

RESEARCH ARTICLE



Retrotransposons in *Betula nana*, and interspecific relationships in the Betuloideae, based on inter-retrotransposon amplified polymorphism (IRAP) markers

Neha Samir Roy¹ · Sung-Il Lee^{1,2} · Kabwe Nkongolo³ · Nam-Soo Kim^{1,4} 

Received: 3 December 2017 / Accepted: 9 January 2018

© The Genetics Society of Korea and Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

The Betulaceae family comprises two subfamilies, Betuloideae and Corylaceae. The subfamily Betuloideae contains two genera, *Alnus* Mill. and *Betula* L. Twenty putative long terminal repeat (LTR) retrotransposons were mined from 171 scaffolds containing 5,208,995 bp of dwarf birch (*Betula nana*) genome sequences. Five retrotransposons were finally selected after filtering the retrotransposon canonical features and nucleotide similarities between left and right LTR sequences. Of the five retroelements, three elements were found to be *Ty1/Copia* retrotransposons; identity of the other two elements could not be ascertained due to sequence undetermined 'N' bases in the sequence database. Inter-retrotransposon amplified polymorphism (IRAP) analysis, based on the LTR sequences of the mined LTR-retrotransposons, produced 179 discernible IRAP bands among the *Alnus* and *Betula* genera. Sequence analysis revealed no size homoplasy among the homologous IRAP bands. Phylogenetic and principle coordinate analysis, based on the band sharing among the taxa, showed the species in two different genera were clearly separated. The subgenera in each genus of *Alnus* and *Betula* were also distinguishable from the IRAP profiles. In the genus *Betula*, the species in subgenus *Betula* showed mixed clustering between species. This is incongruent with the phylogeographical distribution of the species.

Keywords Betuloideae · *Betula* · *Alnus* · LTR-retrotransposons · Inter-retrotransposon amplified polymorphism (IRAP)

Introduction

Betulaceae is a relatively small birch family that comprises 150 species in six genera; *Corylus*, *Ostryopsis*, *Ostrya*, *Carpinus*, *Alnus* and *Betula* (Furrow 1990; Chen et al. 1999). It is generally accepted that the Betulaceae can be sub-divided into two subfamilies, Betuloideae (*Alnus* and

Betula) and Corylaceae (*Corylus*, *Ostryopsis*, *Ostrya*, and *Carpinus*) (Furrow 1990; Forest et al. 2005; Grimm and Renner 2013). In the subfamily Betuloideae, only two genera were recognized; *Alnus* Mill. and *Betula* L. (Savard et al. 1993). Based on the palynological records in extant and fossil evidence, Crane (1989) estimated that the two genera had diverged about 65–80 million years ago (mya), a conclusion also supported by the nucleotide sequence variations in *rbcL* and rDNA internal transcribed spacer (ITS), and 18S rRNA genes (Savard et al. 1993). Interestingly, all *Alnus* species are actinorhizal, whereas all *Betula* species are nonactinorhizal (Benson and Silverster 1993; Savard et al. 1993).

The genus *Alnus* Mill. contains about 29–35 woody species that are distributed throughout the Northern Hemisphere (Chen and Li 2004; Ren et al. 2010). *Alnus* species have a symbiotic relationship with nitrogen-fixing actinomycete *Frankia*, and grow fast, making some *Alnus* species (i.e., *A. rubra*) useful in reforestation (Stein 2014). Furrow (1990) divided the genus *Alnus* into four subgenera, *Alnobetula*, *Alnus*, *Clethropsis*, and *Crematostogyne*, based on morphological and phenological characters. However,

Neha Samir Roy and Sung-Il Lee have contributed equally to this work.

✉ Neha Samir Roy
kimnamsu@kangwon.ac.kr

¹ Department of Molecular Bioscience, Kangwon National University, Chuncheon 24341, South Korea

² Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongseup 580-155, South Korea

³ Department of Biology, Laurentian University, Sudbury, ON, Canada

⁴ Institute of Bioscience and Biomedical Sciences, Kangwon National University, Chuncheon 24341, South Korea

molecular phylogenetic analysis with rDNA ITS sequence variations revealed that the *Alnus* species can be grouped into three subgenera, in which the eastern Asian subgenus *Crematostogyne* was grouped into the subgenus *Alnus* (Chen and Li 2004). Ren et al. (2010) demonstrated species identification in the genus *Alnus* using four barcoding markers (*rbcL*, *matK*, *trnH-psbA*, rDNA ITS) in which the species in the subgenera *Alnus* and *Alnobetula* were grouped together in each clade, but other species did not form a single clade in the subsequent phylogenetic analysis among 23 *Alnus* species in the world.

It is estimated that the genus *Betula* L. contains about 30–120 species, and they are distributed across boreal and temperate climate zones of the Northern Hemisphere (De Jong 1993; Chen and Li 2004; Grimm and Renner 2013; Koropachinskii 2013). Apart from their natural habitats, *Betula* species play important roles in horticulture and forestry (Ashburner and McAllister 2013). However, pollen of birch trees is a major allergen, causing hay fever in residents of Northern Europe (Ferreira et al. 1996). Selection and breeding efforts are underway to breed hypoallergenic birch trees (Schenk et al. 2008). Although several authors attribute difficulties in phylogeny and taxonomy of the genus *Betula* to extensive hybridization and introgression among the *Betula* species (Atkinson 1992; Jäviäinen et al. 2004; Thomson et al. 2015), new species have been described recently (Zeng et al. 2008, 2014; McAllister and Rushforth 2011). Nevertheless, the species in the genus *Betula* are classified into five subgenera, *Betula*, *Betulaster*, *Betulenta*, *Chamaebetula*, and *Neurobetula*, based on morphology, biochemical characters, and chromosome numbers (De Jong 1993; Jukunen-Titton et al. 1996; Keiänanen et al. 1999).

Transposable elements (TEs) are ubiquitous in all eukaryotes, and are classified into two classes, depending on the transposition mode—‘copy-and-paste’ class 1 retrotransposons and ‘cut-and-paste’ class 2 transposons (Wicker et al. 2007). The ‘copy-and-paste’ transposition mode enables high copy accumulation of the class 1 retrotransposons in the genome. As a result, hordes of long terminal repeat (LTR)-retrotransposons are found in plant genomes (Bennetzen and Kellogg 1997). Although the TEs are potentially mobile, most are stable at their current genomic locations, unless they are induced. This allows TEs to be used as a molecular genetic marker in diverse genetic applications (Kalendar and Schulman 2006; Kalendar et al. 2011; Roy et al. 2015). The unidirectional transposition nature of class 1 retrotransposons provides a great advantage in inferring pedigrees and phylogenies because the ancestral site retains the original copy (Lee et al. 2015). Inter-retrotransposon amplified polymorphism (IRAP) is a unique molecular marker system based on the retrotransposon sequences (Kalendar and Schulman 2006; Kalendar et al. 2011). It amplifies genomic

regions between LTR-retrotransposons by utilizing two LTR-sequence complementary primers in PCR reactions.

The current report contains LTR-retrotransposons mined from the complete genome sequence of *Betula nana* (Wang et al. 2013) and species relationship based on the on the IRAP markers among the species in the subfamily Betuloideae.

Materials and methods

Plant growth conditions and DNA extraction

The seeds of the *Alnus* and *Betula* species were obtained from Dr. Dale Simpson at the Natural Resources Canada, National Tree Seed Centre, Fredericton, NB, Canada. All the accessions were Canadian wild accessions, and are shown in Table 1. Plant DNA was extracted from leaves of young seedlings grown in pots containing a mixture of three parts sand, 1 part peat-moss, and 1 part perlite. Tissues from three plants were pooled in each accessions. DNA extraction was performed according to the protocol of Souza et al. (2012) without modification.

Retrotransposons mining

The LTR retrotransposons were retrieved from the whole genome sequence of *Betula nana* (dwarf birch) (GenBank: CAOK00000000.1) (Wang et al. 2013) using LTR_finder (http://tlife.fudan.edu.cn/ltr_finder/) and BLAST (Basic local alignment search tool) program at National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>). Search format was restricted by starting with TG at 5′ end and ending with CA at 3′ end. LTR length was set from 100 to 3000 bp, and the distance between LTRs was set from 4000 to 20,000 bp. The sequence similarities between left and right LTRs were set to be higher than 80%, and the target site duplication (TSD) sequences were to be 4–6 bp. The ORFs in the sequences between left and right LTRs were determined by BLASTX program, with a window size of 1000 bp.

IRAP analysis

IRAP primers were complementary to the 5′ or 3′ LTR sequences with a length between 21 and 24 nucleotides, using PrimerSeq (<http://primerseq.sourceforge.net/>). The primer sequences are shown in Table 2. For IRAP reactions, polymerase chain reaction (PCR) was performed in a 25-μL reaction mixture that contained DNA (100 ng) with 0.5 μM primer mix, 1×PCR buffer and 5U Platinum High Fidelity Taq polymerase (Invitrogen, USA). PCR amplification was conducted using initial denaturation at 94 °C

Table 1 The list of *Alnus* and *Betula* species in the analysis

Code	Species	Seed lot	Provenance	Prov.	Ctry	Lat.	Long.	Elev (m)
1	<i>A. rubra</i>	8370265	Salmon River	BC	CAN	50.20000	− 125.75000	304
2	<i>A. rubra</i>	8470585	Port Hardy	BC	CAN	50.61666	− 127.25000	37
3	<i>A. rubra</i>	9370005	Port Clements	BC	CAN	53.58333	− 132.00000	10
4	<i>A. rubra</i>	9470344	Hanna Valley	BC	CAN	53.25000	− 132.25000	100
5	<i>A. viridis</i> ssp. <i>crispa</i>	20001057	Noonan	NB	CAN	45.98333	− 66.43333	50
6	<i>A. viridis</i> ssp. <i>crispa</i>	20011113	Indian Falls Depot	NB	CAN	47.38334	− 66.33334	225
9	<i>A. viridis</i> ssp. <i>sinuata</i>	7971840	McKendrick Pass	BC	CAN	54.83333	− 126.75000	1190
10	<i>A. viridis</i> ssp. <i>sinuata</i>	8370896	Green Mountain	BC	CAN	49.05000	− 124.35000	1100
13	<i>B. allenghaniensis</i>	20001106	Warren Lake	NS	CAN	46.71667	− 60.41667	30
14	<i>B. allenghaniensis</i>	20021124	Mavis Mills	NB	CAN	49.41667	− 66.61667	179
15	<i>B. allenghaniensis</i>	20041101	Prosser Brook	NB	CAN	45.80000	− 64.91667	410
16	<i>B. allenghaniensis</i>	20061118	Williamsdale	NS	CAN	45.59282	− 63.89773	160
17	<i>B. allenghaniensis</i>	20062403	Sherbrooke	QC	CAN	45.34629	− 71.97470	269
18	<i>B. allenghaniensis</i>	20070119	Gallants	NL	CAN	48.67844	− 58.18196	123
19	<i>B. allenghaniensis</i>	20070120	Gallants	NL	CAN	48.67504	− 58.20837	150
20	<i>B. allenghaniensis</i>	20091142	Nig Pond	NS	CAN	45.94161	− 60.45198	185
21	<i>B. allenghaniensis</i>	20091158	Baddeck	NS	CAN	46.19970	− 60.84272	385
22	<i>B. allenghaniensis</i>	20091196	Richmond	PE	CAN	46.48991	− 60.00280	25
23	<i>B. allenghaniensis</i>	7030880	Pembroke	ON	CAN	45.81667	− 77.11667	
24	<i>B. cordifolia</i>	20001069	Ravine Big Gulch	NB	CAN	47.85000	− 66.45000	350
25	<i>B. cordifolia</i>	20001075	Budworm City	NB	CAN	47.63334	− 66.60000	225
26	<i>B. cordifolia</i>	20070109	Hawkes Bay	NL	CAN	50.55151	− 57.29319	65
28	<i>B. lenta</i>	9230109	St. Catharines	ON	CAN	43.16667	− 79.16666	
29	<i>B. minor</i>	20070138	Bay d'Espoir	NL	CAN	47.86456	− 55.57967	300
30	<i>B. occidentalis</i>	8570219.3	Adam's lake	BC	CAN	50.96667	− 119.68330	
32	<i>B. papyrifera</i>	20021118	North Twion Lake	NS	CAN	44.83333	− 64.66666	280
33	<i>B. papyrifera</i>	20021119	First Eel Lake	NB	CAN	44.86666	− 67.65000	160
34	<i>B. papyrifera</i>	20021120	First Eel Lake	NB	CAN	45.86666	− 67.65000	160
35	<i>B. papyrifera</i>	20041068	Riverdale	PE	CAN	46.23333	− 63.36666	75
36	<i>B. papyrifera</i>	20041083	Prosser Brook	NB	CAN	45.80000	− 64.88333	410
37	<i>B. papyrifera</i>	20061089	Semiwagan	NB	CAN	46.79744	− 65.58814	91
38	<i>B. papyrifera</i>	20062401	Duparquet	QC	CAN	48.46000	− 79.39000	
39	<i>B. papyrifera</i>	20062402	Guerin	QC	CAN	47.75000	− 79.35000	
40	<i>B. papyrifera</i>	20070101	Gallants	NL	CAN	48.67726	− 58.19256	160
42	<i>B. papyrifera</i>	20070107	Hawkes Bay	NL	CAN	50.55151	− 57.29319	65
43	<i>B. papyrifera</i>	9470306	Eaglet Lake	BC	CAN	54.10000	− 122.35000	615
44	<i>B. papyrifera</i>	9470307	Amanita Lake	BC	CAN	54.00000	− 121.78330	615
45	<i>B. papyrifera</i>	9810083.3	Valleyfield	PE	CAN	46.11660	− 62.75000	60
46	<i>B. populifolia</i>	20001078	Porten Settlement	NB	CAN	45.96667	− 67.53333	150
47	<i>B. populifolia</i>	20011151	Little Lake	NB	CAN	45.60000	− 66.80000	150
48	<i>B. populifolia</i>	20021111	Foxley River	PE	CAN	46.71667	− 64.05000	5
49	<i>B. populifolia</i>	20081110	Newmarket	NB	CAN	45.83155	− 66.97514	130
50	<i>B. populifolia</i>	9710016.3	Afton Road	PE	CAN	46.38334	− 62.93333	25

for 3 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 45 s, and 72 °C for 1 min, and a final extension at 72 °C for 5 min. Amplified PCR products were resolved on 6% denaturing polyacrylamide gel.

Sequencing of the IRAP bands

For verifying the sequence identity, the amplified bands corresponding to the desired DNA fragments were cut out from

Table 2 IRAP primer sequences

Primer name	Forward	Reverse
Retro-1	CTCGCTTGCCCTCTACCTCTC	CAGACCTTTCTATTTTCGACA
Retro-4	TGCCACTCTAATTTGATCTTCTG	CTGCCTTAAATCTCAAAGCA
Retro-5	TGACACATCAATTGTGAGATT	
Retro-6	CTATGTTCTTCGAGTTGTTAGC	CTGTGAATACTACCTTCACA
Retro-7	TGTTAGAATAATATATTAGGG	TTCACGTAACCCATTATAACA

the gel using a sterile scalpel, dissolved in 5 µL of distilled water and re-amplified with the respective IRAP primers used and same PCR cycle as the above. Size of the products was confirmed by agarose gel electrophoresis. The amplified fragments were purified using ‘GenElute™ gel extraction kit’ (SIGMA, St. Louis, USA), and successively sequenced using BigDye Terminator Version 3.1 Kit (Applied Bioscience) on an ABI 3730 sequencer (Invitrogen).

Phylogenetic analysis

The amplified IRAP bands were recorded as binary codes: 1 for presence and 0 for absence. The faint or orphan bands were excluded in recordings to avoid ambiguity. Phylogenetic analysis and principal coordinate analysis (PCoA) were done using Genalex version 6.5 software package (Peakall and Smouse 2012). Cluster analysis and Neighbor joining tree was constructed using DARwin 6.0 (Perrier and Jacquemoud-Collet 2006).

Results

Retrotransposons

The whole genome sequence of *Betula nana* (GenBank: CAOK00000000.1) contains 564,011,153 bp in 551,915 scaffolds, from which 171 scaffolds containing 5,208,995 bp were randomly selected for mining LTR retrotransposons. The lengths of the 171 scaffolds ranged from 20 to 150 kb in which many sequence undetermined ‘N’ bases were present. The LTR finders selected 20 putative LTR retrotransposons and filtered them to seven LTR-retrotransposons, of which

five retrotransposons were selected on the basis of a greater than 80% nucleotide similarity between left and right LTRs (Table 3; Fig. 1).

While the five retrotransposons had LTRs at both ends which were flanked with 5–6 bp TSDs, sequence characterization in the internal regions between LTRs was incomplete due to the presence of ambiguous ‘N’ bases. The seven identified retrotransposons were labelled as *BnRetro* (*Betula nana* retrotransposons) elements. Of the five elements selected, *BnRetro-5*, *BnRetro-6*, and *BnRetro-7*, were revealed to be *Ty1/Copia* retrotransposons by the gene order of GAG-INT-RT (Voytas and Boeke 2002). However, the ORFs in these three elements were incomplete due to sequence decay. Ambiguous ‘N’ bases in the other two elements did not permit whether they were *Ty1/Copia* or *Ty3/Gypsy* elements. The *BnRetro-1* carried 790 bp of serine/threonine kinases, interleukin-1 receptor associated kinases (STKc_IRAK) gene while *BnRetro-2* carried 825 bp of Cu-oxidase gene between LTR sequences, respectively. However, many ‘N’ bases within the genes prevented identification of complete gene sequences.

IRAP marker analysis

Nine combinations of IRAP primers produced 179 discernible amplicons. *Alnus* species had a total number of 114 bands and *Betula* species yielded 145 discernible bands, respectively. Eighteen IRAP loci were shared among all accessions in all species in both genera in Betuloideae. While eight IRAP loci were present in all accessions of *Alnus* genus, nine IRAP loci were present in all accession of *Betula* genus. Number of unique bands was 17 in *A. rubra*, 9 in *A. viridis*, and 6 in *B. alleghaniensis*. However, the

Table 3 Molecular features of the LTR-retrotransposons from *Betula nana*

	Total length (bp)	5’LTR (bp)	3’LTR (bp)	LTR similarity (%)	Target Site duplication	Gag (bp)	Int (bp)	RT (bp)	RH (bp)
<i>BnRetro1</i>	13,994	660	651	84.26	ACACCA	–	–	–	–
<i>BnRetro4</i>	8174	246	246	93.50	GTCCA	–	–	–	–
<i>BnRetro5</i>	10,064	143	147	90.98	GACTA	351	348	825	692
<i>BnRetro6</i>	7927	824	804	91.28	TTGGT	756	360	640	–
<i>BnRetro7</i>	5779	140	140	97.86	GTCTT	276	754	834	744

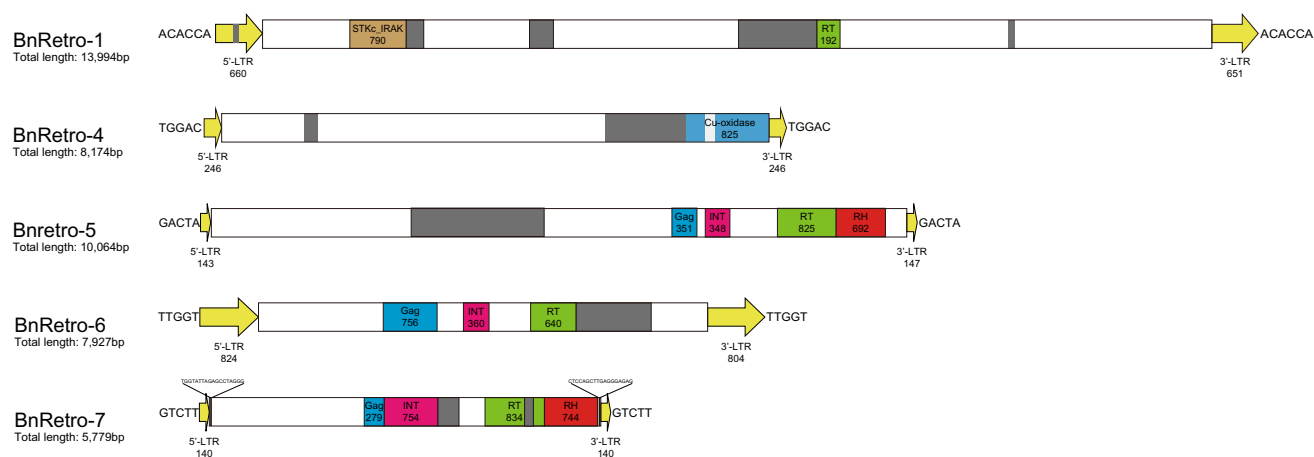


Fig. 1 Structures of the LTR-retrotransposons from *Betula nana*. The numbers in the LTR-retrotransposons are numbers of nucleotides

Table 4 Total band patterns of *Alnus* and *Betula* species

Population	<i>A. rub</i> (4)	<i>A. vir</i> (4)	<i>B. all</i> (11)	<i>B. cor</i> (3)	<i>B. len</i> (1)	<i>B. min</i> (1)	<i>B. occ</i> (1)	<i>B. pap</i> (11)	<i>B. pop</i> (5)
No. bands	90	75	126	123	123	123	123	124	102
No. unique bands	17	9	6	0	0	0	0	0	0
Mean He	—	—	0.155	—	—	—	—	0.170	—
Number of polymorphic loci	—	—	93	—	—	—	—	100	—
Percentage polymorphism	—	—	51.96%	—	—	—	—	55.87%	—
Proportion of genetic diversity among population (Gst)	—	—	0.29	—	—	—	—	0.128	—
Gene flow (Nm*)	—	—	1.19	—	—	—	—	3.98	—

other species showed no private bands (Table 4). Nucleotide sequencing analysis was carried out to check whether the IRAP bands at the same position in the PAGE gel represent the same sequences. We sequenced 24 bands that were derived from five IRAP loci in which the same size IRAP bands were identical in nucleotide sequences (Fig. 2). Thus, they were homologous rather than homoplasy bands, which is important for inferring species relationship and phylogenetic relations among the analyzed *Alnus* and *Betula* species.

Phylogenetic and principal coordinate analyses

Phylogenetic dendrogram revealed that the two genera in the Beuloideae formed separate clusters (Fig. 3). In the cluster of *Alnus* species, both *A. rubra* and *A. viridis* accessions formed their own subcluster. The *A. viridis* accessions were two subspecies, *A. viridis* ssp. *sinuata* and *A. viridis* ssp. *crispa*, which were separated from each other. There are two clusters in the *Betula* species; one with the subgenus *Betula* including *B. papyrifera*, *B. cordifolia*, *B. occidentalis*, and *B. populifolia*, and another with subgenus *Betulenta* including *B. alleghaniensis* and *B. lenta*. While the four *Betula* species in the subgenus *Betula* were mixed into subcluster, all

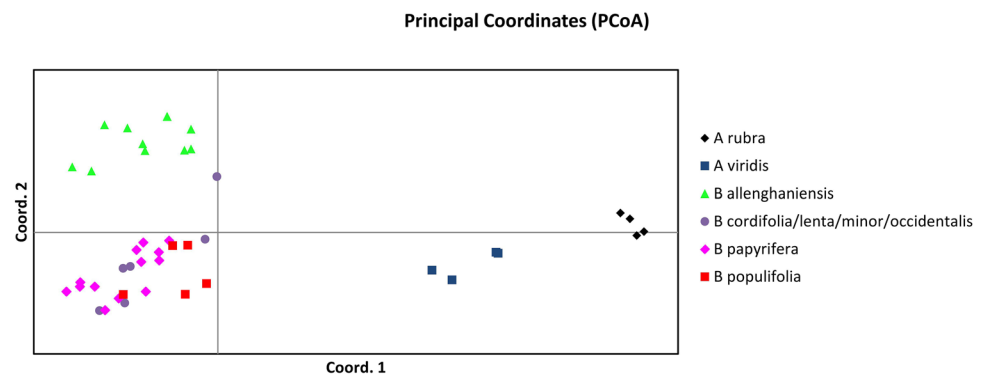
accessions of *B. alleghaniensis* in the subgenus *Betulenta* formed a subcluster together.

The PCoA analysis corroborated the phylogenetic analysis (Fig. 4). The group I of *Alnus* species, and group II of *Betula* species in phylogenetic analysis were separated in coordinate I; coordinate 2 separated the two subgroups of the genus *Betula*. In combination, coordinates 1 and 2 enable a discrimination of all three clusters in the phylogenetic dendrogram.

Discussion

The subfamily Betuloideae contains many important species, both as timber trees in the wild and as attractive ornamentals in the landscaped back yards of homes in the Northern hemisphere (Ashburner and McAllister 2013; Koropachinskii 2013). However, genetic studies on them have lagged behind that of other plant species. This study identifies LTR-retrotransposons from *B. nana* (dwarf birch), and reports on interspecific relationship among subgenera species of the subfamily Betuloideae, using LTR-retrotransposon derived molecular markers.

Fig. 4 Principle coordinate analysis based on the IRAP profiles



LTR-retrotransposons are ubiquitous and major components in plant genomes (Bennetzen and Kellogg 1997; Lee and Kim 2014). The estimated genome size of *B. nana* was ~450 Mbp (Wang et al. 2013), which is only 8% of the average genome size (5.8 Gbp) of angiosperms (Bennett and Leitch 2011). The number of LTR-retrotransposons is known to be positively correlated with genome sizes in plant species with large genomes (Bennetzen and Kellogg 1997; Lee and Kim 2014). Apart from the current study, there are no reports on TE in the Betuloideae. The 20 putative LTR-retrotransposons in 5.2 Mbp of genome sequences of *B. nana* are comparatively lower than the number of LTR-retrotransposons in other reports in plant species. For example, Lee et al. have isolated 12 *Ty1*-Copia elements out of 148 Kb that were derived from fosmid library of *Lilium longiflorum* (Lee et al. 2013). The genome size of *L. longiflorum* is 35 Gbp (Lee and Kim 2014), ~80 times larger than the genome of *B. nana*. Even if we account for the genome size differences between two species, twenty LTR-retrotransposons in 5.2 Mbp in *B. nana* is still comparatively low so that the small genomes of *Betula* species may harbor low amount of LTR-retrotransposons. The five isolated *BnRetro* elements might be retrotransposon fossils because they do not carry canonical retrotransposon features (Voytas and Boeke 2002). The STKc-IRAK domain in *BnRetro-1* and Cu-oxidase gene in *BnRetro-2* are interesting. STKc-IRAK domain families are receptor-like protein kinases in plants that are involved in plant defense and development (Stahl et al. 2013). Copper oxidase is also involved in a variety of phenomena in plant physiology (Coleman 1974). Retroviruses can transfer host genes by transduction. LTR-retrotransposons can carry host genes between LTRs and transpose the gene(s) in the genome (Cordaux and Batzer 2009). Although it is not clear that the genes encoding STKc-IRAK domain and cu-oxidase in *BnRetro-1* and *BnRetro-2* are trans duplicated genes, analysis of the function of these genes may render some interesting inferences on the adaptation of *B. nana*.

Transposable elements have been proven to be useful molecular markers in plant breeding, genetics, and ecology (reviewed in Roy et al. 2015, references therein). IRAP

amplifies the genomic region between retrotransposons by utilizing primers that are complementary to the 3' end of a LTR sequence (Kalendar et al. 1999). In fingerprinting analysis of annual and perennial *Helianthus* species, Vukich et al. (2009) demonstrated high genetic diversity among the *Helianthus* accessions and isolated a species-specific insertion of a *Copia*-like element. Recently, Lee et al. (2016) showed that the molecular genetic diversities measured by IRAP were higher than those values measured by simple sequence repeat polymorphism (SSRP) in natural populations of diploid and triploid accessions of *Lilium lancifolium*. We analyzed nine species in the subfamily Betuloideae using IRAP profiles; 10% of the IRAP bands were present in all the accessions analyzed in both genera. Molecular and fossil-based divergence analyses estimated that the *Alnus* and *Betula* genera diverged about 65–80 mya (Crane 1989; Forest et al. 2005). The conservation of about 10% of the sequences of IRAP bands is striking. Nucleotide sequencing analysis of the IRAP bands that migrated to similar positions in the polyacrylamide gel revealed that the sequences were almost identical among species from both genera (Fig. 2). This would be indicative of homology, and not homoplasy. In our analysis the homologous bands between taxa were different with a few single nucleotide substitutions or one or a few bases deletion/additions. Since they do not encode proteins, it is not clear why the noncoding sequences have been conserved for such a long time.

As expected, the two genera in the Betuloideae formed separated clusters in phylogenetic analysis and principal coordinates analysis (PCoA) (Figs. 3, 4). In the *Alnus* cluster, two *Alnus* species, *A. rubra* (red alder) and *A. viridis* (green alder), were included in the current study. While *A. rubra* was classified into the subgenus *Alnus*, *A. viridis* was included into the subspecies *Alnobetula*, based on the phenological characters (Furrow 1990). Our results of the clustering patterns are corroborated by the results of nuclear rDNA ITS sequences (Chen and Li 2004) and nuclear and chloroplast barcoding genes (Ren et al. 2010). *A. rubra* grows from southeast Alaska to central coast California, nearly along the coast within 200 km (Biota of

North America Program 2013, <http://bonap.net/MapGallery/County/Alnus%20rubra.png>). Although it is not on the list of endangered species, genetic diversity was low to implicate problems with long term sustainability of these populations (Xie 2008). Its habitat distributes in low altitude, and it is replaced by the smaller Sitka alder (*A. viridis* ssp. *sinuata*) in marginal habitat in high mountains. In our study, the four accessions of *A. rubra* and two accessions of *A. viridis* ssp. *sinuata* were collected in their natural habitats in British Columbia, Canada. The collection sites of the *A. rubra* were under 300 m while those of *A. viridis* ssp. *sinuata* were above 1000 m in altitude (Table 1). *A. viridis* is distributed widely in the cooler regions of the Northern Hemisphere (Chen and Li 2004). Three *A. rubra* subspecies grow in North America—*A. viridis* ssp. *sinuata*, *A. rubra* ssp. *crispa*, and *A. viridis* ssp. *fruticosa* (http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=233500041). Only two subspecies, *A. viridis* ssp. *sinuata* and *A. rubra* ssp. *crispa*, were included in our study, and they formed separate subclusters from each other. *A. viridis* ssp. *sinuata* grows in Western North America while *A. rubra* ssp. *crispa* grows in Northeastern North America.

Difficulties in phylogeny and taxonomy of *Betula* species have been attributed to extensive hybridization and introgression among the *Betula* species (Atkinson 1992; Jäviäinen et al. 2004; Thomson et al. 2015). The species used in our analysis were mainly from two subgenera—*Betula* which contains bark birches (including *B. papyrifera*, *B. cordifolia*, *B. populifolia*, *B. occidentalis*), and *Betulenta* which contains species with generally dark, non-exfoliating bark (including *B. alleghaniensis* and *B. lenta*) (DeJong 1993). The species in the two subgenera were clearly separated in the phylogenetic analysis based on the IRAP profiles, which was congruent with the amplified fragment length polymorphism (AFLP) marker analysis (Schenk et al. 2008). However, our results differed in some species with the phylogenetic tree based on the chloroplast *matK* sequences in which the *B. papyrifera* and *B. populifolia* were separated in different groups (Jäviäinen et al. 2004). This might be due to cytoplasmic introgression, as the authors have speculated. In our analysis, the *Betula* species were mostly from eastern Canada except for the one accession of *B. occidentalis* and two accessions of *B. papyrifera* that were from British Columbia, Canada. While the *B. occidentalis* distributes from the west coast to Northern Ontario, the distribution of *B. papyrifera* is much wider, ranging from the west coast of Canada to the east coast, including Newfoundland (Ashburner and McAllister 2013). The two accessions of *B. papyrifera* from British Columbia formed a subcluster with an accession of *B. papyrifera* from Newfoundland and an accession of *B. cordifolia* from New Brunswick, implying that geographic distance might not be related with the genetic

relatedness among *B. papyrifera*. The highly mixed sub-clustering among the different *Betula* subgenus species can be backed up by other studies of wide hybridization and introgression between species (Atkinson 1992; Jäviäinen et al. 2004; Thomson et al. 2015). The two species, *B. lenta* and *B. alleghaniensis*, in the subgenus *Betulenta*, distribute mainly in the Northeastern regions in Canada (Ashburner and McAllister 2013). In a phylogeographical structure analysis of the North American birches, Thomson et al. (2015) demonstrated that there were three phylogeographical groups for *B. papyrifera* and four groups for *B. alleghaniensis*. Moreover, the species shared regional haplotypes in chloroplast microsatellites and *psbA-trnH* intergenic spacer sequences, asserting genetic introgression between the distantly related two species growing in sympatric regions. This is in direct contrast to our study with IRAP profiles. Phylogeographically distant accessions were placed closely in the subclusters of the subgenus *Betula* in our phylogenetic analysis. Nine of the 145 IRAP bands were shared by all accessions in the genus *Betula*, but the band sharing may not be a result of introgression. Also, in both phylogenetic and PCoA analysis, the *B. alleghaniensis* accessions were distinct from the *B. papyrifera* accessions even among the sympatric regional accessions. Sequence analysis verified that the shared bands are homologous bands. IRAP producing retrotransposons might antedate species divergence among the genus *Betula*. The next step in our analysis will be to use more accessions of each species for genetic diversity. This may provide a clearer insight into the genetic make-up among the species in the subfamily Betuloideae.

Acknowledgements This work was carried out with a sabbatical grant to NSK from Kangwon National University.

Compliance with ethical standards

Conflict of interest Neha Samir Roy, Sung-Il Lee, Kabwe Nkongolo and Nam-Soo Kim all declare that they do not have a conflict of interest.

Research involving with human and animal participants This research did not utilize any human or animal subjects. The plant materials used are not on the list of endangered species.

References

- Ashburner K, McAllister HA (2013) The genus *Betula*: a taxonomic revision of birches. Kew Publishing, London
- Atkinson MD (1992) *Betula pendula* Roth (*B. verrucosa* Ehrh.) and *B. pubescens* Ehrh. J Ecol 80:837–840
- Bennett MD, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann Bot 107:467–590

- Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genome obesity? *Plant Cell* 9:1509–1514
- Benson DR, Silverster WB (1993) Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol Rev* 57:293–319
- Chen Z, Li J (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. *Int J Plant Sci* 165:325–335
- Chen ZD, Manchester SR, Sun HY (1999) Phylogeny and evolution of the Betulaceae as inferred from DNA sequence, morphology, and paleobotany. *Am J Bot* 86(8):1168–1181
- Coleman JE (1974) Structure and mechanisms of copper oxidases. *Adv Chem* 136:267–304. <https://doi.org/10.1021/ba-1974-0136.ch011>
- Cordaux R, Batzer MA (2009) The impact of retrotransposons on human evolution. *Nat Rev Genet* 10:691–703
- Crane PR (1989) Early fossil history and evolution of Betulaceae. In: Crane PR, Baltimore S (eds) *Evolution, systematics, and fossil history of Hamamelidae*, vol 2. Clarendon, Oxford, pp 85–116
- De Jong PC (1993) An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. In: *Proceedings of the IDS Betula symposium*, October 1992, p 204. International Dendrology Society, Richmond
- Ferreira F, Hirtenlehner K, Jilek A, Godnik-Cvar J, Breiteneder H, Grimm R, Horffmann-Sommergruber K, Kraft D, Breitenbach M, Rheinberger H-J, Ebner C (1996) Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med* 183:599–609
- Forest F, Savolainen V, Chase MW, Lupia R, Bruneau A, Crane PR (2005) Teasing apart from molecular—versus fossil-based error estimates when dating phylogenetic trees: a case study in the birch family (Betulaceae). *Syst Bot* 30:118–133
- Furlow JJ (1990) The genera *Betulaceae* in the southeastern United States. *J Arnold Arb* 71:1–61
- Grimm GW, Renner SS (2013) Harvesting *Betulaceae* sequences from GenBank to generate a new chronogram for the family. *Bot J Linn Soc* 172:465–477
- Jäviäinen P, Palmé A, Morales L, Lännepää M, Keinänen M, Sopanen T, Lascoux M (2004) Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear *ADH* and chloroplast *MatK* sequences. *Am J Bot* 91:1834–1845
- Julkunen-Titto R, Rousi M, Bryant J, Sorsa S, Keinänen M, Sikanen H (1996) Chemical diversity of Betulaceae species: comparison of phenolics and terpenoids in northern birch stems. *Trees* 11:16–22
- Kalendar R, Schulman AH (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat Protoc* 1:2478–2484
- Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor Appl Genet* 98:704–711
- Kalendar R, Flavell AJ, Ellis T, Sjakste T, Moisy C, Schuman A (2011) Analysis of plant diversity with retrotransposons-based molecular markers. *Heredity* 106:520–530
- Keinänen M, Julkunen-Titto R, Rousi M, Tahvanainen J (1999) Taxonomic implications of phenolic variation in leaves of birch (*Betula* L.) species. *Biochem Syst Ecol* 27:243–254
- Koropachinskii IY (2013) Natural hybridization and taxonomy of birches in North Asia. *Contemp Probl Ecol* 6:459–480
- Lee SI, Kim NS (2014) Transposable elements and genome size variations in plants. *Genom Inform* 12:87–97
- Lee SI, Park KC, Son JH, Hwang YJ, Lim KB, Song YS, Kim JH, Kim NS (2013) Isolation and characterization of novel Ty1-copia-like retrotransposons from lily. *Genome* 56:495–503
- Lee SI, Kim JH, Park KC, Kim NS (2015) LTR-retrotransposons and inter-retrotransposons amplified polymorphism (IRAP) analysis in *Lilium* species. *Genetica* 143:343–352
- Lee SI, Nguyen XT, Kim JH, Kim NS (2016) Genetic diversity and structure analyses on the natural populations of diploids and triploids of tiger lily, *Lilium lancifolium* Thunb., from Korea, China, and Japan. *Genes Genom* 38:467–477
- McAllister HA, Rushforth K (2011) *Betula ashburneri*. *Curtis's Bot Magazine* 28:111–118. <https://doi.org/10.1111/j.1467-8748.2011.01735.x>
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier X, Jacquemoud-Collet J (2006) DARwin software <http://darwin.cirad.fr/>
- Ren BQ, Xiang XG, Chen ZD (2010) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. *Mol Ecol Res* 10:594–605
- Roy NS, Choi JI, Lee SI, Kim NS (2015) Marker utility of transposable elements for plant genetics, breeding, and ecology: a review. *Genes Genom* 37:141–151
- Savard L, Michaud M, Bousquet J (1993) Genetic diversity and phylogenetic relationships between birches and alders using ITS, 18S rRNA, and *rbcL* gene sequences. *Mol Phylogenet Evol* 2(2):112–118
- Schenk MF, Thienpont CN, Koopman WJM, Gilissen LJWJ., Smulders MJM (2008) Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genet Genom* 4:911–924
- Souza HA, Muller LA, Brandao RL, Lovato MB (2012) Isolation of high quality and polysaccharide-free DNA from leaves of *Dimorphandra mollis* (Leguminosae), a tree from the Brazilian Cerrado. *Genet Mol Res* 11(1):756–764
- Stahl Y, Grabowski S, Bleckmann A, Kühnemuth R, Weidtkamp-Peters S, Pinto KG, Kirschner GK, Schmid JB et al (2013) Moderation of *Arabidopsis* root stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 receptor kinase complexes. *Curr Biol* 23:362–371
- Stein WI (2014) Reforestation systems compared on coastal clearcuts: 10 year results. *TreeSearch. Res Pap. PNW-RP-601*. USDA, Forest Service, Pacific, Portland
- Thomson AM, Dick CW, Dayanandan S (2015) A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by introgression than by shared biogeographical history. *J Biogeogr* 42:339–350
- Voytas DF, Boeke JD (2002) *Ty1* and *Ty5* of *Saccharomyces cerevisiae*. In *Mobile DNA II*. Craig NL, Craigie R, Gellert M, Lambowitz AM (eds), ASM Press, Washington, pp 631–662
- Vukich M, Schulman AH, Giordani T, Natali I, Kalendar R, Cavallini (2009) Genetic variability in sunflower (*Helianthus annuus* L.) and in the *Helianthus* genus as assessed by retrotransposon based molecular markers. *Theor Appl Genet* 11:1027–1038
- Wang N, Thomson M, Bodles WJ, Crawford RM, Hunt HV, Featherstone AW, Pellicer J, Buggs RJ (2013) Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol Ecol* 22(11):3098–3111
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B et al (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982
- Xie CY (2008) Ten-year results from red alder (*Alnus rubra* Bong.) provenance-progeny testing and their implications for genetic improvement. *New For* 36:273–284
- Zeng J, Li JH, Chen ZD (2008) A new species of *Betula* section *Betulaster* (Betulaceae) from China. *Bot J* 156:523–528
- Zeng J, Ren BQ, Zhu JY, Chen ZD (2014) *Betula hainanensis* (Betulaster, Betulaceae), a new species from Hainan Island, China. *Ann Bot Fennici* 51:399–402