Theor Appl Genet (2010) 121:985–1000
DOI 10.1007/s00122-010-1367-9

ORIGINAL PAPER

Phylogenetic inferences in *Avena* based on analysis of *FL intron2* sequences

Yuan-Ying Peng · Yu-Ming Wei · Bernard R. Baum · Ze-Hong Yan · Xiu-Jin Lan · Shou-Fen Dai · You-Liang Zheng

Received: 8 December 2009 / Accepted: 15 May 2010 / Published online: 1 June 2010
© Springer-Verlag 2010

Abstract The development and application of molecular methods in oats has been relatively slow compared with other crops. Results from the previous analyses have left many questions concerning species evolutionary relationships unanswered, especially regarding the origins of the B and D genomes, which are only known to be present in polyploid oat species. To investigate the species and genome relationships in genus *Avena*, among 13 diploid (A and C genomes), we used the second intron of the nuclear gene *FLORICAULA*/*LEAFY* (*FL int2*) in seven tetraploid (AB and AC genomes), and five hexaploid (ACD genome) species. The *Avena FL int2* is rather long, and high levels of variation in length and sequence composition were found. Evidence for more than one copy of the *FL int2* sequence was obtained for both the A and C genome groups, and the degree of divergence of the A genome copies was greater than that observed within the C genome sequences. Phylogenetic analysis of the *FL int2* sequences resulted in topologies that contained four major groups; these groups reemphasize the major genomic divergence between the A and C genomes, and the close relationship among the A, B, and D genomes. However, the D genome in hexaploids more likely originated from a C genome diploid rather than the generally believed A genome, and the C genome diploid *A. clauda* may have played an important role in the origination of both the C and D genome in polyploids.

Communicated by P. Heslop-Harrison.

Y.-Y. Peng and Y.-M. Wei contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1367-9) contains supplementary material, which is available to authorized users.

Y.-Y. Peng · Y.-M. Wei · Z.-H. Yan · X.-J. Lan · S.-F. Dai · Y.-L. Zheng (✉)
Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Chengdu City 611130, Sichuan, People's Republic of China
e-mail: ylzheng@sicau.edu.cn

Y.-Y. Peng · B. R. Baum (✉)
Agriculture and Agri-Food Canada, Eastern Cereal and Oilseed Research Centre, Ottawa, ON K1A 0C6, Canada
e-mail: bernard.baum@agr.gc.ca

Y.-L. Zheng
Ministry of Education Key Laboratory for Crop Genetic Resources and Improvement in Southwest China, Sichuan Agricultural University, Ya'an 625014, Sichuan, People's Republic of China

Introduction

The genus *Avena* L. (Poaceae) belongs to the tribe Aveneae and forms a distinct polyploid series ranging from diploid through tetraploid to hexaploid with a basic chromosome number of seven. Baum (1977) presented a taxonomic treatment of *Avena* with seven sections and 27 species. Diploid species have either the A or C genome, tetraploids have either the AC or AB genome, and hexaploids have the ACD genome designation. The classification of these cytologically distinct genomes was based on their karyotypes and the pairing behavior of their hybrids (Rajhathy and Thomas 1974; Baum 1977; Thomas 1992). The A genome is structurally different from the C genome. Diploid species with the A genome have, in general, isobrachial chromosomes, while the C genome diploid species have mostly subterminal chromosomes. Subscripts describing A and C genome diploids indicate structural differentiation and rearrangements of chromosomes. Minor

genic differences have been reported between the A, B, and and D genomes (Oinuma 1952; Leggett and Markhand 1995; Katsiotis et al. 1997; Linares et al. 1998; Irigoyen et al. 2001). It is important to note that no diploid species containing either the B or the D genomes are currently known in the oat collections.

The evolution of the genus *Avena* remains a matter of debate despite numerous observations made of chromosome pairing (Rajhathy 1966; Rajhathy and Thomas 1974; Baum 1977; Thomas 1992; Leggett and Thomas 1995), genomic in situ hybridization (GISH; Chen and Armstrong 1994; Jellen et al. 1994a), in situ hybridization using molecular probes (Murai and Tsunewaki 1987; Fominaya et al. 1995; Katsiotis et al. 1996; Linares et al. 1998; Irigoyen et al. 2001), the use of molecular markers (Sánchez de la Hoz and Fominaya 1989; O'Donoughue et al. 1995; Alicchio et al. 1995; Ronald et al. 1997; Kianian et al. 1999; Jin et al. 1999; Nocelli et al. 1999; Li et al. 2000a, b; Loskutov and Perchuk 2000; Drossou et al. 2004; Fu and Williams 2008; Li et al. 2009), and the comparison of nucleotide sequences (Cheng et al. 2003; Irigoyen et al. 2006; Nikoloudakis et al. 2008; Nikoloudakis and Katsiotis 2008; Peng et al. 2008).

The application of different molecular techniques has provided further information concerning the relationships between the *Avena* genomes. There is now good evidence of a close relationship between the A and D genomes (Chen and Armstrong 1994; Jellen et al. 1994a; Leggett and Markhand 1995; Linares et al. 1996, 1998; Loskutov 2008), as well as between the A and B genomes (Leggett and Markhand 1995; Katsiotis et al. 1997). Molecular probes differentiated the D genome (Linares et al. 1998) and the B genome (Irigoyen et al. 2001) from the A genome. Species relationships have been largely confirmed using various molecular techniques; however, phylogenetic information for oat species remains scarce. Initial research suggested the involvement of *A. strigosa* in the A genome origin of the tetraploid and hexaploid oats (Chen and Armstrong 1994; Jellen et al. 1994a, b; Leggett and Markhand 1995; Linares et al. 1996). Later studies proposed that *A. canariensis* (Li et al. 2000b; Loskutov 2008) or *A. weistii* (Li et al. 2000b; Fu and Williams 2008) were the A genome progenitors. More recently, the possibility that *A. longiglumis* is the donor of the A genome in AC tetraploids and ACD hexaploids was suggested (Rodionov et al. 2005; Nikoloudakis et al. 2008). As for the AB genome species, either *A. hirtula*, *A. wiestii* (Rajhathy and Thomas 1974; Irigoyen et al. 2006), or *A. strigosa* (Fominaya et al. 1988a; Irigoyen et al. 2001, 2006) could be the progenitor of the AABB tetraploids. All of the diploid C genome species have been proposed as being the putative donor of the C genome in the hexaploids (Rajhathy and Thomas 1974; Chen and Armstrong 1994; Jellen et al. 1994a, b; Cheng et al. 2003; Nikoloudakis and Katsiotis 2008). There is no information concerning the origin of the D genome because no D genome diploid species has ever been identified. Many studies have suggested that the D genome may have originated from an A genome diploid, since there is close relationship between these two genomes (Ladizinsky and Zohary 1968; Ladizinsky 1988; Katsiotis et al. 1996; Leggett 1996; Drossou et al. 2004). However, this is still a hypothesis because no such A genome diploid has ever been found. Thus, the research regarding the origin of different genomes in the polyploidy species has been inconclusive.

One of the advantages of using molecular data (especially nucleotide sequences) over morphological data in a phylogenetic study is that a larger number of characters can be incorporated (Hillis 1987). During the past decade, a remarkable number of new sources of data and phylogenetic hypotheses have been accumulated using molecular characters at various taxonomic levels in plant systematics (Soltis and Soltis 1998). DNA sequences of the chloroplast genome and nuclear ribosomal DNA region are widely used at the different levels of molecular plant systematics (e.g., Wendel 2000; Soltis et al. 2003; Liu et al. 2006). However, these markers exhibit low levels of sequence variation in many lineages of angiosperms, resulting in unresolved relationships or weak support (Potter et al. 2000; Roalson et al. 2001; Baldwin et al. 2001). In particular, the internal transcribed spacers (ITS1 and ITS2) are sometimes unsuitable for phylogenetic studies, due to high sequence divergence (Wilson 2003), extensive length variation between copies (Liston et al. 1996), paralogy problems (Baker et al. 2000), or lack of resolving power (Whitcher and Wen 2001). Therefore, in plant systematics, there is a need for more nuclear gene sequence information in order to resolve phylogenetic relationships on the generic and specific level and to provide independent gene trees next to chloroplast phylogenies. Some alternative nuclear genes, perhaps suitable for phylogenetic studies of angiosperms at lower taxonomic levels, have recently been examined as a source of additional characters (e.g., Mason-Gamer et al. 1998; Small et al. 1998; Bailey and Doyle 1999; Mason-Gamer 2001; Tank and Sang 2001). These studies not only showed that the nuclear genes provided additional characters, but also proved that they are phylogenetically useful; i.e., they often contain higher percentages of phylogenetically informative characters than either ITS or chloroplast DNA data.

LEAFY is a homeotic gene that regulates the establishment of the identity of the floral meristem as well as flowering time in *Arabidopsis* (Blázquez et al. 1997; Weigel 1995). It was first described as *FLORICAULA* in *Antirrhinum majus* L. (Coen et al. 1990), then as *LEAFY* in *Arabidopsis thaliana* (L.) Heynh. (Schultz and Haughn 1991).

Although the various orthologues have different names in different genera [e.g., *NFL* in *Nicotiana tabacum* L. (Kelly et al. 1995) and *WFL* in wheat (Shitsukawa et al. 2006)], in this paper we will maintain the name *FLO/LFY* for all orthologs of *FLORICAULA* and *LEAFY*.

The structure of this gene is relatively simple and well conserved in seed plants. There are three exons and two introns, and the locations of the two introns are well conserved (Frohlich and Meyerowitz 1997; Archambault and Bruneau 2001). In contrast to the conserved coding regions of *FLO/LFY*, the second intron may contain a phylogenetic signal on the infrageneric level. However, between genera it might be too variable: Frohlich and Meyerowitz (1997) surveyed the size of this intron from 12 species of seed plants and found that it ranges from 88 bp in *Peperomia* (Piperaceae) to 7,946 bp in *Platanus* (Platanaceae). They suggested that the second intron of *LEAFY* might, therefore, have evolved at a high rate, making it useful for phylogenetic reconstructions of closely related species.

The development and application of molecular methods in oat has been relatively slow compared to other crops. Presently, there is little evidence at the nucleotide level to suggest how *Avena* species evolved. The analyses of chloroplast and nuclear ITS sequence data left many questions concerning infrageneric relationships in *Avena* unanswered, particularly concerning the origins of the B and D genomes. Therefore, we investigated the utility of using the second intron of *FLO/LFY* (hereafter abbreviated as *FL int2*) to clarify the genome origins of allopolyploid species of *Avena*, as well as the relationships amongst those genomes. The phylogenetic utility of *FL int2* within the *Avena* genus was also investigated.

Materials and methods

Plant materials and DNA isolation

Twenty-four *Avena* species, representing 13 diploid, 7 tetraploid, and four hexaploid taxa with different genomic combinations (A, C, AB, AC, and ACD), were included in this study (Table 1). All seed materials were provided by Agriculture and Agri-Food Canada (AAFC) or the National Small Grains Collection, Agriculture Research Service, United States Department of Agriculture (USDA, ARS), except for the three accessions of *A. insularis*, which were kindly provided by Dr. Rick Jellen, Brigham Young University, Provo, UT, USA. Genomic DNA was extracted from leaf samples of single plants following the modified CTAB procedure (Doyle and Doyle 1987), or, for a few samples, using a DNeasy extraction kit from QIAGEN.

FL int2 sequence amplification and sequencing

Primers used for amplifying *FL int2* in this study involved several trials and a set of primers from the literature (Flint2 F1, Flint2 R1; Grob et al. 2004). Information for all primers used in this study is presented in Table 2. A schematic drawing of the *FLO/LFY* gene and the location of the primers relative to the position of the second intron are depicted in Fig. 1. Using these multiple primers, we amplified the region from the second to the third exon of *FL int2* from the different species. PCR amplification was carried out in a reaction volume of 50 µl containing approximately 0.2 µg template DNA; 2.5 units Takara Exo Taq™ polymerase with high fidelity (Takara Bio, Inc., Kyoto, Japan); 0.3 µM each primer; 200 nm each of dATP, dCTP, dGTP, and dTTP (Takara Bio, Inc., Japan); 1.5 mM $MgCl_2$ and 1× PCR buffer. PCR amplifications were carried out in a PTC-240 thermocycler (Genetic Technologies, MJ Research, USA). The PCR reaction conditions were: 95°C for 5 min, followed by 30 cycles of 94°C for 30 s, 65°C for 1 min, and 73°C for 2 min, with a final extension at 72°C for 10 min. Amplified products were directly visualized by gel electrophoresis in 1% agarose gels. Images were photographed and captured using a Gel Doc 2000™ system (Bio-Rad, USA). The bands of interest were purified from gels using the QIAquick Gel Extraction Kit (QIAGEN). The purified PCR products were cloned in the pCR4 TOPO vector (Invitrogen) according to the TOPO TA Cloning Kit protocol. Transformants were plated on LB agar (with blue/white screening ability) containing ampicillin. Clones carrying inserts were identified using blue/white colony selection. PCR was used to double check that positive clones contained an insert. Selected colonies were incubated overnight in LB broth containing ampicillin. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit from QIAGEN, following the manufacturer's instructions. 264 clones of 53 accessions belong to 24 *Avena* species were first sequenced using the universal M13 forward and reverse primers. Because the length of the *LF int2* sequence is too long to obtain using the M13 primers, we adopted the primer walking sequencing method which is using the end of the sequenced strand as a primer for the next part of the long *LF int2* sequence. All these fragments are then assembled as a whole *LF int2* sequence. Both strands of the templates were sequenced. The number of clones sequenced per species is noted in Table 1. Nucleotide sequences of 210 different *FL int2* haplotypes were submitted to GenBank at NCBI (National Center for Biological Information, Bethesda, MD, USA), and the GenBank accession number for each is listed in Table 1.

Table 1 Plant materials of *Avena* L. used in the present study

Species	2*n*	Haplome	Accession no.	Origin	No. of sequenced clones/haplotypes	GeneBank accession no.
A. abyssinica	28	AB	PI 411163	Seraye, Eritrea	8/6	GQ979781–GQ979786
			PI 411173	Tigre, Ethiopia		
			PI 411359	Eritrea		
A. agadiriana	28	AB	CN 25856	Africa: Morocco	7/4	GQ979787–GQ979790
			CN 25869	Africa: Morocco		
A. atlantica	14	As	CN 25864	Africa: Morocco	9/8	GQ979717–GQ979724
A. barbata	28	AB	PI 337802	Izmir, Turkey	12/11	GQ979791–GQ979794
			PI 367322	Beja, Portugal		
A. brevis	14	As	Ciav 9113	Europe	11/9	GQ979725–GQ979733
A. canariensis	14	Ac	CN 23029	Canary Islands, Spain	4/2	GQ979734–GQ979735
			CN 26172	Canary Islands, Spain		
A. clauda	14	Cp	CN 19205	Iran	12/11	GQ979903–GQ979913
			CN 21378	Greece		
			CN 24695	Turkey		
A. damascena	14	Ad	CN 19457	Syria	8/6	GQ979736–GQ979741
A. eriantha	14	Cp	Ciav 9050	England, United Kingdom	7/5	GQ979914–GQ979918
A. fatua	42	ACD	PI 447299	Gansu, China	15/13	GQ979838–GQ979850
			PI 544659	South Dakota, United States		
			PI 560776	Van, Turkey		
A. hirtula	14	As	CN 21674	Corsica, France	6/4	GQ979742–GQ979745
A. hispanica	14	As	CN 25778	Portugal	9/7	GQ979746–GQ979752
			CN 25787	Portugal		
A. insularis	28	AC	Sn	Sicily, Mt. Bubonia, Italy	17/17	GQ979804–GQ979820
			6-B-22	Sicily, Gela, Italy		
			INS-4	Sicily, Gela (reselection), Italy		
A. longiglumis	14	Al	Ciav 9087	Oran, Algeria	10/10	GQ979753–GQ979762
			Ciav 9089	Libya		
A. lusitanica	14	As	CN 25885	Morocco	8/8	GQ979763–GQ979770
			CN 26441	Spain		
A. maroccana	28	AC	Ciav 8330	Morocco	12/12	GQ979821–GQ979832
			Ciav 8331	Khemisset, Morocco		
A. murphyi	28	AC	CN 21989	Spain	8/5	GQ979833–GQ979837
			CN 25974	Morocco		
A. occidentalis	42	ACD	CN 23036	Canary Islands, Spain	24/20	GQ979851–GQ979870
			CN 25942	Morocco		
			CN 26226	Canary Islands, Spain		
A. sativa	42	ACD	PI 258663	Krym, Ukraine	27/20	GQ979871–GQ979890
			PI 258666	Armenia		
			PI 258677	Ankara, Turkey		
			PI 258726	Sakhalin, Russian Federation		
			PI 258734	Kirov, Russian Federation		
			PI 636013	Heves, Hungary		
			PI 636073	South Australia, Australia		
A. sativa (hulless)	42	ACD	Ciav 9009	Ontario, Canada	9/6	GQ979891–GQ979896
A. sterilis	42	ACD	PI 411503	Alger, Algeria	12/6	GQ979897–GQ979902
			PI 411656	Tigre, Ethiopia		
A. strigosa	14	As	PI 158246	Lugo, Spain	4/2	GQ979771–GQ979772

Table 1 continued

Species	2*n*	Haplome	Accession no.	Origin	No. of sequenced clones/haplotypes	GeneBank accession no.
A. vaviloviana	28	AB	PI 412761	Eritrea	6/2	GQ979802–GQ979803
			PI 412766	Shewa, Ethiopia		
A. ventricosa	14	Cv	CN 21405	Algeria	9/8	GQ979919–GQ979926
			CN 39706	Azerbaijan		
A. wiestii	14	As	PI 299112	Chile	10/8	GQ979773–GQ979780
			Ciav 9053	Ontario, Canada		

The seed materials of *A. insularis* were obtained from Dr. Eric N. Jellen, Brigham Young University, Provo, UT, USA. The accessions with CN were obtained from Agriculture and Agri-Food Canada (AAFC), and the remaining accessions were obtained from USDA, ARS, National Small Grain Research Facility, National Small Grains Collection

Table 2 The sequence of primers used in *LF int2* PCR amplification

Primer name	Position of primer	Sequence	Reference
2_1F	Exon II	agaacgggctBgactacctSttcc	The present paper
2_1R	Exon III	ttSYtSgcgtaccKgaacacctgg	
2_2F	Exon II	aacgggctBgactacc	The present paper
2_2R	Exon III	cttgggcttgttgatgtag	
Flint2 F1	Exon II	cttccaYctStacgagcagtg	Grob et al. (2004)
Flint2 R1	Exon III	tcttgggcttgttgatgtagc	
T1F	Exon II	gagcacccBttcgtggtSacVgag	The present paper
TR	Exon III	cacgtagtgccgcatcttgggctt	
T2F	Exon II	ttcgtggtSacVgagcccggcgaggt	The present paper
TR	Exon III	cacgtagtgccgcatcttgggctt	
TLFYF	Exon II	agcagtgccgcStcttcctg	The present paper
TLFYR	Exon III	gtaccKgaacacctggttSgtcac	

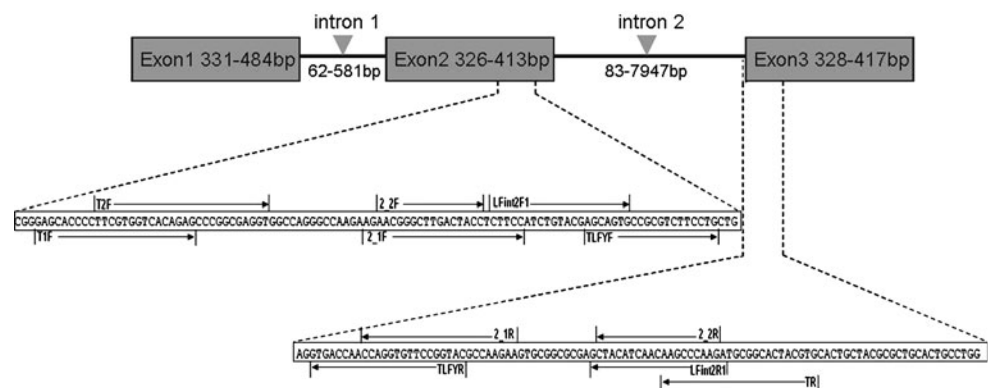

Fig. 1 Schematic drawing of the *FLO/LFY* gene. The minimum and maximum lengths of the exons and introns as found in Angiosperms are given according to Grob et al. (2004). The amplified region is drawn below, including the sequence information, the position of the second intron in *Avena* and the location of the primers used in this study

Phylogenetic analysis

The 210 *FL int2* sequences were verified, corrected, and the complementary strands assembled using the program 'DNAMAN', version 4.0 (Lynnon Biosoft Company). To confirm if the sequences were *FLO/LFY* homologs, the coding regions were compared with the *FLO/LFY* gene sequences available in GenBank using BLAST searches (Altschul et al. 1990). To confirm that the sequence matrix only included the second intron region, the exon regions and exon/intron boundaries were identified. Sequences were aligned using the program 'ClustalW' (Thompson et al. 1994), then optimized using 'SeaView', version 4.1 (Galtier et al. 1996). The alignments were then manually refined using the program 'GeneDoc' (Nicholas et al. 1997).

A first series of analyses was performed using the second intron regions of all haplotypes in all taxa (210 sequences). In this first series, gaps were treated as missing values. Phylogenetic analyses were conducted using the

distance methods as implemented in 'PAUP*', version 4.0 (Swofford 1998). The matrix of aligned sequences was very large and prohibitive for MP and ML analyses of the entire data set; therefore, we next carried out a neighbor-joining (NJ) analysis using the Jukes-Cantor and Kimura two-parameter distance estimates (Kimura 1980; Saitou and Nei 1987).

In order to obtain a concise phylogram that could clearly reflect the species relationships in each clade, a second series of analyses was performed using only the noncoding sequences from the reduced matrix. Based on the result of the NJ analysis of the whole sequence matrix, we carried out the alignment of variable sequences from different clones for every species. Then, upon visual examination of the alignments, we determined groups of similar sequences in every clade among the 210 sequences in the NJ tree. The sequences in a clade were not necessarily identical, but sufficiently close in pattern to determine sequence types found in each species. Subsequently, we calculated the consensus sequence for each type found within each species. In so doing, we reduced the 210 sequences to a 65 sequences matrix, with each species represented by one or more sequence types. Phylogenetic analyses based on the consensus sequence types of each species were conducted using parsimony and distance methods using PAUP* 4.0 (Swofford 1998). Maximum parsimony (MP) analyses were performed by heuristic search, tree bisection reconnection (TBR) branch swapping, and RANDOM stepwise addition with 1,000 replicates.

To find the evolutionary model that best fits given the data, we subjected the alignment of the consensus sequences to jModelTest (Posada 2008) version 0.1.1. The parameters of the best fitting model were then used as input for a maximum likelihood analysis. To carry out a bootstrap analysis with heuristic search, it was necessary to change the search criterion from "likelihood" to "distance". We set the "NegBrLen" to "Prohibit" and subsequently calculated the likelihood trees in memory then used the Shimodaira and Hasegawa (1999) test to select for the best tree. Finally, we subjected the consensus data to an ML analysis with heuristic search, setting the number of replicates to 100, swapping to TBR but limited to 100 while imposing the limits separately for each addition-sequence replicate.

Results

Characteristics of *FL int2* sequences in *Avena*

We were able to amplify and sequence *FL int2* regions from the different *Avena* species successfully using the multiple primers listed in Table 2. The sequences have been deposited in Genbank and the accession numbers are GQ979717 to GQ979926. BLAST searches, looking at the coding region of each sequence, confirmed the cloned sequences to be homologous to the *FLO/LFY* orthologues found in GenBank. The *Avena Flint2* sequences are rather long and demonstrate a high level of variation in length and sequence composition. Different clones ranged in length from 1,484 to 3,281 bp. These lengths are equal to or greater than the *FLO/LFY* second intron sequences found in *Zea mays* (Bomblies and Doebley 2005).

There were some difficulties in amplifying and sequencing the *FL int2* of *Avena*. In most species, diploid or polyploid, more than one copy of *FL int2* was found, and in some cases, amplification of *FL int2* yielded two products of different length. Evidence for more than one copy of *FL int2* was obtained for both the A and C genome groups. Two very different *FL int2* sequences, A-1 and A-2 groups, were obtained from the A genome. Sequences differing by various numbers of nucleotide substitutions and indels were amplified from the majority of A genome diploids. The majority of the sequences from polyploids containing the C genome fell into more than one C genome subgroup, and in each subgroup, the sequences from the same accession were closely related, but not identical, indicating the presence of several copies of *FL int2* in the C genome. However, the degree of divergence of the A genome sequences was greater than that seen amongst the C genome sequences.

Phylogenetic analysis of the entire set of *FL int2* sequences

The *FL int2* sequences of *Avena*, including all the different clones from all the individuals sampled (210 new *FL int2* sequences from 53 accessions of 24 species), were subjected to phylogenetic analysis. The sequences clustered into four major groups (Fig. 2). Groups I (62 sequences) and II (50 sequences), hereafter referred to as the A genome group (although it may include the B genome group), contained sequences from the AA genome diploids, the AABB tetraploids, two AACC tetraploids (*A. maroccona* and *A. insularis*), and the AACCDD hexaploids *A. sativa* and *A. fatua*. Because all species belonging to these clades contain the A genome, we named these two clades A-1 and A-2.

Group III, together with the outgroup referred to as the C genome group, contained 82 sequences from the CC genome diploids *A. clauda* and *A. ventricosa*, and all of the AACC tetraploid and AACCDD hexaploids included in this study. The sequences in the C genome group were clearly divided into three subgroups. One subgroup contained sequences from the C genome diploid *A. clauda* and all the polyploids containing the C genome. Unexpectedly,

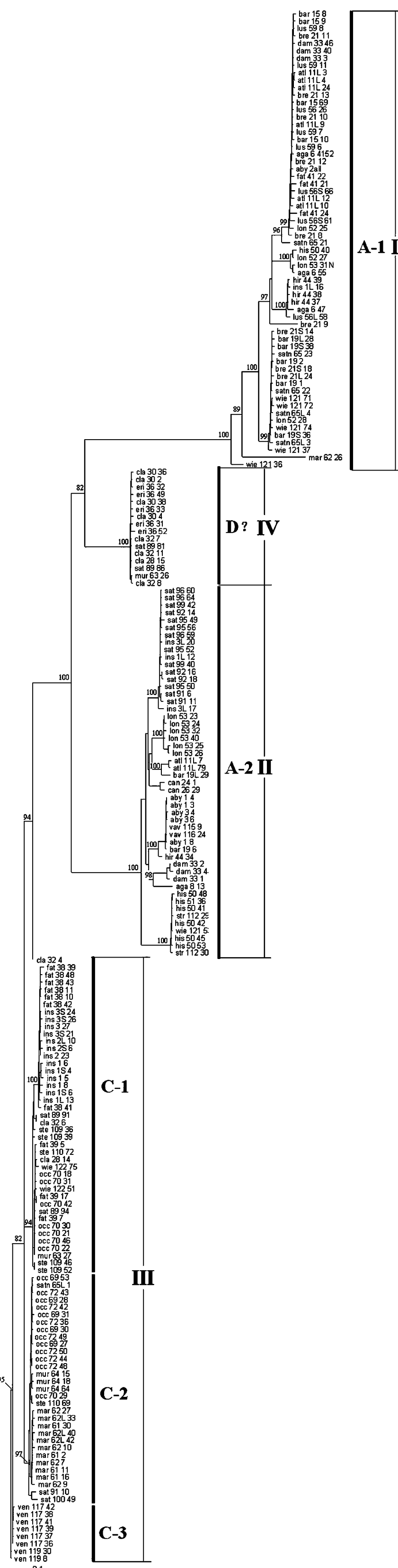

Fig. 2 Neighbor joining gene tree inferred from the *FL int2* sequences. *Numbers* above branches indicate bootstrap values above 50%

two sequences from the previously assigned A genome species *A. wiestii* were found in this clade. The second subgroup contained another (C-type of) sequence from the AACC and AACCDD species. Thus, we labeled these two C genome branches as C-1 and C-2. The third subgroup only contained sequences from the Cv genome species *A. ventricosa*.

Between the two A genome clades A-1 and A-2, there is a different clade supported with 100% bootstrap value containing sequences from the CC genome diploids *A. clauda* and *A. eriantha*, the AACC genome species *A. murphyi*, and the hexaploid *A sativa*. This was designated group IV. It is noteworthy that, although group IV contains16 sequences from the C genome species that differ from A genome-type clades, it also shows a close relationship between them. Since the D genome is close related to A genome, and since the previous suggested D genome donor species *A. clauda* and *A. murphyi* are together in this group, we tentatively deduced that this clade represents D genome sequences, and labeled it as "D?" because it is not certain.

Phylogenic analysis based on consensus sequences

We obtained 65 consensus sequences in which each species contained one or more sequence types. The model that best fits the given data was found to be TVM + G, selected by both the Akaike Information Criterion (AIC, Akaike 1974) and AICc (AIC corrected) using the following parameters: nucleotide frequencies $A = 0.2890$, $C = 0.2167$, $G = 0.2129$, $T = 0.2815$; rate matrix $AC = 0.9736$, $AG = 26065$, $AT = 0.9659$, $CG = 1.2796$, $CT = 2.6065$, $GT = 1.0$; Gamma distribution with shape $= 1.814$. By examining the summary of the AICc values and Akaike weights for the models optimized on the NJ tree, we found that 19 out of 56 models received good support from the data (Table S1). The model averaging gene trees computed as strict and majority rule (Figs. S1, S2) displayed the same general topology as the ML gene tree (Fig. 3b).

The resulting MP tree and the ML trees in the different analyses all included four main groups that are in general accordance with the tree based on the entire set of 210 DNA sequences, and, so, we also labeled these groups as A-1, A-2, C-1, C-2, and D (Fig. 3). In addition, there is a subclade in A-2 with high bootstrap support value that only contains A genome diploids and AB genome tetraploids, which has been labeled as "B?".

One of the A genome groups (A-1) was divided into four subgroups, including either two sequences from *A. brevis* and *A. maroccana* (MP tree) or *A. wiestii* and *A. maroccana* (ML tree). The latter two sequences are sufficiently differentiated from other sequences in the four subclades as

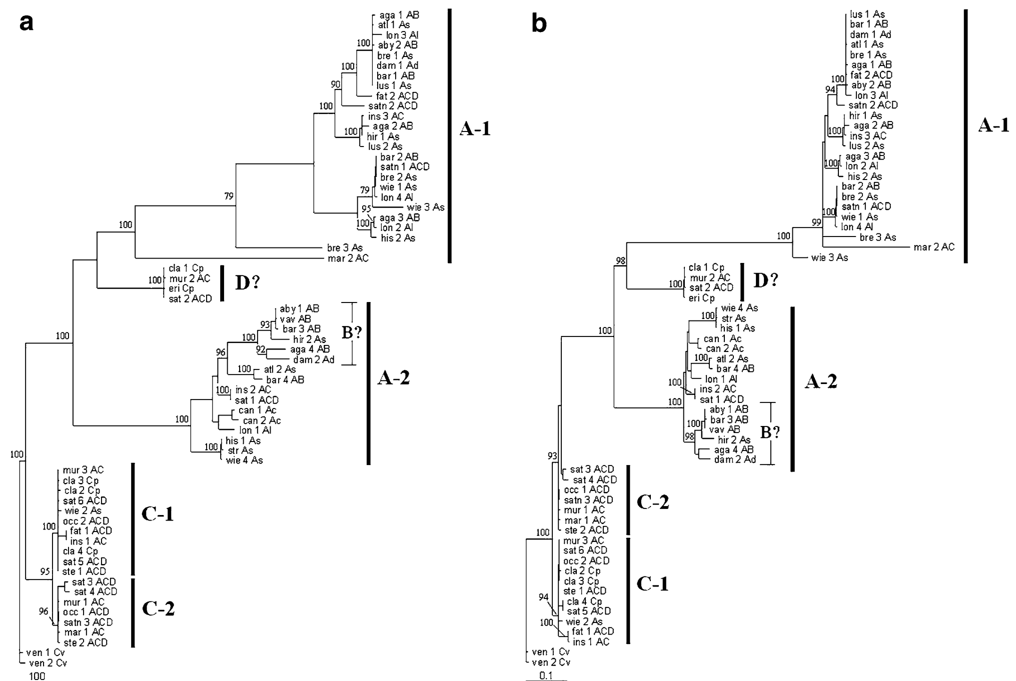

Fig. 3 Phylogenetic trees inferred from the reduced *FL int2* gene sequence matrix. **a** Majority-rule consensus tree of most parsimonious trees (Tree length = 3202, CI = 0.77, RI = 0.96), **b** maximum likelihood gene tree inferred from the reduced matrix, *numbers* above branches indicate bootstrap values above 50%

to be referred to as two distinct branches. The A genome diploid sequences were interspersed with tetraploid and hexaploid sequences, each subclade containing several AA diploids together with one to three AABB species, one AC genome tetraploid species, and one or two AACCDD species. Thus, it is not clear which A genome diploid contributed to the evolution of the different polyploids. Only *A. longiglumis* showed a close relationship with *A. agadiriana*, which is an AB genome species.

The A genome *FL int2* sequences of the A-2 clade could be clearly subdivided. One large subgroup, that we have designated as "B?", comprises the A genome diploid *A. hirtula* and the AABB species *A. abyssinica*, *A. barbata*, *A. vaviloviana*, *A. damascena,* and *A. agadiriana*. The rest of the species in the A-2 clade did not clearly cluster; however, the close relationships between *A. atlantica* and *A. barbata*, *A. insularis* and *A. sativa*, and among the three As genome diploid species *A. hispanica*, *A. strigosa,* and *A. wiestii,* should not be discounted, because these species clustered to form the smaller subclades with 100% bootstrap support values.

The C genome sequences were also divided into subgroups. There were two major subgroups, one of which comprised the C genome diploid *A. clauda* and all the AACC and AACCDD polyploids included in this study. The second major subgroup contained most of the polyploids that carry the C genome. The sequences from the Cv genome species *A. ventricosa* differed from most of C genome sequences and can be referred to as a distinct sequence. The D genome sequences included sequences from the Cp genome species *A. eriantha* and *A. clauda*, the AC genome tetraploid *A. murphyi*, and the AACCDD species *A. sativa*. The sequences in the D clade were not subdivided.

Discussion

The genus *Avena* comprises species with different degrees of ploidy (2×, 4× and 6×) and diverse genomic composition (A, B, C, and D), and includes cultivated oats, which contain the A, C, and D genomes (Thomas 1992). The evolved allopolyploids may have had structural changes in their chromosomes, which could have caused their partial homology (Baum et al. 1973; Rajhathy 1966). In addition, the presence of a genetically controlled regulatory mechanism in *Avena* that prevents homoeologous chromosome pairing (Ladizinsky 1973; Rajhathy and Thomas 1974; Ladizinsky 1974), along with the chromosomal rearrangements among the oat genomes (Chen and Armstrong 1994; Jellen et al. 1994a; Jiang and Gill 1994; Fominaya et al. 1995; O'Donoughue et al. 1995; Leggett and Markhand 1995; Linares et al. 1996, 1998), could lead to an underestimation of the actual relationships between different species at the cytological level, and complicate the unraveling of the phylogeny and the identification of ancestral species for *Avena* polyploids.

The origin of the A genome in polyploids

The A genome origin of polyploids is a most complex and controversial problem. The numerous translocations and

Table 3 Evidences about the A genome progenitor of *Avena* polyploid

Species	For or against	Morphology	Chromosome pairing behavior	Karyotyping	GISH	Molecular evidence
As genome						
A. strigosa	For		Kihara and Nishiyama (1932), Rajhathy and Morrison (1960), Marshall and Myers (1961)	Rajhathy and Thomas (1974), Fominaya et al. (1988a, b), Linares (1992), Leggett and Thomas (1995)	Chen and Armstrong (1994), Jellen et al. (1994a, b), Leggett and Markhand (1995)	Linares (1996, 1998), Irigoyen et al. (2001, 2006)
	Against		Ladizinsky and Zohary (1968), Ladizinsky (1998), Leggett (1998)	Jellen et al. (1993), Jellen and Gill (1996)		Linares et al. (1998)
A. wiestii	For			Rajhathy and Thomas (1974)		Li et al. (2000b), Irigoyen et al. (2006), Fu and Williams (2008)
A. hirtula	For			Rajhathy and Thomas (1974)		Irigoyen et al. (2006)
Ac genome						
A. canariensis	For	Ladizinsky and Zohary (1971), Baum et al. (1973), Craig et al. (1974)	Thomas (1992)			Li et al. (2000b), Nikoloudakis et al. (2008)
Ad genome						
A. damascena	For					Li et al. (2009)
Al genome						
A. longiglumis	For					Rodionov et al. (2005), Nikoloudakis et al. (2008)

unknown levels of homology among the various diploid genomes (Leggett and Thomas 1995) and high affinity between the chromosomes of the A, B, and D genomes complicates the unraveling of the phylogeny and the identification of the ancestral species giving rise to *Avena* polyploids. All potential hybrids between the A genome diploids and the hexaploids have been produced (Leggett and Thomas 1995); however, none of these hybrid combinations exhibited the level of chromosome homology that would suggest one of them to be the donor of the A genome.

The A-genome diploids were originally subdivided into the As genome (*A. strigosa*) and modified A-genome species (Ac, Ad, Al, and Ap) according to their karyotypes (Rajhathy and Thomas 1974; Leggett 1987). Considerable work has been done in an attempt to determine which of the A genome diploids was the progenitor of the polyploid species of *Avena*; however, many of the results are contradictory. Table 3 provides a summary of the evidence accumulated for and against any one particular A genome species being said progenitor. Our results, together with this other accumulated evidence, seem to suggest that various A genome diploid species might have been involved in the evolution of the A genome in polyploid oats. The A genome *FL int2* sequences could be clearly divided into two groups. Each group included almost all of the AA and AABB genome species used in this study, as well as some of the AACC tetraploids and AACCDD hexaploids. The A genome diploid sequences were interspersed among the tetraploid and hexaploid sequences, meaning these data are not informative in the search for the A genome ancestor(s) of polyploidy *Avena*. This justifies the suggestion that tetraploid AABB, AACC, and hexaploid AACCDD species could have obtained the A genome from any of the diploid A genome species included in this study (Cheng et al. 2003).

The results of the present study also call into question the previous subdivision of A genome species, as do studies using random amplified polymorphic DNA (RAPDs) and restriction fragment length polymorphisms (RFLPs; Nocelli et al. 1999; Li et al. 2000b). Other AsAs genome species share similar characters with the mode AsAs species *A. strigosa*, but some of them also show specific traits that differ from *A. strigosa*. The diploid species *A. atlantica* yielded fertile offspring when crossed with *A. strigosa*, so it was classified within the As genome group. Later, this was confirmed by karyological studies (Leggett 1987), while *A. atlantica* is regarded as a wild analog of *A. strigosa* instead of *A. hirtula* (Rajhathy and Morrison 1959). However, in the present research, *A. atlantica* and *A. hirtula* are dispersed from the genome subclade containing the three AsAs species *A. hispanica*, *A. strigosa,* and *A. wiestii*. RAPD and RFLP analysis (Nocelli et al. 1999), as well as a study based on the polymorphisms of satellite, minisatellite and microsatellite DNA (Li et al. 2000b), support our result. In addition, studies of rDNA sequences (Nikoloudakis et al. 2008) and structural divergence (Badaeva et al. 2005), as well as RAPD and AFLP analyses (Drossou et al. 2004), also showed that *A. hirtula* was separate from the other As genome species. In addition, two clones from As-genome diploid species *A. wiestii* are found in the C genome clade unexpectedly, which indicate this species is different from other A genome species to some extent. This is verified by the ITS sequence data, because *A. wiestii* (GenBank accession no. EU833830–EU833837) had novel ITS sequences which were different from the rest A genome species. But why *A. wiestii* is different from other A genome species remains unknown. Perhaps an intensive study with more population material of this species based on morphological, cytological, and molecular evidence would help clarify this question.

Considering the non-As genome species, almost complete chromosome pairing is documented in *A. damascena* × *A. longiglumis* hybrids (Nocelli et al. 1999), but, even though these species are morphologically very similar, they are almost geographically isolated and differ karyotypically by at least three translocations (Leggett 1984a). The low level of genetic similarity between *A. damascena* and *A. longiglumis* reflects their reproductive isolation (Rajhathy and Baum 1972; Leggett 1984a). Leggett and Thomas (1995) reported that *A. damascena* is more closely related to the *A. strigosa* group than *A. longiglumis*, as did Drossou et al. (2004). The meiotic behavior of hybrids between As genome diploids and *A. damascena* (Leggett 1989) and *A. longiglumis* (Rajhathy and Thomas 1974) supports these results, but other studies have reported that *A. longiglumis* has genetic similarities with *A. strigosa* (Linares et al. 1998; Li et al. 2000b; Loskutov 2001). In AFLP study based on 163 accesion of 25 Avena speices, the As genome species *A. lusitanica* showed more related to *A. damascena* with the Ad genome and *A. longiglumis* with the Al genome (Fu and Williams 2008). Ac-genome species *A. canariensis* shared higher genetic similarity with *Avena wiestii* (AsAs) (Li et al. 2000b). Thus, the previous subdivision of A-genome species appears to be in need of review.

The origin of the B genome in AABB tetraploids

Derivation from the A versus B genome in the tetraploid could not be determined. It was possible that sequences may have been derived from both A and B genomes for AABB tetraploids and indicate multiple copies of each genome, or tetraploids of AABB origin could be expected to carry only A genome specific sequences, since all AABB

species belong to the A genome clades, and it is considered that the B genome was derived from the A genome as most research has suggested (Oinuma 1952; Sadasivaiah and Rajhathy 1968; Fominaya et al. 1988b; Leggett and Markhand 1995; Leggett and Thomas 1995; Katsiotis et al. 1996; Leggett 1996; Katsiotis et al. 1997). However, the A and B genomes have been distinguished using an A genome-specific repetitive sequence from *A. strigosa* (Irigoyen et al. 2001). In our result, there is a subclade in the A-2 clade that includes only the AABB species and two A genome diploids, that may represent the B genome group. If this is true, the B genome of the AABB species *A. abyssinica*, *A. barbata* and *A. vaviloviana* [the interfertile *barbata* group (Rajhathy and Thomas 1974)] should be derived from the diploid *A. hirtula.* That *A. hirtula* is closely related to the AABB tetraploids is supported by the work of others (Rajhathy and Thomas 1974; Irigoyen et al. 2006). The work of Peng et al. (2008) also found that the B genome units were embedded as a branch within the A genome units. That the AABB species *A. agadiriana* is different from the other three is in accordance with the ITS sequence results of Nikoloudakis et al. (2008).

The origin of the C genome in AACC and AACCDD species

There was little evidence favoring any of the three diploid species included in previous studies as being the progenitor of the C genome in AACC tetraploids or in AACCDD hexaploid species. It is noteworthy that, during most of the studies using molecular markers and nucleotide sequences, it has been observed that all *Avena* allopolyploids cluster with the A-genome diploids, and the dominance of A-genome over C-genome rRNA sequences complicates phylogenies (Jellen et al. 1994b; Fominaya et al. 1995; Yang et al. 1999; Shelukhina et al. 2007; Peng et al. 2008; Nikoloudakis and Katsiotis 2008; Li et al. 2009). Therefore, identification of the C-genome species contributing to the speciation of cultivated oats was not possible. Furthermore, there is no information based on cytoplasmic sequences available that would permit us to unravel the evolution of the AACC and the AACCDD polyploid species because the plastid genome is maternally inherited in grasses (Ge et al. 1999), and many studies have suggested that the C genome taxon was not the maternal donor of the polyploids (Nishiyama and Yabuno 1975; Leggett 1984b; Nishiyama 1984; Murai and Tsunewaki 1987; Rines et al. 1988; Nikoloudakis and Katsiotis 2008; Li et al. 2009). As a result, the ancestry of the C genome in polyploids remains uncertain.

On the basis of karyotype, the three diploid C genome species have been separated into two genome types: Cv and Cp (Leggett and Thomas 1995). Both types have been proposed as the putative donors of the hexaploid C genome (Rajhathy and Thomas 1974; Chen and Armstrong 1994; Jellen et al. 1994a, b). Murai and Tsunewaki (1987) analyzed cpDNA of the most representative *Avena* species using RFLPs, and, although both *A. eriantha* and *A. clauda* were assigned the Cp genome, they, along with the Cv genome diploid *A. ventricosa,* belong to different cpDNA types. Numerous studies have shown that the C genome is essentially unchanged through all ploidy levels, thus being considered one of the basic genomes in oats (Li et al. 2000b) and referred to the easy detection of the C genome donor of polyploidy species. In the present study, the Cp genome diploid *A. clauda* was shown to share higher genetic similarity with the AC-genome tetraploids and the hexaploids than with the A-genome diploids, as all of the AACC and AACCDD species were found to be included in the C genome clade. *A. ventricosa* formed an outgroup, leading us to the conclusion that *A. clauda* should be considered the C genome donor of the AACC and AACCDD species rather than *A. ventricosa*, as is generally believed (Rajhathy 1966; Fu and Williams 2008; Nikoloudakis and Katsiotis 2008). This conclusion is supported by Cheng et al. (2003) who also proposed that *A. clauda* was the C genome progenitor of AACC tetraploids and is justified by chromosome pairing results that showed that *A. eriantha* is unlikely to have participated in the formation of the tetraploid *A. maroccana* (Leggett 1998).

Tentative presumption of the D genome origin in AACCDD species

The cultivated oat *A. sativa* is a natural allohexaploid that contains three genomes (A, C, and D). It has been suggested that its evolution involved two distinct steps. The first step involved the establishment of a tetraploid (AACC) by the hybridization of two diploid species (AA and CC) followed by doubling of the chromosome number. The second step involved hybridization of this tetraploid with a third diploid species to form a hexaploid by the doubling of the chromosomes of the resulting triploid hybrid (Rajhathy 1991; Thomas 1992).

Although the identification of the diploid species contributing to the evolution of the hexaploids is still controversial, evidence from studies of chromosome pairing in hybrids between hexaploid and tetraploid oat (Ladizinsky and Zohary 1971; Ladizinsky and Fainstein 1977; Ladizinsky 1974, 1998), isoenzyme analysis (Sánchez de la Hoz and Fominaya 1989), and molecular studies (Drossou et al. 2004; Fu and Williams 2008; Nikoloudakis et al. 2008)favors the involvement of the tetraploid oats *A. maroccana, A. murphyi,* and *A. insularis* (Ladizinsky 1998).

Most research supports the hypothesis that an AACC tetraploid hybridized with another diploid species with

subsequent chromosome doubling to form a hexaploid, yet no diploid species containing the D genome have been found. There is some evidence that the A and D genomes are very closely related (Ladizinsky and Zohary 1968; Ladizinsky 1988; Chen and Armstrong 1994; Leggett et al. 1996; Leggett and Markhand 1995; Linares et al. 1996, 1998; Li et al. 2000a, b; Loskutov 2008). The close relationship of the A and D genomes together with the absence of DD genome diploid species supports the hypothesis that an A genome diploid species could have been the donor of both the A and D genomes of hexaploid oat (Ladizinsky and Zohary 1968; Ladizinsky 1988; Leggett 1996; Loskutov 2008). If this is true, there should be one A-genome diploid species sharing higher genetic similarity with the hexaploid species but lower genetic similarity with the AC genome tetraploid species. No such A genome diploid was found in previous studies.

Linares et al. (1998) differentiated the A and D genomes at the diploid level based on the southern blot analysis, with *A. canariensis* and *A. damascena* as putative D-genome diploids. However, their probe picked up repeated sequences; when the copies are low they will not be picked up under the assays conditions outlined in their paper. Furthermore, the reverse was not carried out and could not have been carried out because there was no known D-genome-specific probe. In the present study, we were able and fortunate to find a possible clue of the D genome origin based on DNA sequencing. In the phylogenetic analyses of the *FL int2* sequences, the two A genome products behaved quite differently: all phylogenetic analyses of *FLint2* failed to place the two types of A genome sequences as sisters. Instead, the assigned D genome sequences clade was found between the two A genome groups. This demonstrates that D genome sequences are related to the A genome sequences to a certain extent. However, the diploid species belonging to the D genome clade are the C genome diploids *A. clauda* and *A. eriantha*, as opposed to any of the A genome diploids. The AC genome species *A. murphyi* also belongs to the D genome clade. This result is in general agreement with the 5S rRNA gene diversity study that found the short D1 class unit that was assigned to the D genome not only from all hexaploids, but also from the diploid *A. clauda* and tetraploid *A. murphyi* (Peng et al. 2008). The *FL int2* sequences of *A. clauda* are scattered through the C and D clades, as was found in the study of 5S rRNA gene diversity (Peng et al. 2008) and in the study using ccSSR markers (Li et al. 2009). *A. clauda* is more divergent than the other two C genome species *A. eriantha* and *A. ventricosa*, as demonstrated by the clustering of some *A. clauda* accessions with AA, AABB, and AACCDD species (Li et al. 2009). Many studies have suggested that *A. murphyi* is different from the other two AACC species *A. maroccana* and *A. insularis* (Ladizinsky 1999; Shelukhina et al. 2007; Peng et al. 2008) and that it shows the closest relationship with the hexaploid species (Li et al. 2000a, b; Cheng et al. 2003).

Information concerning the differentiation of the 5S rDNA unit classes (Peng et al. 2008) and the interpretation of the possible origin of the D genome in the present study shed new light on the evolution of the genus *Avena*. The genome assignment of *Avena* species, which is mainly based on karyotypes and chromosome pairing in interspecific hybrids, may need further verification based on a comprehensive revision using combined evidence. Because there is a genetically controlled regulatory mechanism that prevents homoeologous chromosome pairing in the *Avena* (Ladizinsky 1973, 1974; Rajhathy and Thomas 1974), and because of the numerous translocations and unknown levels of homoeology among the various diploid genomes (Chen and Armstrong 1994; Jellen et al. 1994a; Fominaya et al. 1995; O'Donoughue et al. 1995; Leggett and Markhand 1995; Leggett and Thomas 1995; Linares et al. 1996, 2000), all could lead to an underestimation of the actual relationships at the cytological level between the different genomes.

Phylogenetic utility of *FL int2*

FLO/LFY has been reported as a single copy gene in most diploid angiosperms studied to date (Hofer et al. 1997; Theissen 2000; Frohlich and Parker 2000; Schlüter et al. 2007). Frohlich and Parker (2000) suggested that *FLO/LFY* was duplicated on the stem lineage leading to seed plants, but that one copy was lost in angiosperms, making it a single-copy gene in diploid angiosperms.

Although no diploid bearing functional duplicated copies is known, pseudogenes have been found in some individuals (Southerton et al. 1998; Frohlich and Parker 2000); and in polyploids, the copy number can vary to different extents (Kelly et al. 1995; Frohlich and Parker 2000; Wada et al. 2002). Considering that most nuclear genes are members of multigene families (Vision et al. 2000; Martin and Burg 2002), and differences between two products within 1 taxon could be attributed to differences in the intron only as some other studies indicated (Grob et al. 2004), establishing the number of copies of *FLO/LFY* in the genus *Avena* is an important consideration in the future work.

Although we expected to find a single copy of *FLO/LFY* in *Avena*, tree topology and the degree of divergence between the *FL int2* clones suggest conclusively that more than one copy of the second intron region is present in most *Avena* species. The multiple products gave us very important phylogenetic information that neither chloroplast nor rDNA sequences could have. As our data show, results from phylogenetic analysis of *FL int2* are useful in percentage of informative characters, overall homoplasy

levels, number of well-supported clades in the consensus trees, and resolution of ingroup relationships within *Avena*. The present study justifies using the *FL int2* intron as a phylogenetic character and confirms that it can be a valuable tool for phylogenetic studies at lower taxonomic levels (Ahearn et al. 2001; Hoot and Taylor 2001; Oh and Potter 2003, 2005; Archambault and Bruneau 2004; Grob et al. 2004; Howarth and Baum 2005; Schlüter et al. 2007).

In conclusion, the present study extends the knowledge of the phylogenetic relationships within the genus *Avena* and that of the genome origin of allopolyploids, particularly as pertains to the origin of the B and D genomes. The results reemphasize the major genomic divergence between the A and C genomes, and the close relationship among the A, B and D genomes. However, our data suggest that the D genome in hexaploids may have originated from a previously assigned C genome diploid rather than from the generally believed A genome. The necessity of reviewing the genome differentiation and assignment of *Avena* species is suggested, and the important role of *A. clauda* in the evolution of the C and D genomes in polyploids is discussed. Further research in oat with large mapping populations, genome-specific primers, and (or) large-fragment clones will be important to clarify the phylogentic relationship of *Avena* speices and the polyploid origin.

Acknowledgments We thank Agriculture and Agri-Food Canada (AAFC), the National Small Grains Collection, ARS, USDA, and Dr. Eric N. Jellen, Brigham Young University, Provo, Utah, USA, for kindly providing the Avena material. We are very grateful to Mrs Charlene Wight, Eastern Cereal and Oilseed Research Centre, for her most valuable help and comments on the manuscript. We also wish to thank two anonymous reviewers for helpful comments that improved the manuscript. This work was supported by the Earmarked fund for Modern Agro-industry Technology Research System (NYCYTX-14). Yuan-Ying Peng was supported by the Excellent Doctoral Dissertation Fund of Sichuan Agricultural University (YBPY0802).

References

Ahearn KP, Johnson HA, Weigel D, Wagner DR (2001) *NFL1*, a *Nicotiana tabacum LFY*-like gene, controls meristem initiation and floral structure. Plant Cell Physiol 42:1130–1139

Akaike H (1974) A new look at the statistical model identification. IEEE T Automat Contr 19:716–723

Alicchio R, Aranci L, Conte L (1995) Restriction fragment length polymorphism based phylogenetic analysis of *Avena* L. Genome 38:1279–1284

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Archambault A, Bruneau A (2001) How useful is the *LEAFY* gene for the phylogeny reconstruction in the Caesalpinioideae? Am J Bot 88(Suppl):97

Archambault A, Bruneau A (2004) Phylogenetic utility of the *LEAFY/FLORICAULA* gene in the caesalpinioideae (Leguminosae): gene duplication and a novel insertion. Syst Bot 29:609–626

Badaeva E, Loskutov I, Shelukhina O, Pukhalsky V (2005) Cytogenetic analysis of diploid *Avena* L. species containing the as genome. Russ J Genet 41:1428–1433

Bailey CD, Doyle JJ (1999) Potential phylogenetic utility of the low-copy nuclear gene pistillata in dicotyledonous plants: comparison to nrDNA ITS and trnL intron in Sphaerocardamum and other Brassicaceae. Mol Phylogenet Evol 13:20–30

Baker WJ, Hedderson TA, Dransfield J (2000) Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA rps16 intron sequence data. Mol Phylogenet Evol 14:195–217

Baldwin BG, Preston RE, Wessa BL, Wetherwax M (2001) A biosystematic and phylogenetic assessment of sympatric taxa in Blepharizonia (Compositae-Madiinae). Syst Bot 26:184–194

Baum BR (1977) Oats: wild and cultivated, a monograph of the genus *Avena* L. (Poaceae). Minister of supply and services Canada. Agriculture Canada, Ottawa, Ontario

Baum BR, Rajhathy T, Sampson DR (1973) An important new diploid Avena species discovered on the Canary Islands. Can J Bot 51:4759–4762

Blázquez MA, Soowal LN, Lee I, Weigel D (1997) *LEAFY* expression and flower initiation in Arabidopsis. Development 124:3835–3844

Bomblies K, Doebley JF (2005) Molecular evolution of *FLORICAULA/LEAFY* orthologs in the Andropogoneae (Poaceae). Mol Biol Evol 22:1082–1094

Chen Q, Armstrong K (1994) Genomic in situ hybridization in *Avena sativa*. Genome 37:607–612

Cheng DW, Armstrong KC, Drouin G, McElroy A, Fedak G, Molnar SD (2003) Isolation and identification of Triticeae chromosome 1 receptor-like kinase genes (*Lrk10*) from diploid, tetraploid, and hexaploid species of the genus *Avena*. Genome 46:119–127

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R (1990) Floricaula: a homeotic gene required for flower development in antirrhinum majus. Cell 63:1311–1322

Craig IL, Murray BE, Rajhathy T (1974) *A. canariensis*: morphological and electrophoretic polymorphism and relationship to the *A. magna-A. murphyi* complex and *A. sterilis*. Can J Genet Cytol 16:677–689

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Drossou A, Katsiokis A, Leggett JM, Loukas M, Tsakas S (2004) Genome and species relationships in genus *Avena* based on RAPD and AFLP molecular markers. Theor Appl Genet 109:48–54

Fominaya A, Vega C, Ferrer E (1988a) Giemsa C-banded karyotypes of *Avena* species. Genome 30:627–632

Fominaya A, Vega C, Ferrer E (1988b) C-banding and nucleolar activity of tetraploid *Avena* species. Genome 30:633–638

Fominaya A, Hueros G, Loarce Y, Ferrer E (1995) Chromosomal distribution of a repeated DNA sequence from C-genome heterochromatin and the identification of a new ribosomal DNA locus in the *Avena* genus. Genome 38:548–557

Frohlich MW, Meyerowitz EM (1997) The search for flower homeotic gene homologs in basal Angiosperms and Gnetales: a potential new source of data on the evolutionary origin of flowers. Int J Plant Sci 158:S131–S142

Frohlich MW, Parker DS (2000) The mostly male theory of flower evolutionary origins: from genes to fossils. Syst Bot 25:155–170

Fu YB, Williams DJ (2008) AFLP variation in 25 Avena species. Theor Appl Genet 117:333–342

Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA 96:14400–14405

Grob GBJ, Gravendeel B, Eurlings MCM (2004) Potential phylogenetic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae). Mol Phylogenet Evol 30:13–23

Hillis DM (1987) Molecular versus morphological approaches to systematics. Annu Rev Ecol Syst 18:23–42

Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N (1997) UNIFOLIATA regulates leaf and flower morphogenesis in pea. Curr Biol 7:581–587

Hoot SB, Taylor WC (2001) The utility of nuclear ITS, a *LEAFY* homolog intron, and chloroplast *atpB-rbcL* spacer region data in phylogenetic analyses and species delimitation in *Isoetes*. Am Fern J 91:166–177

Howarth DG, Baum DA (2005) Gencalogical wbidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodenraceae) in the Hawarran islands. Evolution 59:948–961

Irigoyen ML, Loarce Y, Linares C, Ferrer E, Leggett M, Fominaya A (2001) Discrimination of the closely related A and B genomes in AABB tetraploid species of *Avena*. Theor Appl Genet 103:1160–1166

Irigoyen ML, Ferrer E, Loarce Y (2006) Cloning and characterization of resistance gene analogs from *Avena* species. Genome 49:54–63

Jellen EN, Gill B (1996) C-banding variation in the Moroccan oat species *Avena agadiriana* (2n = 4x = 28). Theor Appl Genet 92:726–732

Jellen EN, Phillips RL, Rines HW (1993) C-banded karyotypes and polymorphisms in hexaploid oat accessions (*Avena.* spp.) using Wright's stain. Genome 36:1129–1137

Jellen EN, Philips RL, Rines HW (1994a) Chromosomal localization and polymorphisms of ribosomal DNA in oat (*Avena* spp.). Genome 37:23–32

Jellen EN, Gill BS, Cox TS (1994b) Genomic in situ hybridization differentiates between A/D- and C-genome chromatin and detects intergenomic translocations in polyploid oat species (genus *Avena*). Genome 37:613–618

Jiang J, Gill BS (1994) Different species-specific chromosome translocations in *Triticum timopheevii* and *T. turgidum* support the diphyletic origin of polyploid wheats. Chromosome Res 2:59–64

Jin H, Domier LL, Kolb FL, Brown CM (1999) Identification of quantitative loci for tolerance to barley yellow dwarf virus in oat. Phytopathology 88:410–415

Katsiotis A, Schmidt T, Heslop-Harrison JS (1996) Chromosomal and genomic organization of Ty1-*copia*-like retrotransposon sequences in the genus *Avena*. Genome 39:410–417

Katsiotis A, Loukas M, Heslop-Harrison JS (1997) The close relationship between the A and B genomes in *Avena* L. (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Ann Bot 79:103–109

Kelly AJ, Bonnlander MB, Meeks-Wagner DR (1995) *NFL*, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. Plant Cell 7:225–234

Kianian SF, Egli MA, Phillips RL, Rines HW, Somers DA, Gengenbach BG, Wesenberg DM, Stuthman DD, Fulcher RG (1999) Association of a major groat oil content QTL and an acetyl-COA carboxylase gene in oat. Theor Appl Genet 98:884–894

Kihara H, Nishiyama I (1932) The genetics and cytology of certain cereals III. Different compatibility in reciprocal crosses of Avena, with reference to tetraploid hybrids between hexaploid and diploid species. Jpn J Bot 6:245–305

Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Ladizinsky G (1973) Genetic control of bivalent pairing in the *Avena strigosa* polyploid complex. Chromosoma 42:105–110

Ladizinsky G (1974) Genome relationships in the diploid oats. Chromosoma 47:109–117

Ladizinsky G (1988) Biological species and wild genetic resources in *Avena*. In: Mattsson B, Lyhagen R, Svalof AB (eds) Proceedings of the 3rd international oat conference, Lund, Sweden, 4–8 July, 1988

Ladizinsky G (1998) A new species of *Avena* from Sicily, possibly the tetraploid progenitor of hexaploid oats. Genet Resour Crop Evol 45:263–269

Ladizinsky G (1999) Cytogenetic relationships between *Avena insularis* (2n = 28) and both *A. strigosa* (2n = 14) and *A. murphyi* (2n = 28). Genet Resour Crop Evol 46:501–504

Ladizinsky G, Fainstein R (1977) Intergression between the cultivated hexaploid oat *A. sativa* and the tetraploid wild *A. magna* and *A. murphyi*. Can J Genet Cytol 19:59–60

Ladizinsky G, Zohary D (1968) Genetic relationships between the diploids and tetraploids in the series Eubarbatae of *Avena*. Can J Genet Cytol 10:68–81

Ladizinsky G, Zohary D (1971) Notes on species delimitation, species relationships and polyploidy in *Avena*. Euphytica 20:380–395

Leggett JM (1984a) Morphology and metaphase chromosome pairing in three Avena hybrids. Can J Genet Cytol 26:641–645

Leggett JM (1984b) Cytoplasmic substitutions involving six *Avena* species. Can J Genet Cytol 26:698–700

Leggett JM (1987) Interspecific hybrids involving the recently described taxon *Avena atlantica*. Genome 29:361–364

Leggett JM (1989) Interspecific diploid hybrids in *Avena*. Genome 32:346–348

Leggett JM (1996) Using and conserving *Avena* genetic resources. In: Scoles GJ, Rossnagel BG (eds) Barley chromosome coordinators' workshop at the V international oat conference & VII international barley genetics symposium, Saskatoon, Saskatchewan, Canada, 31 July 1996. University of Saskatchewan, pp 128–132

Leggett JM (1998) Chromosome and genomic relationships between the diploid species *Avena strigosa*, *A. eriantha* and the tetraploid *A. maroccana*. Heredity 80:361–363

Leggett JM, Markhand GS (1995) The genomic structure of *Avena* revealed by GISH. In: Brandham PE, Bennett MD (eds) Kew chromosome conference IV. HMSO, UK, pp 133–139

Leggett JM, Thomas H (1995) Oat evolution and cytogenetics. In: Welch RW (ed) The oat crop: production and utilization. Chapman and Hall, London, UK

Li CD, Rossnagel BG, Scoles GJ (2000a) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. Theor Appl Genet 101:1259–1268

Li CD, Rossnagel BG, Scoles GJ (2000b) Tracing the phylogeny of the hexaploid oat *Avena sativa* with satellite DNAs. Crop Sci 40:1755–1763

Li WT, Peng YY, Wei YM, Baum BR, Zheng YL (2009) Relationship among Avena species as revealed by consensus chloroplast simple sequence repeat (ccSSR) markers. Genet Resour Crop Evol 56:465–480

Linares C, Vega C, Ferrer E, Fominaya A (1992) Identification of C-banded chromosomes in meiosis and the analysis of nucleolar activity in *Avena byzantina* C. Koch cv 'Kanota'. Theor Appl Genet 83:650–654

Linares C, Gónzalez J, Ferrer E, Fominaya A (1996) The use of double FISH to physically map the positions of 5S rDNA genes in relation to the chromosomal location of 18S–5.8S–26S rDNA and a C genome specific DNA sequence in the genus Avena. Genome 39:535–542

Linares C, Ferrer E, Fominaya A (1998) Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. Proc Natl Acad Sci USA 95:12450–12455

Linares C, Irigoyen ML, Fominaya A (2000) Identification of C-genome chromosomes involved in intergenomic translocations in *Avena sativa* L., using cloned repetitive DNA sequences. Theor Appl Genet 100:353–360

Liston A, Robinson WA, Oliphant JM, Alvarez-Buylla ER (1996) Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst Bot 21:109–120

Liu Q, Ge S, Tang H, Zhang X, Zhu G, Lu BR (2006) Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. New Phytol 170:411–420

Loskutov IG (2001) Interspecific crosses in the genus *Avena* L. Rus J Genet 37:467–475

Loskutov IG (2008) On evolutionary pathways of Avena species. Genet Resour Crop Evol 55:211–220

Loskutov IG, Perchuk IN (2000) Evaluation of interspecific diversity in *Avena* genus by RAPD analysis. Oat Newsletter 46 (http://wheat.pw.usda.gov/ggpages/oatnewsletter/v46/)

Marshall HG, Myers WM (1961) A cytogenetic study of certain interspecific *Avena* hybrids and the inheritance of resistance in diploid and tetraploid varieties to races of crown rust. Crop Sci 1:29–34

Martin AP, Burg TM (2002) Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. Syst Biol 51:570–587

Mason-Gamer RJ (2001) Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. Syst Bot 26:757–768

Mason-Gamer RJ, Weil CF, Kellogg EA (1998) Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol Biol Evol 15:1658–1673

Murai K, Tsunewaki K (1987) Chloroplast genome evolution in the genus *Avena*. Genetics 116:613–621

Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Embnew News 4:1–4

Nikoloudakis N, Katsiotis A (2008) The origin of the C-genome and cytoplasm of *Avena* polyploids. Theor Appl Genet 117:273–281

Nikoloudakis N, Skaracis G, Katsiotis A (2008) Evolutionary insights inferred by molecular analysis of the ITS1–5.8S-ITS2 and IGS Avena sp. sequences. Mol Phylogenet Evol 46:102–115

Nishiyama I (1984) Interspecific cross-incompatibility system in the genus *Avena*. Bot Mag 97:219–231

Nishiyama I, Yabuno T (1975) Meiotic chromosome pairing in two interspecific hybrids and a criticism of the evolutionary relationship of diploid *Avena*. Jpn J Genet 50:443–451

Nocelli E, Giovannini T, Bioni M, Alicchio R (1999) RFLP- and RAPD-based genetic relationships of seven diploid species of *Avena* with the A genome. Genome 42:950–959

O'Donoughue LS, Kianian SF, Rayapati PJ, Penner GA, Sorrells ME, Tanksley SD et al (1995) A molecular linkage map of cultivated oat. Genome 38:368–380

Oh S-H, Potter D (2003) Phylogenetic utility of the second intron of *LEAFY* in *Neillia Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. Mol Phylogenet Evol 29:203–215

Oh S-H, Potter D (2005) Molecular phylogenetic systematics and biogeography of tribe Neillieae (Rosaceae) using DNA sequences of cpDNA, rDNA and *LEAFY*. Am J Bot 92:179–192

Oinuma T (1952) Karyomorphology of cereals. Biol J Okayama Univ 1:12–71

Peng YY, Wei YM, Baum BR, Zheng YL (2008) Molecular diversity of 5S rDNA gene and genomic relationships in genus *Avena* (Poaceae: Aveneae). Genome 51:137–154

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256

Potter D, Luby JJ, Harrison RE (2000) Phylogenetic relationships among species of Fragaria (Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences. Syst Bot 25:337–348

Rajhathy T (1966) Evidence and an hypothesis for the origin of the C genome of hexaploid Avena. Can J Genet Cytol 8:774–779

Rajhathy T (1991) The chromosomes of *Avena*. In: Gupta PK, Tsuchiya T (eds) Chromosome engineering in plants: genetics, breeding, evolution. Elsevier Science Publishers, The Netherlands, pp 447–465

Rajhathy T, Baum BR (1972) *Avena damascena*: a new diploid oat species. Can J Genet Cytol 14:645–654

Rajhathy T, Morrison JW (1959) Chromosome morphology in the genus *Avena*. Can J Bot 37:372–377

Rajhathy T, Morrison JW (1960) Genome homology in the genus *Avena*. Can J Genet Cytol 2:278–285

Rajhathy T, Thomas H (1974) Cytogenetics of oats (*Avena* L.). Misc Publ Genet Soc Can, Ottawa

Rines HW, Gengenbach BG, Boylan KL, Storey KK (1988) Mitochondrial DNA diversity in oat cultivars and species. Crop Sci 28:171–176

Roalson EH, Columbus JT, Friar EA (2001) Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and trnT-L (cpDNA) region sequences: assessment of subgeneric and sectional relationships in Carex with emphasis on section Acrocystis. Syst Bot 26:318–341

Rodionov AV, Tyupa NB, Kim ES, Machs EM, Loskutov IG (2005) Genomic configuration of the autotetraploid oat species *Avena macrostachya* inferred from comparative analysis of ITS1 and ITS2 sequences: on the oat karyotype evolution during the early events of the *Avena* species divergence. Russ J Genet 41:518–528

Ronald PS, Penner GA, Brown PD, Brule-Babel A (1997) Identification of RAPD markers for percent hull in oat. Genome 40:873–878

Sadasivaiah RS, Rajhathy T (1968) Genome relationships in tetraploid *Avena*. Can J Genet Cytol 10:655–669

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction phylogenetic trees. Mol Biol Evol 4:406–425

Sánchez de la Hoz P, Fominaya A (1989) Studies of isozymes in oat species. Theor Appl Genet 77:735–741

Schlüter PM, Gudrun K, Stuessy TF, Paulus HF (2007) A screen of low-copy nuclear genes rebeals the *LFY* gene as phylogenetically informative in closely related species of orchids (*Ophrys*). Taxon 56:493–504

Schultz EA, Haughn GW (1991) *LEAFY*, a homeotic gene that regulates inflorescence development in Arabidopsis. Plant Cell 3:771–781

Shelukhina OYU, Badaeva ED, Loskutov IG, Pukhal'sky VA (2007) A comparative cytogenetic study of the tetraploid oat species with the A and C genomes: *A. insularis*, *A. maroccana* and *A. murphyi*. Russ J Genet 43:613–626

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Shitsukawa N, Takagishi A, Ikari C, Takumi S, Murai K (2006) *WFL*, a wheat *FLORICAULA/LEAFY* ortholog, is associated with spikelet formation as lateral branch of the inflorescence meristem. Genes Genet Syst 81:13–20

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301–1315

Soltis DE, Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis DE, Soltis PS, Doyle JJ (eds)

Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Boston, pp 1–42
Soltis DE, Soltis PS, Tate JA (2003) Advances in the study of polyploidy since plant speciation. New Phytol 161:173–191
Southerton SG, Strauss SH, Oliver MR, Hercourt RL, Decroocq V, Zhu X, Llewellyn DJ, Dennis ES (1998) *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. Plant Mol Biol 37:897–910
Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Mass
Tank DC, Sang T (2001) Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: evolution and implications in Paeonia (Paeoniaceae). Mol Phylogenet Evol 19:421–429
Theissen G (2000) Plant breedings: FLO-like meristem identity genes: from basic science to crop plant design. In: Progress in botany, Springer, Berlin, 61:167–183
Thomas H (1992) Cytogenetics of *Avena*. In: Marshall HG, Sorrells ME (eds) Oat science and technology, monograph 33, agronomy series. ASA and CSSA, Madison, Wisconsin, pp 473–508
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117
Wada M, Cao Q, Nobuhiro K, Soejima J, Masuda T (2002) Apple has two orthologs of *FLORICAULA/LEAFY* involved in flowering. Plant Mol Biol 49:567–577
Weigel D (1995) The genetics of flower development: from floral induction to ovule morphogenesis. Annu Rev Genet 29:19–39
Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249
Whitcher IN, Wen J (2001) Phylogeny and biogeography of Corylus (Betulaceae): inferences from ITS sequences. Syst Bot 26:283–298
Wilson CA (2003) Phylogenetic relationships in Iris series Californicae based on ITS sequences of nuclear ribosomal DNA. Syst Bot 28:39–46
Yang Q, Hanson L, Bennett MD, Leitch IJ (1999) Genome structure and evolution in the allohexaploid weed *Avena fatua* L. (Poaceae). Genome 42:512–518