Heredity (2009) 102, 266–273

www.nature.com/hdy

ORIGINAL ARTICLE

Evolutionary fate of rhizome-specific genes in a non-rhizomatous *Sorghum* genotype

CS Jang[1,2], TL Kamps[1,3], H Tang[1], JE Bowers[1], C Lemke[1] and AH Paterson[1]
[1]*Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA and* [2]*Plant Genomics Lab, Department of Applied Plant Sciences Technology, Kangwon National University, Chuncheon, Korea*

What is the fate of organ-specific genes after the organ is lost? For *Sorghum propinquum* and *Sorghum halepense* genes that were previously shown to have rhizome-enriched expression, we have conducted comparative analysis of both coding regions and regulatory sequences in *Sorghum bicolor* (non-rhizomatousness) and *S. propinquum* (rhizomatousness). Most genes with rhizome-enriched expression appear to have similar numbers of paralogous copies in both genotypes, with only three of 24 genes studied showing significant differences in copy numbers. We detected no greater propensity for mutation in *S. bicolor* than in *S. propinquum* of genes with rhizome-enriched expression in the latter. Several *cis*-acting regulatory elements, particularly an Myb-binding core (AACGG) that is involved in the regulation of the mitotic cyclin, were more abundant in promoters of *S. propinquum* than in non-rhizomatous *S. bicolor* or *Oryza sativa* (rice). We suggest that many genes with rhizome-enriched expression in *S. propinquum* may serve multiple functions, with partial loss of some of these functions in *S. bicolor* but ongoing purifying selection acting to preserve the remaining functions. Expressed genes in polyploid *S. halepense* rhizomes appeared to be more frequently derived from the *S. propinquum* than the *S. bicolor* progenitor, but there was some evidence of formation of novel alleles and 'recruitment' of *S. bicolor* genes to rhizome-enriched expression in *S. halepense*, suggesting that polyploidy may have offered new evolutionary potential to *S. halepense*.
Heredity (2009) **102,** 266–273; doi:10.1038/hdy.2008.119; published online 12 November 2008

Keywords: evolutionary fate; polyploidy; rhizome; sorghum

Introduction

Rhizomes, modified subterranean stems that are diageotropic (for example, orient their growth perpendicular to the force of gravity), are organs of fundamental importance to plant competitiveness and invasiveness, playing two contrasting roles in agriculture. As a primary means of dispersal, rhizomes are an important component of 'weediness' of many of our most noxious weeds, including Johnsongrass (*Sorghum halepense* L. Pers.), Bermuda grass (*Cynodon dactylon* L. Pers.), purple nutsedge (*Cyperus rotundus*), quack grass (*Agropyron repens*) and cogon grass (*Imperata cylindrica*).

By contrast, rhizomes are also a valuable asset in establishment and persistence of dense, productive stands of forage and turfgrasses cultivated on more than 60 million acres in the southern United States alone (Burton, 1989), including *Cynodon* spp. (bermudagrass), *Paspalum* spp. (bahia and dallisgrass), *Pennisetum*/*Cenchrus* spp. (buffelgrass) and many others. The expansion of agriculture to provide plant biomass for production of fuels or chemical feedstocks will require greater utilization of marginal lands to make production of low per-unit value biomass economical.

Correspondence: Professor AH Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA.
E-mail: paterson@uga.edu
[3]*Current address: Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA.*
Received 6 May 2008; revised 14 October 2008; accepted 19 October 2008; published online 12 November 2008

The *Sorghum* genus has become a model for dissecting the molecular control of rhizomatousness (Paterson *et al*., 1995; Hu *et al*., 2003; Jang *et al*., 2006). *S. halepense* L. Pers. ($2N = 2X = 40$) (Johnsongrass) is one of the world's most noxious weeds (Holm *et al*., 1977), in part because it produces extensive rhizomes (subterranean stems that confer perenniality and also provide for clonal propagation) that make it difficult and expensive to eradicate. *S. halepense* is native to western Asia, but has been introduced and has naturalized in tropical and warm temperate climates worldwide (Holm *et al*., 1977). Cytological, morphological and molecular genetic data suggest that *S. halepense* is a naturally formed tetraploid hybrid derivative of *Sorghum bicolor*, an annual, polytypic African grass species which includes cultivated sorghum; and *Sorghum propinquum*, a perennial native to moist habitats in southeast Asia (Celarier, 1958; Doggett, 1976; Paterson *et al*., 1995).

Rhizomatousness appears to be ancestral within the *Saccharinae* clade. All members of the cultivated species, *S. bicolor*, are non-rhizomatous. However, close relatives are rhizomatous (*S. propinquum*), as is the ancestral form of a sister species, *Saccharum spontaneum*. These observations suggest that the loss of rhizomes in *S. bicolor* has been within the past ~1 million years since its divergence from a common ancestor shared with *S. propinquum* (Feltus *et al*., 2004). As all *S. bicolor* genotypes known, both wild and cultivated, are non-rhizomatous, the trait was presumably lost early in the radiation of *S. bicolor*.

What is the evolutionary fate of a gene that loses its organ? The evolutionary fate in non-rhizomatous

genotypes of genes that formerly contributed to rhizome development is of interest both from a basic and a practical standpoint. From a basic standpoint, rhizomes are an excellent example of a case of 'organ loss' that can be genetically manipulated, perhaps shedding light on basic principles that may apply to understanding of the evolution of morphology of other organisms, for example, the fate of tail-specific genes in humans. From an applied standpoint, better understanding of the fates of rhizome-specific genes would shed some light on alternative models for how rhizomatousness was lost. For example, the elimination of rhizomes by the progressive shutdown of many genes may show a very different 'signature' in its impact on variation of gene sequences between rhizomatous and non-rhizomatous genotypes than an abrupt macro-mutation in one or a small number of genes. Several lines of evidence indicated that an abrupt macro-mutation in one or a small number of regulatory genes was responsible for other striking morphological modifications during crop domestication. For instance, selection in the regulatory region of the *teosinte branched1* gene appears to have contributed substantially to the transformation of maize from the inflorescence morphology of the wild grass teosinte, with long branches with tassels (Wang *et al*., 1999; Clark *et al*., 2006), to the short branches typical of cultivated maize. Recently, Li *et al*. (2006) reported that reduced shattering of the mature inflorescence associated with rice domestication was caused in part by human selection of an amino acid substitution in the DNA-binding domain of the *sh4* gene. R*hz2* and *Rhz3* might be targets for such macro-mutations affecting rhizomatousness (Hu *et al*., 2003). However, little is known on the fate of many genes involved in the molecular pathway after the genotype has lost the ability of rhizomatousness.

The nature of tetraploid *S. halepense* raises questions about the relative roles of diploid *S. propinquum* and *S. bicolor* alleles in the growth and development of Johnsongrass rhizomes. Johnsongrass (*S. halepense*), with nearly worldwide distribution, is clearly more invasive than *S. propinquum*. Could this be partly due to recruitment of *S. bicolor* genes into rhizome development? To what degree has polyploid formation increased the potential for invasiveness of Johnsongrass?

Previously, we reported the functional classification, genomic organization, putative *cis*-acting regulatory elements, relationship to quantitative trait locus (QTL), of *Sorghum* genes with rhizome-enriched expression (Jang *et al*., 2006). However, the evolutionary fate in non-rhizomatous *S. bicolor* of genes with rhizome-enriched expression in its sister *S. propinquum* and possibly ancestral Saccharinae remains unknown. Herein, we have conducted comparative analysis of both coding regions and regulatory sequences for 54 rhizome tip (RT)-enriched genes isolated using bacterial artificial chromosome (BAC) libraries from *S. bicolor* and *S. propinquum*, seeking to shed new light on the evolution of rhizomatousness and the fates of genes whose organ is lost.

Materials and methods

Isolation of candidate BACs with rhizome-enriched genes

Previously, genes with rhizome-enriched expression were identified from both genotypes, *S. propinquum* and *S. halepense*, which produce abundant rhizomes (Jang *et al*., 2006). The 30 genes from *S. propinquum* and *S. halepense*, respectively, that showed the greatest enrichment of expression in the RT relative to mature rhizome internodes (RMIs) and pooled aboveground (AG) tissues were selected. Six clones that were enriched in RT relative to RMI were also RT-enriched relative to AG, resulting in a total of 54 clones that were further analyzed herein. Overgo probes were designed from the sequences of each of 54 RT-enriched genes by BLAST comparison against other plant species to identify the most-conserved 40 bp sequences as described (Bowers *et al*., 2005). For gene sequences with no matches to any other plant species, arbitrary 40 bp sequences were used as overgo probes. Individual overgo probes were radioactively labeled (Yuksel and Paterson, 2005) and then hybridized to seven BAC filters including 40 957 BACs of *S. propinquum* and 69 545 BACs of *S. bicolor* as described (Bowers *et al*., 2005).

Candidate BACs of both genotypes possessing the respective alleles of rhizome-enriched genes were selected based on the hybridization data, also using an FPC database (http://www.plantgenome.uga.edu/). To confirm that positive BAC clones contained the correct locus and allele, gene-specific primer sets were designed using the Primer3 program (Rozen and Skaletsky, 2000) from cDNA sequences of each gene (data not shown) and then were used for PCR with BAC DNAs as templates. The 25 μl PCR reaction mixtures contained 10 mM Tri-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM of each dNTP, 1.0 μM of each of gene specific primers, 0.5 U of *Taq* polymerase and 20 ng of each template DNA. The PCR program was as follows: 5 min at 94 °C, 35 cycles 1 min at 94 °C, 45–60 s at each annealing temperature and 1 min 72 °C with a 5 min final extension step at 72 °C. The PCR products were directly sequenced using each gene-specific primer after 15 min of incubation with 0.1 U of shrimp alkaline phosphatase (Roche, Basel, Switzerland) and 0.1 U of exonuclease I NEB; Ipswich, MA, USA) at 37 °C, followed by incubation at 80 °C.

Shotgun libraries

Each BAC clone was inoculated into a 3 ml culture of Luria–Bertani (LB) medium with 12.5 μg ml^{-1} chloramphenicol, shaking at 250 r.p.m. for 4 h at 37 °C. A total of 50 μl of culture were transferred into 50 ml of the same antibiotic LB medium and then grown for 14–16 h. The cell pellets were suspended after adding 2 ml of 10 mM ethylenediaminetetraacetic acid (EDTA). The solutions were added 4 ml of 0.2 N NaOH, 1% SDS immediately followed by keeping tubes on ice after adding 3 ml of 1.875 M potassium acetate, 11.5% acetic acid. After centrifuging for 15 min at 12 000 r.p.m., the solutions were filtered twice through Miracloth and 9 ml of cold isopropanol added for DNA precipitation. The DNA pellets were dissolved in TE (10 mM Tris-HCl pH 7.6, 50 mM EDTA), 1.15 ml of 7.5 M potassium acetate solution added, and frozen at −80 °C for 30 min. After ethanol precipitation, DNA pellets were dissolved in 700 μl of 50 mM Tris-HCl, 50 mM EDTA and then treated by 7 U of RNase A (Sigma, St Louis, MO, USA) for 1 h at 50 r.p.m. at 37 °C. The solutions were extracted twice with phenol and precipitated with isopropanol/ethanol followed by re-suspending in 40 μl of TE. BAC DNAs

were sheared by using a Hydroshear (Gene Machines Inc., Ann Arbor, MI, USA) with parameters as follows; 200 µl DNA volume, 20 cycles and speed code 12. The sheared DNAs were blunted-ended by using the End-It DNA End-Repair kit (Epicentre biotechnologies, Madison, WI, USA) in accordance with the manufacturer's instruction. The repaired DNAs were separated on a 1% agarose gel, sliced with size ranges of 3–4 kbp using a razor and extracted using a QIAEX II gel extraction kit (Qiagen, Valencia, CA, USA). The extracted DNAs were dephosphorylated by shrimp alkaline phosphatase (Roche) with incubation of 1 h at 37 °C. Ligation, transformation and blue/white screening were carried out using Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer's instruction.

Picking, hybridization and sequencing

A set of 768 clones derived from each BAC clone were picked using a QBOT (Genetix, New Milton, UK). Individual membranes containing 9216 clones with two replicates (that is, clones from six different BAC clones) were prepared as described by Jang *et al*. (2006). Probe labeling, hybridization and detection were conducted as described above. Twenty or fewer subclones per BAC clone were rearrayed into 96-well microtiter plates. Plasmid preparation and sequencing were performed by Jang *et al*. (2006). Trace files were processed using phred (score > 20), followed by phrap assembly into contigs by clustering a minimum continuous 100 bp (Ewing and Green, 1998; Ewing *et al*., 1998). Assembled sequences were visualized and manually edited using Consed (Gordon *et al*., 1998).

Sequence data analysis

Gene structures of the assembled sequences were predicted by either of two methods. BLAST analysis was performed with the rhizome-enriched cDNA sequences against the nonredundant protein databases. Sequences of cDNAs and deduced amino acids were aligned to corresponding genomic sequences, thereby predicting gene structures using the NAP (Huang and Zhang, 1996) and the GAP2 (Huang, 1994) programs. Alternatively, gene structures were predicted by FGENESH gene prediction software (http://sun1.softberry.com/berry.phtml) with the training set for monocot plants. Orthologs of *Oryza sativa* corresponding to rhizome-enriched genes were retrieved from rice pseudomolecules (Version 4) using BLASTx analysis ($E < e^{-25}$). The deduced amino-acid sequences of *S. propinquum*, *S. bicolor* and *O. sativa* corresponding to each of rhizome-enriched genes were aligned with the ClustalW program (http://www.ebi.ac.uk/Tools/clustalw/index.html) with default parameters, then manually edited using the BioEdit software (Hall, 1999). Synonymous and nonsynonymous substitutions per site (K_s and K_a) were measured by using the PAML package with the Nei–Gojobori method (Nei and Gojobori, 1986).

To uncover putative *cis*-acting regulatory elements located in the upstream regions of orthologs of rhizome-enriched genes, the identified 1-kb sequences were analyzed by the signal scan search in the PLACE (http://www.dna.affrc.go.jp/PLACE, Higo *et al*., 1999) database. The confidence limits for a binomial proportion ($P = 95\%$) were calculated according to standard methods (Snedecor and Cochran, 1980) and used to evaluate differences between orthologs in frequencies of *cis*-elements.

Results

BAC clones corresponding to RT-enriched genes from *S. propinquum* and *S. bicolor*

By hybridization of gene-specific overgos, we have identified the BAC clones and number of contigs (genetic loci) at which family members of each gene are represented in the physical maps of both *S. bicolor* and *S. propinquum* (Bowers *et al*., 2005). A total of 24 genes of *S. proqinquum* were anchored to only one locus, whereas 17 genes occurred at 2–10 loci and 9 genes at more than 10 loci. Four genes were not anchored to any locus, although several nonoverlapping overgoes were used for screening. For *S. bicolor*, a total of 26 genes were anchored to only one locus, 17 to 2–10 loci, 8 to more than 10 loci and 3 genes exhibited no anchor locus. The copy numbers for most RT-enriched genes were similar in *S. bicolor* and *S. propinquum* (Figure 1). One gene (RT/RMI19) of unknown function showed remarkably higher abundance in *S. bicolor* (282 loci) than *S. propinquum* (one locus). In contrast, two genes, RT/RMI26 and RT/RMI27, showed substantially higher abundance (38 and 96 loci) in *S. propinquum* than *S. bicolor* (3 and 1 loci, respectively).

Because most genes appear to be present as multiple paralogs in a plant genome, it can be difficult to determine whether truly orthologous loci are being compared between genotypes (Newbury and Paterson, 2003). In order to identify truly orthologous loci of each RT-enriched gene from the two species studied, we chose candidate BAC clones based on the following three criteria: (1) identification of linkage groups which candidate BACs anchored to, in both genomes; (2) synteny of genetic markers between candidate BAC clones in both genomes and (3) chromosomal location of best-hit rice orthologs corresponding to each of the RT-enriched genes. As shown in Supplementary Tables 1 and 2, a total of 20 out of 54 pairs were anchored to same

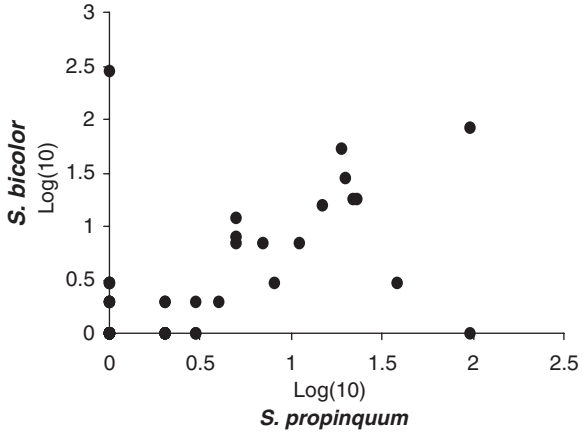

Figure 1 Comparison of putative copy numbers between genotypes, *S. propinquum* and *S. bicolor*. Copy numbers were estimated with the locus matched with more than two different BAC clones through FPC database when probing with single overgo.

linkage groups in both genomes whereas 18 pairs were not anchored to any linkage group of either one of both genomes. Three pairs were anchored to different linkage groups.

Genomic sequences corresponding to RT-enriched genes

The candidate BACs were sheared by a hydroshear and the fragments ranging in size of from 2 to 3 kbp were cloned. In order to get genomic sequences corresponding to RT-enriched genes, BACs were subcloned (see section 'Materials and methods') and the subclones screened for up to 20 matches with the overgoes used in screening of the candidate BAC clone. An average of 40 sequencing reactions (that is, two clones × 10-fold sequencing reactions × both directions) per clone were conducted, making it possible to assemble about 5–6 kbp of contiguous sequence, including most of each RT-enriched gene. However, genes with long transcribed regions, for example RT/AG01 (and RT/RMI01) oligosaccharyl transferase STT3 protein with the transcribed region of 5237 bp including 23 introns, required an additional cycle of hybridization with new overgoes designed near the 5′ proximity of the assembled region.

In RT-enriched genes obtained from RT/AG, full-length coding region sequences of 11 allele pairs of both genotypes were assembled, with the remainder not obtained in either one or both genotypes (Supplementary Table 1). Putative promoter sequences of 1 kbp or more of upstream sequence were obtained for eight gene pairs, whereas either one or both three genes pairs had less than 1 kbp of upstream sequences.

A total of 16 full-length coding regions corresponding to RT-enrich genes from the comparison of RT/RMI were developed from both genotypes (Supplementary Table 2). One kbp or more of upstream sequences were obtained from both genotypes for 12 of these, whereas the other four included less than 1 kbp upstream sequence from one or both.

Genome origin of RT-enriched genes in *S. halepense*

The genome organization of *S. halepense*, that is, a polyploid derived from interspecific hybridization of *S. bicolor* and *S. propinquum*, raises the question of whether there are striking differences in the abundance of transcripts from the respective diploids in *S. halepense* rhizomes. The respective diploid genomic DNAs corresponding to RT-enriched genes could allow us to determine the origins of transcripts in *S. halepense* rhizome. We used sequence-aligned scores produced by the ClustalW program (http://www.ebi.ac.uk/cgi-bin/clustalw2) and/or BLAST 2 sequences (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) as a criterion to define the origin of the RT-enriched transcripts. Among the 24 gene pairs for which we obtained full-length transcribed regions from both genomic DNAs (Table 1), 12 were screened with overgoes of transcripts in *S. halepense* rhizomes. Six transcripts (50%) in *S. halepense* rhizome were either identical (one cases) or significantly more similar (five, and the criteria for determining significance) to the genomic DNAs derived from *S. propinquum* than *S. bicolor*. One of the remaining six appeared to be a possible 'hybrid' transcript, more closely matching *S. propinquum* along the N-terminal part and *S. bicolor* along the C-terminal part although each of these matches was imperfect. The origin of the remainder could not be determined, either due to the identity of *S. propinquum* and *S. bicolor* DNAs (1 case), or to similar degrees of divergence from both. By contrast, most transcripts from *S. propinquum* rhizomes exhibited identity (100%) to its genomic DNAs.

Evolutionary fate of RT-enriched genes in *S. bicolor*

The comparison of *S. bicolor* alleles of RT-enriched genes to those of *S. propinquum* might allow us to better understand the means by which rhizomatousness has been naturally shut off. Out of 24 RT-enriched genes for which their coding and upstream sequences were completely assembled, we detected no propensity for mutation in coding regions (for example, premature stop codon and striking amino-acid changes) in the non-rhizomatousness genotype as compared with those of the rhizomatousness genotype.

We then evaluated the mode of selection acting on RT-enriched genes in *S. bicolor* by the calculation of K_a/K_s values as the divergence of *S. bicolor* and *S. propinquum* (Table 2). Contrary to our expectations, RT/AG-enriched transcripts showed K_a/K_s values of 0.0000–0.3678, suggesting purifying selection. Genetic distances between rice–sorghum orthologs corresponding to RT-enriched genes was investigated to see if either branch (*S. bicolor* or *S. propinquum*) evolved faster. We detected no evidence of asymmetric evolution between *O. sativa*—*S. bicolor* and *O. sativa*—*S. propinquum* (Table 2).

Similarly, for RT/RMI-enriched genes, a range of K_a/K_s values of 0.0000–0.5565 suggested purifying selection after divergence of *S. bicolor*–*S. propinquum*. Again, K_a/K_s values showed no significant differences between *O. sativa*—*S. bicolor* and *O. sativa*—*S. propinquum*, reflecting the symmetric evolution of *S. bicolor* and *S. propinquum* alleles. Three RT/RMI-enriched genes, all of unknown function, exhibited no correspondence to any rice gene.

Orthologs corresponding to rhizome-enriched genes in other taxa

BLAST *n* analysis ($E<e^{-25}$) were performed with coding sequences of RT-enriched genes against expressed sequence tag (EST) databases of each of five major crops including *Sorghum*, *Saccharum*, *Zea*, *Triticum*, *Hordeum* and *Oryza* as well as *Arabidopsis* as an outgroup. Of 24 RT-enriched genes, expressed sequence tag (EST) frequencies of 21 genes showed no obviously biases over the five major crops, showing their orthologs in most of the taxa. However, there is rare or none of EST frequency in *Arabidopsis* with the exception of RT/RMI02 (0.021%). Curiously, three genes, that is, RT/RMI23, RT/RMI24 and RT/RMI26, each of which exhibited no correspondence to the rice genome, evidenced highest frequencies abundance in *Sorghum* and were also found in closely related *Saccharum* and *Zea* but were absent from cDNA resources for the remaining, more distantly related taxa (Figure 2).

Comparison of putative *cis*-acting regulatory element

The discovery that changes in the promoter regions of genes related to domestication may be more important than changes in the coding regions (Wang *et al*., 1999) raised the question of whether changes in promoters of many genes related to rhizomatousness might be found.

Table 1 Genome origins of transcripts in *S. halepense* rhizomes

No. of genes[a]	*No. of BAC clones*[b]	*cDNA origin*[c]		*Similarity of S. halepense cDNA*[d]		
		S. propinquum	*S. halepense*	*S. propinquum*	*S. bicolor*	*ND*[e]
54	24	12	12	6	0	6

[a]Indicated diversifying selection genes that were RT-enriched genes in both *S. propinquum* and *S. halepense*, that are rhizomatous, described as cited.
[b]Gene pairs for which we obtained full-length transcribed regions from both genomic DNAs.
[c]Indicated transcript origins of the overgo probes utilized in BAC library screening to identify full-length transcribed regions.
[d]Similarity of cDNA sequences from *S. halepense* to its respective diploid progenitors (*S. propinquum* and *S. bicolor*) was evaluated by sequence alignments using ClustalW and/or BLAST 2 sequences.
[e]Indicated that similarity could not be determined, as described in text.

Table 2 Evaluation of K_a/K_s value between *S. propinquum* genes with rhizome-enriched expression patterns and their orthologous genes of *S. bicolor* or *O. sativa*

Clone ID	*S. propinquum vs S. bicolor*			*S. propinquum vs O. sativa*			*S. bicolor vs O. sativa*		
	K_a	K_s	K_a/K_s	K_a	K_s	K_a/K_s	K_a	K_s	K_a/K_s
High RT/AG group									
RT/AG01	0.0000	0.0105	0.0000	0.0241	0.4313	0.0559	0.0241	0.4309	0.0560
RT/AG02	0.0000	0.0707	0.0000	0.0113	0.3738	0.0301	0.0113	0.4120	0.0273
RT/AG05	0.0010	0.0144	0.0679	0.2675	0.7202	0.3713	0.2676	0.7115	0.3761
RT/AG06	0.0023	0.0068	0.3361	0.1501	0.4625	0.3245	0.1501	0.4680	0.3208
RT/AG07	0.0219	0.0595	0.3678	0.1859	0.4231	0.4393	0.1785	0.4132	0.4319
RT/AG10	0.0030	0.0203	0.1490	0.0750	0.4987	0.1503	0.0739	0.4820	0.1532
RT/AG11	0.0029	0.0203	0.1407	0.0846	0.6690	0.1264	0.0824	0.6554	0.1258
RT/AG16	0.0010	0.0266	0.0391	0.0516	0.5870	0.0880	0.0522	0.5628	0.0927
RT/AG19	0.0013	0.0240	0.0525	0.3319	0.6701	0.4952	0.3324	0.6686	0.4971
RT/AG20	0.0000	0.0000	NA	0.1021	0.3145	0.3247	0.1021	0.3145	0.3247
RT/AG22	0.0026	0.0161	0.1585	0.1469	0.3489	0.4209	0.1462	0.3516	0.4159
RT/AG25	0.0097	0.0423	0.2301	0.0744	1.0657	0.0699	0.0745	1.0789	0.0690
High RT/RMI group									
RT/RMI01	0.0000	0.0105	0.0000	0.0241	0.4313	0.0559	0.0241	0.4309	0.0560
RT/RMI02	0.0000	0.0707	0.0000	0.0113	0.3738	0.0301	0.0113	0.4120	0.0273
RT/RMI03	0.0060	0.0114	0.5213	0.0721	0.7505	0.0961	0.0682	0.7310	0.0933
RT/RMI04	0.0023	0.0068	0.3361	0.1501	0.4625	0.3245	0.1501	0.4680	0.3208
RT/RMI06	0.0010	0.0266	0.0391	0.0516	0.5870	0.0880	0.0522	0.5628	0.0927
RT/RMI10	0.0024	0.0233	0.1073	0.1603	0.7085	0.2263	0.1612	0.6874	0.2345
RT/RMI11	0.0093	0.0349	0.2661	0.0989	0.7297	0.1355	0.0926	0.6903	0.1342
RT/RMI16	0.0030	0.0618	0.0493	0.2219	0.4475	0.4960	0.2188	0.4586	0.4771
RT/RMI18	0.0214	0.0546	0.2274	0.1873	0.6438	0.2909	0.1840	0.6315	0.2914
RT/RMI20	0.0000	0.0094	0.0000	0.2122	0.3329	0.6374	0.2122	0.3185	0.6661
RT/RMI23	0.0163	0.0294	0.5565	—	—	ND	—	—	ND
RT/RMI24	0.0063	0.0349	0.1813	—	—	ND	—	—	ND
RT/RMI26	0.0052	0.0188	0.3121	—	—	ND	—	—	ND
RT/RMI27	0.0000	0.0044	0.0000	0.0262	0.5112	0.0513	0.0262	0.5114	0.0513
RT/RMI28	0.0037	0.0148	0.2944	0.5030	0.5854	0.2971	0.2971	0.5091	0.5835
RT/RMI30	0.0017	0.0163	0.1037	0.2573	0.5220	0.4929	0.2598	0.5541	0.4689

Abbreviations: AG, aboveground; NA, not applicable; ND, not determined; RMI, rhizome internode; RT, rhizome tip.

To evaluate putative *cis*-acting regulatory elements, upstream regions of *S. bicolor* and *S. propinquum* genes were analyzed using the PLACE database. The upstream regions of 21 rice orthologs of RT-enriched genes were also retrieved from rice pseudomolecules and used as another control set due to the non-rhizomatous nature of *O. sativa*. Comparison of these regions in this sample of rhizome-enriched genes permits us to infer changes in upstream features associated with the general class of rhizome-enriched genes since the divergence of *S. bicolor* from *S. propinquum* or since the divergence of either of these from rice.

One *cis*-acting regulatory element, an Myb-binding core (AACGG) which is involved in the regulation of the mitotic cyclin, especially as an activator element, found in the promoter of the *Arabidopsis thaliana* cyclin B1 gene (Planchais *et al*., 2002), was significantly more abundant in promoters of *S. propinquum* alleles (70.8%) than those of *S. bicolor* (54.2%) or *O. sativa* (47.6%). Five additional *cis*-acting regulatory elements, such as CRT/DRE motif (Xue, 2003), $(CA)_n$ element in storage protein genes (Ellerstrom *et al*., 1996), TATA box (Grace *et al*., 2004), the CCA1-binding element related to regulation by phytochrome (Wang *et al*., 1997), and pyrimidine box required for gibberellic acid (GA) induction (Cercos *et al*., 1999), were also more abundant in *S. propinquum* than either *S. bicolor* or *O. sativa* alleles (Table 3).

Several *cis*-acting regulatory elements were more abundant in promoters of *S. bicolor* than in *S. propinquum* or *O. sativa* (Table 3). One motif found in promoters of

anaerobic genes (AAACAAA; Mohanty *et al*., 2005) showed significantly higher abundance in *S. bicolor* than the other species. The promoters of *S. bicolor* genes were enriched relative to those of *O. sativa* for three additional *cis*-elements, −300 element (Thomas and Flavell, 1990), −10 promoter element (Thum *et al*., 2001) and TATA box (Grace *et al*., 2004). By contrast, two promoters, polyA signal (O'Nell *et al*., 1990) and pro- or hypo-osmolarity element found in the promoter of proline dehydrogenase (Satoh *et al*., 2002) were more abundant in the promoters of *S. bicolor* than *S. propinquum* alleles.

Discussion

Although mutations of a tiny population of major regulatory gene(s) followed by human selection are well known to be responsible for crop morphological modification during domestication, little is known about the evolutionary fate of additional genes which might be affected by such morphological modifications. We found that the loss of rhizomes in the lineage leading to *S. bicolor* has had very little effect on genes that show rhizome-enriched expression in *S. propinquum*. Indeed, these genes continue to show evidence of purifying selection since the *S. bicolor—S. propinquum* divergence, suggesting that the elimination of rhizomes has not been by the progressive shutdown of many genes.

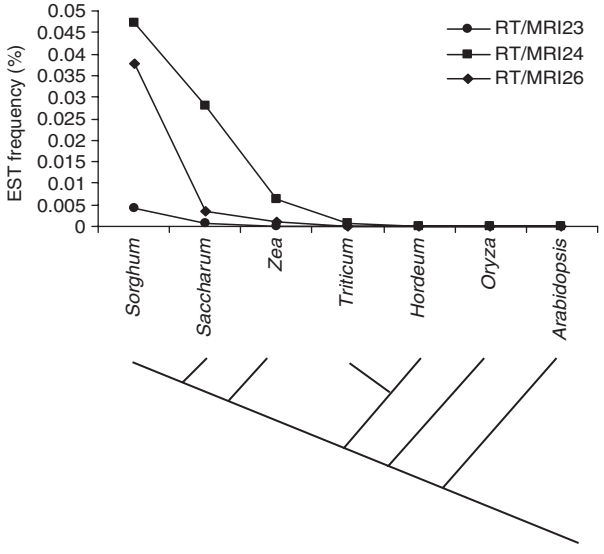

Figure 2 EST frequencies of homologs of three rhizome tip (RT)-enriched genes (RT/RMI23, RT/RMI24 and RT/RMI26), in five major crop plants and *Arabidopsis*.

Some hypotheses might be able to explain our finding that most rhizome-enriched genes have been not mutated in the genotype (for example, *S. bicolor*) that has lost the ability to make rhizomes. The small K_a/K_s ratios observed for the 24 genes suggests that they remain under purifying selection. Indeed, not a single gene among the 24 showed $K_a/K_s>1$, which would be suggestive of diversifying selection; or even $K_a/K_s\sim1$, suggestive of a lack of functional constraint (pseudogenization). An attractive hypothesis is that the RT-enriched genes may serve multiple functions during growth and development of plants, some of which are not in rhizomes. Regulatory, rather than structural, mutations might have been responsible for the loss of rhizomatousness in *S. bicolor*, as suggested by significant differences of one type of *cis*-element between *S. propinquum* and *S. bicolor* as well as *O. sativa*. Planchais *et al*. (2002) reported that one promoter region including the Myb-binding core (AACGG) took part in the cell-cycle-dependent transcriptional regulation of the *A. thaliana* cyclin B1 gene. The *CycB1* gene is localized to lateral root primodia, the base of the first leaf promodium and the shoot meristem, suggesting that its accumulation might be one of the limiting factors for the activation of cell division (Ferreira *et al*., 1994). Significant differences in abundance of the Myb core sequence between the promoters of *S. propinquum* and *S. bicolor* (or *O. sativa*) might reflect loss of one or more subfunctions of RT-enriched alleles in *S. bicolor* due to loss of regulatory motifs important to rhizome-specificity. The core of the $(CA)_n$ element, required for storage organ-specific

Table 3 Summary of selected *cis*-acting regulatory elements located on putative promoter sequences of genes with rhizome-enriched expression patterns and their rice orthologs

		S. propinquum	*S. bicolor*	*O. sativa*
No. of tested clones		24	24	21
Total promoter length (bp)		20749	20749	17749
Cis-elements enriched in S. propinquum				
Myb core	AACGG	70.8 ± 18.2[a]	54.2 ± 19.9	47.6 ± 16.7
CRT/DRE motif	GTCGAC	33.3 ± 18.9	16.7 ± 14.9	19.3 ± 16.7
$(CA)_n$ element	CNAACAC	45.8 ± 19.9	29.2 ± 18.1	21.2 ± 16.7
TATA box	TATATAA	58.3 ± 19.7	45.8 ± 19.9	14.3 ± 15.0
CCA1 binding	AAMAATCT	16.7 ± 14.9	4.2 ± 8.0	9.5 ± 12.6
Pyrimidine box	TTTTTTCC	16.7 ± 14.9	4.2 ± 8.0	2.9 ± 19.3
Cis-elements enriched in S. bicolor				
−300 ELEMENT	TGHAAARK	50.0 ± 20.0	62.5 ± 19.4	38.1 ± 20.1
−10 promoter element	TATTCT	33.3 ± 18.9	45.8 ± 19.9	23.8 ± 18.2
TATA box	TATAAAT	37.5 ± 19.4	54.2 ± 19.9	33.3 ± 20.2
PolyA signal	AATTAAA	12.5 ± 13.2	29.2 ± 18.2	23.8 ± 18.2
PRE	ACTCAT	20.8 ± 16.2	41.7 ± 19.7	28.5 ± 19.3
Anaerobic set	AAACAAA	45.8 ± 19.9	70.8 ± 18.2	42.9 ± 21.2

[a]Percentage of the indicated element family found per putative promoter region ± 95% confidence limits for *P*.

transcription (Ellerstrom *et al*., 1996), also showed lower frequencies in the promoter regions of *S. bicolor* and *O. sativa* than *S. propinquum*.

Another possibility for partial loss of subfunctions of the RT-enriched genes in domesticated genotypes would be that an alternative splice form with rhizome-specificity is lost but the other form(s) still remain due to their importance in other organs. For example, a gene-encoding oligosaccharyl transferase STT3 protein with the highest degree of rhizome-specific expression (both RT/AG and RT/RMI) and located near a QTL which contributes to rhizome length and number, rhizome branching, and RMI number and length (Jang *et al*., 2006) evidenced the longest transcribed region with as many as 23 introns. In its rice ortholog (Os05g44360), two different splice isoforms have been reported (http://www.tigr.org/tigr-scripts/euk_manatee/shared/ORF_infopage.cgi).

Another alternative hypothesis is that rhizome-enriched overexpression of these genes is a 'mistake', that is, that offers no particular fitness advantages to the plant, much like expression of some retroelements (for example, Langille and Clark, 2007), but that their expression at low levels somewhere else in the plant is important.

The aggressive rhizomatousness and widespread international distribution of tetraploid *S. halepense* raised interesting questions—that is, whether alleles from non-rhizomatous *S. bicolor* may have played any role in the aggressive rhizomatousness of *S. halepense*. It is believed that polyploidization can lead to extensive effects on gene expression, as detailed above. Although there is greater abundance of transcripts from *S. propinquum* (50%) than *S. bicolor* in *S. halepense* rhizomes, we do not yet have sufficient information to define the remainders of transcripts of *S. halepense* rhizomes (50%). It remains to be determined, for example, whether different alleles from the respective diploids might have interacted to form a new allele, which could lead to regaining of function even if the *S. bicolor* portion had been nonfunctionalized (Wang *et al*., 2007).

Most RT-enriched genes (51 of 54 genes tested) showed no significant differences in gene copy numbers between *S. propinquum* and *S. bicolor*, with a few exceptions. Three RT-enriched genes did show significant differences in numbers of gene copies, suggesting gene amplification following speciation. Naito *et al*. (2006) reported dramatic amplification of a rice-transposable element designed as *mPing* during domestication, suggesting that the rapid increase represents a potentially valuable source of population diversity. Although the nature of the three genes that are rapidly evolving in copy number is not yet clear, ongoing analysis of the 289 family members of RT/RMI19 in the recently completed genome sequence of *S. bicolor* might provide clues to shed further light on genome evolution following speciation.

The findings that three genes are relatively abundant in the sorghum EST database, and also present in other members of the *Andropogoneae* tribe but not in those of more distant plant taxa (Figure 2) might reflect more gradual amplification of these genes across many millions of years.

Previously, several lines of evidence pointed to GAs as probable key regulators of rhizome gene expression and development (Jang *et al*., 2006). In particular, three *cis*-acting elements related to GA responses were enriched in abundance in the putative promoter regions of rice gene models corresponding to sorghum genes with rhizome-enriched expression. One of these *cis*-elements (TTTTTTCC), the pyrimidine box for GA induction (Cercos *et al*., 1999), also showed higher frequency in the promoter regions of *S. proqinquum* alleles than those of *S. bicolor* or *O. sativa*, suggesting that they could contribute to the difference between these genotypes in the degree of rhizomatousness. Three *cis*-elements found in the previous report exhibited no significant differences in frequencies between the promoter regions of *S. bicolor* and *S. propinquum* and therefore are not likely to contribute to genetic differences between the two, but may still function in rhizome development and/or is important in GA regulation of other plant parts after loss of rhizomes in *S. bicolor*.

Acknowledgements

This work was supported in part by the USDA 'Biology of Weedy and Invasive Species' program (01-35320-10964 to AHP) and NSF Plant Genome Research Program (DBI-9872649; 0115903 to AHP) and the Korean government (MOEHRD, Basic Research Promotion fund; Korea Research Foundation grant no. KRF-2004-214-M01-2004-000-10060-0 to CSJ).

References

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA *et al.* (2005). Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proc Natl Acad Sci USA* **102**: 13206–13211.

Burton GW (1989). Progress and benefits to humanity from breeding warm-season forage grasses. In: Sleper DA, Asay KH, Pedersen JF (eds). *Contributions from Breeding Forages and Turf Grasses*. Crop Science Society of America: Madison, WI. pp 21–29.

Celarier RP (1958). Cytotaxonomic notes on the subsection halepensia of the genus *Sorghum* Bull Torrey Bot. *Club* **85**: 49–62.

Cercos M, Gomez-Cadenas A, Ho THD (1999). Hormonal regulation of a cysteine proteinase gene, *EPB-1*, in barley aleurone layers: *cis*- and *trans*-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. *Plant J* **19**: 107–118.

Clark RM, Wagler TN, Quijada P, Doebley J (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* **38**: 594–597.

Doggett H (1976). Sorghum. In: Simmonds NW (ed). *Evolution of Crop Plants*. Longman: Essex, UK. pp 112–117.

Ellerstrom M, Stalberg K, Ezcurra I, Rask L (1996). Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. *Plant Mol Biol* **32**: 1019–1027.

Ewing B, Green P (1998). Base-calling of automated sequencer traces using phred II error probability. *Genome Res* **8**: 175–185.

Ewing B, Hiller L, Wendl MC, Green P (1998). Base-calling of automated sequencer traces using phred I accuracy assessment. *Genome Res* **8**: 175–185.

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004). An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Res* **14**: 1812–1819.

Ferreira PCG, Hemerly AS, de Almedia Engler J, Van Montagu M, Engler G, Inzé D (1994). Developmental expression of the Arabiodpsis cycline gene *cyc1At*. *Plant Cell* **6**: 1763–1774.

Gordon D, Abajian C, Green P (1998). Consed: a graphical toll for sequence finishing. *Genome Res* **8**: 195–202.

Grace ML, Chandrasekharan MB, Hall TC, Crowe AJ (2004). Sequence and spacing of TATA box elements are critical for accurate initiation from the beta-phaseolin promoter. *J Biol Chem* **279**: 8102–8110.

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

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* **27**: 297–300.

Holm LG, Plucknett DL, Pancho JV, Herberger JP (eds) (1977). *Sorghum halepense* (L.) Pers. In: *The World's Worst Weeds: Distribution and Biology*. University Press of Hawaii: Honolulu, Hawaii. pp 54–61.

Hu FY, Tao DY, Sacks E, Fu BF, Xu P, Li J *et al*. (2003). Convergent evolution of perennially in rice and sorghum. *Proc Natl Acad Sci USA* **100**: 4050–4054.

Huang X (1994). On global sequence alignment. *Comput Appl Biosci* **10**: 227–235.

Huang X, Zhang J (1996). Methods for comparing a DNA sequence with a protein sequence. *Comput Appl Biosci* **12**: 497–506.

Jang CS, Kamps TL, Skinner DN, Schulze SR, Vencill WK, Paterson AH (2006). Functional classification, genomic organization putatively *cis*-acting regulatory elements, and relationship to quantitative trait loci, of *Sorghum* genes with rhizome-enriched expression. *Plant Physiol* **142**: 1148–1159.

Langille MGI, Clark DV (2007). Parent genes of retrotransposition-generated duplicates in *Drosophila melanogaster* have distinct expression profiles. *Genomics* **90**: 334–343.

Li C, Zhou A, Sang T (2006). Rice domestication by reducing shattering. *Science* **311**: 1936–1939.

Mohanty B, Krishnan SP, Swarup S, Bajic VB (2005). Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. *Ann Bot* **96**: 669–681.

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y *et al*. (2006). Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci USA* **103**: 17520–17625.

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

Newbury HJ, Paterson AH (2003). Genomic colinearity and its application in crop plant improvement. In: Newbury HJ (ed). *Plant Molecular Breeding*. Blackwell Publishing: Oxford, UK. pp 60–82.

O'Nell SD, Kumagai MH, Majumdar A, Huand N, Sutliff TD, Rodriguez RL (1990). The alpha-amylase genes in *Oryza sativa*: characterization of cDNA clones and mRNA expression during seed germination. *Mol Gen Genet* **221**: 235–244.

Paterson AH, Schertz KF, Lin Y-R, Liu S-C, Chang Y-L (1995). The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proc Natl Acad Sci USA* **92**: 6127–6131.

Planchais S, Perennes C, Glab N, Mironov V, Inzé D, Bergouniox C (2002). Characterization of *cis*-acting element involved in cell cycle phase-independent activation of Arath;*CycB1*;1 transcription and identification of putative regulatory proteins. *Plant Mol Biol* **50**: 109–127.

Rozen S, Skaletsky HJ (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press: Totowa, NJ. pp 365–386.

Satoh R, Nakashima K, Seki M, Shinozaki K, Yanaguchi-Shinozaki K (2002). ACTCAT, a novel *cis*-acting element for proline- and hypoosmolarity-responsive expression of the *ProDH* gene encoding proline dehydrogenase in *Arabidopsis*. *Plant Physiol* **130**: 709–719.

Snedecor GW, Cochran WG (1980). *Statistical Methods*. Iowa State University Press: Ames, IA. pp 110–121.

Thomas MS, Flavell RB (1990). Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. *Plant Cell* **2**: 1171–1180.

Thum KE, Kim M, Morishiger DT, Eibl C, Koop HU, Mullet JE (2001). Analysis of barely chloroplast psbD light-responsive promoter elements in transplastomic tobacco. *Plant Mol Biol* **47**: 353–366.

Wang R-L, Stec A, Hey J, Livak L, Doebley J (1999). The limits of selection during maize domestication. *Nature* **398**: 236–239.

Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH (2007). Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics* **177**: 1753–1763.

Wang Z-Y, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM (1997). A myb-related to transcription factor is involved in the phytochrome regulation of an *Arabidopsis* Lhcb gene. *Plant Cell* **9**: 491–507.

Xue GP (2003). The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. *Plant J* **33**: 373–383.

Yuksel B, Paterson AH (2005). Construction and characterization of a peanut *Hind*III BAC library. *Theor Appl Genet* **111**: 630–639.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)