

Research Article

Comparative Genomics in Perennial Ryegrass (*Lolium perenne* L.): Identification and Characterisation of an Orthologue for the Rice Plant Architecture-Controlling Gene *OsABCG5*

Hiroshi Shinozuka,^{1,2,3} Noel O. I. Cogan,^{1,2,3}
German C. Spangenberg,^{1,2,3,4} and John W. Forster^{1,2,3,4}

¹ Biosciences Research Division, Department of Primary Industries, Victorian AgriBiosciences Centre, La Trobe University Research and Development Park, 1 Park Drive, Bundoora, VIC 3083, Australia

² Molecular Plant Breeding Cooperative Research Centre, Bundoora, VIC 3083, Australia

³ Dairy Futures Cooperative Research Centre, Bundoora, VIC 3083, Australia

⁴ La Trobe University, Bundoora, VIC 3086, Australia

Correspondence should be addressed to John W. Forster, john.forster@dpi.vic.gov.au

Received 8 April 2011; Accepted 20 June 2011

Academic Editor: Pierre Sourdille

Copyright © 2011 Hiroshi Shinozuka et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Perennial ryegrass is an important pasture grass in temperate regions. As a forage biomass-generating species, plant architecture-related characters provide key objectives for breeding improvement. *In silico* comparative genomics analysis predicted colocation between a previously identified QTL for plant type (erect versus prostrate growth) and the ortholocus of the rice *OsABCG5* gene (*LpABCG5*), as well as related QTLs in other Poaceae species. Sequencing of an *LpABCG5*-containing BAC clone identified presence of a paralogue (*LpABCG6*) in the vicinity of the *LpABCG5* locus, in addition to three other gene-like sequences. Comparative genomics involving five other 5 grass species (rice, *Brachypodium*, sorghum, maize, and foxtail millet) revealed conserved microsynteny in the ABCG5 ortholocus-flanking region. Gene expression profiling and phylogenetic analysis suggested that the two paralogues are functionally distinct. Fourteen additional ABCG5 gene family members, which may interact with the *LpABCG5* gene, were identified through sequencing of transcriptomes from perennial ryegrass leaf, anther, and pistils. A larger-scale phylogenetic analysis of the ABCG gene family suggested conservation between major branches of the Poaceae family. This study identified the *LpABCG5* gene as a candidate for the plant type determinant, suggesting that manipulation of gene expression may provide valuable phenotypes for perennial ryegrass breeding.

1. Introduction

Perennial ryegrass is an economically important temperate pasture grass species and a diploid ($2n = 2x = 14$) member of the Poaceae family which includes other major cereal crops such as rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), sorghum (*Sorghum bicolor* L.), and maize (*Zea mays* L.) [1, 2]. Due to superior herbage digestibility and grazing tolerance, perennial ryegrass has been a primary target for molecular breeding among forage and turf crops [3]. Following establishment of a whole-genome genetic map consisting of 7 linkage groups (LGs) [4, 5], a number of quantitative trait loci (QTLs) related to

variation for herbage yield and quality have been identified [6–9]. Plant morphological traits contribute to such variation in pasture grass species and are largely controlled by genetic factors [6, 10]. For example, evaluation of a one-way pseudotestcross genetic mapping population obtained a broad sense heritability (H) value of 0.73 for the herbage fresh weight character, and significant correlations were observed between this and other traits such as plant height and tiller number [6]. Improvement of forage productivity and quality may hence be accomplished through breeding-based manipulation of plant architecture-related traits.

Comparative genomics provides a strategy for utilisation of sequence colinearity between two or more related species

for molecular genetic studies [11]. The completion of whole-genome sequencing studies for the model grass species *Brachypodium distachyon* (L.) Beauv, in addition to those of the cereals rice, sorghum, and maize, revealed high-levels of gene order conservation between Poaceae species despite lineage divergences dating over 40 million years ago (MYA) [11–13]. Such gene-based microsynteny can expedite comparative genetics approaches for gene identification and trait dissection. For example, a candidate gene for the semidwarfing (*sdw1/denso*) mutation of barley was identified through comparative mapping with rice [14]. Physical mapping of another barley semi-dwarf locus (*sdw3*) has also been progressed using microsynteny between the genome sequences of rice, *Brachypodium sylvaticum* (Huds.), P. Beauv., and sorghum [15]. Recently, perennial ryegrass has been subjected to comparative genetic analysis. Assignment of gene-associated genetic markers permitted comparison between the genetic linkage maps of perennial ryegrass, rice, wheat, and oat (*Avena sativa* L.), macrosynteny between orthologous LGs of each species [4]. Candidate genes for heading date, herbage quality, and disease resistance QTLs were also identified through comparative approaches [7, 8, 16, 17]. A partial physical map was generated in order to support map-based cloning strategies for the identification of the perennial ryegrass gametophytic self-incompatibility (SI) loci through establishment of microcolinearity with the rice genome in the target regions [18].

Restriction of tiller number per plant is a key objective to maximise grain yield in rice breeding. A member of the ATP-binding cassette protein subfamily G (ABCG), gene family OsABCG5, was identified as a tiller number determinant [19]. Although a total of 50 rice gene loci are predicted to encode ABCG proteins, none had been functionally characterised until relatively recently [19, 20]. A mutant genotype with a single nonsynonymous substitution in the OsABCG5 gene resulted in phenotypes with significantly reduced height and tiller number as compared to nonmutant comparators [21]. Further characterisation suggested that the gene controls formation of tiller buds independently of the plant growth regulator auxin, and that transcripts were localised in vascular cells and epidermis of older leaves, as well as the root endodermis, pericycle, and stele [19]. Further study of OsABCG5 has potential for application in the breeding of elite high-yielding rice varieties. Conversely in pasture grasses, for which profuse tillering is a desirable character, identification and manipulation of the orthologous gene could be highly valuable.

The ABCG gene family encodes proteins which have been extensively studied in humans as multidrug resistance transporters and disorder-related molecules [22]. ABCG proteins, which are composed of nucleotide-binding (NBD) and transmembrane domains (TMD), are divided into half- and full-size groups [20]. Half-size ABCG transporters are known to function as homodimers and/or heterodimers [22]. In the model plant *Arabidopsis thaliana* (L.) Heynh., the AtABCG12 protein, which is a half-size transporter, forms a heterodimer with the AtABCG11 protein to export cuticular

lipids [23]. Since the OsABCG5 gene belongs to the half-size gene cluster, the translated product may also require a partner ABCG transporter to form a heterodimer [19].

Transcriptome sequencing based on massively parallel DNA sequencing technologies is an emerging method for systematic discovery of genes associated with specific developmental stages or environmental conditions [24, 25]. A large quantity of sequence information has been collected from nonmodel species, and the resulting data sets may be used for further genetic studies [26]. The method is particularly suitable for survey of complex gene families in which differential expression may occur, such as those encoding ABCG proteins. As an example, perennial ryegrass, candidate genes for selenium transport have been identified through use of transcriptome sequencing [27].

In this study, *LpABCG5*, the putative orthologue of the OsABCG5 gene in perennial ryegrass, has been identified and characterised. The *LpABCG5* gene colocalises with a previously identified QTL for plant architecture, and allelic diversity at the locus may be responsible for quantitative variation for this trait. Sequencing of a bacterial artificial chromosome (BAC) clone containing the *LpABCG5* sequence revealed the presence of a closely adjacent paralogous gene, designated *LpABCG6*. Additional ABCG gene family candidates have been identified through transcriptome sequencing. Computational analysis indicates that the molecular function of OsABCG5 gene orthologues is broadly conserved within the Poaceae family. Modification of *LpABCG* gene expression may provide novel phenotypes for the breeding of both pasture-type and turf-type ryegrasses.

2. Materials and Methods

2.1. In Silico Genomic Analysis. DNA sequence information was obtained from the NCBI (<http://www.nlm.nih.gov/>), *Brachypodium distachyon* (<http://www.brachypodium.org/>), GrainGene (<http://www.gramene.org/>), and Phytozome (<http://www.phytozome.net/>) databases. Gene structure prediction was performed using the FGENESH program with the parameter setting for monocot plants (<http://linux1.softberry.com/berry.phtml>). Phylogenetic analysis was performed using the CLUSTALW program (<http://www.genome.jp/tools/clustalw/>). Nonsynonymous and synonymous nucleotide substitution rates (K_a and K_s , resp.) were calculated using the Synonymous Non-synonymous Analysis Program (SNAP; <http://www.hiv.lanl.gov/>) [28, 29]. Regression lines were deduced using the substitution rates from the ABCG5 genes and the following estimated elapsed times since evolutionary divergence: 53 MYA for species in the Panicoideae (sorghum and maize) and those in Ehrhartoideae and Pooideae subfamilies of the Poaceae; 47 MYA for Ehrhartoideae (rice) and Pooideae (*Brachypodium*, barley and perennial ryegrass), 35 MYA between *Brachypodium* and the other Pooideae species; 21 MYA between barley and perennial ryegrass, and 12 MYA between sorghum and maize. The elapsed times were calculated as averages between the minimum and maximum values of estimated periods since divergence [12]. The time of

divergence between barley and perennial ryegrass was estimated from a previous study involving a closely related Poaceae grass species, tall fescue (*Festuca arundinacea* Schreb. syn. *Lolium arundinaceum*) [30].

2.2. Genetic Linkage Analysis. PCR primers (ABC5.f: 5'-TCGTGCTCAAGTTCACCGAC-3', ABC5.r: 5'-GGA-GAACATGAGCGTCTCTCT-3') were designed based on consensus sequence information from the *OsABC5* gene and putative orthologues from barley, wheat, and *Brachypodium*. The p150/112 sib-ship, which is a one-way pseudotestcross population, was used for genetic mapping [5, 31]. PCR amplicons from the population were sequenced using the ABI 3730xl prism automatic sequencer (Applied Biosystems at present Applera, Foster City Calif, USA) to identify single nucleotide polymorphisms (SNPs) between haplotypes from the multiply-heterozygous parental genotype (C3) [4, 5]. *LpABC5* locus-specific primers (*LpABC5.f*: 5'-ATCAGGAAGGAGAGCCTCCA-3', *LpABC5.r*: 5'-ATG-ATGGTGTGGCCGCGTT-3') and a single nucleotide primer extension (SNUPe) primer (SNP-*LpABC5*: 5'-GCA-AGGAGAAGAAGAA-3') were designed to detect adenosine (A-) guanine (G) substitution SNP, and members of the p150/112 sib-ship were genotyped by the SNUPe assay with use of SNUPe premix (GE Healthcare, Chalfont St. Giles, Buckinghamshire, UK) and the MegaBACE 4000 instrument (GE Healthcare). The JoinMAP 3.0 application was used for genetic map construction [32].

2.3. BAC Clone Sequencing. The perennial ryegrass BAC-based genomic library [3, 33] was PCR-screened using *LpABC5* gene-specific primers. A Sanger shotgun sequencing library was prepared using the Large-Construction Kit (QIAGEN, Hilden, Germany) and TOPO Shotgun Subcloning Kit (Invitrogen, Carlsbad, Calif, USA). The resulting *E. coli* colonies with 1-2 kb DNA inserts were used for multiple displacement amplification (MDA) reactions with the TempliPhi kit (GE Healthcare) [34]. The inserted DNA was sequenced using the ABI 3730xl sequencer, and sequence reads were assembled using Sequencher software (Gene Codes, Mich, USA).

2.4. mRNA Extraction and cDNA Synthesis. Root, crown, leaf, inflorescence, anther, and pistil tissues were harvested from specific perennial ryegrass genotypes (Aurora₆ (AU₆) and Impact₀₄) which were established at DPI-Hamilton (Victoria, Australia) [35]. The RNeasy Plant Mini Kit (QIAGEN) was used for the total RNA purification process, and contaminating genomic DNA was eliminated by the on-column DNase digestion method using DNase I (QIAGEN). A total of 150 ng extracted RNA samples were used for cDNA synthesis with the SMART PCR cDNA synthesis Kit (Clontech, Terra Bella Avenue, Calif, USA).

2.5. Digital Gene Expression Profiling. cDNA samples from root, crown, leaf, and inflorescence tissues of the AU₆ genotype were used for gene expression profiling. PCR amplification was performed using 5 ng aliquots of cDNAs as DNA

templates in concert with gene-specific PCR primer pairs for the *LpABC5*, *LpABC6* (*LpABC6.f*: 5'-TAACGCTCA-ACGGGACG-3', *LpABC6.r*: 5'-TCGTGCGCCGATG-ATGGTGT-3'), and perennial ryegrass glyceraldehydes 3-phosphate dehydrogenase (GAPDH) gene (*LpGAPDH.f*: 5'-TGGTGCCAAGAAGGTCATCAT-3', *LpGAPDH.r*: 5'-GACCATCAACAGTCTTGG-3') [36]. PCR amplicons were visualised on 2.5% (w/v) agarose gels stained with SYBR Safe (Invitrogen).

2.6. Transcriptome Analysis Using the GS FLX Platform. RNA samples derived from leaves of the Impact₀₄ genotype and anthers and pistils of the AU₆ genotype were prepared for cDNA library construction with the modified 3'-cDNA synthesis primer (5'-AAGCAGTGGTATCAACGC-AGAGTCGCAGTCGGTACTTTTTCTTTTTV-3') and the Advantage 2 PCR Kit (Clontech) [37]. The cDNA libraries were normalised using the TRIMMER cDNA Normalization Kit (Evrogen, Miklukho-Maklaya st, Moscow, Russia) and sequenced using the GS FLX pyrosequencing instrument with the Titanium chemistry (Roche Diagnostic, Basel, Switzerland) following the product manual. The successful sequence reads were assembled through use of the Newbler program (Roche Diagnostics).

3. Results and Discussion

3.1. In Silico Comparative Genetics and Empirical Mapping of the *LpABC5* Gene. Through *in silico* mapping (Figure 1), macrosyntentic relationships were identified between a segment of rice chromosome 3 and a region of perennial ryegrass LG4 previously shown to contain a QTL for plant type. These results were consistent with previous studies [4]. A total of 4 functionally associated loci (xcdo20, xcdo1387, xcdo38, and xcdo938) were identified close to the QTL, based on use of oat cDNAs used as hybridisation probes to detect restriction fragment length polymorphisms (RFLPs) [4]. DNA sequences similar to the oat cDNAs were identified in the 1.8–12.4 Mb region of rice chromosome 3. Conservation of the same gene order was also revealed in chromosome Bd5 of *B. distachyon* (62.5–73 Mb) (Figure 1). The rice and *Brachypodium* *ABC5* orthologues were confirmed as locating in the interval between the cdo1387 and cdo38 ortholoci of each species.

PCR amplification based on the Poaceae *ABC5* consensus primers generated DNA fragments c. 250 bp in length from the C3 parental genotype, and subsequent analysis revealed sequences derived from two types of amplicon. The majority class showed the highest similarity among rice unigenes to the *OsABC5* sequence, while the minor class bore the same relationship to the paralogous *OsABC6* gene (Table 1) [20]. The respective sequences were designated *LpABC5* and *LpABC6*. An SNP marker for the *LpABC5* sequence was designed, and a total of 47 genotypes of the p150/112 population were successfully genotyped, resulting in a segregation ratio of 23:24, close to 1:1 expectation. The *LpABC5* locus was mapped to the interval between the cdo1387 and cdo38 markers on LG4, providing increased

TABLE 1: Identity and classification of rice gene sequences employed for comparative genomic analysis.

TIGR UI*	RAP-DB UI**	Other name	Reference
LOC_Os03g17350	Os03g0281900	OsABCG5	ABC transporter-related domain containing protein [19, 20]
LOC_Os03g17370	Os03g0282100	OsABCG6	Putative ABC transporter family protein [20]
LOC_Os03g17300	Os03g0281500		Resistance protein candidate
LOC_Os03g17310	Os03g0281600		Ca ²⁺ -ATPase
LOC_Os03g17340	Os03g0281800		Sulfated surface glycoprotein 185 precursor
LOC_Os03g17410	Os03g0282300		Conserved hypothetical protein

* Unique identifier in TIGR database.

** Unique identifier in the rice annotation project database (RAP-DB).

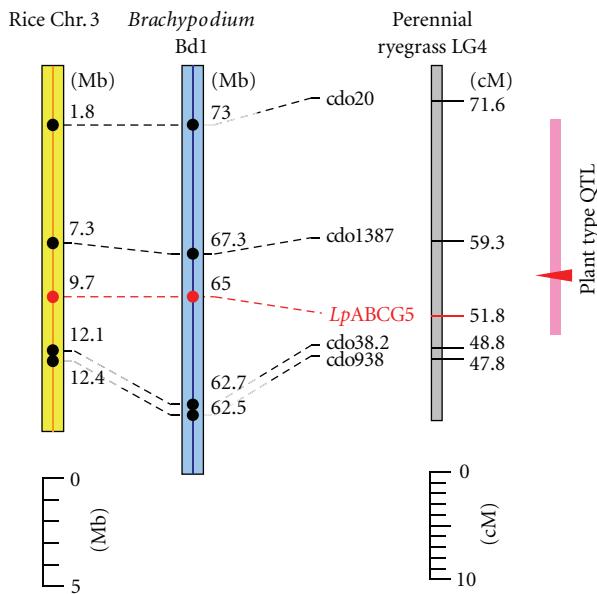


FIGURE 1: *In silico* comparative and empirical genetic mapping of the Poaceae ABCG5 locus. The gray bar represents the plant type QTL-containing region of perennial ryegrass LG4, and the yellow and blue bars show putative orthologous regions in rice and *Brachypodium*, respectively. The positions of the QTL interval and maximum LOD value are indicated with the pink line and red triangle, respectively. The putative orthologs are connected with the broken lines, and the locations of loci are shown on the right side of maps.

confidence for the predicted comparative relationships with model species in this region of the genome.

The plant type phenotypic character was previously evaluated in the p150/112 population based on semiquantitative visual score scale from 1 (most erect) to 9 (most prostrate) [6]. The broad sense heritability for this trait was 0.83, and the maximum log-of-odds (LOD) value for the QTL probability distribution was observed in the cdo1387-cdo38 interval marker loci (Figure 1) [6]. Plant type variation in the p150/112 population was significantly correlated with both plant height (maximum length in cm from the base to the top of the plant) and tiller number, revealing a tendency for prostrate individuals to exhibit short plant height and low tiller numbers [6]. Rice plants with the putative nonfunctional *OsABCG5* allele also displayed reduction of

both plant height and tiller number phenotypic characters [19, 21]. Colocation between the *LpABCG5* candidate gene and a plant type QTL suggests the possibility of functional conservation between rice and perennial ryegrass ABCG5 orthologues [19, 21]. Although tiller number and plant height phenotypic characters were also examined in the previous study, no significant locus for tiller number was identified, and plant height QTLs were only detected on LGs 1 and 3 [6]. As the QTL analysis was based on comparison of effects due to segregation of two parental alleles within a biparental population [38], it is possible that effects of the specific *LpABCG5* alleles on tiller number and plant height determinates were too small to be detected as significant. Association mapping experiments, which potentially involve a larger number of alleles for each locus, may be capable of detecting larger effects of *LpABCG5* allelic diversity on the relevant characters.

Further support for the hypothesis of gene functional conservation comes from studies of the closely related taxon Italian ryegrass (*Lolium multiflorum* Lam.), for which a plant height QTL was detected in close linkage to the cdo38 and cdo1387 marker loci [10]. In barley, plant height and ear number QTLs have also been located on the long arm of chromosome 4H, which has a macrosyntentic relationship with the lower part of perennial ryegrass LG4 [4, 39, 40]. Further comparative analyses are capable of refining the potential causal relationships between *OsABCG5* gene orthologues and plant architecture-related QTLs in these grass species.

3.2. Screening and Sequencing of the *LpABCG5* BAC Clone.

Two candidate BAC clones were identified from the perennial ryegrass genomic library, and one, designated *LpBAC39-B03*, was subjected to further analysis. A total of 653 Sanger reads (>300 bp) were generated by the shotgun sequencing method, and the assembly process created 19 large (>2.5 kb) sequence contigs. The *LpABCG5* and *LpABCG6* genomic sequences were identified in two contigs of 10.4 kb and 7.6 kb length (NCBI acc. no. JN051254 and JN051255), respectively. In other contigs, LOC_Os03g17300-, LOC_Os03g17310, and LOC_Os03g17340-like sequences (NCBI acc. no. JN051256, JN051257 and JN051258), all predicted to be conserved within this genomic region, were identified by the BLASTN search (Table 1).

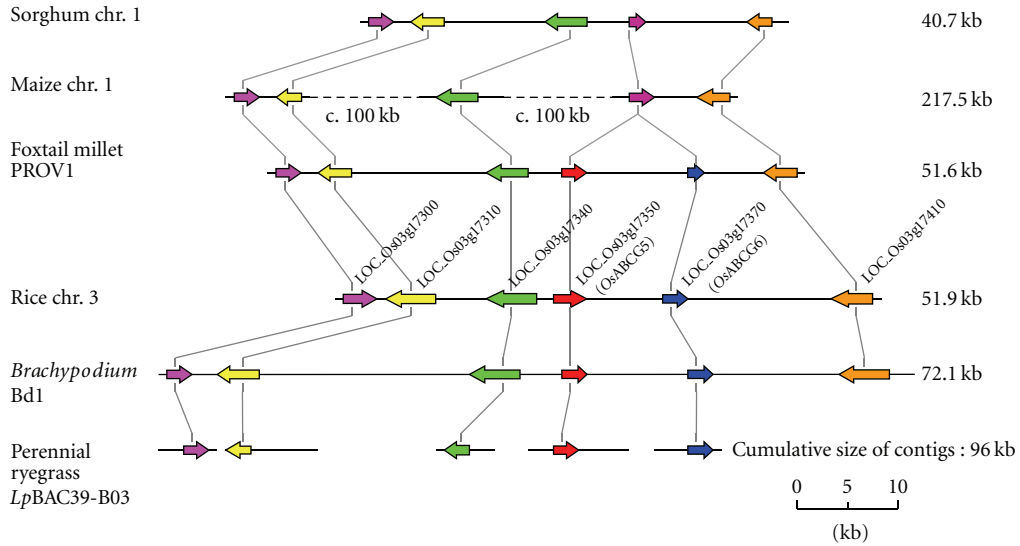


FIGURE 2: Comparative physical map for the vicinity of the Poaceae ABCG5 locus. The horizontal black lines represent genomic sequences of Poaceae species. The arrows represent gene (like) sequences, indicating direction of transcription, and putative orthologues are shown in complementary colours; ABCG5 orthologues in red; ABCG6 orthologues in blue; and ABCG gene-like sequences in sorghum and maize in purple. Physical lengths (in kb) are shown on the right side of maps.

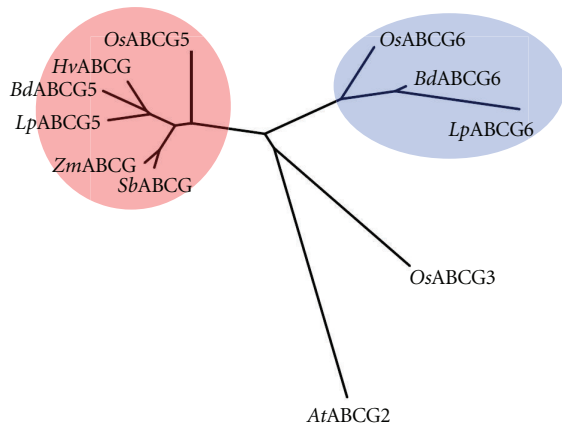


FIGURE 3: Phylogenetic trees of ABCG5 and ABCG6 orthologues. Clusters including the *OsABCG5* and *OsABCG6* genes are shown in red and blue, respectively. The barley, *Brachypodium*, sorghum, and maize ABCG genes are indicated with *Hv*, *Bd*, *Sb*, and *Zm* prefixes, respectively.

A comparative physical map for the vicinity of the ABCG5 locus was generated, based on genomic sequence information from perennial ryegrass, rice, *Brachypodium*, sorghum, maize, and foxtail millet (*Setaria italica* (L.) P. Beauv.) (Table 1, Figure 2). Only one ABCG gene-like sequence was identified in the corresponding regions of sorghum and maize, while two ABCG gene-like sequences were found in the other species. The relative order of the two paralogues was conserved for those species in which contiguous sequence assembly was completed. These results support classification of the *LpABCG5* and *LpABCG6* genes as orthologues of corresponding rice genes. Protein

coding regions composed of single exons were predicted for the *LpABCG5* (2382 bp) and *LpABCG6* (1977 bp) genes, respectively.

3.3. Phylogenetic Analysis for ABCG5 and ABCG6 Genes.

A phylogram was constructed to include the ABCG5 and ABCG6 (like) genes of monocot species and of *Arabidopsis* (Figure 3). The *LpABCG5* sequence showed the highest sequence similarity to the *AtABCG2* and *OsABCG3* genes, following exclusion of the *OsABCG5* and *OsABCG6* genes. For barley, maize, and sorghum, only single cDNA sequence database accessions were identified. In the tree, the genes from barley, sorghum and maize were clustered together with the ABCG5-like sequences from rice, *Brachypodium*, and perennial ryegrass, while the rice, *Brachypodium*, and perennial ryegrass ABCG6-like sequences formed a separate cluster. The *OsABCG3* and *AtABCG2* genes, which were classified in the same group as *OsABCG5* and *OsABCG6* in the previous study, were not clustered with either ABCG5 or ABCG6 genes [20]. This result suggests that the ABCG-like genes observed in sorghum and maize are orthologues of *OsABCG5*, but not *OsABCG6*. In the previous report, few putative orthologous pairs were identified when ABCG gene family members in *Arabidopsis* and rice were subjected to phylogenetic analysis, suggesting a taxon-specific functional diversification scenario for the gene family [20]. In contrast, the results of the present study suggest functional conservation of each ABCG5 and ABCG6 orthologous group within the Poaceae family. A relatively large genetic distance between the ABCG5 and ABCG6 clusters may reflect distinct molecular functions for the two groups.

3.4. Evolutionary Genetic Analysis. Due to the presence of insertion/deletion (indel) polymorphism, K_a and K_s values

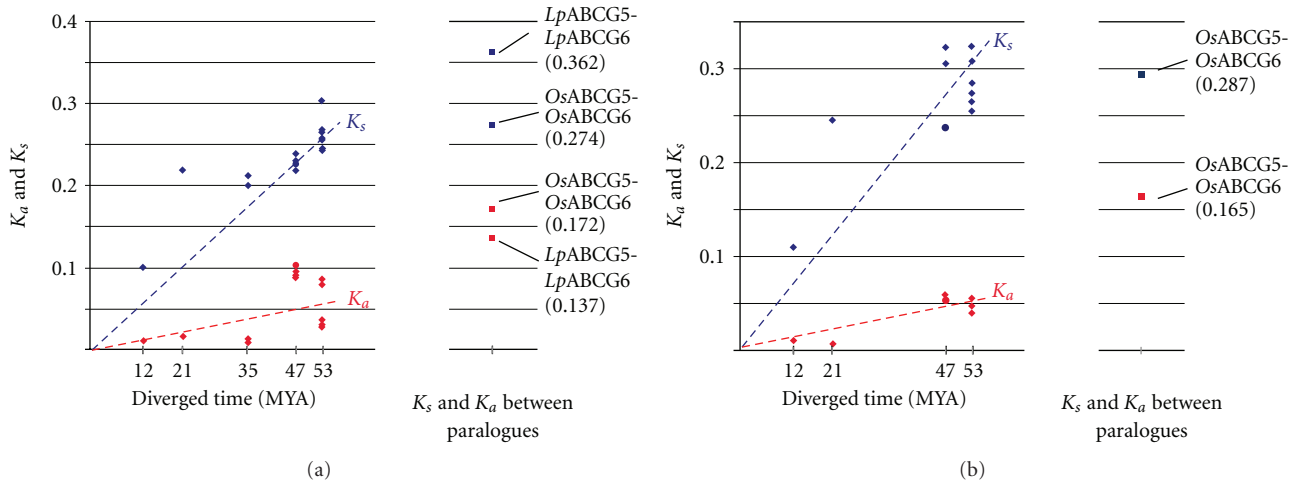


FIGURE 4: K_a and K_s values for the NBD (a) and TMD (b) peptide-coding regions. The K_a (red) and K_s (blue) values are plotted according to the estimated elapsed times since lineage divergence. The diamond (ABCG5) and circle (ABCG6) symbols denote substitution ratios between orthologous combinations, while square symbols denote substitution ratios between paralogous combinations. Broken lines represent regression plots.

TABLE 2: Putative perennial ryegrass ABCG genes identified through transcriptome sequencing. Highest-matching genes in rice are shown as candidate orthologues, with E values indicating level of similarity with the perennial ryegrass sequences.

Sequence name	Tissue	Candidate rice orthologue		E value
		TIGR UI	RAP-DB UI	
LpABCG1_cand	Leaf	LOC_01g03144	Os01g0121600	$3.00E-72$
LpABCG9_cand	Leaf	LOC_04g44610	Os04g0528300	$4.00E-37$
LpABCG12_cand	Leaf	LOC_05g13520	Os05g0222200	$2.00E-29$
LpABCG16_cand	Leaf, Anther	LOC_06g51460	Os06g0731200	$3.00E-129$
LpABCG18_cand	Leaf	LOC_08g07010	Os08g0167000	$5.00E-36$
LpABCG25_cand	Pistil	LOC_10g30610	Os10g0442900	$5.00E-66$
LpABCG26_cand	Anther	LOC_10g35180	Os10g0494300	$8.00E-104$
LpABCG27_cand	Leaf	LOC_11g07600	Os11g0177400	$6.00E-54$
LpABCG28_cand	Anther	LOC_11g22350	Os11g0416900	$8.00E-68$
LpABCG30_cand	Leaf	LOC_12g22284	Os12g0411700	$2.00E-78$
LpABCG31_cand	Anther	LOC_01g08260	Os01g0177900	$6.00E-24$
LpABCG41_cand	Leaf	LOC_02g32690	Os02g0528900	$5.00E-48$
LpABCG45_cand	Anther	LOC_08g43120	Os08g0544400	$1.00E-42$
LpABCG47_cand	Pistil	LOC_09g16380	Os09g0333000	$4.00E-40$

could not be adequately determined when whole coding region sequences were used as query data. DNA sequences for the NBD (795 bp, including a 6 bp indel) and TMD (597 bp) peptide-coding regions of the ABCG proteins were used as individual queries, due to relatively high sequence conservation. The K_a and K_s values between orthologous sequences were plotted according to estimated elapsed time (MYA) since divergence of species (Figure 4) [12, 30]. The means and standard deviations for the most distantly related ABCG5 orthologue pairs, which diverged when the BEP (Bambusoideae, Ehrhartoideae, and Pooideae) and PACC (Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae) clades (representing groups of subfamilies) branched (45–60 MYA), were 0.047 ± 0.023 (K_a , NBD), 0.046 ± 0.006 (K_a , TMD), 0.244 ± 0.018 (K_s , NBD), and 0.284 ± 0.026 (K_s , TMD). The values for the comparison of

paralogous genes (e.g., OsABCG5 and OsABCG6) were also examined (Figure 4). Both K_a and K_s values for paralogues were higher than the average values obtained for orthologue comparisons. The substitution rates between the *At*ABCG2 and monocot ABCG5 genes were also analysed. K_a values of 0.185 ± 0.005 (NBD) and 0.181 ± 0.007 (TMD), and K_s values of 0.902 ± 0.021 (NBD) and 0.845 ± 0.005 (TMD) were obtained, all higher than those obtained from comparison of ABCG5 and ABCG6 paralogues. These comparisons of nucleotide substitution frequencies suggest that the two Poaceae ABCG paralogues resulted from duplication of the ancestral gene, which occurred prior to divergence of the BEP and PACC clades. Although the common ancestor of contemporary grass species possessed two copies, the ABCG6 gene was probably deleted from the genome of the common ancestor of sorghum and maize after lineage

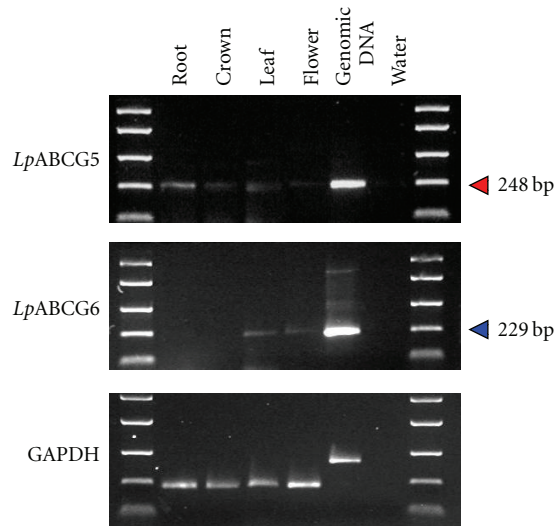


FIGURE 5: Gene expression profiling for the *LpABCG5* and *LpABCG6* genes. The fragment sizes are shown with triangles. The EasyLadder I (BIOLINE) was used as a size standard. Genomic DNA (AU₆) and distilled water were used as experimental controls. The PCR primers for the GAPDH gene were designed across intron sequence, in order to identify spurious products due to genomic DNA contamination in the cDNA samples.

divergence within the Panicoideae subfamily, as two genes are present in foxtail millet (supertribe Paniceae), but only the ABCG5 gene was detected in sorghum and maize (supertribe Andropogoneae) (Figure 2).

3.5. Gene Expression Analyses. PCR assays using cDNA templates indicated that the *LpABCG5* gene is expressed in all examined tissues, while the expression of *LpABCG6* gene is restricted to leaves and inflorescences (Figure 5). Expression of the *OsABCG5* gene was confirmed in rice roots and basal shoots by Northern hybridization analysis, but no signal was detected from either leaf blades or panicles [19]. As different detection technologies were used in each study, a conclusion of differential gene expression between species is not definitive. Further detailed analysis using common methodology would be required for comparison of expression patterns between the orthologous genes. However, the differences in expression patterns of the *LpABCG5* and *LpABCG6* genes do support the hypothesis, derived from the phylogenetic study, that the two genes are functionally distinct.

3.6. Transcriptome Sequencing Analysis. Totals of c. 57,000, 63,000 and 116,000 reads were obtained from pyrosequencing analysis of perennial ryegrass anther-, pistil-, and leaf-derived cDNA libraries, of which the average lengths were c. 233, 291 and 300 bp, respectively. After assembly, 1,315, 2,151, and 5,867 contigs were generated and c. 46,000, 48,000, and 26,000 reads remained as singletons in each library. A total of 14 fragments were identified as sharing sequence similarity with members of the rice ABCG gene family (Table 2). As both *LpABCG5* and *LpABCG6* genes are expressed in leaf and inflorescence tissues, some of the newly

identified ABCG-like genes might be interacting partners of the two genes, due to coexpression pattern. However, orthologues for the *OsABCG3*, *OsABCG10*, *OsABCG11*, and *OsABCG21* genes, which have been described as candidate heterodimer partners of the *OsABCG5* gene, were not identified [19]. Sequence reads representing the *LpABCG5* and *LpABCG6* genes were not found within the datasets, suggesting that the transcriptome analysis was still at insufficient depth to identify all mRNA types. A larger-scale sequencing analysis is hence required to obtain a full set of expressed ABCG genes from perennial ryegrass leaf and inflorescence tissues.

3.7. Phylogenetic Analysis for the ABCG Gene Family in Poaceae. Totals of 43 and 58 ABCG gene-like coding sequences (CDSs) were identified from *Brachypodium* and sorghum, respectively. A total of 167 sequences were subjected to phylogenetic analysis. Preliminary analysis indicated that the ABCG orthologues can be divided into 3 subgroups: two of half-size and one of full-size transporters (Figure 6). Of 14 perennial ryegrass sequences obtained from transcriptome sequencing, 7 were classified into the first group, while 4 and 3 were categorised in the second and third groups, respectively. The *LpABCG5* and *LpABCG6* sequences clustered with their orthologues in the second group. In the phylograms, 7 ABCG-like fragments derived from transcriptome sequencing clustered with putative rice orthologues identified by BLASTN similarity search, while the other 7 were more distantly related to the predicted orthologues. The *LpABCG18* and *LpABCG 27* candidate sequences were relatively closely related to the *LpABCG5* and *LpABCG6* genes, respectively. However, the analysis suggested the presence of more closely related paralogues in perennial ryegrass, such as the orthologue of LOC_Os01g61940 (*OsABCG3*).

The majority of half-size transporter genes formed small clades with 2-3 sequences from other species, suggesting mutual orthologous relationships (Figures 6(a) and 6(b)). A minority of genes, in contrast, formed such clades in the full-size transporter group (Figure 6(c)), suggesting different patterns of evolutionary divergence between the two types of ABCG genes. A total of 31 clades composed of putative orthologues were identified in the phylograms. Some paralogues that are closely adjacent in physical terms were also near-neighbours in sequence terms (e.g., LOC_Os09g29660 and LOC_Os09g29670), suggesting the possibility of relatively recent local gene duplication events. This analysis suggests that the common ancestor of Poaceae species possessed over 30 ABCG-like genes and that further expansion of the gene family occurred during subsequent lineage diversification.

4. Conclusions

In this study, colocation between the *LpABCG5* gene and the plant type QTL on perennial ryegrass LG4 was demonstrated, as predicted by the comparative approach. Further characterisation suggested that ABCG5 orthologues are functionally conserved as plant architecture-controlling

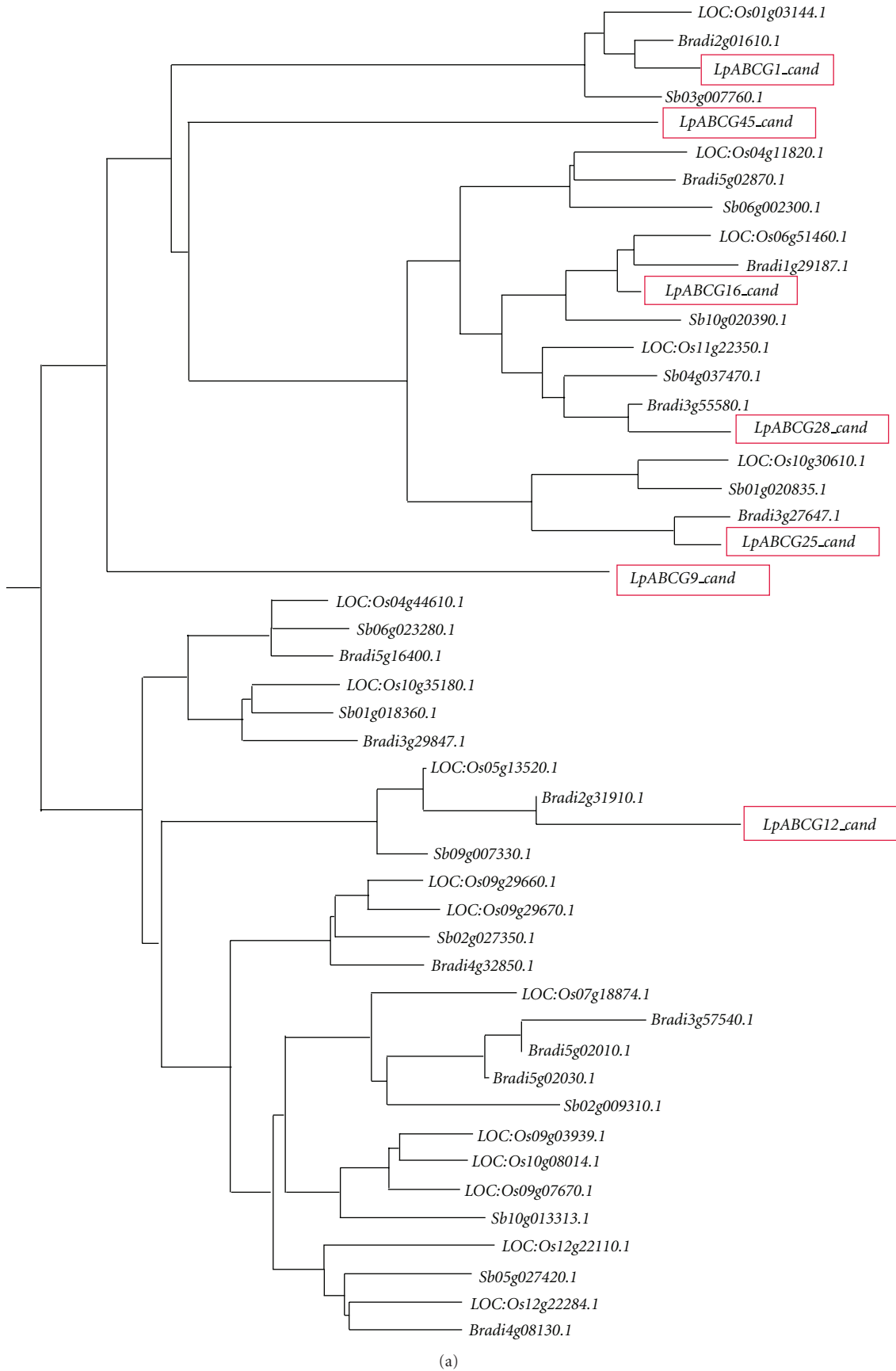


FIGURE 6: Continued.

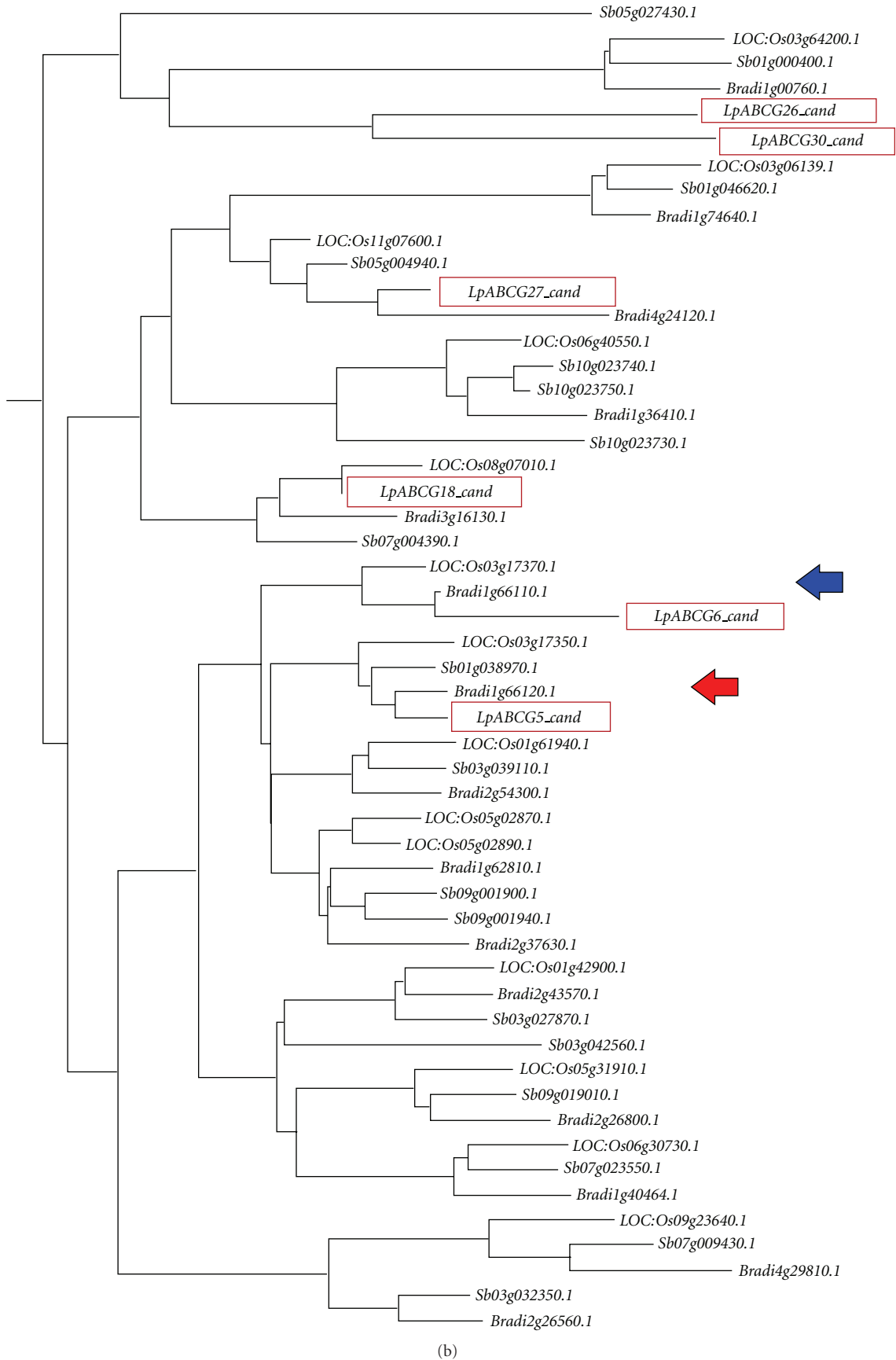


FIGURE 6: Continued.

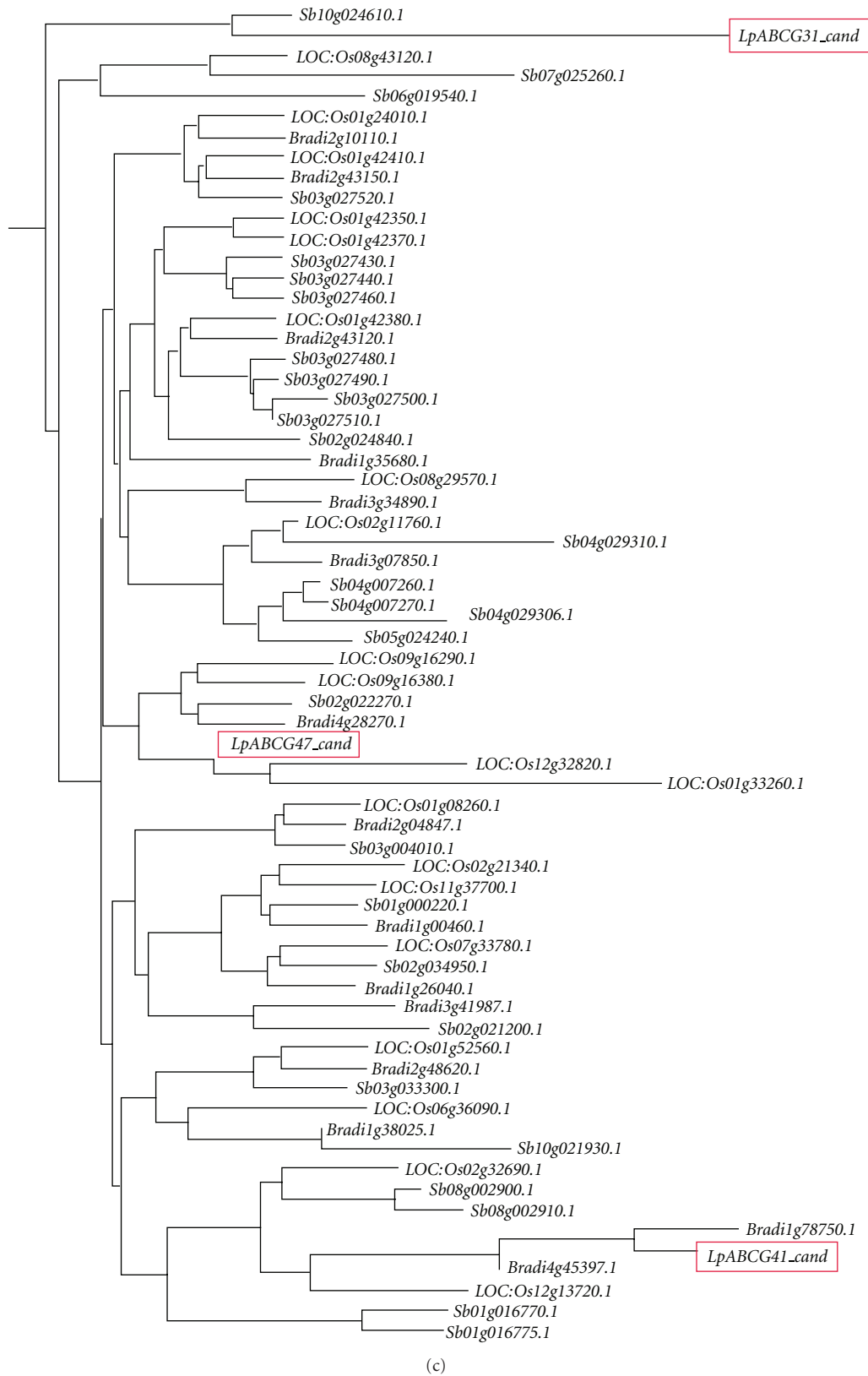


FIGURE 6: Phylogenetic trees of the ABCG gene family of the Poaceae species. The rice half- ((a) and (b)) and full-size (c) ABCG transporter genes are separately clustered. The rice, *Brachypodium*, sorghum, and perennial ryegrass genes are indicated with *LOC:Os*, *Bradi*, *Sb*, and *Lp* prefixes, respectively. ABCG transporter sequences from perennial ryegrass are highlighted with red boxes. The clusters including the ABCG5 and ABCG6 genes are indicated with the red and blue arrows, respectively (b).

genes in grass and cereal species. The *LpABCG6* gene, which is the most similar paralogue to the *LpABCG5* gene at the DNA sequence level, was also identified, and additional ABCG gene candidates were identified through deep-sequencing of tissue-specific cDNA libraries. The other ABCG genes may be related to the molecular function of the *LpABCG5* gene through heterodimer interaction. Phylogenetic analysis has further clarified the processes of ABCG gene family diversification within the Poaceae family, revealing the presence of both taxonomically conserved and species-specific genes.

Comparative genomics approaches have identified a number of conserved genes as candidates for agronomically important traits, including *LpABCG5* as a candidate for plant type variation, as shown in this study. Additional strategies, however, may be required for further understanding of such complex traits, as species-unique genes may be involved in the mechanisms. Although a reduced plant size has been a major objective for crop breeding for several decades [41, 42], the emerging demand for grass crop species, such as sorghum and maize, as renewable energy resources requires an alternative breeding strategy to deliver increases in plant size [43]. The known phenotype of putative loss-of-function mutations in *OsABCG5* suggests the possibility that manipulation of the orthologues in biomass species, such as by transgenic overexpression, may generate desirable characteristics [19, 21].

Acknowledgments

This work was supported by funding from the Victorian Department of Primary Industries and the Molecular Plant Breeding Cooperative Research Centre (MPB CRC).

References

- [1] P. W. Wilkins, "Breeding perennial ryegrass for agriculture," *Euphytica*, vol. 52, no. 3, pp. 201–214, 1991.
- [2] K. M. Devos and M. D. Gale, "Comparative genetics in the grasses," *Plant Molecular Biology*, vol. 35, no. 1-2, pp. 3–15, 1997.
- [3] J. W. Forster, N. O. I. Cogan, M. P. Dobrowolski, M. G. Francki, G. C. Spangenberg, and K. F. Smith, "Functionally-associated molecular genetic markers for temperate pasture plant improvement," in *Plant Genotyping II: SNP Technology*, R. J. Henry, Ed., pp. 154–187, CABI Press, Wallingford, Oxford, UK, 2008.
- [4] E. S. Jones, N. L. Mahoney, M. D. Hayward et al., "An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes," *Genome*, vol. 45, no. 2, pp. 282–295, 2002.
- [5] E. S. Jones, M. P. Dupal, J. L. Dumsday, L. J. Hughes, and J. W. Forster, "An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.)," *Theoretical and Applied Genetics*, vol. 105, no. 4, pp. 577–584, 2002.
- [6] T. Yamada, E. S. Jones, N. O. I. Cogan et al., "QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass," *Crop Science*, vol. 44, no. 3, pp. 925–935, 2004.
- [7] N. O. I. Cogan, K. F. Smith, T. Yamada et al., "QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.)," *Theoretical and Applied Genetics*, vol. 110, no. 2, pp. 364–380, 2005.
- [8] S. Sim, K. Diesburg, M. Casler, and G. Jung, "Mapping and comparative analysis of QTL for crown rust resistance in an Italian x perennial ryegrass population," *Phytopathology*, vol. 97, no. 6, pp. 767–776, 2007.
- [9] A. Pearson, N. O. I. Cogan, R. C. Baillie et al., "Identification of QTLs for morphological traits influencing waterlogging tolerance in perennial ryegrass (*Lolium perenne* L.)," *Theoretical and Applied Genetics*, vol. 122, no. 3, pp. 609–622, 2011.
- [10] M. Inoue, Z. Gao, and H. Cai, "QTL analysis of lodging resistance and related traits in Italian ryegrass (*Lolium multiflorum* Lam.)," *Theoretical and Applied Genetics*, vol. 109, no. 8, pp. 1576–1585, 2004.
- [11] A. J. Windsor and T. Mitchell-Olds, "Comparative genomics as a tool for gene discovery," *Current Opinion in Biotechnology*, vol. 17, no. 2, pp. 161–167, 2006.
- [12] J. P. Vogel, D. F. Garvin, T. C. Mockler et al., "Genome sequencing and analysis of the model grass *Brachypodium distachyon*," *Nature*, vol. 463, no. 7282, pp. 763–768, 2010.
- [13] P. S. Schnable, D. Ware, R. S. Fulton et al., "The B73 maize genome: complexity, diversity, and dynamics," *Science*, vol. 326, no. 5956, pp. 1112–1115, 2009.
- [14] Q. J. Jia, J. J. Zhang, S. Westcott et al., "GA-20 oxidase as a candidate for the semidwarf gene *sdw1/denso* in barley," *Functional and Integrative Genomics*, vol. 9, no. 2, pp. 255–262, 2009.
- [15] G. T. H. Vu, T. Wicker, J. P. Buchmann et al., "Fine mapping and syntenic integration of the semi-dwarfing gene *sdw3* of barley," *Functional and Integrative Genomics*, vol. 10, no. 4, pp. 509–521, 2010.
- [16] I. P. Armstead, L. B. Turner, A. H. Marshall, M. O. Humphreys, I. P. King, and D. Thorogood, "Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics," *New Phytologist*, vol. 178, no. 3, pp. 559–571, 2008.
- [17] H. Shinozuka, H. Hisano, R. C. Ponting et al., "Molecular cloning and genetic mapping of perennial ryegrass casein protein kinase 2 α -subunit genes," *Theoretical and Applied Genetics*, vol. 112, no. 1, pp. 167–177, 2005.
- [18] H. Shinozuka, N. O. I. Cogan, K. F. Smith, G. C. Spangenberg, and J. W. Forster, "Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.)," *Plant Molecular Biology*, vol. 72, no. 3, pp. 343–355, 2010.
- [19] N. Yasuno, I. Takamura, S. I. Kidou et al., "Rice shoot branching requires an ATP-binding cassette subfamily G protein," *New Phytologist*, vol. 182, no. 1, pp. 91–101, 2009.
- [20] P. J. Verrier, D. Bird, B. Burla et al., "Plant ABC proteins—a unified nomenclature and updated inventory," *Trends in Plant Science*, vol. 13, no. 4, pp. 151–159, 2008.
- [21] N. Yasuno, Y. Yasui, I. Takamura, and K. Kato, "Genetic interaction between 2 tillering genes, reduced culm number 1 (*rcn1*) and tillering dwarf gene *d3*, in rice," *Journal of Heredity*, vol. 98, no. 2, pp. 169–172, 2007.
- [22] S. Velamakanni, S. L. Wei, T. Janvilisri, and H. W. van Veen, "ABCG transporters: structure, substrate specificities and physiological roles—a brief overview," *Journal of Bioenergetics and Biomembranes*, vol. 39, no. 5-6, pp. 465–471, 2007.
- [23] H. E. McFarlane, J. J. H. Shin, D. A. Bird, and A. L. Samuels, "Arabidopsis ABCG transporters, which are required

- for export of diverse cuticular lipids, dimerize in different combinations,” *Plant Cell*, vol. 22, no. 9, pp. 3066–3075, 2010.
- [24] S. Marguerat and J. Bähler, “RNA-seq: from technology to biology,” *Cellular and Molecular Life Sciences*, vol. 67, no. 4, pp. 569–579, 2010.
- [25] F. Ozsolak and P. M. Milos, “RNA sequencing: advances, challenges and opportunities,” *Nature Reviews Genetics*, vol. 12, no. 2, pp. 87–98, 2011.
- [26] S. Zeng, G. Xiao, J. Guo et al., “Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim,” *BMC Genomics*, vol. 11, no. 1, article 94, 2010.
- [27] S. L. Byrne, K. Durandeu, I. Nagy, and S. Barth, “Identification of ABC transporters from *Lolium perenne* L. that are regulated by toxic levels of selenium,” *Planta*, vol. 231, no. 4, pp. 901–911, 2010.
- [28] B. Korber, “HIV signature and sequence variation analysis,” in *Computational and Evolutionary Analysis of HIV Molecular Sequences*, A. G. Rodrigo and G. H. Learn, Eds., pp. 55–72, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [29] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, NY, USA, 1987.
- [30] C. Kim, H. Tang, and A. H. Paterson, “Duplication and divergence of grass genomes: integrating the chloridoids,” *Tropical Plant Biology*, vol. 2, no. 1, pp. 51–62, 2009.
- [31] E. Ritter, C. Gebhardt, and F. Salamini, “Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents,” *Genetics*, vol. 125, no. 3, pp. 645–654, 1990.
- [32] J. W. van Ooijen and R. E. Voorrips, *JoinMap® 3.0, Software for the Calculation of Genetic Linkage Maps*, Plant Research International, Wageningen, The Netherlands, 2001.
- [33] G. C. Spangenberg, J. W. Forster, D. Edwards et al., “Future directions in the molecular breeding of forage and turf,” in *Molecular Breeding for the Genetic Improvement of Forage Crops and Turf*, M. O. Humphreys, Ed., pp. 83–97, Wageningen Academic Publishers, Wageningen, The Netherlands, 2005.
- [34] F. B. Dean, S. Hosono, L. Fang et al., “Comprehensive human genome amplification using multiple displacement amplification,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 8, pp. 5261–5266, 2002.
- [35] M. J. Faville, A. C. Vecchies, M. Schreiber et al., “Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.),” *Theoretical and Applied Genetics*, vol. 110, no. 1, pp. 12–32, 2004.
- [36] Y. Tu, S. Rochfort, Z. Liu et al., “Functional analyses of caffeic acid O-methyltransferase and Cinnamoyl-CoA-reductase genes from perennial ryegrass (*Lolium perenne*),” *Plant Cell*, vol. 22, no. 10, pp. 3357–3373, 2010.
- [37] E. Meyer, G. V. Aglyamova, S. Wang et al., “Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx,” *BMC Genomics*, vol. 10, article 219, 2009.
- [38] S. A. Flint-Garcia, J. M. Thornsberry, and S. B. Edward, “Structure of linkage disequilibrium in plants,” *Annual Review of Plant Biology*, vol. 54, pp. 357–374, 2003.
- [39] M. von Korff, H. Wang, J. Léon, and K. Pillen, “AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*),” *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1221–1231, 2006.
- [40] M. von Korff, S. Grando, A. Del Greco, D. This, M. Baum, and S. Ceccarelli, “Quantitative trait loci associated with adaptation to Mediterranean dryland conditions in barley,” *Theoretical and Applied Genetics*, vol. 117, no. 5, pp. 653–669, 2008.
- [41] P. Hedden, “The genes of the green revolution,” *Trends in Genetics*, vol. 19, no. 1, pp. 5–9, 2003.
- [42] T. Sakamoto and M. Matsuoka, “Identifying and exploiting grain yield genes in rice,” *Current Opinion in Plant Biology*, vol. 11, no. 2, pp. 209–214, 2008.
- [43] M. G. Salas Fernandez, P. W. Becraft, Y. Yin, and T. Lübberstedt, “From dwarves to giants? Plant height manipulation for biomass yield,” *Trends in Plant Science*, vol. 14, no. 8, pp. 454–461, 2009.