Conservation Genetics Resources
https://doi.org/10.1007/s12686-021-01207-6

TECHNICAL NOTE

Applying targeted genotyping by sequencing with a new set of nuclear and plastid SNP and indel loci for *Quercus robur* and *Quercus petraea*

Bernd Degen[1] · Celine Blanc-Jolivet[1] · Svetlana Bakhtina[2] · Ruslan Ianbaev[2] · Yulai Yanbaev[2] · Malte Mader[1] · Sabine Nürnberg[3] · Hilke Schröder[1]

Received: 12 October 2020 / Accepted: 16 April 2021

Abstract

We used Double Digest Restriction site associated DNA sequencing (ddRAD) and Miseq to develop new geographically informative nuclear and plastid SNP and indel loci in *Quercus robur and Q. petraea*. Genotypes derived from sequence data of 95 individuals and two pools of 20 individuals each of *Q. robur* and *Q. mongolica* covering the distribution range of the species, were analysed to select geographically informative and polymorphic loci within Germany and Russia. We successfully screened a selected set of 431 nuclear single nucleotide polymorphism (nSNP), six nuclear Indel, six mitochondrial single nucleotide polymorphism (mtSNP) and ten chloroplast single nucleotide polymorphism (cpSNP) loci with a SeqSNP genotyping platform on 100 individuals *Quercus petraea* from 10 locations in Germany, 100 individuals *Quercus robur* from ten locations in Germany and 100 individuals *Quercus robur* from ten locations in Russia. The newly developed loci are useful for species identification and genetic studies on the genetic diversity and genetic differentiation of *Quercus robur* and *Quercus petraea* in Europe.

Keywords *Quercus robur* · *Quercus petraea* · Single nucleotide polymorphism · Indel · Targeted genotyping by sequencing

In face of climate change both species *Quercus robur* and *Quercus petraea* are considered as alternatives compared to more drought sensitive tree species in Europe (Albert et al. 2018). The selection of the right oak provenances is essential for the re-forestation programs (Wilkinson et al. 2017; George et al. 2020). Gene markers and especially large sets of SNPs play an important role to distinguish and identify different tree provenances (Blanc-Jolivet et al. 2018; Pettenkofer et al. 2020). In frame of a Russian- German cooperation project, we are developing a suitable set of SNPs and Indels for this purpose. Here we present an enlarged set of markers compared to our first study (Blanc-Jolivet et al. 2020).

✉ Bernd Degen
bernd.degen@thuenen.de

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

[2] Bashkir State Agrarian University, 50-letiya Oktyabrya str.-34, 450001 Ufa, Russia

[3] LGC, Ostendstrasse 25, 12459 Berlin, Germany

For SNP and Indel discovery, we used leaf or cambium material from 95 *Q. robur* and *Q. petraea* trees originating from all Europe including Ukraine and Russia for the nuclear SNPs and additional 40 individuals from Europe (*Q. robur*) and Far East Russia (*Q. mongolica*) (Schroeder et al. 2016) to check for plastid SNPs (Suppl. 1). For the selection of nuclear SNPs derived from Double Digest Restriction site associated DNA sequencing (ddRAD) (Peterson et al. 2012), we used the same samples and data described in Blanc-Jolivet et al. (2020). From the 3648 loci filtered for their ability to be included in a design (50 bp flanking length around the target SNP and maximum two SNPs in the flanking regions), 484 nuclear loci were selected. Discriminant analysis was conducted grouping the samples per species, per country within species, and state within Germany. Loci with the highest contributions were identified according to each grouping strategy. Further, samples from Germany and Russia were analysed separately to select loci with both high expected heterozygosity and positive fixation index (F_{is}). Combining the best loci for each grouping strategy allows the development of a multipurpose set of loci, which may distinguish between *Q. robur* and *Q. petraea*, show

Published online: 21 April 2021

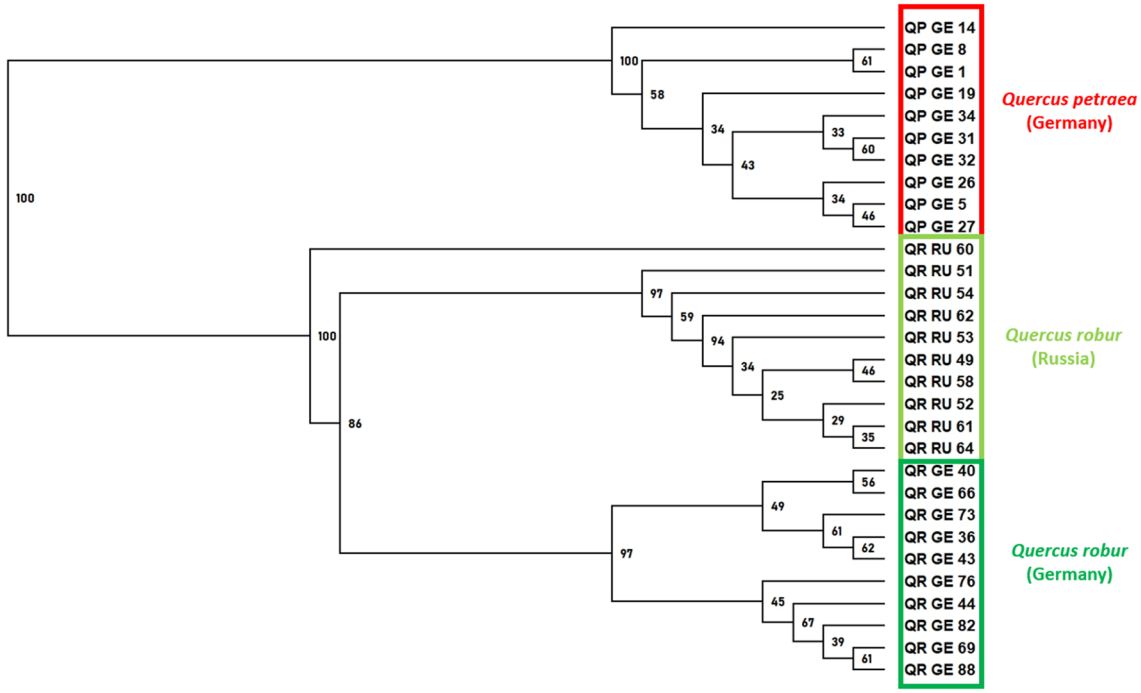

Fig. 1 UPGM-cluster analysis based on matrix of genetic distances (Gregorius 1984) among the 437 nuclear SNPs/indels

differentiation among countries and have enough intrapopulation genetic diversity for population genetics purposes. We used the packages vcfR, poppR, adegenet and hierfstat in R 3.6.0 to conduct the analysis (Goudet 2005; Kamvar et al. 2014; Knaus and Grünwald 2017).

Additionally, to the SNPs produced by the above described ddRAD, we used data from a previously performed MiSeq analysis with a *Q. mongolica* individual as a reference and two pooled DNA samples that included 20 individual specimens each of European *Q. robur* and Asian *Q. mongolica*. The *Q. robur* specimens were sampled from ten geographically-widespread populations in Europe and *Q. mongolica* specimens were sampled from 11 geographically-widespread populations in Far East Russia, China and South Korea (bold in Suppl. 1). The MiSeq analysis, assembly, variation detection and accession numbers for all the data of this analysis is in detail described in Schroeder et al. (2016). Though, originally this data was produced to discriminate within populations of *Q. robur* or *Q. mongolica*, respectively, we finally chose ten chloroplast and six mitochondrial SNPs (Suppl. 3) from this previous study because these SNPs turned out to be also helpful for discriminating *Q. robur* and *Q. petraea*.

From a total of 559 loci (518 nuclear loci, 34 chloroplast loci, seven mitochondrial loci), a set of 479 loci could be designed for targeted genotyping by sequencing (SeqSNP assay). SeqSNP is a targeted genotyping by sequencing (targeted GBS) service, which allows for genotyping of SNPs and small insertions/deletions using a single primer enrichment technology (Anonymous 2019).

We choose to test our newly developed markers on 100 *Q. petraea* from 10 locations in Germany and 200 *Q. robur* from 10 locations in Germany and ten locations in Russia (Suppl. 2). The locations were selected from different regions of the natural distribution range in the countries. All samples were run on Illumina NextSeq 500/550 platforms at LGC. We estimated for each locus the percentage of amplification, observed heterozygosity (H_o), effective number of alleles (A_e), fixation indices (F_{is}, F_{st}) according to Weir and Cockerham (1984); Gregorius (1987) and average differentiation (delta) among sampling locations (Gregorius 1987). A final set of 453 loci (437 nuclear SNPs/indels + 10 chloroplast SNPs + six mitochondrial SNPs) was selected from the screened 479 loci. (Suppl. 3). The criteria for the final selection were polymorphism, an amplification rate above 85% and average F_{IS}-values for the nuclear markers between − 0.3 and 0.3 (Suppl. 4).

We computed genetic distances (Gregorius 1984) among the allele frequencies at the 437 nuclear SNPs/indels of the individuals at the 30 locations and entered this data into an UPGM-cluster analysis (Fig. 1) using the program PAST 4.3 (Hammer et al. 2001). The dendrogram showed that the developed new SNP and Indel markers for *Q. robur* and *Q. petraea* were useful to distinguish between species and populations at the European level.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-021-01207-6.

Acknowledgements The study was co-financed by grant WKF-WF04-22WC4111 01 (German Federal Ministry of Food and Agriculture & German Federal German Ministry for the Environment, Nature Conservation and Nuclear Safety) and for the study on the Russian pedunculate oak samples by the Grant No 19-16-00084 from the Russian Science Foundation (Project members: S. Bakhtina, R. Ianbaev, Y. Yanbaev and B. Degen). We are grateful to Stefan Jencsik, Laura Schulz and Vivian Kuhlenkamp for sampling in provenance trial and technical assistance with laboratory work. Further, we are very thankful for the helpful comments of three reviewers on a former version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Albert M, Nagel RV, Sutmoller J, Schmidt M (2018) Quantifying the effect of persistent dryer climates on forest productivity and implications for forest planning: a case study in northern Germany. For Ecosyst 5:21

Anonymous (2019) SeqSNP targeted GBS as alternative for array genotyping in routine breeding programs. LGC Biosearch Technologies. https://biosearch-cdn.azureedge.net/assetsv6/seqsnp-tgbs-alternative-genotyping-routine-breeding-programs.pdf.

Blanc-Jolivet C, Yanbaev Y, Kersten B, Degen B (2018) A set of SNP markers for timber tracking of *Larix* spp. in Europe and Russia. Forestry 91:614–628

Blanc-Jolivet C, Bakhtina S, Yanbaev R, Yanbaev Y, Mader M, Guichoux E, Degen B (2020) Development of new SNPs loci on *Quercus robur* and *Quercus petraea* for genetic studies covering the whole species' distribution range. Conserv Genet Resour 12(4):597–600

George JP, Theroux-Rancourt G, Rungwattana K, Scheffknecht S, Momirovic N, Neuhauser L, Weissenbacher L, Watzinger A, Hietz P (2020) Assessing adaptive and plastic responses in growth and functional traits in a 10-year-old common garden experiment with pedunculate oak (*Quercus robur* L.) suggests that directional selection can drive climatic adaptation. Evol Appl 13:2422–2438

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186

Gregorius HR (1984) A unique genetic distance. Biom J 26:13–18

Gregorius HR (1987) The relationship between the concepts of genetic diversity and differentiation. Theor Appl Genet 74:397–401

Hammer Ø, Harper DA, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:9

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17:44–53

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

Pettenkofer T, Finkeldey R, Muller M, Krutovsky KV, Vornam B, Leinemann L, Gailing O (2020) Development of novel *Quercus rubra* chloroplast genome CAPS markers for haplotype identification. Silvae Genet 69:78–85

Schroeder H, Cronn R, Yanbaev Y, Jennings T, Mader M, Degen B, Kersten B (2016) Development of molecular markers for determining continental origin of wood from White oaks (*Quercus* L. sect. Quercus). PLoS ONE 11:e0158221

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38:1358–1370

Wilkinson M, Eaton EL, Morison JIL (2017) Variation in the date of budburst in *Quercus robur* and *Q. petraea* across a range of provenances grown in Southern England. Eur J For Res 136:1–12