QTL and candidate genes phytoene synthase and carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

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