TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2014) 38: 1229-1238

doi:10.3906/bot-1405-80

Plastid and nuclear genomic resources of a relict and endangered plant species: *Chamaedaphne calyculata* (L.) Moench (Ericaceae)

Monika SZCZECIŃSKA[1,*], Angelika GOMOLIŃSKA[1], Piotr SZKUDLARZ[2], Jakub SAWICKI[1,3]

[1]Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland
[2]Department of Plant Taxonomy, Adam Mickiewicz University, Poznań, Poland
[3]Department of Biology and Ecology, University of Ostrava, Ostrava, Czech Republic

Received: 26.05.2014 • **Accepted:** 08.10.2014 • **Published Online:** 17.11.2014 • **Printed:** 28.11.2014

Abstract: *Chamaedaphne calyculata*, one of the largest shrubs in the family Ericaceae, is native to boreal zones of the northern hemisphere. The complete nucleotide sequence of the *Chamaedaphne* plastid genome was determined in the present study. The genome is a circular double-stranded DNA molecule 176,744 bp in length, that includes the typical large single-copy (LSC), small single-copy, and 2 inverted repeats (IR) regions. The *C. calyculata* plastid genome contains 113 genes, excluding the second IR region. A comparative analysis revealed that the plastid genome organization of the *Chamaedaphne* plastome is almost identical to that of *Vaccinium macrocarpon*, the first sequenced chloroplast genome from the family Ericaceae. The most profound changes were observed in regions IRa and IRb, which additionally contain *rpl*23 and *rps*14 genes and duplications of *trnf*M-CAU and *trn*G-UCC in LSC. Pairwise identity of chloroplast sequences in *C. calyculata* and *V. macrocarpon* was 79.6% and the p-distance was 0.032. A large number of nuclear and plastid microsatellites that could be useful for population genetics studies and phylogeographic research were also identified in the study.

Key words: Plastid genome, sequencing by synthesis, Ericaceae, *Chamaedaphne,* SSR markers, comparative genomics

1. Introduction

The evolutionary rates of organelle genomes vary significantly between taxonomic groups. Animal mitochondrial genomes constitute an excellent model for phylogenomic studies (Boore, 1999), whereas plant mitochondrial genomes are characterized by low variation in coding regions even between distant taxa (Duminil, 2014). Plastid genomes have a more dynamic structure, gene content, nucleotide sequences in coding and noncoding regions, and informative signals that are useful for the study of plant phylogenies. Typically, a circular plastid genome consists of 2 inverted repeats (IRa and IRb) separated by 2 regions of unique DNA: large (LSC) and small (SSC) single copy regions (Palmer and Stein, 1986; Jansen et al., 2005).

Plant chloroplast phylogenomics is a rapidly developing branch of science that employs next-generation sequencing methods. Initially, it was applied only to studies of higher taxa, but it is increasingly often used in analyses of taxa belonging to the same family or even genus (Zhang et al., 2011; Yang et al., 2013).

The family Ericaceae comprises around 3000 species belonging to 100 genera that are widely distributed around the world, excluding Australia. The largest 2 genera of the family Ericaceae are *Rhododendron*, which is most abundant in Asia (Merev and Yazuf, 2000), and *Erica*, which is characterized by the greatest species diversity in Southern Africa (Güvenç and Kendir, 2012). Ericaceae is a family of shrubs and climbers that occupy acidic, poor nutrient status habitats and overcome nutrient limitation through symbiosis with mycorrhizal fungi that mobilize nutrients in recalcitrant organic matter. Many species belonging to this family are endemic and threatened; the more that is known about them, the more they can be helped. One such species is *Chamaedaphne calyculata* (Kruszelnicki, 2001). *C. calyculata* (L.) Moench (Ericaceae), commonly known as leatherleaf, is a perennial shrub species reproducing sexually and also vegetatively by rhizomes (Figure 1). It is a circumboreal plant species distributed in North America, Asia, and the boreal and subarctic zones of Europe (Meusel et al., 1978). The species inhabits mostly unforested and raised bogs with acidic soils and high moisture content (i.e. *Sphagnetum magellanici* communities), but it is also found in *Vaccinio uliginosi–Pinetum* marshy coniferous forests (Kloss, 1996; Kruszelnicki, 2001).

In Central Europe *C. calyculata* is a rare and endangered species. It is considered to be a postglacial relict whose populations most probably spread from the southeastern

* Correspondence: monika.szczecinska@uwm.edu.pl

Figure 1. *Chamaedaphne calyculata* (L.) Moench located in Masurian Landscape Park.

refuge in Siberia (Kruszelnicki, 2001). Most of the Central European populations of this plant are small and isolated geographically. Moreover, the number of individuals per population has been declining steadily in the recent years due to an unstable level of ground waters, which hinders the preservation of *C. calyculata* habitats (Szkudlarz, 1995; Kloss, 1999; Kruszelnicki, 2001).

The effective protection of rare species lies in the conservation of their genetic resources. We previously estimated the level of genetic diversity of the Central European population of *C. calyculata* using nonspecific markers (ISSR and ISJ) (Szczecińska et al., 2009). In this study, we describe the complete plastid genome of *C. calyculata*. Due to maternal inheritance, details about plastid genome maybe useful for addressing a number of important biological questions about this plant species.

Our analysis led to the identification of microsatellite loci markers for *C. calyculata.* These markers are widely used in ecological studies and can also be used for investigating the genetic diversity among populations. The popularity of those markers stems from their near-ubiquity, high level of polymorphism, codominance and multiallelic variations. The genomic resources presented here should be useful for the analysis of population genetics and phylogeography, and will be also useful for conservation genetic studies of this rare and endangered species. The obtained complete plastid genome of *C. calyculata* will be the second completely identified plastid genome of the family Ericaceae after *Vaccinium macrocarpon* (Fajardo et al., 2013) and will shed some light on the levels of plastid genome variation in the family Ericaceae.

2. Materials and methods

2.1. Development of the genomic library and the sequencing procedure

The genomic library was developed based on the DNA extracted for a previous study (Szczecińska et al., 2009). A single individual from one of the largest populations of *C. calyculata* from the Masurian Landscape Park was used in the analysis. DNA was isolated from 40 mg of dry leaf tissue using the DNeasy Plant extraction kit (Qiagen). DNA quantity was estimated with the use of the Qubit fluorometer system (Invitrogen, USA) and the Quant-IT ds-DNA BR Assay kit (Invitrogen).

A genomic library for MiSeq sequencing was developed with the use of the Nextera XT Kit; 1 ng of DNA was used in the procedure described in the Nextera XT protocol (Illumina).

The number and correctness of libraries were verified with the use of primers whose sequences are given in the Sequencing *Library qPCR* Quantification Guide (Illumina). PCR reactions were performed in 20 µL of reaction mixture containing 3 µL of library genomes, 1.0 µM of each primer, 1.5 mM $MgCl_2$, 200 µL M dNTP (dATP, dGTP, dCTP, dTTP), 1X PCR buffer, and 1 U OpenExTaq polymerase (OpenExome). PCR reactions were performed under the following thermal conditions: (1) initial denaturation, 5 min at 94 °C, (2) denaturation, 30 s at 94 °C, (3) annealing, 30 s at 52 °C, (4) elongation, 1 min at 72 °C, and final elongation at 7 min at 72 °C. Stages 2–4 were repeated 34 times. The products of the PCR reaction were separated in the QIAxcel capillary electrophoresis system (Qiagen). Electrophoresis was performed using the QIAxcel High Resolution Kit with the 15–1000-bp alignment marker (Qiagen) and the 25–1000-bp DNA size marker (Qiagen). Standard OL500 settings were used as the electrophoresis program. Validated libraries were pooled according to the Nextera XT protocol. Genomic libraries were sequenced using the Miseq 500v2 cartridge that supported the acquisition of 2 × 250 bp pair-end reads. The resulting reads were preliminarily assembled using the Velvet de novo assembler (Zerbino and Birney, 2008) implemented in Illumina BaseCloud service.

2.2. Plastid genome assembly and annotation

The obtained contigs were mapped in the *V. macrocarpon* genome with the use of medium sensitivity settings implemented in Geneious 6.0.1 software (Drummond et

al., 2011). The longest mapped contigs of 6–12 kb were used to reconstruct the *C. calyculata* genome. All pair-end reads were mapped in the above sequence with the use of a modified medium sensitivity algorithm (with minimum overlap of 100 bp and 99% overlap identity) with 100–1000 iterations. Obtained large contigs were assembled de novo using Geneious 6.0.1 (Drummond et al., 2011) default settings. The assembled plastid genome did not have any gaps and the overlap size of the large contigs varied from 705 bp to 3670 bp. The flow chart for the in silico construction of *Chamaedaphne* plastid genome is presented in Figure 2.

The 4 junctions between the single-copy segments and the IRs were confirmed using PCR-based product sequencing of the preliminary assembled genomes. Purified PCR products were sequenced in both directions using ABI BigDye 3.1 Terminator Cycle Kit (Applied Biosystems) and then visualized using an ABI Prism 3130 Automated DNA Sequencer (Applied Biosystems). Sequences obtained using the Sanger method were aligned with the assembled genomes using Geneious 6.01 assembly software to determine if there were any differences.

Genome annotation was performed with the use of Geneious (Drummond et al., 2011), CpGavas (Liu et al., 2012), and DOGMA applications (Wyman et al., 2004). The assembled and annotated plastid genome of *C. calyculata* is available in GenBank (KJ463365). A circular genome map of the *C. calyculata* plastid genome was drawn in Organellar Genome DRAW (Lohse et al., 2007). Genome alignment was performed in Mauve version 2.3.1 (Darling et al., 2010).

The aligned genomes were used to calculate p-distance in MEGA 5.2 (Tamura et al., 2001) to evaluate the utility of complete plastid genome sequencing for phylogenomics research and species identification. Gaps were excluded from the analysis. The number of indels and substitutions in the noncoding regions of the plastid genome were calculated manually.

2.3. Microsatellite markers

Chloroplast microsatellite markers were identified independently for 3 plastid genome subunits (LSC, SSC, and IRa-b) in the Websat (Martins et al., 2009) application with default settings.

A procedure tested in previous studies (Sawicki et al., 2012; Szczecińska et al., 2012, 2013) was applied to nuclear microsatellites. Nuclear microsatellite regions were identified in the pool of contigs that had not been mapped in the plastid genome in the msatcommander application (Faircloth, 2008). Eight or more dinucleotide repeats, 6 or more trinucleotide repeats, and 4 or more tetranucleotide, pentanucleotide, and hexanucleotide repeats were taken into account.

3. Results and discussion

3.1. Sequencing results

A total of 2,901,406 pair-end reads (2 × 250 bp) from individual plants were obtained from a single run of the MiSeq sequencer. The *C. calyculata* plastid genome was assembled using 89,539 reads to produce an average coverage of 126.7-fold.

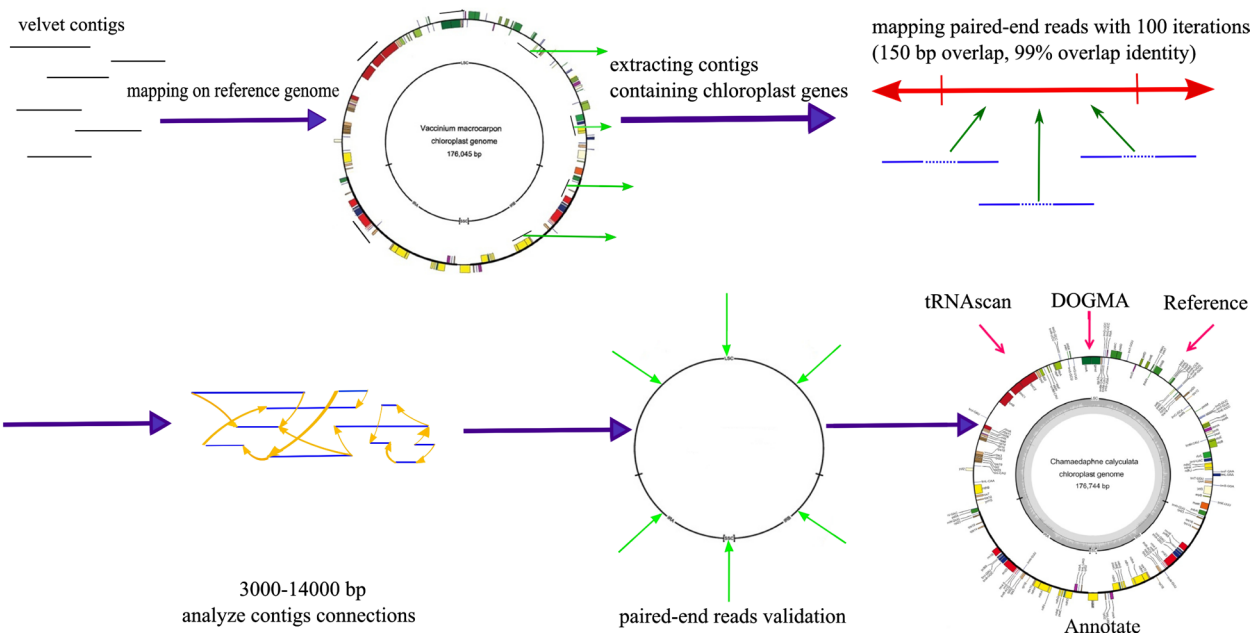

Figure 2. The scheme for in silico assembly and annotation of *C. calyculata* plastid genome.

3.2. Structure of the plastid genome

The complete *C. calyculata* plastid genome has the length of 176,744 base pairs (Figure 3), including an LSC region of 106,673 unique base pairs, separated from the SSC region of 2249 bp by 2 IRs of 33,911 bp each. The IR occurs between *trn*H-GUG and *rpl*32. The plastome encodes 113 genes, excluding the second IR region, corresponding to 79 protein-coding genes, 4 rRNAs, and 30 tRNAs (Table 1). Five protein-coding genes have introns and 4 are pseudogenes. The potential protein-coding genes are *mat*K in LSC and *ycf*68 in IR. Genes found in the plastid genome of *C. calyculata* are given in Table 2.

3.3. Comparative genomics

Chamaedaphne calyculata is the second species in the family Ericaceae for which the complete sequence of the chloroplast gene has been determined. Both sequenced genomes belong to the subfamily Vaccinioideae, which comprises 3 tribes: Vaccinieae, Andromedeae, and Gaultherieae (Kron et al., 2002). Despite considerable phylogenetic distance between *C. calyculata* and *V. macrocarpon* (Bush et al., 2009), the structure of the plastid genome is relatively similar in both taxa. The most profound changes were observed in regions IRA and IRB, which additionally contain *rpl*23 and *rps*14 genes. The *psb*A gene is present in 2 different copies, one highly similar (99%) to those found in the *Vaccinium* genome and the second 270 bp shorter. It should be noted that *rpl*23 and *rps*14 genes are still present in LSC. The LSC region of the chloroplast of *C. calyculata* in comparison with *V. macrocarpon* genome underwent only one structural change, a duplication of *trnf*M-CAU and *trn*G-UCC genes.

The annotation of the *C. calyculata* genome also revealed the presence of the protein-coding *rpl*20 gene, which was not reported in *V. macrocarpon*, but the alignment

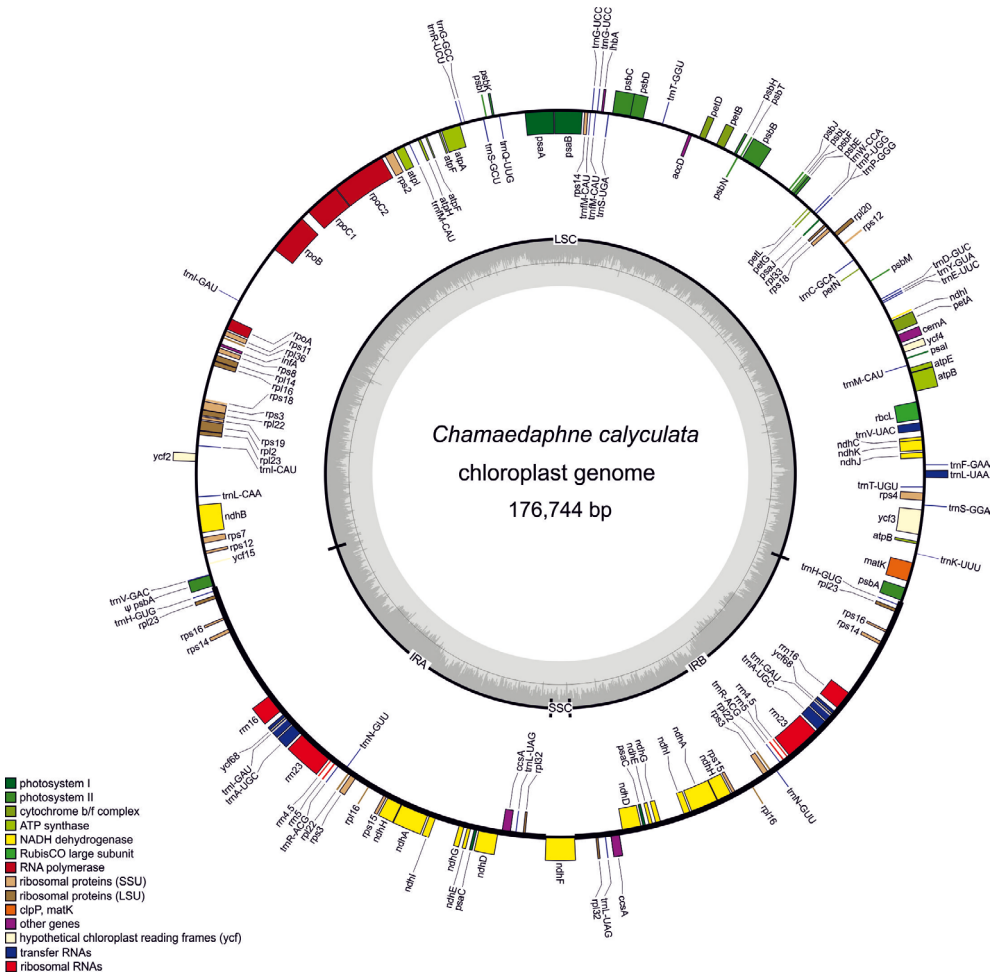

Figure 3. Gene map of the plastid genome of *Chamaedaphne calyculata*. The inner circle indicates the large single-copy region, inverted repeat regions, and small single-copy region. Genes inside and outside the outer circle are transcribed in a counterclockwise and clockwise direction, respectively. The genes are color-coded based on their function. The inner circle visualizes G/C content.

Table 1. Gene distribution by plastid region in *Chamaedaphne calyculata*.

	LSC	IR	SSC	Total
ORF	61 (including *mat*K)	17 (including *ycf*68)	1	79
tRNA	24	6	0	30
rRNA	0	4	0	4
Total	85	27	1	113

Table 2. Genes found in the *Chamaedaphne calyculata* ([a]genes located in single-copy regions; [b]pseudogene; [c]gene located in the inverted repeats; [d]2 copies, 1 functional and 1 pseudogene).

Gene names	Type of gene
accD[a, b]	Acetyl-CoA carboxylase
atpA[a] , atpB[a], atpE[a], atpF[a], atpH[a], atpI[a]	ATP synthase subunits
petA[a], petB[a], petD[a], petG[a], petL[a], petN[a]	Cytochrome b/f complex subunit
ccsA [c]	Cytochrome c protein
cemA[a]	Envelope membrane protein
matK[a]	Maturase K
ndhA[c], ndhB[a], ndhC[a], ndhD[c], ndhE[c], ndhF[a], ndhG [c], ndhH[c], ndhI[a, c, d], ndhJ[a], ndhK[a]	NADH-dehydrogenase subunits
psaA[a] ,psaB[a], psaC[c], psaI[a], psaJ[a]	Photosystem I subunits
psbA[a, d] , psbB[a], psbC[a], psbD[a], psbE[a], psbF[a], psbH[a], psbI[a], psbJ[a], psbK[a], psbL[a], psbM[a], psbN[a], psbT[a]	Photosystem II subunits
rrn16[c], rrn23[c], rrn4.5[c], rrn5[c]	Ribosomal RNAs
rpoA[a], rpoB[a], rpoC1[a], rpoC2[a]	RNA polymerase subunits
rbcL[a]	Rubisco large subunits
rps11[a], rps12[a], rps14[c], rps15[c], rps16[b, c], rps18[a], rps19[a], rps2[a], rps3[a, c, d], rps4[a], rps7[a], rps8[a]	Small ribosomal protein units
rpl14[a], rpl16[a,c] , rpl2[a], rpl20[a], rpl22[a, c, d], rpl23[a,c], rpl32[c], rpl33[a], rpl36[a]	Large ribosomal protein units
ycf15[a], ycf2[a, b], ycf3[a], ycf4[a], ycf68[b, c]	Hypothetical genes
trnA-UGC[c], trnC-GCA[a], trnD-GUC[a], trnE-UUC[a], trnF-GAA[a],trnG-GCC[a], trnG-UUC[a], trnH-GUG[c], trnI-CAU[a], trnI-GAU[c], trnK-UUU[a], trnL-CAA[a], trnL-UAA[a], trnL-UAG[c], trnM-CAU[a], trnN-GUU[c], trnP-GGG[a], trnP-UGG[a], trnQ-UUG[a], trnR-ACG[c], trnR-UCU[a], trnS-GCU[a], trnS-GGA[a], trnS-UGA[a], trnT-GGU[a], trnT-UGU[a], trnV-GAC[a], trnV-UAC[a], trnW-CCA[a], trnY-GUA[a]	Transfer RNAs
infA[a], lhbA[a]	Other genes (translational initiation factor A and photosystem II protein Z, respectively)

of both genomes clearly indicates that *rpl*20 was present between the *rps*12 and *rps*18 genes in the LSC region of the *Vacciunium* plastid genome. A similar situation was found in the case of 2 transfer RNA genes, *trn*G-GCC and *trn*K-UUU, which were found between *psb*L and *trn*R-UCU and *mat*K and *atp*B, respectively. The noted structural changes are relatively small in comparison with those in the family Pinaceae, where many inversions have occurred (Wu et al., 2011). In other plant families, genome stability was similar to that observed in the family Ericaceae (Yang et al., 2013). Comparative plastid genomics of *Helianthus annuus* and *Lactuca sativa* of the family Asteraceae also did not reveal any changes in the gene order or composition (Timme et al., 2007). Pairwise identity of chloroplast sequences in *C. calyculata* and *V. macrocarpon* was 79.6% and the p-distance was 0.032. The noted values are much higher in comparison with those of other families subjected to comparative genomic analysis. In Bambusoideae (Zhang et al., 2011), the average p-distance was determined to be 0.009; however, plastid genomes of the monocots seem to be more conserved (Kim and Kim, 2013).

Nucleotide variability was observed mostly in noncoding regions (Figure 4). Pairwise identity for different gene groups ranged from 83.16% for *acc*D to 99.62% for ribosomal RNA sequences (Table 3). Mean pairwise identity for coding regions was 96.6%. The nucleotide variation in the gene coding Acetyl-CoA carboxylase (*acc*D) is not surprising, since it has proven its usefulness many times in plant barcoding studies (Newmaster et al., 2008; Krawczyk et al., 2014). In coding regions, the total number of nonsynonymous and synonymous mutations was similar at 354 and 375, respectively (Figure 5). The number of nonsynonymous/synonymous mutations differed between gene families. A much higher number

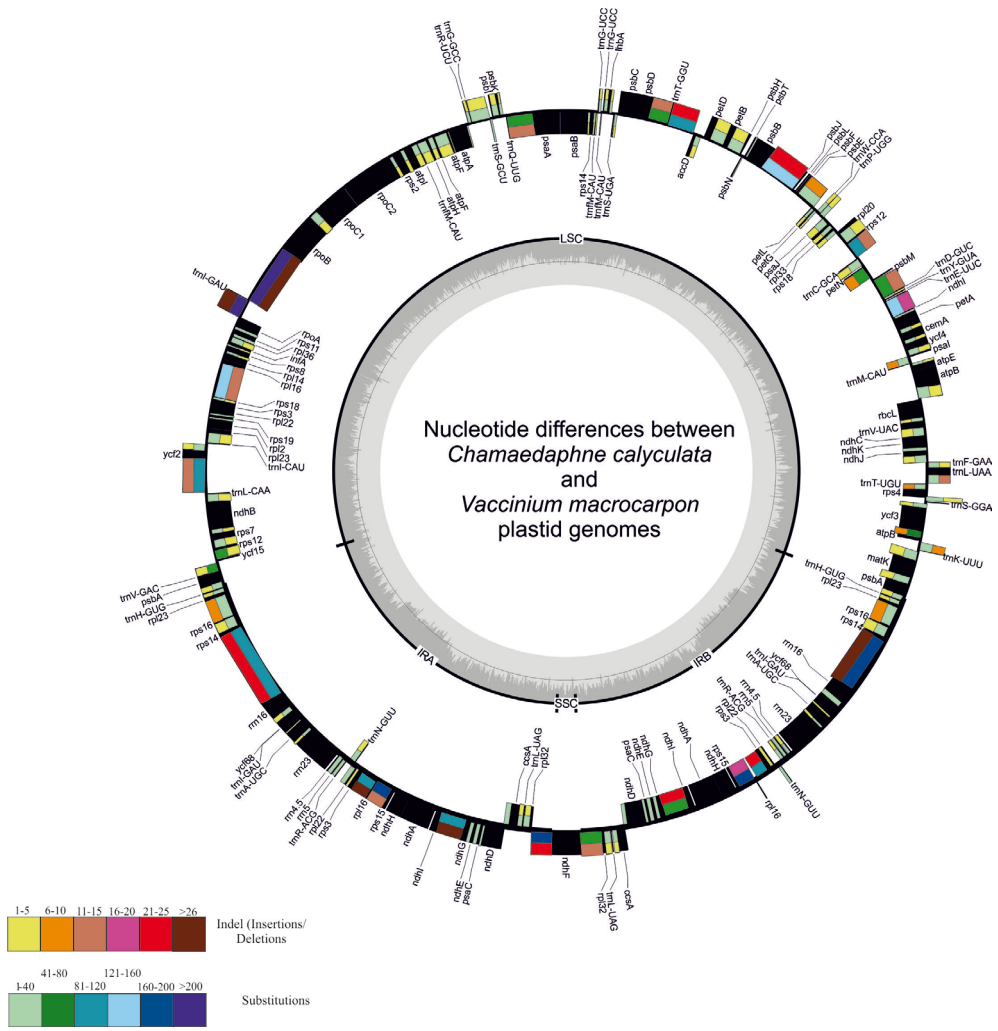

Figure 4. Nucleotide variation of noncoding regions of plastid genomes of *C. calyculata* and *V. macrocarpon*. Coding regions are marked in black. The inner circle visualizes G/C content.

Table 3. Pairwise identity for different gene groups.

Type of gene	Percentage %
Acetyl-CoA carboxylase	83.16
ATP synthase subunits	97.97
Cytochrome b/f complex subunit	98.78
Cytochrome c protein	98.53
Envelope membrane protein	98.41
Maturase K	96.72
NADH-dehydrogenase subunits	96.14
Photosystem I subunits	98.40
Photosystem II subunits	98.92
Ribosomal RNAs	99.62
RNA polymerase subunits	93.51
Rubisco large subunits	97.62
Small ribosomal protein units	97.08
Large ribosomal protein units	96.50
Hypothetical genes	94.86
Transfer RNAs	98.88
Other genes (translational initiation factor A and photosystem II protein Z, respectively)	97.09

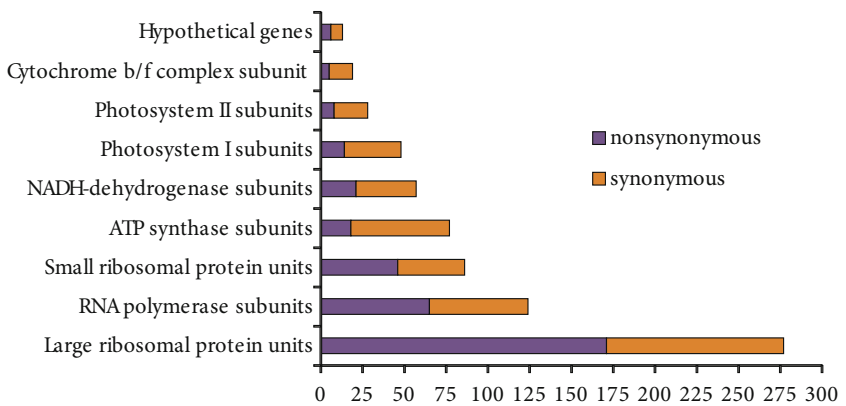

Figure 5. The number of synonymous/nonsynonymous mutations in each of gene groups.

of synonymous mutations was found in genes encoding cytochrome, photosystem I and II, NADH dehydrogenase, and ATP synthase proteins. Proteins coded by these gene families are crucial for photosynthesis, the fundamental biological process of almost every plant species. A prevalence of nonsynonymous mutations was noted in genes encoding proteins of a large ribosomal subunit, which seems to be under lower selection pressure.

3.4. Microsatellite mining in the chloroplast and nuclear genomes

Fifty-four simple sequence repeats (microsatellites (SSR)) were identified in the *C. calyculata* plastid genome (with the second IR segment excluded) with the use of WebSat web software (Martins et al., 2009). Thirty-one of those were mononucleotide repeats of 10 or more bases (A or T), 5 were dinucleotide repeats with 5 to 7 copies (all AT),

4 were trinucleotide repeats with 4 to 5 copies (GAA, CTT, TTA, TTC), 11 were tetranucleotide repeats with 3 to 4 copies (TTTC, CAAA, TTTA, TTCT, TTCA, ATTC, AATA, AATC), 2 were pentanucleotide repeats with 3 copies (GATCC, AATAA), and 2 were hexanucleotide repeats with 3 copies (TTTCTT, TTTGAT) (Table 4).

The total number of SSRs identified in the plastid genome of *C. calyculata* was lower (54) than in *V. macrocarpon*, where 62 plastid SSRs were found (Fajardo et al., 2013). Among the types of repeats, only mononucleotides were more common in the plastid genome of *Chamaedaphne* than *Vaccinium* (31 versus 23). The distribution of SSR among parts of plastid genomes was similar in the case of single copy regions. Both species have only 1 SSR in the SSC and 38 and 42 in the LSC of *Vaccinium* and *Chamaedaphne*, respectively. However, *V. macrocarpon* has twice as many SSRs in the IR regions as *C. calyculata* (22 vs. 11). Except SSRs located in the coding regions of plastid genomes, the locations of the remaining microsatellites is different in both species. A similar pattern was revealed in the family Asteraceae. *Lactuca sativa* and *Helianthus annuus* shared only 6 SSR regions out of 68 identified for each species (Timme et al., 2007).

The plastid genomes of the species of the family Ericaceae contains similar numbers of microsatellite regions in comparison to other dicot taxa. Plastid SSR mining revealed 61 regions containing repeats in *Olea europaea* (Besnard et al., 2011). Much lower numbers of microsatellite regions were identified in the monocot plastid genomes. The number of SSR motifs in the plastid genomes of the order Liliales ranged from 9 in *Lilium* to 21 in *Smilax* (Kim and Kim, 2013). Since the available literature data are limited, it is too early to make any general conclusion about the reasons for this phenomenon.

In the assembled data, 30,618 of 448,353 contigs not mapped on the plastid genome contained microsatellite regions. There was a prevalence of dinucleotide and trinucleotide repeats at 32,323 and 2379 regions, respectively. The number of regions with trinucleotides and tetranucleotides did not confirm the decreasing trend in the number of regions containing longer repeats, because the number of regions containing trinucleotide repeats was smaller (683) than the number of regions with tetranucleotide repeats. Hexanucleotide repeats were the least frequent; they were found in 415 regions. A list of 100 selected SSR motifs and their GenBank numbers are presented in Table 5. Listed SSRs were selected on the basis of contig coverage, which must be higher than 10-fold.

Nuclear SSR sequences reported in this study can be used to develop effective conservation programs for endangered populations of *C. calyculata*. NGS is the most popular method of developing SSR markers, and it significantly reduces the time and cost of the procedure (Szczecińska et al., 2012). In studies analyzing the polymorphism of SSR markers isolated by the NGS method, the percentage of polymorphic loci in the total number of tested loci in species from various taxonomic groups ranged from 20% to 50% (Sawicki et al., 2012; Szczecińska et al., 2012, 2013).

The complete plastid genome of *C. calyculata* points to the structural stability of cpDNA in the subfamily Vaccinioideae. A direct comparison of plastid genomes in *C. calyculata* and *V. macrocarpon* revealed pairwise identity of 79.6%. The main differences were observed in noncoding regions, and pairwise identity for coding regions was determined to be 96.6%. The described plastid and nuclear microsatellite loci will contribute

Table 4. Location and length distribution of *Chamaedaphne calyculata* plastid simple sequence repeats found in the large single-copy region, inverted regions, and the small single-copy region. Repeated sequences were classified by type, based on the number of repeated nucleotides (mono-, di-, tri-, tetra-, penta-, and hexanucleotides).

Type of repeat	Repeats found	Number of motif repeats	LSC/IR/SSC	Genes/pseudogenes
Mono-	31	10–13	27/4/0	1/3
Di-	5	5–7	4/1/0	2/0
Tri-	4	4–5	4/0/0	1/0
Tetra-	11	3–4	6/4/1	1/1
Penta-	2	3	0/2/0	0/0
Hexa-	2	3	1/0/0	1/0

Table 5. Identified nuclear SSR markers of *Chamedaphne calyculata*.

Locus	Motif	Number of repeats	GenBank number	Locus	Motif	Number of repeats	GenBank number
ChC-ssr1	CT	28	KJ417329	ChC-ssr51	AG	27	KJ417379
ChC-ssr2	GTGTT	4	KJ417330	ChC-ssr52	AT	20	KJ417380
ChC-ssr3	CT	14	KJ417331	ChC-ssr53	CT	22	KJ417381
ChC-ssr4	AT	17	KJ417332	ChC-ssr54	CT	17	KJ417382
ChC-ssr5	CT	17	KJ417333	ChC-ssr55	ATCCG	5	KJ417383
ChC-ssr6	AG	20	KJ417334	ChC-ssr56	AATT	5	KJ417384
ChC-ssr7	AG	22	KJ417335	ChC-ssr57	CT	16	KJ417385
ChC-ssr8	CT	17	KJ417336	ChC-ssr58	AT	15	KJ417386
ChC-ssr9	AC	8	KJ417337	ChC-ssr59	CT	23	KJ417387
ChC-ssr10	AT	11	KJ417338	ChC-ssr60	AG	15	KJ417388
ChC-ssr11	CT	13	KJ417339	ChC-ssr61	ATTT	5	KJ417389
ChC-ssr12	AG	18	KJ417340	ChC-ssr62	CTT	6	KJ417390
ChC-ssr13	AG	8	KJ417341	ChC-ssr63	ATTT	5	KJ417391
ChC-ssr14	AAG	6	KJ417342	ChC-ssr64	ATTTT	5	KJ417392
ChC-ssr15	CT	31	KJ417343	ChC-ssr65	ACT	8	KJ417393
ChC-ssr16	CT	8	KJ417344	ChC-ssr66	AAG	6	KJ417394
ChC-ssr17	AAG	6	KJ417345	ChC-ssr67	ACAT	5	KJ417395
ChC-ssr18	AAG	9	KJ417346	ChC-ssr68	ATTT	6	KJ417396
ChC-ssr19	CT	10	KJ417347	ChC-ssr69	AGT	6	KJ417397
ChC-ssr20	AG	40	KJ417348	ChC-ssr70	GTT	6	KJ417398
ChC-ssr21	ATT	7	KJ417349	ChC-ssr71	AAACT	4	KJ417399
ChC-ssr22	AG	24	KJ417350	ChC-ssr72	ATT	7	KJ417400
ChC-ssr23	AG	18	KJ417351	ChC-ssr73	AAG	13	KJ417401
ChC-ssr24	AG	8	KJ417352	ChC-ssr74	ACTCT	4	KJ417402
ChC-ssr25	AG	11	KJ417353	ChC-ssr75	CTT	8	KJ417403
ChC-ssr26	GTT	6	KJ417354	ChC-ssr76	CTGAGT	4	KJ417404
ChC-ssr27	AG	10	KJ417355	ChC-ssr77	AAAT	5	KJ417405
ChC-ssr28	CT	16	KJ417356	ChC-ssr78	ATTT	5	KJ417406
ChC-ssr29	AG	21	KJ417357	ChC-ssr79	ATTT	5	KJ417407
ChC-ssr30	AG	10	KJ417358	ChC-ssr80	AAG	9	KJ417408
ChC-ssr31	AG	8	KJ417359	ChC-ssr81	AAG	26	KJ417409
ChC-ssr32	ATTT	5	KJ417360	ChC-ssr82	GGTTT	4	KJ417410
ChC-ssr33	AT	21	KJ417361	ChC-ssr83	ATCT	7	KJ417411
ChC-ssr34	GAT	7	KJ417362	ChC-ssr84	AAG	6	KJ417412
ChC-ssr35	AG	20	KJ417363	ChC-ssr85	AAG	6	KJ417413
ChC-ssr36	AC	9	KJ417364	ChC-ssr86	ACT	14	KJ417414
ChC-ssr37	CT	23	KJ417365	ChC-ssr87	GCCGAT	4	KJ417415
ChC-ssr38	CTGT	5	KJ417366	ChC-ssr88	ATTTT	4	KJ417416
ChC-ssr39	CT	17	KJ417367	ChC-ssr89	CTT	7	KJ417417
ChC-ssr40	AAAT	7	KJ417368	ChC-ssr90	ATTT	5	KJ417418
ChC-ssr41	AC	18	KJ417369	ChC-ssr91	ACT	10	KJ417419
ChC-ssr42	GT	8	KJ417370	ChC-ssr92	CTTTT	5	KJ417420
ChC-ssr43	AGT	15	KJ417371	ChC-ssr93	GTT	11	KJ417421
ChC-ssr44	AG	29	KJ417372	ChC-ssr94	ATCT	5	KJ417422
ChC-ssr45	AG	25	KJ417373	ChC-ssr95	AACCG	4	KJ417423
ChC-ssr46	GGT	8	KJ417374	ChC-ssr96	CTTTTT	4	KJ417424
ChC-ssr47	AG	25	KJ417375	ChC-ssr97	ATTTTT	4	KJ417425
ChC-ssr48	AAAT	5	KJ417376	ChC-ssr98	CCTCTT	5	KJ417426
ChC-ssr49	AAAT	5	KJ417377	ChC-ssr99	AAGTGT	5	KJ417427
ChC-ssr50	CTGGTT	5	KJ417378	ChC-ssr100	CAGTGT	8	KJ417428

to evaluations of genetic diversity within and among populations, including in phylogeographic studies. SSR markers can also be used to develop effective conservation programs for endangered populations of *C. calyculata*.

Acknowledgment

This work was made possible through financial support from Polish State Committee for Scientific Research (KBN grant N303 807840).

References

Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011). Genomic profiling of plastid DNA variation in the Mediterranean olive tree. BMC Plant Biol. 11: 80 doi:10.1186/1471-2229-11-80.

Boore JL (1999). Animal mitochondrial genomes. Nucleic Acids Res 27: 1767–1780.

Bush CM, Lu L, Fritsch PW, Li DZ, Kron KA (2009). Phylogeny of Gaultherieae (Ericace: Vaccinioideae) based on DNA sequence data from matK, ndhF, and nrITS. Int J Plant Sci 170: 355–364.

Darling AE, Mau B, Perna NT (2010). ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5: e11147.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S et al. (2011). Geneious v5.4. Available from http://www.geneious.com/.

Duminil J (2014). Mitochondrial genome and plant taxonomy. Methods Mol Biol 1115: 121–140.

Faircloth B (2008). MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Res 8: 92–94.

Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K et al. (2012). Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content and rearrangements revealed by next generation sequencing. Tree Genet Genomes 9: 489–498.

Güvenç A, Kendir G (2012). The leaf anatomy of some *Erica* taxa native to Turkey. Turk J Bot 36: 253–262.

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ et al. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol 395: 348–384.

Kim JS, Kim JH (2013). Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two lilies (*Lilium longiflorum* and *Alstroemeria aurea*). PLoS ONE 8: e68180. doi:10.1371/journal.pone.0068180.

Kloss M (1996). Plant succession on *Chamaedaphne calyculata* (L.) Moench mire in the Kampinos Forest. Pol Ecol Stud 22: 129–13.

Kloss M (1999). *Chamaedaphne calyculata* – endangered species of marshy coniferous forest and peat bogs. Las Polski 10: 16–17.

Krawczyk K, Szczecińska M, Sawicki J (2014). Evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L. (Lamiaceae). Mol Ecol Res 14: 272–285.

Kruszelnicki J (2001). *Chamaedaphne calyculata* (L.) Moench– Chamedafne północna. In: Kaźmierczakowa R, Zarzycki K, editors. Polish Red Data Book of Plants. Pteridophytes and Flowering Plants. Kraków, Poland: Polish Academy of Sciences, W. Szafer Institute of Botany, Institute of Nature Conservation, pp. 283–285.

Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics doi: 10.1186/1471-2164-13-715.

Lohse M, Drechsel O, Bock R (2007). OrganellaGenomeDRAW (OGDRAW) - a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52: 267–274.

Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2009). WebSat - A Web Software for MicroSatellite Marker Development. Bioinformation 3: 282–283.

Merev N, Yavuz H (2000). Ecological wood anatomy of Turkish *Rhododendron* L. (Ericaceae). Intraspecific variation. Turk J Bot 24: 227–237.

Meusel H, Jäger E, Rauschert S, Weinert E (1992). *Vergleichende* Chorologie der zentraleuropäischen Flora. Bd III. G. Jena, Germany: Fischer Verlag.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008). Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Res 8: 480–490.

Palmer JD, Stein DB (1986). Conservation of chloroplast genome structure among vascular plants. Curr Genet 10: 823–833.

Sawicki J, Kwaśniewski M, Szczecińska M, Chwiałkowska K, Milewicz M, Plášek V (2012). Isolation and characterization of simple sequence repeats (SSR) markers from the moss genus *Orthotrichum* using a small throughput pyrosequencing machine. Int J Mol Sci 13: 7586–7593.

Szczecińska M, Sawicki J, Wąsowicz K, Hołdyński Cz (2009). Genetic variation of the relict and endangered population of *Chamaedaphne calyculata* (Ericaceae) in Poland. Dendrobiology 62: 23–33.

Szczecińska M, Kwaśniewski M, Sawicki J, Chwiałkowska K, Szandar K, Pisarek W (2012). Development of microsatellite markers using pyrosequencing in *Galium trifidum* (Rubiaceae), a rare species in Central Europe. Int J Mol Sci 13: 9893–9899.

Szczecińska M, Kwaśniewski M, Chwiałkowska K, Sawicki J (2013). Isolation and characterization of microsatellite loci in *Pulsatilla patens* (L.) Mill. (Ranunculaceae) a rare and endangered plant species in Europe. Conserv Genet Resour 5: 421–423.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2001). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM (2011). Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. Genome Biol Evol 3: 309–319.

Wyman SK, Jansen RK, Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255.

Yang JB, Yang SX, Li HT, Yang, J, Li DZ (2013). Comparative chloroplast genomes of *Camellia* species. PLoS ONE 8: e73053. doi:10.1371/journal.pone.0073053.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821–829.

Zhang YJ, Ma PF, Li, DZ (2011). High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS ONE 6: e20596. doi:10.1371/journal.pone.0020596.