Theor Appl Genet (2005) 111: 1347–1356
DOI 10.1007/s00122-005-0064-6

ORIGINAL PAPER

Hui Wei · Yan Fu · Rajeev Arora

Intron-flanking EST–PCR markers: from genetic marker development to gene structure analysis in *Rhododendron*

Received: 6 April 2005 / Accepted: 26 July 2005 / Published online: 16 September 2005

Abstract With a long-term goal of constructing a linkage map of *Rhododendron* enriched with gene-specific markers, we utilized *Rhododendron catawbiense* ESTs for the development of high-efficiency (in terms of generating polymorphism frequency) PCR-based markers. Using the gene-sequence alignment between *Rhododendron* ESTs and the genomic sequences of *Arabidopsis* homologs, we developed 'intron-flanking' EST–PCR-based primers that would anneal in conserved exon regions and amplify across the more highly diverged introns. These primers resulted in increased efficiency (61% vs. 13%; 4.7-fold) of polymorphism-detection compared with conventional EST–PCR methods, supporting the assumption that intron regions are more diverged than exons. Significantly, this study demonstrates that *Arabidopsis* genome database can be useful in developing gene-specific PCR-based markers for other non-model plant species for which the EST data are available but genomic sequences are not. The comparative analysis of intron sizes between *Rhododendron* and *Arabidopsis* (made possible in this study by aligning of *Rhododendron* ESTs with *Arabidopsis* genomic sequences and the sequencing of *Rhododendron* genomic PCR products) provides the first insight into the gene structure of *Rhododendron*.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-0064-6

Communicated by O. Savolainen

H. Wei · R. Arora (✉)
Department of Horticulture, Iowa State University, Ames, IA, 50011 USA
E-mail: rarora@iastate.edu
Tel.: +1-515-2940031
Fax: +1-515-2940730

Y. Fu
Interdepartmental Genetics Graduate Program, Department of Genetics, and Developmental and Cellular Biology, Iowa State University, Ames, IA, 50011 USA

Keywords *Rhododendron* · Intron · EST–PCR marker · Phylogeny · Comparative genomics · Cold hardiness

Introduction

Rhododendrons are among the most important woody landscape plants. Broad-leafed members of this family belong to genus *Rhododendron*, the largest in the heath family (i.e. Ericaceae) comprising ~1,000 species. Among them, over 800 species are distributed throughout the Northern Hemisphere, ranging from tropical to polar climates and varying widely in their cold hardiness (Leach 1961; Sakai et al. 1986).

The primary goal of most *Rhododendron* breeders working in cold climates has been to combine the ornamental diversity found in less cold-hardy tropical taxa with the superior cold hardiness of several North American and Asian species. However, limited understanding of the genetic control of freezing tolerance has hampered the efforts in cold- hardiness breeding. A goal of our research is to identify molecular markers linked to genes, which control cold-hardiness in *Rhododendron* (Lim et al., 1999). Placing cold-hardiness genes on a linkage map of rhododendrons (once available) could allow for "marker-assisted" selection of cold-hardy genotypes.

Various types of molecular markers including randomly amplified polymorphic DNA (RAPD) (Iqbal et al. 1995), amplified fragment length polymorphism (AFLP) (Escaravage et al. 1998; De Riek et al. 1999; Pornon et al. 2000), and restriction fragment length polymorphism (RFLP) (Milne and Abbott 2000; Dunemann et al. 1999) have been employed for DNA fingerprinting in *Rhododendron*. Using RAPD, RFLP, and microsatellite markers, Dunemann et al. (1999) constructed a genetic linkage map for a *Rhododendron* population. However, to our knowledge, the PCR-based markers derived from expressed sequence tags (ESTs), i.e. EST–PCR markers, have not been reported for DNA fingerprinting and/or genetic mapping in *Rhododendron*.

ESTs are generated from single-pass sequencing of randomly picked cDNA clones (Adams et al. 1991). The EST approach and subsequent gene-expression profiling (cDNA microarrays) have proven to efficiently identify genes and analyze their expression during different developmental stages, or under various environmental stresses (Fowler and Thomashow 2002; Milla et al. 2002; Bhalerao et al. 2003; Dubos and Plomion 2003; Dhanaraj et al. 2004; Wei et al. 2005). ESTs are also useful for providing markers for genome mapping (Boguski and Schuler 1995; Hudson et al. 1995; Picoult-Newberg et al. 1999; Eujayl et al. 2002): since they target specific genes, EST-derived markers are particularly useful for QTL mapping. Recently, our group has released *Rhododendron* ESTs using cDNA libraries derived from non-hardened and cold-hardened leaf tissues (Wei et al. 2005), which provide an opportunity to obtain EST-based markers for use in *Rhododendron* populations.

A potentially significant, but relatively unexplored, application of EST datasets is to conduct a comparative analysis of the genomes between less-intensively studied, non-model species (e. g. *Rhododendron*) and well-studied model plants such as *Arabidopsis*. Bioinformatic tools coupled with PCR and sequencing technologies can allow us to use the coding sequence information (i.e. ESTs) of a given species, such as *Rhododendron*, in conjunction with *Arabidopsis* genome database to learn about the non-coding regions (i.e. introns) of *Rhododendron* genes vis-a-vis *Arabidopsis*.

Choi et al. (2004) recently used *Medicago truncatula* ESTs and their alignment with the genomic sequences of *Arabidopsis* homologs to design intron-targeting primers for developing EST-based markers to construct a genetic map of *M. truncatula*. Such EST-specific primers allow the amplification of genomic DNA across intron regions producing the PCR products that exhibit size or presence/absence polymorphisms. The basic assumption for this strategy is that introns contain more DNA polymorphisms than exons: non-coding regions (introns) evolve much faster than the coding regions (exons) (Small et al. 2004). Therefore, intron-targeting strategy of primer design is expected to yield higher polymorphism frequency (and therefore more efficiency) than other EST–PCR-based conventional strategies.

Specifically, the objectives of the current study were 3-fold. First, because of the scarcity of *Rhododendron* genomic sequences, the alignment of *Rhododendron* ESTs (i.e. exons) with *Arabidopsis* genomic sequences (exon + intron) was used to predict the exon/exon junction sites in *Rhododendron* ESTs; such information was then used to design intron-flanking primers. The second objective was to use these primers in genomic PCR of several *Rhododendron* species and compare the efficiency of polymorphism-detection using intron-flanking primers with those designed without involving intron-flanking. The third objective was to compare the gel- and sequencing-based data on *Rhododendron* intron size with their *Arabidopsis* counterparts to draw inferences about comparative genomic structure, particularly introns, of the two genera.

Considering the far phylogenetic distance between *Rhododendron* and *Arabidopsis* (which is much more than that between *Medicago* and *Arabidopsis* in the study of Choi et al. 2004), the approach used in the current study is anticipated to have a wider application for the development and efficient use of EST–PCR-based markers (for genetic mapping) among the less-studied yet economically important plant species, which may only have EST sequences but lack genomic sequence information; *Medicago* and *Arabidopsis* belong to eurosids clade within rosids whereas *Rhododendron* falls in the asterids group (APG 2003). The sequencing results of genomic PCR products obtained in this study provide the first insight into the comparative analysis of genomic structure of two distantly related plant species, i.e. *Rhododendron* and *Arabidopsis*.

Materials and methods

Generation, assembly and annotation of ESTs

5′ and 3′ ESTs were generated from two cDNA libraries of the leaf tissues (summer-collected, non-cold-hardened and winter-collected, cold-hardened, respectively) of *Rhododendron catawbiense* 'catalgla' (Wei et al. 2005). All ESTs were deposited into GenBank EST database, assembled, and annotated using the same procedures as previously reported (Wei et al. 2005). The sequences of the unique transcripts after assembly are available upon request.

Sequence comparison between *Rhododendron* ESTs and *Arabidopsis* genes

The full *Arabidopsis* gene set (AGI) with introns and UTRs (Version: 3/Feb/2004) was downloaded from TAIR web site (http://www.arabidopsis.org) and formatted into a database for local BLASTN search using assembled rhododendron transcripts as queries. The output file, generated from the E-value cutoff for BLASTN alignment as 1e-4, was used for the further analyses.

Strategy for developing EST–PCR Primers

Three distinct groups of EST–PCR primer pairs were designed to compare their PCR success rates and polymorphism-generating efficiency. The group 1 primers were designed by the conventional method, i.e. not targeting any specific region within ESTs. Second and third groups of primers were designed based on the 'exon/exon' junction site information, inferred from the pairwise alignments between the *Rhododendron (R. catawbiense)* ESTs, and their *Arabidopsis* homologs'

genomic sequences, i.e. exons (corresponding to *Rhododendron* ESTs) + introns. The strategy used for designing group 2 and 3 primers is illustrated in Fig. 1. Basically, as per this pair-wise alignment strategy, the predicted 'exon/exon' junctions within *Rhododendron* ESTs should essentially correspond to 'exon/intron' junctions (i.e. splice sites) within predicted *Rhododendron* genomic sequences (Fig. 1; step 1); throughout the paper, the terms "exon/exon junction" and "exon/intron junction" are used to refer to ESTs and genomic sequence, respectively.

The forward and reverse primers in group 2 were designed from the sequence within a particular *Rhododendron* exon (Fig. 1; step 2), whereas group 3 primers were designed based on the EST sequence that flanks at least two exon/exon junctions (Fig. 1; step 2). These primers are therefore expected to span the 'predicted' rhododendron intron regions in a PCR, hence termed as 'intron-flanking' primers. Importantly, to avoid the situation where primers might cross (instead of just flanking) the predicted exon/exon junction sites potentially resulting in a failed genomic PCR, 25-bp primers at the 5′ and 3′ ends of the predicted exon/exon junction sites were excluded from the primer design.

All primers were designed using Primer3 (Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/cgi-bin/primer3) with the default settings. The primers used for amplification or sequencing are listed in Supplementary Table 1.

Plant materials and genomic DNA extraction

Six species of genus *Rhododendron* were used in this study. Among them, four (*R. catawbiense*, originated from NE America; *R. ponticum*, W. Europe; *R. arbor-*

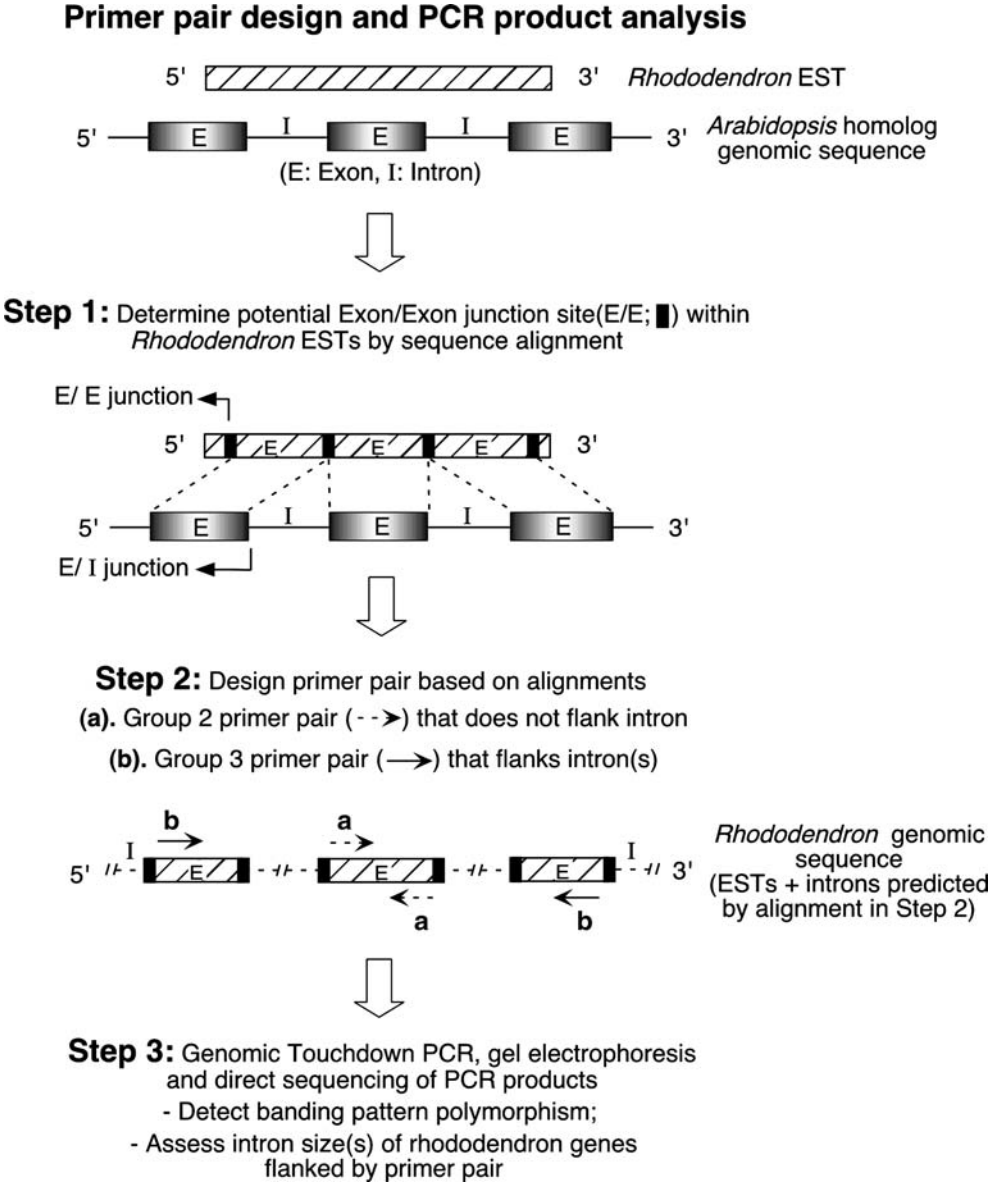

Fig. 1 Schematic representation of the designing of EST-specific primers that do not flank or do flank introns (designated as group 2 or group 3 primers in the text, respectively) and their use in amplifying PCR product from *Rhododendron* genomic DNA. The primer pairs in both groups were designed to avoid exon/intron junction. The sizes of intron(s) flanked by the Group 3 primer pair (*solid arrow*) can be calculated as: "genomic PCR product size"—"EST (cDNA) length between forward and reverse primers". Genomic PCR product size can be determined from gel-estimation and / or sequencing of PCR bands. Intron sizes flanked by Group 2 primer pair (*dotted arrow*) equal zero, since they do not amplify across intron(s)

eum, S. Asia; *R. dichroanthum*, S. China) belong to the subgenus *Hymenanthes*, which are described as elepidote (E) that lack scales on their abaxial leaf surface. Another two (*R. minus* var minus, NE America; *R. keiskei*, Japan) belong to the subgenus of *Rhododendron*, which are described as lepidote (L) that have these scales. Young leaves from the spring growth were sampled in May 2004 from potted plants being maintained at the Horticulture Experimental Station, Iowa State University. Leaves were rinsed in ice-cold distilled and deionized water and stored in −80°C freezer until DNA extraction. Two grams of leaf tissue was ground in liquid N_2 and genomic DNA was extracted using the CTAB method described by Doyle and Doyle (1990). DNA quantification was performed by gel electrophoresis analysis.

Touchdown PCR amplification and sequencing of genomic PCR products

PCR amplification of genomic DNA was carried out by "touchdown PCR" procedure (Don et al. 1991), which was slightly modified (to minimize nonspecific amplification) as given below: PCR mixtures were incubated for 2 min at 92°C, followed by 10 cycles of denaturation at 94°C for 30 s, annealing for 30 s at selected temperatures (see below) and elongation at 72°C for 1 min and another 24 cycles of 94°C for 30 s, 61°C for 30 s and 72°C for 1 min followed by a final 10 min extension at 72°C. The annealing temperature was decreased by 0.8°C per cycle during the first 10 cycles from 69 to 61°C to increase the specificity of amplification. PCR products were separated by agarose gel electrophoresis; band sizes were determined against "I kb plus DNA ladder" (Invitrogen Life Technologies, Carlsbad, CA, USA).

To test the PCR specificity, selected PCR bands were gel-purified using QIAquick spin columns (QIAGEN USA, Valencia, CA, USA) and sequenced from both directions. The obtained genomic sequences were aligned with the corresponding EST sequences via GeneSeqer (Schluter et al. 2003) using *Arabidopsis* parameters to assess the PCR amplification specificity and determine the exon/intron structure.

Detection and evaluation of PCR band pattern polymorphism

Size and presence/absence (+/-) polymorphisms for genomic PCR bands were analyzed by 1.2% (w/v) agarose gel electrophoresis. A given EST-specific marker was considered polymorphic, for which at least one *Rhododendron* species' PCR banding pattern was different from the other five with regard to either size or the presence *vs*. absence. The degree of polymorphism was calculated using the polymorphic information content (PIC; Anderson et al. 1993), which is also referred to as "diversity index" (Milbourne et al. 1997):

$$\mathrm{PIC} = 1 - \sum_{i}^{n} p_i^2$$

where p_i is the frequency of an individual genotype (PCR product banding pattern) generated by a given EST–PCR primer pair.

The PIC value reflects the discriminatory power of a marker by taking into account not only the number of banding patterns generated by a primer pair but also the relative frequencies of these banding patterns. In the current study, PIC values ranged from 0 (no polymorphism) to 0.83 (most polymorphism, i.e. each of the six species having a different banding pattern).

Results

EST generation, assembly and gene identification

We have previously generated a total of 862 5′ ESTs (GenBank accession Nos. CV014938 – 015799) from two cDNA libraries prepared from the non-cold-hardened and cold-hardened leaf tissues (leaf freezing tolerance of −53°C and −7°C, respectively) of *Rhododendron catawbiense* 'Catagla' (Wei et al. 2005); comparative analysis of these ESTs was used to identify major genes involved in cold acclimation of *Rhododendron*. In the current study, we included an additional 378 3′ ESTs (GenBank accession Nos. DN161677–162054) from the same libraries. These 1,240 ESTs were assembled into 814 unique transcripts (172 contigs and 642 singletons). The BLASTX search (E-value: 1e-4) of PIR-NREF protein database results show that ~79% (639/814) had significant protein hits. These hits included several candidate genes potentially involved in the development of cold tolerance in *Rhododendron*, including early light-induced proteins (ELIP), chlorophyll a/b-binding protein, and trehalose 6-phosphate synthase, etc., as previously described (Wei et al. 2005).

Sequence comparison between *Rhododendron* and *Arabidopsis* genes

The BLASTN search returned 503 pair-wise alignments (out of 814 used in the analysis) between *Rhododendron* ESTs and *Arabidopsis* genomic sequences with E-value less than 1e-4. Distribution plots for the alignment similarity (Fig. 2a) and length (Fig. 2b) show a single mode (i.e. peak) average alignment similarity and median alignment length were 86 ± 4% and 86 bp, respectively.

Specificity of touchdown PCR and accuracy of gel-based size estimation of PCR bands

To test the specificity of genomic PCR in this study, we selected seven genomic PCR bands produced by the

intron-flanking EST–PCR primers (each of these primers generated a single band for *R.catawbiense*) and sequenced them after gel-purification. Our analysis of these sequences revealed that they perfectly matched the corresponding EST sequences as determined by GeneSeqer (nucleotide data not shown), confirming that the ESTs were indeed derived from the amplified genomic sequences and that the PCR specificity was satisfactory.

Sequencing of the seven randomly selected genomic PCR bands also enabled us to obtain their precise sizes and compare them with those estimated from the gel-based analysis. Results indicated that the gel-based size-estimation of the PCR bands was essentially similar (i.e. within $2.8 \pm 1.1\%$) to that derived from sequencing analysis (data not shown), highlighting the validity of gel-based size-estimation, which provides the basis for analyzing PCR band size polymorphisms (see "Polymorphism rate comparison" section).

It is noteworthy that out of the 43 intron-flanking primers that resulted in PCR amplification (Table 1), two primers (C24, C42) generated more than one (i.e. two or three) bands using *R. catawbiense* genomic DNA as the template; this resulted in a rate of approximately 5% for primers generating multiple PCR bands. This ratio was comparable with a previous report where 4% (336 out of 7718) of EST–PCR primers amplified two to three DNA fragments using rice genomic DNA as the template (Wu et al. 2002). In our study, since primers C24 and C42 repetitively amplified multiple DNA fragments from several *Rhododendron* species (data not shown), the multiple bands in PCR amplifications were unlikely to be the result of erroneous amplification. These multiple bands may be due to paralogous sequences with high similarity in the genome; alternatively, these primer pairs may span a region of introns with high variability. It may also be caused by the rearrangements of specific genes during the evolution process. Future studies involving sequencing of individual bands and isolation of their full-length genes would provide insight into some of these questions.

PCR success rate comparison

As described in "Methods", three types of EST–PCR primers were used in this study for *Rhododendron* genomic PCR amplification and subsequent polymor-

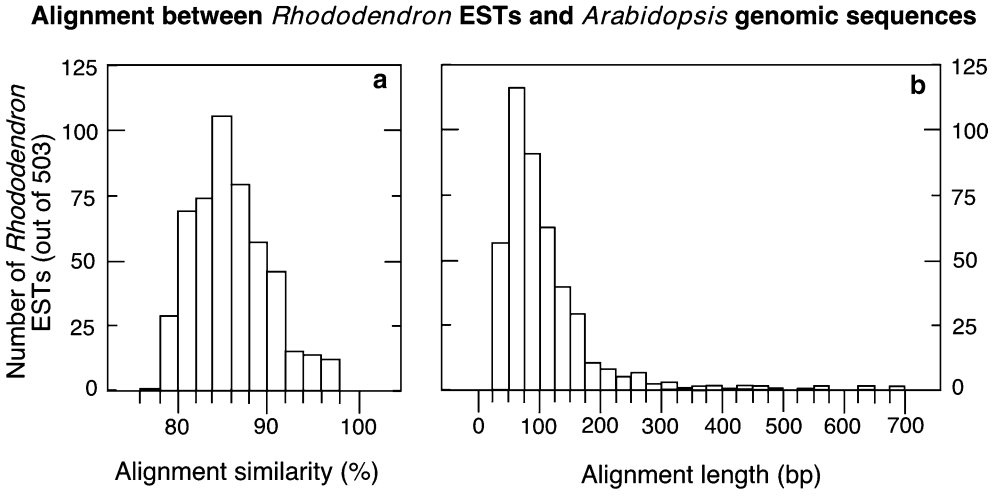

Fig. 2 Histograms for the similarity score (**a**) and the length (**b**) of all 503 alignments between *Rhododendron* ESTs and *Arabidopsis* genes

Table 1 The genomic PCR success rate, PCR band polymorphism rate, overall polymorphism-generating efficiency, and average polymorphism information content (PIC) using (group 1) primers that do not target specific regions within ESTs, (group 2) primers that do not flank introns, and (group 3) intron(s)-flanking primers, respectively

	Primers not targeting specific EST regions (Group 1)	Primers based on alignment between *Rhododendron* ESTs and *Arabidopsis* genomic sequences	
		Primers *not* flanking intron(s) (Group 2)	Intron-flanking primers (Group 3)
No. of primer pairs	8	8	44
PCR success rate	63% (5/8)	100% (8/8)	98% (43/44)
Polymorphism rate	20% (1/5)	0% (0/8)	63% (27/43)
Overall Polymorphism-generating efficiency[a]	13%(1/8)	0%(0/8)	61%(27/44)
Average PIC	0.10	0.00	0.27

[a]Overall polymorphism-generating efficiency is calculated as "PCR success rate" × "Polymorphism rate"

Table 2 Comparative size-analysis of *Rhododendron* introns and corresponding *Arabidopsis* introns based on *R. catawbiense* genomic PCR products amplified by using intron-flanking primer pairs (designated as 'Group 3' in the text)

Primer ID	*R. catawbiense* (R.c.) genes	*A. thaliana* genes (A. t.) homolog genes		Ratio of intron size (R.c./A.t.)
	Size of primer-spanned intron(s) (bp)[a]	ID	Size of corresponding intron(s) (bp)[b]	
C1	735	At5g28840	163	4.5
C2	633	At2g39730	281	2.3
C3	2855	At1g30110	490	5.8
C5	649	At5g58330	151	4.3
C6	536	At3g54890	178	3.0
C7	86	At5g13930	86	1.0
C8	1393	At3g26060	185	7.5
C10	1377	At5g67370	176	7.8
C11	645	At4g00550	180	3.6
C12	537	At2g13360	119	4.5
C13	123	At5g64040	92	1.3
C14	660	At5g05690	85	7.8
C15	935	At5g25760	82	11.4
C16	360	At2g40490	101	3.6
C17	1000	At3g55620	473	2.1
C18	193	At3g24170	314	0.6
C20	568	At2g22780	154	3.7
C22	102	At3g47470	169	0.6
C23	343	At5g54270	64	5.4
C24	283	At3g14415	171	1.7
C25	2019	At1g65930	619	3.3
C26	181	At1g72370	79	2.3
C27	1223	At5g44110	163	7.5
C29	1998	At5g46110	518	3.9
C32	2049	At1g60070	397	5.2
C33	354	At1g78870	201	1.8
C34	884	At2g36060	199	4.4
C35	580	At4g27130	124	4.7
C36	384	At3g04120	89	4.3
C39	1312	At2g02760	420	3.1
C40	1025	At5g01530	287	3.6
C41	628	At2g21330	372	1.7
C42	777	At1g18080	593	1.3
C43	1616	At3g60190	525	3.1
C44	677	At1g20630	282	2.4
C47	841	At1g60470	181	4.6
C48	124	At1g52890	97	1.3
C57	493	At3g50980	301	1.6
C58	382	At2g04350	268	1.4
C59	1599	At2g32260	696	2.3
C60	1498	At1g68020	98	15.3
C62	416	At1g65930	262	1.6
C63	149	At3g22840	144	1.0
Average ± SE (bp)				3. 8 ± 0.4

[a]Length of the primer-spanned intron(s) in *Rhododendron* genomic PCR products was calculated as described in Fig. 1 (i.e. "genomic PCR product size" – "EST length between primer pair")[b]The calculation of the sizes of corresponding *Arabidopsis* intron(s) was based on the alignment of designed primer pair onto *Rhododendron* EST, and the pair-wise alignment between *Rhododendron* EST and the genomic sequence of their *Arabidopsis* homologue which would indicate which *Arabidopsis* intron(s) correspond to the *Rhododendron* intron(s) spanned by the primer pairs

phism detection. For the eight primers of group 1, designed by a conventional method, i.e. not targeting specific regions within ESTs, five primers produced amplifications; the PCR success rate was 63% (Table 1). For the eight primers of group 2, designed from the same exon (thus flanking no introns), all the primers successfully resulted in PCR amplification with the success rate of 100% (Table 1). For the 44 primers of group 3 (designed using the intron-flanking strategy), 43 primers produced PCR amplification with the success rate of 98% (Table 1). Overall, the PCR success rate of the primer pairs designed based on the alignment between *Rhododendron* ESTs and *Arabidopsis* genomic sequences (which include Group 2 and 3 primer pairs) was 35∼37% higher than that obtained by group 1 ("control") primers (Table 1). All PCR results are also summarized in Supplementary Table 1.

Polymorphism rate comparison

With a long-term goal of constructing a *Rhododendron* genetic map of EST-based genetic markers, the designed primers and the genomic DNA from 6 *Rhododendron* species as templates were used to examine PCR band pattern polymorphism. As described in "Methods", a given EST–PCR marker was considered polymorphic when at least PCR product band pattern of one species was different from the other five with regard to either size or the presence vs. absence of polymorphism. The calculated PIC values reveal the degree of diversity in polymorphism across the surveyed species.

For the group 1 EST–PCR primers targeting no specific EST regions, the polymorphism rate was 20% (1/5) (Table 1). The PIC value for the sole primer (Primer ID. A4) that resulted in polymorphism is 0.78, with an overall average of 0.10 (0.78/8) (Table 1). In contrast, for the group 2 EST–PCR primer pairs, specifically designed not to span introns, the polymorphism rate was zero (Table 1)—gel-based sizes of the all the eight genomic PCR products (one band/PCR product) equaled the EST lengths between forward and reverse primers (i.e., expected PCR band size using cDNA as template) (Supplementary Table 1), confirming that the group 2 primers indeed did not span any intron.

In contrast, for the group 3 intron-flanking EST–PCR primers, the polymorphism rate was 63% (27/43), i.e. three times the rate when using primers designed by a conventional method (Table 1). The overall polymorphism detection efficiency (PCR success rate × polymorphism rate) of intron-flanking primers was 4.7-fold (61% vs. 13%) compared to the conventional method (Table 1). Accordingly, the average PIC value for intron-flanking primers was 0.27, nearly 3-fold compared to that of group 1 primers designed by conventional method (Table 1). The PIC values for all primers and their genomic PCR product polymorphism can be found in Supplementary Table 1. None of the group 2 primer pairs (primer ID. B1-B8) resulted in polymorphism.

Comparative analysis of intron-size between *Rhododendron* and *Arabidopsis* genes

Data indicate that the gel-estimated sizes of all genomic PCR products amplified by intron-flanking primer pairs are larger than the corresponding EST lengths between the forward and reverse primers (i.e. the expected PCR product when using cDNA as a template) (see Supplementary Table 1). As described in Fig. 1, the difference between the former and the latter values represents the size of *Rhododendron* intron(s) flanked by the primers. Meanwhile, the pair-wise alignment between *Rhododendron* ESTs and the genomic sequence of their *Arabidopsis* homologs would indicate which *Arabidopsis* intron(s) correspond to the *Rhododendron* intron(s) spanned by the primer pairs, thus allowing a comparative analysis of intron sizes between *Rhododendron* genes and *Arabidopsis* homologs as shown in Table 2.

Intron length data for 43 genes indicated that, with only four exceptions (Primer ID. Nos. C7, C63, C18, and C22), intron(s) of *Rhododendron* genes were substantially larger than that of corresponding *Arabidopsis* introns with an overall intron size ratio (*R.c./A.t.*) of 3.8 ± 0.4 (Table 2); *Rhododendron* intron(s) amplified by primers C7 and C63 were approximately of the same size as corresponding *Arabidopsis* introns, whereas those amplified by C18 and C22 primers were smaller.

Discussion

Intron-flanking EST–PCR markers: PCR success rate and polymorphism-generating efficiency

EST databases of various plant species are becoming a valuable source of PCR-based gene-specific markers for DNA fingerprinting and gene mapping. However, as described below, the potential utility of PCR-based EST markers (in terms of their ability to generate high frequency of genomic PCR band polymorphism) in many non-model plants has not been fully exploited due to the unavailability of genomic DNA sequence database for these species. It is generally believed that intron regions are more divergent than exons (Choi et al. 2004). It follows, therefore, that the EST–PCR primers that are designed to anneal to conserved regions of exons and amplify across intron regions should result in relatively higher rate of genomic PCR band polymorphism, making the gene mapping more efficient. The current study employs alignment of *Rhododendron* ESTs with the genomic sequence of *Arabidopsis* homologs to predict the potential exon/intron junction sites in *Rhododendron* genomic sequence, and uses this information to design 'intron-flanking' EST-specific PCR primers that amplify intron regions in genomic PCR (Fig. 1).

In the current study, conventionally designed primers (group 1) did not target specific regions within ESTs. Their genomic PCR success rate and PCR band polymorphism rate were 63% (5/8) and 20% (1/5), respectively (Table 1). These observations are consistent with the study of Temesgen et al. (2001), who used EST–PCR markers designed from cDNA clones of loblolly pine libraries using the similar method and noted these two values to be 58–76 and 5.3%, respectively.

In contrast, groups 2 and 3 primers resulted in higher genomic PCR success (100 and 98%, respectively) (Table 1). These primers were designed by taking into account the exon/intron junction site information obtained by aligning *Rhododendron* ESTs with genomic sequences of *Arabidopsis* homologs (Fig. 1). High genomic PCR success rate of these primers can be attributed to the avoidance of the exon/intron junction sites—if a 20-nucleotide long primer is designed for crossing the exon/intron junction site, the

genomic PCR will fail because of the inefficiency in primer–template alignment.

Our results indicate that the 'intron-flanking' EST–PCR primers (group 3) resulted in the highest genomic PCR band polymorphism (63 vs. 0–20% for the other two primer types) (Table 1). When coupled with high PCR success rate, these primers were the most efficient group in terms of overall polymorphism-generating ability (61 vs. 0–13% for the other two primer types) (Table 1). The observed high PCR band polymorphism using intron-flanking strategy support our assumption that introns are relatively less conserved and prone to generate more PCR band-polymorphisms than exons.

It is noteworthy that PCR amplifications using EST-specific primers are generally followed by restriction enzyme digestion, heteroduplex analysis, denaturing gradient gel electrophoresis, or single-stranded conformational polymorphism (SSCP) gels (Harry et al. 1998; Plomion et al. 1999; Cato et al. 2001; Temesgen et al. 2001; Rowland et al. 2003b). In the current study, however, genomic PCR products were directly electrophoresed on agarose gels for polymorphism screening. It is anticipated that the polymorphism-generating efficiency of EST-specific markers can be further augmented by using the above analyses after PCR amplification. In other words, the actual polymorphism rate using intron-flanking strategy could potentially be even higher than what we observed in the current study.

Comparative intron analysis of *Rhododendron* and *Arabidopsis* genes

Although ESTs are referred as "poor man's genome", they can be invaluable resource for predicting exon/intron boundaries and as the foundation sequences for genomic analysis (Rudd 2003). The approach described in the current study further extends the utility of ESTs from their use in gene mapping to gene structure analysis.

Analysis of the data presented in Table 2 revealed that the intron(s) lengths spanned in *Rhododendron*'s genomic PCR products were 3.8-fold the corresponding *Arabidopsis* introns. This is the first report on the comparison of intron sizes between *Rhododendron* genes and *Arabidopsis* homologs. It is worth noting that our conclusions about their relative intron sizes are consistent with the difference in genome sizes of these two distant species: *Rhododendron* and *Arabidopsis* genomes are reported to be 0.55 – 0.75 pg DNA/1C (i.e. ~550–750 Mbp; Vainola 2000) and 0.16 pg DNA/1C (i.e. ~160,Mbp; Bennett et al. 1997), respectively; and therefore *Rhododendron* genome is approximately four fold that of *Arabidopsis*.

The relatively larger size of genome and introns of *Rhododendron*, a highly outcrossing woody plant, compared with that of *A. thaliana* is consistent with the similar genomic comparison between *A. lyrata* (a self-incompatible outcrossing member of Brassicaceae) and *A. thaliana* (highly self-fertilizing). *A. lyrata*'s genome size as well as intron lenghts are substantially greater than that of *A. thaliana* (Wright et al. 2002), suggesting that a decrease in the sizes of noncoding regions is likely in *A. thaliana*.

Sizes of individual introns and nucleotide sequences at exon/intron junctions of *Rhododendron* genes

Sequencing of seven *R. catawbiense* genomic PCR bands and their alignments with corresponding ESTs were used to determine the size of individual introns. The results indicated that the sizes of 12 introns (resulting from seven genomic PCRs due, presumably, to the spanning of >1 introns by some primer pairs) ranged from 81–1,485 bp; most of them between 81 and 555. The observed lower range of *Rhododendron* intron size (81 bp) is similar to the reported minimal size (78 bp) for the introns spanned by EST–PCR markers of *M. truncatula* (Choi et al. 2004). It is generally believed that a certain minimum intron size (~70 bp) is required for correct splicing of introns (Deutsch and Long 1999): intron-splicing is believed to remove introns from eukaryotic precursor mRNA such that the adjacent exon sequences join together to form mature mRNA for translation. Furthermore, analysis of the nucleotide sequences at exon/intron junctions demonstrated that, for all the introns surveyed, the first two nucleotides at the 5′ end were "gt" and the last two at the 3′ end were "ag". Such observation is consistent with reported canonical "gt–ag" intron splice-site in both plant and animal genes (Senapathy et al. 1990; Simpson et al. 1993; Thanaraj 1999).

Application of *Rhododendron* EST–PCR markers for fingerprinting and genetic mapping

Various molecular makers have been used for DNA fingerprinting and genetic mapping (for review, see Liu and Corder 2004), which can be divided into hybridization-based *vs.* PCR-based markers; the latter are generally more cost-efficient. The PCR-based markers used in woody plant genome mapping include RAPD markers, simple sequence repeat (SSR) markers, and EST markers (Komulainen et al. 2003; http://dendrome.ucdavis.edu/Gen_Page_body).

EST–PCR markers are derived from gene-coding regions as opposed to AFLP and RAPD markers, which are derived from non-expressed sequences. EST-based markers are, therefore, more likely to be conserved across populations and species and thus applicable to those species for which EST sequences may be lacking. The observed high PCR success rates of EST–PCR based primers across the six *Rhododendron* species in the current study confirm that although the EST sequences were derived from only one species (*R. catawbiense*), they are highly conserved across other five species

screened. Similarly, a recent study by Rowland et al. (2003a) showed that the EST–PCR primers designed from the blueberry (*Vaccinium* sp.) ESTs can be used to amplify genomic DNA from two other Ericaceous species, i.e. rhododendrons and cranberries.

As described in 'Results', of the total 814 *Rhododendron* transcripts, 311 did not have significant alignments with *Arabidopsis* genes using the E-value of 1e-4 as a cut-off. This observation may have two ramifications: first, the current study used exclusively those transcripts that had relatively high similarity (503/814; E-value < 1e-4) with *Arabidopsis* genes, which are likely to be more conserved across plant species. It follows that even greater polymorphisms in genomic PCR product banding patterns could be achieved if *Rhododendron* ESTs with lower similarity score (with corresponding *Arabidopsis* genomic sequences) are used to design EST–PCR primers. Second, for the *Rhododendron* transcripts that have a lower or no similarity with *Arabidopsis* genes, the genomic sequences of phylogenetically closer species (such as other Ericaceous plants) could potentially be used in pair-wise alignments to predict exon/intron junction sites and design intron-flanking primer pairs using a similar approach as described in Fig. 1.

Interestingly, our sequencing data of the primer C60 PCR products (i.e. trehalose-6-phospate synthase) showed that the sequences of the two exons (199 bp) were conserved across the six *Rhododendron* species (data not shown); however, the single intron between these two exons varied in length (1,485–1,492) with several bases being different across the screened species. With more information on the genomic sequences of *Rhododendron* becoming available in the future, it is likely that sequence polymorphism (within and/or between species) would be revealed and can be utilized to design other types of genetic markers.

One of the goals of our laboratory is to construct an EST-marker-based linkage map of *Rhododendron*, and to determine the location of specific cold-induced/cold-hardiness-associated genes as well as leaf-freezing tolerance QTL on this map. Toward this end, our group is developing F_2-mapping populations from a 'super' cold-hardy parent (*R. catawbiense*; leaf freezing tolerance of ~ -53°C; Lim et al. 1999) crossed with a less-hardy *R. ponticum* parent (leaf freezing tolerance of ~ -24°C; unpublished data); both species were surveyed in the current study. Encouragingly, among the EST-specific markers in the current study that generated polymorphisms particularly in these two species, some are cold-acclimation-related genes, such as trehalose-6-phospate synthase (Primer C60), early light-induced proteins (ELIPs; Primer C63), and chlorophyll a/b-binding protein CP29 (Primer C40) (Wei et al. 2005). In future studies, if EST–PCR markers (for these or other candidate genes) are found linked to cold-hardiness QTL, it is possible that the genes themselves, from which the EST–PCR markers were derived, contribute to the cold-hardiness trait. Such a knowledge could enable breeders in "marker-assisted" selection of cold-hardy rhododendron genotypes.

Acknowledgments We are grateful to Dr. Lisa J. Rowland, Dr. Anik L. Dhanaraj and Wencai Yang for helpful discussions and suggestions.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Anderson JA, Churchill GA, Autroque JE, Tanksley SD, Sorrells ME (1993) Optimising parental selection for genetic linkage maps. Genome 36:181–186

APG (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of the flowering plants: APG II. Bot J Linnean Soc 141:399–436

Bennett MD, Cox AV, Leitch IJ (1997) Angiosperm DNA c-values database. http://www.rbgkew.org.uk/cval/database1.html

Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J, Jansson S (2003) Gene expression in autumn leaves. Plant Physiol 131:430–442

Boguski MS, Schuler GD (1995) Establishing a human transcript map. Nat Genet 10:369–371

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102:296–306

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun J-H, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

De Riek J, Dendauw J, Mertens M, De Loose M, Heursel J, Van Bockstaele E (1999) Validation of criteria for the selection of AFLP markers to assess the genetic variation of a breeders' collection of evergreen azaleas. Theor Appl Genet 99:1155–1165

Deutsch M, Long M (1999) Intron–exon structures of eukaryotic model organisms. Nucleic Acids Res 27:3219–3228

Dhanaraj AL, Slovin JP, Rowland LJ (2004) Analysis of gene expression associated with cold acclimation in blueberry floral buds using expressed sequence tags. Plant Sci 166:863–872

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dubos C, Plomion C (2003) Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster Ait.*) roots. Plant Mol Biol 51:249–262

Dunemann F, Kahnau R, Stange I (1999) Analysis of complex leaf and flower characters in Rhododendron using a molecular linkage map. Theor Appl Genet 98:1146–1155

Escaravage N, Queestiau S, Pornon A, Doche B, Taberlet P (1998) Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population from AFLP markers. Mol Ecol 7:975–982

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14:1675–1690

Harry DE, Temesgen B, Neale DB (1998) Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. Theor Appl Genet 97:327–336

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH et al (1995) An STS-based map of the human genome. Science 270:1945–1954

Iqbal MJ, Paden DW, Rayburn AL (1995) Clonal stability of RAPD markers in three *Rhododendron* species. J Environ Hort 13:43–46

Komulainen P, Brown GR, Mikkonen M, Karhu A, Garcia-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *P. taeda*. Theor Appl Genet 107:667–678

Leach DG (1961) Rhododendrons of the world and how to grow them. Charles Scribner's Sons, New York

Lim C-C, Krebs SL, Arora R (1999) A 25 kD dehydrin in association with age- and genotype-dependent leaf freezing tolerance in *Rhododendron*. Theor Appl Genet 99:912–920

Liu ZJ, Corder JF (2004) DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238:1–37

Marian CO, Krebs SL, Arora R (2004) Dehydrin variability among *Rhododendron* spp: A 25 kDa dehydrin is highly conserved and associated with cold acclimation across a wide array of species. New Phytol 161:773–780

Milbourne D, Meyer R, Bradshow JE, Baird E, Bonar N, Powel JW, Morgante M, Andre C, Hanafey M, Vogel M, Tingey S, Rafalski A (1997) The utility of RFLP, RAPID, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 3:225–238

Milla MAR, Butler E, Huete AR, Wilson CF, Anderson O, Gustafson JP (2002) Expressed sequence tag-based gene expression analysis under aluminum stress in rye. Plant Physiol 130:1706–1716

Milne RI, Abbott RJ (2000) Origin and evolution of invasive naturalized material of *Rhododendron ponticum* L. in the British Isles. Mol Ecol 9:541–556

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167–174

Plomion C, Hurme P, Frigerio JM, Ridolfi M, Pot D, Pionneau C, Avila C, Gallardo F, David H, Neutelings G, Campbell M, Canovas FM, Savolainen O, Bodenes C, Kremer A (1999) Developing SSCP markers in two *Pinus* species. Mol Breed 5:21–31

Pornon A, Escaravage N, Thomas P, Taberlet P (2000) Dynamic of genotypic structure in clonal *Rhododendron ferrugineum* (Ericaceae) population. Mol Ecol 9:1099–1111

Rowland LJ, Dhanaraj AL, Polashock JJ, Arora R (2003a) Utility of blueberry-derived EST–PCR primers in related Ericaceae species. HortSci 38:1428–1432

Rowland LJ, Mehra S, Dhanaraj AL, Ogden EL, Slovin JP, Ehlenfeldt MK (2003b) Development of EST–PCR markers for DNA fingerprinting and genetic relationship studies in blueberry (*Vaccinium*, section *Cyanococcus*). J Am Soc Hort Sci 128:682–690

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8:321–329

Sakai A, Fuchigami L, Weiser CJ (1986) Cold hardiness in the genus *Rhododendron*. J Am Soc Hortic Sci 111:273–80

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93:10614–10619

Schlueter SD, Dong Q, Brendel V (2003) GeneSeqer@PlantGDB: gene structure prediction in plant genomes. Nucleic Acids Res 31:3597–3600

Senapathy P, Shapiro MB, Harris NL (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome projects. Methods Enzymol 13:252–278

Simpson CG, Leader DJ, Brown JWS (1993) Plant intron sequences. In: Croy RRD (ed) Plant molecular biology labfax. BIOS Scientific Publishers, Oxford, pp 183–251

Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 102:664–675

Thanaraj TA (1999) A clean data set of EST-confirmed splice sites from *Homo Sapiens* and standards for clean-up procedures. Nucleic Acids Res 27:2627–2637

Vainola A (2000) Polyploidization and early screening of *Rhododendron* hybrids. Euphytica 112:239–244

Wei H, Dhanaraj AL, Rowland LJ, Fu Y, Krebs SL, Arora R (2005) Comparative analysis of expressed sequence tags from cold acclimated and non-acclimated leaves of *Rhododendron catawbiense* Michx. Planta 221:406–416

Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. Mol Biol Evol 19:1407–1420

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J et al. (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535