

Article

Complete Chloroplast Genomes of *Fagus sylvatica* L. Reveal Sequence Conservation in the Inverted Repeat and the Presence of Allelic Variation in NUPTs

Bartosz Ulaszewski ^{1,*}, Joanna Meger ¹, Bagdevi Mishra ^{2,3}, Marco Thines ^{2,3,4} and Jarosław Burczyk ¹

¹ Department of Genetics, Faculty of Biological Sciences, Kazimierz Wielki University, Chodkiewicza 30, 85-064 Bydgoszcz, Poland; warmbier@ukw.edu.pl (J.M.); burczyk@ukw.edu.pl (J.B.)

² Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany; bagdevi.mishra@senckenberg.de (B.M.); marco.thines@senckenberg.de (M.T.)

³ Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, Goethe University, Max-von-Laue-Str. 13, 60483 Frankfurt am Main, Germany

⁴ LOEWE Centre for Translational Biodiversity Genomics, Georg-Voigt-Str. 14-16, 60325 Frankfurt am Main, Germany

* Correspondence: ulaszewski@ukw.edu.pl

Abstract: Growing amounts of genomic data and more efficient assembly tools advance organelle genomics at an unprecedented scale. Genomic resources are increasingly used for phylogenetic analyses of many plant species, but are less frequently used to investigate within-species variability and phylogeography. In this study, we investigated genetic diversity of *Fagus sylvatica*, an important broadleaved tree species of European forests, based on complete chloroplast genomes of 18 individuals sampled widely across the species distribution. Our results confirm the hypothesis of a low cpDNA diversity in European beech. The chloroplast genome size was remarkably stable ($158,428 \pm 37$ bp). The polymorphic markers, 12 microsatellites (SSR), four SNPs and one indel, were found only in the single copy regions, while inverted repeat regions were monomorphic both in terms of length and sequence, suggesting highly efficient suppression of mutation. The within-individual analysis of polymorphisms showed >9k of markers which were proportionally present in gene and non-gene areas. However, an investigation of the frequency of alternate alleles revealed that the source of this diversity originated likely from nuclear-encoded plastome remnants (NUPTs). Phylogeographic and Mantel correlation analysis based on the complete chloroplast genomes exhibited clustering of individuals according to geographic distance in the first distance class, suggesting that the novel markers and in particular the cpSSRs could provide a more detailed picture of beech population structure in Central Europe.

Keywords: complete chloroplast genome; microsatellite; SNP; indel; heteroplasmy; population genomics; European beech



Citation: Ulaszewski, B.; Meger, J.; Mishra, B.; Thines, M.; Burczyk, J. Complete Chloroplast Genomes of *Fagus sylvatica* L. Reveal Sequence Conservation in the Inverted Repeat and the Presence of Allelic Variation in NUPTs. *Genes* **2021**, *12*, 1357. <https://doi.org/10.3390/genes12091357>

Received: 24 May 2021

Accepted: 27 August 2021

Published: 29 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chloroplasts not only play a key role in photosynthesis but also other metabolic processes of green plants [1]. The generally maternal inheritance of the chloroplast genome in Angiosperms and relatively conserved gene content and order has made chloroplast genomes a valuable resource for phylogenetic and evolutionary studies [2,3]. Plant chloroplast genomes are mostly between 120 kb and 160 kb in length and usually have a quadripartite circular structure comprising of two regions of inverted repeats A and B (IR-A/IR-B), separated by a large single-copy (LSC) region and a small single-copy (SSC) region [4]. Due to next-generation sequencing approaches, sequencing of chloroplast genomes for dozens or hundreds of individuals is now achievable through whole genome sequencing [5,6]. Chloroplast genome sequences have helped to elucidate the phylogenetic relationships

and evolutionary history of many tree genera, including *Acer* [7], *Prunus* [8], *Populus* [9,10], *Quercus* [11,12] and *Pinus* [13]. With an increased availability of whole chloroplast sequences, numerous studies demonstrated the presence of variation among individuals within species, which includes SNPs, indels, inversions, translocations, copy number variations and also IR expansion, gene loss and intron retention [14]. The level of this variability is usually considered low, both in terms of composition, as well as in terms of the degree of variation within the different regions [14–17]. Interestingly, the individuals themselves have been reported to show variations of their chloroplast genomes through heteroplasmy—which can occur as a result of independent mutations or biparental inheritance of organelles in one organism [18]. While chloroplast genomes can be a good tool for phylogeographic analyses, such studies are currently limited to only few conifers [16,17].

European beech (*Fagus sylvatica* L.) is ecologically and economically one of the most important broadleaved tree species in Europe [19]. There are several molecular studies that evaluated genetic diversity and structure of European beech using chloroplast DNA [20–27]. Most of these studies showed a low level of chloroplast diversity and a rather homogeneous genetic structure in Central Europe, but none of them exploited the full potential resolution offered by complete chloroplast genomes. Thus, there is still lack of comparative analyses based on complete chloroplast genomes, which would allow to identify novel chloroplast polymorphisms and to detect genetic structure of European beech at a regional scale. Recently, Mishra and coauthors [26] evaluated three complete chloroplast genomes of beech from areas glaciated during the Weichselian glacial maximum and found a very low genetic variation with only two SNP and three indel positions. This raised the question of whether the low variation found was due to genetic impoverishment by founder effects at the leading edge of the recolonization or if chloroplast genetic diversity is generally low in European beech. To clarify this, there is a need to assess genetic diversity of complete chloroplast genome of European beech sampled from a wider range, which was the aim of the current study.

Here, we report 16 newly sequenced and assembled complete chloroplast genomes of *F. sylvatica* and perform comparative genomic analyses of the new sequences with the two recently published chloroplast genomes: Bhaga (MW531753) and Jamy (MW537046) [26]. The aim of our study was to identify potentially highly variable markers in chloroplast genome of *F. sylvatica* suitable for phylogeographic studies as a useful genetic resource for developing chloroplast-based genetic markers (SNPs and SSRs) for large-scale population studies.

2. Materials and Methods

2.1. DNA Isolation and Sequencing

Details regarding DNA isolation and sequencing of Bhaga and Jamy individuals are given in Mishra and coauthors [26]. The remaining 16 individuals representing a wide range of the species distribution were collected from Siemianice provenance trial [28]. Detailed geographic locations are presented in Table 1 and Figure S1. DNA was isolated from leaves with a GeneMATRIX Plant & Fungi DNA Purification Kit (EURx, Poland), after storing them in the dark for 48 h. Genomic library preparation and sequencing was done by an external service provider (IGA Technology Services s.r.l.) with Illumina HiSeq 250 device in 125-bp PE mode. The obtained reads were purified from adapters and trimmed with Trimmomatic [29]. Raw reads were deposited in SRA under the accession numbers listed in Table 1.

Table 1. Origin of sampled individuals and sequencing data volume.

No.	Origin or Individual Name	Country	Longitude	Latitude	Number of Read Pairs	NCBI Accession Number	SRA Accession Number
1	Bhaga	Germany	51.169167 N	8.963056 E	[26]	MW531753	N/A
2	Jamy	Poland	53.586019 N	18.935019 E	[26]	MW537046	SAMN08948264
3	Gdańsk	Poland	54.383262 N	18.516724 E	3,777,769	MW566769	SAMN18917950
4	Foret des Colettes	France	46.183328 N	2.949992 E	4,899,373	MW566771	SAMN18917951
5	Limitaciones	Spain	42.818059 N	2.249663 W	6,210,877	MW566772	SAMN18917952
6	Glorup	Denmark	55.184748 N	10.681238 E	20,891,953	MW566770	SAMN18917953
7	Lopuchówko	Poland	52.583300 N	17.083339 E	5,114,816	MW566774	SAMN18917954
8	Hasbruch	Germany	53.120708 N	8.4302740 E	4,650,347	MW566776	SAMN18917955
9	Bieszczady NP	Poland	49.117093 N	22.579103 E	3,046,013	MW566773	SAMN18917956
10	Eisenach	Germany	50.087605 N	10.106152 E	4,461,792	MW566778	SAMN18917957
11	Morbach	Germany	50.740891 N	6.980116 E	5,833,195	MW566784	SAMN18917958
12	Ehingen	Germany	48.399106 N	9.500861 E	5,632,928	MW566775	SAMN18917959
13	Veneto	Italy	46.133489 N	12.216683 E	7,741,036	MW566783	SAMN18917960
14	Cesky Krumlov	Czechia	48.850035 N	14.250406 E	7,853,097	MW566777	SAMN18917961
15	Brzeziny	Poland	51.836489 N	19.601247 E	7,349,714	MW566779	SAMN18917962
16	Smolenice	Slovakia	48.485171 N	17.372687 E	5,072,400	MW566782	SAMN18917963
17	Fantanele	Romania	46.416750 N	26.466475 E	6,584,825	MW566780	SAMN18917964
18	Fläming	Germany	52.133389 N	12.583406 E	7,423,489	MW566781	SAMN18917965

2.2. Chloroplast Genome Assemblies and Annotation

Methods describing assembly and annotation of chloroplast genomes of Bhaga and Jamy individuals are presented in Mishra and coauthors [26]. All the remaining 16 chloroplast genomes were generated using the same protocol: Illumina reads were used for de novo assembly using NOVOPlasty v 4.2. [30,31] with seed sequence NCBI: AY453092.1 [32] and Bhaga chloroplast genome for guiding the program in both inverse repeat regions. After assembling, the sequences were manually checked, in case of presence of ambiguous nucleotides manual curation was done with the assistance of reads mapped to a genome with bwa-mem [33] and visualization of the results in Tablet software [34]. Coding sequences and RNA elements annotation was done with GeSeq ChloroBox [35] using chloroplast genomes of *F. crenata* (NC_041252; [36]), *F. engleriana* (KX852398; [37]), *F. japonica* (MT762294; [38]) and *F. sylvatica* (NC_041437; [39]) as references. Postprocessed annotated genomes were uploaded to GenBank, for accession numbers see Table 1.

2.3. Assessment of Genome Variation

REPuter was employed to identify four types of large repeating sequences (reverse, forward, complement and palindromic) with a minimum repeat size of 30 bp, hamming distance equal to 3 maximum computed repeats set to 50 [40]. Identification of chloroplast simple sequence repeats (cpSSR) was done using MISA [41]. The minimum number of repeat units was set to eight, six, five, five, three, and three, for mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides, respectively. For assessment of variance between the 18 studied chloroplast genomes alignments were done with MAFFT v 7.450 [42], as implemented in Unipro UGENE [43]. After this variations among the genomes were highlighted using the Bhaga chloroplast genome as reference.

2.4. Detection of Heteroplasmy

Potential chloroplast genome variation within individuals (heteroplasmy) was assessed with mapping of reads of each individual with bwa [33] to extracted SSC, LSC and IR regions of the genomes. The marker calling was done with FreeBayes [44] with 0.02 minor allele frequency (MAF) and depth of 200x thresholds for variant detection to

avoid sequencing error [45]. To verify the origin of the markers reads were mapped to chromosome 10 of the *Fagus sylvatica* nuclear assembly [46].

2.5. Phylogenetic Analysis

Phylogenetic analysis was done with on the basis of the dataset of the 18 *Fagus sylvatica* assemblies to which the complete chloroplast genomes of *F. crenata* (NC_041252; [36]), *F. japonica* (MT762294; [38]) and *F. engleriana* (KX852398; [37]) before alignment as described above. Phylogenetic reconstructions were done using IQ-TREE with the GTR+I+R model [47–49] and 1000 bootstrap replicates. The resulting phylogram was edited with Figtree 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>, accessed on 21 March 2021) with rooting to midpoint and proportional transformation of branches.

The phylogenetic distance matrix obtained from IQ-TREE was also used to test the relationship between genetic and geographic distances among individuals. The correlation was calculated with PASSaGE 2 [50] using Mantel's test [51] for a global assessment and Mantel's correlogram to search for significance within 10 equally paired distance classes (the largest class excluded). All tests were performed with 10,000 permutations.

3. Results

3.1. Assembly Size Variance and Genome Annotation

Read coverage of each of 16 new assemblies varied from 86x to 625x with average value of 302x and 284x median. Chloroplast genome structure was stable throughout the studied sequences, and assembly sizes varied from 158,391 bp to 158,464 bp, with highest length variance observed in the large single copy region (LSC) (87,634–87,706 bp). The small single copy regions (SSC) differed only by 4 bp (19,010–19,013 bp), while both inverted repeat regions (IR-A/IR-B) were identical in length (25,873 bp) (Table 2).

Table 2. Statistics for main chloroplast genome elements: LSC - large single copy region, SSC—small single copy region, IR-A/IR-B—invited repeat regions A and B.

Origin or Individual Name	Read Coverage	NCBI Accession Number	Total Size (bp)	Main Genome Elements		
				LSC (bp)	SSC (bp)	IR-A/IR-B (bp)
Bhaga	-	MW531753	158,458	87,702	19,010	25,873
Jamy	-	MW537046	158,462	87,705	19,011	25,873
Gdańsk	253x	MW566769	158,456	87,699	19,011	25,873
Colettes	498x	MW566771	158,391	87,634	19,011	25,873
Limitaciones	491x	MW566772	158,461	87,704	19,011	25,873
Glorup	356x	MW566770	158,461	87,704	19,011	25,873
Lopuchówko	212x	MW566774	158,461	87,704	19,011	25,873
Hasbruch	267x	MW566776	158,462	87,705	19,011	25,873
Bieszczady NP	211x	MW566773	158,426	87,669	19,011	25,873
Eisenach	105x	MW566778	158,456	87,699	19,011	25,873
Morbach	350x	MW566784	158,463	87,706	19,011	25,873
Ehingen	91x	MW566775	158,446	87,689	19,011	25,873
Veneto	625x	MW566783	158,463	87,706	19,011	25,873
Cesky Krumlov	300x	MW566777	158,462	87,705	19,011	25,873
Brzeziny	521x	MW566779	158,462	87,705	19,011	25,873
Smolenice	86x	MW566782	158,430	87,674	19,010	25,873
Fantanele	157x	MW566780	158,462	87,705	19,011	25,873
Fläming	306x	MW566781	158,464	87,705	19,013	25,873

Similarly to the previously published *F. sylvatica* chloroplast assemblies [26,39] each of the 16 new genomes had an identical set of 131 annotated elements: 83 protein coding genes, 8 rRNAs and 40 tRNAs. Total share of coding elements differed across main elements of the genome, with 51% in LSC, 72.1% in SSC and 59.1% in IR regions.

3.2. Repeat elements and SNPs

Large repeating sequence (LRS) assessment showed that sixteen genomes had 32 LRSs >30 bp: 16 forward, 13 reverse, one complement and two palindromic. Two assemblies (Gdańsk and Glorup) had 31 LRSs as a result of the loss of one palindromic match, and one chloroplast genome (Ehingen) had 33 LRSs with an additional reverse match compared to the 16 previously mentioned assemblies. Analysis of cpSSR using MISA detected a total number of 138 markers, out of which only 4 of 97 mononucleotide cpSSRs and 8 of 35 complex cpSSRs were polymorphic. All discovered dinucleotide and pentanucleotide cpSSRs were found to be monomorphic (Table 3).

Table 3. General characteristics of chloroplast microsatellite markers in 18 *F. sylvatica* individuals.

	Mononucleotide	Dinucleotide	Pentanucleotide	Complex	Total
Monomorphic	93	2	4	27	126
Polymorphic	4	-	-	8	12
Total	97	2	4	35	138

In the SSC region we found four polymorphic markers: two mononucleotide and two complex repeats. Variation of the SSR mononucleotide T at position 12,538 occurred due to the absence of a microsatellite in one of the individuals (Fantanelle). The LSC region had eight polymorphic markers with six complex and two mononucleotide cpSSR (Table 4). Marker ratio, reflecting the number of individuals associated with a particular variant, showed that in eight sites an alternative nucleotide was present, while three cpSSR loci had two, and one locus had five variants.

Table 4. Basic information of polymorphic chloroplast microsatellites; marker ratio—number of individuals associated with a particular marker variant; region types: LSC—Large Single Copy; SSC—Small Single Copy.

No.	Starting Position (bp) *	Type	Region	Marker Ratio	Flanking Annotation
1	4363	Complex	SSC	17/1	ndhA (exon II) ↔ ndhA (exon I)
2	8012	Complex	SSC	16/1/1	psaC ↔ ndhD
3	11,476	Mononucleotide (A)	SSC	17/1	trnL ↔ rpl32
4	12,583	Mononucleotide (T)	SSC	17/0 **	rpl32 ↔ ndhF
5	46,142	Complex	LSC	16/1/1	matK ↔ trnQ
6	46,952	Complex	LSC	11/2/2/1/1/1	matK ↔ trnQ
7	50,589	Mononucleotide (A)	LSC	17/1	trnG (exon I) ↔ trnG (exon II)
8	55,923	Complex	LSC	16/2	atpH ↔ atpI
9	70,097	Complex	LSC	16/2	rpoB ↔ trnC
10	92,043	Mononucleotide (A)	LSC	16/2	trnG (exon II) ↔ trnG (exon I)
11	105,126	Complex	LSC	12/5/1	ycf4 ↔ cemA
12	107,580	Complex	LSC	17/1	petA ↔ psbJ

* according to the Bhaga reference; ** marker absent in an individual

Alignment of the 18 chloroplast genomes revealed four SNPs and one indel located in noncoding regions. The first SNP (pos. 12,587) was associated with a mononucleotide cpSSR, the alternative variant was present in an individual from Fantanele, shortening the repetitive sequence to 7 bp. Variants for second (pos. 46,985) and fourth (pos. 112,198) SNP, as well as the indel (pos. 80,558) were present in only one individual; however, in the third SNP (pos. 71,204) 50% of individuals had the alternative variant (Table 5).

Table 5. Summary of the variant sites detected in the 18 chloroplast genomes, region types: LSC—Large Single Copy; SSC—Small Single Copy.

No.	Position (bp) *	Marker Type	Region	Consensus	Alternative	Area	Marker Ratio	Flanking Annotation
1	12,587	SNP	SSC	T	C	noncoding	17/1	rpl32 ↔ ndhF
2	46,985	SNP	LSC	G	A	noncoding	17/1	tRNA-K ↔ tRNA-Q
3	71,204	SNP	LSC	G	T	noncoding	9/9	tRNA-C ↔ petN
4	80,558	Indel	LSC	T	-	noncoding	17/1	psbZ ↔ tRNA-G
5	112,198	SNP	LSC	A	C	noncoding	17/1	psaJ ↔ rpl3

* the position (bp) is referred to the Bhaga genome

3.3. Within Individual Polymorphisms

Within individual polymorphism related to chloroplast genome was found in all 16 tested individuals. After filtering with and minimum depth >200x and MAF of 0,02 a total number of 9028 markers were detected in all analyzed regions.

The average depth of each base at a variant position was lower in single copy elements 349x and 360x in LSC and SSC respectively, while both IR regions had 477x depth (Table 6). However, the average alternative variant's depth was very similar across the main genome regions from 16.1x in SSC, 18.4–18.5x in IRs and 18.7x in LSC. This suggests that these variants represent the nuclear encoded plastome sequences (NUPTs), as these values are similar to the average coverage of 17x found on chromosome 10 of the complete nuclear genome. This chromosome was selected for comparative analysis due to lowest NUPT detection in the whole assembly.

Table 6. Summary statistics of within individual polymorphisms detected in regions of the 16 chloroplast genome assemblies. LSC - large single copy region, SSC—small single copy region, IR-A/IR-B—involved repeat regions A and B.

	LSC	SSC	IR-A	IR-B
Avg. variant depth	349x	360x	477x	477x
Avg. alternative var. depth	18.7x	16.1x	18.4x	18.5x
Number of unique positions	5348	1161	1257	1262
SNP	76.8%	80.9%	83.7%	84.1%
Indel	10.2%	8.8%	9.6%	9.4%
Complex	8.2%	6.2%	3.1%	3.1%
MNP	0.2%	0.3%	0.8%	0.7%
Mix	4.6%	3.9%	2.7%	2.7%
Coding	48.6%	67.3%	62.9%	63.1%
Non-coding	51.4%	32.7%	37.1%	36.9%

Among the variant positions majority of them were SNPs (76.8–84.1%), the remaining share was associated to indels (8.8–10.2%), complex markers (3.1–8.2%) and MNPs (0.2–0.7%). In 2.7–4.6% of sites a mix of variants was detected e.g., a SNP and an indel called at a specific position in the same individual.

Markers detected in this study were found both in coding (48.6–67.3%) and non-coding areas (32.7–51.4%). The size of the contribution in each of these parts was related to the size of coding and non-coding regions of the main genome element (Figure 1).

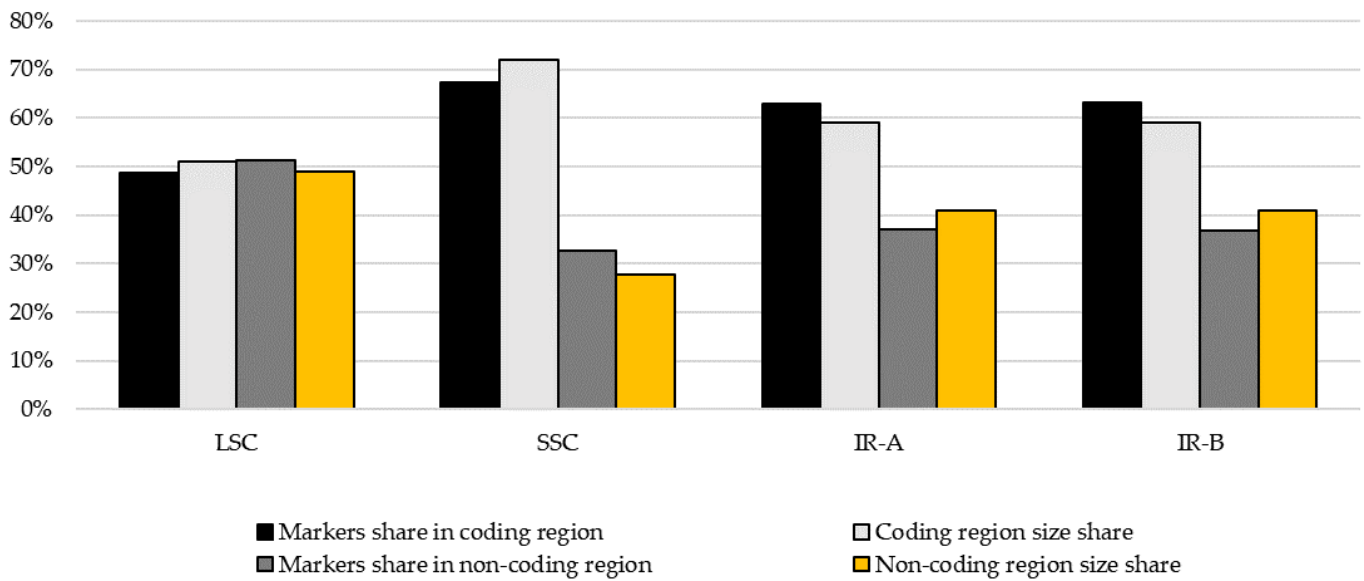


Figure 1. Share or markers in coding and non-coding regions in relation to the size share of coding and non-coding elements in main genome regions: LSC—large single copy region, SSC—small single copy region, IR-A/IR-B—inverted repeat regions A and B.

3.4. Phylogenetic Analysis

The complete chloroplast genome sequences of 18 *F. sylvatica* individuals, as well as *F. crenata*, *F. japonica* and *F. engleriana*, were used to for a phylogenetic reconstruction based on maximum likelihood method and 1000 bootstrap replicates (Figure 2).

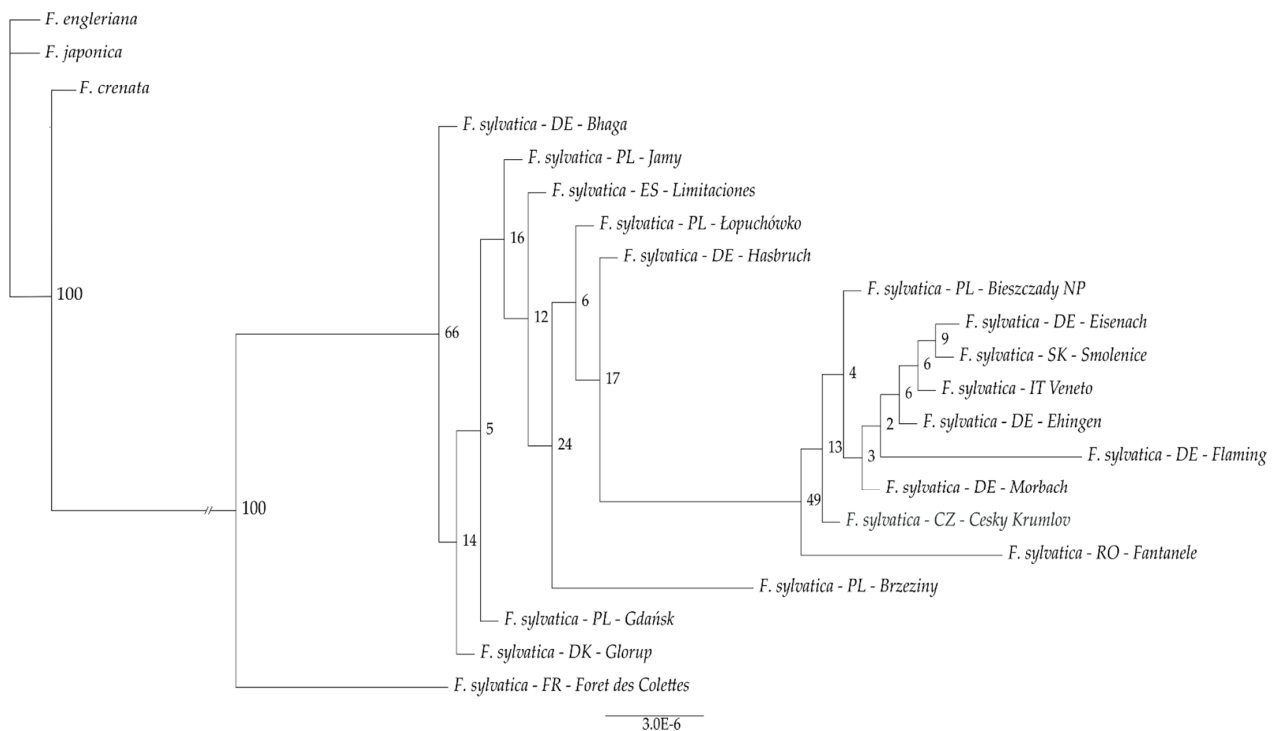


Figure 2. Phylogenetic relationships among 18 *F. sylvatica* individuals, as inferred using Maximum Likelihood, with *F. crenata*, *F. japonica*, and *F. engleriana* as outgroup. Numbers on nodes indicate percentages of bootstrap support from 1000 bootstrap replicates, the genetic distance between the *F. sylvatica* individuals and the outgroup was shortened by 0.04.

While a global Mantel's test did not reveal significant relationship ($r = 0.2190$; $p = 0.1861$) between the phylogenetic and geographic distances for the 18 individuals, clustering of similar individuals was confirmed with Mantel's correlogram within the first distance class (<250 km; $r = 0.286$; $p = 0.011$) (Table 7).

Table 7. Summary of Mantel's test statistics calculated within consecutive distance classes.

Class	Boundry max (km)	Number of Pairs	Mantel r	p
1	250	11	0.286	0.011
2	500	31	0.106	0.361
3	750	46	0.121	0.144
4	1000	27	−0.016	0.760
5	1250	15	−0.004	0.900
6	1500	11	−0.023	0.374

4. Discussion

Complete chloroplast genomes have helped to reveal species relationships [9,11,12], but also allow to measure divergence within populations [16,17]. Growing genomic resources for European beech provide tools to extend our knowledge on this critically important forest tree species. Our results support previous hypotheses suggesting low genetic diversity of the beech plastome [23,52,53].

The total genome length variation ($158,428 \pm 37$ bp) and the presence of polymorphisms were associated exclusively with Single Copy (SC) regions, while the pairs of IR regions were monomorphic both in terms of the length (25,873 bp) and in terms of nucleotide sequence. Stability of IR and variability of SC regions was also present in sequences of *F. japonica* (MT762294; MT762295), the results suggest a powerful gene conversion mechanism in *Fagus* species. Our study revealed 138 cpSSR in *F. sylvatica* out of which 126 were monomorphic. This group included universal cpSSRs: ccmp4, ccmp7, ccmp10, commonly used to assess phylogeny and relationship in eudicot species [54,55]. Magri and coauthors [23] using these markers concluded that Central Europe beech populations generally can be considered as a homogeneous group. The 12 polymorphic microsatellite markers discovered in this study, when applied for a higher number of individuals and populations, could, however, potentially provide a more detailed phylogeographic picture. Our phylogeographic analysis supports this assumption due to significant clustering of individuals over a relatively short distance <250 km.

Additional source of variation was found in SNPs (4) and indel (1), all located in noncoding regions, but their position is not in line with results obtained based on reduced representation genomic libraries presented by Meger and coauthors [27]. Heteroplasmy is well reported in plants with known biparental inheritance of chloroplasts, even though in some species (e.g., *Passiflora*) it can occur at the seedling and embryo but not at the mature developmental stages [56]. In beech, due to maternal inheritance of organelles [3], heteroplasmy can exclusively be caused by mutations. The evidence of multiple integrations of organelle DNA integration with the nuclear genome in beech [46] and detection of within-individual polymorphisms of cpDNA-related sequences presented in this study suggest that assessing beech diversity with chloroplast related SNPs due to a large occurrence of nuclear encoded plastid DNA (NUPTs) can lead to uncertain results and should be taken with caution [57].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12091357/s1>, Figure S1: Distribution map of the individuals origin sites and the *Fagus sylvatica* species range.

Author Contributions: Conceptualization, B.U., J.B.; methodology, B.U., J.M.; investigation, B.U., J.M.; data curation, B.U., B.M.; writing—original draft preparation, B.U., J.M.; writing—review and editing, B.U., J.M., B.M., M.T., J.B.; 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 Center, 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. 008/RID/2018/19), and the German Science Foundation (Grant No. Thi1362/18-1). 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: MW566769, MW566771, MW566772, MW566770, MW566774, MW566776, MW566773, MW566778, MW566784, MW566775, MW566783, MW566777, MW566779, MW566782, MW566780, MW566781.

Acknowledgments: We would like to thank Władysław Barzdajn from Poznań University of Life Sciences and the staff of Experimental Forest District in Siemianice. We also to thank our lab team members: Ewa Sztupecka and Katarzyna Meyza for their outstanding job in fieldwork and DNA isolations. authors gratefully acknowledge the permission of the office of the National Park Kellerwald-Edersse to sample the reference individual Bhaga.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Douglas, S.E. Plastid evolution: Origins, diversity, trends. *Curr. Opin. Genet. Dev.* **1998**, *8*, 655–661. [[CrossRef](#)]
- Wu, Z.Q.; Ge, S. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol. Phylogenet. Evol.* **2012**, *62*, 573–578. [[CrossRef](#)]
- Birky, C.W., Jr. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Ann. Rev. Genet.* **2001**, *35*, 125–148. [[CrossRef](#)] [[PubMed](#)]
- Wicke, S.; Schneeweiss, G.M.; de Pamphilis, C.W.; Muller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)] [[PubMed](#)]
- Straub, S.C.; Parks, M.; Weitemier, K.; Fishbein, M.; Cronn, R.C.; Liston, A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am. J. Bot.* **2012**, *99*, 349–364. [[CrossRef](#)]
- Bock, D.G.; Andrew, R.L.; Rieseberg, L.H. On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* **2014**, *23*, 4899–4911. [[CrossRef](#)] [[PubMed](#)]
- Xia, X.; Yu, X.; Fu, Q.; Zheng, Y.; Zhang, C. Complete chloroplast genome sequence of the three-flowered maple, *Acer triflorum* (Sapindaceae). *Mitochondrial DNA Part B* **2020**, *5*, 1859–1860. [[CrossRef](#)]
- Xue, S.; Shi, T.; Luo, W.; Ni, X.; Iqbal, S.; Ni, Z.; Huang, X.; Yao, D.; Shen, Z.; Gao, Z. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic. Res.* **2019**, *6*, 89. [[CrossRef](#)]
- Zong, D.; Gan, P.; Zhou, A.; Li, J.; Xie, Z.; Duan, A.; He, C. Comparative analysis of the complete chloroplast genomes of seven *Populus* species: Insights into alternative female parents of *Populus tomentosa*. *PLoS ONE* **2019**, *14*, e0218455. [[CrossRef](#)]
- Du, S. The complete chloroplast genome sequence of *Populus wilsonii* based on landscape design, and a comparative analysis with other *Populus* species. *Mitochondrial DNA B Resour.* **2020**, *5*, 2716–2718. [[CrossRef](#)]
- Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative Analysis of the Complete Chloroplast Genomes of Five *Quercus* Species. *Front. Plant. Sci.* **2016**, *7*, 959. [[CrossRef](#)]
- Liu, X.; Chang, E.; Liu, J.; Jiang, Z. Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. *J. For. Res.* **2021**. [[CrossRef](#)]
- Asaf, S.; Khan, A.L.; Khan, M.A.; Shahzad, R.; Lubna; Kang, S.M.; Al-Harrasi, A.; Al-Rawahi, A.; Lee, I.J. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS ONE* **2018**, *13*, e0192966. [[CrossRef](#)] [[PubMed](#)]
- Sabir, J.S.; Arasappan, D.; Bahieldin, A.; Abo-Aba, S.; Bafeel, S.; Zari, T.A.; Edris, S.; Shokry, A.M.; Gadalla, N.O.; Ramadan, A.M.; et al. Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. *PLoS ONE* **2014**, *9*, e94158. [[CrossRef](#)] [[PubMed](#)]
- Li, F.W.; Harkess, A. A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **2018**, *6*, e1030. [[CrossRef](#)] [[PubMed](#)]
- Bondar, E.I.; Putintseva, Y.A.; Oreshkova, N.V.; Krutovsky, K.V. Siberian larch (*Larix sibirica* L.) chloroplast genome and development of polymorphic chloroplast markers. *BMC Bioinform.* **2019**, *20*, 38. [[CrossRef](#)] [[PubMed](#)]
- Chen, S.; Ishizuka, W.; Hara, T.; Goto, S. Complete Chloroplast Genome of Japanese Larch (*Larix kaempferi*): Insights into Intraspecific Variation with an Isolated Northern Limit Population. *Forests* **2020**, *11*, 884. [[CrossRef](#)]
- Frey, J.E. Genetic flexibility of plant chloroplasts. *Nature* **1999**, *398*, 115–116. [[CrossRef](#)]
- Packham, J.R.; Thomas, P.A.; Atkinson, M.D.; Degen, T. Biological Flora of the British Isles: *Fagus sylvatica*. *J. Ecol.* **2012**, *100*, 1557–1608. [[CrossRef](#)]
- Demesure, B.; Comps, B.; Petit, R.J. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* **1996**, *50*, 2515–2520. [[CrossRef](#)]

21. Sebastiani, 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](#)]
22. 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](#)]
23. Magri, D.; Vendramin, G.G.; Comps, B.; Dupanloup, I.; Geburek, T.; Gomory, D.; Latalowa, M.; Litt, T.; Paule, L.; Roure, J.M.; et al. A new scenario for the quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytol.* **2006**, *171*, 199–221. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Papageorgiou, A.C.; Tsiripidis, I.; Mouratidis, T.; Hatziskakis, S.; Gailing, O.; Eliades, N.G.H.; Vidalis, A.; Drouzas, A.D.; Finkeldey, R. Complex fine-scale phylogeographical patterns in a putative refugial region for *Fagus sylvatica* (*Fagaceae*). *Bot. J. Linn. Soc.* **2014**, *174*, 516–528. [[CrossRef](#)]
26. Mishra, B.; Ulaszewski, B.; Ploch, S.; Burczyk, J.; Thines, M. 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. *Forests* **2021**, *12*, 180. [[CrossRef](#)]
27. Meger, J.; Ulaszewski, B.; Vendramin, G.G.; Burczyk, J. Using reduced representation libraries sequencing methods to identify cpDNA polymorphisms in European beech (*Fagus sylvatica* L.). *Tree Genet. Genomes* **2019**, *15*, 7. [[CrossRef](#)]
28. Barzdajn, W.; Rzeznik, Z. Wstepne wyniki miedzynarodowego doswiadczenia proweniencyjnego z bukiem (*Fagus sylvatica* L.) serii 1993/1995 w Lesnym Zakladzie Doswiadczalnym Siemnianice. *Sylvan* **2002**, *146*, 149–164.
29. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
30. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2017**, *45*, e18. [[CrossRef](#)]
31. Dierckxsens, N.; Mardulyn, P.; Smits, G. Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genom. Bioinform.* **2020**, *2*, lqz011. [[CrossRef](#)]
32. 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](#)]
33. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997v1.
34. Milne, I.; Bayer, M.; Cardle, L.; Shaw, P.; Stephen, G.; Wright, F.; Marshall, D. Tablet—Next generation sequence assembly visualization. *Bioinformatics* **2010**, *26*, 401–402. [[CrossRef](#)] [[PubMed](#)]
35. 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](#)]
36. Worth, J.R.P.; Liu, L.; Wei, F.J.; Tomaru, N. The complete chloroplast genome of *Fagus crenata* (subgenus *Fagus*) and comparison with *F. engleriana* (subgenus *Engleriana*). *PeerJ* **2019**, *7*, e7026. [[CrossRef](#)]
37. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bai, G.; Yang, J.; Zhao, G. Plastid Genome Comparative and Phylogenetic Analyses of the Key Genera in *Fagaceae*: Highlighting the Effect of Codon Composition Bias in Phylogenetic Inference. *Front. Plant Sci.* **2018**, *9*, 82. [[CrossRef](#)] [[PubMed](#)]
38. Yang, J.; Takayama, K.; Youn, J.S.; Pak, J.H.; Kim, S.C. Plastome Characterization and Phylogenomics of East Asian Beeches with a Special Emphasis on *Fagus multinervis* on Ulleung Island, Korea. *Genes* **2020**, *11*, 1338. [[CrossRef](#)]
39. Mader, M.; Schroeder, H.; Schott, T.; Schoning-Stierand, K.; Leite Montalvao, A.P.; Liesebach, H.; Liesebach, M.; Fussi, B.; Kersten, B. Mitochondrial Genome of *Fagus sylvatica* L. as a Source for Taxonomic Marker Development in the Fagales. *Plants* **2020**, *9*, 1274. [[CrossRef](#)] [[PubMed](#)]
40. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
41. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
42. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
43. Okonechnikov, K.; Golosova, O.; Fursov, M.; Team, U. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **2012**, *28*, 1166–1167. [[CrossRef](#)]
44. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.
45. Schirmer, M.; D’Amore, R.; Ijaz, U.Z.; Hall, N.; Quince, C. Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinform.* **2016**, *17*, 125. [[CrossRef](#)]
46. Mishra, B.; Ulaszewski, B.; Meger, J.; Pfenninger, M.; Gupta, D.K.; Wötzel, S.; Ploch, S.; Burczyk, J.; Thines, M. A chromosome-level genome assembly of the European Bee (*Fagus sylvatica* L.) reveals anomalies for organelle DNA integration, repeat content and distribution of SNPs. *bioRxiv* **2021**. [[CrossRef](#)]
47. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.
48. Soubrier, J.; Steel, M.; Lee, M.S.; Der Sarkissian, C.; Guindon, S.; Ho, S.Y.; Cooper, A. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **2012**, *29*, 3345–3358. [[CrossRef](#)]

49. Yang, Z. A space-time process model for the evolution of DNA sequences. *Genetics* **1995**, *139*, 993–1005. [[CrossRef](#)] [[PubMed](#)]
50. Rosenberg, M.S.; Anderson, C.D. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis; Version 2. *Methods Ecol. Evol.* **2011**, *2*, 229–232. [[CrossRef](#)]
51. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220.
52. Magri, D. Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J. Biogeogr.* **2008**, *35*, 450–463. [[CrossRef](#)]
53. 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, *Fagus sylvatica* L. *J. Biogeogr.* **2017**, *44*, 2475–2487. [[CrossRef](#)]
54. Heuertz, M.; Fineschi, S.; Anzidei, M.; Pastorelli, R.; Salvini, D.; Paule, L.; Frascaria-Lacoste, N.; Hardy, O.J.; Vekemans, X.; Vendramin, G.G. Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Mol. Ecol.* **2004**, *13*, 3437–3452. [[CrossRef](#)]
55. Weising, K.; Gardner, R.C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **1999**, *42*, 9–19. [[CrossRef](#)] [[PubMed](#)]
56. Shrestha, B.; Gilbert, L.E.; Ruhlman, T.A.; Jansen, R.K. Clade-Specific Plastid Inheritance Patterns Including Frequent Biparental Inheritance in *Passiflora* Interspecific Crosses. *Int. J. Mol. Sci.* **2021**, *22*, 2278. [[CrossRef](#)] [[PubMed](#)]
57. Scarcelli, N.; Mariac, C.; Couvreur, T.L.; Faye, A.; Richard, D.; Sabot, F.; Berthouly-Salazar, C.; Vigouroux, Y. Intra-individual polymorphism in chloroplasts from NGS data: Where does it come from and how to handle it? *Mol. Ecol. Resour.* **2016**, *16*, 434–445. [[CrossRef](#)] [[PubMed](#)]