Conservation Genetics Resources
https://doi.org/10.1007/s12686-022-01276-1

TECHNICAL NOTE

A novel and diverse set of SNP markers for rangewide genetic studies in *Picea abies*

Malte Mader[1] · Céline Blanc-Jolivet[1] · Birgit Kersten[1] · Heike Liesebach[1] · Bernd Degen[1]

Received: 3 November 2021 / Accepted: 29 April 2022

Abstract

We used Double Digest Restriction site associated DNA sequencing (ddRADseq), exome sequencing (exome-seq) and targeted genotyping by sequencing (GBS) to develop new geographically informative nuclear SNP markers in *Picea abies*. This set of 518 loci consists of 397 loci specifically designed for the geographic differentiation of populations and 121 loci of adaptive markers for drought stress which all were identified from 26 samples in 23 populations distributed over Central Europe. This set of novel markers represents a valuable basis to study the geographic population structure and genetic differentiation of *Picea abies* in its natural distribution range as well as outside of its native range with a focus on Central Europe.

Keywords *Picea abies* · Single nucleotide polymorphism · Targeted genotyping by sequencing · Geographic origin

Norway spruce (*Picea abies* (L.) H.Karst) is one of the most economically and ecologically important timber species in Europe with a distribution range from Central Europe to Northern Europe and western Russia (Caudullo et al. 2016). In Central Europe, a significant part of the present days distribution of Norway Spruce is outside the natural species range and has been created by artificial regeneration using distant seed sources (Jansen et al. 2017).

Several studies have investigated the geographic population structure of the natural species distribution in Europe with mitochondrial markers (Tollefsrud et al. 2008) and nuclear SNP markers (Tollefsrud et al. 2009; Dering et al. 2012; Scalfi et al. 2015). Chen et al. (2019) also studied the population structure and transfer of forest reproduction material with a focus on Northern Europe based on markers from exome sequencing. Additionally, the studies from Azaiez et al. (2018) and Depardieu et al. (2021) identified adaptive variation in conifers with respect to population genomics and drought.

The genetic assignment of seed origin with a broad set of informative genetic markers is key to distinguish autochthonous from planted populations (Blanc-Jolivet & Liesebach 2015). With our new set of SNP markers, we want to fill this gap with focus on Central Europe.

Based on ddRADseq (Peterson et al. 2012) a screening of putative nuclear SNPs separating populations by geographic origin on 26 samples from 23 populations (Table 1) was performed. DNA was extracted according to (Dumolin et al. 1995). Sequencing was done by Floragenex Inc. One of the 26 samples was used for an assembly with the software velvet (Zerbino & Birney 2008) (minimum contig length: 200 bp, minimum contig coverage: 6×, expected contig length: 600 bp) resulting in 112,422 contigs with a N50 contig length of 160 bp and with a total assembly length of about 20 Mbp. Read mappings against the contigs were done with bowtie (Langmead et al. 2009). A screening for SNPs with samtools (Li 2011) identified a set of 45,044 SNPs, which was filtered by selecting all variants without neighboring SNPs within 50 bp of flanking sequence. We further restricted the selected loci with calling rates > 90% and minor allele frequencies higher than 1%. In order to select loci potentially showing geographical structure, discriminant analysis of principal components (DAPC) was conducted to detect the optimal number of genetic groups. Loci with highest allele contribution were selected [function var.contr in R package Adegenet, (Jombart 2008)]. Additionally, we grouped the data per country and selected the loci with the highest average differentiation [$D_j > 0.2$, (Gregorius and Roberds 1986)] using GDA_NT 2021 (Degen 2021). Finally we selected loci interesting for population genetics purposes,

✉ Malte Mader
malte.mader@thuenen.de

[1] Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Grosshansdorf, Germany

Published online: 30 May 2022

Table 1 26 samples used for the genetic screening with ddRADseq and the development of drought stress markers. Extreme drought stress behavior was evaluated by analyzing tree ring patterns of a total of 628 trees from a provenance trail

ID	Locality	Country[a]	Longitude	Latitude	Drought
PCABI_440	Reutte	AT	10.72	47.48	Resistant
PCABI_441	Thiersee	AT	12.08	47.58	Resistant
PCABI_442	Knittelfeld	AT	14.82	47.2	Resistant
PCABI_443	Lessach	AT	13.8	47.23	Senstitive
PCABI_444	Zeltschach	AT	14.45	46.95	Resistant
PCABI_445	Eisenkappel	AT	14.62	46.48	Resistant
PCABI_446	Vorau	AT	15.88	47.4	Senstitive
PCABI_447	Crespato	CH	8.63	46.52	Resistant
PCABI_448	Crespato	CH	8.63	46.52	Resistant
PCABI_449	Crespato	CH	8.63	46.52	Resistant
PCABI_450	Cavagnago	CH	8.88	46.42	Senstitive
PCABI_451	Luhacovice	CZ	17.75	49.17	Senstitive
PCABI_452	Nove Hrady	CZ	14.7	48.67	Senstitive
PCABI_453	Frantiskovy	CZ	12.33	50.13	Resistant
PCABI_454	Ruzomberok	SK	19.3	49.08	Senstitive
PCABI_455	Cervena	CZ	20.2	48.85	Senstitive
PCABI_456	Cervena	CZ	20.2	48.85	Resistant
PCABI_457	Hrabusice	SK	20.18	48.98	Resistant
PCABI_458	Koenigsbronn	DE	10.13	48.73	Senstitive
PCABI_459	Mengen	DE	9.32	48.03	Senstitive
PCABI_460	Rungstock	DE	13.32	50.65	Senstitive
PCABI_461	Rebenstein	DE	13.2	49.05	Senstitive
PCABI_462	Tegernsee	DE	11.78	47.72	Senstitive
PCABI_463	Buerkkszent	HU	20.63	48.07	Resistant
PCABI_464	Val di Fiem	IT	11.47	46.28	Resistant
PCABI_465	Cosna	RO	25.17	47.33	Senstitive

[a]Two-letter country code according to ISO-3166-1 ALPHA-2

showing high average intra-country polymorphism (within population gene diversity, $H_s > 0.2$) but without excess of heterozygosity (fixation index, $F_{is} > 0$) using the basic.stats function of the R package hierfstat (Goudet 2005). Altogether, this resulted in a final set of 397 variants (Table S2).

A set of adaptive SNPs for drought stress has also been delimited as we expect that these SNPs could be used to analyze geographic population structure if clinal adaptation occurred (Li et al. 2019). These markers have been developed based on exome-seq of candidate genes identified by RNA-Seq (sequencing done by GATC Biotech AG) in a drought stress experiment (six ramets of a clone originated from tissue culture, three plants as control and drought stress each during flushing period). Read mappings were done with the STAR aligner (Dobin et al. 2013), reads were counted with htseq-count (Anders et al. 2015) and differential gene expression analysis was done with Deseq2 using the Wald test (Love et al. 2014). Differentially expressed genes ($|\log_2$fold change$| > 2$, FDR adjusted p-value < 0.05) assigned to significantly enriched drought stress-related Gene Ontology terms [gene set enrichment analysis with Cytoscape using the plugin Bingo Shannon et al. 2003; Maere et al. 2005)], were consequently examined with exome-seq of 26 phenotyped individuals (Table 1) (exome-seq done by GATC Biotech AG). All original 88,478 SNPs were based on the reference genome of *Picea abies* (Nystedt et al. 2013) (gene only scaffolds). SNPs were filtered and selected using SAS statistics software (9.4 TS Level 1M5), looking for significant differences in allele or genotype frequencies between the resistant and sensitive groups. This step also removed all SNPs with near-complete heterozygous genotypes that were likely to be paralogous SNPs. For genotyping in a first run with 96 samples (Tables 1 and S1), we used targeted GBS applying single primer enrichment technology (Barchi et al. 2019; Scaglione et al. 2019). Probe design for the selected 777 potentially adaptive SNPs and sequencing for 500 SNPs with most specific probes were done by LGC Genomics GmbH (service SeqSNP). 319 SNPs fulfilled the conditions with calling rates > 90% and minor allele frequencies higher than 1% and could be recommended for further use. 121 of these SNPs have been added to the preselected set originating from ddRADseq (Table S2).

In total, 518 loci (397 loci from ddRADseq and 121 loci from exome-seq) represent a solid basis for further population genetic applications allowing to explore geographic population structure and adaptive genetic differentiation in the natural distribution range as well as outside of the native range of *Picea abies*. For further studies all SNPs, including flanking sequences (Table S3), as well as functional annotation of the adaptive SNPs are provided (Table S4).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-022-01276-1.

Acknowledgements We thank Stefan Jencsik, Markus Hartmann, Vivian Kuhlenkamp, Marie-Fee August, Petra Hoffmann and Susanne Hoppe for technical assistance in the field and in the laboratory. We thank Dietrich Ewald who kindly provided the 6 ramets of a clone for the drought stress experiment.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the grants WKF-WF04-22WC4111 01 and 22WB40 8301 (German Federal Ministry of Food and Agriculture & German Federal German Ministry for the Environment, Nature Conservation and Nuclear Safety).

Data availability All reference contigs used in the ddRADseq analysis are deposited at OSF (Open Science Framework; http://osf.io) within the following project: https://osf.io/yspgv/ (https://doi.org/10.17605/OSF.IO/YSPGV).

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638

Azaiez A, Pavy N, Gérardi S, Laroche J, Boyle B, Gagnon F, Mottet M-J, Beaulieu J, Bousquet J (2018) A catalog of annotated high-confidence SNPs from exome capture and sequencing reveals highly polymorphic genes in Norway spruce (*Picea abies*). BMC Genomics 19:942. https://doi.org/10.1186/s12864-018-5247-z

Barchi L, Acquadro A, Alonso D, Aprea G, Bassolino L, Demurtas O, Ferrante P, Gramazio P, Mini P, Fortis E, Scaglione D, Toppino L, Vilanova S, Diez MJ, Rotino GL, Lanteri S, Prohens J, Giuliano G (2019) Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. Front Plant Sci. https://doi.org/10.3389/fpls.2019.01005

Blanc-Jolivet C, Liesebach M (2015) Tracing the origin and species identity of *Quercus robur* and *Quercus petraea* in Europe: a review. Silvae Genet 64:182–193. https://doi.org/10.1515/sg-2015-0017

Bolte A, Sanders T, Natkhin M, Czajkowski T, Chakraborty T, Liesebach H, Kersten B, Mader M, Liesebach M, Lenz C, Lautner S, Löffler S, Kätzel R (2021) Coming from dry regions Norway spruce seedlings suffer less under drought. Project Brief. Thünen Institute, vol 16, pp 1–2. https://doi.org/10.3220/PB1623066406000.

Caudullo G, Tinner W, de Rigo D (2016) *Picea abies* in Europe: distribution, habitat, usage and threats. In: European Atlas of forest tree species. Publications Office of the European Union, Luxembourg

Chen J, Li L, Milesi P, Jansson G, Berlin M, Karlsson B, Aleksic J, Vendramin GG, Lascoux M (2019) Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. Evol Appl 12:1539–1551. https://doi.org/10.1111/eva.12801

Degen B (2021) GDA_NT 2021—genetic data analysis and numerical tests. https://www.thuenen.de/en/fg/software/gda-nt-2021/. Accessed 21 Apr 2022

Depardieu C, Gérardi S, Nadeau S, Parent GJ, Mackay J, Lenz P, Lamothe M, Girardin MP, Bousquet J, Isabel N (2021) Connecting tree-ring phenotypes, genetic associations and transcriptomics to decipher the genomic architecture of drought adaptation in a widespread conifer. Mol Ecol 30:3898–3917. https://doi.org/10.1111/mec.15846

Dering M, Misiorny A, Lewandowski A, Korczyk A (2012) Genetic and historical studies on the origin of Norway spruce in Białowieża Primeval Forest in Poland. Eur J for Res 131:381–387. https://doi.org/10.1007/s10342-011-0510-8

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor Appl Genet 91:1253–1256. https://doi.org/10.1007/BF00220937

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Gregorius HR, Roberds JH (1986) Measurement of genetic differentiation among subpopulations. Theor Appl Genet 71:826–834. https://doi.org/10.1007/BF00276425

Jansen S, Konrad H, Geburek T (2017) The extent of historic translocation of Norway spruce forest reproductive material in Europe. Ann for Sci 74:56. https://doi.org/10.1007/s13595-017-0644-z

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li LL, Chen J, Lascoux M (2019) Clinal variation in growth cessation and FTL2 expression in Siberian spruce. Tree Genet Genomes. https://doi.org/10.1007/s11295-019-1389-7

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. https://doi.org/10.1186/s13059-014-0550-8

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448–3449. https://doi.org/10.1093/bioinformatics/bti551

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theissen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. https://doi.org/10.1038/nature12211

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Scaglione D, Pinosio S, Marroni F, Di Centa E, Fornasiero A, Magris G, Scalabrin S, Cattonaro F, Taylor G, Morgante M (2019) Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. Ann Bot Lond 124:543–551. https://doi.org/10.1093/aob/mcz054

Scalfi M, Mosca E, Di Pierro EA, Troggio M, Vendramin GG, Sperisen C, La Porta N, Neale DB (2015) Micro- and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway spruce. PLoS ONE 9:e115499. https://doi.org/10.1371/journal.pone.0115499

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

Tollefsrud MM, Kissling R, Gugerli F, Johnsen O, Skroppa T, Cheddadi R, van Der Knaap WO, Latalowa M, Terhurne-Berson R, Litt T, Geburek T, Brochmann C, Sperisen C (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. Mol Ecol 17:4134–4150. https://doi.org/10.1111/j.1365-294X.2008.03893.x

Tollefsrud MM, Sonstebo JH, Brochmann C, Johnsen O, Skroppa T, Vendramin GG (2009) Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. Heredity 102:549–562. https://doi.org/10.1038/hdy.2009.16

Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107