Development of new SNPs loci on *Quercus robur* and *Quercus petraea* for genetic studies covering the whole species' distribution range

Celine Blanc-Jolivet¹ · Svetlana Bakhtina² · Ruslan Yanbaev² · Yulai Yanbaev² · Malte Mader¹ · Erwan Guichoux³ · Bernd Degen¹

Received: 17 February 2020 / Accepted: 28 February 2020
© The Author(s) 2021, corrected publication 2021

Abstract

We used double digest restriction site associated DNA sequencing (ddRAD) to develop new geographically informative nuclear SNP loci in *Quercus robur* and *Quercus petraea*. Genotypes derived from sequence data of 95 individuals covering the distribution range of the species were analysed to select geographically informative and polymorphic loci within Russia and Germany. We successfully screened a selected set of 119 loci on a MassARRAY® iPLEX™ platform on 190 individuals from 19 locations in Russia. The newly developed loci will be useful for genetic studies over the whole distribution range of both species.

Keywords *Quercus robur* · *Quercus petraea* · Single nucleotide polymorphism · MassARRAY

Quercus robur (L.) and *Quercus petraea* (Matt.) are very common tree species in many forests in Europe. Their distribution range from Scandinavia to the Iberian Peninsula, and from Great Britain towards Eastern and South-Eastern Europe (San-Miguel-Ayanz et al. 2016). The distribution edge of *Quercus robur* goes a few hundred kilometres more to the East up to the South-Ural region in Russia. Both species are economically significant and provide high quality hardwood for timber construction, furniture and barrels.

Quercus robur and *Q. petraea* have been genetically very well studied over their West-European distribution range, but there are still some gaps especially to the East (Degen et al. 2019). As gene markers, chloroplastic RFLPs, microsatellites and recently SNPs have been applied to study genetic differentiation and population genetic processes like gene

flow, mating system and the impact of post glacial re-colonisation and human seed transfer (Konig et al. 2002; Petit et al. 2002; Buschbom et al. 2011; Gerber et al. 2014). Although large sets of SNPs including thousand to millions of SNPs have been developed for the two oak species (Leroy et al. 2019; Lepoittevin et al. 2015), there is still a need for smaller sets of highly geographically informative SNPs that can be cost-effectively genotyped.

For SNP discovery, we used leaf or cambium material from 95 *Q. robur* and *Q. petraea* trees originating from all Europe, Ukraine and Russia (Table 1). The sampling size was more intensive in Germany to catch within-country variation. DNA was extracted according to Dumolin et al. (1995). Double Digest Restriction site associated DNA sequencing (ddRAD) (Peterson et al. 2012) was conducted on all samples to detect SNPs in the nuclear genome (Florigenex, Portland, USA). Among the 26,074 putative SNP loci obtained by ddRADseq, only those with a minimum flanking region of 50 bp around the SNP and a maximum of two neighbour SNPs were selected for further analysis (3648 loci). Further cleaning of the data included the removal of loci with more than 10% missingness and a minor allele frequency lower than 1%. Data was grouped by species, country, and state within Germany to conduct discriminant analysis and to detect the loci with the highest contribution. This allowed a selection of SNP loci with a geographical signal. We additionally looked among samples from Germany

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12686-020-01141-z>) contains supplementary material, which is available to authorized users.

✉ Celine Blanc-Jolivet
celine.blanc-jolivet@thuenen.de

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

² Bashkirian Agrarian University, 50-letiya Oktyabrya Str.-34, Ufa, Russia 450001

³ INRAE, Univ. Bordeaux, BIOGECO, 33610 Cestas, France

Table 1 *Quercus robur* and *Q. petraea* samples used for double digest restriction site associated DNA sequencing (ddRAD)

Species	Country	Region	Latitude	Longitude	N
<i>Quercus petraea</i>	Austria	BL, Klostermarienber	47.4	16.6	3
<i>Quercus petraea</i>	Austria	NÖ, Mistelbach	48.6	16.6	3
<i>Quercus petraea</i>	Belgium	Fagnes	50.1	4.4	2
<i>Quercus petraea</i>	Belgium	Hainaut Gent	51	4.2	3
<i>Quercus petraea</i>	Denmark	Arhus, Löndal Naes	56.1	9.6	3
<i>Quercus petraea</i>	France	Allier ONF Moulin	45.6	2.9	3
<i>Quercus petraea</i>	France	Vienne ONF Poitiers	46.6	0.2	3
<i>Quercus petraea</i>	Germany	BB, Müncheberg	52.6	14.1	3
<i>Quercus petraea</i>	Germany	BY, Ebrach	49.9	10.5	2
<i>Quercus petraea</i>	Germany	BY, Riedenburg	48.9	11.7	2
<i>Quercus petraea</i>	Germany	HE, Bad Harsfeld	50.9	9.7	2
<i>Quercus petraea</i>	Germany	HE, Wolfgang	50.2	9.1	2
<i>Quercus petraea</i>	Germany	NI, Göhrde	53.1	10.9	1
<i>Quercus petraea</i>	Germany	NI, Lappwald	52.2	10.9	1
<i>Quercus petraea</i>	Germany	NI, Lüß	53.3	10.3	2
<i>Quercus petraea</i>	Germany	RP, Cochem	50.1	7.1	1
<i>Quercus petraea</i>	Germany	RP, Elmstein_Nord	49.4	8	2
<i>Quercus petraea</i>	Germany	RP, Johanniskreuz	49.3	7.8	1
<i>Quercus petraea</i>	Great Britain	Glouc., Sutton Bottom	51.8	2.5	3
<i>Quercus petraea</i>	Poland	Kalisz Syców	51.2	17.9	3
<i>Quercus petraea</i>	Turkey	Bolu	40.9	31.7	3
<i>Quercus robur</i>	Austria		–	–	3
<i>Quercus robur</i>	Belarus		53.6	24.7	3
<i>Quercus robur</i>	Finland		60.0	23.5	3
<i>Quercus robur</i>	France		47.7	1.9	3
<i>Quercus robur</i>	Germany	Bavaria	–	–	3
<i>Quercus robur</i>	Germany	Freiburg	–	–	3
<i>Quercus robur</i>	Germany	Mecklenburg-Vorpommern	–	–	3
<i>Quercus robur</i>	Germany	North Rhine-Westphalia	51.8	7.9	3
<i>Quercus robur</i>	Germany	Rhineland-Palatinate	–	–	3
<i>Quercus robur</i>	Germany	Schleswig-Holstein	–	–	3
<i>Quercus robur</i>	Germany	Thuringia	–	–	3
<i>Quercus robur</i>	Hungary		48.4	20.6	3
<i>Quercus robur</i>	Poland		51.5	22.8	2
<i>Quercus robur</i>	Russia		54.4	57.1	3
<i>Quercus robur</i>	Slovakia		–	–	3
<i>Quercus robur</i>	Ukraine		49.1	38.0	3

Most of the samples from Europe were collected from provenance trials located in Germany. Geographic coordinates are given in decimal degrees

and Russia to select loci with both high expected heterozygosity and positive F_{is} to avoid parapatric loci. All analyses were conducted in R 3.6.0 using, among others, the packages vcfR, poppr, adegenet and hierfstat. A final selection of 168 loci was used to design four MassARRAY® iPLEX™ multiplexes (Assay Design Suite v2.0 [Agena Bioscience™, San Diego, USA]), which included a total of 130 loci (Supplementary Material S1).

We choose to test our newly developed markers on 190 *Quercus robur* from 19 locations in its Russian

distribution range (Table 2). All samples were run on a MassARRAY® iPLEX™ platform (Agena Bioscience™, San Diego, USA) using the iPLEX™ GOLD chemistry. Genotyping was conducted with Typer Viewer v.4.0.24.71 (Agena Bioscience™, San Diego, USA). We estimated for each locus the percentage of amplification, observed heterozygosity (H_o), within population gene diversity (H_s) (Nei 1987), F_{is} , F_{st} (Weir and Cockerham 1984) and average differentiation among sampling locations (Gregorius 1987). Significance levels for F_{is} and F_{st} were tested with

Table 2 Spatial location of the 19 Russian populations included in the screening of 130 SNP loci

Population	Latitude	Longitude	N	H _o	H _e	F _{is}
Boguchar	49.596	40.283	10	0.26	0.245	−0.039
Ilisch	55.363	54.478	10	0.246	0.247	0.006
Kamyshin	50.124	45.195	10	0.279	0.272	−0.027
Kir_1	57.636	50.156	10	0.237	0.247	0.06
Kst_2	58.482	41.550	10	0.267	0.231	−0.122
Ling_4	59.265	31.392	10	0.279	0.26	−0.064
Lipetsk	53.184	39.239	10	0.281	0.262	−0.063
Mar	56.177	47.933	10	0.253	0.228	−0.096
Orel	52.582	35.270	10	0.291	0.272	−0.067
Orenburg_1	52.370	54.918	10	0.284	0.259	−0.076
Penza	53.023	43.670	10	0.277	0.256	−0.065
Pskov	56.194	29.301	10	0.289	0.265	−0.09
Samara_2	54.125	50.346	10	0.289	0.262	−0.085
Tolyatti	53.305	49.534	10	0.264	0.219	−0.166
Tvr_2	56.847	35.916	10	0.259	0.256	−0.01
Udm_2	56.637	53.745	10	0.262	0.256	−0.028
Vld_2	55.516	41.941	10	0.27	0.256	−0.035
Vlg_1	59.228	39.913	10	0.27	0.261	−0.034
Volgograd	48.458	43.159	10	0.281	0.253	−0.101

N is the sampling size, H_o and H_e the observed and expected heterozygosities (Hartl and Clark 1997), F_{is} the intra-population fixation index

10,000 randomizations (Supplementary Material S2). Analyses were conducted with GDA_NT (Degen, unpublished) and Fstat (Goudet 1995). A total of 119 markers was usable. Among those, four loci showed an amplification lower than 80% and 24 loci were not polymorphic in the screened Russian samples. Additionally, nine loci significantly deviated from HWE in the tested samples. Differentiation was low ($F_{st}=0.03$) but significant (Goudet et al. 1996).

We successfully developed new SNP markers for *Q. robur* and *Q. petraea*, which will be useful for population genetic studies at the European level. Additional screening will be needed on samples from other regions, to address whether a single set of SNP markers will be sufficient to cover the whole distribution range of these species and be useful to track the origin of forest reproductive material.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12686-020-01141-z>) contains supplementary material, which is available to authorized users.

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. Yanbaev, Y. Yanbaev and B. Degen). Genotyping was performed at the Genome

Transcriptome Facility of Bordeaux (Grants from ANR-10-EQPX-16), with the help of Marie Massot. We are grateful to Stefan Jencsik for his work with sampling in provenance trial and with laboratory work.

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/>.

References

- Buschbom J, Yanbaev Y, Degen B (2011) Efficient long-distance gene flow into an isolated relict oak stand. *J Hered* 102:464–472
- Degen B, Yanbaev R, Yanbaev Y (2019) Genetic differentiation of *Quercus robur* in the South-Ural. *Silvae Genet* 68:111–115
- 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
- Gerber S, Chadoeuf J, Gugerli F, Lascoux M, Buiteveld J, Cottrell J, Dounavi A, Fineschi S, Forrest LL, Fogelqvist J, Goicoechea PG, Jensen JS, Salvini D, Vendramin GG, Kremer A (2014) High rates of gene flow by pollen and seed in oak populations across Europe. *PLoS ONE* 9:16

- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* 86:485–486
- Goudet J, Raymond M, deMeeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* 144:1933–1940
- Gregorius HR (1987) The relationship between the concepts of genetic diversity and differentiation. *Theor Appl Genet* 74:397–401
- Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer Associates, Sunderland MA
- König AO, Ziegenhagen B, van Dam BC, Csai KL, Coart E, Degen B, Burg K, de Vries SMG, Petit RJ (2002) Chloroplast DNA variation of oaks in western Central Europe and genetic consequences of human influences. *For Ecol Manag* 156:147–166
- Lepointevin C, Bodenes C, Chancerel E, Villate L, Lang T, Lesur I, Boury C, Ehrenmann F, Zelenica D, Boland A, Besse C, Garnier-Gere P, Plomion C, Kremer A (2015) Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. *Mol Ecol Res* 15:1446–1459
- Leroy T, Rougemont Q, Dupouey JL, Bodenes C, Lalanne C, Belser C, Labadie K, Le Provost G, Aury JM, Kremer A, Plomion C (2019) Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytol.* <https://doi.org/10.1111/nph.16039>
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- 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
- Petit RJ et al (2002) Chloroplast DNA variation in European white oaks—phylogeography and patterns of diversity based on data from over 2600 populations. *For Ecol Manag* 156:5–26
- San-Miguel-Ayán J, De Rigo D, Caudullo G, Durrant TH, Mauri A (2016) European atlas of forest tree species. Publications Office of the European Union
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358–1370

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