

## Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland

Michał Chmielewski<sup>(1)</sup>, Katarzyna Meyza<sup>(1)</sup>, Igor J Chybicki<sup>(1)</sup>, Artur Działuk<sup>(1)</sup>, Monika Litkowiec<sup>(2)</sup>, Jarosław Burczyk<sup>(1)</sup>

Assessing the distribution of chloroplast DNA (cpDNA) haplotype variation is useful for studying the phylogeography of angiosperms. In the last two decades the cpDNA phylogeography of white oaks in Europe has been extensively studied, mostly based on the PCR-RFLP technique. However, PCR-RFLPs have low mutation rates and are primarily useful for reconstructing patterns at large geographical scales and lack resolution at fine spatial scales. Here we evaluate the usefulness of chloroplast microsatellites (cpSSR) as an alternative to PCR-RFLPs in Polish oak populations which have been underrepresented in previous studies. Eighty-five cpSSR haplotypes were detected using 14 cpSSR loci and a broad collection of 6680 trees sampled throughout Poland. Haplotype diversity was significantly lower in *Q. petraea* ( $H_e = 0.798$ ) than in *Q. robur* ( $H_e = 0.820$ ). Only 17 haplotypes (H01-H17) were found in 13 or more individuals, comprising together 97.9% of the sample. Most frequent cpSSR haplotypes were related to PCR-RFLP haplotypes, establishing the cross-references between the two marker systems. There was significant concordance between the matrices of genetic distances obtained by PCR-RFLP haplotypes and cpSSR haplotypes. Phylogenetic relationships among cpSSR haplotypes supported the existence of the three predominant maternal lineages of oaks in Poland: Iberian (7.8%), Apennine (20.6%) and Balkan (65.5%). The results are discussed with regards to the usefulness of cpSSR markers for phylogeographic studies.

**Keywords:** Chloroplast Microsatellites, PCR-RFLP, White Oaks, Phylogeography

### Introduction

Studies of genetic diversity of chloroplast DNA (cpDNA) can provide important insights into phylogenetic and historical relationships among populations of angiosperms (Hewitt 1999, Petit et al. 2005). For example, the geographic distribution of intraspecific cpDNA variants has been widely used to define post-glacial migration routes and to identify zones of secondary contact (Hewitt 1999, Petit et al. 2002a). Additionally, cpDNA has found a practical application in marker-based tracking of forest reproductive material and timber (Degen et al. 2013). Both of these aspects are important for management and the conservation of forest genetic resources (Neophytou & Michiels 2013).

The most widespread species of Euro-

pean white oaks, pedunculate oak (*Quercus robur L.*) and sessile oak (*Quercus petraea* [Matt.] Libel.) grow naturally across a large part of Europe and are important components of forest ecosystems, with high economic value. The application of the PCR-RFLP technique has allowed the detection of multiple cpDNA haplotypes with specific geographic distributions across Europe, grouped into three major lineages originating from the locations of the primary glacial refugia (Iberia, Apennines, Balkans (Dumolin-Lapègue et al. 1997, Bordács et al. 2002, Fineschi et al. 2002, König et al. 2002, Olalde et al. 2002, Petit et al. 2002a, 2002b, 2005, Buitenveld & Koelewijn 2006, Gailing et al. 2007a, Moldovan & Sofletea 2010). However, these studies focused mostly on Western or Southern

European populations, and populations from Central and Eastern Europe (Poland, Baltic Countries, Slovakia and Czech Republic) have received relatively less attention (Csaikl et al. 2002, Tutkova-van Loo & Burg 2004). Within the Baltic region (54 locations from Estonia, Latvia, Lithuania and Poland) Csaikl et al. (2002) found 13 haplotypes. On the other hand, Dering et al. (2008), studying 78 locations in Poland (each consisting of 1 to 12 individuals) detected only 6 PCR-RFLP haplotypes representing the three major maternal lineages. Neither of these studies uniformly covered the distribution of both oak species in Poland, introducing a possible bias to the observed distribution of cpDNA variation in the region.

CpDNA variation can be alternatively mapped with microsatellite sequences (cpSSR - Deguilloux et al. 2004). Several universal cpSSR loci have been identified for angiosperm taxa (Weising & Gardner 1999). In addition, a number of loci have been designed specifically for oaks (Deguilloux et al. 2003) or in general for Fagaceae (Sebastiani et al. 2004). Since then, several phylogeographic studies have been done in oaks employing cpSSR markers (Grivet et al. 2006, Magri et al. 2007, Pakkad et al. 2008, Tovar-Sánchez et al. 2008). Microsatellites differ from PCR-RFLP markers because of their mutation mechanism and rate (Estoup et al. 2002). Therefore, they may reveal incomplete concordance with

□ (1) Institute of Experimental Biology, Department of Genetics, Kazimierz Wielki University of Bydgoszcz, Chodkiewicza 30, 85-064 Bydgoszcz (Poland); (2) Institute of Dendrology, Polish Academy of Sciences, ul. Parkowa 5, 62-035 Kórnik (Poland)

@ Jarosław Burczyk (burczyk@ukw.edu.pl)

Received: Feb 12, 2015 - Accepted: Jun 01, 2015

**Citation:** Chmielewski M, Meyza K, Chybicki IJ, Działuk A, Litkowiec M, Burczyk J (2015). Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. iForest 8: 765-771. - doi: 10.3832/ifor1597-008 [online 2015-07-19]

Communicated by: Andrea Piotti

PCR-RFLP when studying the genetic structure of populations. Nonetheless, direct relationships between cpSSR and PCR-RFLP haplotypes is expected since they are fully linked in the chloroplast genome, as previously reported (Deguilloux et al. 2004, Gailing et al. 2007b, Neophytou & Michiels 2013). However, PCR-RFLPs have low mutation rates and are primarily useful for reconstructing patterns at large geographical scales, while they lack the resolution needed for fine spatial scales studies.

Laboratory techniques for analyses of cpSSR markers are less demanding than for PCR-RFLP variants. CpSSRs, being usually amplified as short DNA fragments, are easily detected even in partially degraded DNA, and they are less sensitive to inhibitory substances present in plant tissues (Gailing et al. 2007a). They are relatively cost-effective and are easily multiplexed for high-throughput studies. However, cpSSR markers seem to be more sensitive to size homoplasy (Estoup et al. 2002, Vachon & Freeland 2011), which could complicate phylogenetic inferences. However, the level of homoplasy within species is usually considered low enough to allow population studies (Navascués & Emerson 2005).

Here, we assembled a large collection of *Q. petraea* and *Q. robur* trees (n=6680) obtained by different sampling schemes to achieve a representative coverage of the study region. Such a large sample enables the identification of cpSSR haplotype variants existing in Central Europe and can provide a reference for future studies. Therefore, in this paper the identification of polymorphic cpSSR loci within the study region was used to assess the level of genetic diversity in species and populations of oaks. We detected numerous multilocus haplotypes whose relatedness was examined. Additionally, we related the most frequent cpSSR haplotypes to the haplotypes identified in previous studies based on the PCR-RFLP technique. The protocol for an optimized high-throughput genotyping procedure is described for future research.

## Material and methods

### Plant material

A total of 6680 oak trees were sampled throughout Poland (Fig. S1 in Appendix 1). The samples included individuals from various sampling schemes: 2512 trees from certified seed stands (approx. 20 individuals from each of 132 forest stands), a set of 599 plus trees (phenotypically selected trees utilized in forest tree improvement programs), a separate collection of 3310 trees consisting of up to 5 individuals per species from 420 forest districts where the species were present (uniformly scattered sample), and a subset of 259 old trees (usually older than 250 years) growing individually or in road alleys, some being considered as national nature monuments. Overall, the collection comprised 3938 indi-

viduals of *Quercus robur* and 2742 individuals of *Q. petraea*. The species status of individuals was determined based on leaf morphology at the time of sampling, and whenever possible (e.g., for plus trees and seed stands) also verified in the State Forest Reproductive Material Office.

### DNA isolation

Total genomic DNA was extracted from 50 mg of dried leaves. Plant material was ground in a Mixer Mill MM301 (Retsch, Haan, Germany). DNA isolation followed a CTAB protocol (Doyle & Doyle 1990). The amount and quality of DNA was evaluated and adjusted to a concentration of 10 ng/ $\mu$ l using DNA calculator (Biophotometer<sup>®</sup>, Eppendorf, Hamburg, Germany).

### Chloroplast microsatellites

For the initial screening of polymorphism at cpSSR loci we used a representative subsample of individuals consisting of DNA from 282 plus trees (198 *Q. robur*, 84 *Q. petraea*) from 64 Forest Districts in Poland, widely distributed across the country. We selected 35 cpSSR loci that had been used in population genetic studies of forest trees, especially in oaks (see Table S1 in Appendix 1), including 8 markers developed for dicotyledonous angiosperms (Weising & Gardner 1999), 16 made specifically for oaks (Deguilloux et al. 2003) and 11 created for Fagaceae (Sebastiani et al. 2004). The loci were divided into four marker sets, each consisting of 6-11 loci, and subjected to PCR-multiplex protocols (Table S1 in Appendix 1). However, among the 35 pre-selected cpSSR loci only 14 (*ccmp4*, *cmcs12*, *cmcs5*, *cmcs6*, *cmcs7*, *cmcs8*, *cmcs9*, *ucd1*, *ucd4*, *ucd5*, *udt1*, *udt3*, *udt4*, *ukk4*) were polymorphic in the test subsample. The set of 14 loci was subjected to PCR-multiplex optimization in order to establish a single PCR reaction. Finally the optimized reaction mix contained: 1x PCR Buffer, 1.25 mM MgCl<sub>2</sub>, 0.25 mM dNTP-Mix, 0.25 U Taq polymerase (Taq PCR Core Kit<sup>®</sup>, Qiagen, Hilden, Germany), 0.5 mg/ml BSA, 0.2  $\mu$ M *ucd1*, 0.125  $\mu$ M *udt3*, 0.075  $\mu$ M *cmcs8*, 0.15  $\mu$ M *cmcs9*, 0.1  $\mu$ M *ccmp4*, 0.125  $\mu$ M *cmcs7*, 0.05  $\mu$ M *ucd5*, 0.05  $\mu$ M *ucd1*, 0.125  $\mu$ M *cmcs5*, 0.075  $\mu$ M *cmcs12*, 0.05  $\mu$ M *udt1*, 0.15  $\mu$ M *ukk4*, 0.25  $\mu$ M *udt4*, 0.15  $\mu$ M *cmcs6* (concentration refers to that of both primers) and 10 ng of template DNA. Primers from these loci were fluorescently labeled (see Table S1 in Appendix 1). PCR conditions were as follows: an initial denaturation step at 95°C for 5 min, followed by 8 touchdown cycles at 94°C for 30 s, 1:30 s at 56 °C (-1 °C/cycle), 1 min at 72 °C, 24 cycles at 94 °C for 30 s, 45 s at 48 °C, 1 min at 72 °C, and a final extension at 72 °C for 10 min. The amplification products were separated using an ABI 3130XL sequencer (Applied Biosystems, Foster City, CA, USA), with LIZ600 as an internal size standard. The identification of alleles based on their size was determined using GENEMAPPER<sup>®</sup> software v. 4.0 (Applied Biosystems). The accuracy of genotyping ba-

sed on the 14-plex PCR reaction was verified for 16 randomly chosen individuals, by repeating single PCR reactions of each locus. This indicated full concordance of allele sizes determined in multiplex and singleplex reactions. Finally, the 14-plex PCR protocol was used to analyze all 6680 samples.

### PCR-RFLP haplotypes

PCR-RFLP haplotypes were also determined on a subset of 282 individuals in order to relate cpSSR haplotypes to the PCR-RFLP haplotypes identified in earlier studies (Csaiik et al. 2002, Petit et al. 2002b, Dering et al. 2008). PCR-RFLP variants were assayed following the protocol described by Dering et al. (2008) and references therein. The interpretation of the restriction pattern followed Petit et al. (2002b).

### Data analysis

For analysis of cpSSR data, each species was treated as single population. Allelic frequencies, number of alleles and genetic diversity were calculated using the FSTAT software (Goudet 2001). Haplotypes were determined as a combination of different microsatellite variants across the cpSSR loci. We used the software HAPLOTYPE ANALYSIS<sup>®</sup> ver. 1.05 (Eliades & Eliades 2009) to estimate haplotype frequencies and the following genetic diversity measures: number of haplotypes (A), effective number of haplotypes ( $N_e$ ), haplotypic richness ( $H_r$  – El Mousadik & Petit 1996), genetic diversity ( $H_e$ ), and the mean genetic distance between individuals ( $D_{sh}^2$  – Goldstein et al. 1995).

The phylogeny of the established cpSSR haplotypes was inferred based on the median-joining method (Bandelt et al. 1999). Maximum-parsimony analysis was conducted using the software NETWORK ver. 4.6.1.2 (Fluxus Technology Ltd, <http://www.fluxus-engineering.com>). Additionally, phylogenetic relationships were assessed based on genetic distances between individual haplotypes (Goldstein et al. 1995), using the neighbor-joining algorithm, implemented in the software package TREEVIEW (Page 1996).

After establishing the correspondence between PCR-RFLP and cpSSR haplotypes based on the subset of 282 plus trees, the matrix of genetic distances between PCR-RFLP haplotypes presented in Kremer et al. (2002), was compared to the matrix of genetic distances between cpSSR haplotypes (Goldstein et al. 1995) using a Mantel test implemented in the R package “ecodist” (Goslee & Urban 2007). The significance of the Mantel test was evaluated with 100 000 random permutations.

## Results

### Genetic diversity

Overall, we detected 45 alleles among 6680 oak individuals based on 14 cpSSR loci. The number of alleles per locus ranged

from 2 (*cmcs7*, *μcd1*, *μcd5*, *μkk4*) to 5 (*cmcs6* - Tab. 1), with an average of  $3.214 \pm 0.261$  (SE). Genetic diversity of individual loci ranged from 0.001 to 0.525, with a mean of  $0.281 \pm 0.052$ . Although the number of alleles found was greater in *Q. robur*, the genetic diversity was generally higher in *Q. petraea* (Tab. 1). Genetic diversity at individual loci revealed high interspecific consistency ( $r=0.946$ ;  $p<0.001$ ), indicating that these loci can be equally informative about genetic diversity in both species.

#### Haplotype diversity

The use of 14 cpSSR loci resulted in the detection of 85 different haplotypes across both species. The number of haplotypes was higher in *Q. robur* (67) than *Q. petraea* (47 - Tab. 2). Haplotypes were ranked based on their observed frequencies across both species and labels (H01, H02, etc.) were assigned according to their frequency in descending order, i.e., starting from the most frequent haplotype. The frequency spectrum of the haplotypes was highly skewed, with H49-H85 found in just one individual, whilst each of H34-H48 occurred in just two. Seventeen haplotypes (H01-H17) were found in 13 or more individuals; 97.9% of the collection had one of these 17 haplotypes. Frequency dropped off rapidly for haplotypes lower than H10 in the ranking. Full details of all multilocus haplotypes are presented in Appendix 1 (Table S2). The 17 most frequent haplotypes could be characterized using a subset of 5 cpSSR loci only, as shown in Tab. 3. Some 25 haplotypes were shared by both species, with 39 and 21 (mostly rare haplotypes) found exclusively in *Q. robur* and *Q. petraea*, respectively. The effective number of haplotypes was 5.543 and 4.936 in *Q. robur* and *Q. petraea*, respectively, with a mean of 5.367 estimated across both species. The effective number of haplotypes was significantly lower in *Q. petraea* than in *Q. robur* ( $p < 0.001$  after permutation test with individual haplotypes taken as units). Haplotype diversity was lower in *Q. petraea* ( $H_e = 0.798$ ) than in *Q. robur* ( $H_e = 0.820$  -  $p < 0.001$  after permutation test). The mean genetic distances between individuals within species ( $D_{sh}^2$ ) were comparable (Tab. 2).

#### Relationship between PCR-RFLP and cpSSR haplotypes

Haplotypes based on cpSSR loci were compared to the PCR-RFLP haplotypes in the subset of 282 trees. Six different PCR-RFLP haplotypes belonging to the three major maternal lineages (Apennine, Iberian, Balkan) were found, with each haplotype present at least in 3 individuals (Tab. 4). The cpSSR haplotypes H02 and H07 could be assigned to the PCR-RFLP haplotypes RFLP1 and RFLP2, respectively, both belonging to the Apennine lineage C (Petit et al. 2002b). The haplotype H05 represents most likely the haplotype RFLP12, which is part of the Iberian maternal linea-

**Tab. 1** - Genetic diversity of cpSSR loci based on *Quercus robur* and *Q. petraea*. (A): number of alleles; ( $H_e$ ): genetic diversity; (SE): standard error.

Locus	<i>Quercus robur</i> (n=3938)		<i>Quercus petraea</i> (n=2742)		Both species (n=6680)	
	A	$H_e$	A	$H_e$	A	$H_e$
<i>cmcs12</i>	3	0.002	1	0.000	3	0.001
<i>cmcs5</i>	3	0.300	3	0.401	3	0.346
<i>cmcs6</i>	5	0.399	5	0.462	5	0.429
<i>cmcs7</i>	2	0.003	1	0.000	2	0.001
<i>cmcs8</i>	3	0.028	3	0.032	3	0.030
<i>μdt1</i>	4	0.405	4	0.445	4	0.421
<i>μdt3</i>	4	0.447	4	0.494	4	0.472
<i>μdt4</i>	4	0.394	3	0.523	4	0.454
<i>ccmp4</i>	3	0.299	2	0.175	3	0.252
<i>cmcs9</i>	4	0.395	3	0.523	4	0.454
<i>μcd1</i>	2	0.083	2	0.018	2	0.057
<i>μcd4</i>	4	0.525	3	0.506	4	0.525
<i>μcd5</i>	2	0.138	2	0.183	2	0.157
<i>μkk4</i>	2	0.278	2	0.395	2	0.331
Mean	3.214	0.264	2.714	0.296	3.214	0.281
SE	0.261	0.048	0.304	0.058	0.261	0.052

ge B. Finally, the most frequent Balkan maternal lineage A consisted of three RFLP haplotypes: RFLP4 (related mostly to cpSSR haplotypes H03 and H06), RFLP7 (H01) and RFLP5 (related mostly to H04, H06, H10 - Tab. 4). Notably, while RFLP1, RFLP2 and RFLP12 haplotypes were connected in general to single cpSSR haplotypes, the Balkan PCR-RFLP haplotypes were related to several different cpSSR haplotypes. In particular, haplotype RFLP5 appeared highly heterogeneous and was related to nine different cpSSR haplotypes, which included quite divergent haplotypes such as H04, H06, H08, H09, H10 and H12 (the effective number of cpSSR haplotypes within RFLP5 haplotype was 2.889 - see Tab. 4). The spatial distribution of PCR-RFLP haplotypes generally followed the distributions observed in earlier studies (Csaikl et al. 2002, Dering et al. 2008, Chmielewski et al., in preparation).

We used a Mantel test to compare the matrix of genetic distances between PCR-

**Tab. 2** - Measures of haplotypic diversity for *Q. robur* and *Q. petraea* based on 14 cpSSR loci, and the subset of 5 loci (*ccmp4*, *cmcs5*, *cmcs6*, *μcd4*, *μdt1*; see the text).

Parameter	<i>Quercus robur</i>	<i>Quercus petraea</i>	Both species	Both species (subset of 5 loci)
Number of individuals (N)	3938	2742	6680	6680
Number of haplotypes (A)	67	47	85	43
Effective number of haplotypes ( $N_e$ )	5.543	4.936	5.367	5.267
Genetic diversity ( $H_e$ )	0.820	0.798	0.814	0.810
Mean genetic distance between individuals ( $D_{sh}^2$ )	2.589	2.629	2.618	1.924

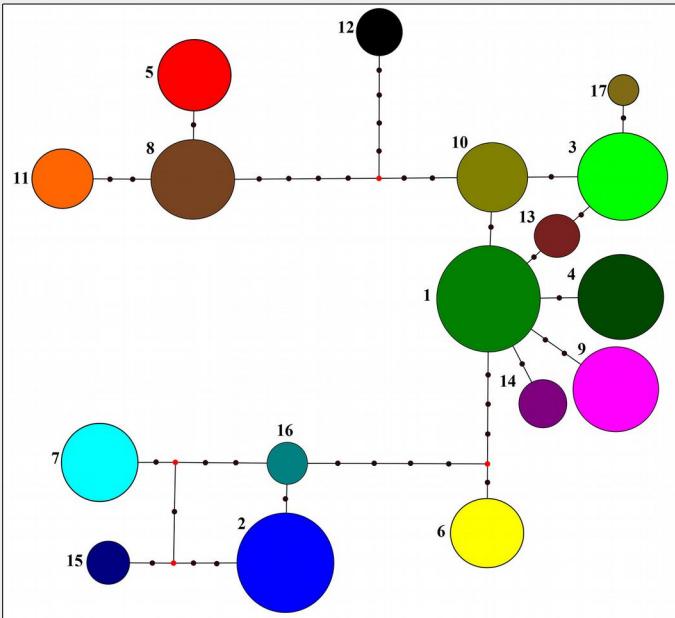
**Tab. 3** - Genotypes of the most frequent haplotypes based on the subset of 5 loci (*ccmp4*, *cmcs5*, *cmcs6*, *μcd4*, *μdt1*). Allele names are given based on their size (bp).

Haplotype	<i>ccmp4</i>	<i>cmcs5</i>	<i>cmcs6</i>	<i>μcd4</i>	<i>μdt1</i>	Count	Freq
H01	147	200	79	114	96	2219	0.3281
H02	146	198	79	114	97	1179	0.1749
H03	147	200	78	114	97	1169	0.1731
H04	147	200	79	115	96	767	0.1135
H05	147	200	81	114	97	302	0.0442
H06	147	197	79	114	96	245	0.0366
H07	146	198	79	115	95	176	0.0261
H08	147	200	80	114	97	168	0.0245
H09	147	201	79	114	96	124	0.0178
H10	147	200	79	114	97	109	0.0154
H11	147	199	80	114	97	59	0.0088
H12	148	200	79	114	96	28	0.0037
H13	147	200	78	114	96	22	0.0027
H14	146	200	79	114	96	18	0.0025
H15	146	197	79	115	97	18	0.0024
H16	146	198	79	114	96	16	0.0024
H17	146	200	78	114	97	13	0.0019
<i>Sum</i>						6632	0.9904

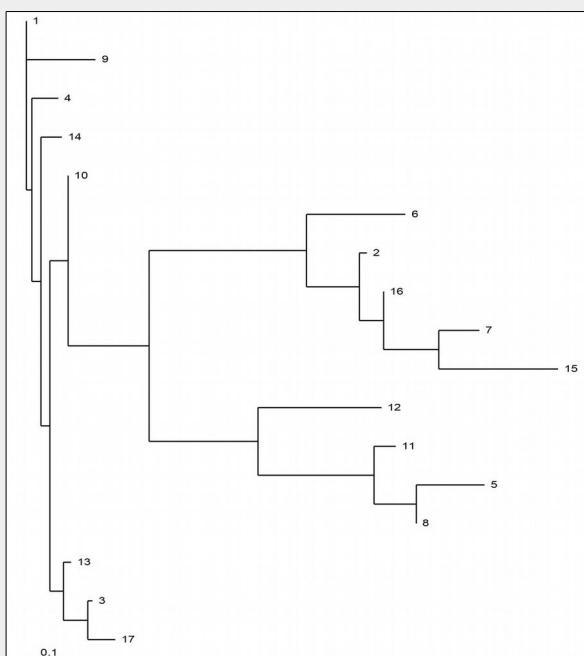
**Tab. 4** - Number of individuals with specific combinations of PCR-RFLP and cpSSR haplotypes. (\*): indicate the most frequent PCR-RFLP haplotype within a given cpSSR haplotype. ( $AE_{cpSSR}$ ): effective number of PCR-RFLP haplotypes within a given cpSSR haplotype; ( $AE_{RFLP}$ ): effective number of cpSSR haplotypes within a given PCR-RFLP haplotype.

Haplo	RFLP1	RFLP2	RFLP4	RFLP5	RFLP7	RFLP12	Count	$AE_{cpSSR}$
H01	-	-	1	3	95*	-	99	1.085
H02	42*	-	-	-	-	-	42	1.000
H03	1	-	31*	-	-	-	32	1.064
H04	-	-	-	51*	1	-	52	1.039
H05	-	-	-	1	-	3*	4	1.600
H06	-	-	5	11*	-	-	16	1.753
H07	1	7	1	-	-	-	9	1.588
H08	-	-	-	3*	1	-	4	1.600
H09	-	-	-	5*	-	-	5	1.000
H10	-	-	-	12*	-	-	12	1.000
H12	-	-	1	2*	-	-	3	1.800
H25	-	-	-	4*	-	-	4	1.000
Count	44	7	39	92	97	3	282	-
$AE_{RFLP}$	1.096	1.000	1.538	2.889	1.042	1.000	-	-

**Fig. 1** - Maximum parsimony phylogenetic network among the 17 most frequent haplotypes ( $n \geq 13$ ) for 14 cpSSR loci. Circle sizes are proportional to the haplotype frequency. Red dots (median vectors) and black dots represent missing or not sampled haplotypes.



**Fig. 2** - Neighbor-joining tree of cpDNA haplotypes for the 17 most frequent haplotypes ( $n \geq 13$ ) based on pairwise genetic distances (Goldstein et al. 1995) estimated from 14 cpSSR loci (2000 bootstraps).



RFLP haplotypes presented in Kremer et al. (2002) to the matrix of genetic distances between cpSSR haplotypes related to the respective PCR-RFLP haplotypes, as established in this study based on Goldstein et al. (1995), e.g., the distance between RFLP1 and RFLP7 was compared to the distance between H01 and H02 (see Tab. 4). The test revealed a significant correlation between the two types of genetic distances (Pearson's  $r = 0.570$ ,  $p < 0.05$ ), suggesting that phylogenetic relationships estimated by cpSSR haplotypes were fairly similar to those obtained by PCR-RFLP haplotypes.

#### Phylogenetic relationship among cpSSR haplotypes

The maximum parsimony haplotype network based on 14 cpSSR loci for the 17 most frequent haplotypes ( $n \geq 13$ ) is presented in Fig. 1. The haplotypes clustered in three major clades. The largest and centrally located group includes 8 haplotypes: H01, H03, H04, H09, H10, H13, H14, H17, with a cumulative frequency of 65.5%. This group is related to the Balkan maternal lineage A, based on the above-mentioned relationships between PCR-RFLP and cpSSR variants. Another group consisted of 4 haplotypes H02, H07, H15 and H16 (cumulative frequency: 20.6%) representing the Apennine maternal lineage C; while 3 haplotypes, H05, H08 and H11 (cumulative frequency: 7.8%) were assigned to the Iberian lineage B. The haplotypes H06 and H12 seem to be intermediate between the lineages A and C and between lineages A and B, respectively (Fig. 1). The inferred phylogenetic relationships between haplotypes were confirmed when measured based on genetic distances (Goldstein et al. 1995) using the neighbor-joining algorithm (Fig. 2).

Each of the rare cpSSR haplotypes (H18-H85) might be related to one of the most frequent haplotypes (H1-H17 – Fig. S2 in Appendix 1). When we assign all haplotypes to one of five major haplotype groups (concentrated around most frequent haplotype lineages H01, H02, H05, H06, H12), the number of different haplotypes within each group is proportional to the number of individuals observed within haplogroups ( $r = 0.978$ ,  $p = 0.004$ ). Similarly, when we consider nine haplogroups (given that the largest major haplotype lineage is divided into smaller haplogroups H01, H03, H04, H09, H10), the number of haplotypes within such haplogroups is still proportional to the number of individuals observed within haplogroups ( $r = 0.772$ ,  $p = 0.015$ ). This suggests that the number of different haplotypes is related to the overall sample size of a haplogroup.

#### Optimization

The use of all 14 cpSSR loci resulted in the detection of 85 different haplotypes over the two species analyzed. However, as mentioned above, the majority of haplotypes were found in just one or two individ-

duals, and only 17 haplotypes (H01-H17) were found in 13 or more individuals. Therefore, we optimized the genotyping procedure in order to minimize the number of cpSSR loci to be used for screening, whilst still being able to detect the 17 most frequent haplotypes. We tested various combinations of loci and found that the 17 most frequent haplotypes could be characterized using just 5 loci (*ccmp4*, *cmc55*, *cmc56*, *μcd4*, *μdt1*). With these five loci we were able to detect 43 different haplotypes (Tab. 3). The optimized set of loci will be valuable for future research and for practical, marker-based forest management (e.g., tracking of forest reproductive material or gene conservation).

## Discussion

Studies using cpSSR markers have usually reported higher levels of genetic variation than studies based on the PCR-RFLP technique (Provan et al. 1999). Chloroplast microsatellites have already been used for studying phylogeography in oaks (Grivet et al. 2006, Magri et al. 2007, Pakkadi et al. 2008), revealing patterns of variation similar to PCR-RFLP markers (Deguilloux et al. 2004, Gailing et al. 2007a). However, recent studies of cpSSR diversity in European white oaks based on chloroplast polymorphisms indicated a smaller number of haplotypes compared with this study (Grivet et al. 2006, Neophytou & Michiels 2013).

In this paper, 14 cpSSR loci were analyzed to identify 85 distinct haplotypes on a large number of individuals sampled across Poland. Such a large number of cpSSR haplotypes contrasts with the amount of PCR-RFLP haplotypes found earlier in the region (Csaikl et al. 2002, Dering et al. 2008). However, the majority of haplotypes were rare, with frequencies below 0.001. The number of cpSSR haplotypes was larger for *Q. robur* than for *Q. petraea* (67 vs. 47, respectively), reflecting the overall trend found across Europe based on PCR-RFLP markers (23 vs. 17 - Petit et al. 2002b). Sharing of haplotypes between the two oak species confirms their close phylogenetic relationship and/or ability to hybridize (Dumolin-Lapègue et al. 1997, Petit et al. 2002b).

To date, only a few studies have attempted to relate chloroplast haplotypes based on cpSSRs and PCR-RFLP markers. Defining such relatedness is necessary if new phylogeographic studies based on cpSSRs are to be compared to earlier works carried out using PCR-RFLPs. Deguilloux et al. (2004) found almost complete redundancy between the haplotypes identified using the two types of markers in French oak populations. Other studies related haplotypes from the two marker systems only partially (Gailing et al. 2007b, Neophytou & Michiels 2013).

The analysis of the phylogenetic relationships among the 17 most frequent haplotypes ( $n > 13$ ) and their associations to PCR-RFLP haplotypes confirmed the existence of the three major clades related to the

main maternal lineages previously described in this area (Petit et al. 2002a, Csaikl et al. 2002, Dering et al. 2008). The most frequent one appeared to be the Balkan maternal lineage A (nearly 64%), followed by the Apennine lineage C (21%) and the Iberian lineage B (8%). The cpSSR haplotypes H06 and H12 were found to be intermediate between different clades (H06 intermediate between Balkan and Apennine, H12 intermediate between Balkan and Iberian lineages - Fig. 1), but both were assigned to haplotype RFLP5 which represents the Balkan lineage. While PCR-RFLP haplotypes of the Apennine and Iberian lineages were largely monotypic as for the number of cpSSR haplotypes found, the Balkan lineage included several cpSSR haplotypes. In particular, haplotype RFLP5 appeared highly heterogeneous as it was related to seven cpSSR haplotypes. Previously, Petit et al. (2002b) and Bordács et al. (2002) suggested that this haplotype might have persisted in both the Apennine and Balkan refugia. In Slovakia, Tutkova-van Loo & Burg (2004) found that haplotype RFLP5 could be resolved as three related variants (5A, 5B, and 5C).

We have found a large number of rare haplotypes (with frequency  $< 0.001$ ), which were the result of mutations in one of the 14 cpSSR loci. Petit et al. (2002b) recorded numerous rare PCR-RFLP haplotypes; however, these haplotypes were not confirmed by the reference laboratory. Rare haplotypes are generally of minor importance in phylogeographic studies, but they might be useful for identifying specific populations when showing an increased local frequency. An assessment of the spatial distribution of rare haplotypes (local or widely dispersed) might provide insights into the spread dynamics of new cpDNA variants.

The high mutation rates of cpSSR regions increase the potential for homoplasy. This may cause problems in phylogenetic studies, especially when performed at a very large spatial scale (Doyle et al. 1998, Provan et al. 2001, Estoup et al. 2002). However, such problem should be negligible for studies at small geographical scales. Furthermore, homoplasy level within species is usually considered as low enough to allow population genetic analysis (Navascués & Emerson 2005). In our study, we found similar relatedness patterns among PCR-RFLP and cpSSR haplotypes. A given cpSSR haplotype was related mostly to a single dominant PCR-RFLP haplotype, suggesting that our data set was affected by homoplasy to a limited degree. Similarly, a given PCR-RFLP haplotype was mostly related to a single dominant cpSSR haplotype, indicating some cryptic diversity within PCR-RFLP haplotypes. A distinct exception, haplotype RFLP5 (see above), needs further investigation.

The number of distinct haplotypes detected in the study area is expected to depend on two main factors. Firstly, it should be related to the effective size of a

migrating population. Secondly, it may depend on the migration distance from the refugium. As a result of long-distance migration, when effective population size decreases dramatically (bottleneck), rare haplotypes tend to be lost. On the other hand, new haplotypes might be generated through mutation. The theory predicts that the number of different variants in a sample is a function of the effective size and the mutation rate (Kimura & Ohta 1975). Assuming that mutation rate is constant for a given locus, the number of haplotypes in our sample should be determined mainly by the effective population size. As a consequence, the prevalence of specific haplotypes centered on haplotype H01 can be explained by historically-determined, unbalanced representation of specific maternal lineages (refugia). It should be noticed that our sampling was fairly uniform across the country (Fig. S1 in Appendix 1), thus the number of haplotypes within the maternal lineages was not affected by the clustered distribution of lineages (Dering et al. 2008).

We found that as few as five cpSSR loci were necessary to characterize most frequent haplotypes in white oaks, indicating some redundancy among loci. This set of five loci could serve as a basis for standard assessment of haplotypic diversity for gene conservation or forest reproductive material (FRM) traceability surveys. However, using the established multiplex of 14 loci may be helpful in detecting new local haplotypes, which could be used for genetic fingerprinting of local FRM or timber.

## Conclusions

Chloroplast microsatellites, with their amenability to multiplex PCR and automation can be considered markers of choice for population genetic studies in white oaks. Our study confirmed the relationship between haplotypes established with cpSSR and PCR-RFLP polymorphisms, making possible comparisons with previous studies. CpSSR markers have the potential to detect a relatively larger number of haplotypes, which in turn might cause problems in interpretation of results. Nevertheless, occasional rare haplotypes can be easily "binned" with related, more frequent haplotypes to simplify analysis. In-depth analysis of the spatial distribution of cpSSR haplotypes of the two oak species should provide a detailed picture of the current distribution of maternal lineages. This will be helpful in reconstructing the migratory routes of the species and potentially would allow the detection of local population differentiation.

## Acknowledgements

We thank Dr. Stephen Cavers and two anonymous reviewers for valuable comments and suggestions, which significantly improved the paper. The study was supported by the Ministry of Science and Higher Education of the Republic of Poland

(Grant No: NN309706040). The authors are thankful to Ewa Sztupecka for laboratory assistance, and to Lukasz Kubera and the personnel of State Forests of Poland for their help in collecting oak samples in the field.

MC, JB planned the research and wrote the article. MC, IJC and JB contributed to statistical analyses. JB obtained funding and organized the sample collection. AD organized uniformly scattered sample collection. ML performed PCR-RFLP analyses. MC, KM and IJC optimized cpSSR workflow and designed PCR multiplexes. MC and KM performed cpSSR laboratory analyses.

## References

- Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37-48. - doi: [10.1093/oxfordjournals.molbev.a026036](https://doi.org/10.1093/oxfordjournals.molbev.a026036)
- Bordács S, Popescu F, Slade D, Csaikl UM, Lesur I, Borovics A, Kézdy P, König AO, Gömöry D, Brewer S, Burg K, Petit RJ (2002). Chloroplast DNA variation of white oaks in northern Balkans and in the Carpathian Basin. *Forest Ecology and Management* 156 (1-3): 197-209. - doi: [10.1016/S0378-1127\(01\)00643-0](https://doi.org/10.1016/S0378-1127(01)00643-0)
- Buiteveld J, Koelewijn HP (2006). CpDNA haplotype variation reveals strong human influence on oak stands of the Veluwe forest in the Netherlands. *Forest Ecology and Management* 228: 160-167. - doi: [10.1016/j.foreco.2006.03.017](https://doi.org/10.1016/j.foreco.2006.03.017)
- Csaikl UM, Burg K, Fineschi S, König AO, Mátyás G, Petit RJ (2002). Chloroplast DNA variation of white oaks in the alpine region. *Forest Ecology and Management* 156: 131-145. - doi: [10.1016/S0378-1127\(01\)00639-9](https://doi.org/10.1016/S0378-1127(01)00639-9)
- Degen B, Ward SE, Lemes MR, Navarro C, Cavers S, Sebbenn AM (2013). Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. *Forensic Science International: Genetics* 7 (1): 55-62. - doi: [10.1016/j.fsgen.2012.06.003](https://doi.org/10.1016/j.fsgen.2012.06.003)
- Deguilloux MF, Dumolin-Lapègue S, Gielly L, Grivet D, Petit RJ (2003). A set of primers for the amplification of chloroplast microsatellites in *Quercus*. *Molecular Ecology Notes* 3: 24-27. - doi: [10.1046/j.1471-8286.2003.00339.x](https://doi.org/10.1046/j.1471-8286.2003.00339.x)
- Deguilloux MF, Pemonge MH, Petit RJ (2004). Use of chloroplast microsatellites to differentiate oak populations. *Annals of Forest Science* 61: 825-830. - doi: [10.1051/forest:2004078](https://doi.org/10.1051/forest:2004078)
- Dering M, Lewandowski A, Ufnalski K, Kedzierska A (2008). How far to the east was the migration of white oaks from the Iberian refugium? *Silva Fennica* 42 (3): 327-335. - doi: [10.14214/sf.240](https://doi.org/10.14214/sf.240)
- Doyle J, Doyle J (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Doyle JJ, Morgante M, Tingey SV, Powell W (1998). Size homoplasy in chloroplast microsatellite of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Molecular Biology and Evolution* 15: 215-218. - doi: [10.1093/oxfordjournals.molbev.a025916](https://doi.org/10.1093/oxfordjournals.molbev.a025916)
- Dumolin-Lapègue S, Demesure B, Fineschi S, Corre VL, Petit RJ (1997). Phylogeographic structure of white oaks throughout the European continent. *Genetics* 146: 1475-1487. [online] URL: <http://www.genetics.org/content/146/4/1475.short>
- El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* [L.] Skeels) endemic to Morocco. *Theoretical and Applied Genetics* 92: 832-839. - doi: [10.1007/BF00221895](https://doi.org/10.1007/BF00221895)
- Eliades NG, Eliades DG (2009). HAPLOTYPE ANALYSIS: software for analysis of haplotype data. *Forest Genetics and Forest Tree Breeding*, Georg-August University Goettingen, Germany. [online] URL: <http://www.uni-goettingen.de/en/134935.html>
- Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591-1604. - doi: [10.1046/j.1365-294X.2002.01576.x](https://doi.org/10.1046/j.1365-294X.2002.01576.x)
- Fineschi S, Taurchini D, Grossoni P, Petit RJ, Vendramin GG (2002). Chloroplast DNA variation of white oaks in Italy. *Forest Ecology and Management* 156: 103-114. - doi: [10.1016/S0378-1127\(01\)00637-5](https://doi.org/10.1016/S0378-1127(01)00637-5)
- Gailing O, Wachter H, Heyder J, Schmitt H-P, Finkeldey R (2007a). Chloroplast DNA analysis in oak stands (*Quercus robur* L.) in North Rhine-Westphalia with presumably Slavonian origin: is there an association between geographic origin and bud phenology? *Journal of Applied Botany and Food Quality* 81: 165-171. [online] URL: <http://pub.jki.bund.de/index.php/JABQ/article/view/2164>
- Gailing O, Wachter H, Schmitt H-P, Curtu A-L, Finkeldey R (2007b). Characterization of different provenances of Slavonian pedunculate oaks (*Quercus robur* L.) in Munsterland (Germany) with chloroplast DNA markers: PCR-RFLPs and chloroplast microsatellites. *Allgemeine Forst und Jagdzeitung* 178: 85-85. - [online] URL: <http://citeseerx.ist.psu.edu/view/doc/download?doi=10.1.1.459.3749&rep=rep&type=pdf#page=3>
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471. [online] URL: <http://www.genetics.org/content/139/1/463.short>
- Goslee SC, Urban DL (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22 (7): 1-19.
- Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Web site. [online] URL: <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Grivet D, Deguilloux MF, Petit RJ, Sork VL (2006). Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology* 15: 4085-4093. - doi: [10.1111/j.1365-294X.2006.03083.x](https://doi.org/10.1111/j.1365-294X.2006.03083.x)
- Hewitt GM (1999). Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society* 68: 87-112. - doi: [10.1111/j.1095-8312.1999.tb01160.x](https://doi.org/10.1111/j.1095-8312.1999.tb01160.x)
- Kimura M, Ohta T (1975). Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Sciences USA* 72 (7): 2761-2764. - doi: [10.1073/pnas.72.7.2761](https://doi.org/10.1073/pnas.72.7.2761)
- König AO, Ziegenhagen B, Van Dam BC, Csaikl UM, 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. *Forest Ecology and Management* 156: 147-166. - doi: [10.1016/S0378-1127\(01\)00640-5](https://doi.org/10.1016/S0378-1127(01)00640-5)
- Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducoussou A, König AO, Lowe AJ, Munro RC, Petit RJ, Stephan BR (2002). Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management* 156: 75-87. - doi: [10.1016/S0378-1127\(01\)00635-1](https://doi.org/10.1016/S0378-1127(01)00635-1)
- Magri D, Fineschi S, Bellarosa R, Buonamici A, Sebastiani F, Schirone B, Simeone MC, Vendramin GG (2007). The distribution of *Quercus* suber chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology* 16: 5259-5266. - doi: [10.1111/j.1365-294X.2007.03587.x](https://doi.org/10.1111/j.1365-294X.2007.03587.x)
- Moldovan IC, Sofletea N (2010). Chloroplast DNA diversity of oak species in Eastern Romania. *Notulæ Botanicæ Horti Agrobotanici Cluj-Napoca* 38: 302-307.
- Navascués M, Emerson C (2005). Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. *Molecular Ecology* 14: 1333-1341. - doi: [10.1111/j.1365-294X.2005.02504.x](https://doi.org/10.1111/j.1365-294X.2005.02504.x)
- Neophytou C, Michiels HG (2013). Upper Rhine valley: a migration crossroads of middle European oaks. *Forest Ecology and Management* 304: 89-98. - doi: [10.1016/j.foreco.2013.04.020](https://doi.org/10.1016/j.foreco.2013.04.020)
- Oalde M, Herrán A, Espinel S, Goicoechea PG (2002). White oaks phylogeography in the Iberian Peninsula. *Forest Ecology and Management* 156: 89-102. - doi: [10.1016/S0378-1127\(01\)00636-3](https://doi.org/10.1016/S0378-1127(01)00636-3)
- Page RDM (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.
- Pakkad G, Ueno S, Yoshimaru H (2008). Genetic diversity and differentiation of *Quercus semiserrata* Roxb. in northern Thailand revealed by nuclear and chloroplast microsatellite markers. *Forest Ecology and Management* 255: 1067-1077. - doi: [10.1016/j.foreco.2007.10.021](https://doi.org/10.1016/j.foreco.2007.10.021)
- Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, Van Dam B, Deans JD (2002a). Identification of refugia and postglacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49-74. - doi: [10.1016/S0378-1127\(01\)00634-X](https://doi.org/10.1016/S0378-1127(01)00634-X)
- Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J, Van Dam B, Deans JD, Dumolin-Lapègue S, Fineschi S (2002b). Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* 156: 5-26. - doi: [10.1016/S0378-1127\(01\)00645-4](https://doi.org/10.1016/S0378-1127(01)00645-4)
- Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689-701. - doi: [10.1111/j.1365-294X.2004.02410.x](https://doi.org/10.1111/j.1365-294X.2004.02410.x)

Provan J, Russell JR, Booth A, Powell W (1999). Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Molecular Ecology* 8: 505-511. - doi: [10.1046/j.1365-294X.1999.00545.x](https://doi.org/10.1046/j.1365-294X.1999.00545.x)

Provan J, Powell W, Hollingsworth PM (2001). Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution* 16: 142-147. - doi: [10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)

Sebastiani F, Carnevale S, Vendramin GG (2004). A new set of mono- and dinucleotide chloroplast microsatellites in Fagaceae. *Molecular Ecology Notes* 4: 259-261. - doi: [10.1111/j.1471-8286.2004.00635.x](https://doi.org/10.1111/j.1471-8286.2004.00635.x)

Tovar-Sanchez E, Mussali-Galante P, Esteban-Jimenez R, Pinero D, Arias DM, Dorado O, Oyama K (2008). Chloroplast DNA polymorphism reveals geographic structure and introgression in the *Quercus crassifolia* x *Quercus crassipes* hybrid complex in Mexico. *Botany* 86: 228-239. - doi: [10.1139/B07-128](https://doi.org/10.1139/B07-128)

Tutkova-van Loo M, Burg K (2004). Haplotype diversity of white oak species in Slovakia and the Czech Republic: results from PCR-RFLP analysis and phylogeographic interpretations. *Forest Genetics* 2: 125-138. [online] URL: [http://www.tuzvo.sk/files/fg/volumes/2004/FG11-2\\_12-5-137-str127\\_133.pdf](http://www.tuzvo.sk/files/fg/volumes/2004/FG11-2_12-5-137-str127_133.pdf)

Vachon N, Freeland JR (2011). Phylogeographic inferences from chloroplast DNA: quantifying the effects of mutations in repetitive and non-repetitive sequences. *Molecular Ecology Resources* 11: 279-285. - doi: [10.1111/j.1755-0998.2010.02921.x](https://doi.org/10.1111/j.1755-0998.2010.02921.x)

Weising K, Gardner RC (1999). A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42 (1): 9-19. - doi: [10.1139/g98-104](https://doi.org/10.1139/g98-104)

## Supplementary Material

### Appendix 1

**Tab. S1** - Characteristics of 35 preselected chloroplast microsatellites loci identified in a screening panel of 282 samples of *Q. robur* & *Q. petraea*.

**Tab. S2** - Genotypes of all haplotypes based on 14 chloroplast microsatellites (*cmcs12*, *cmcs5*, *cmcs6*, *cmcs7*, *cmcs8*, *μdt1*, *μdt3*, *μdt4*, *ccmp4*, *cmcs9*, *μcd1*, *μcd4*, *μcd5*, *μkk4*).

**Fig. S1** - Map of oaks from which samples were collected in Poland (green dots - *Quercus robur*, red dots - *Quercus petraea*).

**Fig. S2** - A possible maximum parsimony haplotype tree for all 85 haplotypes established based on 14 cpSSR loci. Grey circles indicate haplotypes H18-H85 detected in 7 or fewer individuals. Haplotypes H01-H17 indicated with the same colours as in Fig. 1.

Link: [Chmielewski\\_1597@suppl001.pdf](mailto:Chmielewski_1597@suppl001.pdf)