



A set of nuclear SNP loci derived from single sample double digest RAD and from pool sequencing for large-scale genetic studies in the European beech *Fagus sylvatica*

Celine Blanc-Jolivet¹ · Malte Mader¹ · Heike Liesebach¹ · Birgit Kersten¹ · Bernd Degen¹

Received: 13 October 2021 / Accepted: 22 February 2022 / Published online: 8 March 2022
© The Author(s) 2022

Abstract

The large-scale spatial genetic structure of European beech, *Fagus sylvatica*, has been until now poorly studied. We conducted double digest RAD sequencing (ddRADseq) on 54 beech individuals stemming from 36 provenances to discover spatially informative nuclear SNP loci. In addition, two pools derived from 14 early and 14 late flushing individuals each were sequenced with Illumina HiSeq. From an initial amount of 5,464 loci detected by ddRADseq, we selected 559 informative loci. Further 27 additional loci showing significant allelic differences among early and late flushing individuals could be identified after a genotyping on 95 test individuals. The final selection of 578 loci was submitted to probe design for targeted genotyping by sequencing, which yielded 543 loci. The new set of SNP loci should be, after validation on a larger sample size, useful for large-scale genetic studies in this economically-important species.

Keywords *Fagus sylvatica* · European beech · Single nucleotide polymorphism · Genetic structure

European beech, *Fagus sylvatica*, is one of the most significant forest species in Europe, which naturally occurs in temperate regions, from France to the west of Ukraine, and from northern Spain to southern Sweden (EUFORGEN network, <http://www.euforgen.org/species/fagus-sylvatica/>). The first genetic studies on beech dates back to the 1980s (Mueller-Starck 1985) and the co-dominant isoenzyme markers, followed later by SSRs (Pastorelli et al. 2003), allowed a better knowledge of the mating system and gene flow (Lander et al. 2021), population structure (de Lafontaine et al. 2013) and response to biotic and abiotic stress (Mueller and Gailing 2019) in this species. Nowadays, the availability of high-throughput sequencing technologies facilitates the sequencing of full-genomes as well as the development and screening of single nucleotide polymorphism (SNP loci). In beech, as for many other European forest species, many researchers focused on polymorphism in candidate genes to detect selection processes due to climatic changes, in particular adaptation to drought (Cuervo-Alarcon et al. 2018). However,

only a few large-scale genetic inventories describing the spatial genetic structure along the distribution of beech are available (Magri et al. 2006; Magri 2008; Postolache et al. 2021), although a reference genome is published (Mishra et al. 2018). In this study, we conducted a double digest restriction-site associated DNA sequencing (ddRADseq) on European beech samples covering most of its distribution range and from several regions within Germany, to detect spatially informative polymorphisms. We additionally searched for loci associated with early or late bud burst, as potentially spatially differentiated polymorphisms.

We extracted DNA from fresh leaf or cambium material (Dumolin et al. 1995) for 54 beech trees taken from the provenance trial Schädtebek established in 1995. Our samples covered 36 provenances and 18 countries, focusing on Germany (26 trees) (Table 1). Through this unbalanced sampling, we expected to discover informative SNP loci showing differentiation among stands within Germany, as well as spatially-informative loci within Europe. First, ddRADseq (Peterson et al. 2012) was applied to discover nuclear SNP loci. We used the published beech reference genome (v1.2, Mishra et al. 2018) to map our reads and conduct variant calling (library preparation, ddRADseq and bioinformatic conducted by Floragenex, Portland, USA). A total of 5464 loci passed all quality filters using

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

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

Table 1 Beech provenances and their sampling size used for SNP discovery

Country	Population	Province	Latitude	Longitude	N
Austria	Neuberg-Muerzsteg		47.75	15.47	1
Bulgaria	Ribaritza		42.92	24.27	2
Croatia	Opatija		45.33	14.30	1
Czech Republic	Kladaska		50.03	12.62	2
Denmark	Fyn		55.13	10.88	1
France	F.D De Foueres		48.38	-1.17	1
France	F.D De Lagast		44.15	2.63	1
France	F.D De Verrier		47.20	6.25	1
Germany	Bovenden	NI	51.50	9.83	2
Germany	Buedingen	HE	50.28	9.12	2
Germany	Ebeleben	TH	51.33	10.50	2
Germany	Giengen	BW	48.62	10.25	2
Germany	Gransee	BB	53.00	13.17	3
Germany	Hermeskeil	RP	49.65	6.95	2
Germany	Herrenberg	BW	48.67	9.00	2
Germany	Jesberg	HE	51.03	9.10	2
Germany	Lensahn	SH	54.20	10.75	2
Germany	Oberhaus	NI	51.67	10.83	1
Germany	Schluechtern	HE	50.33	9.67	2
Germany	Spangenberg	HE	51.13	9.67	2
Germany	Tharandt	SN	50.95	13.57	2
Great Britain	Westfield 2002		57.67	-3.42	1
Hungary	Magyaregregy		46.22	18.35	1
Italy	Veneto		46.13	12.22	1
Luxemburg	Heinerscheid		50.08	6.12	1
Netherlands	Elspeet		52.28	5.80	1
Poland	Brzeziny		51.83	19.60	1
Poland	Krynica		49.42	20.90	1
Poland	Ladek Zdroj		50.25	16.83	2
Romania	Beius-Bihor		46.68	22.27	2
Slovakia	Muran		48.75	20.07	1
Slovenia	Postojna		45.75	14.32	1
Spain	Anguiano		42.25	-2.75	1
Spain	Limitationes		42.82	-5.25	2
Sweden	Ryssberget		56.08	14.60	1
Switzerland	Oberwil		47.17	7.45	1

the “stringent” criteria (calling rate > 90% and flanking regions available). However, only 2988 loci were informative (minor allele frequency > 0.01). Discriminant analysis of principal components (DAPC, “dapc” in R-package “Adegenet”, Jombart and Ahmed 2011) was conducted with four putative clusters, which arranged the samples in a west–east spatial pattern (Fig. S1). The 63 loci with the highest variable contributions were selected. The same procedure was applied separately on the 26 samples from

Germany with three clusters and yielded 39 loci. To select loci potentially with high polymorphism and exclude SNPs probably resulting from the merging of paralog loci (no excess of heterozygotes), we grouped the data per country, and the data of the German samples per province. We estimated average H_s (within-population gene diversity), as well as F_{is} and A_r (allelic richness) per locus (“basicstat” in R-package “hierfstat”, Goudet 2005) for groups with at least two individuals. We filtered 325 and 305 loci, respectively for the complete dataset grouped by countries and the German provenances ($F_{is} > 0$ and $H_s > 0.4$, Table S1). A total of 559 unique loci were finally selected, and we visualized the effect of locus selection on the differentiation among provenances (PCA, “dudi.pca” in R-package “ade4”, Dray and Dufour 2007; Fig. S2).

Among the 54 individuals sequenced with ddRADseq, a subset of 14 early and 14 late flushing was separately evaluated. From the 2996 putative informative SNPs detected in this subset of samples, 486 were selected for their genetic differences between the two groups. We additionally conducted an Illumina HiSeq 150 bp Paired-end pooled sequencing (formerly GATC, Konstanz, Germany) of two groups (poolseq, 14 early and 14 late flushing individuals) with 84× coverage each. 6800 top SNPs showed the highest allele frequency difference estimates (Ries et al. 2016) between the two pools. Applying a genome scan technique (Soyk et al. 2017), nine genomic regions were identified as top scaffolds showing an enrichment of top SNPs. A total of 1006 loci were submitted to probe design to select 500 loci for targeted genotyping by sequencing (SeqSNP, LGC Genomics GmbH, Berlin, Germany). Interestingly, ddRAD and poolseq showed sticking differences in probe specificities (Table S2). Genotyping was conducted on 95 test individuals and 27 loci were identified as potentially differentiated among early and late flushing individuals. Surprisingly, eight from these 27 loci were already included in the set of 559 loci identified from the ddRAD data at all 54 individuals (potentially spatially informative and/or high diversity loci) (Table S1). Altogether, we ended with a selection of 578 loci.

From the PCA and DACP analyses, we expect that our set of loci will be useful to disentangle genetic structure over beech’s distribution range. Indeed, the spatial grouping suggested by our data fits the findings of Postolache et al. (2021). A final set of 543 loci could be designed for SeqSNP genotyping and will be used for large genetic inventories.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12686-022-01256-5>.

Acknowledgements We thank Mirko Liesebach for providing material information, and Stefan Jencsik, Vivian Kuhlenskamp, Marie-Fee August, Petra Hoffmann and Susanne Hoppe for technical assistance in the field and in the laboratory.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Grants WKF-WF04-22WC4111 01 and WKF-22WC4092 10 (German Federal Ministry of Food and Agriculture & German Federal German Ministry for the Environment, Nature Conservation and Nuclear Safety).

Data availability The datasets resulting from ddRADseq (in vcf format) under the stringent, standard and relaxed criteria are available at <https://osf.io/35pca/>.

Declarations

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

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

- Cuervo-Alarcon L, Arend M, Muller M, Sperisen C, Finkeldey R, Krutovsky KV (2018) Genetic variation and signatures of natural selection in populations of European beech (*Fagus sylvatica* L.) along precipitation gradients. *Tree Genet Genomes* 14:84. <https://doi.org/10.1007/s11295-018-1297-2>
- de Lafontaine G, Ducouso A, Lefevre S, Magnanou E, Petit RJ (2013) Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. *Mol Ecol* 22:4397–4412. <https://doi.org/10.1111/mec.12403>
- Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *J Stat Softw* 22:1–20. <https://doi.org/10.18637/jss.v022.i04>
- Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor App 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>
- Jombart T, Ahmed I (2011) Adegnet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Lander TA, Klein EK, Roig A, Oddou-Muratorio S (2021) Weak founder effects but significant spatial genetic imprint of recent contraction and expansion of European beech populations. *Heredity* 126:491–504. <https://doi.org/10.1038/s41437-020-00387-5>
- Magri D (2008) Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J Biogeogr* 35:450–463. <https://doi.org/10.1111/j.1365-2699.2007.01803.x>
- Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latałowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit RJ, de Beaulieu JL (2006) A new scenario for the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol* 171(1):199–221. <https://doi.org/10.1111/j.1469-8137.2006.01740.x>
- Mishra B, Gupta DK, Pfenninger M, Hickler T, Langer E, Nam B, Paule J, Sharma R, Ulaszewski B, Warmbier J, Burczyk J, Thines M (2018) A reference genome of the European beech (*Fagus sylvatica* L.). *Gigascience*. <https://doi.org/10.1093/gigascience/gy063>
- Mueller-Starck G (1985) Genetic-differences between tolerant and sensitive beeches (*Fagus sylvatica* L) in an environmentally stressed adult forest stand. *Silvae Genet* 34:241–247
- Müller M, Gailing O (2019) Abiotic genetic adaptation in the Fagaceae. *Plant Biol* 21:783–795. <https://doi.org/10.1111/plb.13008>
- Pastorelli R, Smulders MJM, Van't Westende WPC, Vosman B, Gianini R, Vettori C, Vendramin GG (2003) Characterization of microsatellite markers in *Fagus sylvatica* L. and *Fagus orientalis* Lipsky. *Mol Ecol Notes* 3:76–78. <https://doi.org/10.1046/j.1471-8286.2003.00355.x>
- 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*. <https://doi.org/10.1371/journal.pone.0037135>
- Postolache D, Oddou-Muratorio S, Vajana E, Bagnoli F, Guichoux E, Hampe A, Le Provost G, Lesur I, Popescu F, Scotti I, Piotti A, Vendramin GG (2021) Genetic signatures of divergent selection in European beech (*Fagus sylvatica* L.) are associated with the variation in temperature and precipitation across its distribution range. *Mol Ecol* 30:5029–5047. <https://doi.org/10.1111/mec.16115>
- Ries D, Holtgräwe D, Viehöver P, Weisshaar B (2016) Rapid gene identification in sugar beet using deep sequencing of DNA from phenotypic pools selected from breeding panels. *BMC Genomics* 17:236. <https://doi.org/10.1186/s12864-016-2566-9>
- Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jiménez-Gómez JM, Lippman ZB (2017) Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nat Genet* 49:162. <https://doi.org/10.1038/ng.3733>

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