Sequence Analysis of Bacterial Artificial Chromosome Clones from the Apospory-Specific Genomic Region of *Pennisetum* and *Cenchrus*[1][W][OA]

Joann A. Conner, Shailendra Goel[2], Gunawati Gunawan, Marie-Michele Cordonnier-Pratt, Virgil Ed Johnson, Chun Liang[3], Haiming Wang[4], Lee H. Pratt, John E. Mullet, Jeremy DeBarry, Lixing Yang, Jeffrey L. Bennetzen, Patricia E. Klein, and Peggy Ozias-Akins*

Department of Horticulture, University of Georgia, Tifton, Georgia 31793–0748 (J.A.C., S.G., G.G., P.O.-A.); Department of Plant Biology (M.-M.C.-P., C.L., H.W., L.H.P.), Department of Genetics (J.D., L.Y., J.L.B.), and Office of Research Services (V.E.J.) University of Georgia, Athens, Georgia 30602; and Department of Plant Biology Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, Texas 77843 (J.E.M., P.E.K.)

Apomixis, asexual reproduction through seed, is widespread among angiosperm families. Gametophytic apomixis in *Pennisetum squamulatum* and *Cenchrus ciliaris* is controlled by the apospory-specific genomic region (ASGR), which is highly conserved and macrosyntenic between these species. Thirty-two ASGR bacterial artificial chromosomes (BACs) isolated from both species and one ASGR-recombining BAC from *P. squamulatum*, which together cover approximately 2.7 Mb of DNA, were used to investigate the genomic structure of this region. Phrap assembly of 4,521 high-quality reads generated 1,341 contiguous sequences (contigs; 730 from the ASGR and 30 from the ASGR-recombining BAC in *P. squamulatum*, plus 580 from the *C. ciliaris* ASGR). Contigs containing putative protein-coding regions unrelated to transposable elements were identified based on protein similarity after Basic Local Alignment Search Tool X analysis. These putative coding regions were further analyzed in silico with reference to the rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*) genomes using the resources at Gramene (www.gramene.org) and Phytozome (www.phytozome.net) and by hybridization against sorghum BAC filters. The ASGR sequences reveal that the ASGR (1) contains both gene-rich and gene-poor segments, (2) contains several genes that may play a role in apomictic development, (3) has many classes of transposable elements, and (4) does not exhibit large-scale synteny with either rice or sorghum genomes but does contain multiple regions of microsynteny with these species.

Apomixis is a naturally occurring mode of asexual reproduction in angiosperms that leads to embryo and seed formation without a requirement for meiosis or fertilization of the egg. The application of apomixis in plant breeding could have tremendous impact by providing a mechanism for hybrid progeny to avoid segregation and fix heterosis. Although apomixis has been reported in over 300 species in at least 35 angiosperm families, 75% of the apomictic species reported have been from the Poaceae, Rosaceae, and Asteraceae families. Apomixis occurs almost exclusively in polyploid genotypes. Classification of the apomictic mode of reproduction as either sporophytic or gametophytic is based on the developmental origin of the cell from which the embryo is derived (Nogler, 1984; Koltunow, 1993). In adventitious embryony, a sporophytic process common in citrus, embryos develop through mitotic division of somatic cells of the ovule. Sexual reproduction is usually not disrupted, leading frequently to polyembryony (Koltunow et al., 1995). In gametophytic apomixis, common to grasses, an unreduced embryo sac can be formed through two different mechanisms, either diplospory or apospory. In diplospory, the unreduced embryo sac is derived from the megaspore mother cell, which either circumvents meiosis or begins meiosis, but subsequently reverts back to a state that enables mitosis to occur. In apospory, the unreduced embryo sac is initiated from nearby nucellar cells. In both diplospory and apospory, one cell of the unreduced embryo sac will become the egg cell and begin to divide mitotically to form the embryo that is genetically identical to the maternal plant. In order to form a fully mature seed, parthenogenetic development of the egg cell must be accompanied by endosperm formation.

[1] This work was supported by the National Science Foundation (grant no. 0115911) and the University of Georgia Experiment Station.

[2] Present address: Department of Botany, University of Delhi, Delhi, India 110007.

[3] Present address: Department of Botany, Miami University, Oxford, OH 45056.

[4] Present address: Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA 30602.

* Corresponding author; e-mail pozias@uga.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Peggy Ozias-Akins (pozias@uga.edu).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.119081

The endosperm develops autonomously from the central cell in some apomicts, but most require pollination and probably fertilization of the central cell (pseudogamy).

A single dominant ''locus'' is required for apomeiosis and parthenogenesis in the aposporous grasses *Pennisetum/Cenchrus* (Sherwood et al., 1994; Ozias-Akins et al., 1998), *Brachiaria* (Pessino et al., 1997), and *Paspalum* (Martinez et al., 2001) and in the diplosporous grass *Tripsacum* (Grimanelli et al., 1998). In all cases, penetrance of the trait varied among the apomictic siblings. Inheritance of diplospory in the Asteraceae is more complex, with two loci being required for diplosporous embryo sac development and parthenogenic development of the seed in both *Erigeron* (Noyes and Rieseberg, 2000) and *Taraxacum* (van Dijk et al., 1999). In all mapping populations studied except *Taraxacum*, suppression of recombination has been noted either through the identification of multiple linked markers or through the lower recombination rate in apomicts among apomixis-linked markers compared with homologous or syntenic regions of sexual relatives.

In *Pennisetum* and *Cenchrus*, many markers specific to apospory were found to be hemizygous. Given the hemizygosity, as well as genetic data suggesting that the region responsible for apospory may be physically large, the term ASGR (for apospory-specific genomic region) was coined to describe the locus in *Pennisetum/Cenchrus*. Ninety-nine bacterial artificial chromosome (BAC) clones containing molecular markers (SCARs, RFLPs, and AFLPs) showing total genetic linkage to the aposporous phenotype have been identified from BAC libraries constructed from the apomictic polyhaploid line MS228-20 and the apomictic *Cenchrus ciliaris* line B-12-9 (Roche et al., 2002; Akiyama et al., 2004; Goel et al., 2006). These BACs will be referred to as ASGR-BACs. One BAC clone containing an AFLP marker segregating approximately 2 cM from the aposporous phenotype in *Pennisetum squamulatum* has also been identified (Goel et al., 2006) and will be referred to as the ASGR-recombinant BAC. The ASGR and ASGR-recombinant BAC clones were fingerprinted and grouped based on fingerprinted contig (FPC) analysis (Roche et al., 2002; Goel et al., 2006). These analyses demonstrated that portions of the ASGR are composed of duplicated regions and that the ASGR between the two species is highly conserved.

Subsets of the ASGR-BAC clones and the ASGR-recombinant BAC have been used for fluorescence in situ hybridization (FISH) analysis in both species. In *P. squamulatum*, FISH analysis has confirmed that the ASGR is physically large (approximately 50 Mb) and is located near the telomere on the short arm of the ASGR-carrier chromosome. The ASGR has also been categorized as hemizygous and heterochromatic in nature (Goel et al., 2003, 2006; Akiyama et al., 2004). A centromere-derived repeat signal found at the distal end of the ASGR in *P. squamulatum* suggests that a possible inversion of the ASGR-carrier chromosome arm occurred. In *C. ciliaris*, the ASGR-carrier chromosome is approximately 20 Mb larger than its presumed homo(eo)logous chromosomes (Akiyama et al., 2005), and the ASGR is located near the centromere in a region that is hemizygous and heterochromatic (Goel et al., 2003, 2006; Akiyama et al., 2004). The physical size of the ASGR within *C. ciliaris* has not been determined due to the current lack of BACs containing recombining, flanking markers. In both species, the ASGR contains an approximately 13 Mb ''low-copy'' region identified by ASGR-BAC clones that give distinct FISH signals. The low-copy region is flanked on both sides by a ''high-copy'' region identified by ASGR-BAC clones that give an intense dispersed FISH signal. In *P. squamulatum*, this high-copy repeat signal is found only at the ASGR. In *C. ciliaris*, the high-copy repeat signal flanks the low-copy ASGR region and can also be identified on all other chromosomes (Akiyama et al., 2004, 2005). An *Opie-2*-like retrotransposon has been identified that can mimic the high-copy repeat signal given by BAC clones (Akiyama et al., 2004). Comparison of BAC order across the low-copy region of the ASGR between the two species shows that the ASGR is inverted but macrosyntenic. Macrosynteny outside of the ASGR between the species is unlikely, as the BAC clone found to be recombinant in *P. squamulatum* does not localize to the ASGR-carrier chromosome in *C. ciliaris* (Goel et al., 2006). Detailed work on the UGT197 ASGR-BAC clones, which contain the ASGR-specific SCAR marker UGT197 (Ozias-Akins et al., 1998), determined that on a smaller scale segments of similar gene order exist between the two apomictic species and that the segments could be found multiple times within the ASGR in both species. This work also identified homology and colinearity of the UGT197 ASGR-BACs to a centromere-proximal region of chromosome 11 in rice (*Oryza sativa*; Gualtieri et al., 2006).

Other than the *Opie-2*-like retrotransposon and the limited sequences generated from the UGT197 ASGR-BACs, very little is known about the sequence composition of the ASGR in *Pennisetum/Cenchrus* and whether the colinearity between the UGT197 ASGR-BACs and rice chromosome 11 would extend throughout the ASGR. To further investigate this region, 32 ASGR-BACs, 18 from *P. squamulatum* and 13 from *C. ciliaris*, in addition to one ASGR-recombinant BAC located approximately 2 cM from the ASGR in *P. squamulatum*, were shotgun cloned and sample sequenced at approximately 0.5× coverage. Some additional targeted sequences were generated from BACs containing putative protein coding regions (PPCRs) identified by BLASTX analysis of the sample sequences to multiple protein databases. We generated approximately 2.5 Mb of data from 4,521 high-quality sequencing reads. After Phrap assembly, 1,341 sequence contigs (730 contigs from the *P. squamulatum* ASGR, 581 contigs from the *C. ciliaris* ASGR, and 30 contigs from the *P. squamulatum* ASGR-recombining BAC), covering approximately 1.0 Mb, were obtained.

Twenty-five *C. ciliaris* and 23 *P. squamulatum* ASGR-PPCRs were discovered through similarity to known proteins along with five ASGR-recombinant PPCRs from *P. squamulatum*. The protein function of the ASGR-PPCRs identified varied widely, with many showing similarity to proteins with functional domains that are known to bind DNA and/or alter DNA transcription and hence could be involved in the apomictic pathway. The strongest candidate ASGR-PPCR identified was the *ASGR-BABY BOOM* (*BBM*)*-like* genes. This study demonstrates that the colinearity previously identified between the UGT197 SCAR-containing BACs and rice chromosome 11 does not extend throughout the ASGR; instead, multiple small regions of shared synteny to the rice and sorghum (*Sorghum bicolor*) genomes exist throughout the ASGR.

RESULTS

ASGR Sample and Targeted Sequencing

Shotgun libraries were constructed from 32 ASGR-BAC clones isolated in our laboratory and from one ASGR-recombining BAC clone (Table I). The 13 BACs designated c__ were isolated from the *C. ciliaris* line B-12-9 library, while the 20 BACs designated p__ were isolated from the polyhaploid apomict library (Roche et al., 2002). In all, approximately 2.7 Mb from the ASGR was shotgun cloned (Table I). With the exception of ASGR-BACs p102 and c004/c014, we chose to sample sequence BACs not considered orthologs between the two species. BACs were considered orthologous between the two species if FPC analysis (settings: tolerance 7, stringency 10^{-12}) ordered them within the same contig. Additional data from this work (see "Materials and Methods") allowed us to modify the FPC grouping of Goel et al. (2006) slightly by combining FPC contigs 7 and 10 and by placing the singleton p801 within FPC contig 6.

Shotgun subclone libraries were sample sequenced at a depth of approximately 0.5× coverage. A total of 2,055 high-quality sequences were obtained from the 33 BAC subclone libraries. Sequence reads derived from an individual BAC clone or from a group of overlapping BAC clones were given a unique Phrap group identification number as shown in Table I. These grouped BAC-derived sequences were then assembled into sequence contigs by Phrap. During Phrap assembly of individual sequences into contigs, two numeric identifiers were added to the Phrap group identification number to generate a unique name for each sequence contig. When an FPC group contained BACs sequenced from both species, the sequences from each species were grouped separately (Phrap assemblies 11 and 22). All generated sequence contigs were analyzed for homology to other known proteins using BLASTX against an internal PIR_NREF database at FUNGEN (www.fungen.org). If a sequence contig contained a BLASTX e-value hit of $\leq 10^{-6}$ to a protein unrelated to transposable elements and if the sequence was generated from a BAC subclone library produced at the University of Georgia (UGA; Table I), targeted sequencing data were generated (see "Materials and Methods"). After 2,466 additional targeted high-quality sequences were generated, the random and targeted sequences were assembled again by Phrap into 1,341 sequence contigs ranging in size from 100 to 8,521 bp. Each sequence contig was given a Uniscript name for database reference. All high-quality reads have been deposited in GenBank dbGSS ED544199 to ED548719. Individual sequences can also be accessed through the FUNGEN ASGR database at http://asgr.uga.edu. A detailed description of sequence processing and analysis can be found in Cordonnier-Pratt et al. (2004).

Putative Protein-Coding Regions on ASGR and ASGR-Recombining BACs

Eighty-seven sequence contigs (6.5%) contained similarity to proteins from multiple species unrelated to transposable elements upon BLASTX analysis to the UniprotTrEMBL database (version 7) with an e-value hit of $\leq 10^{-6}$. These sequence contigs were tagged for further analysis. The most significant hit for these 87 sequence contigs can be found in Supplemental Table S1. BLASTX results for all 1,341 sequence contigs against the UniprotTrEMBL database can be accessed through the FUNGEN ASGR database at http://asgr.uga.edu. Given that sample sequencing was combined with targeted sequencing to increase coverage of the protein-coding region sequences, we expected that we could generate multiple but not necessarily overlapping sequence contigs within the same Phrap group that would be associated with the same protein-coding region. Indeed, of the 87 contigs that had BLASTX e-values of $<10^{-6}$ to the UniprotTrEMBL database, only 54 unique protein hits were identified. To make the analysis of the 87 sequence contigs containing PPCRs more uniform, each sequence contig was individually reanalyzed using the BLASTX program at Gramene (version 25.0; http://www.gramene.org/) against The Institute for Genomic Research (TIGR) rice gene model protein database. Twelve sequence contigs were removed from further analysis due to similarity to repetitive elements based on Gramene classification, and one sequence contig was removed for lack of significant (e-value $\leq 10^{-6}$) similarity to a rice protein. The remaining 74 sequence contigs with PPCRs are identified in Tables II to IV. To simplify comparison with the rice protein, if multiple sequence contigs from the same Phrap grouping had corresponding best hits to the same rice protein, the sequence contig with the most significant hit was used in further comparisons with rice. If a sequence contig contained two PPCRs based on similarity to separate rice proteins, both protein similarities are described and ordered numerically as they appear on the sequence contig. Tables II and III identify the ASGR-BAC

Table I. *ASGR and ASGR-recombinant BACs analyzed*

Asterisks indicate BAC orthologs between the two species. Shotgun libraries were created and sample sequenced at TAMU (underlined) or UGA (boldface). n/a, Not assayed.

Phrap Group Identification No.	BACs Sequenced	FISH Signal Location within ASGR	Approximate Size (kb) of BAC Insert Cloned into Shotgun Libraries	Molecular Marker Used for BAC Isolation	No. of High-Quality Sequences Submitted to GenBank GSS Database	No. of Phrap Contigs Generated	No. of Putative Protein Coding Regions
10	c002	Low copy	112	Q8M	47	31	4
11*	c004* and **c014***	Low copy	145	Q8M	432	61	6
22*	p102*	Low copy	133	Q8M	71	50	2
12	**c018**	n/a	109	Q8M walk	83	46	2
13	c100 and **c111**	Low copy	195	UGT197	639	116	6
14	**c1000**	Low copy	79	HHU27	267	38	2
15	c108	Low copy	70	UGT197	75	52	2
16	c205	n/a	110	C4	100	53	0
17	c201	n/a	121	C4	52	45	0
18	c501 and **c522**[a]	n/a	145	O7M	282	82	2
19	**c801**	n/a	95	M02	155	56	1
20	p002 and p003 and **p004**	Low copy	175	A14	169	93	2
21	**p1000**	Low copy	88	py503	178	75	1
23	**p104**	Low copy	85	Q8M	229	55	3
24	**p1100**	High copy	82	px299	41	36	0
25	**p1200**	Low copy	82	pa265	109	35	3
27	p201 and p207 and p208	Low copy	145	UGT197	460	69	2
28	p301	Low copy	125	A10	24	13	0
29	**p303**	Low copy	75	A10	132	47	2
30	p506	Low copy	85	O7M	62	54	1
31	p602	High copy	82	X18	83	64	0
32	p708	n/a	107	R13	68	40	1
33	**p800** and **p801**[a]	High copy	125	U12H	374	53	6
34	p900	High copy	82	W10M	68	46	0
26	**p1300**[b]	Outside ASGR	82	pq355	321	30	5

[a]Modified FPC data of BACs based on current work. [b]ASGR-recombinant BAC.

sequence contigs containing nonrepetitive element PPCRs derived from *C. ciliaris* and *P. squamulatum*, respectively. Table IV identifies the ASGR-recombinant BAC sequence contigs containing nonrepetitive element PPCRs derived from *P. squamulatum*. The predicted number of amino acids contained within the ASGR- and ASGR-recombinant PPCRs (Tables II–IV, column 6) was determined by counting the number of amino acids within the PPCRs that aligned with the corresponding rice protein from the BLASTX output. The predicted size of the corresponding rice protein, as determined by Gramene, is shown in parentheses after the Gramene protein description. The overall percentage similarity of the PPCR (Tables II–IV, column 7) to the aligned regions of the rice protein was based on calculating amino acid length and percentage similarity of all BLASTX alignments for each sequence contig.

Duplication of ASGR and ASGR-Recombinant PPCRs

Previous analysis of ASGR-BAC clones suggested that regions within the ASGR regardless of species were duplicated. Sample and targeted gene sequencing allowed for the analysis of sequence contigs with PPCR duplications within a Phrap grouping. Intra-Phrap group PPCR duplications were considered verified if the PPCR was present in two or more sequence contigs and a nucleotide polymorphism was detected either by RFLP analysis or by having two or more sequences for each sequence contig as well as an error P value of less than 0.0001 for a putative nucleotide polymorphism. Four intra-Phrap group duplications were identified. Sequence contigs 14-2-24 and 14-2-25 are 99.6% identical over 6,593 bp; sequence contigs 33-2-33 and 33-2-35 are 99.7% identical over 1,180 bp; sequence contigs 11-2-38 and 11-2-30 are 99.1% identical over 570 bp; while sequence contigs 11-2-38 and 11-2-36 are 99.8% identical over 1,000 bp. As full-length sequencing of the PPCRs was not always accomplished, it is unknown whether the intra-Phrap group duplications identified would extend over the complete gene or whether we are finding remnants of duplication. While Gualtieri et al. (2006) was able to distinguish two copies of the *ASGR-BBM-like* gene within Phrap group 27 by fragment hybridization, the initially generated sequences from this study did not allow for the identification of two distinct copies of the *ASGR-BBM-like* gene until further sequence analysis was done (see below).

Table II. *Putative protein-coding regions identified in C. ciliaris ASGR sequence contigs*

Underlined sequence contigs contain PPCRs with 80% or greater amino acid coverage compared with the rice protein.

Sequence Contig	Uniscript Identifier	Lowest BLASTX e-Value[a]	TIGR Locus Identifier[b]	TIGR Protein Description and Predicted Size (in Amino Acids)	Predicted No. of Amino Acids in the PPCR[c]	Predicted Amino Acid Identity (%) of the PPCR to the Corresponding TIGR Protein[d]
10-0-18	0_18	10^{-8}	LOC_Os04g46490.1	Aquaporin TIP5.1, putative, expressed (269)	45	62
10-0-35	0_35	10^{-9}	LOC_Os02g56100.1	Ribonucleoside-diphosphate reductase large subunit, putative, expressed (810)	32	84
10-0-36	0_36	10^{-18}	LOC_Os06g37620.1	ATP-binding protein, putative, expressed (806)	102	55
10-2-2	2_2	10^{-22}	LOC_Os06g37610.1	Cyclopropane-fatty acyl-phospholipid synthase, putative, expressed (322)	66	84
11-2-12	2_20	10^{-7}	LOC_Os02g42060.1	Two-component response regulator ARR3, putative, expressed (147)	26	92
11-2-31	2_39	10^{-62}	LOC_Os03g56920.1	Conserved hypothetical protein (678)	364	38
11-2-34	2_42	10^{-27}	LOC_Os02g42270.1	T-cell activation protein phosphatase 2C-like protein, putative (315)	285	41
11-2-35	2_43	10^{-54}	LOC_Os01g24240.1	Hypothetical protein (755)	345	33
11-2-36	2_44					
11-2-38	2_46	10^{-54}	(1) LOC_Os06g35700.1	(1) Reticuline oxidase precursor, putative, expressed (528)	173	67
11-2-30	2_38	10^{-32}	(2) LOC_Os04g44280.1	(2) OsRR5, rice type A response regulator, expressed (134)	98	69
11-2-36	2_44					
12-2-16	2_62	10^{-34}	LOC_Os06g37610.1	Cyclopropane-fatty acyl-phospholipid synthase, putative, expressed (322)	89	81
12-0-25	0_103					
12-1-1	0_117					
12-2-6	2_52	10^{-9}	LOC_Os02g34430.1	BRASSINOSTEROID INSENSITIVE1-associated receptor kinase 1 precursor, putative, expressed (506)	43	79
13-2-1	2_63	10^{-26}	LOC_Os06g05790.1	Transferase, putative, expressed (600)	106	66
13-2-3	2_65	10^{-17}	LOC_Os11g19210.1	β-D-Xylosidase, putative, expressed (782)	74	66
13-0-5	0_122					
13-2-35	2_97	10^{-12}	LOC_Os06g34260.1	Hypothetical protein (113)	74	57
13-2-40	2_102	10^{-10}	LOC_Os07g13240.1	Hypothetical protein (164)	128	37
13-2-54	2_116	10^{-54}	LOC_Os11g18690.1	β-Xylosidase, putative (793)	211	63
13-2-60	2_122	10^{-71}	LOC_Os11g19060.1	Protein BBM1, putative, expressed (564)	226	70
13-2-61	2_123					
13-2-56	2_118					
14-0-9	0_207	10^{-8}	LOC_Os07g13470.1	Hydrolase/protein Ser/Thr phosphatase, putative, expressed (486)	87	46
14-2-25	2_149	10^{-68}	LOC_Os02g06340.1	EH domain-containing protein 1, putative, expressed (543)	148	89
14-2-24	2_148					
14-2-22	2_146					
15-0-12	0_235	10^{-40}	LOC_Os11g19210.1	β-D-Xylosidase, putative, expressed (782)	155	59
15-2-1	2_150					
15-2-15	2_164	10^{-9}	LOC_Os01g65150.1	Expressed protein (537)	101	49

(*Table continues on following page.*)

Table II. (*Continued from previous page.*)

Sequence Contig	Uniscript Identifier	Lowest BLASTX e-Value[a]	TIGR Locus Identifier[b]	TIGR Protein Description and Predicted Size (in Amino Acids)	Predicted No. of Amino Acids in the PPCR[c]	Predicted Amino Acid Identity (%) of the PPCR to the Corresponding TIGR Protein[d]
18-2-37	2_232	10^{-57}	LOC_Os10g38710.1	Glutathione *S*-transferase GSTU6, putative, expressed (233)	224	57
18-2-16	2_211					
18-2-39	2_234	10^{-48}	LOC_Os03g01210.1	Domain of unknown function DUF614-containing protein, expressed (254)	226	56
19-2-19	2_253	10^{-135}	LOC_Os04g31120.1	Influenza virus NS1A-binding protein isoform 3, putative, expressed (375)	332	74

[a]Results based on BLASTX from http://www.gramene.org/Multi/blastview compared with the peptide (TIGR gene model) database (version 25). [b]The Tigr_gene translation identifier of the most significant BLASTX hit for the boldface sequence contigs. [c]The number of amino acids from the sequence contig that aligned to the Tigr_gene translation identifier protein based on the BLASTX outputs. [d]Similarity based on calculating the amino acid length and percentage similarity of all BLASTX alignments for each PPCR to the Tigr_gene translation identifier.

Analysis of potential duplicate sequence contigs of PPCRs between different Phrap groupings within a species by sample sequence analysis alone is harder to assess due to the complexity of the region. For example, sequence contigs 10-2-2 and 12-2-16 contain similarity to the same rice gene and show 100% sequence similarity over the length of the 10-2-2 sequence. The c018 BAC was isolated using ASGR-specific primers derived from an end clone sequence of ASGR-BAC c001. BAC c001 is contained in the same FPC contig as the sequenced c002 BAC found in Phrap group 10. However, as FPC did not group the c018 BAC with the BACs from the c001/c002 group, and as only one PPCR was identified that was similar between the two BACs, it is unclear without extensive BAC sequencing whether this PPCR is a true duplication or the same gene on overlapping BACs. The duplication of the β-D-xylosidase-like protein on sequence contigs 15-0-12 and 13-2-3 reported by Gualtieri et al. (2006) was also identified by this sequence analysis.

Whole gene sequencing data for the analysis of PPCR conservation between the two apomictic species was not generated in this study except for the *ASGR-BBM-like* genes (see below). With the exception of Phrap groups 11 and 22, we chose to sample sequence BACs not considered orthologs between the two species to increase overall coverage across the ASGR. Additionally, if the same PPCR was identified between the two species, only one of the species was chosen for targeted gene sequencing.

ASGR-BBM-like Genomic Sequences

We chose to fully sequence the *ASGR-BBM-like* genes from both species. Four ASGR-BACs (c100, c102, p203, and p207) were identified that contained distinct copies of the *ASGR-BBM-like* genes. C102, considered by FPC analysis to be orthologous to the Phrap 27 BACs, and p203 were not previously sample sequenced. The *ASGR-BBM-like* genes were sequenced and contain approximately 300 bp upstream of the predicted start codon to the predicted stop codon. The *PsASGR-BBM-like1* gene derived from BAC p203 contains 3,826 bp (EU559280); the *PsASGR-BBM-like2* gene derived from BAC p207 contains 3,832 bp (EU559277); the *CcASGR-BBM-like1* gene derived from BAC c102 contains 3,835 bp (EU559278); and the *CcASGR-BBM-like2* gene contains 3,856 bp (EU559279). The four genes were aligned by ClustalW2, and the alignment is shown in Supplemental Figure S1. Table V shows the percentage of consensus positions of the four *ASGR-BBM-like* genes with each other using global alignment. The two *ASGR-BBM-like* genes from *P. squamulatum* share 99.8% identity with each other and differ only at two positions within the first predicted intron (see ClustalW2 alignment). The two *C. ciliaris ASGR-BBM-like* genes share 98.6% identity with each other. The differences between the two *C. ciliaris ASGR-BBM-like* genes are found in both the predicted coding and noncoding regions. When analyzed across the species, the *CcASGR-BBM-like1* gene is more similar to both *P. squamulatum ASGR-BBM-like* genes than to the *CcASGR-BBM-like2* gene, confirming the FPC analysis. Unlike the *P. squamulatum ASGR-BBM-like* genes, which are found on overlapping BACs in an FPC contig, the *C. ciliaris ASGR-BBM-like* genes are identified on separate FPC contigs.

The four *ASGR-BBM-like* genes were analyzed for the potential to be transcribed using the rice gene prediction program at RiceGAAS (http://ricegaas.dna.affrc.go.jp/usr/). All *ASGR-BBM-like* sequences were predicted to contain seven exons. *PsASGR-BBM-like1*, *PsASGR-BBM-like2*, and *CcASGR-BBM-like1* sequences were predicted to encode a 542-amino acid protein containing two AP2 domains. The predicted exons for these genes are highlighted in red in Sup-

Table III. *Putative protein-coding regions identified in P. squamulatum ASGR sequence contigs*

Underlined sequence contigs contain PPCRs with 80% or greater amino acid coverage compared with the rice protein.

Sequence Contig	Uniscript Identifier	Lowest BLASTX e-Value[a]	TIGR Locus Identifier[b]	TIGR Protein Description and Predicted Size (in Amino Acids)	Predicted No. of Amino Acids in the PPCR[c]	Predicted Amino Acid Identity (%) of the PPCR to the Corresponding TIGR Protein[d]
20-0-44	0_500	10^{-15}	LOC_Os03g55620.2	Ser/Thr protein kinase 12, putative, expressed (279)	40	93
20-2-16	2_269	10^{-29}	LOC_Os06g33220.1	Hypothetical protein (167)	149	48
21-2-30	2_331	10^{-14}	LOC_Os01g36950.3	Nitrogen-rich protein, putative, expressed (309)	127	54
22-0-5	0_584	10^{-27}	LOC_Os06g35700.1	Reticuline oxidase precursor, putative, expressed (528)	93	68
22-0-25	0_604	10^{-11}	LOC_Os02g42060.1	Two-component response regulator ARR3, putative, expressed (147)	95	37
23-0-3	0_620	10^{-42}	LOC_Os06g37620.1	ATP-binding protein, putative, expressed (806)	153	67
23-2-23	2_369	10^{-102}	(1) LOC_Os02g44080.1	(1) Aquaporin TIP2.1, putative, expressed (248)	248	81
			(2) LOC_Os04g46490.1	(2) Aquaporin TIP5.1, putative, expressed (269)	109	82
25-0-6	0_704	10^{-34}	LOC_Os11g48040.1	Mitochondrial uncoupling protein 3, putative, expressed (301)	127	65
25-2-14	2_389	10^{-41}	LOC_Os01g72990.1	ATP-binding protein, putative, expressed (952)	258	78
25-2-15	2_390					
25-2-3	2_378	10^{-18}	LOC_Os02g53150.1	Expressed protein (316)	201	29
27-2-35	2_440	10^{-9}	LOC_Os01g67410.1	AP2/EREBP transcription factor BBM, putative, expressed (692)	51	82
27-2-53	2_458	10^{-84}	LOC_Os11g19060.1	Protein BBM1, putative, expressed (564)	317	58
27-2-51	2_456					
27-2-49	2_454					
29-2-18	2_483	10^{-23}	LOC_Os04g01240.1	Expressed protein (206)	64	84
29-2-25	2_490	10^{-31}	LOC_Os04g01230.1	Phosphoglycerate mutase-like protein, putative, expressed (213)	121	79
29-2-22	2_487					
29-0-22	0_826					
29-2-6	2_471					
30-2-7	2_498	10^{-14}	LOC_Os03g01210.2	Domain of unknown function DUF614-containing protein, expressed (240)	55	62
32-0-22	0_954	10^{-10}	LOC_Os04g19960.1	Expressed protein (237)	65	49
33-2-29	2_561	10^{-30}	LOC_Os04g31030.1	Nitrate-induced NOI protein, expressed (224)	117	59
33-2-31	2_563	10^{-17}	LOC_Os08g41570.1	Hypothetical protein (618)	282	27
33-2-32	2_564	10^{-151}	LOC_Os04g31040.1	Violaxanthin deepoxidase, putative, expressed (468)	382	77
33-2-34	2_566	10^{-41}	(1) LOC_Os04g31050.1	(1) Expressed protein (408)	172	68
		10^{-165}	(2) LOC_Os04g31070.1	(2) Acyl-desaturase, chloroplast precursor, putative, expressed (390)	370	89
33-2-35	2_567	10^{-111}	LOC_Os04g31000.1	Spermine/spermidine synthase, putative, expressed (750)	366	72
33-0-11	0_973					
33-0-12	0_974					
33-1-2	0_1002					
33-2-24	2_556					
33-2-27	2_559					
33-2-33	2_565					

[a]Results based on BLASTX from http://www.gramene.org/Multi/blastview compared with the peptide (TIGR gene model) database (version 25). [b]The Tigr_gene translation identifier of the most significant BLASTX hit for the boldface sequence contig. [c]The number of amino acids from the sequence contig that aligned to the Tigr_gene translation identifier protein based on the BLASTX outputs. [d]Similarity based on calculating the amino acid length and percentage similarity of all BLASTX alignments for each PPCR to the Tigr_gene translation identifier.

Table IV. *Putative protein-coding regions identified in P. squamulatum ASGR-recombinant sequence contigs*

Underlined sequence contigs contain PPCRs with 80% or greater amino acid coverage compared with the rice protein.

Sequence Contigs Containing PPCRs	Uniscript Identifier	Lowest BLASTX e-Value[a]	TIGR Locus Identifier[b]	TIGR Protein Description and Predicted Size (in Amino Acids)	Predicted No. of Amino Acids in the PPCR[c]	Predicted Amino Acid Identity (%) of the PPCR to the Corresponding TIGR Protein[d]
26-1-3	0_768	10^{-23}	LOC_Os08g18200.1	Hypothetical protein (515)	130	47
26-2-10	2_400	10^{-122}	LOC_Os07g26630.1	Aquaporin PIP2.4, putative, expressed (306)	306	87
26-2-11	2_401					
26-2-12	2_402	10^{-38}	LOC_Os08g37320.1	ATP synthase D chain, mitochondrial, putative, expressed (169)	149	77
26-2-14	2_404	10^{-43}	(1) LOC_Os07g26610.1	(1) Phosphoglucomutase/ phosphomannomutase family protein, putative, expressed (617)	154	84
26-2-3	2_393	10^{-25}	(2) LOC_Os09g04990.1	(2) Cytoskeletal protein, putative, expressed (787)	134	55

[a]Results based on BLASTX from http://www.gramene.org/Multi/blastview compared with the peptide (TIGR gene model) database (version 25). [b]The Tigr_gene translation identifier of the most significant BLASTX hit for the boldface sequence contig. [c]The number of amino acids from the sequence contig that aligned to the Tigr_gene translation identifier protein based on the BLASTX outputs. [d]Similarity based on calculating the amino acid length and percentage similarity of all BLASTX alignments for each PPCR to the Tigr_gene translation identifier.

plemental Figure S1. The PsASGR-BBM-like1 and PsASGR-BBM-like2 predicted proteins are identical and 99.3% identical to the predicted CcASGR-BBM-like1 protein. The *CcASGR-BBM-like2* sequence was predicted to encode a smaller 489-amino acid protein also containing two AP2 domains. Splice site changes between *CcASGR-BBM-like1* and *CcASGR-BBM-like2* caused the removal of the third exon in the CcASGR-BBM-like2 predicted protein, and a potential stop codon in the seventh exon was removed by the addition of an intron in the ORF. The predicted exons from *CcASGR-BBM-like2* are highlighted in blue in Supplemental Figure S1. The alignment of the predicted ASGR-BBM-like proteins can be seen in Supplemental Figure S2.

The genomic sequences and predicted proteins for the four *ASGR-BBM-like* genes were compared against the rice genomic sequence and the TIGR gene model databases at http://www.gramene.org/multi/blastview. The most significant hits for all protein and genomic DNA *ASGR-BBM-like* queries were, in order of most to least significant, LOC_Os11g19060, LOC_Os02g40070, LOC_Os04g55970, and LOC_Os01g67410. To further verify that LOC_Os11g19060 is the most similar rice gene to the *ASGR-BBM-like* genes, the above rice hits were aligned using both genomic DNA sequences and predicted protein sequences against the *ASGR-BBM-like* genes using ClustalW. In both types of alignments, LOC_Os11g19060 was the most similar to the *ASGR-BBM-like* sequences.

Transposable Elements in the ASGR

A total of 303 sequence contigs (23%) had predicted coding regions with similarity to known or predicted transposable elements based on BLASTX analysis using the TrEMBL database (version 7). Forty-three sequence contigs contained similarity to transposases from type II transposable elements. The remaining 235 sequence contigs had similarity to various proteins from type I retrotransposable elements.

All 1,341 sequence contigs were scanned using RepeatMasker (A.F.A. Smit, R. Hubley, and P. Green, unpublished data; current version, open-3.1.8) at http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker against the rice repeat database (RM database version 20061006) using the default settings. A total of 468 sequence contigs (33%) were identified that contained sequences with similarity to transposable elements: 362 sequence contigs to type I and 106 sequence contigs to type II. Within the type I elements, 304 hits were classified as long terminal repeat retrotransposons. When the analysis was separated by species, similar percent-

Table V. *Comparison of sequence similarity between ASGR-BBM-like genes*

	CcASGR-BBM-like1	*CcASGR-BBM-like2*	*PsASGR-BBM-like1*	*PsASGR-BBM-like2*
CcASGR-BBM-like1	100%	98.6%	98.6%	98.8%
CcASGR-BBM-like2		100%	99.2%	99.1%
PsASGR-BBM-like1			100%	99.8%
PsASGR-BBM-like2				100%

ages of repetitive elements were found. The 580 *C. ciliaris* sequence contigs contain 505,489 bp of sequence and identified 175 type I and 55 type II elements. The 730 *P. squamulatum* ASGR sequence contigs, containing 452,641 bp of sequence, identified 178 type I and 48 type II elements. The 30 *P. squamulatum* ASGR-recombinant sequence contigs covering 46,504 bp had nine and three type I and type II elements identified, respectively. Supplemental Table S2 shows the percentage of nucleotides masked by RepeatMasker for each Phrap grouping for type I and type II elements as well as the percentage masked for repeats and low complexity DNA.

A total of 2,033 individual sequences (1,221 from *Pennisetum* and 814 from *Cenchrus*) generated from the 0.5× shotgun sequences were analyzed using the Assisted Automated Assembler of Repeat Families (A.A.A.R.F.) algorithm (J. DeBarry and J.L. Bennetzen, unpublished data). The A.A.A.R.F. program takes sequence overlaps from shotgun sample sequence data sets and walks them out to create pseudomolecular "builds" representing the most abundant repeat families within randomly sequenced data sets. The A.A.A.R.F. algorithm was applied to both species separately and to a combined data set that consisted of sequences pooled together from both species. Individual tests produced a number of builds for each species (118 for *Pennisetum* and 74 for *Cenchrus*). During build construction for individual species, 474 sequences were used from the *Pennisetum* data and 321 from the *Cenchrus* data set. The number of sequences used to create the pseudomolecules accounts for 38.8% and 39.5% of the *Pennisetum* and *Cenchrus* data sets respectively, indicating that these sequences were repetitive within the data sets. The builds were compared with known repeats from the TIGR plant repeat database (Ouyang and Buell, 2004) by BLAST. Many small partial elements (<2 kb) and two larger repeat elements (>2 kb) were identified using A.A.A.R.F. One large element was a type II transposon with highest homology to a putative CACTA transposon, while the other was an *Opie-2*-like element. Other identified type I homologues were to RIRE1 and *Houba/Osr13* of rice plus *Huck* and a long interspersed nuclear element from maize (*Zea mays*; data not shown). Furthermore, when A.A.A.R.F. was applied to the pooled data, builds representing both the *Opie-2*-like element family and the putative CACTA transposons were constructed, each composed of sequences from both species' data sets. This indicates that these repeat families are shared between the two species in the ASGR region.

The *Opie-2*-like retrotransposon partial sequence (AY375366) previously identified in our laboratory to give a dispersed signal across the high-copy region of the ASGR was compared by BLAST with the combined species A.A.A.R.F. output and with the ASGR sequence contigs. The AY375366 clone, generated from the Phrap 31 BAC, has the last 467 nucleotides containing an open reading frame, similar to the 5′ end of the gag protein of *Opie-2* from maize. Neither the A.A.A.R.F. outputs nor the ASGR sequence contigs contained a hit that spanned the AY375366 sequence. Rather, pieces of the AY375366 sequence were found in two A.A.A.R.F. outputs and multiple Phrap 31 contigs.

All sequence contigs were screened for possible Helitrons (Kapitonov and Jurka, 2001) within the ASGR. By structural criteria and coding potential, 21 Helitrons were predicted, 11 from *C. ciliaris* and 10 from *P. squamulatum*, accounting for about 1% of the genome in each species. Five sequence contigs from *C. ciliaris* (18-2-36, 10-0-16, 10-0-22, 10-0-39, and 18-0-35) and three from *P. squamulatum* (29-2-26, 29-0-26,and 29-2-24) exhibited sequences similar to Helitron helicases (Supplemental Table S3).

PPCR Colinearity with the Rice and Sorghum Genomes

All sequence contigs listed in boldface in Tables II to IV were analyzed by BLASTN against the rice genomic sequence database at Gramene (version 25). If a sequence contig contained both a PPCR and repetitive sequences, the repetitive sequence was removed prior to BLASTN analysis. Sequence removal was based on the BLASTX alignment of the sequence contig PPCR to the respective rice proteins. Additionally, sequence contigs with more than one PPCR were broken into separate PPCR components for analysis. Thirty-two ASGR sequence contigs and four ASGR-recombinant sequence contigs were placed on the rice genome based on the BLASTN analysis using the difference between the best BLASTN e-value score for the contig and the next highest hit (Table VI) if the sequence contig hit multiple duplicate genes in the same region on the same rice chromosome or if the total number of significant (e-value $\leq 10^{-6}$) genomic hits for the sequence contig was less than three. All but four sequence contigs had the most significant BLASTN match correspond to the same genomic region of rice as the most significant BLASTX match. Within these four cases, two sequence contigs (25-0-6 and 30-2-7) had very little difference in e-value separating the top two BLASTX hits. The other two sequence contigs (18-2-39 and 23-2-23) had BLASTN and BLASTX results that differed dramatically due to the predicted splicing of the rice genes. Eight ASGR sequence contigs and one ASGR-recombining sequence contig produced more than three hits with very similar e-values to multiple genomic regions and therefore could not be placed on the rice genome. Close proximity (<0.1 Mb apart) of sequence contigs placed on the rice genome with sequence contigs with multiple similar hits was not found.

As shown in Table VI, the ASGR sequence contigs analyzed showed highest similarity to genomic regions on seven different rice chromosomes. Five regions of colinearity (genes located <0.1 Mb apart in rice) were identified for ASGR sequence contigs and one region for the ASGR-recombining sequence con-

Table VI. *Highest similarity of ASGR and ASGR-recombinant PPCR sequences to the rice genome*

Single asterisk indicates identified regions of microsynteny (<0.1Mb) between sequence contig PPCRs and the rice genome. Two asterisks indicate regions of synteny reported by Gualtieri et al. (2006).

Sequence Contigs Containing PPCRs	Lowest BLASTN e-Value[a]	TIGR Gene Identifier	Rice Chromosome	Chromosome Location (Mb)	Next Lowest BLASTN e-Value and Corresponding Location Information (Chromosome, Mb)
11-2-35	10^{-15}	LOC_Os01g24080	1	13.6	10^{-9} (7, 1.7)
	10^{-12}	LOC_Os01g24240			
21-2-30	10^{-24}	LOC_Os01g36950	1	20.9	10^{-16} (5, 30)
15-2-15	10^{-30}	LOC_Os01g65150	1	38.1	10^{-25} (1, 38.1)
	10^{-30}	LOC_Os01g65169	1	38.2	
	10^{-27}	LOC_Os01g65190	1	38.1	
25-2-14	10^{-125}	LOC_Os01g72990	1	42.3	10^{-46} (12, 8.7)
25-0-6	10^{-102}	LOC_Os01g74640[b]	1	43.6	10^{-30} (11, 28.4)
14-2-25	10^{-129}	LOC_Os02g06340	2	3.2	10^{-106} (6, 28.7)
10-0-35	10^{-21}	LOC_Os02g56100	2	34.3	10^{-16} (6, 3.5)
11-2-34	10^{-41}	LOC_Os02g42270	2	25.4	10^{-5} (12, 13.4)
	10^{-41}	LOC_Os2g42250			
20-0-44	10^{-25}	LOC_Os03g55620	3	31.6	None
29-2-18*	10^{-34}	LOC_Os04g01240	4	0.17	None
29-2-25*	10^{-69}	LOC_Os04g01230	4	0.17	None
32-0-22	10^{-22}	LOC_Os04g19960	4	11.1	None
33-2-35*	10^{-213}	LOC_Os04g31000	4	18.4	10^{-55} (10, 10.1)
33-2-29*	10^{-33}	LOC_Os04g31030	4	18.4	None
33-2-32*	10^{-263}	LOC_Os04g31040	4	18.4	10^{-57} (4, 11.8)
33-2-34* (gene 1)	0	LOC_Os04g31050	4	18.4	10^{-133} (1, 40.5)
33-2-34* (gene 2)		LOC_Os04g31070			
19-2-19*	10^{-168}	LOC_Os04g31120	4	18.4	10^{-144} (2, 18.0)
12-2-6	10^{-18}	LOC_Os04g35080	4	21.1	10^{-9} (3, 1.4)
	10^{-14}	LOC_Os02g34430	2	20.6	
11-2-38 (gene 2)	10^{-62}	LOC_Os04g44280	4	26.0	10^{-53} (2, 25.3)
23-2-23* (gene 2)	10^{-57}	LOC_Os04g46490	4	27.4	10^{-24} (6, 13.4)
					10^{-24} (4, 27.4)
23-2-23* (gene 1)	10^{-149}	LOC_Os04g46530[b]	4	27.4	10^{-136} (2, 26.6)
13-2-1	10^{-45}	LOC_Os06g05790	6	2.6	10^{-15} (1, 8.8)
11-2-38 (gene 1)	10^{-68}	LOC_Os06g35700	6	20.8	10^{-48} (6, 20.8)
12-2-16*	10^{-42}	LOC_Os06g37610	6	22.3	None
23-0-3*	10^{-63}	LOC_Os06g37620	6	22.3	10^{-11} (2, 6.6)
18-2-37	10^{-55}	LOC_Os10g38710	10	20.3	10^{-42} (10, 20.3)
18-2-39*	10^{-53}	LOC_Os10g41070[b]	10	21.6	10^{-30} (3, 0.1)
30-2-7*	10^{-20}	LOC_Os10g41083[b]	10	21.7	10^{-13} (3, 0.01)
*CcASGR-BBM-like1***	10^{-154}	LOC_Os11g19060	11	10.8	10^{-86} (2, 24.3)
13-2-54**	10^{-90}	LOC_Os11g18690	11	10.5	10^{-77} (11, 10.9)
15-0-12**	10^{-85}	LOC_Os11g19210		11.0	
	10^{-83}	LOC_Os11g18730		10.6	
26-2-10*[c]	10^{-169}	LOC_Os07g26630	7	15.4	10^{-146} (4, 25.9)
	10^{-169}	LOC_Os07g26640			
26-2-14* (gene 1)[c]	10^{-85}	LOC_Os07g26610	7	15.3	None
26-2-12[c]	10^{-82}	LOC_Os08g37320	8	23.4	10^{-5} (12, 3.7)
	10^{-82}	LOC_Os08g37310	8	23.4	
26-2-14 (gene 2)[c]	10^{-49}	LOC_Os09g04990	9	2.7	10^{-7} (1, 2.7)

[a]Results based on BLASTN at http://www.gramene.org/Multi/blastview compared with the Rice Genomic Sequence database. [b]Lowest BLASTN results that differ from the lowest BLASTX results in Table II. [c]Sequences from ASGR-recombinant BAC.

tigs. Rice homologs for sequence contigs 29-2-18/29-2-25 and 12-2-16/23-0-3 are colinear on rice chromosomes 4 and 6, respectively. All five rice homologs identified from the sequence contigs 33-2-35, 33-2-29, 33-2-32, and 33-2-34 (genes 1 and 2) are located in a colinear manner on chromosome 4. Sequence contig 19-2-19 contains a rice homolog near the rice homologs found in Phrap group 33 sequence contigs. Supplemental Figure S3 compares these sequence contigs with the rice genome using the Artemis Comparison Tool. The sequence contigs from the ASGR-recombinant BAC showed highest sequence similarity to regions on three different rice chromosomes but identified colinear genes on rice chromosome 7.

Two separate sorghum BAC libraries were probed with amplicons covering sections of ASGR-PPCRs plus one potential flanking ASGR, marker HHU27 (Jessup et al., 2002; Goel et al., 2006), to establish possible colinearity with the sorghum genome. The sorghum BAC libraries probed included the SB_BBc filters (Clemson University Genomics Institute [CUGI]) and the 052-SOR-H3 filters (Texas A&M University [TAMU]), both derived from SorghumBTx623. Included in the probing were four PPCRs from sequence contigs 33-2-32, 33-2-34 (PPCR 2), 33-2-35, and 19-2-19 showing colinearity with the rice genome on chromosome 4, two PPCRs from sequence contigs 13-2-54 and 13-2-56 showing synteny on chromosome 11 (Gualtieri et al., 2006), and two PPCRs from sequence contig 11-2-38 (PPCR 1 and 2) that do not show synteny in the rice genome. Supplemental Table S4 indicates the primers used for each PPCR amplicon, the ASGR-BAC from which the amplicon was derived, and the address for each sorghum clone hit. No two probes identified the same BAC from the SB_BBc library. One BAC from the 052-SOR-H3 library hybridized to both the 19-2-19 and 33-2-34 (PPCR 2) probes. Each hybridizing BAC was then analyzed for placement in an FPC contig. SB_BBc filter data were analyzed at the Comparative Saccharinae Genomics Resource (http://cggc.agtec.uga.edu/). The 052-SOR-H3 data were provided by P.E.K. (TAMU). Approximately 73% of sorghum BAC clones from library 052_SOR_H3 and approximately 37% of sorghum BAC clones from library SB_BBc identified as hybridizing with the ASGR-PPCR probes were placed into 59 and 25 different FPC contigs, respectively. Of these FPC contigs, 44% from 052_SOR_H3 and 52% from SB_BB filters have been mapped to a sorghum chromosome. The PPCR probe from sequence contig 13-2-56 and the HHU27-like probe identified two BAC clones each (c0029N21/c0002F04 and c0057K09/c0077M17, respectively) from the SB_BBC library that are located on FPC contig 564, chromosome 4 (according to the nomenclature of Kim et al. [2005]), at approximately 65 cM. PPCR probes from sequence contigs 19-2-19 and 33-2-34 (gene 2) identified BACs located on FPC contig 2,408 of the SB_BBc library (not anchored) and HICF contig 8,567 located on chromosome 4 at approximately 92 cM of the 052_SOR_H3 library.

An in silico BLASTN analysis using identical sequences as the rice BLASTN was conducted on September 11, 2007, using the SORprelim.fasta.masked100 database at http://www.phytozome.net/search.php?show=blast. Table VII shows the best hit, the corresponding sorghum super contig, and its position on the contig. Twenty-seven sequence contigs containing ASGR-PPCRs and four sequence contigs containing ASGR-recombinant PPCRs were placed on sorghum supercontigs. In total, four regions of microsynteny could be identified for the ASGR-PPCRs and one for the ASGR-recombining PPCRs. Three of the ASGRs and the one ASGR-recombining region of microsynteny in sorghum were also present in rice.

DISCUSSION

Sequence Analysis of ASGR and ASGR-Recombinant BAC Shotgun Libraries

We sample sequenced 32 ASGR-BAC clones chosen to maximize coverage across the ASGR in both species based on FPC analysis (Goel et al., 2006), *Hin*dIII restriction, hybridization fingerprinting (Gualtieri et al., 2006), and FISH hybridization patterns (Goel et al., 2003, 2006; Akiyama et al., 2004, 2005). We also chose to sample sequence a BAC clone that is closely linked to the ASGR in *P. squamulatum* for comparison. Through sample and targeted gene sequencing, approximately 1 Mb of sequence linked to the ASGR from both species and approximately 46 kb of sequence recombinant to the ASGR in *P. squamulatum* have been generated. Our sequencing represents approximately 2% of the entire ASGR based on the predicted size of the region in *P. squamulatum*.

The analysis of all sequence contigs generated in this study identified 24 *C. ciliaris* sequence contigs containing 25 PPCRs, 21 *P. squamulatum* sequence contigs containing 23 PPCRs, and four *P. squamulatum* recombinant sequence contigs containing five PPCRs based on BLASTX similarity to the TIGR gene model database. Excluding sequence contigs related to transposable elements, the gene density of BACs within the ASGR was quite variable. Of the six sample-sequenced ASGR-BACs that failed to provide evidence of PPCRs based on BLASTX analysis, four are physically located in the high-copy region of the ASGR. While a greater number of PPCRs were identified on BACs located within the low-copy region of the ASGR, the highest gene density (one PPCR per 21 kb) was identified from Phrap group 33, whose BACs physically map at the edge of one high-copy flanking region toward the low-copy region of the ASGR. The ASGR-recombinant BAC had the highest gene density of approximately one gene every 16 kb.

Forty different rice proteins were identified when sequence contigs containing ASGR-PPCRs were combined from both species. Using gene ontology terms identified in the rice proteins containing the highest BLASTX hit to the ASGR-PPCRs, there are four ASGR-PPCRs predicted to encode proteins with functional domains known to bind or alter DNA structure and two ASGR-PPCRs with catalytic kinase domains. Nine ASGR-PPCRs had similarity to hypothetical or expressed proteins in rice. One could postulate a role for any of these ASGR-PPCRs in the apomictic developmental pathway. Mutational load in the sequenced regions of the ASGR was not greatly useful for discarding ASGR-PPCRs as nonfunctional. Using the gene prediction program at RiceGAAS, the four *ASGR-BBM-like* genes and the eight ASGR-PPCRs with at least 80% amino acid coverage compared with the corresponding rice protein were all predicted to encode potentially expressed transcripts, even though the sequence contig 11-2-34 and the *CcASGR-BBM-like2* gene contained

Table VII. *Highest similarity of ASGR and ASGR-recombinant PPCR sequences to the preliminary masked sorghum genome*

Asterisk indicates identified regions of microsynteny (<0.2 Mb) between sequence contig PPCRs and the sorghum supercontigs.

Sequence Contigs Containing PPCRs	Lowest BLASTN e-Value[a]	Sorghum Supercontig No.	Hit Location on the Supercontig (Mb)	Next Lowest BLASTN e-Value and Corresponding Sorghum Supercontig No.
14-2-25	10^{-61}	1	1.760	10^{-23} (154)
21-2-30	10^{-30}	19	7.775	10^{-13} (59)
20-0-44	10^{-35}	27	0.682	10^{-28} (16)
33-2-35*	10^{-114}	32	1.737	10^{-60} (98)
33-2-29*	10^{-101}	32	1.952	None
33-2-32*	0	32	1.956	10^{-23} (62)
33-2-34* (gene 1)	10^{-23}	32	1.970	None
33-2-34 (gene 2)	0	32	3.077	10^{-37} (11)
11-2-12*	10^{-24}	33	3.534	10^{-12} (18)
11-2-38* (gene 2)	10^{-29}	33	3.535	10^{-22} (195)
23-0-3*	10^{-56}	34	2.076	None
12-2-16*	10^{-79}	34	2.080	None
11-2-38 (gene 1)	10^{-52}	34	4.826	10^{-33} (89)
18-2-39	10^{-7}	37	5.522	None
12-2-6	10^{-16}	46	3.014	10^{-7} (93)
10-0-35	10^{-31}	57	2.308	10^{-10} (13)
13-2-54	10^{-44}	72	2.778	10^{-34} (38)
15-0-12				
32-0-22	10^{-14}	74	2.335	None
25-2-14	10^{-65}	99	0.042	10^{-7} (1)
29-2-25*	10^{-69}	99	0.357	None
29-2-18*	10^{-62}	99	0.366	None
19-2-19	10^{-110}	100	2.092	10^{-13} (17)
23-2-23* (gene 1)	10^{-180}	131	1.021	10^{-90} (110)
23-2-23* (gene 2)	10^{-59}	131	1.022	10^{-42} (49)
18-2-37	10^{-27}	135	0.368	10^{-19} (77)
11-2-35	10^{-86}	735	0.007	10^{-10} (10)
	10^{-86}	132	0.648	
26-2-14 (gene 2)[b]	10^{-48}	27	6.145	None
26-2-14* (gene 1)[b]	10^{-64}	46	1.202	None
26-2-12*[b]	10^{-62}	46	1.312	10^{-47} (108)
26-2-10*[b]	0	46	1.291, 1.282,	10^{-90} (85)
	0	33	1.208	

[a]Results based on BLASTN at http://www.phytozome.net/search.php?show=blast using the preliminary masked sorghum genomic database. [b]Sequences from ASGR-recombinant BAC.

potential stop codon mutations when compared with the corresponding rice protein. The combination of these results suggests that even if full sequencing of the ASGR was accomplished, too many potentially functional ASGR-PPCRs would be identified for gene-by-gene analysis, assuming that the apomictic pathway is controlled by genes located within the ASGR and not by epigenetic factors such as noncoding RNAs or heterochromatin structure.

Of the PPCRs identified in the study, we did choose to fully sequence the *ASGR-BBM-like* genes as the best potential candidate gene. *BBM* originally was designated as a transcript induced in microspore cultures of *Brassica napus* (*BnBBM*) undergoing somatic embryogenesis. Two copies of *BBM* were identified in *B. napus* and are orthologous to a single gene in Arabidopsis (*Arabidopsis thaliana*). Overexpression of *BnBBM* in Arabidopsis results in the formation of ectopic embryos on leaves (Boutilier et al., 2002) and is sufficient to induce spontaneous somatic embryogenesis. While the exact function of BBM is still unclear, the data suggest a role for BBM in either the induction and/or the maintenance of embryo development. Data from another BBM-like gene, *EgAP2-1*, isolated from *Elaeis guineensis* (oil palm), suggest that it also may play a role in oil palm zygotic and somatic embryo development (Morcillo et al., 2007). The complete gene sequences and gene prediction of the four *ASGR-BBM-like* genes suggests that they are potentially functional. Their similarities to other *BBM* genes suggest a possible role in the induction and/or maintenance of the unreduced embryo in the aposporous embryo sac.

Further study of gene expression and, more importantly, functional knockdown of the *ASGR-BBM-like* genes will be needed to confirm a possible role in the apomictic developmental pathway.

The duplication of sequence contigs containing PPCRs at the ASGR was not unexpected given our previous FPC and comparative mapping results (Roche et al., 2002; Goel et al., 2006; Gualtieri et al., 2006). Data generated by the shotgun sequencing approach does not allow us to estimate the true extent and/or the type of duplications at the ASGR. However, gene duplication events are not rare even within the smallest plant genomes. The Arabidopsis genome is composed of duplications derived from ancient polyploidization and also contains 17% of its genes arranged as tandem repeats (Arabidopsis Genome Initiative, 2000; Vision et al., 2000; Bowers et al., 2003b). Rice has a comparable percentage of tandem gene families, approximately 14%. When the rice genome was scanned for tandem repeat genes, allowing for interruptions within the repeat of genes, then approximately 29% of the rice genes were found to be organized in this manner (International Rice Genome Sequencing Project, 2005). For maize, approximately one-third of the genes are organized in tandem arrays (Messing et al., 2004).

Transposable Elements at the ASGR

Classification of repetitive elements ranged from 23% to 40% of the sequences generated, depending on the analysis used. As in other higher plants, retroelements constituted more of the DNA in the studied regions than any other repeat class, but even the most abundant element (a previously reported *Opie-2*-like long terminal repeat retrotransposon) contributed only slightly more than 1% of the sequenced DNA. Helitrons (Kapitonov and Jurka, 2001) were also found to be abundant in the sequenced regions, with all of the families present providing about 1% of the genomic DNA.

The haploid genome sizes of *P. squamulatum*, *C. ciliaris*, and the sexual diploid *Pennisetum glaucum* are approximately 5,150 Mb, approximately 1,500 Mb, and approximately 1,950 Mb, respectively (Roche et al., 2002). The genomic complexity of the two apomicts is not known; however, two studies have looked at genome-wide complexity in *P. glaucum*. These studies placed the amount of repetitive DNA at 54% and 69% using identical C_ot conditions (Wimpee and Rawson, 1979; Deshpande and Ranjekar, 1980). Sorghum (approximately 700–772 Mb) is composed of about 31% highly repetitive and fold-back DNA and 41% middle repetitive DNA (Sorghum Genomics Planning Workshop Participants, 2005). In maize (approximately 2,500 Mb), >80% of the genome is composed of repetitive DNA (Vitte and Bennetzen, 2006).

Therefore, it is somewhat puzzling that the sequence contigs derived from the ASGR, a highly heterochromatic region in both *P. squamulatum* and *C. ciliaris*, are not showing a larger percentage of transposable elements using multiple bioinformatics programs. It was also surprising that the frequency of transposable elements and other repeats was not dramatically different between the large *P. squamulatum* genome and the smaller *C. ciliaris* genome in the ASGR. Perhaps the repeats in these two genomes are so divergent that they are not detected as homologous by the informatic techniques employed (although the informatic screening tends to be much more sensitive than C_ot analysis). More likely, the ASGR region may be unusual in its repeat properties compared with the rest of the *P. squamulatum* or *C. ciliaris* genomes.

ASGR Synteny to the Rice and Sorghum Genomes

Comparative genetic mapping of members of the Poaceae family, including rice, maize, barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), and pearl millet (*P. glaucum*) has shown a conservation of gene and marker order between the genomes despite an up to 35-fold difference in genome size and 50 to 80 million years of evolution (Moore et al., 1995; Devos and Gale, 2000). The macrosynteny initially observed has been shown on the microsyntenic scale to be disrupted by small chromosomal rearrangements, such as translocations, duplications, and insertions (Bennetzen and Ramakrishna, 2002; Bennetzen and Ma, 2003). Polyploidization followed by differential gene loss of some but not all duplicate genes (Adams and Wendel, 2005) as well as the movement of genic sequences by transposons such as Helitrons and Pack-MULEs (Bennetzen, 2005) can explain much of the deviation in colinearity seen among relatively closely related plants. Even with the discovery of interrupted microsynteny between Poaceae genomes, comparative mapping between species can be useful. Regions of extensive conservation of gene order have been discovered between chromosome 3 of sorghum and chromosome 1 of rice (Klein et al., 2003), between the Hardness locus of *Triticum monococcum* and rice chromosome 12 (Chantret et al., 2004), and even genome-wide synteny with genomic regions undergoing recombination between sorghum and rice (Bowers et al., 2005).

Comparative mapping studies using RFLPs have also been attempted, with limited success, in apomicts. Comparative mapping has been used to identify regions of synteny between the distal part of the long arm of rice chromosome 12 and the apomixis locus in *Paspalum* (Pupilli et al., 2001, 2004). In *Tripsacum*, RFLP markers mapped to maize chromosome 6 L were completely linked to diplospory (Leblanc et al., 1995; Grimanelli et al., 1998). In *Brachiaria*, the apomixis locus is flanked by markers from maize chromosome 5 (Pessino et al., 1997, 1998). In *C. ciliaris*, markers associated with linkage group D (i.e. chromosome 6) of sorghum were mapped near the apospory locus (Jessup et al., 2002; Bowers et al., 2003a).

Genomic sequence data from other apomicts is also limited. A BAC from the apomictic controlling locus (ACL) in *Paspalum simplex* was sequenced to approximately 99% completion and organized into 20 contigs.

Four of the contigs, containing approximately 13 kb of sequence, did not show any predicted peptides. Nine additional contigs from this BAC contained similarity to 13 repetitive elements. Four genes, unrelated to transposable elements and containing significant hits to rice proteins, were identified through BLASTP alignment of FGENESH predicted proteins to the rice protein database at TIGR. Synteny of the ACL of *P. simplex* with the distal end of rice chromosome 12 was confirmed at the sequence level for the *PsEXS* and *PsPKD* genes (Calderini et al., 2006). One nonsyntenic ACL gene was similar to a FAR1 family protein found on rice chromosome 3 and may actually be a transposable element. The fourth ACL protein showed similarity to a hypothetical rice protein on chromosome 12 located on a BAC approximately 0.5 Mb from *PsEXS* and *PsPKD*.

With 95% of the rice genome completely sequenced and annotated (International Rice Genome Sequencing Project, 2005), an in silico approach was used to determine whether microsynteny or macrosynteny exists between the ASGR in *Pennisetum/Cenchrus* and rice. Gualtieri et al. (2006) used this same approach to focus on one set of BACs containing SCAR marker UGT197 or a UGT197 walking probe. When analyzed for gene content, synteny to rice chromosome 11 was shown. When more regions of the ASGR were analyzed, macrosynteny between the ASGR and a specific region of the rice genome was not preserved. However, including the region on chromosome 11 found by Gualtieri and reanalyzed in this study, regions of microsynteny can be established between ASGR sequence contigs containing PPCRs and the rice genome.

As previous articles presented data tentatively finding synteny of the ASGR in *C. ciliaris* and sorghum chromosome D (or chromosome 6; Jessup et al., 2002; Bowers et al., 2003a), two sorghum BAC libraries were hybridized with 11 ASGR-PPCR probes. As with rice, a limited number of ASGR-PPCR probes hybridized to BACs contained in the same FPC contig. Three ASGR probes and the HHU27-like probe hybridized to BACs placed on FPC contigs anchored on chromosome D (or chromosome 6). From the results of additional analyses of the sequence contigs containing PPCRs relative to the public sorghum data available, along with the BAC probing data, it is unlikely that large regions of colinearity between the ASGR and the sorghum genome will be found.

The many chromosomal rearrangements that our study has uncovered in the ASGR compared with the rice and sorghum genomes indicate that colinearity with other grasses will be a tenuous resource for ASGR analysis. Relative to other grass genome comparisons of microsynteny (Bennetzen, 2005), it appears that the ASGRs in *P. squamulatum* and *C. ciliaris* are exceptionally unstable. The lack of overall synteny identified at the locus may be biased by our limited sequencing of the region. However, it is also possible that this genomic instability might be exhibited all across these two genomes or that the lack of recombination at the ASGR leads to the reduction in synteny, as was identified when genome-wide synteny between sorghum and rice was compared with recombination rates (Bowers et al., 2005).

CONCLUSION

Apomixis is a fascinating developmental process leading to the clonal propagation of the maternal plant through seed. The ability to harness this potential in food crops could significantly alter agricultural practices. Through genomic technologies, we have sample sequenced and analyzed a small portion of the genomic region required for apomixis in *Pennisetum* and *Cenchrus*, two related apomictic genera. Our analyses have identified 40 potentially transcribed genes, four of which contain domains known to bind or alter DNA and two that contain similarity to kinase domains and are thus ''apomixis gene candidates.'' Overall gene density across the ASGR was very low (approximately one per 61 kb), although gene-rich subregions were identified. Regions of microsynteny with the rice and sorghum genomes were identified, suggesting that a narrowly defined ASGR region could use genomic colinearity with rice and/or sorghum as a tool to assist the discovery of the apomixis ''gene(s).''

MATERIALS AND METHODS

Grouping of ASGR-BAC Clones for Sequence Analysis

Ninety-nine ASGR-BAC clones isolated using 17 ASGR molecular markers totally linked to the apospory phenotype in either and/or both apomictic species were previously analyzed by fingerprinting (Goel et al., 2006) and *Hin*dIII restriction and hybridization fingerprinting (Gualtieri et al., 2006). PCR products containing genic regions were radiolabeled with ^{32}P, probed onto filters containing the 99 ASGR-linked BAC clones, and washed at high stringency (0.1× SSC, 0.1% SDS at 65°C). Probe homologies to BACs were identified, and the FPC fingerprinting data were reanalyzed at a tolerance of 7 and a cutoff of 10^{-10}. BAC groupings at 10^{-10} were considered accurate if they contained the same ASGR marker and at least two genic probe sequences in common.

Shotgun Library Construction and Sequencing

Twenty BAC shotgun libraries (Table I, underlined) were generated at TAMU, and their construction and sequencing are outlined by Gualtieri et al. (2006). Fifteen BAC shotgun libraries (Table I, boldface) were generated in the Ozias laboratory (UGA) and sequenced in the Pratt laboratory (UGA) using the following protocol. Starter cultures were generated by inoculation with a single BAC colony into 5 mL of Luria broth medium with 25 μg mL^{-1} chloramphenicol and incubated overnight at 37°C on a rotary shaker. Large-scale cultures were generated by inoculation with 500 μL of the starter culture and incubated into 500 mL of the same medium with shaking for 14 h at 37°C. BAC DNA was purified using the Large Construct Kit (Qiagen). BAC DNA (approximately 10 μg) was randomly sheared by passage through a GeneMachines Hydroshear DNA shearing device (Genomic Instrumentation Services) at a speed code setting of 10 for 20 cycles to produce fragments of 2.0 to 4.0 kb. DNA fragments were blunt-end repaired using a combination of T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase. DNA was purified using the Qiaquick PCR purification kit (Qiagen). Purified fragments were ligated into the dephosphorylated *Eco*RV site of pBluescript SK− (Stratagene). Ligation reactions were used for transformation of DH10B competent cells (Stratagene). For each library generated, random transformants were picked and placed onto two or three 384-well plates containing freeze broth and 100 μg mL^{-1} ampicillin. The bacteria were grown overnight at 37°C and stored at −80°C. For sample sequencing, one 384-well plate per library was thawed and the first quadrant replicated onto a 96-well plate

containing freeze broth and 100 μg mL^{-1} ampicillin. The plates were sent to the Pratt laboratory (UGA) for DNA extraction and sequencing. Protocols for DNA extraction and sequencing were described by Pratt et al. (2005).

Targeted Gene Sequencing

Additional gene sequences were generated by radioactive labeling of rice (*Oryza sativa*; http://www.genome.arizona.edu/orders/) or sorghum (*Sorghum bicolor*; http://www.fungen.org/Projects/Sorghum/Clone%20requests.htm) cDNA inserts containing high similarity to the ASGR-PPCR of interest followed by hybridization to filters containing the shotgun library of interest. Hybridization and washes were conducted at low stringency (55°C hybridization; 2× SSC, 0.1% SDS at 55°C wash). If a rice or sorghum cDNA was not available, PCR probes were generated from each end of the genic sequence of interest. The PCR fragments were labeled and hybridized to filters containing the shotgun library of interest. Hybridization and washes were conducted at high stringency (65°C hybridization; 0.2× SSC, 0.1% SDS at 65°C wash). Hybridizing clones not previously sequenced were sequenced using both the T7 and M13REV primers.

Sequence Processing

Each sequencing reaction was given a unique name and processed for quality using the MAGIC processing pipeline and database (Liang et al., 2006). Sequences were screened for the multiple cloning site in the vector, for *Escherichia coli* genomic DNA, and for BAC vector DNA. A total of 4,521 sequences (71% sequencing success rate) having a Q16VS $>$ 100 bp (average read, 554 bp) were submitted to GenBank under the dbGSS database accession numbers ED544199 to ED548719.

Assembly of Sequences and Initial BLASTX Analysis

Individual sequencing reads derived from individual BAC clones or from a group of overlapping BAC clones were given a unique Phrap group identification number and then assembled into sequence contigs using Phrap. During Phrap assembly, two additional numeric identifiers were added to the Phrap group identification number to generate a unique name for each sequence contig. We used BLASTX with the parameters gapopen=11, expect=10.0, gapext=1, and allowgaps=yes to BLAST the consensus sequences of the sequence contigs against the Swiss-Prot and TrEMBL databases. These databases make up the Uniprot Knowledgebase database (UniprotKB). Version 7 (release date February 6, 2006) of the databases was used. More information about the Uniprot databases may be found at http://www.uniprot.org.

Completed *ASGR-BBM-like* Genomic Sequences

Beginning with the sequences generated from sample and targeted gene sequencing, complete *ASGR-BBM-like* gene genomic structures from c100, c102, p203, and p207 were generated. These four BACs contain individual copies of the *ASGR-BBM-like* genes based on restriction mapping and hybridization (Gualtieri et al., 2006). PCR amplicons covering the predicted coding region of the *ASGR-BBM-like* genes were derived from the four BACs using standard PCR conditions and a proofreading enzyme. The amplicons were purified with a Qiagen PCR clean up kit and sequenced on a Beckman CEQ8000 sequencer using various primers across the *ASGR-BBM-like* gene. The individual BAC sequences were analyzed using VectorNTI advanced 10 programs (Invitrogen) and submitted to GenBank with accession numbers EU559277 (*PsASGR-BBM-like2* from p207), EU559278 (*CcASGR-BBM-like1* from c102), EU559279 (*CcASGR-BBM-like2* from c100), and EU559280 (*PsASGR-BBM-like1* from p203).

Hybridization to BAC Filters

Sorghum library filter 052-SOR-H3 from TAMU (http://hbz.tamu.edu/cgi-bin/htmlassembly?bacs) and sorghum library filter SB_BBc from CUGI (https://www.genome.clemson.edu/cgi-bin/orders?andpage=productGroupandservice=bacrcandproductGroup=26) were hybridized at 55°C following the protocol from CUGI (http://www.genome.clemson.edu/) and washed under low stringency (2× SSC, 0.1% SDS at 55°C). Probes were labeled PCR products covering PPCRs from sequence contigs and amplified from the corresponding BACs listed in Supplemental Table S2. Sorghum BAC addresses were called, and potential orthologous BACs were identified and anchored to FPC contigs and chromosomes using information at http://www.stardaddy.uga.edu/fpc/bicolor/WebAGCoL/WebFPC/ and through personal communication with Patricia Klein at TAMU.

In Silico Alignment of ASGR Genic Contigs with Rice

Putative gene contigs were analyzed with BLASTN and BLASTX programs at Gramene release 25.0 (http://www.gramene.org/) using the Genomic Sequence and Rice PEP_TIGR databases, respectively, during the month of September 2007. For both analyses, the distant homologies BLAST parameters used were -E:10; -B:100; -filter:seg; -W:3; -hitdist:40; matrix: BLOSUM62; T:15 (for BLASTX) and -E:10; -B:100; -filter:dust; -W:9, -M:1; -N:-1; -Q:2; -R:1 (for BLASTN). Gene prediction was performed using the RiceGAAS Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/).

Comparison of Rice Genomic DNA with ASGR Sequence Contigs Using the Artemis Comparison Tool

The rice genomic sequence from chromosome 4 (18330000 to 18425000) was exported from Gramene.org. This sequence is derived from the TIGR5 build. Double ACT version 2 at http://www.hpa-bioinfotools.org.uk/pise/double_act.html was used to produce a BLASTN analysis for the rice sequence and the ASGR sequence contigs 33-2-35, 33-2-29, 33-2-32, 33-2-34, and 19-2-19 using a cutoff score of 10. The resulting output was visualized using the Artemis Comparison Tool program, release 7 (Carver et al., 2005). A score of 49 was used for Supplemental Figure S3.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers ED544199 to ED548719 and EU559277 to EU559280.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Genomic alignment of the *ASGR-BBM-like* genes.

Supplemental Figure S2. Protein alignment of the predicted ASGR-BBM-like coding region.

Supplemental Figure S3. Artemis comparison of rice chromosome 4 at 18.3 Mb with ASGR sequence contigs.

Supplemental Table S1. Sequence contigs most significant BLASTX hit to the TrEMBL database.

Supplemental Table S2. RepeatMasker results broken down by Phrap grouping.

Supplemental Table S3. Details of sequence contigs with similarity to the Helitron database.

Supplemental Table S4. Sequence contig-derived probes for sorghum filters and sorghum BAC hits.

ACKNOWLEDGMENTS

We thank Evelyn Morgan and Anne Bell for technical support and Benji Adair for computational support in Tifton.

Received March 17, 2008; accepted May 25, 2008; published May 28, 2008.

LITERATURE CITED

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol **8:** 135–141

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in *Pennisetum squamulatum* reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol **134:** 1733–1741

Akiyama Y, Hanna WW, Ozias-Akins P (2005) High-resolution physical mapping reveals that the apospory-specific genomic region (ASGR) in

Cenchrus ciliaris is located on a heterochromatic and hemizygous region of a single chromosome. Theor Appl Genet **111:** 1042–1051

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev **15:** 621–627

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol **6:** 128–133

Bennetzen JL, Ramakrishna W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol **48:** 821–827

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AAM, Miki BLA, et al (2002) Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. Plant Cell **14:** 1737–1749

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, et al (2003a) A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics **165:** 367–386

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, 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

Bowers JE, Chapman BA, Rong J, Paterson AH (2003b) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Calderini O, Chang SB, de Jong H, Busti A, Paolocci F, Arcioni S, de Vries SC, Abma-Henkens MHC, Klein Lankhorst RM, Donnison IS, et al (2006) Molecular cytogenetics and DNA sequence analysis of an apomixis-linked BAC in *Paspalum simplex* reveal a non pericentromere location and partial microcolinearity with rice. Theor Appl Genet **112:** 1179–1191

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005) ACT: the Artemis Comparison Tool. Bioinformatics **21:** 3422–3423

Chantret N, Cenci A, Sabot F, Anderson O, Dubcovsky J (2004) Sequencing of the *Triticum monococcum* hardness locus reveals good microcolinearity with rice. Mol Genet Genomics **271:** 377–386

Cordonnier-Pratt MM, Liang C, Wang H, Kolychev DS, Sun F, Freeman R, Sullivan R, Pratt LH (2004) MAGIC database and interfaces: an integrated package for gene discovery and expression. Comp Funct Genomics **5:** 268–275

Deshpande VG, Ranjekar PK (1980) Repetitive DNA in three Gramineae species with low DNA content. Hoppe Seylers Z Physiol Chem **361:** 1223–1233

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell **12:** 637–646

Goel S, Chen Z, Akiyama Y, Conner JA, Basu M, Gualtieri G, Hanna WW, Ozias-Akins P (2006) Comparative physical mapping of the apospory-specific genomic region in two apomictic grasses: *Pennisetum squamulatum* and *Cenchrus ciliaris.* Genetics **173:** 389–400

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Physical evidence that a single hemizygous chromosomal region is sufficient to confer aposporous embryo sac formation in *Pennisetum squamulatum* and *Cenchrus ciliaris.* Genetics **163:** 1069–1082

Grimanelli D, Leblanc O, Espinosa E, Perotti E, Gonzales de Leon D, Savidan Y (1998) Mapping diplosporous apomixis in tetraploid Tripsacum: one gene or several genes? Heredity **80:** 33–39

Gualtieri G, Conner JA, Morishige DT, Moore LD, Mullet JE, Ozias-Akins P (2006) A segment of the apospory-specific genomic region is highly microsyntenic not only between the apomicts *Pennisetum squamulatum* and buffelgrass, but also with a rice chromosome 11 centromeric-proximal genomic region. Plant Physiol **140:** 963–971

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793–800

Jessup RW, Burson BL, Burow GB, Wang YWCC, Li Z, Paterson AH, Hussey MA (2002) Disomic inheritance, suppressed recombination, and allelic interactions govern apospory in buffelgrass as revealed by genome mapping. Crop Sci **42:** 1688–1694

Kapitonov V, Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA **98:** 8714–8719

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor.* Genetics **169:** 1169–1173

Klein PG, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J **34:** 605–621

Koltunow AM (1993) Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell **5:** 1425–1437

Koltunow AM, Soltys K, Nito N, McClure S (1995) Anther, ovule, seed, and nuclear embryo development in *Citrus sinensis* cv. Valencia. Can J Bot **73:** 1567–1582

Leblanc O, Grimanelli D, Gonzalez-de-Leon D, Savidan Y (1995) Detection of the apomictic mode of reproduction in maize-Tripsacum hybrids using maize RFLP markers. Theor Appl Genet **90:** 1198–1203

Liang C, Sun F, Wang H, Qu J, Freeman RM, Pratt LH, Cordonnier-Pratt M-M (2006) MAGIC-SPP: a database-driven DNA sequence processing package with associated management tools. BMC Bioinformatics **7:** 115

Martinez EJ, Urbani MH, Quarin CL, Ortiz JPA (2001) Inheritance of apospory in bahiagrass, *Paspalum notatum.* Hereditas **135:** 19–25

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

Moore G, Devos KM, Wang Z, Gale MD (1995) Grasses, line up and form a circle. Curr Biol **5:** 737–739

Morcillo F, Gallard A, Pillot M, Jouannic S, Aberlenc-Bertossi F, Collin M, Verdeil JL, Tregear JW (2007) *EgAP2-1,* an *AINTEGUMENTA-like* (*AIL*) gene expressed in meristematic and proliferating tissues of embryos in oil palm. Planta **226:** 1353–1362

Nogler GA (1984) Gametophytic apomixis. *In* BM Johri, ed, Embryology of Angiosperms. Springer, Berlin, pp 475–518

Noyes RD, Rieseberg LH (2000) Two independent loci control agamospermy (apomixis) in the triploid flowering plant *Erigeron annuus.* Genetics **155:** 379–390

Ouyang S, Buell CR (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res (Database issue) **32:** D360–D363

Ozias-Akins P, Roche D, Hanna WW (1998) Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus which may have no allelic form in sexual genotypes. Proc Natl Acad Sci USA **95:** 5127–5132

Pessino SC, Evans C, Ortiz JPA, Armstead I, do Valle CB, Hayward MD (1998) A genetic map of the apospory-region in Brachiaria hybrids: identification of two markers closely associated with the trait. Hereditas **128:** 153–158

Pessino SC, Ortiz J, Leblanc O, do Valle CB, Hayward MD (1997) Identification of a maize linkage group related to apomixis in Brachiaria. Theor Appl Genet **94:** 439–444

Pratt LH, Liang C, Shah M, Sun F, Wang H, Reid SP, Gingle A, Paterson AH, Wing R, Dean R, et al (2005) Sorghum expressed sequence tags provide a milestone set of 16,801 clusters and identify signature genes for drought, pathogenesis and etiolation. Plant Physiol **139:** 869–884

Pupilli F, Labombarda P, Caceres ME, Quarin CL, Arcioni S (2001) The chromosome segment related to apomixis in *Paspalum simplex* is homoeologous to the telomeric region of the long arm of rice chromosome 12. Mol Breed **8:** 53–61

Pupilli F, Martinez EJ, Busti A, Calderini O, Quarin CL, Arcioni S (2004) Comparative mapping reveals partial conservation of synteny at the apomixis locus in *Paspalum* spp. Mol Genet Genomics **270:** 539–548

Roche DR, Conner JA, Budiman MA, Frisch D, Wing R, Hanna WW, Ozias-Akins P (2002) Construction of BAC libraries from two apomictic grasses to study the microcolinearity of their apospory-specific genomic regions. Theor Appl Genet **104:** 804–812

Sherwood RT, Berg CC, Young BA (1994) Inheritance of apospory in buffelgrass. Crop Sci **34:** 1490–1494

Sorghum Genomics Planning Workshop Participants (2005) Toward sequencing the sorghum genome. A U.S. National Science Foundation-sponsored workshop report. Plant Physiol **138:** 1898–1902

van Dijk PJ, Tas ICQ, Falque M, Bakx-Schotman T (1999) Crosses between sexual and apomictic dandelions (Taraxacum). II. The breakdown of apomixis. Heredity **83:** 715–721

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science **290:** 2114–2117

Vitte C, Bennetzen JL (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA **103:** 17638–17643

Wimpee CF, Rawson JRY (1979) Characterization of the nuclear genome of pearl millet. Biochim Biophys Acta **562:** 192–206