A Circular Chloroplast Genome of *Fagus sylvatica* Reveals High Conservation between Two Individuals from Germany and One Individual from Poland and an Alternate Direction of the Small Single-Copy Region

Bagdevi Mishra 1,2, Bartosz Ulaszewski 3, Sebastian Ploch 1, Jaroslaw Burczyk 3 and Marco Thines 1,2,4,*

Abstract: Chloroplasts are difficult to assemble because of the presence of large inverted repeats. At the same time, correct assemblies are important, as chloroplast loci are frequently used for biogeography and population genetics studies. In an attempt to elucidate the orientation of the single-copy regions and to find suitable loci for chloroplast single nucleotide polymorphism (SNP)-based studies, circular chloroplast sequences for the ultra-centenary reference individual of European Beech (*Fagus sylvatica*), Bhaga, and an additional Polish individual (named Jamy) was obtained based on hybrid assemblies. The chloroplast genome of Bhaga was 158,458 bp, and that of Jamy was 158,462 bp long. Using long-read mapping on the configuration inferred in this study and the one suggested in a previous study, we found an inverted orientation of the small single-copy region. The chloroplast genome of Bhaga and of the individual from Poland both have only two mismatches as well as three and two indels as compared to the previously published genome, respectively. The low divergence suggests low seed dispersal but high pollen dispersal. However, once chloroplast genomes become available from Pleistocene refugia, where a high degree of variation has been reported, they might prove useful for tracing the migration history of *Fagus sylvatica* in the Holocene.

Keywords: cpDNA; conservation genetics; European Beech; population genetics

1. Introduction

The European Beech is a dominant species in Germany and other parts of Central Europe [1]. It is valued for its wood and provides crucial resources for a number of species that depend on it for survival [2]. Due to climate change, the European beech has recently experienced drought stress in several areas [3,4], which probably marks a transition to oak (*Quercus robur* s.l.) in drier regions of Germany [5]. However, succession from oak- to beech-dominated forest stands is also observed, in particular in north-eastern beech distribution limits [6]. There are several population genetics studies for European Beech available [7–12], and some of them are focused on chloroplast-based genetic diversity [9,13–16]. However, analyses considering SNPs in the complete chloroplast genome are so far lacking, even though such a dataset might help understanding maternal gene flow within local stands and understanding population dynamics and phylogeography in conjunction with past and present climate change [17,18].

Potential single nucleotide polymorphisms (SNPs) within the cpDNA of beech were recently investigated with the aid of RADseq methods [16]; however, the identification of...
reliable SNPs is challenging due to the presence of cpDNA sequences in nuclear genomes of several plants. Therefore, using complete chloroplast genomes is the ultimate way to assess cpDNA diversity within the species, allowing for the detection of all kinds of variation including SNPs, indels, or even structural variants. Recently, a complete chloroplast genome sequence for European beech was published [19], but no evaluation regarding the variability of the complete chloroplast genome was done. Therefore, it was the aim of this study to provide two additional chloroplast sequences from parts of the beech distribution area glaciated during the ice ages, which are known for their low chloroplast haplotype diversity [9,10,18], with the objective to search for highly variable regions suitable for population genetics, as previously done for East Asian beech [20]. For this, the ultra-centenary reference individual Bhaga from the National Park Kellerwald-Edersee (Germany) and an individual from the Jamy Nature Reserve (Poland), named Jamy, were used, and their chloroplast genome sequences were compared to the chloroplast sequence recently published for another German individual [19].

2. Materials and Methods

2.1. Initial Draft Assembly

For assembling the chloroplast of the reference individual of European beech, Bhaga [21], from the National Park Kellerwald-Edersee, genomic short (Illumina) and long (PacBio) reads were used. The PacBio reads were double corrected by applying first an Illumina correction using proovread [22] followed by self-correction using Canu [23]. PacBio reads were blasted against the completely assembled chloroplast of Arabidopsis thaliana [24], Eucalyptus grandis [25], and Fagus sylvatica [19]. Three individual BLAST searches were conducted on the PacBio reads at query length match cut-offs of 50%, 70%, and 90% and with default values for all BLAST parameters. The reads obtained at each cut-off level were used separately to calculate chloroplast assemblies using Canu. All three assemblies resulted in a longest contig with the same length, while several small contigs differed, depending on the dataset used. However, in none of the cases was the longest contig, which represents the chloroplast, circular and thus carefully refined with additional analyses (see below) to obtain a circular chloroplast genome.

The chloroplast genome of the Polish individual (named Jamy) originating from the Jamy Nature Reserve (18°56′07.07″ E, 53°35′9.67″ N) was built with 15,718,780 (PE 150-bp) Illumina reads, using a de novo assembler NOVOPlasty v 4.2 [26] with the sequence with the GenBank accession number AY042393.1 [27] as seed and the Bhaga chloroplast genome as the reference for guiding the assembler in inverted repeat regions. This resulted in a 158,462 bp circularised assembly with an average 481× coverage. The final chloroplast sequences have been deposited under the GenBank accession numbers MW531753 for Bhaga and MW537046 for Jamy.

2.2. Manual Curation

Illumina paired-end reads were used to check the consistency of the coverage of the reference assembly, as a lack of coverage or coverage highly deviating from the average coverage would indicate misassembled or duplicate regions. Subsequently, a self-BLAST of the sequence was done to identify the inverted repeat regions. The assembled sequence was cut at the points of potential misassembly (i.e., regions of lower coverage and the borders to the repeat regions and subsequently re-joined using bridging PacBio reads. Finally, Pacbio reads were aligned over the final sequence using Blasr [28] to verify coverage continuity. At this stage, the sequence was circularised with evidence from long read mapping. The accuracy of the circular chloroplast sequence was further improved by using Illumina paired reads using Pilon [29]. The chloroplast genome of the Polish individual was verified by alignment to the reference genome using MUMmer [30].
2.3. Annotation of Genic Regions

GeSeq ChloroBox [31] was used for annotation and graphical representation. For the annotation, the MPI-MP reference set of chloroplasts of land plants (CDS + rRNA), available from the ChloroBox server, was used. The graphical representation indicating the gene locations, the Large Single Copy region (LSC), the Small Single Copy region (SSC), and the two Inverse Repeat regions (IRA and IRB), was built using OGDraw [31].

2.4. Comparison of Chloroplast Assemblies

For the comparison the two chloroplast sequences reported in this study obtained from Bhaga (Germany) and Jamy (Poland), and the one previously published from another German individual (GenBank accession MK598696) [19], were aligned using Mummer [32] and inspected for the occurrence of SNPs. To verify the inversion found for the small single-copy region as compared to previously published genomes, long-read alignment onto the chloroplast configuration published by [19] and the one inferred in this study was done using Geneious R1 (Biomatters, Auckland, New Zealand).

3. Results

3.1. General Features of the Chloroplast Assembly of Bhaga

The circularised chloroplast genome for the *Fagus sylvatica* reference individual Bhaga was 158,458 bp long. It contains two inverted repeat regions (IRA and IRB in Figure 1) of 25,873 bp each, along with a Small Single-Copy (SSC in Figure 1) region of 19,010 bp and a Large Single-Copy (LSC in Figure 1) region of 87,702 bp. In our assembly, all these four subunits are adjacent to each other and constitute a circular chloroplast. The functional annotation of the chloroplast genomes revealed 88 protein-coding gene locations, out of which 75 are single-exon genes, nine genes have two exons, three genes have three exons, and one gene has six exons. There are eight rRNA gene locations and 40 tRNA locations. Gene locations on both strands are shown in Figure 1.

3.2. Comparison of Chloroplast Assemblies

In the earlier assembly of a *Fagus sylvatica* chloroplast genome from Germany (GenBank accession number MK598696) [19], there is a stretch of eight bases that is not assigned to any of the subunits. Alignment of the chloroplast genome published by Mader et al. [19] and the one from Bhaga revealed that the SSC locations of the previous and our assembly are inversely matching to each other with one gap and one mismatch. The gap is within a polyA stretch in the non-genic region and the mismatch is a C to T transversion which does not change the amino acid in the protein product of the gene ccsA. The inversion found in our study was confirmed with long-read mapping (Figure 2). The other sequence stretches were directly matching to each other with one mismatch and three deletions. The mismatch is a G to T transversion that does not change the amino acid in the protein product of the gene *atpE*. At one place, there is a gap of two bases in a homopolymer of C, which is a non-genic location, and another gap is in a homopolymer of T inside the genic location of *rpl23* gene, but which does not affect an amino acid change.

The chloroplast genome of the Jamy individual is identical in length to the chloroplast genome published by [19] and in comparison to the Bhaga genome showed only three indels located in three intergenic regions: C between *psaC* and *ndhD*; TT between *matK* and tRNA-Gln; and T between *psbK* and *psgfd- bI* genes.
Figure 1. Circular chloroplast of *Fagus sylvatica* (Bhaga) with gene locations on both strands, along with an indication of the locations of the inverted repeat regions and the large and the small single-copy regions.

Figure 2. Confirmation of the chloroplast assembly by aligning illumina-corrected PacBio reads >25 kb onto (A) the chloroplast configuration inferred in this study from Bhaga (Germany) and (B) the configuration previously published [19] from another German individual. Long reads spanning over the inverted repeats on both sides are highlighted in red. Note that there are two reads spanning the inverted repeat regions in (A), confirming this orientation, while the alternate orientation lacks the necessary support by a second read spanning IRA in (B).
4. Discussion

As one of the most important forest trees of central Europe, the European beech has been intensely studied throughout the past decades with respect to its ecology [33–35], phenology [36,37], and population genetics [11,38]. The latter studies are usually based on fingerprinting methods [18,38], but recently, SNP-based studies in beeches have been reported [11,12,16,39,40]. Especially for maternal inheritance, which can inform about seed dispersal rates, phylogeography, migration, and tracking of historic and current human-mediated seed transfer, chloroplast-based analyses are useful. Using manual curation to account for assembly errors, a circular chloroplast genome sequence of the *Fagus sylvatica* reference individual, Bhaga, from the German National Park Kellerwald-Edersee [21], could be obtained. A comparison to the chloroplast genome of a Polish beech individual, Jamy, and the recently published chloroplast genome of [19] from another German individual revealed surprisingly low variation, with only two SNP and three indel positions. Interestingly, the order of the chloroplast elements of Bhaga was found to be different from the previously published chloroplast genome of beech (MK598696) [19]. However, the structure of our chloroplast genome has been verified by long-read alignment, providing evidence of the accuracy of this structure. Whether there is a structural difference between the genome reported previously by Mader et al. [19] or if there was an assembly artefact needs to be clarified by future studies. In contrast to the low variation in beech, in the annual plant *Microthlaspi erraticum*, already several SNPs were encountered in the comparatively small *trnL-F* gene in central Europe [41], probably reflecting a higher dispersal ability. The high uniformity of the chloroplast genome is in line with earlier studies [9,18], but is contrasted by the significant variation observed in the nuclear genome of *Fagus sylvatica* in central Europe [11,35,42,43]. A reason for the low amount of chloroplast variation might be that central Europe was colonised primarily from small but quickly expanding refugial populations [9,18], resulting in a limited number of chloroplast haplotypes. As beech seeds are not frequently dispersed over long distances [44,45], the forerunners of the colonisation would inherit the same chloroplast type during postglacial colonisation of central Europe, while the non-maternally inherited genome content could diversify due to pollen influx. In line with the low genetic diversity in the chloroplast genomes, Meger et al. [19], using RADseq methods, identified only eight SNPs located in the beech chloroplast genome on a metapopulation dataset, which also included samples from the northern Carpathians.

Additional complete chloroplast genomes from a broad range of beech populations including glacial refugia will be the key to test the hypothesis of low seed dispersal rates and efficient gene-flow by pollen dispersal, despite previous assumptions of rather low pollen dispersal rates [46]. If the hypothesis will be confirmed and more variation in chloroplast diversity will be confirmed in chloroplast genomes from refugia in future studies, the chloroplast genome might offer a high-resolution tool to track the migration of *Fagus sylvatica* in the Holocene.

Author Contributions: Conceptualization, M.T.; formal analysis, B.M., M.T., B.U., J.B.; resources, M.T., S.P.; data curation, B.M., B.U.; writing—original draft preparation, M.T., B.M.; writing—review and editing, J.B., B.U., M.T., B.M.; visualization, B.M., M.T.; supervision, M.T., J.B.; project administration, M.T., J.B.; funding acquisition, M.T., J.B. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the National Science Centre, Poland (2012/04/A/NZ9/00500), the Polish Ministry of Science and Higher Education under the program “Regional Initiative of Excellence” in 2019–2022 (Grant No. 006/RID/2018/19), and the German Science Foundation (Grant No. Th1362/18-1). Additional support by the LOEWE initiative in the framework of the Centre for Translational Biodiversity Genomics (TBG) is gratefully acknowledged.

Data Availability Statement: The chloroplast genome sequences have been deposited in GenBank under the accession numbers MW531753 for Bhaga and MW537046 for Jamy.

Acknowledgments: The authors gratefully acknowledge the permission of the office of the National Park Kellerwald-Edersee to sample the reference individual Bhaga.
Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References
1. Giesece, T.; Hickler, T.; Kunkel, T.; Sykes, M.T.; Bradshaw, R.H. Towards an understanding of the Holocene distribution of Fagus sylvatica L. J. Biogeogr. 2007, 34, 118–131. [CrossRef]
2. Brunet, J.; Fritz, O.; Richnau, G. Biodiversity in European beech forests—a review for sustainable forest management. Ecol. Bull. 2010, 53, 77–94.
3. Jump, A.S.; Hunt, J.M.; Peñuelas, J. Rapid climate change-related growth decline at the southern range edge of Fagus sylvatica. Glob. Chang. Biol. 2006, 12, 2163–2174. [CrossRef]
4. Gessler, A.; Keitel, C.; Kreuzwieser, J.; Matyssek, R.; Seiler, W.; Rennenberg, H. Potential risks for European beech (Fagus sylvatica L.) in a changing climate. Trees 2007, 21, 1–11. [CrossRef]
5. Albert, R.E.I.F.; Xystrakis, F.; Gaertner, S.; Sayer, U. Floristic change at the drought limit of European beech (Fagus sylvatica L.) in the Postglacial period. J. Biogeogr. 2007, 34, 118–131. [CrossRef]
6. Sandurska, E.; Ulaszewski, B.; Burczyk, J. Genetic Insights into Ecological Succession from Oak- (Quercus robur L.) to Beech- (Fagus sylvatica L.) Dominated Forest Stands. Acta Biol. Crac. Bot. 2017, 59, 23–33. [CrossRef]
7. Comps, B.; Thiébaut, B.; Paule, L.; Merzeau, D.; Letouzey, J. Allozymic variability in beechwoods (Fagus sylvatica L.) over central Europe: Spatial differentiation among and within populations. Heredity 1990, 65, 407–417. [CrossRef]
8. Leonardi, S.; Menozzi, P. Genetic variability of Fagus sylvatica L. in Italy: The role of postglacial recolonization. Heredity 1995, 75, 35–44. [CrossRef]
9. Demesure, B.; Comps, B.; Petit, R.J. Chloroplast DNA Phylogeography of the Common Beech (Fagus sylvatica L.) in Europe. Evolution 1996, 50, 2515. [CrossRef]
10. Sjölund, M.J.; González-Díaz, P.; Moreno-Villena, J.J.; Jump, A.S. Understanding the legacy of widespread population translocations on the post-glacial genetic structure of the European beech wood (Fagus sylvatica L. & B. J. Biogeogr. 2017, 44, 2475–2487. [CrossRef]
11. Cuervo-Alarcon, L.; Arend, M.; Müller, M.; Sperisen, C.; Finkeldey, R.; Krutovsky, K.V. Genetic variation and signatures of natural selection in populations of European beech (Fagus sylvatica L.) along precipitation gradients. Tree Genet. Genomes 2018, 14, 84. [CrossRef]
12. Müller, M.; Lopez, P.A.; Papageorgiou, A.C.; Sisiripidis, I.; Gailing, O. Indications of Genetic Admixture in the Transition Zone between Fagus sylvatica L. and Fagus orientalis L. Diversity & Burd. 2019, 11, 90. [CrossRef]
13. Sebastiano, F.; Carnevale, S.; Vendramin, G.G. A new set of mono-and dinucleotide chloroplast microsatellites in Fagaceae. Mol. Ecol. Notes 2004, 4, 259–261. [CrossRef]
14. Vettori, C.; Vendramin, G.G.; Anzidei, M.; Pastorelli, R.; Paffetti, D.; Giannini, R. Geographic distribution of chloroplast variation in Italian populations of beech (Fagus sylvatica L.). Theor. Appl. Genet. 2004, 109, 1–9. [CrossRef]
15. Hatziskakis, S.; Papageorgiou, A.C.; Gailing, O.; Finkeldey, R. High chloroplast haplotype diversity in Greek populations of beech (Fagus sylvatica L.). Plant Biol. 2009, 11, 425–433. [CrossRef] [PubMed]
16. Mader, M.; Liesebach, B.; Paule, L.; Rennenberg, H. The plastid and mitochondrial genomes of Eucalyptus grandis. BMC Genom. 2019, 20, 132. [CrossRef] [PubMed]
26. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2017, 45, e18.

27. Manos, P.S.; Stanford, A.M. The historical biogeography of Fagaceae: Tracking the Tertiary history of temperate and subtropical forests of the northern hemisphere. *Int. J. Plant Sci.* 2001, 162, S77–S93. [CrossRef]

28. Chaisson, M.J.; Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinform.* 2012, 13, 238. [CrossRef]

29. Walker, B.J.; Abeel, T.; Shen, T.; Priest, M.; Abouelliel, A.; Sakhkhumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 2014, 9, e112963. [CrossRef] [PubMed]

30. Marçais, G.; Delcher, A.L.; Phillippy, A.M.; Coston, R.; Salzberg, S.L.; Zimin, A.V. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* 2018, 14, e1005944. [CrossRef]

31. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017, 45, W6–W11. [CrossRef] [PubMed]

32. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* 2004, 5, R12. [CrossRef] [PubMed]

33. Barna, M. Adaptation of European beech (*Fagus sylvatica* L.) to different ecological conditions: Leaf size variation. *Pol. J. Ecol.* 2004, 52, 35–45.

34. Maringer, J.; Ascoli, D.; Küffer, N.; Schmidtkein, S.; Conedera, M. What drives European beech (*Fagus sylvatica* L.) mortality after forest fires of varying severity? *For. Ecol. Manag.* 2016, 368, 81–93. [CrossRef]

35. Rey, F.; Gobet, E.; Schwörer, C.; Wey, O.; Hafer, A.; Tinner, W. Causes and mechanisms of synchronous succession trajectories in primeval Central European mixed *Fagus sylvatica* forests. *J. Ecol.* 2019, 107, 1392–1408. [CrossRef]

36. Gauzere, J.; Delzon, S.; Davi, H.; Bonhomme, M.; De Cortazar-Atauri, I.G.; Chuine, I. Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: *Fagus sylvatica* and *Quercus petraea*. *Agric. For. Meteorol.* 2017, 244, 9–20. [CrossRef]

37. Krajmerová, D.; Hrivnáková, M.; Ditmarová, L.; Jannicka, G.; Kmeť, J.; Kurjak, D.; Gomóry, D. Nucleotide polymorphisms associated with climate, phenology and physiological traits in European beech (*Fagus sylvatica* L.). *New For.* 2017, 48, 463–477. [CrossRef]

38. Sjölund, M.J.; González-Díaz, P.; Moreno-Villena, J.J.; Jump, A.S. Gene flow at the leading range edge: The long-term consequences of isolation in European beech (*Fagus sylvatica* L. Kuhn). *J. Biogeogr.* 2019, 46, 2787–2799. [CrossRef]

39. Capblancq, T.; Morin, X.; Gueguen, M.; Renaud, J.; Lobreaux, S.; Bazin, E. Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *J. Ecol. Biol.* 2020, 33, 783–796. [CrossRef]

40. Tsukamoto, M.; Akada, S.; Matsuda, S.; Jouyu, H.; Kisanuki, H.; Tomaru, N.; Torimaru, T. Assessments of fine-scale spatial patterns of SNPs in an old-growth beech forest. *Heredity* 2020, 125, 240–252. [CrossRef]

41. Ali, T.; Schmucker, A.; Runge, F.; Solovyeva, I.; Nigrelli, L.; Paul, J.; Buch, A.-K.; Xia, X.; Ploch, S.; Orren, O.; et al. Morphology, phylogeny, and taxonomy of *Microthlaspi* (Brassicaceae: Coluteocharpeae) and related genera. *Taxon* 2016, 65, 79–98. [CrossRef]

42. Pfueß, A.R.; Frank, A.; Heiri, C.; Lalagüe, H.; Vendramin, G.G.; Oddou-Muratorio, S. Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytol.* 2016, 210, 589–601. [CrossRef]

43. Comps, B.; Gömöry, D.; Letouzey, J.; Thiebaut, B.; Petit, R.J. Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* 2001, 157, 389–397.

44. Martínez, I.; González-Taboada, F. Seed dispersal patterns in a temperate forest during a mast event: Performance of alternative dispersal kernels. *Oecologia* 2009, 159, 389–400. [CrossRef] [PubMed]

45. Bialozyt, R.; Bradley, L.R.; Bradshaw, R.H. Modelling the spread of *Fagus sylvatica* and *Picea abies* in southern Scandinavia during the late Holocene. *J. Biogeogr.* 2012, 39, 665–675. [CrossRef]

46. Oddou-Muratorio, S.; Boncompas, A.; Klein, E.K.; Chybicki, I.J.; Vendramin, G.G.; Suyama, Y. Comparison of direct and indirect genetic methods for estimating seed and pollen dispersal in *Fagus sylvatica* and *Fagus crenata*. *For. Ecol. Manag.* 2010, 259, 2151–2159. [CrossRef]