Molecular Evolution of the *teosinte branched* Gene Among Maize and Related Grasses

Lewis Lukens[1] and John Doebley[2]

Department of Plant Biology, University of Minnesota at Minneapolis St. Paul

Several authors have proposed that changes in a small number of regulatory genes may be sufficient for the evolution of novel morphologies. Recent analyses have indicated that *teosinte branched1* (*tb1*), a putative bHLH transcription factor, played such a role during the morphological evolution of maize from its wild ancestor, teosinte. To address whether or not *tb1* played a similar role during the evolution of the Andropogoneae, the tribe to which maize belongs, and to examine the rate and pattern of *tb1* evolution within this tribe, we analyzed *tb1*-like sequences from 23 members of the Andropogoneae and five other grasses. Our analysis revealed that the TB1 protein evolves slowly within three conserved domains but rapidly outside these domains. The nonconserved regions of the gene are characterized by both a high nonsynonymous substitution rate and frequent indels. The ratio of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S) was not significantly greater than 1.0, providing no evidence for positive selection. However, the d_N/d_S ratio varied significantly among lineages and was high compared with those of other plant nuclear genes. Variation in the d_N/d_S ratio among the Andropogoneae could be explained by unequal levels of purifying selection among lineages. Consistent with this interpretation, the rate of nonsynonymous substitution differed along several lineages, while the synonymous substitution rate did not differ significantly. Finally, using *tb1,* we examined phylogenetic relationships within the Andropogoneae. The phylogeny suggests that the tribe underwent a rapid radiation during its early history and that the monoecious Andropogoneae are polyphyletic.

Introduction

Molecular genetic analyses have demonstrated that transcription factors often act as switches between discrete developmental programs, encouraging the view that novel morphologies may arise from changes in this class of genes (Goodrich, Carpenter, and Coen 1992; Carroll 1994; Doebley and Lukens 1998). Differences in the expression patterns of transcription factors may alter the timing and location of developmental events and thereby generate morphological variation. Indeed, changes in the level and localization of transcription factor mRNA have been correlated with phenotypic differences within and between species (Doebley, Stec, and Hubbard 1997; Stern 1998; Cubas, Vincent, and Coen 1999). The analysis of DNA sequence variation provides another means of investigating the contribution of transcription factors to morphological evolution. For example, naturally occurring alleles of the MADS box transcription factor, *cauliflower,* possess an excess of intraspecific nonsynonymous substitutions, and this variation is associated with effects on floral morphology (Purugganan and Suddith 1998). Analyses of coding sequence variation can also provide evidence for past episodes of positive selection (e.g., Sutton and Wilkinson 1997) and variation in the rate of sequence evolution (e.g., Yang and Nielsen 1998).

The *teosinte branched1* (*tb1*) gene belongs to the TCP gene family whose members encode putative basic helix-loop-helix DNA-binding proteins that appear to play a role in organ growth. In rice, two members of the TCP family, PCF1 and PCF2, have been shown to act as DNA-binding proteins and to participate in cell cycle regulation (Kosugi and Ohashi 1997). Another TCP gene is *cycloidea* of the snapdragon, which is required for the formation of a bilaterally symmetrical flower (Luo et al. 1996). A naturally occurring hypermethylated allele of *cycloidea* cosegregates with a radially symmetric floral phenotype in Linaria, a relative of the snapdragon (Cubas, Vincent, and Coen 1999). *tb1* itself corresponds to a quantitative trait locus that distinguishes maize from its wild ancestor teosinte in plant architecture, and the maize allele of *tb1* shows evidence of having undergone a selective sweep during maize domestication (Wang et al. 1999). Maize plants introgressed with the teosinte allele of *tb1* have long side branches like teosinte (Doebley, Stec, and Gustus 1995; Lukens and Doebley 1999).

The involvement of TCP gene family members in the morphological evolution of two distantly related flowering plants (maize and Linaria) suggests that these genes may have frequently contributed to morphological evolution. To investigate this possibility, we chose to analyze the evolution of *tb1*-like genes among members of the grass tribe Andropogoneae. This tribe includes maize and a variety of other grasses, such as sorghum and sugarcane. The morphology of the tribe is remarkably diverse, most notably in floral and inflorescence characteristics (Clayton 1987). The tribe likely originated within the past 30 Myr and underwent a rapid radiation (Mason-Gamer, Weil, and Kellogg 1998; Spangler et al. 1999). If *tb1* was involved in the morphological diversification of the tribe, the pattern of sequence evolution in *tb1* may contain evidence for past episodes of positive selection.

To examine whether the pattern of DNA sequence variation in *tb1*-like genes is consistent with a history

Table 1
Names and Origins of Analyzed Taxa, Including Collection Numbers

Sample	Origin	Source[a]	Collection
Bambusoideae, Oryzeae			
Oryza sativa L.	United States	USDA	CIor 1344
Panicoideae, Andropogoneae			
Andropogon gerardii Vitman	United States	LL-UMC	—
Bothriochloa odorata (Lisboa) A. Camus[b]	India	USDA	PI 301632
Capillipedium parviflorum (R. Br.) Stapf[c]...........	India	USDA	PI 301782
Chionachne koenigii (Spreng.) Thwaites............	India	NBPGR	Chio-6
Chrysopogon fulvus (Spreng.) Chiov.[c].	Pakistan	USDA	PI 199241
Cleistachne sorghoides Benth.[d]	Malawi	ICRISAT	IS 14340
Coelorachis selloana (Hack.) Camus[c]	Brazil	USDA	PI 309987
Coix lacryma-jobi L.............................	Brazil	USDA	PI 320865
Cymbopogon distans (Nees ex Steud.) J. F. Watson ..	India	USDA	PI 271552
Elionurus tripsacoides Humb. & Bonpl. ex Willd. ...	Texas	JD	646
Heteropogon contortus (L.) P. Beauv. ex Roem. & Schult.[e]..	United States	USDA	PI 216429
Ischaemum afrum (J. F. Gmel.) Dandy	South Africa	USDA	PI 364923
Phacelurus digitatus (Sibth. & Sm.) Griseb.[c]........	Turkey	USDA	PI 206746
Saccharum officinarum L.........................	N/A	LL-UMG	—
Sorghastrum nutans (L.) Nash	United States	LL-UMC	—
Sorghum bicolor (L.) Moench.....................	Ethiopia	USDA	PI 302105
	Uganda	ICRISAT	18803
Sorghum australiense Garber & Snyder[f]............	Australia	KS	85-261
Sorhum matarakense Garber & Snyder[f]	Australia	KS	85-242
Sorghum nitidum (Vahl) Pers.[f]	Australia	KS	85-248
Tripsacum cundinamarce de Wet & Timothy.........	Colombia	DT	63-6
Zea diploperennis Iltis Doebley & Guzman	Mexico	RG	1120
Zea mays L......................................	United States	JD	Inbred W22
Panicoideae, Arundinelleae			
Arundinella hirta (Thunb.) Tanaka[c]	Japan	USDA	PI 246756
Danthoniopsis dinteri (Pilg.) C. E. Hubb.[c]	South Africa	USDA	PI 207548
Panicoideae, Paniceae			
Panicum virgatum L.[g]............................	Papua New Guinea	USDA	PI 356816
Pennisetum glaucum (L.) R. Br...................	United States	NSL	97490

[a] DT = Dr. David Timothy; ICRISAT = International Crops Research Institute for Semi-Arid Tropics; JD = Dr. John Doebley; KS = Dr. Keith Shertz; LL = Dr. Lewis Lukens (University of Minnesota Campus [UMC] or Greenhouse [UMG]); NBPGR = National Bureau of Plant Genetic Resources, New Delhi; NSL = National Seed Laboratory; RG = Dr. R. Guzman; USDA = United States Department of Agriculture.

[b] Identified as *Bothriochloa bladhii* (Retz.) S. T. Blake by Spangler et al. (1999).

[c] Voucher specimen preserved at the Harvard University Herbaria (A or GH); identification verified by E. A. Kellogg.

[d] Voucher specimen preserved at the Iowa State University Herbarium (ISC).

[e] Voucher specimen at the Harvard University Herbaria (A/GH) identified by E. A. Kellogg as *Hyparrhenia rufa* (Nees) Stapf.

[f] Voucher specimen preserved at the University of Minnesota Herbarium (MIN).

[g] Voucher specimen at the Harvard University Herbaria (A/GH) identified by E. A. Kellogg as *Panicum repens* L.

of past positive selection, we analyzed 27 *tb1*-like genes from within the tribe Andropogoneae and the subfamily to which it belongs, Panicoideae. We also compared these genes with an additional *tb1*-like sequence from rice (*Oryza sativa*), a more distantly related grass. We addressed several questions. How variable is *tb1* among these grasses both overall and within specific regions of the gene? Does either the *tb1* gene overall or any specific region within it possess the signature of positive selection as measured by an excess of nonsynonymous substitutions per nonsynonymous site (d_N) relative to synonymous substitutions per synonymous site (d_S)? Does the d_N/d_S ratio vary among lineages, suggesting lineage-specific episodes of positive selection? Has *tb1* evolved in a clocklike fashion at both synonymous and nonsynonymous sites as expected for a strictly neutral gene? In general, our analyses suggest that although some parts of the coding region have evolved rapidly relative to others, the pattern of evolution is consistent with models that do not invoke positive selection. Similarly, although the d_N/d_S ratio varies among lineages, our analyses suggest that this is more likely a function of relaxed constraint within some lineages than a function of positive selection.

Materials and Methods
Plant Material

tb1-like sequences were isolated from members of the tribe Andropogoneae of the subfamily Panicoideae of the grass family, Poaceae (table 1). The taxa were selected for broad representation of the Andropogoneae, with the inclusion of rice and several nonandropogonoid members of the panicoid subfamily as outgroups.

Nucleic Acid Isolation, Southern Transfer, and Membrane Hybridization

Genomic DNA extractions, restriction digestions, electrophoresis, and Southern hybridizations were per-

formed as previously described (Doebley and Stec 1993) except that membranes were washed under less stringent conditions following hybridization with a radiolabeled probe. Membranes were washed three times for 30 min at 65°C in 0.5 × SSC and 0.2% SDS. Cloned restriction fragments of *tb1* (H3.4 or NH0.8) from maize were used as probes (Doebley, Stec, and Hubbard 1997).

Library Construction

Phage libraries for rice, sorghum, *Zea diploperennis,* and *Tripsacum cundinamarce* were constructed from size-selected genomic DNA fragments. Genomic DNAs were digested with *Eco*RI, *Hind*III, or *Sst*I, and Southern blot analyses were performed. These blots revealed only one strongly hybridizing band for each of these species when hybridized with the radiolabeled maize *tb1* probe (see *Results*). Based on these results, the genomic DNAs were digested again and electrophoresed (0.8% 1 × TAE low-melting-point agarose gel), and fragments the size of the *tb1*-hybridizing band were excised from the gel. The selected DNA fragments were purified from the agarose plug by agarase digestion (Boehringer) followed by nucleic acid precipitation. The size-selected genomic DNA fragments were cloned into Lambda ZapII or Lambda DashII vectors (Stratagene).

Between 150,000 and 1,600,000 plaques were screened from each library using the plaque lift procedure as described by Sambrook, Fritsch, and Maniatis (1989). Probes and hybridization conditions were identical to those described above for Southern blotting. Areas of the plate that contained plaques which strongly hybridized to the probe were excised, and the selected phage population was screened two additional times, until a plaque containing a single phage genotype was obtained. For Lambda DashII, lambda DNA was harvested and the insert was subcloned into pBluescript using standard plating, harvesting, and subcloning techniques (Sambrook, Fritsch, and Maniatis 1989). For Lambda ZapII, autoexcision was used to isolate plasmids containing the cloned insert as per the manufacturer's instructions (Stratagene).

Primers, PCR, and Subcloning

PCR with oligonucleotide primers was used to amplify *tb1*-like genes from genomic DNA of taxa within the tribes Andropogoneae, Arundinelleae, and Paniceae. Primer names and sequences were JD96 (TCCCAT-CAGTAAAGCACATG), JD73 (GCTCTTGGCAGT-AGTAGTTGC), and JD75 (GTATCCTCCTCCTCC-GTTGC). JD96 was the forward primer and was used with either JD75 or JD73 as the reverse primer. Amplification was achieved using ≈50 ng genomic DNA as template with the following components: 50 μl PCR Supermix (BRL Life Technologies), 10 pmol of each primer, and 5 μl DMSO. Reactions were heated to 95°C for 1 min prior to cycling. Cycling parameters consisted of 36 cycles at 95°C for 1 min, between 42°C and 51°C for 1 min, and 75°C for 3 min. After the cycles were complete, samples were incubated at 75°C for 10 min and cooled to 4°C. PCR reaction products were electro-phoresed on agarose gels, and PCR products that were of the expected size (approximately 1,100 bp) were cloned using the TA or TOPO-TA cloning kits (Invitrogen). Ligation, transformation, and colony selection were carried out according to the manufacturer's instructions, and plasmids were purified from *Escherichia coli* using Qiaprep columns (Qiagen).

DNA Sequencing and Alignment

The coding and complementary strands of all plasmid inserts were sequenced using an ABI 373A automated DNA sequencing machine, with the exception of the nucleotide sequence of *Zea mays tb1,* which was previously reported (Doebley, Stec, and Hubbard 1997). Sequence data were edited with Sequencher 3.0 (Gene Codes Corp.), and sequences were aligned by eye using Se-Al, Sequence Alignment Editor, version 1.1 (Rimbaut 1996). Sequences have been deposited in GenBank (accession numbers AF322117–AF322143).

Phylogenetic and Evolutionary Analyses

Phylogenetic analyses were performed using both the maximum-likelihood and the parsimony methods of PAUP*, version 4.0 (Swofford 1999). For maximum-likelihood analyses, unless stated otherwise, the gamma shape parameter alpha was 0.5, base frequencies were determined empirically from the sequence, and the transition : transversion ratio was set at the default value of 2.0. To find all shortest trees and to identify multiple tree islands, 100 heuristic maximum-parsimony searches were done with tree bisection-reconnection (TBR) branch swapping and random order of taxon addition. Bootstrap support for nodes was estimated using the parsimony criterion with 1,000 bootstrap replicates. Different tree topologies were compared using the Templeton (1983) test and the Kishino-Hasegawa test (Kishino and Hasegawa 1989).

We used a sliding-window analysis in the WINA computer program to examine the ratio of nonsynonymous to synonymous substitution rates along the *tb1*-like sequences (Nei and Gojobori 1986; Ina et al. 1994). The ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site corrected for multiple hits (d_N/d_S) was estimated between two sequences for 60-bp windows at 3-bp intervals. Gaps were ignored in this analysis. MEGA was used to calculate pairwise d_S and d_N values for the sequence overall (Kumar, Tamura, and Nei 1993). Six of the 27 panicoid grasses were excluded from both the MEGA and the WINA analyses, since they differed by ≤0.020 nt per site from another sequence and thus contributed little additional information to the analysis.

The Macintosh version of PAML (Yang 1998) was used to compare strictly neutral and positive selection models of coding sequence evolution. In the strictly neutral model, the coding sequences were constrained to have evolved with d_N/d_S ratios of 0.0 (nonsynonymous substitutions do not occur) or 1.0 (nonsynonymous substitutions occur at the same rate per nonsynonymous site

as synonymous substitutions per synonymous site). These values reflect the strictly neutral expectation that almost all nonsynonymous mutations are either deleterious or neutral (Kimura 1983). In the positive-selection model, two d_N/d_S values were constrained to be 1.0 or 0.0, as above, but a third d_N/d_S statistic was inferred from the data and could take on any value. If the predicted third d_N/d_S ratio was significantly greater than 1.0, nonsynonymous changes were inferred to occur more frequently than synonymous substitutions at some codons, and there was evidence for positive selection. The difference between a strictly neutral model (d_N/d_S = 1.0 or 0.0) and a positive-selection model may be determined with a likelihood ratio test (Nielsen and Yang 1998).

PAML was also used to test the strictly neutral prediction that for a given gene phylogeny, the d_N/d_S ratio would be the same for all lineages. The likelihoods of two models were again compared. In the first model, a single d_N/d_S ratio is estimated and constrained to fit all lineages. In the second model, d_N/d_S statistics were estimated for every lineage. A likelihood ratio test was used to determine whether one of these two models explained the data better than the other (Yang and Nielsen 1998).

For both the coding and the lineage analyses, a total of 19 sequences were analyzed using PAML. With the exception of Arundinella, sequences from taxa not classified within the Andropogoneae (Panicum, Danthoniopsis, and Pennisetum) and Andropogoneae sequences that were very similar to other sequences (*Zea diploperennis, Capillipedium parviflorum, Chionachne koenigii, Sorghum australiense,* and *Sorghum bicolor* from Ethiopia) were excluded from the analysis. The transition : transversion ratio and the nucleotide frequencies at each codon position were estimated from the data. The maximum-likelihood tree generated by PAUP (see below) was used in all PAML analyses.

The program Codrates was used to detect both synonymous and nonsynonymous substitution rate variation along different lineages (Muse and Gaut 1994). Codrates examines substitution rates in trees with three taxa in which two sequences have diverged more recently than the third or outgroup sequence. The likelihoods of two models were compared. In the first model, the lineages from the most recent common ancestor to the two related taxa were constrained to have the same substitution rate. In the second model, these lineages could have different substitution rates. A likelihood ratio test was used to test which of these models best explained the data. As in the PAML analysis, Arundinella and 18 representative sequences from within the Andropogoneae were included.

Results
Cloning of *tb1*-like Sequences

As a first step, we wished to establish whether *tb1* existed as a multigene family or as a single-copy gene in the grasses. Southern blot analysis of genomic DNAs from nine members of the grass family using a *tb1* probe
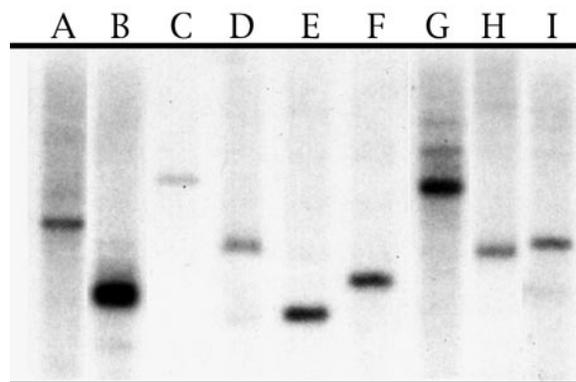

Fig. 1.—Southern blot of genomic DNAs from several taxa restricted with *Hin*dIII and hybridized with a radiolabeled ^{32}P maize genomic *tb1* clone (NH0.8). The blot shows that *tb1* often hybridizes strongly to only one genomic fragment size. The letters above the lanes refer to the following taxa: A—*Oryza sativa* (rice); B—*Saccharum officinarum* (sugarcane); C—*Coix aquatica* Roxb.; D—*Coix lacryma-jobi*; E—*Sorghum bicolor* ssp. *bicolor*; F—*Sorghum bicolor* ssp. *arundinaceum* (Desv.) de Wet & Harlan; G—*Tripsacum dactyloides* (L.) L.; H—*Zea luxurians* (Durieu and Ascherson) R. Bird; and I—*Zea mays* ssp. *mays*.

revealed one strongly hybridizing fragment in most grasses, and no grass had more than three clearly hybridizing fragments (fig. 1). This result suggests that lambda cloning would likely yield orthologous sequences, especially since our size-selected libraries targeted the most intensely hybridizing fragments for those species with more than one fragment. Similarly, PCR-based isolation of *tb1*-like sequences would be unlikely to yield a complex mixture of paralogous sequences given the low copy number of *tb1* among the grasses. Nevertheless, given the dynamic nature of plant genomes, orthology cannot be assured on the basis of these data.

The sequence for the maize allele of *tb1* had been reported previously (Doebley, Stec, and Hubbard 1997). *tb1*-like sequences were isolated from an additional 27 members of the tribes Oryzeae, Andropogoneae, Arundinelleae, and Paniceae (table 1). *tb1*-like genes were first cloned from *O. sativa* (rice), *S. bicolor, Z. diploperennis,* and *Tripsacum cundinarmarce* using lambda libraries. The clones from all four taxa contained a *tb1*-like open reading frame. PCR primers were designed based on regions of high nucleotide similarity among these aligned sequences, and these primers were used to isolate *tb1*-like sequences from 23 additional taxa. Amplification reactions yielded a single product of the expected size in all cases. The lengths of *tb1*-like sequences varied between 1,023 and 1,167 bp, with a mean length of 1,066 bp. Excluding rice and six other sequences (see *Materials and Methods*), the remaining 21 sequences have a mean identity of 92.8% at the nucleotide level with a range of 84.2%–97.8% identity (table 2).

Despite the high similarity among *tb1*-like sequences, the alignment among all panicoid grasses revealed a remarkably variable structure with numerous indels. We identified a total of 91 indels between 3 and 24 nt in length. Most of these were microsatellite-like repeats of a codon such as $(GCC)_N$. Among all sequences, five

Table 2
Mean Identities and d_N/d_S Ratios for *tb1*-like Genes

Taxon[a]	No.[b]	Identity (%)	d_N	d_S	d_N/d_S	$d_N/d_S > 1.0$[c]
All	210	92.8 (84.2–97.8)	0.054 (0.012–0.138)	0.157 (0.046–0.386)	0.387 (0.128–1.259)	4
Cla.	20	93.5 (86.8–95.7)	0.041 (0.018–0.092)	0.171 (0.123–0.358)	0.242 (0.128–0.499)	0
Cpa	20	92.0 (84.2–94.4)	0.077 (0.053–0.138)	0.116 (0.053–0.326)	0.783 (0.422–1.259)	4
Iaf	20	92.8 (85.6–94.9)	0.065 (0.043–0.120)	0.118 (0.062–0.314)	0.608 (0.329–1.077)	1
Pvi.	20	92.2 (89.5–93.6)	0.050 (0.035–0.081)	0.202 (0.167–0.263)	0.250 (0.168–0.438)	0

NOTE.—Taxa with d_N/d_S ratios that were within the top or bottom 10% of all values are included in the table. The means and ranges (in parentheses) of values are given for all pairwise comparisons. *Oryza sativa* and six other taxa were not included (see *Materials and Methods*).

[a] Cla = *Coix lacryma-jobi*; Cpa = *Capillipedium parviflorum*; Isf = *Ischaemum afrum*; Pvi = *Panicum virgatum*.

[b] Number of pairwise tests whose results are reported in each row.

[c] Number of pairwise comparisons per row for which d_N/d_S was estimated to be greater than 1.0.

regions had a high level of variation in the number of microsatellite-like repeats. Residues 49–53 and 116–120 had 2–5 and 1–9 alanine repeats, respectively (fig. 2). Residues 211–213 had 1–6 repeats of glycine; residues 304–305 had 1–7 repeats of histidine; and residues 306–309 had 2–7 repeats of serine. Despite the large number of indels, not one indel altered the frame of the coding sequence.

Structure of *tb1*-like Genes

Sequence alignments identified three domains that are highly conserved and lack indels (fig. 2). One of these domains, the TCP domain, is predicted to form a noncanonical basic helix-loop-helix (bHLH) structure and was previously used to define the TCP gene family to which *tb1* belongs (Cubas et al. 1999). Among the panicoid grasses, the basic region of the TCP domain (residues 121–140) which includes a putative nuclear localization signal is more highly conserved than the HLH region (residues 146–179). The second conserved domain, the R domain, also contains several basic arginine and lysine residues that are shared among all panicoids (fig. 2). This domain was previously identified in both monocots and dicots (Cubas et al. 1999). We identified a third conserved domain among the grasses which had significant (E < 0.001) similarity only to *tb1* sequences in a BLAST 2.1 search of GenBank. This domain was not identified in other members of the TCP family (Cubas et al. 1999). This novel domain lies close to the amino terminus of the predicted protein and contains a large number of irregularly spaced proline and serine residues. Like the TCP and R domains, this domain can be readily aligned at the amino acid level between the panicoid grasses and rice (fig. 3), indicating its conservation across the grasses. We have termed this novel domain the SP domain because of the high frequency of these two amino acids.
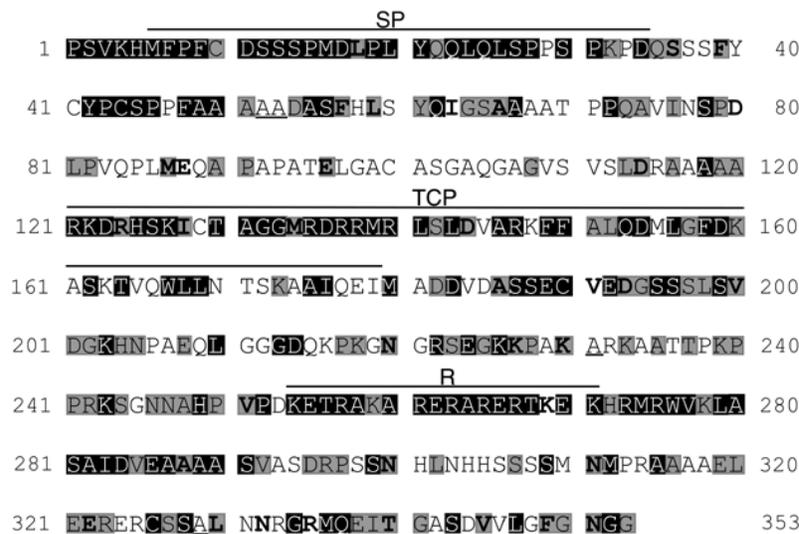

FIG. 2.—Majority-rule consensus of the predicted *tb1* amino acid sequences of 27 panicoid grasses. Identical amino acids are in black boxes. Sites which share >90% identity are in gray boxes. Residues that are shared by fewer than 50% of all taxa are underlined, and amino acids that are similar in charge or hydrophobicity are represented in bold. The novel SP domain (see text) and the previously identified TCP and R domains are identified.

```
                                          SP
                          _____
maize   1  HMFPFCDSSSPMDLPLYQQLQLSPSSPKTDQSSSFYCYPCSPPFAAADAS
rice       SMLPFFDSPSPMDIPLYQQLQLTPPSPKPDHHHHHHSTFFYYH-HHPPPS


maize  51  FPLSYQIGSAAAADATPPQAVINSPDLPVQALMDHAPAPATELGACASG-
rice       PSFPSFPSPAAATIASPSPAM-HPFMDLELEPHGQQLAAAEEDGAGGQGV

                                                        TCP
                                         _____
maize 101  AEGS--GASLDRAAAAARKDRHSKICTAGGMRDRRMRLSLDVARKFFALQ
rice       DAGVPFGVDGAAAAAAARKDRHSKISTAGGMRDRRMRLSLDVARKFFALQ

           _____
maize 151  DMLGFDKASKTVQWLLNTSKSAIQEIMADDASSECVEDGSSSLSVDGKHN
rice       DMLGFDKASKTVQWLLNMSKAAIREIMSDDASSVCEEDGSSSLSVDGKQQ


maize 201  -NPAEQLGGGGDQKPKGNCRG--EGKKPAKASKAAATPKPPRKSANNAHQ
rice       QHSNPADRGGGAGDHKGAAHGHSDGKKPAKPRRAAANPKPPRRLAN-AHP

                              R
                   _____
maize 251  VPDKETRAKARERARERTKEKHRMRWVKLASAIDVE----AAAASGPSDR
rice       VPDKESRAKARERARERTKEKNRMRWVTLASAISVEAATAAAAAGEDKSP


maize 301  PSSNNLSHHSSLSMNMPCAAAELEERERCSSALSNRSAGRMQEITGAS--
rice       TSPSNNLNHSSSTNLVSTELEDGSSSTRHNGVGVSGG--RMQEISAASEA


maize 351  -DVVLGFGNGG
rice       SDVIMAFANGG
```
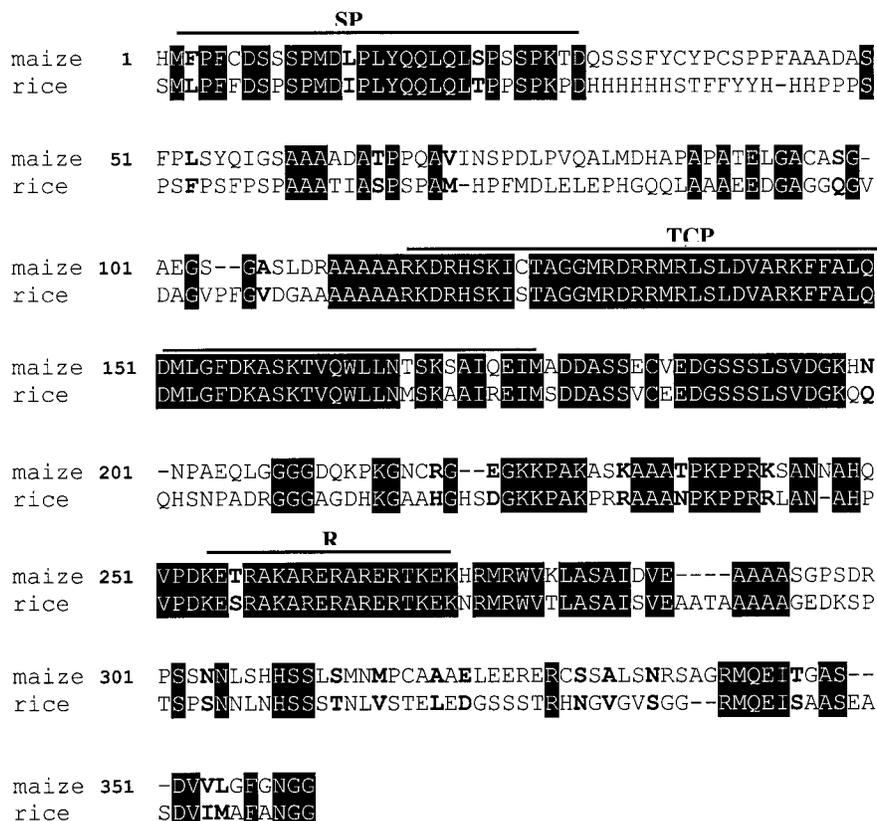
FIG. 3.—Alignment of the *tb1* predicted amino acid sequences for maize and rice. The SP, TCP, and R domains are identified. Identical amino acids are shaded, and amino acids of similar charges or hydrophobicities are shown in bold.

tb1 Phylogeny

We used likelihood and parsimony methods to construct the *tb1* gene phylogeny. To reduce inappropriate comparisons among nonhomologous nucleotide positions, regions of the gene that could not be unambiguously aligned were removed. The final data set had a total of 1,122 bp of aligned coding sequence including gaps. Of the 1,122 bp, 430 sites were variable and 211 were phylogenetically informative. Maximum-parsimony analysis resulted in three equally parsimonious trees on a single island of length 723 (consistency index [CI] = 0.76, retention index [RI] = 0.66). Heuristic searches using maximum-likelihood criteria described above generated one tree (ln $L = -5,595$) for each of 10 independent replicates with random input order of the taxa (fig. 4). More complex models with transition/transversion (ts/tv) frequencies estimated from the data (estimated ts/tv = 1.05) and with rate variation estimated among nucleotide sites (estimated gamma shape parameter alpha = 0.67) resulted in the same topology.

Parsimony and maximum-likelihood analyses produced different topologies; however, the areas of disagreement lacked statistical support. For example, in all parsimony trees, sequences from the genus Sorghum were polyphyletic, whereas Sorghum was monophyletic in the maximum-likelihood analysis. However, the maximum-likelihood phylogeny required only three more steps (726 vs. 723) than does the maximum parsimony tree, a difference that was not statistically significant (Templeton test; $P = 0.48$). Similarly, the most parsimonious phylogenies had a log likelihood of $-5,600$ which was not significantly different from the maximum-likelihood tree (ln $L = -5,595$) in a comparison using the Kishino-Hasegawa test ($P = 0.76$). The parsimony- and maximum-likelihood–generated trees were not significantly different despite different topologies because of weak support for internal nodes. All inferred phylogenies had short, internal branches with low bootstrap support.

Despite the poor resolution of the internal branches, several clades were well supported. The monophyly of the Andropogoneae was supported with 85% bootstrap support. Tripsacum and Zea formed a clade with 100% bootstrap support, as did Chionachne and Coelorachis, *S. australiense* and *Sorghum matarakense,* the two *S. bicolor* sequences, and Bothriochloa and Capillipedium. There was somewhat weaker support for the following groups: (1) a Chionachne/Coelorachis and Zea/Tripsacum clade (71%), (2) an Andropogon/Heteropogon clade (85%), and (3) a *Sorghum nitidum/S. bicolor* clade (74%).

The relationship of Coix to other Andropogoneae has been a question of some interest among grass systematists (Kellogg and Birchler 1993; Spangler et al. 1999). Coix, which is monoecious, has been placed with the other monoecious Andropogoneae (Zea, Tripsacum, and Chionachne) in a separate tribe (Maydeae) distinct from the Andropogoneae (Watson and Dallwitz 1992).
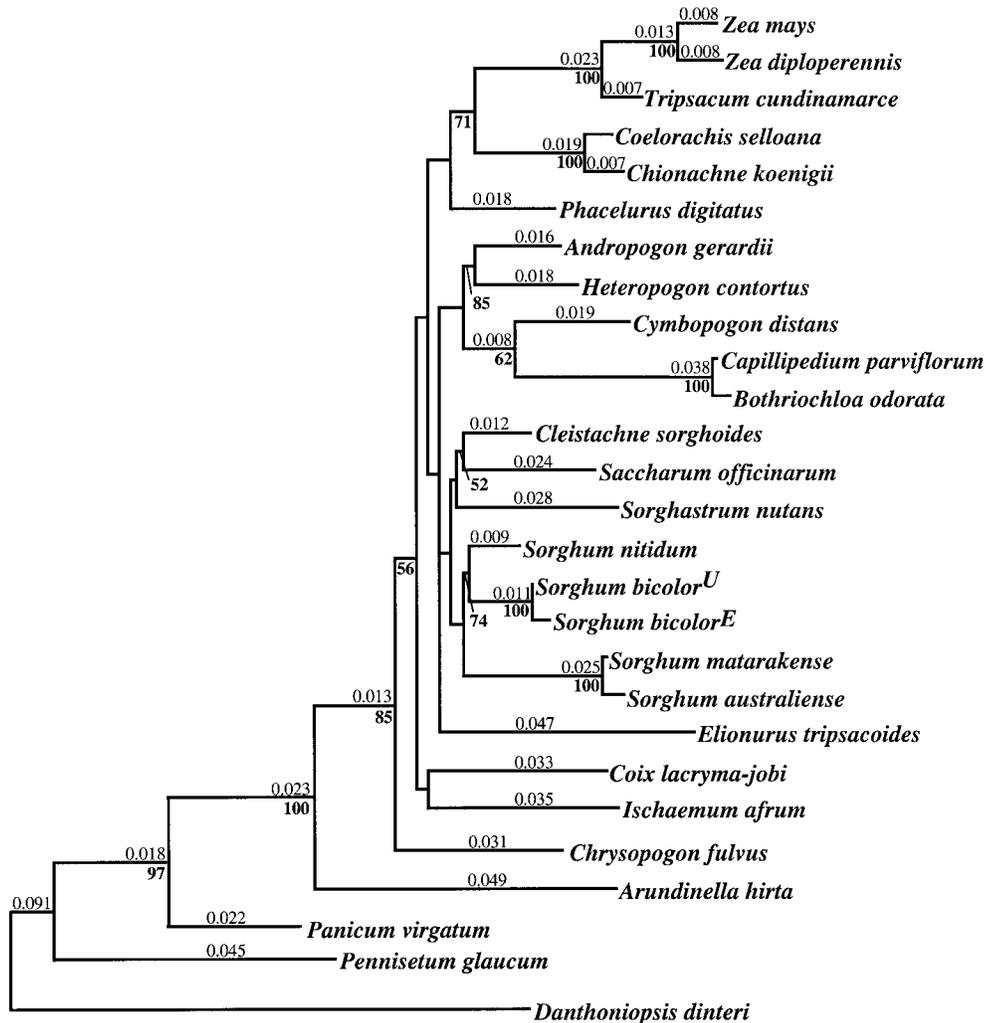
Fig. 4.—Maximum-likelihood phylogeny of *tb1*-like sequences. Branch lengths >0.005 substitutions per site are indicated above each branch. Bootstrap values are shown below a node if the node is present in more than 50% of the bootstrap replicate analyses. *Sorghum bicolor*[U] refers to the Ugandan isolate of *S. bicolor*. *Sorghum bicolor*[E] refers to the Ethiopian isolate.

In our phylogeny, Coix was sister to Ischaemum and well separated from the other monoecious taxa. Thus, our *tb1*-like data did not support a placement of Coix with the other monoecious species in the Maydeae. Indeed, trees in which the monoecious genera (including Coix) were forced to be monophyletic were 25 steps longer and statistically worse (Kishino-Hasegawa—$P < 0.0001$; Templeton—$P < 0.0001$) than the most parsimonious tree.

Testing for Positive Selection

Regulatory loci have been hypothesized to evolve quickly and acquire novel functions in their variable domains as a result of positive selection, while their conserved domains maintain basic functions such as DNA binding (Purugganan and Wessler 1994; Purugganan 1998). To detect the signature of such positive selection in *tb1*, a sliding-window analysis of mean pairwise d_N/d_S ratios was performed. Different regions of *tb1* exhibited greatly different substitution patterns, with the lowest ratios in the SP domain, the 5' region of the TCP

domain, and the R domain (fig. 5). The overall lowest mean d_N/d_S value was in the SP domain with an inferred ratio for nucleotides 1–60 of 4.13×10^{-4}. In contrast, the highest mean d_N/d_S ratios were observed in the regions immediately 5' (1.24) and 3' of the TCP domain (0.994).

A d_N/d_S value that is significantly greater than 1.0 is considered strong evidence for positive selection. Using the method of Nielsen and Yang (1998) to examine the entire *tb1* sequence, we tested for evidence of positive selection for nonsynonymous substitutions at some codons while allowing other codons to be invariant or neutral. This analysis placed codons into one of three categories: (1) fully neutral ($d_N/d_S = 1$); (2) invariant ($d_N/d_S = 0$); or (3) slightly deleterious, those with nonsynonymous substitutions occurring less frequently (per site) than synonymous substitutions ($d_N/d_S = 0.18$). This analysis provided no evidence of codons with d_N/d_S ratios greater than 1.0 and suggested that there was not a subset of codons within *tb1* under positive selection. To bias our analysis of positive selection toward quickly
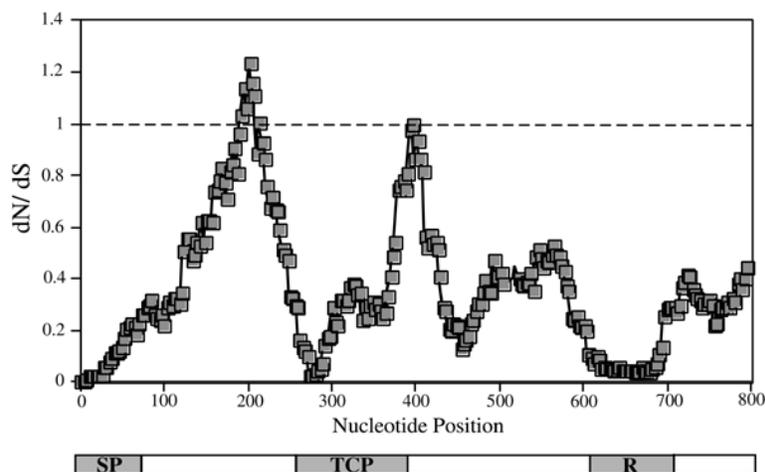
FIG. 5.—Sliding-window analysis of the estimated ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) along the *tb1* coding sequence. The mean d_N/d_S ratio for all pairwise sequence comparisons at each window is depicted as a square on the graph. The d_N/d_S estimate was computed for windows of 60 nt with 3-nt intervals between each pair of windows. Conserved domains are indicated below the graph.

evolving regions, we next analyzed only the most variable region (amino acids 83–119) of *tb1*. For some codons of this region, the nonsynonymous substitution rate per nonsynonymous site exceeded the synonymous substitution rate per synonymous site. Codons in this analysis were classified as either fully neutral ($d_N/d_S = 1$), deleterious ($d_N/d_S = 0$), or potentially under positive selection, with nonsynonymous substitutions occurring more frequently per nonsynonymous site than synonymous substitutions per synonymous site ($d_N/d_S = 4.18$). However, the log likelihood of this positive selection model (ln $L = -402.7$) did not significantly differ from the log likelihood of the simpler, strictly neutral model in which the d_N/d_S ratio was constrained to be either 1.0 or 0.0 (ln $L = -404.8$; $P < 0.12$).

We wished to test whether d_N/d_S varied among lineages as expected if selection intensities on *tb1* differed in some lineages relative to others. For the 21 species examined in the sliding-window analysis, the mean pairwise d_N/d_S ratio was 0.39 (table 2). However, some species appear to deviate greatly from this mean. *Capillipedium parviflorum* and *Ischaemum afrum* had high d_N/d_S ratios relative to the overall mean, while *Coix lacryma-jobi* and *Panicum virgatum* had lower ratios than other species (table 2). The significance of lineage-specific heterogeneity in d_N/d_S was tested by comparing the likelihood of a model in which d_N/d_S was held constant among all lineages with that of a model in which this ratio was allowed to vary among lineages (Yang and Nielsen 1998). The likelihood of these two models differed significantly ($P < 0.016$) with the model allowing the ratio to vary among lineages (ln $L = -3,797$) being favored over the model in which the ratio was constrained (ln $L = -3,824$). Although d_N/d_S ratios varied across lineages from 0 to 0.989 (with 4 of 35 lineages having no predicted synonymous substitutions), the estimated ratio of d_N/d_S for each lineage was still less than 1.0. Thus, the data were consistent with variation in the degree of selective constraint, and there is no need to invoke positive selection.

Rate Variation Between Lineages

To identify specific lineages for which either synonymous or nonsynonymous substitution rates were significantly accelerated or decelerated, the program Codrates was used (Muse and Gaut 1994). Substitution rate heterogeneity for nonsynonymous sites was far more prevalent than that for synonymous sites (fig. 6). Nonsynonymous substitution rates on lineages leading to Zea, Capillipedium, and Ischaemum appeared to be strongly accelerated. In contrast, *S. bicolor* and *S. nitidum* had slower rates of nonsynonymous substitution relative to other taxa. In this analysis, variation in the synonymous substitution rate was rare, and only *Z. mays* had significant synonymous substitution rate variation for 2 of 17 comparisons. For all comparisons, *Arundinella hirta* was used as the outgroup. Using different outgroups to compare rate variation, we found that the overall pattern of rate variation was maintained but that the significance of rate heterogeneity for individual comparisons often changed (data not shown). The fact that synonymous substitution rates did not differ greatly and that nonsynonymous rate variation was common suggests that *tb1*-like genes may be under different levels of functional constraint in different lineages.

Discussion
tb1 in the Grasses

tb1 belongs to the TCP family of putative transcriptional regulators. In both monocots and dicots, the TCP genes form a multigene family that includes two distinct groups: the *tb1-cyc*–like genes and the *pcf*-like genes. One concern in evolutionary analysis of genes belonging to such a multigene family is that one cannot be certain of the orthology of sequences. This can be a particular problem when performing PCR with conserved primers that may amplify nonorthologous gene family members. Two lines of evidence suggest that this is not an overriding problem with the *tb1*-like sequences that we isolated. First, Southern blot analysis of grass

	Zma	Cse	Pdi	Age	Hco	Cdi	Cpa	Sof	Snu	Sni	Sbi	Sma	Etr	Cso	Cla	Cfu	Iaf	Tcu
Zma	--		Zma**							Zma**	Zma**	Zma*		Zma*	Zma*			
Cse		--								Cse*	Cse*							
Pdi			--		Hco*		Cpa*										Iaf*	Tcu*
Age				--														
Hco					--					Hco**	Hco*							
Cdi						--	Cpa**										Iaf *	
Cpa	Zma*						--			Cpa**	Cpa**	Cpa*		Cpa*	Cpa*			
Sof								--		Sof*	Sof*							
Snu									--									
Sni										--			Etr*			Cfu*	Iaf **	Tcu**
Sbi											--						Iaf *	Tcu*
Sma	Zma*											--					Iaf *	
Etr													--					
Cso														--				
Cla															--			
Cfu																--		
Iaf																	--	
Tcu																		--

FIG. 6.—Results of 306 relative-rate tests for synonymous and nonsynonymous substitutions. The upper triangle shows the results of nonsynonymous substitution rate tests, and the lower triangle shows the results of synonymous substitution rate tests. Entries in the figure refer to those pairwise comparisons in which rate equality was rejected. The name of the sequence which was inferred to have evolved more quickly in each pairwise comparison is given as the first letter of the genus name and the first two letters of the species name. The Ugandan *Sorghum bicolor* sequence was used. *Arundinella hirta* was used as the outgroup sequence. One asterisk indicates rejection of rate equality at the 0.05 level; two asterisks indicate rejection at the 0.01 level.

genomic DNAs usually revealed only a single strongly hybridizing restriction fragment and no more than three hybridizing fragments (fig. 1). Second, the *tb1*-like gene phylogeny for the Andropogoneae was similar in several respects to phylogenies reported in previous studies of this tribe of grasses (Kellogg and Birchler 1993; Spangler et al. 1999). Nevertheless, strict orthology cannot be assured.

Structural Domains of *tb1*

The TCP gene family was named after its three founding members: *tb1, cyc,* and *pcf* (Cubas et al. 1999). These genes share the TCP domain, which forms a putative basic helix-loop-helix structure. The TCP domain of the rice PCF transcription factor has been demonstrated to function in DNA binding (Kosugi and Ohashi 1997). The basic region of the domain contains a putative bipartite nuclear localization signal. Additionally, some TCP family members, including *tb1* and *cyc,* share a second domain, termed the R domain, which has no known function but is predicted to form a hydrophilic alpha helix (Cubas et al. 1999).

We identified within the *tb1*-like genes among the grasses a third conserved region that we have named the SP domain. This new domain is near the amino terminus of the protein and is 29 amino acids in length. These 29 amino acids include five serine, five proline, and three glutamine residues. Eukaryotic transcription factors often have distinct activation domains that are characterized by multiple serine/threonine, proline, or glutamine residues (Seipel, Georgiev, and Schaffner 1992; Coustry, Maity, and de Crombrugghe 1995; Dörfler and Busslinger 1996). Thus, it is possible that the conserved SP region represents an activation domain. Functional analyses are necessary to test this putative function. For example, the fusion of the SP sequence with a DNA-binding domain such as the GAL4 DNA binding-domain from yeast should induce transcription of a reporter gene (such as β-galactosidase) flanked by the GAL4 upstream activating sequence.

The conservative nature of the TCP, R, and SP domains is apparent from several lines of evidence. First, comparison of the rice and maize proteins identified these three domains as highly conserved (fig. 3). Second, among the Andropogoneae, these same three domains stood out as conserved relative to the rest of the protein (fig. 2). Third, the sliding-window analysis showed that these three domains had the lowest d_N/d_S ratios (fig. 5). While each of these three domains is conserved relative to the rest of the protein, the HLH region of the TCP domain is somewhat less conserved than the basic region (figs. 2 and 5). The HLH region is likely involved in protein-protein interactions (Cubas et al. 1999), and heterogeneity within this region may indicate that *tb1* has acquired novel binding partners within different lineages. Such a pattern of substitution is similar to that found among homeobox proteins for which domains involved in protein : protein interaction are less conserved than DNA-binding domains (Sharkey, Graba, and Scott 1997).

Outside of their conserved domains, our *tb1* sequences have evolved unusually rapidly. *tb1* sequences had a greater number of indels and a higher d_N/d_S ratio relative to other plant genes (Wolfe, Sharp, and Li 1989; Purugganan 1998). The mean d_N/d_S ratio was 0.39 for all pairwise comparisons of *tb1*. In contrast, six genes compared between maize and wheat or barley had an average rate ratio of 0.08 (Wolfe, Sharp, and Li 1989). The d_N/d_S ratio for *tb1* was also higher than that of the duplicated *adh1* and *adh2* genes. Within the grasses, the maximum d_N/d_S value of the *adh* phylogeny was 0.36 for a single branch (Gaut et al. 1999). An additional level of structural heterogeneity among *tb1* sequences is caused by indels which are present even between taxa that have very few nucleotide differences. The *Z. mays* and *S. bicolor* sequences, for example, share 93.9% nucleotide identity but differ by 12 indels. It has been suggested that indels are important for creating novel variation within the gamete recognition protein bindin (Metz and Palumbi 1996).

The *tb1*-like Gene Phylogeny

The rapid evolution of *tb1* gives these sequences utility in testing phylogenetic hypotheses for the closely related taxa within the Andropogoneae. Although the tribe contains cosmopolitan genera such as Heteropogon and Andropogon, as well as crop plants such as maize, sorghum, and sugarcane, the relationships among taxa within the tribe have been difficult to determine. Early taxonomic studies and more recent analyses have suggested that the Andropogoneae underwent a rapid radiation (Celarier 1956; Spangler et al. 1999). Our phylogeny for *tb1* was consistent with this hypothesis, since most internal branches at the base of the Andropogoneae were short (fig. 4).

Several clades within the *tb1* phylogeny had reasonable bootstrap support and were consistent with morphological or previous phylogenetic analyses. The monophyly of the Andropogoneae with the Arundinelleae as its sister tribe agreed with several previous studies (Kellogg and Birchler 1993; Spangler et al. 1999). Tripsacum appears to be sister to Zea, as recognized over 60 years ago (Mangelsdorf and Reeves 1939). Coelorachis and Chionachne are sister to the Zea/Tripsacum clade, which seems reasonable given that Coelorachis has been proposed to be an intermediate between the monoecious Andropogoneae and other members of the tribe (Reeves and Mangelsdorf 1935). The monophyly of the Australian species *S. matarakense* and *S. australiense* and the close relationship between Bothriochloa and Capillipedium have also been reported elsewhere (de Wet and Harlan 1969; Sun et al. 1994; Spangler et al. 1999).

Two traditional taxonomic groups were not well supported by our data, and their monophyly has been previously challenged by other authors. First, our data argue strongly against the monophyly of the Maydeae (Chionachne, Coix, Tripsacum, and Zea). This conclusion is consistent with the ribosomal gene phylogeny of Buckler and Holtsford (1996). Accordingly, we suggest that this tribal designation be dropped from usage and its members recognized simply as belonging to the Andropogoneae, as previously proposed by Clayton (1987). Second, although Sorghum appears to be monophyletic in our maximum-likelihood tree, our data did not provide statistical support for the monophyly of Sorghum. Thus, our data do not argue against the growing body of evidence suggesting that Sorghum is polyphyletic (Duvall and Doebley 1990; Sun et al. 1994; Spangler et al. 1999).

Molecular sequence data for the chloroplast gene *ndh*F are the only sequence data that have been published to date from a large number of taxa within the Andropogoneae. Spangler et al.'s (1999) *ndh*F phylogenies and our *tb1* phylogeny both support many of the same relationships (see above). However, several specific areas of difference are noteworthy. In the *ndh*F phylogeny, the position of Coelorachis was unresolved; while in the *tb1* phylogeny, Coelorachis was strongly supported as the sister genus to Chionachne. Second, Elionurus was weakly supported as sister to the Trip- sacum/Zea clade in the in the *ndh*F phylogeny, but Elionurus was not sister to these taxa in the *tb1* phylogeny. Finally, Coix was sister to the Cymbopogon/Heteropogon/Andropogon clade in the *ndh*F phylogeny, but its position was unresolved in the *tb1* phylogeny.

No Evidence for Positive Selection

tb1 is a putative transcription factor, and genes of this class have been identified as potential key players in the evolution of new adaptations (Shubin, Tabin, and Carroll 1997; Doebley and Lukens 1998). The role of *tb1* in the evolution of maize (Wang et al. 1999) and the role of a *cyc*-like gene in the evolution of Linaria floral morphology (Cubas, Vincent, and Coen 1999) adds support to this general thesis and specifically identifies TCP genes as potential players in evolution and thus targets of positive Darwinian selection. In this regard, changes in *tb1* gene function over time may occur in several ways, two of which we consider here. First, evolution of *tb1* could involve changes in gene regulation, giving rise to new levels or spatial/temporal patterns of expression. Second, *tb1* evolution could involve changes in protein function such that the TB1 protein would interact with new protein partners or recognize new target sequences. We sought to test the latter of these possibilities by searching for evidence of past positive selection on the *tb1* protein among the andropogonoid grasses.

First, we looked for the signature of positive selection among different codons of *tb1* using the method of Nielsen and Yang (1998). For *tb1* overall, d_N/d_S was 0.39, which is far below the value of 1.0 expected for an unconstrained protein and even further below values such as 3.0 or greater seen in cases of clear positive selection (e.g., Hughes, Ota, and Nei 1990; Swanson and Vacquier 1995; Bishop, Dean, and Mitchell-Olds 2000). We further tested whether different codons within the gene had variable d_N/d_S rate ratios. This analysis indicated that the pattern of evolution within the gene was consistent with a nearly neutral model (Ohta 1992), with no evidence for positive selection. Finally, to bias our analysis in favor of finding evidence for positive selection, the Nielsen and Yang method was applied to the most variable region of the gene (5′ of the TCP domain). Here again, we could not reject the null hypothesis that codons within *tb1* were evolving in a neutral fashion among the Andropogoneae.

We performed a second set of analyses to test for lineage-specific positive selection using the method of Yang and Nielsen (1998). This involved comparing one model under which d_N/d_S was allowed to vary among lineages with a second model under which d_N/d_S was uniform across the phylogeny. In this analysis, there was evidence for variation in d_N/d_S among lineages, violating the expectations of the strictly neutral theory, but the largest value for any lineage was 0.99, which is within the range of values expected for proteins evolving in a nearly neutral fashion.

Both our lineage and our codon tests for positive selection failed to detect any deviation from nearly neu-

tral expectations. We consider two possible interpretations of these results. First, positive selection has indeed occurred, but our test simply lacked sufficient power to detect it. This interpretation cannot be excluded. Second, the TB1 protein has not experienced significant positive selection during the history of these grasses. If the latter is the case, this need not imply that the gene has not been affected by positive selection. Selection could act on 5′ regulatory sequences or other features that change the pattern/level of *tb1* expression. Indeed, in both of the two cases in which *tb1*-like genes have been implicated in the evolution of new morphologies, changes in gene expression were implicated (Cubas, Vincent, and Coen 1999; Wang et al. 1999).

Rate Heterogeneity Among Lineages

We wished to test whether *tb1* had evolved in a clocklike fashion along different lineages within the Andropogoneae, as expected under strict neutrality. Among the 153 tests of synonymous rate variation, only two were significant at the 5% level. One would expect more than seven significant tests by chance alone; thus, these results indicate that the synonymous rate does not vary among these grasses. Among the 153 tests for nonsynonymous rate variation, 30 were significant at either the 1% or the 5% level, indicating that there is extensive variation in the nonsynonymous rate. Examination of these tests shows that three lineages including *C. parviflorum, I. afrum,* and *Z. mays* have accelerated rates, while *S. bicolor* and *S. nitidum* have slower nonsynonymous substitution rates. It is of interest that substitution rate heterogeneity was detected for both synonymous and nonsynonymous sites in *Z. mays*. This result concurs with an earlier analysis of *Adh1,* for which Zea has an accelerated substitution rate relative to other grasses (Gaut and Clegg 1993).

What is the cause of the extensive variation in nonsynonymous rates among lineages? The tests of positive selection discussed above were all negative, and thus there is no evidence that this rate variation resulted from past episodes of positive selection. Excluding positive selection, there are at least two other possible explanations. First, there may have been a relaxation of constraint along some lineages, especially outside of the three conserved domains. Second, in small populations, nonsynonymous substitution rates may be accelerated due to the fixation of slightly deleterious alleles under random genetic drift (Ohta 1992, 1993). Unfortunately, the history of these grasses is too poorly known to make credible estimates of whether some have had historically smaller population sizes than others. One could test the hypothesis that population size has influenced the nucleotide substitution rate along these lineages, since such population size variation should alter the rates for all genes within a lineage in a similar way (Muse and Gaut 1997).

Acknowledgments

LITERATURE CITED

BISHOP, J. G., A. M. DEAN, and T. MITCHELL-OLDS. 2000. Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. Proc. Natl. Acad. Sci. USA **97**:5322–5327.

BUCKLER, E. S., and T. P. HOLTSFORD. 1996. *Zea* systematics: ribosomal ITS evidence. Mol. Biol. Evol. **13**:612–622.

CARROLL, S. B. 1994. Developmental regulatory mechanisms in the evolution of insect diversity. Dev. Suppl. **21**:7–223.

CELARIER, R. 1956. Cytotaxonomy of the Andropogoneae. I. subtribes Dimeriinae and Saccharinae. Cytologia **21**:272–291.

CLAYTON, W. D. 1987. Andropogoneae. Pp. 307–309 *in* T. SODERSTROM, K. HILU, C. CAMPBELL, and M. BARKWORTH, eds. Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

COUSTRY, F., S. MAITY, and B. DE CROMBRUGGHE. 1995. Studies on transcription activation by the multimeric CCAT-binding factor CBF*. J. Biol. Chem. **270**:468–475.

CUBAS, P., N. LAUTER, J. DOEBLEY, and E. COEN. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. Plant J. **18**:215–222.

CUBAS, P., C. VINCENT, and E. COEN. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401**:157–161.

DE WET, J. M. J., and J. R. HARLAN. 1969. Apomixis, polyploidy, and speciation in *Dicanthium*. Evolution **24**:270–277.

DOEBLEY, J., and L. LUKENS. 1998. Transcriptional regulators and the evolution of plant form. Plant Cell **10**:1075–1082.

DOEBLEY, J., and A. STEC. 1993. Inheritance of the morphological differences between maize and teosinte: comparison of the results of two F2 populations. Genetics **134**:559–570.

DOEBLEY, J., A. STEC, and C. GUSTUS. 1995. *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics **141**:333–346.

DOEBLEY, J., A. STEC, and L. HUBBARD. 1997. The evolution of apical dominance in maize. Nature **386**:485–488.

DÖRFLER, P., and M. BUSSLINGER. 1996. C-terminal activating and inhibitory domains determine the transactivation potential of BSAP (Pax-5), Pax-2, and Pax-8. EMBO J. **15**:1971–1982.

DUVALL, M., and J. DOEBLEY. 1990. Restriction site variation in the chloroplast genome of *Sorghum* (Poaceae). Syst. Bot. **15**:472–480.

GAUT, B. S., and M. T. CLEGG. 1993. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135**:1091–1097.

GAUT, B. S., A. S. PEEK, B. MORTON, and M. CLEGG. 1999. Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). Mol. Biol. Evol. **16**:1086–1097.

GOODRICH, J., R. CARPENTER, and E. COEN. 1992. A common gene regulates pigmentation pattern in diverse plant species. Cell **68**:955–964.

HUGHES, A., T. OTA, and M. NEI. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-

binding cleft of class I major histocompatibility complex molecules. Mol. Biol. Evol. **7**:515–524.

INA, Y., M. MIZOKAMI, K. OHBA, and T. GOJOBORI. 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. J. Mol. Evol. **38**:50–56.

KELLOGG, E., and J. BIRCHLER. 1993. Linking phylogeny and genetics: *Zea mays* as a tool for phylogenetic studies. Syst. Biol. **42**:415–439.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

KOSUGI, S., and Y. OHASHI. 1997. PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. Plant Cell **9**:1607–1619.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.0. Pennsylvania State University, University Park.

LUKENS, L., and J. DOEBLEY. 1999. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet. Res. **74**:291–302.

LUO, D., R. CARPENTER, C. VINCENT, L. COPSEY, and E. COEN. 1996. Origin of floral asymmetry in *Antirrhinum.* Nature **383**:794–799.

MANGELSDORF, P. C., and R. G. REEVES. 1939. The origin of maize. Proc. Natl. Acad. Sci. USA **24**:303–312.

MASON-GAMER, R., C. WEIL, and E. A. KELLOGG. 1998. Granule bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. **15**:1658–1673.

METZ, E., and S. PALUMBI. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol. Biol. Evol. **13**:397–406.

MUSE, S., and B. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11**:715–724.

———. 1997. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. Genetics **146**:393–399.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

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

OHTA, T. 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23**:263–286.

———. 1993. Amino acid substitution at the *Adh* locus of Drosophila is facilitated by small population size. Proc. Natl. Acad. Sci. USA **90**:4548–4551.

PURUGGANAN, M. 1998. The molecular evolution of development. BioEssays **20**:700–711.

PURUGGANAN, M., and J. SUDDITH. 1998. Molecular population genetics of the Arabidopsis *cauliflower* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95**:8130–8134.

PURUGGANAN, M., and S. WESSLER. 1994. Molecular evolution of the plant *R* regulatory gene family. Genetics **138**:849–854.

REEVES, R. G., and P. C. MANGELSDORF. 1935. Chromosome numbers in relatives of *Zea mays* L. Am. Nat. **69**:633–635.

RIMBAUT, A. 1996. Se-AL v1.d1: Sequence alignment program. Oxford University, Oxford, England.

SAMBROOK, J., E. M. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SEIPEL, K., O. GEORGIEV, and W. SCHAFFNER. 1992. Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J. **11**:4961–4968.

SHARKEY, M., Y. GRABA, and M. SCOTT. 1997. *Hox* genes in evolution: protein surfaces and paralog groups. Trends Genet. **14**:125–165.

SHUBIN, N., C. TABIN, and S. CARROLL. 1997. Fossils, genes and the evolution of animal limbs. Nature **388**:639–648.

SPANGLER, R., B. ZAITCHIK, E. RUSSO, and E. KELLOGG. 1999. Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndh*F sequences. Syst. Bot. **24**:267–281.

STERN, D. 1998. A role of *ultrabithorax* in morphological differences between Drosophila species. Nature **396**:463–466.

SUN, Y., D. Z. SKINNER, G. H. LANG, and S. H. HULBERT. 1994. Phylogenetic analysis of *Sorghum* and related taxa using transcribed spacers of nuclear ribosomal DNA. Theor. Appl. Genet. **89**:26–32.

SUTTON, K., and M. WILKINSON. 1997. Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45**:579–588.

SWANSON, W. J., and V. VACQUIER. 1995. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc. Natl. Acad. Sci. USA **92**:4957–4961.

SWOFFORD, D. L. 1999. PAUP*: phylogenetic analysis using parsimony. Beta test version 4.0 b2.0. Sinauer, Sunderland, Mass.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage sites with particular reference to the evolution of humans and the apes. Evolution **37**:221–244.

WANG, R. L., A. STEC, J. HEY, L. LUKENS, and J. DOEBLEY. 1999. The limits of selection during maize domestication. Nature **398**:236–239.

WATSON, L., and M. DALLWITZ. 1992. Grass genera of the world. C.A.B. International, Wallingford, England.

WOLFE, K., P. SHARP, and W. LI. 1989. Rates of synonymous substitutions in plant nuclear genes. Mol. Biol. Evol. **29**:208–211.

YANG, Z. 1998. Phylogenetic analysis by maximum likelihood (PAML). Version 1.4. University of California, Berkeley.

YANG, Z., and R. NIELSEN. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46**:409–418.