



OPEN

Dissection of figured wood trait in curly birch (*Betula pendula* Roth var. *carelica* (Mercklin) Hämet-Ahti) using high-throughput genotyping

Rim Gubaev¹, Dmitry Karzhaev^{1,2}, Elizaveta Grigoreva^{1,5}, Kirill Lytkin¹, Elizaveta Safronycheva^{1,2}, Vladimir Volkov^{1,2}, Veronika Nesterchuk², Lidiia Vetchinnikova^{1,3}, Anatoly Zhigunov¹ & Elena Potokina^{1,4}✉

Curly (Karelian) birch is a special variety of *Betula pendula* Roth distributed in the northwestern part of Europe. Karelian birch is well-known for its valuable figured curly wood also known as “wooden marble”. The genetic basis underlying curly wood formation has been debated since last century, however, there was no data about loci responsible for the curly wood trait. In the present study, we analyzed two full-sibs populations derived from experimental crosses of curly birches and segregating for the trait. RADseq genotyping was applied to reveal how many loci are involved in ‘curliness’ formation and to search for genetic variants associated with this trait. One single interval on chromosome 10 was detected containing possible candidate genes. InDel marker BpCW1 was suggested for the first time for marker-assisted selection of trees with curly wood at their earliest stages of development.

Keywords Karelian birch, Curly wood, RADseq, Association mapping, SNPs, Molecular markers, Marker-assisted selection, BpCW1

Curly (Karelian) birch (*Betula pendula* Roth var. *carelica* (Mercklin) Hämet-Ahti) is a special variety of silver birch, which forms small populations scattered within the north-western part of *B. pendula* distribution: Finland, Russia (Karelia), Baltic countries, Belarus, Poland, southern Sweden and Norway^{1,2}. The wood fibers of the Karelian birch are not directed strictly vertically, but at different angles, which leads to the formation of wood “curliness” underlying the special ‘curly’ phenotype. The figured ‘curly’ wood of the Karelian birch is used for manufacturing highly decorative furniture and architectural panels, and it is sometimes called ‘wooden marble’ from Finland³. Because of its exotic nature, figured wood is paid by weight (3–5 € kg⁻¹), rather than by volume like the wood of other trees⁴. Due to the high commercial value, interest in the propagation of curly birch on plantations has increased significantly since the 1980s. In Russia, the total area of forest plantations of Karelian birch by 1986 reached 5500 ha⁵. Unfortunately, in the 90 s, due to illegal logging, about 1.5 thousand curly birches were cut down on the territory of Karelia. In Finland, though, 6500 ha of curly birch stands have been established, and they are currently reaching their rotation age (37 years)³.

The major issue of curly birch propagation on plantations is that the first signs of ‘curliness’ among the trees appear at the age of 8–15 years^{6,7}. Until that time, it is necessary to maintain a large plantation of trees with unpredictable yield, since any stands of trees artificially planted using Karelian birch seeds always have an admixture of ordinary (non-curly) birches. Proportions of ‘curly’ trees among progenies of open-pollinated Karelian birch vary from 2–3%⁸ to 25% (rarely 50%)^{9,10}. 60–70% of curly-wooded offspring appear from controlled crosses of two parents with a ‘curly’ phenotype¹¹.

Non-curly birches are usually removed from the stand at the age of about 10–13 years³ to provide genuine curly trees with sufficient space and light. For this cleaning, a non-destructive visual assessment of curliness is commonly used¹². However, even at this age, tinning often encounters a problem, since detecting curliness by

¹Institute of Forest and Natural Resources Management, Saint Petersburg State Forest Technical University, St. Petersburg, Russia. ²Research Laboratory, Saint Petersburg Forest Research Institute, St. Petersburg, Russia. ³Forest Research Institute of Karelian Research Centre Russian Academy of Sciences, Petrozavodsk, Russia. ⁴Skolkovo Institute of Science and Technology, Moscow, Russia. ⁵Present address: Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria. ✉email: e.potokina@skoltech.ru

indirect morphological features of trees is rather difficult. Thus, understanding the genetic mechanism underlying the figured wood phenomenon of Karelian birch is becoming a truly challenging task, yet it has the potential to provide foresters with trait-specific molecular markers to help recognize curly birch genotypes at their earliest stages of development.

Throughout the long history of studying Karelian birch, various hypotheses have been suggested regarding the nature of the factors that determine curly wood^{12–14}. The heritability of the trait, though, is acknowledged by most researchers. Recently, Kärkkäinen et al.¹¹ have precisely evaluated the phenotypic segregation ratios in several progeny trials obtained from crosses of curly and non-curly birch parent trees. As a result, a simple Mendelian inheritance model for curliness was proposed: a monogenic trait with two alleles, the allele coding for curly phenotype is dominant over the allele coding for normal phenotype and the semi-dominant curly allele is lethal when homozygous¹¹.

With advances in modern high throughput genotyping technologies and well-developed experimental workflow of the Genome-Wide Association Study (GWAS), it is possible to take another step forward in the study of the molecular genetic basis of the curly birch phenotype. In the present study, we used 37,045 SNP markers obtained by RADseq to genotype 192 trees with and without the ‘curliness’ phenotype derived from Karelian birch crosses, to reveal how many loci could be potentially involved in ‘curliness’ formation and to search for genetic variants associated with this trait.

Materials and methods

Plant material and phenotyping

In the study we investigated the full-sib seed progeny of the Karelian birch, growing on the Zaonezhskaya Forest-Seed Plantation, Medvezhyegorsk region (62° 15' N 35° 02' E) in the south-eastern part of Karelia (Russia). 192 trees from two full-sib families derived from two controlled crosses between Karelian birch parents №133 × №134 (88 trees) and №51 × №58a (104 trees) have been analyzed (Supplementary Table S1). The crosses were carried out in 1987 and 2006, respectively, direct and reciprocal crosses were performed. Phenotyping was carried out in June 2022. “Curliness” (curly wood, cw) of phenotyped trees was determined by a nondestructive testing approach: by stem growth form and bark color, but mostly by the presence and shape of bulges on the stem surface. Among these, spherical thickened, finely tuberculate and also without signs of “curliness” are visually distinguished (Fig. 1). According to the type of the surface of the stem, one can roughly estimate the level of the manifestation of wood “curliness”. The most intense patterned texture in wood, as a rule, is formed in Karelian birch trees with a finely tuberculate type of stem surface. The signs of “curliness” in a pronounced form appear on average in the 8–15th year of the plant’s life. It was also established that as the plants develop, after 30–40 years the reverse process of “smoothing” or “swollen” of the previously convex stem surface is observed due to an increase in the thickness of the bark (Fig. 1C). Although these features make it difficult to reliably identify Karelian birch without wood destruction, tree breeding programs make extensive use of this non-destructive approach to detect “curliness” based on the externally visible stem morphology conducted on 10-year-old and older birches^{5,11}. The field studies including collection of plant material and phenotyping on birch plants presented in this paper comply with relevant institutional, national and international guidelines and legislation.

Genotyping procedures

Genotyping-by-sequencing of plant material was performed using double-digested RAD sequencing (RADseq) method¹⁵. The detailed protocol of RADseq library preparation using *HindIII* and *NlaIII* restriction enzymes

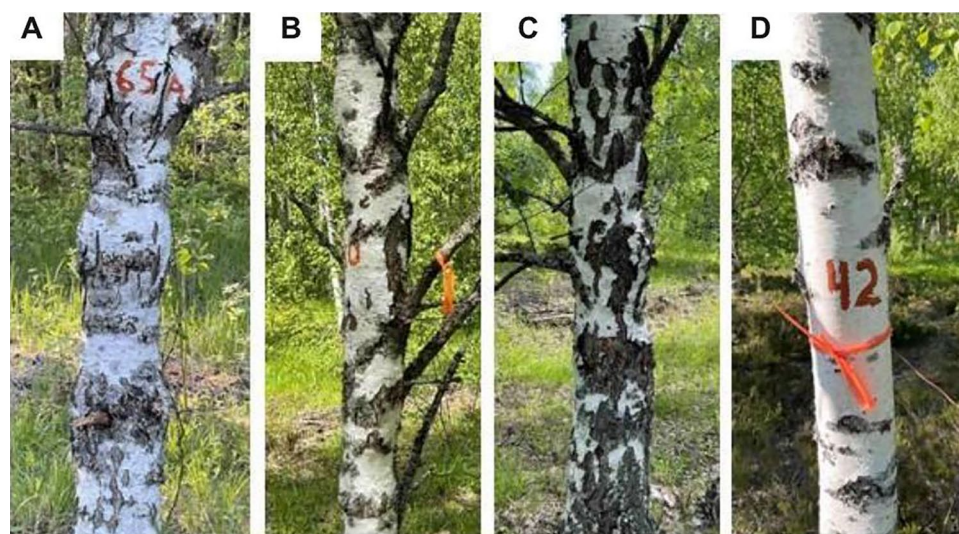


Figure 1. Types of the surface of the stem of the Karelian birch: spherical thickened (A), finely tuberculate (B), “swollen” (C) and without signs of “curliness” (D). Types of stem surface (A), (B) and (C) were considered “curly” phenotype, (D)—as “non-curly” phenotype.

was described previously¹⁶. In short, genomic DNA was isolated from frozen leaves using a modified CTAB method¹⁷. The integrity, quality and concentration of the purified DNA samples were determined by gel electrophoresis, using SPECTROstar Nano spectrophotometer (BMG LABTECH, Ortenberg, Germany) and Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

96 barcode adapters were used for indexing individual DNA samples. Thus, two RADseq libraries were prepared and sequenced separately to genotype 192 plants.

During library preparation the genomic DNA of each sample was digested with *HindIII* and an adapter with individual barcode was ligated to the sticky ends. Next, barcoded samples were mixed in one tube and purified with CleanMag DNA PCR magnetic beads (Evrogen, Moscow) to remove fragments shorter than 150 bp. The second restriction digestion was performed with *NlaIII*, the common adapter was ligated to the overhanging ends of *NlaIII* and purified again with magnetic beads.

The libraries were then enriched using PCR (98 °C for 30 s, 14 cycles of 98 °C for 10 s, 65 °C for 30 s, 72 °C for 15 s, and then 72 °C for 2 min) and purified using magnetic beads. The quality and quantity of two prepared amplicon libraries were determined using Qubit 3.0 Fluorometer and Agilent 2200 TapeStation.

Sequencing was performed using Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) in the paired-end mode, the length of reads was 150 bp. Raw sequence data is available on NCBI SRA under the project number PRJNA997794. Tree numbers, corresponding barcodes and tree phenotypes are summarized in Supplementary Table S1.

Read filtering, alignment, and SNP calling

The obtained reads were first distributed according to the sample barcodes in separate fastq files using the `process_radtags` function in Stacks software¹⁸. The quality of the reads and presence of Illumina adapters was assessed using FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were filtered using the Trimmomatic software¹⁹. Low-quality bases (Phred score < 18) in the reads, as well as reads that correspond to Illumina adapters and short reads (< 45 bp) were removed. The SNP calling procedure was performed using GATK v.4.4.0.0 software²⁰ after aligning reads on the reference genome assembly²¹. Reads alignment was performed using a bowtie2 aligner with a very sensitive flag²². Birch genome²¹ sequence version 1.4c was used as a reference (<https://genomeevolution.org/coge/api/v1/genomes/35080/sequence>). After alignment with bowtie2, reads were sorted within each bam file and next each bam file was indexed using Picard version 3.0 (<https://broadinstitute.github.io/picard/>). Next, for each bam file we generated individual vcf with the HaplotypeCaller function in GATK, then these files were merged using the `CombineGVCFs` function in GATK. Final SNP calling in combined vcf was performed with the GATK's `GenotypeGVCFs` function using the parameter `Max Alternate Alleles = 2` to select only biallelic SNPs. GATK's Variant Filtration function was used to filter SNPs based on the following parameters: Minor Allele Frequency (removing SNPs with frequency > 0.05), Mapping Quality (removing SNPs with score < 40.0), Quality by Depth (removing SNPs with score/depth < 24.0), Fisher Strand Bias (removing SNPs with *p*-value > 60.0), Strand Odds Ratio (removing SNPs with odds ratio > 3.0), and Depth (removing SNPs with cumulative depth < 20.0 across samples). More detailed information on SNP calling and example scripts used in the present study are available at GitHub (https://github.com/RimGubaev/GATK_pipeline_customized). To assess the coverage of the SNPs we estimated the mean number of reads that cover each SNP.

Population structure assessment

In order to find the number of subpopulations within the studied birch cohort, ADMIXTURE software version 1.3 was used²³. The potential number of subpopulations was set from 1 to 15. The principal component analysis was performed using default parameters in the PLINK software version 1.9²⁴. The number of principal components was set to 20. Principal components were estimated using the variance-standardized genetic relationship matrix. Linkage disequilibrium was evaluated using r^2 which in turn was derived from the pairwise comparisons made in the PLINK software version 1.9. The r^2 value was calculated for SNP pairs located within 1000 kb frames across the chromosomes. LOESS (locally estimated scatterplot smoothing) was used to estimate the regression between r^2 and distance.

Association mapping, SNP annotation

To identify genotype–phenotype associations, the GAPIT version 3 software was used²⁵. The mapping was performed using the SUPER (Settlement of MLM Under Progressively Exclusive Relationship) approach²⁶. The first five principal components, as well as the kinship matrix calculated using methods described previously²⁷ were added to the model to account for the population structure and kin relationships, respectively. In order to assume an association, the adjusted *p*-value (Bonferroni correction) of less than 0.00000135 (0.05/37,045) was used, where 37,045 is the number of tests (SNPs). Additionally, false discovery rate (FDR) correction was used according to the Benjamini-Hochberg (BH) procedure. The SNPs that demonstrated BH-corrected *p*-values less than 0.05 were assumed to be associated with the phenotype. To search for the candidate genes, the list of coding sequences was retrieved in the region on chromosome 10 identified by association mapping and restricted by SNPs S10_2623237 and S10_3848247 (± 30 kb). Original annotation in gff format for assembled genome was used to extract potential candidates within identified region on chromosome 10 (https://genomeevolution.org/coge/api/v1/downloads/?gid=35080&filename=Betula_pendula_subsp_pendula_annos1-cds0-id_typename-nu1-upa1-add_chr0.gid35080.gff). The proportion of the explained phenotypic variance was estimated by performing an analysis of the variance of the linear model (`lm()` function) in the R base package.

SNP validation, molecular marker development

Sanger sequencing was applied to validate the high throughput SNP call performed by GATK using RADseq data. Primer pairs flanking the most significant SNPs were designed. (Supplementary Table S2). For the primer design *B. pendula* genome sequence version 1.4c²¹ was used as a reference.

For all the designed primers PCR reaction mix contained 1 × Taq-buffer (2.5 mM Mg²⁺); 200 μM dNTP; 2.5 U Taq DNA polymerase (Evrogen, Moscow); 0.4 μM of each primer; ~ 20 ng of template DNA, sterile distilled water in a final volume of 25 μl. PCR cycling conditions consisted of an initial denaturation step of 95 °C for 3 min, followed by 30 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s, and a final extension cycle at 72 °C for 5 min. PCR products were purified using magnetic particles, followed by termination PCR with BrilliantDye™ Terminator Cycle Sequencing Kit 3.1 using forward and reverse primer. Sanger sequencing was performed with Applied Biosystems 3500 Genetic Analyzer. The DNA sequences were aligned with UniproUGENE software²⁸. The genetic marker BpCW1 described in this paper is the subject of a patent application no. 2023122806 filed by the authors and protected by the Federal Service for Intellectual Property of Russia.

Ethical approval and informed consent

All methods in the present research were performed in accordance with the relevant guidelines and regulations specified in Nature Portfolio journals' Editorial policies. The present research did not involve human or animal participants.

Research involving plants, permission for collection of plant material

The research and field studies on plants presented in this paper comply with relevant institutional, national and international guidelines and legislation. Karelian birch (*Betula pendula* Roth var. *carelica* (Mercklin) Hämet-Ahti) is a protected birch variety in the Republic of Karelia, permission for the collection of the plant specimens was issued by the Ministry of Natural Resources and Environment of the Republic of Karelia.

Results

Phenotyping, genotyping and SNP calling

The phenotypic evaluation was performed for 192 trees from two studied crosses. In the first full-sib population consisting of 88 trees, 70 plants were classified as 'curly' and 18 demonstrated non-curly phenotype. For the second cross (104 trees) 59 and 45 trees demonstrated curly and non-curly phenotypes, respectively (Supplementary Table S1).

The procedure of RADseq library preparation using *HindIII* and *NlaIII* enzyme combination and CleanMag bead 'lower cut' protocol resulted in two final libraries with the amplicon sizes varied from 450 to 600 bp. The RADseq procedure yielded an average of 16,090,513 paired-end sequencing reads for each birch DNA sample. After the read filtering procedures, an average of 9,107,446 reads per plant were mapped to the birch genome. By analyzing alignment files, we identified 3,065,097 biallelic SNPs using the GATK SNP caller. As a result of filtering, 37,045 SNPs remained for further analysis of population structure and association mapping (Supplementary Table S3). The mean number of reads that cover each SNP followed a Poisson distribution with an average coverage of 13.8 reads per SNP.

Population structure analysis revealed clusterization and rapid LD decay

To assess the population structure of the birch cohort we estimated the potential number of K clusters (sub-populations) using ADMIXTURE software²³. Next, we performed a principal component analysis implemented in PLINK software²⁴ to estimate the proportion of genotype variance explained by principal components as well as to perform visualization (Fig. 2A–D). Finally, we assessed a linkage disequilibrium decay using PLINK (Fig. 2E).

The most significant drop in cross-validation (CV) error was observed when binning the birch cohort into two clusters (K = 2) (Fig. 2A,B). This is in concordance with the information that the studied cohort was represented by two populations derived from two independent crosses. Subsequent binning at K = 4, 7, 11 divides each of the two studied populations into different groups (Fig. 2B). However, no correlation between the observed phenotype of the tree wood and clustering at K = 4, 7, 11 were identified. Thus, the observed clustering could be due to random genetic segregation. The first three principal components explained 50.74% of genotypic variance which is in concordance with the strong population clustering (Fig. 2C,D). The first principal component explained the variation between the two studied populations obtained from two different crosses, while the second and the third principal components explained variation within them. Notably 'curly' and 'non-curly' phenotypes were evenly distributed between and within two full-sibs' populations. Linkage disequilibrium (LD) decayed rapidly and was equal to 30.8 kb at $r^2 = 0.25$ and 79.5 kb at $r^2 = 0.15$ (Fig. 2E).

Association mapping revealed the target locus on chromosome 10

To perform association mapping, the SUPER model was applied accounting for population structure and relationships by introducing the first five principal components and kinship matrix. Out of the 37,045 SNPs tested, 3 SNPs (S10_3168885, S10_3465040, S10_3472479) were found to be significant after Bonferroni correction for multiple testing (Fig. 3).

Using a softer FDR threshold ($\alpha = 0.05$) 5 additional SNPs (S10_2623237, S10_3103637, S10_3835235, S10_3848236, S10_3848247) surrounding the above-mentioned SNPs were identified to be significantly associated with the studied trait. 5 out of 8 SNPs were found within the coding sequences annotated for the *B. pendula* reference genome (Table 1). In addition, S10_3465040, detected outside of any coding sequence, was located in

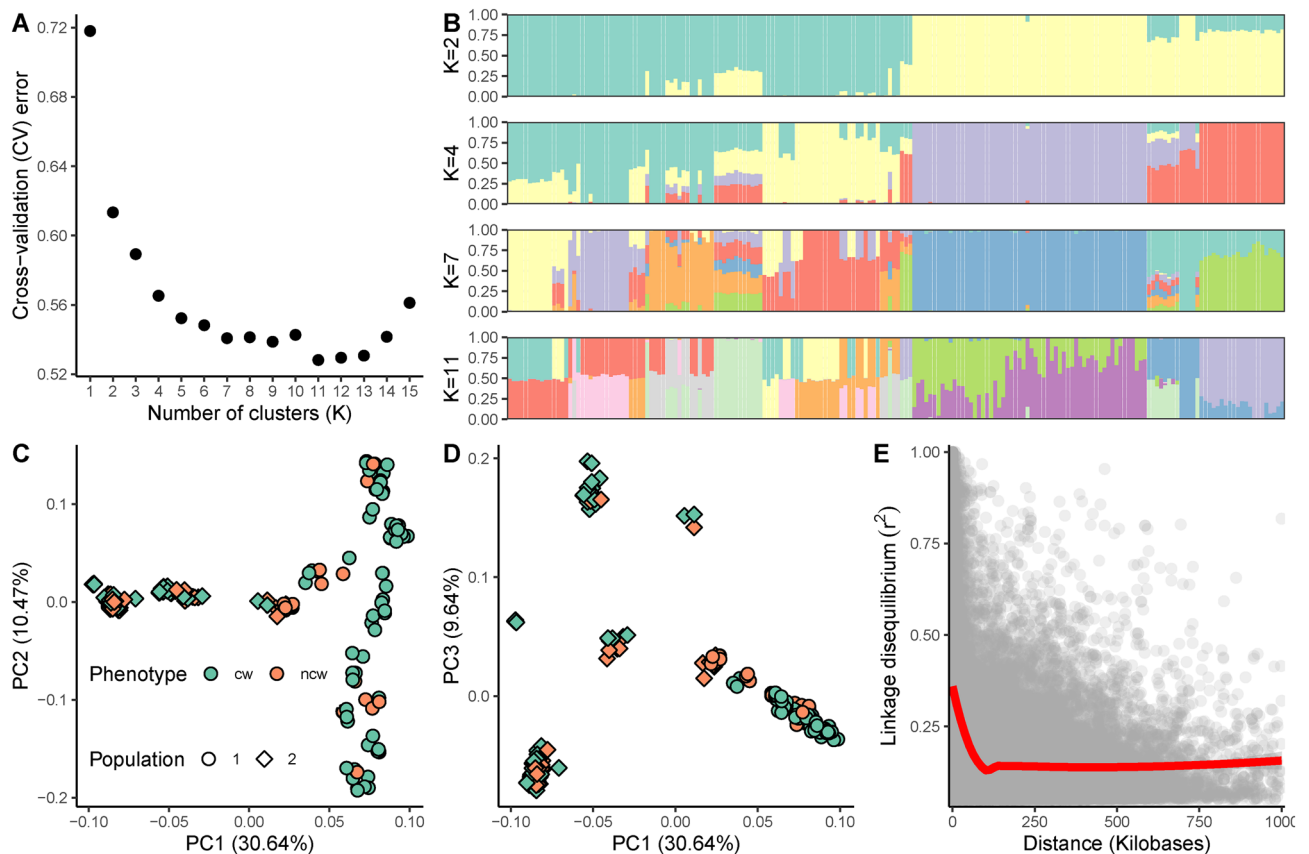


Figure 2. Population structure of the studied Karelian birch cohort. Panels A and B represent the results of the ADMIXTURE software. (A) Estimated cross-validation error value for possible clusters from 1 to 15. The drop in the cross-validation error indicates the optimal number of possible subpopulations. (B) ADMIXTURE bar plots reflecting the subpopulations at $K=2, 4, 7, 11$ each bar corresponds to the birch accession, colors indicate the proportion of subpopulation admixtures in each accession. For $K=2$ green color corresponds to the genetic admixtures from population 2, while yellow color corresponds to the genetic admixtures from population 1. (C) and (D) represent the result of the Principal Component Analysis. Colors correspond to the phenotype of the trees (cw—'curly', ncw—'non-curly') and the shape indicates the corresponding full-sib population according to the plan of the birch plantation. The numbers in brackets correspond to the proportion of genotype variance explained by the principal component. (E) Linkage disequilibrium (LD) decay in the studied cohort of birch trees. Each dot corresponds to the r^2 value between a pair of SNPs. The gray marker corresponds to the 95% confidence interval. The red line represents the LOESS curve.

the 3'UTR region of the Bpev01.c0000.g0109 gene, at 27 bp downstream of the stop codon. The mean number of reads that cover each SNP associated with the 'curly' phenotype varied from 13.07 to 39.43 (Table 1).

At the next step we assessed the effect that the SNPs exerted on the phenotype. To do so the phenotype distribution was visualized (Fig. 4). For one of the most significant SNP (S10_3465040) the 'non-curly' phenotype was associated with 'G' allele in the homozygous state, while the presence of the alternative 'A' allele indicated the curly phenotype. Remarkably, this SNP showed the largest phenotypic variance explained (PVE = 56.28%). That makes the S10_3465040 a good candidate for marker-assisted breeding applied to the curly wood trait.

Since all identified SNPs were located at the 10th chromosome, we assumed the presence of a locus associated with the studied trait (Fig. 3C). Despite the quite large chromosome region that carries SNPs associated with the trait (more than 1200 kb), the marginal SNPs (S10_2623237 and S10_3848247) demonstrated moderate ($r^2 > 0.3$) linkage (Fig. 3D). Hence, we considered this region as potentially carrying a candidate gene or genes associated with the curly wood phenotype.

Detailed analysis of the region restricted by SNP S10_2623237 (3'-end) and SNP S10_3848247 (5'-end) plus an additional 30 kb at both ends, yielded 104 coding sequences (Supplementary Table S4). Among the sequences encoding proteins with known function, there were some genes previously discussed as potential players in curly wood formation. Most of the genes have been annotated as encoding unknown proteins, however, some genes from the list were previously discussed as potential players in curly wood formation.

For example, Bpev01.c0000.g0045 (Bpe_Chr10_2774789-2776265) encoding protein from Flavin-containing monooxygenase_family located approx. 150 kb upstream from the candidate SNP S10_2623237. This coding sequence was recently reported as the *BpYucca5* gene involved in auxin biosynthesis²⁹. The *Yucca* gene family is a key player in IAA (the main natural auxin form) homeostasis. Much of IAA in plants is synthesized from tryptophan with the involvement of *Yucca* flavin-dependent monooxygenases³⁰. It has recently been proposed that

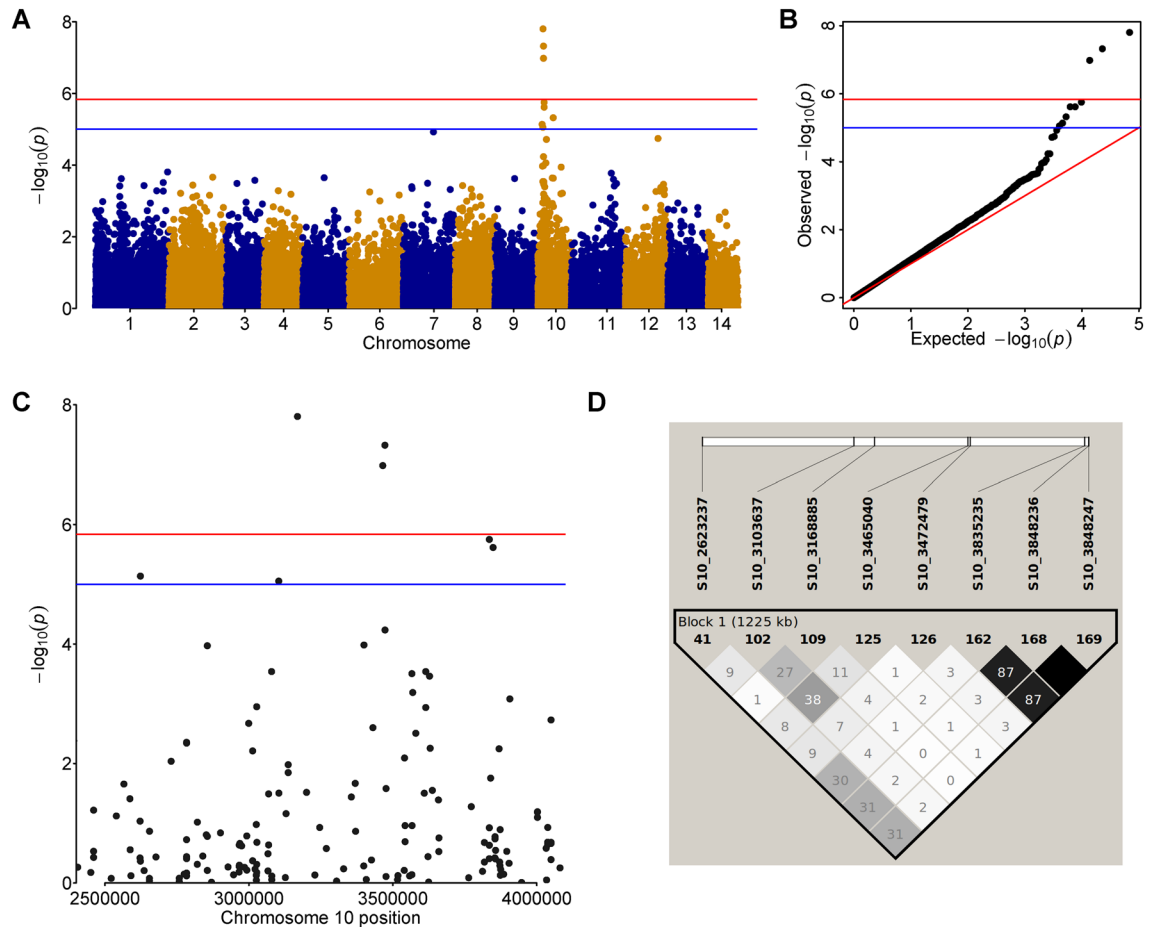


Figure 3. Results of the association mapping of curly wood trait. **(A)** Manhattan plot and respective, **(B)** QQ-plot showing SNP markers associated with the ‘curly’ phenotype of Karelian birch. Each dot corresponds to a single SNP. Blue dots correspond to SNPs from odd chromosomes, yellow dots correspond to SNPs from even chromosomes. **(C)** Manhattan plot for the region of interest on chromosome 10 restricted with SNPs S10_2623237 and S10_3848247. The red line corresponds to the Bonferroni-adjusted significance threshold. The blue line corresponds to the FDR-adjusted significance threshold. Linkage between the associated SNPs **(D)**, numbers in squares correspond to the digits after the comma of the r^2 value between a pair of SNPs.

the development of anomalous wood in Karelian birch may be associated with impaired auxin biosynthesis^{14,29}. Another possibly relevant gene Bpev01.c0000.g0081 encoding Sm-like snRNP protein (Bpe_Chr10_3118250-3127056) is located in 14,613 bp downstream from SNP S10_3103637. Sm-like snRNP proteins are required for mRNA splicing, export, and degradation. It was shown that the homolog of the candidate gene Bpev01.c0000.g0081 in *Arabidopsis* (Sm-like protein (*SAD1*), AT5G48870) affects both ABA sensitivity and drought-induced ABA biosynthesis: ABA-deficiency was detected in *sad1* mutant plants³¹. The authors proposed a critical role for the *SAD1* gene in RNA metabolism: *sad1* mutation affects the decay rate of mRNA for an early component(s) in ABA signaling in drought stress conditions.

SNP validation revealed the first molecular marker for *B. pendula* curly wood trait (BpCW1)

Validation by Sanger sequencing was performed for 4 out of 8 candidate SNPs: three of them were found to be significant by multiple testing, and one additional SNP (S10_3103637) was selected because it was the only SNP identified in an exon sequence (Table 1). PCR fragments obtained with primers flanking each of the four most significant SNPs were re-sequenced for the randomly chosen 16–20 birch trees. The accuracy of genotyping was estimated as 81%, 100%, 86%, 75% for SNPs S10_3103637, S10_3168885, S10_3465040, S10_3472479, respectively (Supplementary Table S5).

The length of PCR products obtained for neighboring SNPs S10_3465040 and S10_3472479 varied significantly in size in the populations due to numerous InDels detected within the amplified fragments. In particular, in the sequence of the PCR fragment containing S10_3472479, at least two InDels of 101 bp and 6 bp were detected by Sanger sequencing. PCR analysis of the entire sample of 192 trees with the primers specific for S10_3472479 revealed at least three more length variants of the PCR products (Supplementary Fig. S6). None of the InDels adjacent to S10_3472479 were significantly linked to the ‘curly’ phenotype. As a consequence, no correlation between the PCR fragment size and the ‘curly’ phenotype was detected (Supplementary Fig. S6).

SNP chromosome_position	Mean SNP coverage	P value	FDR	PVE, %	Coding sequence, putative function	Nearest upstream transcript match	Nearest downstream transcript match
S10_2623237	18.12	7.3E-06	0.0313	16.78	Bpev01.c0000.g0035 unknown protein	Bpev01.c0000.g0034, Alcohol dehydrogenase	Bpev01.c0000.g0036, RmlC-like cupins superfamily protein
S10_3103637	14.15	8.85E-06	0.0337	37.25	Bpev01.c0000.g0077, 3-ketoacyl-CoA thiolase (exon)	Bpev01.c0000.g0076, PB1 domain-containing protein tyrosine kinase	Bpev01.c0000.g0078, mRNA splicing factor thioredoxin-like U5 snRNP
S10_3168885	21.17	1.57E-08	0.0005	25.24		Bpev01.c0000.g0085, Alba DNA/RNA-binding protein LENGTH 405	Bpev01.c0000.g0086, O-Linked N-acetylglucosamine transferase
S10_3465040	25.28	1.04E-07	0.0012	56.28		Bpev01.c0000.g0109, unknown protein	Bpev01.c0000.g0110, Pentatricopeptide repeat-containing protein
S10_3472479	17.34	4.77E-08	0.0008	22.74		Bpev01.c0000.g0110, Pentatricopeptide repeat-containing protein	Bpev01.c0000.g0111, RING/U-box superfamily protein LENGTH 360
S10_3835235	39.43	1.78E-06	0.0139	10.01	Bpev01.c0000.g0140, LanC-like protein 2	Bpev01.c0000.g0139, Plant protein of unknown function (DUF827) LENGTH 1345	Bpev01.c0000.g0141, SNF1-related protein kinase
S10_3848236	13.07	2.42E-06	0.0139	9.03	Bpev01.c0000.g0141, SNF1-related protein kinase	Bpev01.c0000.g0140, LanC-like protein 2	Bpev01.c0000.g0142, Putative Peroxidase
S10_3848247	13.12	2.42E-06	0.0139	9.03	Bpev01.c0000.g0141, SNF1-related protein kinase	Bpev01.c0000.g0140, LanC-like protein 2	Bpev01.c0000.g0142, Putative Peroxidase

Table 1. SNPs associated with the ‘curly’ phenotype in two full-sib populations obtained from crossing Karelian birches according to GWAS results.

At least 3 InDels of 54 bp, 13 bp and 2 bp were detected within the PCR products amplified with primers flanking S10_3465040. Notably, the 54 bp deletion both in the homozygous and heterozygous state correlated perfectly with the “curly” phenotype in the analyzed populations. 54 bp deletion results in a 476 bp amplified fragment compared to the size of 530 bp intact PCR fragment, which makes it easily recognizable by electrophoresis (Fig. 5). Of the 190 trees tested for the presence of a 54 bp deletion with this PCR marker, 174 trees (92%) were identified correctly as ‘curly’ or ‘non-curly’ depending on the presence or absence of the 476 bp PCR-fragment (Supplementary Fig. S7). To our knowledge, this is the first molecular marker proposed for the recognition of *Betula pendula* Roth var. *carelica* with the curly wood phenotype (BpCW1).

Discussion

In this study we attempted to unravel the enigma of genetic factors underlying the phenomenon of curly wood in Karelian birch, which has been debated for over a century³². Although in 1922 it was already shown that the patterned texture of the wood is quite stably transmitted to progenies during seed propagation³³, there were many adherents of the ‘infectious’ or ‘pathological’ hypotheses, which considered curly wood to be the result of exposure to biotic or abiotic stressors^{34–36}. The idea finds support in some modern studies, for example, the influence of a low-temperature stress gradient is assumed, which determines the boundaries of the distribution of Karelian birch³⁷. According to the reports, the suppression of xylem-phloem transport and a decrease in the rate of plant growth occur in parallel with an increasing deficit of heat, moisture, or mineral nutrition. Recently, an ecological and genetic hypothesis has been proposed, according to which the origin of the Karelian birch is associated both with the natural and climatic conditions of its growth and with genetic factors^{38,39}.

Ruden⁴⁰ was the first to postulate that the ‘curly’ trait is monogenic, the gene can be dominant and lethal when homozygous. Later it was proposed that the inheritance of curly wood is due to a series of multiple alleles in one locus or genes in different loci⁴¹. In 2011 a hypothesis about the epigenetic origin of curly wood of the Karelian birch was suggested⁴².

Since the aberrant ‘curly’ phenotype occurs regularly, albeit sporadically within the north-western part of the silver birch distribution area, and is inherited in offspring, the hypothesis that curly birch can be a single-locus mutant of *B. pendula* seems quite plausible. This hypothetical mutation apparently disrupts the normal development of the plant, affecting not only the general habitus but probably also the tree’s longevity. As stated by some reports, the life cycle of the Karelian birch takes approximately 50–60 years^{35,39,43,44}, while the life cycle of the silver birch takes 120–140 years. The impairment of xylogenesis observed in Karelian birch may be due to the loss of function of a particular causative gene. It may also play a role in whether this mutation, which affects the fitness of the tree, is present in a homo- or heterozygous state. Accordingly, Kärkkäinen et al.¹¹ reported that their analysis of crosses studying the inheritance of curly birch phenotype indicates a Mendelian inheritance based on a single semi-dominant gene that is lethal when homozygous.

Consistent with this report, our study based on RADseq and association mapping revealed a single interval on chromosome 10 spanning ~1200 kb, where 8 SNPs were found significantly associated with the trait under study. The fact that one of the SNP markers (S10_3465040) was identified explaining up to 56% of phenotypic variation (Table 1) speaks in favor of the previously suggested monogenic (Mendelian) mode of inheritance for the curly wood trait. This particular SNP S10_3465040 and the 7 kb distant SNP S10_3472479 seem to fall into a

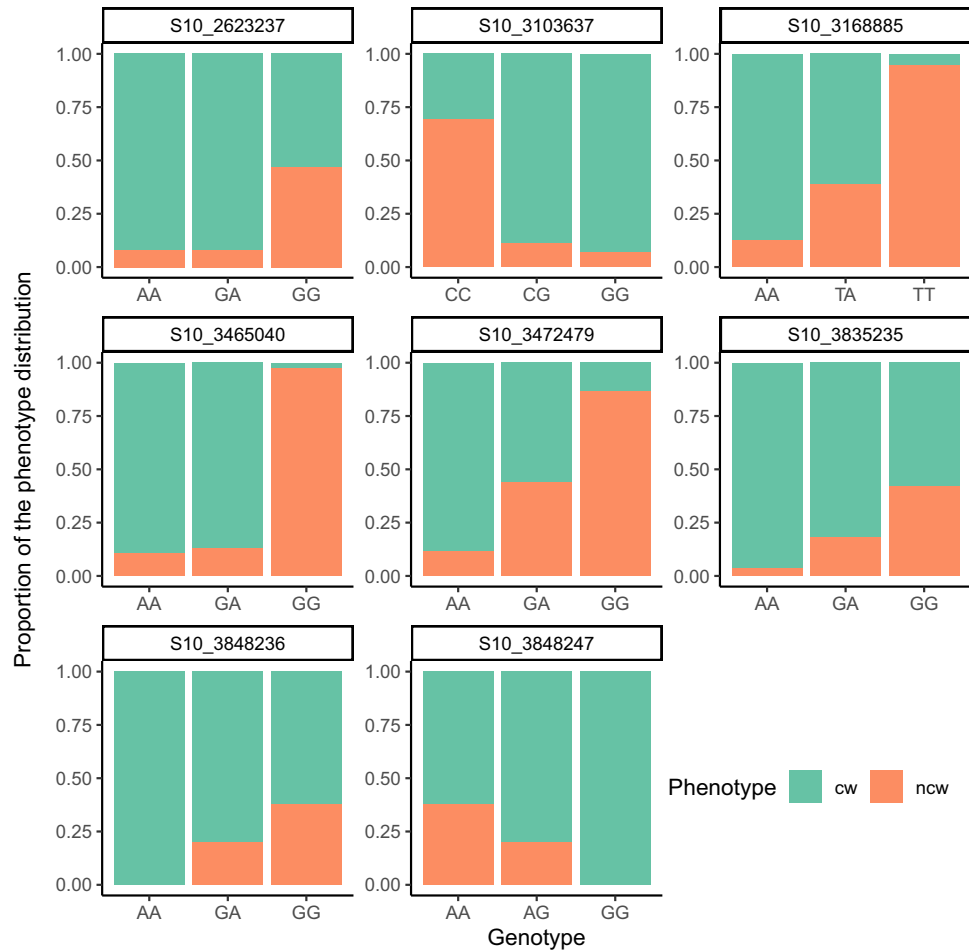


Figure 4. Barplots reflecting the effects of detected SNPs on the segregation of trees according to the “curly” phenotype. Each bar corresponds to a phenotype distribution across one of the three possible genotypes. Each colored bar shows the proportion of the phenotype distribution (cw—curly wood, ncw—non-curly wood) across genotypes.

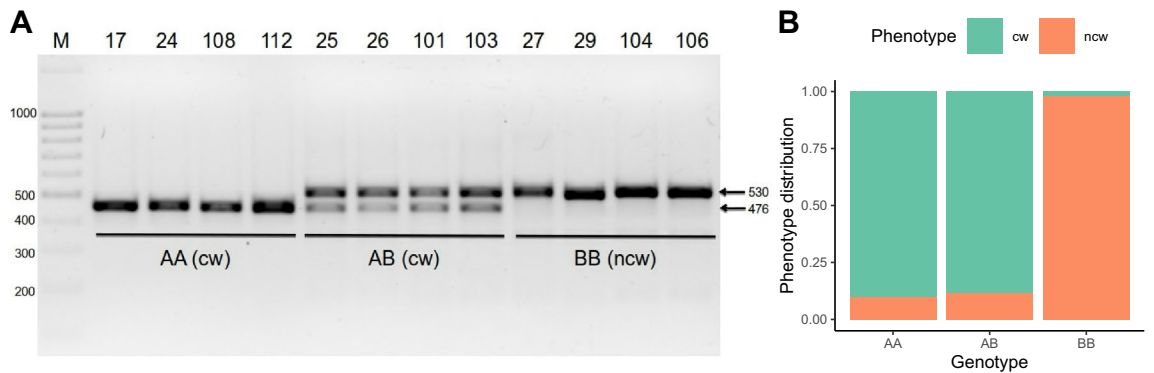


Figure 5. Results of the validation of the 54 bp deletion associated with SNP marker S10_3465040. (A) The result of PCR amplification of DNA fragments containing deletion associated with the curly wood phenotype. Samples 17, 24, 108, 112 and 25, 26, 101, 103 demonstrating curly wood (cw) phenotype carry deletion in homozygous (AA) and heterozygous state (AB), respectively. Samples 27, 29, 104, 106 do not carry deletion (BB) and demonstrate non-curly wood phenotype (ncw). (B) Phenotype (cw, ncw) distribution among genotypes associated with the deletion (AA, AB, BB) for the studied cohort of 192 trees.

region on the chromosome rich in small InDels. Three InDels were found only within 530 bp surrounding SNP S10_3465040, and even more within the 621 bp interval surrounding SNP S10_3472479.

Thus, in the region of interest, we observed a distinct structural characteristic: a high frequency of insertions and deletions of varying sizes. This structural dynamism might hint at the active presence of transposable elements (TEs) within this region. TEs, often termed “jumping genes”, have been recognized for their capacity to move within the genome, and their activation can be influenced by various factors. TEs are not just passive genomic elements; they can play pivotal roles in gene regulation, especially in response to environmental stresses⁴⁵. Furthermore, TEs can influence alternative splicing patterns, potentially serving as alternative splicing start or acceptor sites⁴⁶.

We found that the 54 bp InDel located at a distance of 160 bp from SNP S10_3465040 successfully predicted the “curly” phenotype for 92% of the 190 trees examined. It should be considered that the phenotyping of the “curly” trait was carried out by the method of non-destructive visual assessment of the habitus of the tree, without assessing the cut, so a certain percentage of erroneous phenotyping can be expected. Notably, of the 16 mislabeled trees, 15 were diagnosed as “curly” by the PCR marker, while visual assessment phenotyped these trees as “non-curly”. Only one tree was mislabeled the other way around. It should be considered that by the time of visual assessment (June, 2022), almost half of the phenotyped trees had just reached 16 years of age. Thus, in some of these “mislabeled” individuals the “curly” phenotype may not yet be visually manifested, but will appear at a later stage.

The identified 54 bp InDel does not fall into any coding sequence, although it is located in the 3'UTR region of the Bpev01.c0000.g0109 gene encoding an unknown protein. Blast analysis of the Bpev01.c0000.g0109 sequence revealed just a few hits—in the genomic assemblies of *Quercus*, *Alnus* and *Fagus*, which suggests that it may be specific to the *Fagales* order. Deletions in the 3'UTR region of this unknown gene can dramatically affect the fate of its mRNAs. The 3'-UTR regulatory regions are involved in polyadenylation, stability, transport and mRNA translation in plant cells⁴⁷. When the effect of increasing the length of the 3'UTR on expression of luciferase mRNA was examined in transiently transfected carrot protoplasts, expression increased 18-fold when the length of the 3'-UTR was increased from 7 to 27 bases⁴⁸.

Several studies have highlighted the profound effects of TE insertions, especially within the 3'UTR regions of genes. Such insertions can lead to translational repression, thereby influencing phenotypic outcomes. For instance, in rice, the *Ghd2* gene, a regulator of flowering time, is impacted by such a mechanism⁴⁹. This phenomenon isn't exclusive to plants. In *Drosophila*, a TE insertion has been linked to enhanced survival rates in response to chemicals, due to its effect on the upregulation of the *CG11699* gene⁵⁰. Similarly, in mammals, TEs, particularly SINE repeats, have been shown to modulate gene promoter activity⁵¹. A study in humans demonstrated that the expression of an AcGFP reporter gene is significantly attenuated by inverted Alu repeats in the 3'UTR⁵². Given the proximity of this SNP to the 3'UTR and the known effects of TEs on gene regulation, it can be assumed that TEs may influence the curly wood phenotype of the Karelian birch.

Bpev01.c0000.g0109 can be considered as one of the possible candidate genes, however, linkage disequilibrium (LD) between the S10_3465040 and the putative causal locus located near the identified polymorphism can also be assumed. Rapid LD decay was reported for many arboreal species including pines⁵³, aspens⁵⁴ and willows⁵⁵ due to the high level of genetic diversity (heterozygosity) of these woody species. In our study, though, the identified LD at $r^2=0.15$ was higher (79.5 kb) compared to one revealed in the previous study (23 kb) performed on the 60 genetically diverse birch trees²¹. This could be explained by the fact that the association study was performed using a cohort of related individuals (two populations of full-sibs), which undergo only few recombination events. This may be the reason why the chromosomal interval at LG10, where significantly associated SNPs were detected, stretched to 1200 kb.

The detected interval on chromosome 10 was discussed previously as containing the sequence of the *BpYucca5* gene involved in auxin biosynthesis which is critical for normal xylogenesis in Karelian birch²⁹. The *BpYucca5* gene is listed as one of the 104 coding sequences detected within the interval on chromosome 10. Among the potential candidates, there are also several genes encoding proteins that facilitate the processing, splicing, editing, stability and translation of mRNAs (Supplementary Table S4). Some of them are regulated by abiotic stress factors, such as Sm-like_snRNP_protein (Bpe_Chr10_3118250-3127056) which affects ABA biosynthesis under drought conditions.

Assuming that a causal polymorphism (in which TEs may be involved) can influence gene expression or mRNA stability depending on the environment, it may be explained why the Mendelian inheritance of this valuable trait has long been contested. Indeed, the formation of figured patterns in the wood of Karelian birch can begin at any age: cuts of trees are described in which patterned wood began to form at 8, 15, and even at 25 years. Moreover, there are examples of cross-cuts of trees, where the formation of a patterned wood texture is visible only in a certain sector of the trunk, perhaps, due to the special light conditions for the growth of the tree².

The phenomenon, when the severity of the phenotype caused by the genotype can vary among affected individuals, is known in genetics as “expressivity”. Nowadays, both variable expressivity and incomplete penetrance are mainly investigated in studies focusing on monogenic human diseases. They are described to be caused by a range of factors, including variants in regulatory regions, epigenetics and environmental factors. The exact causes of expressivity are still not well understood, although they most likely lie in the molecular mechanisms governing genetic regulation^{56,57}. Accordingly, in Karelian birch the dominant allele may underlie the cause of figured wood, however, the degree of xylogenesis anomalies and the stage of development at which they occur may depend on environmental conditions. That may explain the reason why the segregation ratio of curly and non-curly progenies in the crossing of Karelian birches varies from study to study^{2,58}.

Our results suggest that the variation of the “curly” phenotype in Karelian birch is linked to the single interval on chromosome 10, where a high frequency of insertions and deletions of different sizes was observed. Unfortunately, the RADseq approach employed in our study doesn't provide a comprehensive resolution of this region.

To determine the true causal polymorphism in this interval, a deeper resequencing effort using more genetically diverse plant material is required. The parents of the crosses in our study were not related to each other, but they originated from the same geographical region—the Republic of Karelia. Thus, the identified SNPs associated with the curly phenotype and the proposed InDel marker should be verified in an independent sample of birch trees segregated for “curliness”. This will provide information on whether the BpCW1 marker is useful to phenotype Karelian birch with other genetic backgrounds.

The discovered PCR marker BpCW1, though, is the first marker suggested for selection at an early stage of development of the Karelian birch seedlings, which later, upon reaching the age of 8–10 years, will most likely show the characteristic features of decorative (patterned) wood.

Identification of genetic determinants of curly wood phenotype and development of corresponding molecular markers can prevent excessive costs and efforts in the maintenance of Karelian birch plantations for industrial purposes and, possibly, reduce the anthropogenic pressure on the natural populations of these trees.

Data availability

The high-throughput sequencing datasets generated during the current study are available in the NCBI's Sequence Read Archive repository, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997794>.

Received: 9 September 2023; Accepted: 22 February 2024

Published online: 01 March 2024

References

- Hagqvist, R. & Mikkola, A. *Visakoivun kasvatus ja käyttö* (Metsäkustannus and Visaseurary, Hämeenlinna, 2008).
- Vetchinnikova, L. V. & Titov, A. F. The Karelian birch: A unique biological object. *Biol. Bull. Rev.* **10**, 102–114 (2020).
- Viherä-Aarnio, A. & Hagqvist, R. Curly birch (*Betula pendula* var. *carelica*), wooden “marble” from finland—soon easily available. In *Proceedings of the International Scientific Conference on Hardwood Processing* (Lahti, Finland, 2017).
- Hynninen, J. *et al.* Silviculture of birch (*Betula pendula* Roth and *Betula pubescens* Ehrh.) in northern Europe. *Forestry* **83**, 103–119 (2010).
- Vetchinnikova, L. V., Titov, A. F. & Kuznetsova, T., Yu. *Karel'skaya bereza: biologicheskie osobennosti, dinamika resursov i vosproizvodstvo = Curly birch: biological characteristics, resource dynamics, and reproduction*. (Karel'skii nauchnyi tsentr RAN, Petrozavodsk, 2013).
- Ermakov, V. *Mekhanizmy adaptatsii berezy k usloviyam Severa [Mechanisms of birch adaptation to the conditions of the North]*. (Izd-vo " Nauka," Leningradskoe otd-nie, 1986).
- Ermakov, V. & Vetchinnikova, L. Growing birch wood with combined grain. *Drevarsky Vysk.* **46**, 37–42 (2001).
- Johnsson, H. Avkommor av Masur bjork (Experiments with Masur Birch). *Sartryck Ur Sven. Skogsvardsforeningens Tidskr.* **1**, 12 (1951).
- Heikinheimo, O. Experiences in the growing of curly birch. *Commun. Inst. Fenn.* **39**, 1–26 (1951).
- Larsen, C. Curly birch. *Dan. Skovforen. Tidsskr.* 33–72 (1940).
- Kärkkäinen, K., Viherä-Aarnio, A., Vakkari, P., Hagqvist, R. & Nieminen, K. Simple inheritance of a complex trait: Figured wood in curly birch is caused by one semi-dominant and lethal Mendelian factor?. *Can. J. For. Res.* **47**, 991–995 (2017).
- Vetchinnikova, L. V. & Titov, A. F. Curly birch: Some secrets remain. *Biol. Bull. Rev.* **13**, 162–174 (2023).
- Novitskaya, L. L. Regeneration of bark and formation of abnormal birch wood. *Trees* **13**, 74–79 (1998).
- Novitskaya, L. L. *et al.* The formation of structural abnormalities in Karelian birch wood is associated with auxin inactivation and disrupted basipetal auxin transport. *J. Plant Growth Regul.* **39**, 378–394 (2020).
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**, e37135 (2012).
- Volynkin, V. *et al.* The assessment of agrobiological and disease resistance traits of grapevine hybrid populations (*Vitis vinifera* L. × *Muscadina rotundifolia* Michx.) in the climatic conditions of Crimea. *Plants* **10**, 1215 (2021).
- Rahimah, A. R., Cheah, S. C. & Rajinder, S. Freeze drying of oil palm (*Elaeis guineensis*) leaf and its effect on the quality of extractable DNA. *J. Oil Palm Res.* **18**, 296–304 (2006).
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: Building and genotyping loci de novo from short-read sequences. *G3 Genes/Genomes/Genetics* **1**, 171–182 (2011).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).
- McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Salojärvi, J. *et al.* Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat. Genet.* **49**, 904–912 (2017).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Lipka, A. E. *et al.* GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
- Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER powerful method for genome wide association study. *PLOS ONE* **9**, e107684 (2014).
- VanRaden, P. M. *et al.* Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16–24 (2009).
- Okonechnikov K., Golosova O., Fursov M. & UGENE Team. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
- Tarelkina, T. V. *et al.* Expression analysis of key auxin biosynthesis, transport, and metabolism genes of *Betula pendula* with special emphasis on figured wood formation in Karelian birch. *Plants* **9**, 1406 (2020).
- Ljung, K. Auxin metabolism and homeostasis during plant development. *Development* **140**, 943–950 (2013).
- Xiong, L. & Zhu, J.-K. Regulation of abscisic acid biosynthesis. *Plant Physiol.* **133**, 29–36 (2003).
- Vetchinnikova, L. V. & Titov, A. F. *Curly birch: major research results and prospects for future research* (Petrozavodsk, 2021).
- Hintikka, T. J. Die „Wisa“-Krankheit der Birken in Finnland. *Z. Für Pflanzenkrankh. Gallenkd.* **32**, 193–210 (1922).
- Atanasoff, D. Virus stem pitting of birch (Wisa and Maser Disease). *Z. Für Pflanzenkrankh. Pflanzenpathol. Pflanzenschutz* **74**, 205–208 (1967).

35. Saks, K. & Bander, V. New data about the origin of Karelian birch. *Tr Inst Ekol Rast Zhivotn* (1975).
36. Vailionis, L. Lietuvos berzu reta. Referat: Die Wisakrankheit in den Wäldern Litauens. *Kaunas Sr Hort Bot Univ* 3, 5–36 (1935).
37. Novitskaya, L. L. *Karelian birch: mechanisms of growth and development of structural anomalies* (Petrozavodsk, 2008).
38. Vetchinnikova, L. V. & Titov, A. F. The origin of the Karelian birch: An ecogenetic hypothesis. *Russ. J. Genet. Appl. Res.* 7, 665–677 (2017).
39. Vetchinnikova, L. V. *Karel'skaya bereza i drugie redkie predstaviteli roda Betula L* (Nauka, 2005).
40. Ruden, T. On speckled birch (“mazer-birch”) and some other forms of curled birch. *Meddelelser Fra Det Nor. Skogforsöksves.* 43, 455–505 (1954).
41. Johnsson, H. Genetic characteristics of *Betula verrucosa* Ehrh. and *B. pubescens* Ehrh. *Ann. For.* 6, 91–133 (1974).
42. Mashkina, O. S., Butorina, A. K. & Tabatskaya, T. M. Karelian birch (*Betula pendula* Roth. Var. *carelica* Merkl.) as a model for studying genetic and epigenetic variation related to the formation of patterned wood. *Russ. J. Genet.* 47, 951–957 (2011).
43. Ermakov, V. Reproduction of Karelian birch by inoculation method. *Lesn. Genet. Sel. Semenovod. For. Genet. Sel. Seed Breed.* 282–293 (1970).
44. Mikkilä, H. *Guide to the Montell Trail in the Punkaharju Experimental Area.* (Metsäntutkimuslaitos, 1996).
45. Reddy, A. S. N., Marquez, Y., Kalyna, M. & Barta, A. Complexity of the Alternative Splicing Landscape in Plants. *Plant Cell* 25, 3657–3683 (2013).
46. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* 22, 1503–1517 (2013).
47. Bernardes, W. S. & Menossi, M. Plant 3' Regulatory Regions From mRNA-Encoding Genes and Their Uses to Modulate Expression. *Front. Plant Sci.* 11, 1252 (2020).
48. Tanguay, R. L. & Gallie, D. R. The effect of the length of the 3'-untranslated region on expression in plants. *FEBS Lett.* 394, 285–288 (1996).
49. Shen, J. *et al.* Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. *Nat. Commun.* 8, 14651 (2017).
50. Mateo, L., Ullastres, A. & González, J. A Transposable Element Insertion Confers Xenobiotic Resistance in *Drosophila*. *PLoS Genet.* 10, e1004560 (2014).
51. Estácio, M. R. H. *et al.* SINE Retrotransposons Cause Epigenetic Reprogramming of Adjacent Gene Promoters. *Mol. Cancer Res.* 10, 1332–1342 (2012).
52. Fitzpatrick, T. & Huang, S. 3'-UTR-located inverted Alu repeats facilitate mRNA translational repression and stress granule accumulation. *Nucleus* 3, 359–369 (2012).
53. Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H. & Neale, D. B. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci.* 101, 15255–15260 (2004).
54. Ingvarsson, P. K. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics* 169, 945–953 (2005).
55. Berlin, S., Fogelqvist, J., Lascoux, M., Lagercrantz, U. & Rönnerberg-Wästljung, A. C. Polymorphism and divergence in two willow species, *Salix viminalis* L. and *Salix schwerinii* E. Wolf. *G3 GenesGenomesGenetics* 1, 387–400 (2011).
56. Miiko, I. Phenotype variability: Penetrance and expressivity. *Nat. Educ.* 1, 137 (2008).
57. Kingdom, R. & Wright, C. F. Incomplete penetrance and variable expressivity: From clinical studies to population cohorts. *Front. Genet.* 13, 920390 (2022).
58. Korovin, V., Novitskaya, L. & Kurnosov, G. Structural abnormalities of the stem in woody plants. *Mosc. State For. Univ. Mosc.* 258 (2003).

Acknowledgements

We thank Dr. Anna Igolkina for the valuable discussion about the role of transposable elements in gene expression and trait manifestation.

Author contributions

E.P., A.Z., L.V. conceptualized the study and designed experiments. E.P., A.Z., and V.V. administered the project. L.V. and A.Z. collected plant material. D.K., E.S. and V.V. performed DNA extraction, library preparation, and high-throughput sequencing. R.G., K.L., D.K. and E.G. performed data analysis and visualization. D.K. and V.N. performed marker verification. E.P., R.G., E.G. and L.V. wrote the main manuscript text. All authors reviewed the manuscript.

Funding

The research was supported by the Russian Science Foundation Grant No. 22-16-00096, <https://rscf.ru/en/project/22-16-00096/>.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-55404-y>.

Correspondence and requests for materials should be addressed to E.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024