

Supplementary Information for:

An intact, but dormant LTR retrotransposon defines a moderately-sized family in white spruce (*Picea glauca*)

Britta Hamberger, Macaire Man Saint Yuen, Emmanuel Buschiazzi, Claire Cullis, Agnes Yuen, Carol Ritland, Jörg Bohlmann and Björn Hamberger

Corresponding author Björn Hamberger

Email: hamberge@msu.edu

This PDF file includes:

Tables S1 to S3

Figs. S1

Other supplementary materials for this manuscript include the following:

Datasets S1

Supplemental Table S1: Summary of sequencing and assembly of plasmid libraries for PGB10 and PGB17.

BAC	Plasmids sequenced (bidirectional)	Contamination ¹	Reads	Contigs ²	Coverage ³	Avg. plasmid insert	Size/estimated size
PGB10	1152	15.1%	2304	11	9x	1104bp	174kb/170kb
PGB17	1152	5%	2304	1	13x	n.d.	102kb/100kb

¹Contaminating sequences were identified using a BLAST search against the non-redundant nucleotide database at NCBI (blast.ncbi.nlm.nih.gov)

²The number of successful reads is given. Reads with masked repeats were assembled by PHRAP into contigs.

³The coverage is calculated from the number of non-redundant nucleotides and the physical size of both BAC clones.

n.d. not determined.

Supplemental Table S2: Features of LTR-REs identified in white spruce genomic BAC sequences.

Annotation	LTR integrity [%] ¹	5'TG, 3'CA	TSR	PBS	PPT	Position [bp]	Length [bp]	Orientation	LTR-flanked region [bp]	RE type
Picnicker1	99.1	yes	yes	yes	yes	62,542	16,083	+	4,755	Ty3-gypsy
Picnicker2 ²	98.4	yes	no	yes	yes	91,487	11,088	-	4,758	Ty3-gypsy
PGB17-3	96.7	yes	yes	yes	yes	80,497	6,686	+	6,259	Ty1-copia
PGB10-1 ³	93.4	yes	yes	yes	yes	17,715	18,086	+	6,909	Ty3-gypsy
PGB10-2	94.4	yes	yes	yes	no	12,894	5,741	+	2,738	Ty3-gypsy
PGB10-3	88.9	yes	yes	no	yes	693	3,265	+	1,955	Ty3-gypsy
PGB09-1 ⁴	100	yes	yes	yes	yes	13,127	8,584	+	3,345	unclassified
PGB09-2	97	yes	no	yes	yes	29,566	12,820	-	11,628	Ty1-copia
PGB09-3 ^{4,5}	100	yes	yes	yes	yes	30,812	6,445	-	5,158	Ty1-copia
PGB09-4 ⁶	96.7	yes	yes	yes	yes	71,257	9,666	-	4,689	Ty3-gypsy
PGB09-5	90.1	yes	no	yes	yes	80,916	16,256	-	15,613	Ty1-copia
PGB08-1	89.7	yes	no	yes	yes	54,701	6,165	+	5,374	Ty3-gypsy
PGB08-2	89.8	yes	yes	yes	yes	266	8,428	-	6,546	unclassified
PGB08-3	95.1	yes	yes	yes	no	27,014	16,645	-	7,089	Ty3-gypsy
PGB08-4	95.4	yes	yes	yes	yes	86,424	6,018	-	3,134	unclassified
PGB08-5	95.7	yes	yes	yes	yes	101,464	6,076	-	5,237	Ty3-gypsy
PGB08-6	97	yes	no	yes	yes	112,290	14,542	+	12,589	Ty3-gypsy
PGB04-1 ³	95.7	yes	no	yes	yes	39,525	8,165	+	7,514	Ty1-copia

PGB04-2 ³	92.2	yes	yes	yes	yes	68,804	10,301	+	5,394	Ty1-copia
PGB02-1	85.7	yes	no	yes	yes	36,467	5,381	+	4,632	Ty1-copia
PGB02-2 ³	89	yes	no	yes	yes	65,563	11,484	-	10,726	Ty1-copia
PGB02-3 ³	91	yes	no	yes	yes	137,987	11,802	+	10,961	Ty3-gypsy
PGB02-4 ³	89.9	yes	yes	yes	yes	141,121	5,429	-	4,971	Ty1-copia

TSR, target side repeat; PBS, primer binding site; PPT, poly-purine tract.

¹ Determined by BLAST comparison of LTRs and given in percent identical nucleotides.

² 3' end of BAC including TSR sequence is truncated, extrapolated full length size 15,765 bp.

³ Internal duplicated region. Typically caused by nested LTR retroelements integrating into an existing element and with a characteristic higher degree of conservation for the secondary event. These loci do not follow this model and may be result of a yet to be elucidated mechanism.

⁴ LTRs in both elements flank highly degraded regions without detectable ORFs indicating that both are ancient or that the observed structure may be result of interlocus gene conversion, a recently described mechanism reducing divergence of LTR elements (Kijima and Innan 2010).

⁵ Nested transposon PGB09-3 located within older PGB09-2.

⁶ Region flanked by LTR in PGB09-5 has insertion of Ty1-copia type ORF without LTR. Likely nested RE.

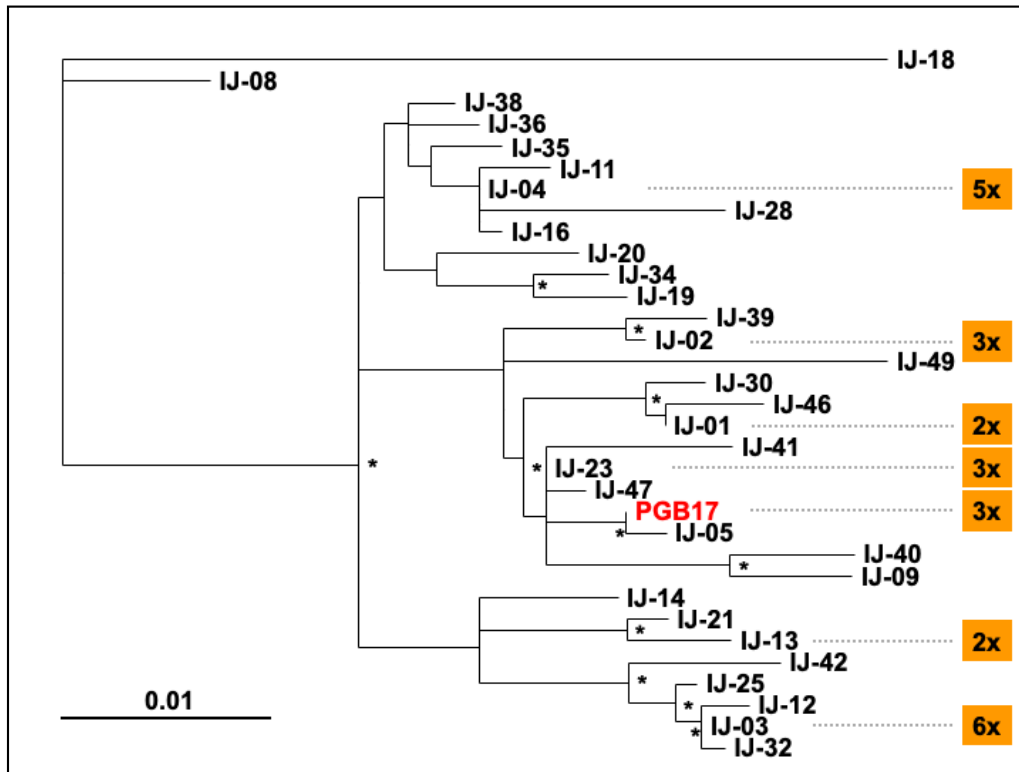
PGB10-1 on contig19 (56,482 bp), PGB10-2 on contig18 (23,287 bp), PGB10-3 on contig16 (15,887 bp) (Figure 1).

PGB08-1 on contig1 (61,169 bp), PGB08-2 on contig2 (8,726 bp), PGB08-3 to PGB08-6 on contig3 (128,238 bp)

Genbank accession PGB02, FJ609174.2; PGB04, FJ60975.1; PGB08, GU059904.1; PGB09, GU059905.1.

Table S3: The geographical origin for white spruce trees (FN, Fort Nelson; PG, Prince George; EK, East Kootenays).

Sample	Location
1 FN6126	+59° 32' 60.00", -126° 28' 58.80"
2 FN6129	+59° 28' 1.20", -126° 19' 1.20"
3 FN6118	+59° 23' 60.00", -125° 47' 60.00"
4 FN6115	+59° 16' 58.80", -125° 25' 58.80"
5 FN6095	+59° 20' 60.00", -124° 52' 58.80"
6 FN6157	+58° 55' 58.80", -120° 34' 58.80"
7 PG29	+53° 31' 12.00", -122° 8' 60.00"
8 PG165	+53° 8' 60.00", -121° 17' 60.00"
9 PG105	+53° 14' 24.00", -122° 2' 24.00"
10 PG160	+53° 0' 36.00", -121° 59' 60.00"
11 PG159	+52° 59' 60.00", -121° 35' 24.00"
12 PG141	+52° 28' 12.00", -122° 1' 12.00"
13 EK25	+50° 5' 24.00", -115° 16' 12.00"
14 EK28	+50° 10' 12.00", -115° 24' 36.00"
15 EK29	+50° 22' 12.00", -115° 19' 48.00"
16 EK49	+50° 20' 24.00", -116° 11' 60.00"
17 EK76	+50° 22' 12.00", -116° 8' 24.00"
18 EK101	+49° 11' 60.00", -115° 31' 48.00"



Supplemental Fig 1. Phylogenetic tree of cloned and sequenced PCR amplicons targeting the Picnicker LTR region on genomic DNA.

1,106 bp were amplified in three independent PCR reactions using a primer combination specific for the left LTR using white spruce genomic DNA of cultivar PG29.

Sequence identity ranges between 80-99%. 60 clones were sequenced bi-directional. 10 reactions had no similarity to the target region. Of the 50 remaining products, only 26 are unique singletons, while 24 cluster with identical sequences. Close to 50% of the reads are redundant. This beginning saturation of genomic clones indicates that the copy number of the Picnicker family in the genome is moderate. In comparison, the inactive pine IFG LTR-RE has an estimated number of 10,000 copies. The bar indicates 0.01 nucleotide changes. Asterisks indicate 80% and higher bootstrap values.