Discovery of MicroRNA169 Gene Copies in Genomes of Flowering Plants through Positional Information

Martín Calviño and Joachim Messing*

Waksman Institute of Microbiology, Rutgers University

*Corresponding author: E-mail: messing@waksman.rutgers.edu.

Accepted: January 18, 2013

Abstract

Expansion and contraction of microRNA (miRNA) families can be studied in sequenced plant genomes through sequence alignments. Here, we focused on miR169 in sorghum because of its implications in drought tolerance and stem-sugar content. We were able to discover many miR169 copies that have escaped standard genome annotation methods. A new miR169 cluster was found on sorghum chromosome 1. This cluster is composed of the previously annotated sbi-*MIR169*o together with two newly found *MIR169* copies, named sbi-*MIR169*t and sbi-*MIR169*u. We also found that a miR169 cluster on sorghum chr7 consisting of sbi-*MIR169*l, sbi-*MIR169*m, and sbi-*MIR169*n is contained within a chromosomal inversion of at least 500 kb that occurred in sorghum relative to *Brachypodium*, rice, foxtail millet, and maize. Surprisingly, synteny of chromosomal segments containing *MIR169* copies with linked bHLH and CONSTANS-LIKE genes extended from *Brachypodium* to dictotyledonous species such as grapevine, soybean, and cassava, indicating a strong conservation of linkages of certain flowering and/or plant height genes and microRNAs, which may explain linkage drag of drought and flowering traits and would have consequences for breeding new varieties. Furthermore, alignment of rice and sorghum orthologous regions revealed the presence of two additional miR169 gene copies (miR169r and miR169s) on sorghum chr7 that formed an antisense miRNA gene pair. Both copies are expressed and target different set of genes. Synteny-based analysis of microRNAs among different plant species should lead to the discovery of new microRNAs in general and contribute to our understanding of their evolution.

Key words: comparative genomics, grasses, synteny, linkage drag, flowering, drought.

Introduction

Several mechanisms have been proposed to explain the evolutionary origin of microRNA (miRNA) genes. For instance, they can be derived from miniature-inverted repeat transposable elements (MITEs) because the inverted repeat with a short internal sequence can be transcribed and form a hairpin structure that can be processed into small RNAs. Indeed, several miRNA genes derived from MITEs have been described in *Arabidopsis* and rice (Piriyapongsa and Jordan 2008). It has also been proposed that miRNA genes can originate from spontaneous mutations in hairpin-like structures in the genome, and several miRNAs in *Arabidopsis* appeared to have originated this way (Fenselau de Felippes et al. 2008). The third and probably the most accepted explanation for the origin of microRNAs is based on the inverted duplication of genes, which when transcribed would form hairpin structures capable of generating small RNAs with perfect complementarity to the parental transcripts (Allen et al. 2004; Axtell and Bowman 2008). Over time, the accumulation of mutations erodes the extensive homology with the parental transcripts and the accuracy of small RNA processing improves, eventually leaving a single segment (the mature miRNA) that retains complementarity (Allen et al. 2004; Axtell and Bowman 2008). This hypothesis is supported with evidence where extended complementarity between plant miRNAs and target mRNAs is more evident in less-conserved and younger loci (Fahlgren et al. 2007).

Duplication of a newly formed miRNA eventually results in the creation of a multigene miRNA family, with evolutionary old and conserved miRNAs having more than one gene copy in the genome, whereas new and thus nonconserved (or species-specific) miRNAs being usually single copy (Allen et al. 2004; Fahlgren et al. 2007; Ma et al. 2010). Similar to protein-coding genes, duplication and subsequent divergence of miRNA gene copies can lead to loss of function (pseudogenes), keep current function (gene redundancy), gain a new

function (neofunctionalization), or acquire a more specialized function (subfunctionalization) (Maher et al. 2006). Consistent with this, diversification in the sequence of duplicated miRNA gene copies was accompanied by changes in spatial and temporal expression patterns (Jiang et al. 2006; Maher et al. 2006). MicroRNA genes that undergo events of tandem duplication result in the formation of paralogous miRNA gene copies located in close proximity to each other on the same chromosome and thus forming miRNA clusters. Recently, Sun et al. (2012) analyzed miRNAs that had amplified through tandem duplication in *Arabidopsis*, poplar (*Populus thricocarpa*), rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*) genomes and found that 248 miRNAs in total belonging to 51 miRNA families arose by tandem duplication. This study showed the importance of tandem duplication events as a major force in the creation of new miRNA gene copies and into the expansion of miRNA families. Interestingly, the average miRNA copy number in tandemly duplicated regions from eudicots *A. thaliana* and *P. thricocarpa* was lower (2.8 copies/tandem) than in monocots *O. sativa* and *S. bicolor* (3.4 copies/tandem), suggesting that tandem duplications might have been more common in rice and sorghum (Sun et al. 2012). Despite this finding, there is a lack of knowledge on the evolutionary fate of miRNA gene clusters across the grass family.

Here, we analyzed the process of tandem duplication that gave rise to *MIR169* gene clusters in sorghum (*S. bicolor* [L.] Moench) and traced its evolutionary path by aligning contiguous chromosomal segments of diploid *Brachypodium*, rice, foxtail millet, and the two homoeologous regions of allotetraploid maize. We have chosen miR169 as an example because of its possible role in stem-sugar accumulation in sorghum besides its previously described role in drought stress response in several plant species. We discovered allelic variation in miR169 expression between grain and sweet sorghum, suggesting that miR169 could also play a role in the sugar content of sorghum stems (Calvino et al. 2011). Although high sugar content in stems is a trait shared by sorghum and sugarcane (Calvino et al. 2008, 2009), this trait seems to be silent in other grasses (Calvino and Messing 2011). This prompted us to investigate the evolution and dynamic amplification of miR169 gene copies in grass genomes. We found that synteny of chromosomal segments containing *MIR169* gene copies was conserved between monocotyledonous species such as *Brachypodium* and sorghum but surprisingly also across the monocot barrier in dicotyledonous species such as grapevine, soybean, and cassava. Furthermore, linkage of *MIR169* copies with a bHLH gene similar to *Arabidopsis* bHLH137 and with a CONSTANS-LIKE gene similar to *Arabidopsis COL14* was conserved in all the grasses examined as well as in soybean and cassava (linkage between *MIR169* and bHLH genes) and grapevine (linkage between *MIR169* and *COL14* genes). We discuss the importance of this finding for breeding crops with enhanced bioenergy traits.

Materials and Methods

DNA Sequences

Rice sequences were downloaded from the Rice Annotation Project Database website (http://rapdb.dna.affrc.go.jp/), whereas *Brachypodium*, foxtail millet, sorghum, maize, grapevine, soybean, and cassava sequences were downloaded from the Join Genome Institute website (www.phytozome.net). MicroRNA sequences were downloaded from the miRBase database (http://www.mirbase.org/).

MIR169 Gene Prediction and Annotation

Stem-loop precursors/hairpin structures from previously annotated *MIR169* genes were used in reciprocal Blastn analysis during the process of creating synteny graphs. Previously known *MIR169* stem-loop precursors were used as query sequences with Blastn. When the corresponding target sequences identified matched a genomic region where there was no any previous annotation of a *MIR169* gene copy, we took a 100–300 bp segment and fed it into an RNA folding program (RNAfold web server: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) to look for signatures of hairpin-like structures typical of microRNAs. Guidelines in microRNA gene prediction were followed as suggested by Meyers et al. (2008).

Experimental Validation of Predicted *MIR169* Genes

We took advantage of our previously sequenced small RNA libraries from sorghum stems (Calvino et al. 2011) and mapped small RNAs to the newly predicted *MIR169*r/s/t/u/v hairpin sequences. To validate the newly predicted *MIR169*s in maize, we used the SOLiD platform to sequence small RNAs derived from endosperm tissue from B73 and Mo17 inbred lines as well as endosperm tissue derived from their reciprocal crosses. Small RNA reads were then mapped to zma-*MIR169*s stem-loop precursor.

Prediction of miR169 Targets

Target prediction was conducted in sorghum for the newly discovered miR169r* and miR169s microRNAs using the Small RNA Target Analysis Server psRNATarget (Dai and Zhao 2011) at http://plantgrn.noble.org/psRNATarget/. In addition to the sorghum genome sequence incorporated into psRNATarget (Sorghum DCFI Gene Index SBGI Release 9) as preloaded transcripts, we also uploaded a FASTA file from phytozome (http://www.phytozome.net/dataUsagePolicy.php?org=Org_Sbicolor) with all sorghum genes coding sequences and used this data set for target prediction as well. Target prediction was conducted for the annotated 21 nt miR169 and for the most abundant small RNA reads different from 21 nt in size that matched the predicted miR169 sequence (miR169 variants).

Estimation of *MIR169* Gene Number in Ancestral Species

To estimate the numbers of *MIR169* genes in ancestral species of the grass family together with gains and losses of *MIR169* copies during grass evolution, we took the parsimony approach as described previously by Nozawa et al. (2012).

Estimation of Substitution Rates in *MIR169* Genes and Ancient Duplication Time

To study the rate of nucleotide substitution in *MIR169* genes, we aligned *MIR169* stem-loop sequences using MUSCLE, available with the MEGA5 software package (Tamura et al. 2011). When we analyzed the gained *MIR169* gene copy that gave rise to sit-*MIR169*h, sbi-*MIR169*v, and zma-*MIR169*s copies (fig. 6*A*: region miR169 cluster on sorghum chr2), we first computed the average (Jukes and Cantor) distance (D_a) between zma-*MIR169*s/sbi-*MIR169*v and zma-*MIR169*s/sit-*MIR169*h gene pairs. The substitution rate (R) was subsequently calculated with the formula $R = D_a/2T$, where T is the divergence time (in this case 26 million years ago [Ma]), when the ancestor of maize and sorghum diverged from foxtail millet. We then calculated the ancient duplication time at which sit-*MIR169*h arose by using the formula $t = d_a/2R$, where t is the divergence time of two sequences and d_a is the average distance between sequences in the miR169 cluster (the average of pairwise distances between sit-*MIR169*h/sit-*MIR169*g and sit-*MIR169*h/sit-*MIR169*f, respectively). A similar rationale was applied for the calculation of the ancient duplication time of sbi-*MIR169*t in the sorghum miR169 cluster 1 (fig. 6*A*).

Rate of Synonymous and Nonsynonymous Substitutions of the bHLH Orthologous Gene Pairs

We used gene exon sequences to estimate synonymous and nonsynonymous substitutions using the MEGA5 program (Tamura et al. 2011). The synonymous and nonsynonymous substitution rate was calculated for a given bHLH orthologous gene pair (*Brachypodium*–rice; *Brachypodium*–foxtail millet; *Brachypodium*–sorghum; and *Brachypodium*–maize), where *Brachypodium* bHLH gene Bradi3g41510 was compared with the HLH gene Bradi4g34870.

Phylogenetic Analysis

Phylogenetic analysis were performed by creating multiple alignments of nucleotide or amino acid sequences using MUSCLE and Clustal_W, respectively, and phylograms were drawn with the MEGA5 program using the neighbor joining (NJ) method (Tamura et al. 2011). Multiple alignments of microRNA 169 stem-loop sequences were improved by removing the unreliable regions from the alignment using the web-based program GUIDANCE (http://guidance.tau.ac.il), and NJ phylogenetic tress were created with 2,000 bootstrap replications, and the model/method used was the maximum composite likelihood.

Results

New *MIR169* Gene Copies in the Rice, Sorghum, and Maize Genomes

A miRNA cluster as defined in the miRBase database (release 19, August 2012) is composed of two or more miRNA gene copies that are located on the same chromosome and separated from each other by a distance of 10 kb or less. The distance set to define a miRNA cluster is arbitrary though, as evidenced by a cluster composed of 16 copies of MIR2118 distributed over an 18-Kb segment on rice chr4 (Sun et al. 2012). The sequencing of the sorghum genome allowed the identification of 17 *MIR169* gene copies, from which five were arranged in two clusters, one located on chr2 (sbi-*MIR169*f and sbi-*MIR169*g) and the other located on chr7 (sbi-*MIR169*l, sbi-*MIR169*m, and sbi-*MIR169*n, respectively (Paterson et al. 2009) (fig. 1 and table 1).

We first analyzed the region containing the *MIR169* cluster on sorghum chr7 because it had the highest number of gene copies. The alignment of sorghum genes flanking *MIR169* copies to the rice genome permitted the identification of a collinear region on rice chr8 also containing a cluster of *MIR169* gene copies (fig. 2). Interestingly, the cluster on rice chr8 was composed of five *MIR169* gene copies, whereas the orthologous cluster on sorghum chr7 contained only three annotated *MIR169* gene copies. Further investigation based on reciprocal Blastn analysis revealed that osa-*MIR169*l and osa-*MIR169*q are orthologous to a region on sorghum chr7, where there was no previous annotation of *MIR169* genes. Indeed, by taking the sorghum DNA segment highly similar to osa-*MIR169*l and osa-*MIR169*q and subjecting it to an RNA folding program (RNAfold: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) to identify hairpin-like structures characteristic of microRNA precursors, we were able to discover two new *MIR169* gene copies in sorghum that we named sbi-*MIR169*r and sbi-*MIR169*s, respectively (fig. 2 and supplementary fig. S1, Supplementary Material online). Independent support for the new annotation of sbi-*MIR169*r and sbi-*MIR169*s was achieved through orthologous alignment of a third species, maize, through zma-*MIR169*e and zma-*MIR169*h gene copies (supplementary fig. S2, Supplementary Material online).

To identify additional *MIR169* gene copies in sorghum that might have arisen by tandem duplication, we took each of the annotated *MIR169* genes and performed Blastn analysis against the sorghum genome to search for new copies located in close proximity to any of the previously annotated ones. Such analysis identified two new *MIR169* copies on sorghum chromosome 1 (chr1) when sbi-*MIR169*o was used as query that we named sbi-*MIR169*t and sbi-*MIR169*u, respectively (supplementary fig. S1, Supplementary Material online). Thus, sbi-*MIR169*o together with sbi-*MIR169*t and sbi-*MIR169*u constituted a new *MIR169* cluster of the sorghum genome (table 1). The segment containing the newly

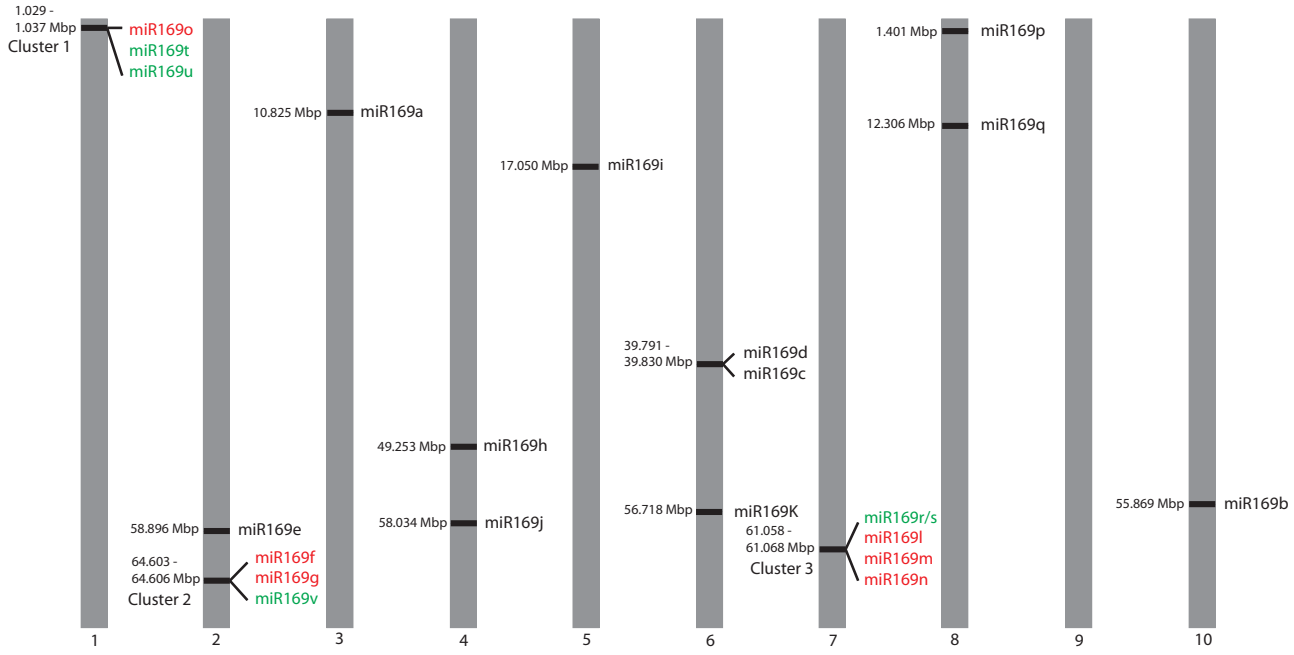

Fig. 1.—Distribution of *MIR169* gene copies in the genome of *Sorghum bicolor* cultivar BTx623. A total of 22 *MIR169* gene copies are shown, with 17 copies previously annotated by the sorghum genome-sequencing consortium (shown in black and red) (Paterson et al. 2009) and with five additional *MIR169* copies described in this study for the first time (shown in green). The evolutionary trajectory of sorghum *MIR169* gene copies arranged in clusters 1, 2, and 3 are described.

Table 1

Summary of *MIR169* Gene Copies Described in This Study

Chromosome	Gene ID[a]	Coordinates[b]	Strand	Distance between Genes Flanking the Cluster[c]
Brachypodium distachyon				
chr1	bdi-*MIR169*k	1,175,425..1,175,598	+	
chr3	bdi-*MIR169*e	43,441,526..43,441,689	+	Cluster 1: bdi-*MIR169*e to bdi-*MIR169*g = 2,960 bp
	bdi-*MIR169*g	43,444,486..43,444,666	+	
Oryza sativa				
chr3	osa-*MIR169*r	35,782,397..35,782,553	+	
chr8	osa-*MIR169*i	26,891,154..26,891,261	+	Cluster 1: osa-*MIR169*i to osa-*MIR169*q = 14,446 bp
	osa-*MIR169*h	26,895,354..26,895,475	+	
	osa-*MIR169*m	26,901,902..26,902,039	+	
	osa-*MIR169*l	26,905,493..26,905,600	+	
	osa-*MIR169*q	26,905,600..26,905,493	−	
chr9	osa-*MIR169*j	19,788,861..19,788,985	+	Cluster 2: osa-*MIR169*j to osa-*MIR169*k = 3,272 bp
	osa-*MIR169*k	19,792,133..19,792,288	+	
Setaria italica				
chr9	sit-*MIR169*o	526,081..525,981	−	
chr2	sit-*MIR169*f	36,921,078..36,921,205	+	Cluster 1: sit-*MIR169*f to sit-*MIR169*h = 3,137 bp
	sit-*MIR169*g	36,923,991..36,924,143	+	
	sit-*MIR169*h	36,924,215..36,924,361	+	
chr6	sit-*MIR169*i	33,994,480..33,994,680	+	Cluster 2: sit-*MIR169*i to sit-*MIR169*s = 8,922 bp
	sit-*MIR169*j	33,997,832..33,997,997	+	
	sit-*MIR169*k	34,001,008..34,001,109	+	
	sit-*MIR169*r	34,003,536..34,003,402	−	
	sit-*MIR169*s	34,003,402..34,003,536	+	

(continued)

Table 1 Continued

Chromosome	Gene ID[a]	Coordinates[b]	Strand	Distance between Genes Flanking the Cluster[c]
Sorghum bicolor				
chr1	sbi-*MIR169*o	1,029,916..1,029,814	–	Cluster 1: sbi-*MIR169*o to sbi-*MIR169*u = 7,321 bp
	sbi-*MIR169*t	1,030,265..1,030,155	–	
	sbi-*MIR169*u	1,037,237..1,037,096	–	
chr2	sbi-*MIR169*f	64,603,670..64,603,817	+	Cluster 2: sbi-*MIR169*f to sbi-*MIR169*v = 3,049 bp
	sbi-*MIR169*g	64,606,503..64,606,654	+	
	sbi-*MIR169*v	64,606,719..64,606,868	+	
chr7	sbi-*MIR169*r	61,058,625..61,058,750	+	Cluster 3: sbi-*MIR169*r to sbi-*MIR169*n = 12,648 bp
	sbi-*MIR169*s	61,058,750..61,058,625	–	
	sbi-*MIR169*l	61,062,736..61,062,640	–	
	sbi-*MIR169*m	61,068,118..61,068,027	–	
	sbi-*MIR169*n	61,071,181..61,071,273	+	
Zea mays				
chr1	zma-*MIR169*l	298,277,019..298,277,107	+	
chr2	zma-*MIR169*j	192,700,339..192,700,489	+	Cluster 1: zma-*MIR169*j to zma-*MIR169*s = 277 bp
	zma-*MIR169*s	192,700,616..192,700,748	+	
chr4	zma-*MIR169*i	47,241,963..47,242,153	+	Cluster 2: zma-*MIR169*i to zma-*MIR169*e = 271,605 bp
	zma-*MIR169*d	47,454,177..47,454,304	–	
	zma-*MIR169*h	47,513,567..47,513,694	+	
	zma-*MIR169*e	47,513,695..47,513,568	–	
chr7	zma-*MIR169*k	135,706,179..135,706,311	–	
Vitis vinifera				
chr1	vvi-*MIR169*y	22,233,573..22,233,820	+	
chr14	vvi-*MIR169*z	25,082,612..25,082,498	–	Cluster 1: vvi-*MIR169*z to vvi-*MIR169*e = 367 bp
	vvi-*MIR169*e	25,082,865..25,082,717	–	
chr17	vvi-*MIR169*x	355,713..355,837	–	
Glycine max				
chr6	gma-*MIR169*w	13,783,352..13,783,225	–	
chr8	gma-*MIR169*x	717,092..717226	+	Cluster 1: gma-*MIR169*o to gma-*MIR169*p = 7,248 bp
	gma-*MIR169*y	724,205..724,340	+	
Manihot esculenta				
scaffold01701	mes-*MIR169*w	436,633..436,794	+	
scaffold09876	mes-*MIR169*y	536,510..536,709	–	

[a]In green are microRNA genes identified in this study.
[b]Chromosomal positions are based on Phytozome annotation for all the species except rice that is based on RAPDB annotation.
[c]Distance within the cluster is calculated from the beginning of the first miRNA gene to the beginning of the last miRNA gene in the cluster.

identified *MIR169* cluster on sorghum chr1 was collinear with an orthologous segment of rice chr3 (fig. 3), although no *MIR169* gene had previously been found in this region. By performing reciprocal Blastn analysis with sbi-*MIR169*o against the rice genome, we could identify the corresponding orthologous *MIR169* copy on rice chr3 that we named osa-*MIR169*r (fig. 3 and supplementary fig. S1, Supplementary Material online). Furthermore, osa-*MIR169*r is contained within a segment that is collinear with an orthologous region of chr1 of a fourth species, *Brachypodium*, corresponding to bdi-*MIR169*k (fig. 3). Comparison between sorghum and maize revealed that the *MIR169* cluster on sorghum chr1 is collinear with a segment on maize chr1 that contains zma-*MIR169*l (supplementary fig. S3, Supplementary Material online). Indeed, sbi-*MIR169*u and zma-*MIR169*l are also orthologous gene copies. Finally, when the cluster on sorghum chr2 containing sbi-*MIR169*f and sbi-*MIR169*g was analyzed, collinearity with the segment on sorghum chr7 containing the sbi-*MIR169*r/s and sbi-*MIR169*l-n cluster revealed the existence of an additional *MIR169* copy on sorghum chr2 that we named sbi-*MIR169*v (fig. 2; supplementary fig. S1, Supplementary Material online; and table 1). Furthermore, the sbi-*MIR169*f/g/v cluster is syntenic with a region on maize chr7 containing zma-*MIR169*k and its homoeologous region on maize chr2 containing zma-*MIR169*j and the newly identified zma-*MIR169*s gene copy (supplementary figs. S1 and S4, Supplementary Material online; table 1).

In summary, by aligning sorghum chromosomal segments containing *MIR169* clusters with orthologous regions of *Brachypodium,* rice, and maize, we were able to identify five additional *MIR169* copies in sorghum and an additional copy in rice and maize, respectively.

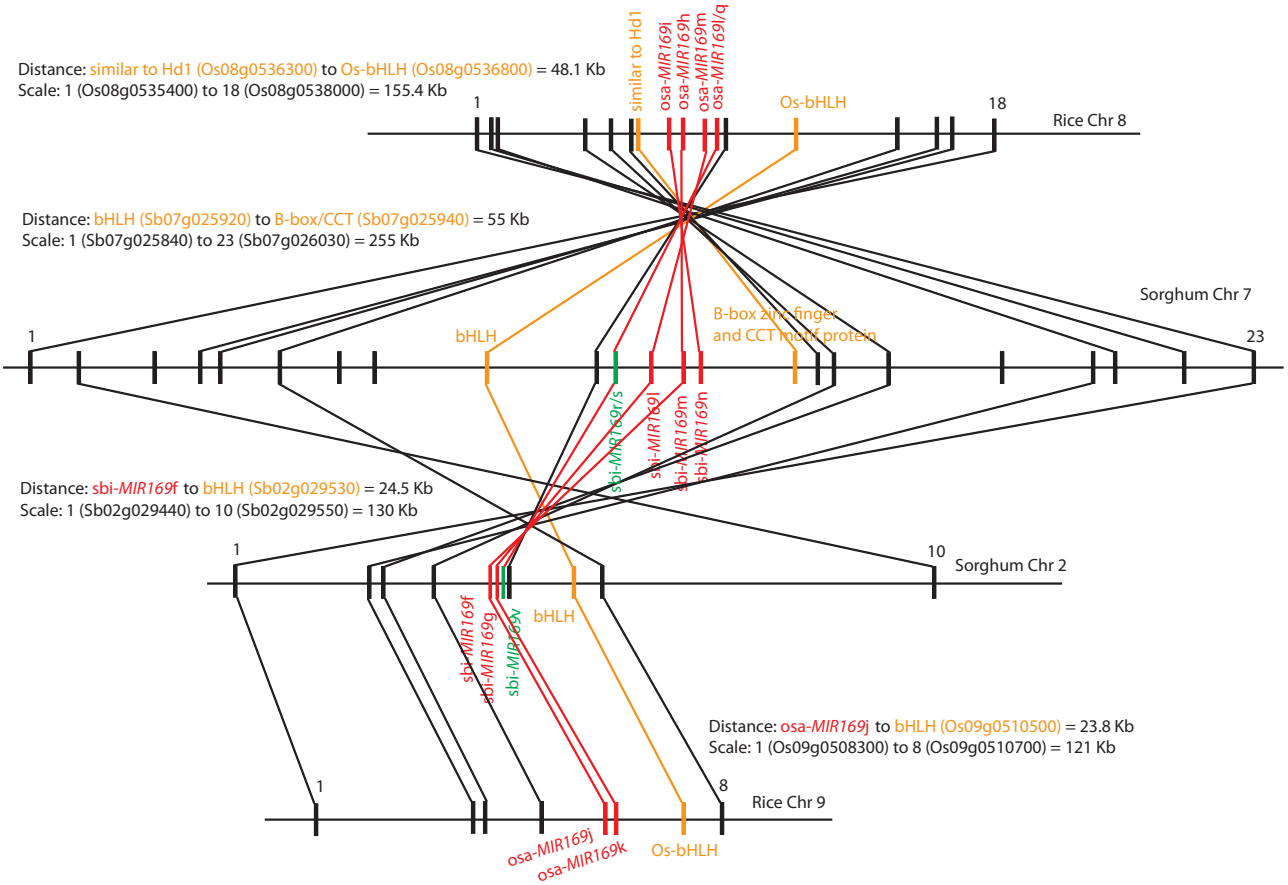

Fig. 2.—Syntenic alignment of rice and sorghum chromosomal segments containing *MIR169* gene clusters. Sorghum *MIR169* gene clusters on chr2 and chr7 together with their flanking protein coding genes were aligned with rice by orthologous gene pairs. Rice and sorghum chromosomes are represented as horizontal lines, whereas genes along the chromosome are represented as rectangle bars. Known *MIR169* gene copies are shown as red bars, whereas new *MIR169* gene copies described in this study are shown as green bars. The bHLH and B-box zinc finger and CCT motif (B-box/CCT) genes are represented as yellow bars. All other protein coding genes in the chromosomal regions under study are represented as black bars. Orthologous gene pairs are indicated as lines connecting bars, with red lines indicating orthology between *MIR169* gene pairs and yellow lines indicating orthology between bHLH and B-box/CCT gene pairs, respectively. All other orthology between rice and sorghum protein coding genes are indicated as black lines connecting black bars. The physical distance between bHLH and B-box/CCT genes and/or between bHLH or B-Box/CCT genes to the flanking *MIR169* copy is indicated. To provide a scale of the chromosomal segments highlighted in the figure, the physical distance between the first and the last gene in the segment is indicated and thus serves as a reference to observe expansion and contraction of genomic regions. An inversion event on sorghum chr7 containing the *MIR169* cluster occurred relative to the orthologous regions on sorghum chr2 and rice chr8 and chr9 respectively.

New *MIR169* Clusters in the Recently Sequenced Foxtail Millet Genome

The recent release of the complete reference genome sequence for foxtail millet (*Setaria italica*) (Bennetzen et al. 2012; Zhang et al. 2012) greatly enhances comparative genomics analysis within the *Poaceae,* with genome sequences available from five species. Foxtail millet provided us with additional information to study syntenic relationships with sorghum because they split from each other approximately 26 Ma (Zhang et al. 2012). Indeed, 19 collinear blocks were found between foxtail millet and sorghum, which comprised approximately 72% of the foxtail millet genome (Zhang et al. 2012). Consequently, we could use sorghum to identify and predict *MIR169* gene copies in the foxtail millet genome. We identified and predicted *MIR169* copies in foxtail millet, collinear with sorghum *MIR169* copies, arranged in clusters on chr1, chr2, and chr7. The sorghum *MIR169* cluster on chr1 was collinear with a segment on chr9 of foxtail millet, from which sit-*MIR169*o was identified as the ortholog of sbi-*MIR169*o (fig. 3; supplementary fig. S1, Supplementary Material online; and table 1). The sorghum *MIR169* copies arranged in cluster on chr7 were collinear with a segment on chr6 from foxtail millet that harbored the newly identified orthologous *MIR169* copies sit-*MIR169*i, sit-*MIR169*j, sit-*MIR169*k, sit-*MIR169*r, and sit-*MIR169*s (fig. 4; supplementary fig. S1, Supplementary Material online; and table 1). Finally,

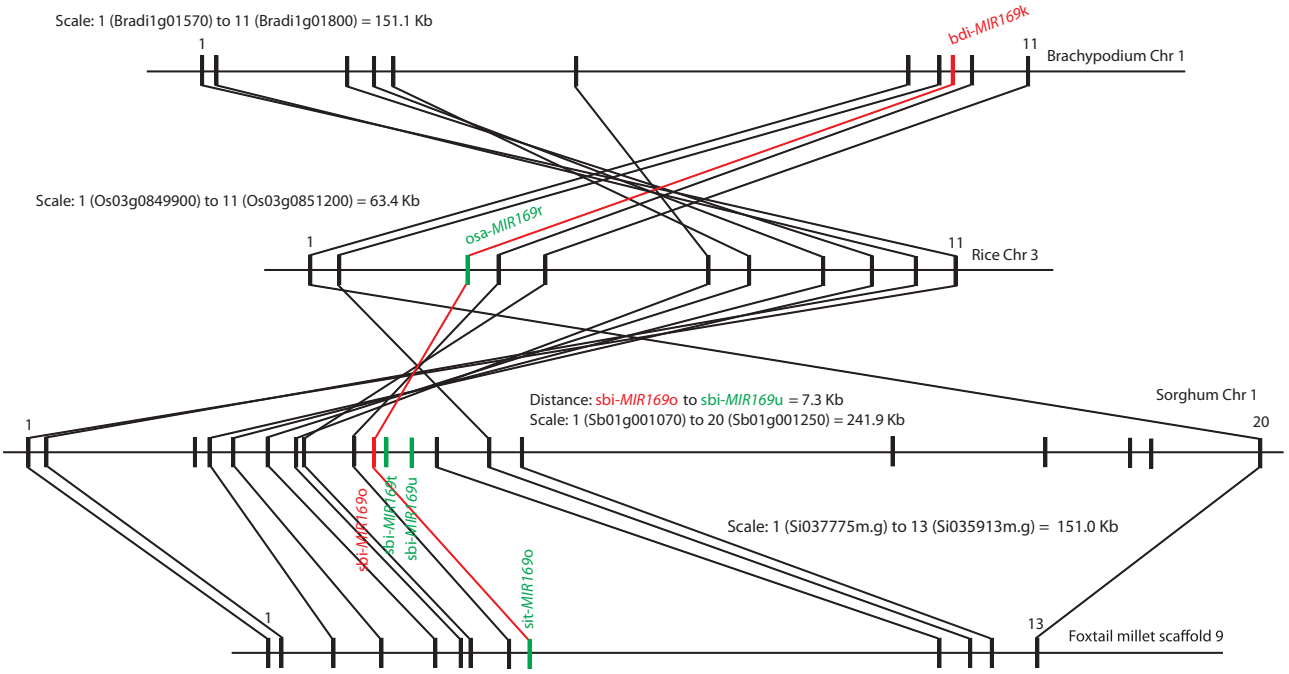

Fig. 3.—Sequence alignment of sorghum *MIR169* cluster on chr1 with orthologous regions from *Brachypodium*, rice and foxtail millet. The sbi-*MIR169*o copy in sorghum allowed the identification of the orthologous osa-*MIR169*r copy in rice and sit-*MIR169*o copy in foxtail millet, respectively. For the region containing sbi-*MIR169*o/t/u on chr1, we could not find sufficient conservation of synteny to identify an orthologous region in sorghum, thus a synteny graph is only shown with sorghum chr1. An inversion event on rice chr3 occurred relative to *Brachypodium*, foxtail millet, and sorghum.

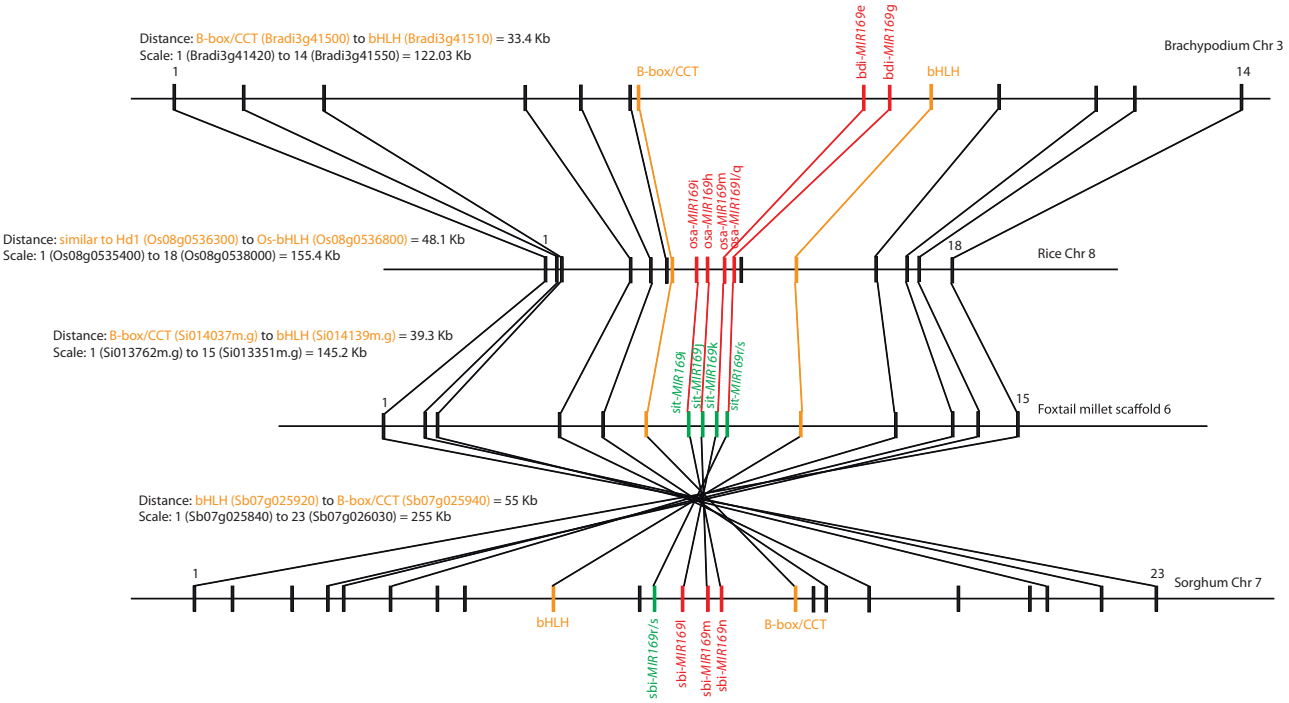

Fig. 4.—Sequence alignment of sorghum *MIR169* cluster on chr7 with orthologous regions from *Brachypodium*, rice, and foxtail millet. Rice and sorghum *MIR169* gene copies were used to identify and annotate five *MIR169* genes in foxtail millet (shown in green). The bHLH and B-box/CCT genes were physically adjacent to *MIR169* gene copies in the four species examined. The region examined on sorghum chr7 expanded relative to the orthologous region from the other three grasses and was inverted only in sorghum.

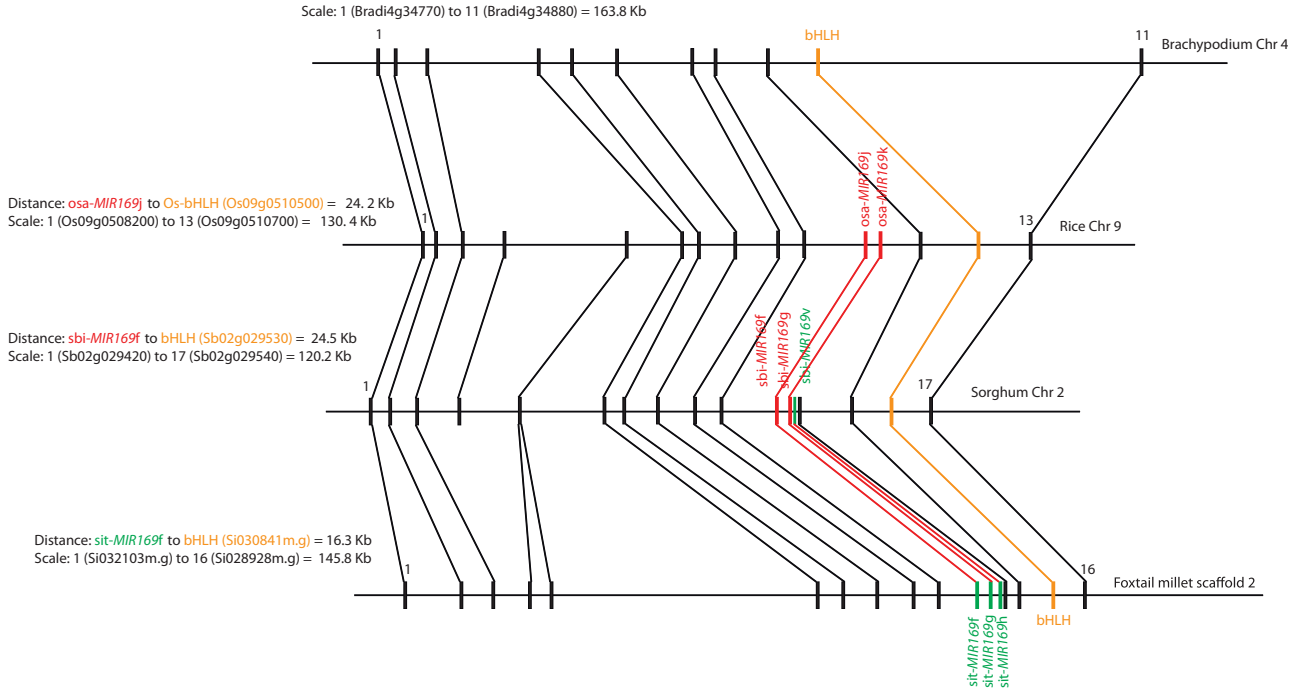

FIG. 5.—Sequence alignment of sorghum *MIR169* cluster on chr2 with orthologous regions from *Brachypodium*, rice, and foxtail millet. *MIR169* gene copies were deleted from *Brachypodium* chr4 but the flanking genes remained. The *MIR169* gene cluster in rice was composed of two copies, whereas in sorghum and foxtail millet, the cluster comprised three copies. The bHLH gene was present in all four grasses and was physically adjacent to *MIR169* gene copies in rice, sorghum, and foxtail millet. Sorghum *MIR169* gene copies were used to identify and annotate the orthologous copies on foxtail millet scaffold 2 (shown in green).

tandem sorghum *MIR169* copies on chr2 were collinear with a segment on foxtail millet chr2 that contained the three newly predicted *MIR169* copies sit-*MIR169*f, sit-*MIR169*g, and sit-*MIR169*h (fig. 5; supplementary fig. S1, Supplementary Material online; and table 1).

In summary, we used sorghum as a reference genome to identify and predict nine *MIR169* gene copies that were collinear with foxtail millet. The prediction of *MIR169* genes in the foxtail millet will greatly facilitate their experimental validation through the sequencing of small RNAs from different tissues and developmental stages.

Gain and Losses of *MIR169* Gene Copies during Grass Evolution

To determine expansion and contraction of the *MIR169* gene clusters, we aligned collinear chromosomal segments of diploid *Brachypodium*, rice, and foxtail millet and the two homoeologous regions of allotetraploid maize. Based on nucleotide substitution rates, the cluster of *MIR169* copies on sorghum chr7 was likely preserved from an ancestral grass chromosome and comprised five *MIR169* gene copies, from which three of them were deleted in *Brachypodium* after the split of *Brachypodium* from the ancestor of rice, foxtail millet, and sorghum (figs. 4 and 6*A* and *B*). The number of *MIR169* genes (five copies per cluster) was unchanged in rice, sorghum, and foxtail millet, whereas in maize, four copies were retained on orthologous homoeologous region on chr4 but none on the homoeologous region on chr1 (supplementary fig. S2, Supplementary Material online, and fig. 6*A*). Although the *MIR169* copies were deleted from maize chr1, the flanking genes remained intact.

In the case of the *MIR169* cluster on sorghum chr2, its evolution can be explained according to two models (fig. 6*A*). In the first one, the ancestor of the grasses had two *MIR169* copies and they were conserved before the split of *Brachypodium* and rice, with *Brachypodium* losing these two *MIR169* copies, where as rice maintained them. An additional copy was gained in the common ancestor of foxtail millet, sorghum, and maize, giving rise to a cluster with three *MIR169* gene copies. Phylogenetic analysis suggested that the new copy in the ancestor of foxtail millet, sorghum, and maize was the ancestral copy that gave rise to sit-*MIR169*h, sbi-*MIR169*v, and zma-*MIR169*s, respectively (fig. 6*C*). We estimated that the time at which this copy arose in the progenitor of foxtail millet, sorghum, and maize was approximately 41.1 Ma (see Materials and Methods for estimation of time of duplication). Alternatively, the common ancestor of the grasses could have had three *MIR169* gene copies, and one copy was lost in the common ancestor of *Brachypodium* and rice, with a subsequent loss of two

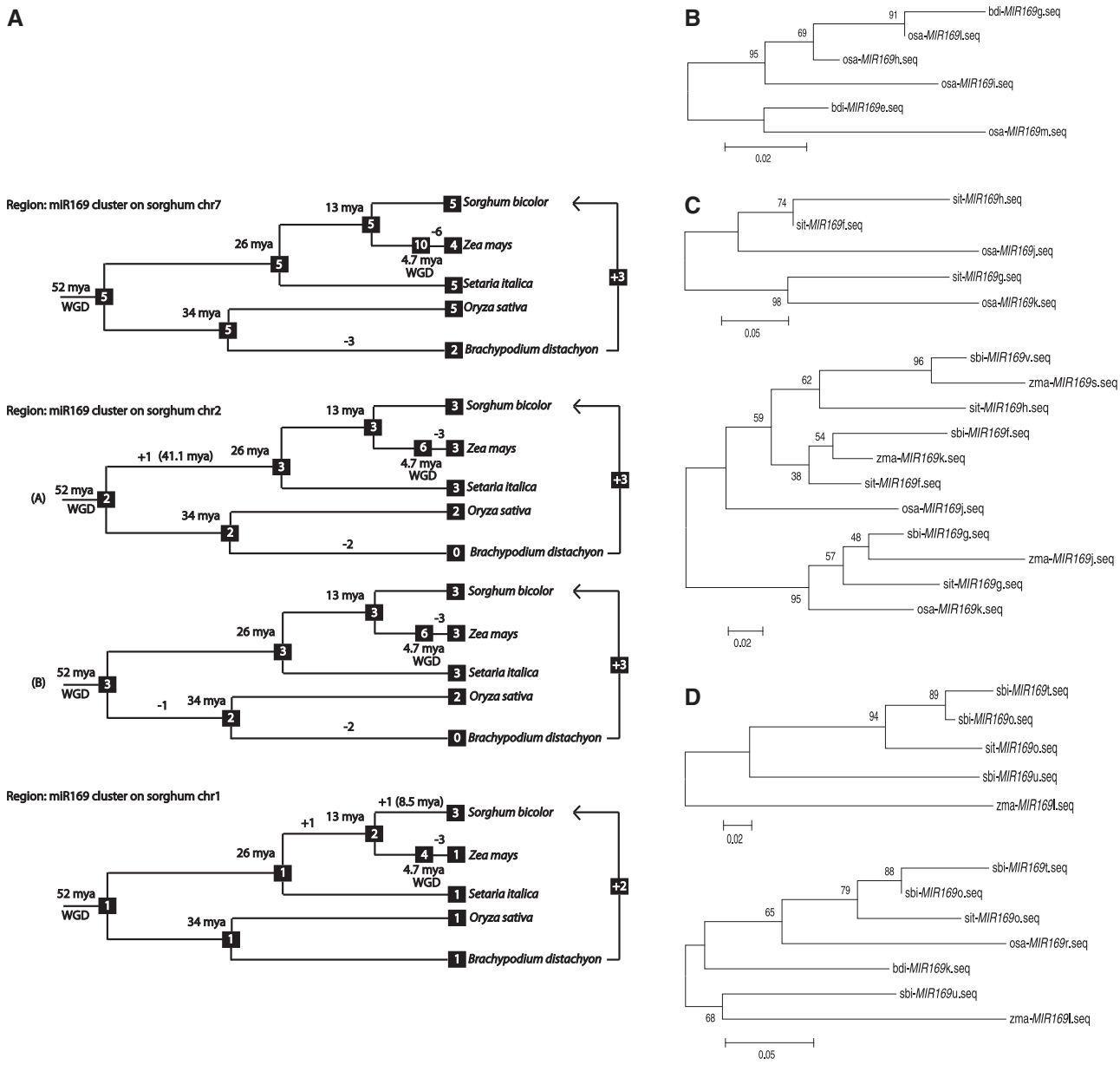

Fig. 6.—Gains and losses of *MIR169* gene copies during grass evolution. (*A*) Phylogenetic distribution of *MIR169* gene copies in ancestral and current species with gain and losses of *MIR169* copy number during grass evolution. Numbers in squares represent the number of *MIR169* gene copies for a given cluster in each species. Numbers along each line represent gains (+) and losses (−) of *MIR169* gene copies. The estimated divergence time for each species is given at each node in the tree according to Paterson et al. (2009), Brachypodium-Sequencing-Initiative (2010), Bennetzen et al. (2012) and Zhang et al. (2012). The gain in *MIR169* copy number of sorghum relative to *Brachypodium* is depicted. Note: WGD in maize is used as a term to represent the allotetraploidy event that took place. NJ phylogenetic trees with bootstrap support are shown depicting the relationships of *MIR169* stem-loop sequences from the grass species shown in (*A*). (*B*) NJ phylogenetic tree with *Brachypodium* (bdi) and rice (osa) *MIR169* stem-loop sequences orthologous to sorghum *MIR169* copies on chromosome 7. (*C*) NJ phylogenetic tree with rice (osa) and foxtail millet (sit) *MIR169* stem-loop sequences (top) and rice, foxtail millet, sorghum (sbi), and maize (zma) *MIR169* stem-loop sequences (bottom) orthologous to *MIR169* copies on sorghum chromosome 2. (*D*) NJ phylogenetic tree depicting the relationship of foxtail millet and maize *MIR169* copies orthologous to sorghum *MIR169* copies on chromosome 1 (top), and *Brachypod*ium, rice, foxtail millet, and maize *MIR169* copies orthologous to sorghum *MIR169* copies on chromosome 1 (bottom).

additional *MIR169* gene copies in *Brachypodium* relative to rice (fig. 6*A*).

Regarding the cluster of *MIR169* copies on sorghum chr1, we favor a model where the ancestor of the grasses had a single *MIR169* copy because *Brachypodium*, rice, and foxtail millet all have a single *MIR169* copy (fig. 6*D*). Thus, the additional two *MIR169* copies present in the sorghum cluster could have arisen by duplication events. Phylogenetic analysis

suggested that the ancestral copy in the cluster was sbi-*MIR169*o, from which sbi-*MIR169*t subsequently duplicated 8.5 Ma (see Materials and Methods) (fig. 6*D*). Thus, sbi-*MIR169*t was acquired specifically in the sorghum lineage. Because sbi-*MIR169*u and zma-*MIR169*l are highly related but distantly related from sbi-*MIR169*o and sbi-*MIR169*t (fig. 6*D*), we postulate that the ancestral copy of sbi-*MIR169*u and zma-*MIR169*l was inserted next to the other *MIR169* gene copies in the progenitor of sorghum and maize. In the maize lineage, diploidization after allotetraploidization led to the deletion of the corresponding orthologous *MIR169* copy from the homoeologous segment on chr5, whereas the flanking genes remained conserved (supplementary fig. S3, Supplementary Material online).

In summary, differences in *MIR169* copy number between clusters from *Brachypodium*, rice, foxtail millet, sorghum, and maize arose by duplication of ancestral *MIR169* genes that were retained or lost during grass evolution. Overall, sorghum gained eight *MIR169* copies relative to *Brachypodium,* three copies relative to rice, two copies relative to foxtail millet, and three copies relative to maize.

Polymorphisms in Chromosomal Inversions Containing *MIR169* Clusters

Through the analysis of three chromosomal regions in sorghum containing *MIR169* clusters and their alignment with the genomes of *Brachypodium*, rice, foxtail millet, and maize, we were able to identify four chromosomal inversions in total, one in rice chr3 containing osa-*MIR169*r (fig. 3); a second on sorghum chr7 containing sbi-*MIR169*r, sbi-*MIR169*s, sbi-*MIR169*l, sbi-*MIR169*m, and sbi-*MIR169*n (fig. 2); a third on maize chr1 containing zma-*MIR169*l (supplementary fig. S3, Supplementary Material online); and the fourth on maize chr7 containing zma-*MIR169*k (supplementary fig. S4, Supplementary Material online), respectively. The inversion on rice chr3 was absent from the corresponding collinear regions on *Brachypodium* chr1*,* sorghum chr1, and foxtail millet chr9 (fig. 3), indicating that the inversion happened after the split of rice from the common ancestor of sorghum and foxtail millet. The region on sorghum chr1 containing sbi-*MIR169*o, sbi-*MIR169*t, and sbi-*MIR169*u that was collinear with the inverted segment on rice chr3 was also collinear with an inverted segment on the homoeologous region of maize chr1 containing zma-*MIR169*l (supplementary fig. S3, Supplementary Material online). However, the inversion did not occur on the homoeologous region on maize chr5, indicating that the inversion occurred after the allotetraploidization event that took place in maize. The inversion on sorghum chr7 containing sbi-*MIR169*r, sbi-*MIR169*s, sbi-*MIR169*l, sbi-*MIR169*m, and sbi-*MIR169*n cluster only occurred in this species (supplementary fig. S2, Supplementary Material online, and fig. 4), suggesting that it took place after the split of sorghum from the common ancestor of sorghum and maize. The *MIR169* cluster on sorghum chr2 was collinear with an inverted region on maize chr7 containing zma-*MIR169*k (supplementary fig. S4, Supplementary Material online). The homoeologous region on chr2 did not exhibit the inversion, suggesting that it took place after the allotetraploidization event that occurred in maize.

In summary, four inversions containing *MIR169* copies were found in total, one in rice, one in sorghum, and two in maize. These inversions were lineage specific as none of them was present in a collinear region in the genome of a second grass species, indicating that these inversions happened after the species were formed.

Validation of Newly Identified *MIR169* Gene Copies in Sorghum and Maize

To experimentally validate the new *MIR169* gene copies found in sorghum through our syntenic analysis among grasses, we mapped previously sequenced small RNAs from sorghum stems (Calvino et al. 2011) to the newly predicted *MIR169*t/u/v/r/s hairpins. Similarly, to validate the newly described zma-*MIR169*s gene copy in maize, we constructed small RNA libraries from endosperm tissue belonging to cultivars B73, Mo17, and their reciprocal crosses (supplementary table S1, Supplementary Material online). Maize endosperm-derived small RNAs were then mapped to the new *MIR169*s hairpin annotated in this study. We could effectively map small RNA reads to the stem-loop sequences of all five predicted microRNA169 in sorghum (with respect to sbi-*MIR169*r/s, see next section). In the case of sbi-*MIR169*t and sbi-*MIR169*u, the most abundant small RNA reads were derived from the miR169* sequence (supplementary fig. S5, Supplementary Material online), although small RNAs derived from the canonical miR169 sequence were also found but in less abundance. The experimental validation of sbi-*MIR169*v was supported with mapping of small RNAs to the corresponding predicted mature miR169v sequence (supplementary fig. S5, Supplementary Material online). Regarding the experimental validation of the predicted zma-*MIR169*s copy in maize, we were able to detect small RNA reads derived from miR169s although their abundance was very low (supplementary fig. S5, Supplementary Material online).

Antisense MicroRNA169 Gene Pairs Generate Small RNAs that Target Different Set of Genes

In rice, osa-*MIR169*l and osa-*MIR169*q were annotated as antisense microRNAs and small RNA reads derived from both strands were identified (Xue et al. 2009). In sorghum, sbi-*MIR169*r, and sbi-*MIR169*s are collinear with osa-*MIR169*l/q (figs. 2 and 4) and are antisense microRNAs as well (supplementary figs. S1 and S6*A*, Supplementary Material online). Despite the lack of Expressed Sequence Tag (EST) evidence for sbi-*MIR169*r and sbi-*MIR169*s annotation, our previously generated small RNA library from sorghum stem tissue

(Calvino et al. 2011) supported the transcription from both strands based on small RNA reads mapped to both sbi-*MIR169*r and sbi-*MIR169*s, respectively (supplementary fig. S6*A*, Supplementary Material online). Similarly, EST evidence supported the transcription from opposite strands in the microRNA antisense pair zma-*MIR169*e/h (ESTs ZM_BFb0354L14.r and ZM_BFb0294A24.f, respectively). Because small RNAs derived from zma-*MIR169*e/h had not been previously reported (miRBase database: release 19, August 2012), we used the SOLiD system to sequence small RNAs from endosperm tissue derived from B73 and Mo17 cultivars and their reciprocal crosses; however, we could not detect small RNA reads derived from them, at least in endosperm tissue. Thus, antisense microRNAs from *MIR169* gene copies are being actively produced in rice and sorghum, and possibly in maize.

With respect to the sbi-*MIR169*r/s antisense gene pair, we found that the small RNA reads mapped to sbi-*MIR169*r were predominantly associated with the miR169r* sequence (supplementary fig. S6*A*, Supplementary Material online). The mature miRNA sequences for sbi-miR169r* and sbi-miR169s differed from each other in seven nucleotides (supplementary fig. S6*B*, Supplementary Material online). Moreover, they would have different set of genes as targets based on their sequences (supplementary figs. S7 and S8, Supplementary Material online). Moreover, the assumption that also microRNA* have functional roles was recently described (Meng et al. 2011; Yang et al. 2011).

Linkage of *MIR169* Gene Copies with Flowering and Plant Height Genes

Based on the alignment of collinear regions containing *MIR169* genes located on sorghum chr2 and chr7, we noticed a tight linkage of *MIR169* copies with two genes encoding a bHLH protein, and a B-box zinc finger and CCT-motif protein that were similar to *Arabidopsis* bHLH137 and CONSTANS-LIKE 14 proteins (figs. 2, 4, and 5 and supplementary figs. S2 and S4, Supplementary Material online). The *Arabidopsis* bHLH137 and *COL14* genes were described to have a role in gibberellin signaling (mutations in genes involved in gibberellin signaling and/or perception affects plant height [Fernandez et al. 2009]) and flowering time, respectively (Griffiths et al. 2003; Wenkel et al. 2006; Zentella et al. 2007). The physical linkage of *MIR169* gene copies to bHLH and COL genes (or any of the two) was present in all the five grasses examined. We hypothesized that the physical association of *MIR169* to either of these flowering and/or plant height genes could be of relevance because of previously reported trade-offs in sorghum between sugar content in stems and plant height and flowering time, respectively (Murray et al. 2008). For breeding purposes, the introgression of a particular gene/phenotype from a specific cultivar into another would consequently also bring in the neighboring gene, a process known as linkage drag. Furthermore, linkage drag between *MIR169* copies and the bHLH and COL genes could also be of ecological importance because a single chromosomal segment comprises genes involved in drought tolerance, sugar accumulation, and flowering. If this is the case, linkage of *MIR169* copies to either bHLH or COL genes could have been preserved even after the monocotyledonous diversification. Indeed, we were able to find collinearity between chromosomal segments containing *MIR169* and bHLH genes from *Brachypodium*, sorghum, soybean, and cassava (fig. 7). Moreover, we found that the physical linkage between *MIR169* and the bHLH gene on sorghum chr7 was retained in collinear regions of soybean chr6 and cassava scaffold 01701, respectively (fig. 7). Similarly, the physical/genetic association of *MIR169* with the bHLH gene from sorghum chr2 was retained in the corresponding collinear regions from soybean chr8 and cassava scaffold 09876 (fig. 8). Interestingly, the linkage between *MIR169* and the COL gene that was present in *Brachypodium* chr3 and sorghum chr7 was broken in the corresponding collinear regions of soybean chr6 and cassava scaffold 01701 (fig. 7). We then compared the two *MIR169* clusters from sorghum chr2 and chr7 with the grapevine genome because grapevine and sorghum are more closely related than sorghum to soybean and cassava, respectively. Our comparison revealed a two-to-three relationship between sorghum and grapevine (fig. 9), and this is consistent with the paleo-hexaploidy event that took place in the grapevine genome (Jaillon et al. 2007). The physical/genetic linkage of *MIR169* copies with the COL gene on sorghum chr7 was preserved in two of the three homoeologous chromosomal segments in grapevine on chr1 and chr14, whereas the third homoeologous segment on chr17 retained the close association of *MIR169* with the bHLH gene.

The finding of microsynteny conservation between monocots and dicots species in chromosomal segments containing *MIR169* gene copies together with bHLH and COL genes is remarkable because the estimated time of divergence between monocots and dicots is approximately 130–240 Ma (Wolfe et al. 1989; Jaillon et al. 2007). Such microsynteny conservation permitted the discovery of new *MIR169* gene copies in soybean (gma-*MIR169*w, gma-*MIR169*x and gma-*MIR169*y), cassava (mes-*MIR169*w and mes-*MIR169*y), and grapevine (vvi-*MIR169*z).

Subfunctionalization of the bHLH Gene in the *MIR169* Cluster of *Brachypodium*

The microsynteny in chromosomal segments containing miR169 gene copies flanked by the bHLH gene among such distantly related species such as *Brachypodium* and cassava suggests that the linkage between miR169 and bHLH resulted from selection because of the divergence from a common ancestor approximately 130–240 Ma. In support of this interpretation, the bHLH gene on *Brachypodium* chr4, where the

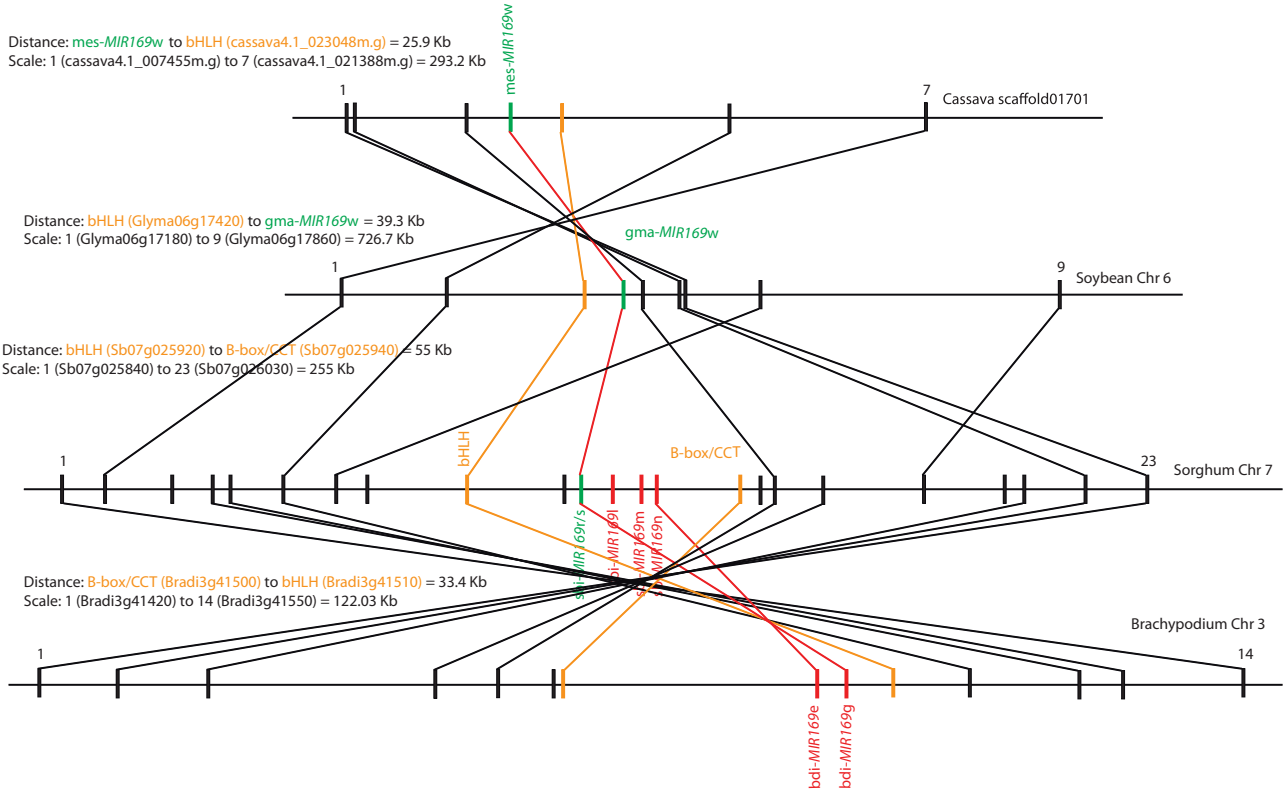

Fig. 7.—Sequence alignment of sorghum *MIR169* cluster on chr7 with orthologous regions from *Brachypodium*, soybean, and cassava. There is conservation of synteny between monocot species *Brachypodium* and sorghum and dicot species soybean and cassava when chromosomal segments containing *MIR169* gene copies and their flanking genes are aligned. Conservation of synteny allowed the identification of new *MIR169* gene copies on soybean chromosome 6 (gma-*MIR169*w) and cassava scaffold 01701 (mes-*MIR169*w), respectively. Physical association on the chromosome between *MIR169* and the flanking bHLH gene was retained in soybean and cassava as well. Notice the inversion on soybean chr6.

miR169 cluster had been deleted, appeared to have undergone subfunctionalization. First, the bHLH copy on *Brachypodium* chr4 involved the loss of the basic domain, which is involved in DNA binding (Toledo-Ortiz 2003) and thus evolved into a HLH protein (supplementary fig. S9*A* and *B*, Supplementary Material online). Because bHLH proteins act as homo- and/or heterodimers, where the basic domain of each bHLH protein binds DNA, HLH proteins homo- or heterodimerize and prevent the binding of the complex to DNA and thus becomes a negative regulator (Toledo-Ortiz 2003). Second, *Brachypodium* has a redundant intact orthologous copy on chr3, also an miR169 cluster next to it (supplementary fig. S9, Supplementary Material online). Third, the synonymous and nonsynonymous substitution rate of the HLH orthologous gene pairs was higher than the synonymous and nonsynonymous substitution rate in the bHLH orthologous gene pairs, respectively (supplementary fig. S9*C*, Supplementary Material online). Fourth, when we run a test for detecting adaptive evolution (calculated as the number of replacement mutations per replacement sites [d*N*] divided by the number of silent mutations per silent site [d*S*]) in the bHLH and HLH coding sequences, we found evidence on purifying selection on the HLH gene sequence (d*N*/d*S* ratio of −4.647).

Conservation of synteny between sorghum and grapevine showed that the linkage between *MIR169* gene copies and the COL gene was maintained in both species. Both COL genes in grapevine, on chr14 and on chr1, lost the B-box and zinc finger domain, whereas the orthologous copy in sorghum retained it (supplementary fig. S10*A* and *B*, Supplementary Material online). Similarly, foxtail millet COL protein lost the B-box and zinc finger domain, whereas *Brachypodium*, rice, and maize retained it. The B-box and zinc finger domain are thought to mediate protein–protein interactions, whereas the CCT domain acts as a nuclear localization signal, with mutations in both domains causing flowering time phenotypes (Griffiths et al. 2003; Wenkel et al. 2006; Valverde 2011). Although the COL gene on grapevine chr14 has been recently identified as a candidate gene for a flowering Quantitative Trait Loci (QTL) (Duchêne et al. 2012), the function of its corresponding orthologous copy on sorghum chr7 remains to be elucidated.

Discussion

We describe the alignment of 25 chromosomal regions with orthologous gene pairs from eight different plant species.

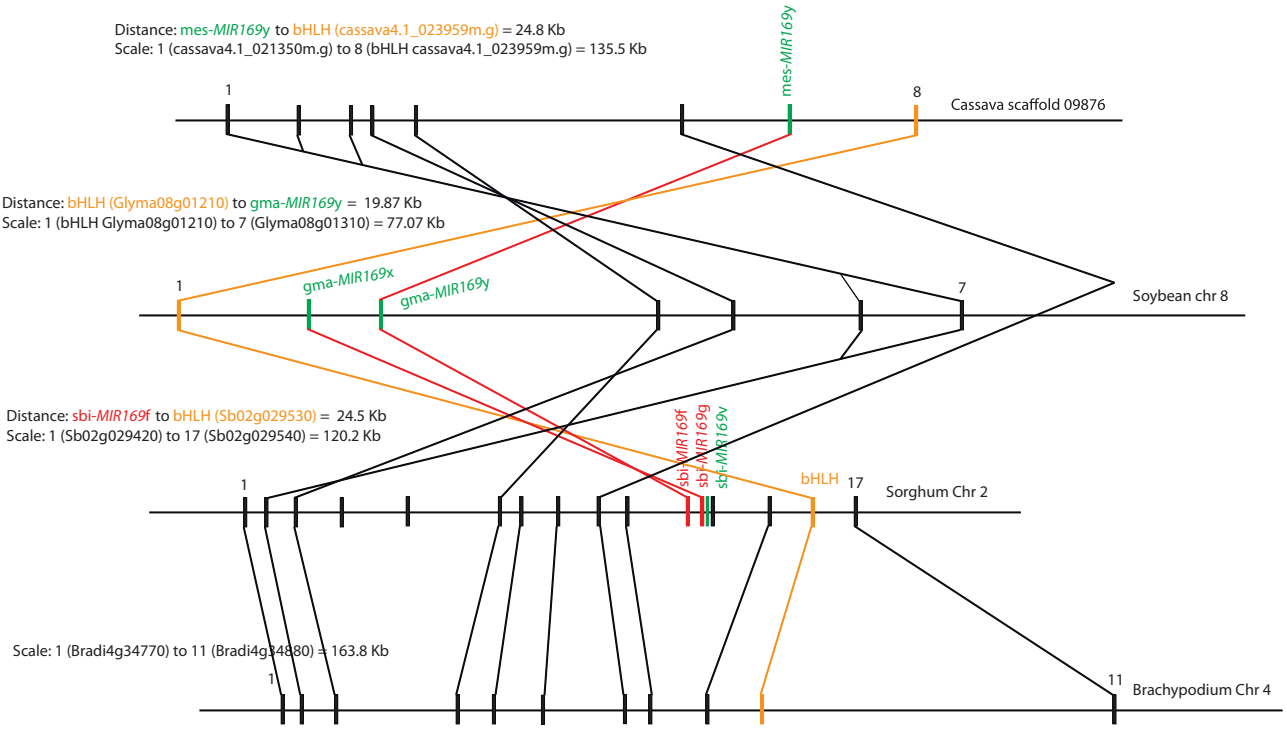

Fig. 8.—Sequence alignment of sorghum *MIR169* cluster on chr2 with orthologous regions from *Brachypodium*, soybean, and cassava. The alignment of sorghum *MIR169* cluster on chr2 with soybean chr8 and cassava scaffold 09876 allowed the identification of two new *MIR169* gene copies in soybean (gma-*MIR169*x and gma-*MIR169*y) and one new copy in cassava (mes-*MIR169*y), respectively. The physical association of *MIR169* gene copies with the bHLH was retained in soybean and cassava. An inversion occurred on soybean chr8.

These regions contain a total of 48 *MIR169* gene copies, from which 22 of them have been described and annotated here for the first time. The alignment of sorghum chromosomal regions containing *MIR169* clusters to their corresponding orthologous regions from *Brachypodium*, rice, foxtail millet, and maize, respectively, allows us not only to better understand the differential amplification of *MIR169* gene copies during speciation but also to identify new *MIR169* gene copies not previously annotated in the rice, sorghum, and maize genomes. Our work highlights the usefulness of this approach in the discovery of microRNA gene copies in grass genomes and surprisingly also in dicotyledonous genomes such as those from grapevine, soybean, and cassava. In addition, collinearity among grasses was used to predict and annotate *MIR169* hairpin structures in the foxtail millet genome de novo, from which no current microRNA annotation was available from the miRBase database (Release 19: August 2012). Our work suggests that synteny-based analysis should complement (whenever possible) homology-based searches of new microRNA gene copies in plant genomes.

Our analysis of *MIR169* gene copies organized in clusters in the sorghum genome revealed that sorghum acquired eight *MIR169* gene copies after *Brachypodium* split from a common ancestor, primarily due to gene losses (up to 5 *MIR169* gene copies) in the *Brachypodium* lineage and new gene copies (up to 3) in the sorghum lineage (fig. 6*A*). We propose that differences in *MIR169* gene copy number between sorghum and *Brachypodium* is based on selective amplification in sorghum. Because diploidization of the maize genome resulted in the deletion of duplicated gene copies after allotetraploidization approximately 4.7 Ma (Messing et al. 2004; Swigonova et al. 2004), also resulted in selective amplification in sorghum. Maize lost more than half, 9 of 16 *MIR169* gene copies, after allotetraploidization. Single gene losses in maize appear to be caused by short deletions that are predominantly in the 5–178 bp size range, with these deletions being approximately 2.3 times more frequent in one homoeologous chromosome than in the other (Woodhouse et al. 2010). This observation is particularly relevant to maize microRNAs genes with average length distributions at the 5′-regions of their primary microRNAs (pri-miRNAs) in the order of 100–300 nt (Zhang et al. 2009). Although we detected chromosome breaks of the *MIR169* neighboring gene *COL14* on the maize homoeologous chr1–chr4 pair (supplementary fig. S2, Supplementary Material online) and the bHLH gene on maize homeologous chr2–chr7 pair (supplementary fig. S4, Supplementary Material online), retention of the bHLH gene copy on both homoeologous regions from chr1 and chr4 was observed (supplementary fig. S2, Supplementary Material online). It has been observed that transcription factors are preferentially retained after whole-genome duplication (WGD) (Xu and Messing 2008; Murat et al. 2010), with a recent study

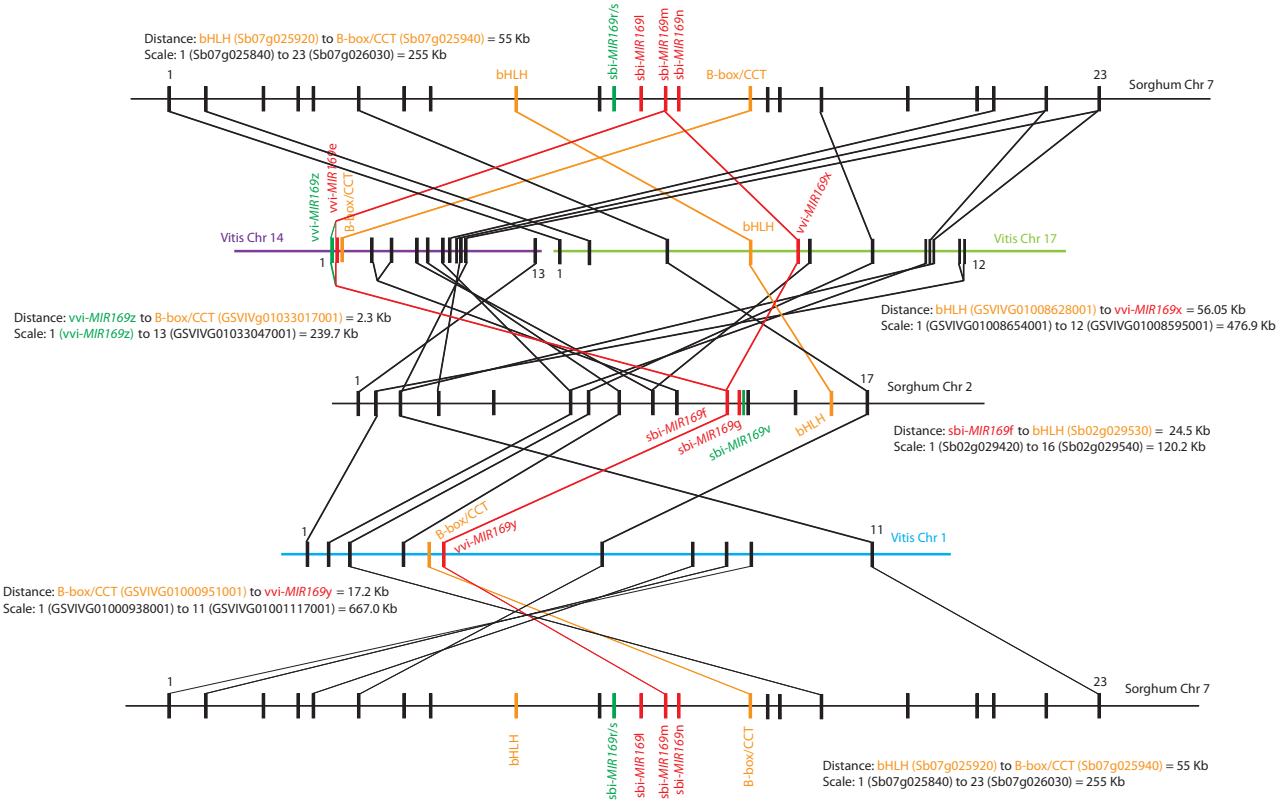

Fig. 9.—Conservation of synteny between sorghum and grapevine chromosomal segments containing *MIR169* gene copies. Sorghum segments containing *MIR169* gene clusters from chr2 and chr7 were aligned to the grapevine genome based on orthologous gene pairs. Because grapevine is a hexopaleo-polyploid, we found a 2:3 chromosomal relationship between sorghum and grapevine. Collinearity allowed the identification of a new *MIR169* copy (vvi-*MIR169*z) in grapevine chr14. Different grapevine chromosomes are represented in colors, whereas sorghum chromosomes are in black. Relative to sorghum chr2, grapevine had an inversion event on chr14 and chr17. The association of *MIR169* with its flanking COL gene was maintained on grapevine chr14 and chr1, whereas the association of *MIR169* with the bHLH gene was maintained on chr1.

showing that from 2,943 sorghum–maize syntenic shared genes, 43% of them were retained as homoeologous pairs in maize, from which transcription factors were 4.3 times more frequently among retained genes than other functions (Woodhouse et al. 2010).

Alignment of sorghum regions containing *MIR169* gene copies on chr2 and chr7 with their respective collinear regions from *Brachypodium*, rice, foxtail millet, and maize revealed the close linkage of *MIR169* gene copies with their flanking *COL14* and bHLH genes in all five grasses examined. Furthermore, collinearity of *MIR169* gene copies with either the *COL14* and/or the bHLH genes extended to dicot species such as grapevine, soybean, and cassava. Previously, it was suggested that conservation of collinearity between monocot and dicot species is rather rare because of the dynamic genomic rearrangements in genomes over 130–240 Ma (Wolfe et al. 1989; Jaillon et al. 2007). Still, conservation of synteny between rice and grapevine was also previously observed (Tang et al. 2010). Therefore, we hypothesized that preservation of collinearity in rare cases was subject to selection even after WGD events. In support of this hypothesis, the pseudofunctionalization and higher protein divergence rate of the HLH gene in *Brachypodium* chr4, where the *MIR169* cluster was deleted, occurred in comparison to the orthologous bHLH copy on chr3 with the *MIR169*e and *MIR169*g copies next to it. Indeed, trade-offs between sugar content and flowering time/plant height were reported in sorghum (Murray et al. 2008). When two genes controlling linked phenotypes are in close proximity on the chromosome for selection to act on both of them, the loss of one gene releases selection pressure on the other gene, allowing it to diverge. On the basis of its similarity to *Arabidopsis* bHLH137, which was postulated as putative DELLA target gene that functions in the GA response pathway (Zentella et al. 2007), we hypothesize that the grass homolog may function either in flowering and/or plant height, which future research will have to confirm. On the other hand, the importance of COL family proteins in the regulation of flowering time is well known (Griffiths et al. 2003; Wenkel et al. 2006). Collinearity between sorghum and grapevine revealed the tight association of *COL14* with vvi-*MIR169*z and vvi-*MIR169*e on grapevine chr14, with the three genes contained within a 2.3 Kb interval.

Furthermore, *COL14* has been recently considered a candidate gene for a flowering QTL in grapevine (Duchêne et al. 2012). With such a short physical distance between a flowering time gene and two *MIR169* gene copies, it is tempting to propose that grapevine breeding for late or early flowering time could have brought different *COL14* alleles together with its neighboring *MIR169* genes, a process known as linkage drag. Interestingly, although we could not find extensive collinearity between sorghum and *Arabidopsis thaliana* as to draw a synteny graph, we did find a close association on chr5 between COL4 gene and ath-*MIR169*b, separated each other 61.7 kb (data not shown).

On the basis of these considerations, we can propose a hypothesis where the linkage of *MIR169* gene copies with the neighboring COL gene could have coevolved (supplementary fig. S11, Supplementary Material online). This hypothesis is based on the findings presented here, together with a previous report describing that CO and COL proteins can interact through their CCT domains with proteins belonging to the NF-Y (HAP) family of transcription factors (Wenkel et al. 2006); specifically, it was described that CO together with COL15 interacted with NF-YB and NF-YC displacing NF-YA from the ternary complex. The mRNAs encoded by the NF-YA gene family are known targets of miR169 (Li et al. 2008). Thus, the association on the chromosome of a COL gene with a *MIR169* gene or gene cluster would ensure that miR169 would reduce the expression of the NF-YA mRNA and thus its protein levels, so that the COL protein can replace NF-YA in the ternary complex and drive transcription of CC AAT box genes. Furthermore, this hypothesis could provide a genetic framework where to test the previously known drought and flowering trade-offs: When plants are exposed to drought stress during the growing season, they flower earlier than control plants under well-watered environments (Franks et al. 2007), with the response being genetically inherited. For this reason, we decided to term our model the "Drought and Flowering Genetic Module Hypothesis."

We can envision a prominent role of linkage drag in breeding sorghum for enhanced biofuel traits such as high sugar content in stems and late flowering time for increased biomass. Under the *MIR169*-bHLH and/or *MIR169*-COL linkage drag model, any breeding scheme in sweet sorghum whose aim is to increase plant biomass through delayed flowering by crossing cultivars with different COL and/or bHLH alleles on either chr7 or chr2, respectively, should take into account the allelic variation at the neighboring *MIR169* gene copies as they may affect sugar content in stems and drought tolerance. The same can be said in breeding sorghum for grain production where the norm is to increase germplasm diversity among grain sorghums through the introduction of dwarf and early flowering genes from a donor line into exotic tall and late flowering lines with African origins (Brown et al. 2008).

On the basis of our results from comparative genomics analysis, we envision that any conservation in collinearity between closely associated genes (in this particular study between a microRNA and a protein-coding gene) controlling related phenotypes that is conserved among several plant species might be subject to linkage drag through breeding, opening a new area of research in genomics assisted breeding. In support of this notion, the early development of conserved ortholog set markers (referred as COS markers) among different plant species (Fulton et al. 2002) highlighted the existence of a set of genes with synteny conservation because of the early radiation of dicotyledonous plants that can be used in mapping through comparative genomics. In addition, conservation in linkage between candidate genes for seed glucosinolate content and SSR markers between *Arabidopsis* and oilseed rape (*Brassica napus* ssp. *napus*) were used in marker-assisted selection in breeding oilseed rape for total glucosinolate content (Hasan et al. 2008).

Supplementary Material

Supplementary figures S1–S11 and table S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org).

Acknowledgments

The authors are thankful to Jian-Hong Xu for providing valuable advice in creating synteny graphs and Yongrui Wu for providing RNA from maize endosperm tissue to prepare small RNA libraries. They thank Pinal Kanabar and the Waksman Genomics Core Facility for the service of SOLiD sequencing of maize small RNAs. This work was supported by the Selman A. Waksman Chair in Molecular Genetics to J.M. and in part by the sponsorship from the Institute of International Education (IIE) and the Fulbright Commission in Uruguay to M.C.

Literature Cited

Allen E, et al. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet. 36: 1282–1290.

Axtell MJ, Bowman JL. 2008. Evolution of plant microRNAs and their targets. Trends Plant Sci. 13:343–349.

Bennetzen JL, et al. 2012. Reference genome sequence of the model plant *Setaria*. Nat Biotechnol. 30:555–561.

Brachypodium-Sequencing-Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768.

Brown PJ, Rooney WL, Franks C, Kresovich S. 2008. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637.

Calvino M, Bruggmann R, Messing J. 2008. Screen of genes linked to high-sugar content in stems by comparative genomics. Rice 1: 166–176.

Calvino M, Bruggmann R, Messing J. 2011. Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. BMC Genomics 12:356.

Calvino M, Messing J. 2011. Sweet sorghum as a model system for bioenergy crops. Curr Opin Biotechnol. 23:1–7.

Calvino M, Miclaus M, Bruggmann R, Messing J. 2009. Molecular markers for sweet sorghum based on microarray expression data. Rice 2: 129–142.
Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 39:W155–W159.
Duchêne E, Butterlin G, Dumas V, Merdinoglu D. 2012. Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. Theor Appl Genet. 124:623–635.
Fahlgren N, et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. PloS One 2:e219.
Fenselau de Felippes F, Schneeberger K, Dezulian T, Huson DH, Weigel D. 2008. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. RNA 14:2455–2459.
Fernandez MGS, Becraft PW, Yin Y, Luebberstedt T. 2009. From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci. 14:454–461.
Franks SJ, Sim S, Weis AE. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc Natl Acad Sci U S A. 104:1278–1282.
Fulton T, Van der Hoeven R, Eannetta N, Tanksley S. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467.
Griffiths S, Dunford RP, Coupland G, Laurie DA. 2003. The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*. Plant Physiol. 131:1855–1867.
Hasan M, et al. 2008. Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). Theor Appl Genet. 116:1035–1049.
Jaillon O, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467.
Jiang D, et al. 2006. Duplication and expression analysis of multicopy miRNA gene family members in *Arabidopsis* and rice. Cell Res. 16:507–518.
Li WX, et al. 2008. The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell 20:2238–2251.
Ma Z, Coruh C, Axtell MJ. 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. Plant Cell 22:1090–1103.
Maher C, Stein L, Ware D. 2006. Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. 16:510–519.
Meng Y, Shao C, Gou L, Jin Y, Chen M. 2011. Construction of microRNA- and microRNA*-mediated regulatory networks in plants. RNA Biol. 8: 1124–1148.
Messing J, et al. 2004. Sequence composition and genome organization of maize. Proc Natl Acad Sci U S A. 101:14349–14354.
Meyers BC, et al. 2008. Criteria for annotation of plant microRNAs. Plant Cell 20:3186–3190.
Murat F, et al. 2010. Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res. 20:1545–1557.
Murray SC, et al. 2008. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48:2165–2179.
Nozawa M, Miura S, Nei M. 2012. Origins and evolution of microRNA genes in plant species. Genome Biol Evol. 4:230–239.
Paterson AH, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556.
Piriyapongsa J, Jordan IK. 2008. Dual coding of siRNAs and miRNAs by plant transposable elements. RNA 14:814–821.
Sun J, Zhou M, Mao Z, Li C. 2012. Characterization and evolution of microRNA genes derived from repetitive elements and duplication events in plants. PloS One 7:e34092.
Swigonova Z, et al. 2004. Close split of sorghum and maize genome progenitors. Genome Res. 14:1916–1923.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.
Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci U S A. 107:472–477.
Toledo-Ortiz G. 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15:1749–1770.
Valverde F. 2011. CONSTANS and the evolutionary origin of photoperiodic timing of flowering. J Exp Bot. 62:2453–2463.
Wenkel S, et al. 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell 18:2971–2984.
Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci U S A. 86:6201–6205.
Woodhouse MR, et al. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol. 8:e1000409.
Xu J-H, Messing J. 2008. Diverged copies of the seed regulatory opaque-2 gene by a segmental duplication in the progenitor genome of rice, sorghum, and maize. Mol Plant. 1:760–769.
Xue L-J, Zhang J-J, Xue H-W. 2009. Characterization and expression profiles of miRNAs in rice seeds. Nucleic Acids Res. 37:916–930.
Yang JS, et al. 2011. Widespread regulatory activity of vertebrate microRNA* species. RNA 17:312–326.
Zentella R, et al. 2007. Global analysis of DELLA direct targets in early gibberellin signaling in *Arabidopsis*. Plant Cell 19: 3037–3057.
Zhang G, et al. 2012. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotechnol. 30:549–554.
Zhang L, et al. 2009. A genome-wide characterization of microRNA genes in maize. PLoS Genet. 5:e1000716.

Associate editor: Eugene Koonin