A Novel Set of Chloroplast SSR Markers for the Genus *Juglans* Reveals Within Species Differentiation

Jeremias Götz[1]*, Oliver Caré[1,2], Winston Beck[3], Oliver Gailing[2], Bernhard Hosius[1], Ludger Leinemann[1,2]

[1] ISOGEN GmbH & Co. KG; Büsgenweg 2; 37077 Göttingen, Germany
[2] Department of Forest Genetics and Forest Tree Breeding; Georg-August University of Göttingen; Büsgenweg 2; 37077 Göttingen, Germany
[3] Division Urban Plant Ecophysiology, Humboldt-Universität zu Berlin, Lentzeallee 55-57, 14195 Berlin, Germany

* Corresponding author: Jeremias Götz, E-mail: jeremias.goetz@isogen.de

Abstract

Juglans is an economically important genus distributed all over the world. Due to its high value, there is a large interest in genetic markers for population genetic analyses and the certification of reproductive material. Using chloroplast genomes of different *Juglans* species, we developed a highly polymorphic and reliable chloroplast SSR marker set of eleven cpSSRs. It was successfully tested in five different Juglans species. Genetic differentiation among species and provenances within species was assessed. We were able to define a total of nine different chloroplast haplotypes using the novel cpSSR markers with up to four different haplotypes within a single *Juglans* species. Due to the cost and labor effective character of SSR markers, when compared to other available methods, such as whole chloroplast genome sequencing or SNP analyses, this novel cpSSR marker set will contribute to future population genetic studies in the *Juglans* genus.

Keywords: *Juglans, chloroplast, haplotype, SSR, variation, genetic structure*

Introduction

Due to its valuable hardwood and edible nuts, *Juglans* is an economically highly important genus distributed across Asia, North America, and Europe. Breeding efforts require genetic markers as a basis for selection and characterization of plant material and the verification of crossing experiments (e.g., Bernard et al., 2018). A fast and cost-effective type of genetic markers are simple sequence repeats (SSRs). A comprehensive set of highly polymorphic nuclear SSRs (nSSRs) is already available (e.g., Dangl et al., 2005; Zhang et al., 2010; Ebrahimi et al., 2019). Organelle SSRs, on the other hand, are still rare, even though they would be desirable in order to expand the range of tools for future investigations. Chloroplast genomes are maternally inherited (via seeds) in most angiosperms, including *Juglans* species. Additionally, they exhibit a low mutation rate as is expected for organelle genomes (e.g., Provan et al., 2001; Souframanien and Reddy, 2015). Hence, chloroplast SSRs (cpSSR) are of particular use to trace historical migration routes based on maternal lines. Due to their slow evolutionary rate, polymorphic cpSSRs within *Juglans* species are expected to be rare or even absent (Ebrahimi et al., 2019).

Ebrahimi et al. (2019) aimed to develop nuclear and organelle SSR markers for the genus *Juglans* to contribute to future genetic analyses. For allele scoring, motifs of two, three, or even more base pairs provide a reliable basis for distinction. Thus, Ebrahimi et al. (2019) focused on di-, tri- and complex motifs in their report and inferred complete absence of polymorphic cpSSRs within *Juglans* species and overall low genetic variation in *Juglans* chloroplasts. However, chloroplast genomes often exhibit variation in short single base pair SSR motifs as for example was observed in *Quercus* (Götz et al., 2022) and *Abies* species (Götz et al., 2024). In our study, we considered 26 chloroplast genomes of five different *Juglans* species, covering various economically important species to develop a reliable cpSSR marker set for the genus *Juglans*. To contribute to future

DOI:10.2478/sg-2024-0012
edited by the Thünen Institute of Forest Genetics

investigations within the *Juglans* genus, we aimed for a marker set that is not only useful to reliably differentiate between different species of the genus, but also exhibits genetic variation within chloroplast genomes of single species. Such cpSSRs are, to our knowledge, so far not available.

Materials and methods

SSR identification and primer design

We aligned a total of 26 chloroplast genomes of five different *Juglans* species (Table S1) using the QIAGEN CLC Workbench 23.03 (Qiagen, Hilden, Germany) to detect chloroplast DNA variation in the *Juglans* genus. All sequences originate from the NCBI database (Sayers et al., 2021) in FASTA format. Manual browsing for regions of variation in short repetitive motives in the alignment revealed potential SSR candidates. Only motifs flanked by conservative nucleotide sequences were regarded potentially reliable markers. The QIAGEN CLC Workbench 23.03 (Qiagen, Hilden, Germany) was also used to design forward and reverse primers in highly conserved upstream and downstream regions around chloroplast SSR candidates with the following boundary setting: Melting temperature of each primer between T_m = 50°C and T_m = 68°C, GC content of 40 – 60 %, and calculated PCR product fragment lengths of 70 to 550 base pairs. Each forward primer had a M13-tail addition to the designed sequence, as proposed by Schuelke (2000). Each reverse primer had a pig-tail addition following Brownstein et al. (1996) to enable cost-effective primer testing and accurate genotyping.

Plant material

Leaves of different *Juglans* species were collected from a provenance collection in Baden-Württemberg, Germany. The planted material is in private property of the members of the German *Juglans* enthusiast community "IG Nuss". Altogether, the sample set included five different Juglans species, originating from several countries (Table 1). In case of *J.* × *intermedia*, different varieties (NG38, MJ209, NG23, RENI) were included instead of natural populations.

DNA extraction

Genomic DNA was extracted from leaf tissue. We used approximately 50 mg fresh or 25 mg dried material cut into small pieces, frozen in liquid nitrogen, and ground in a MM300 ball mill (Retsch, Haan, Germany) for two times two minutes at 20 Hz. The DNeasyTM 96 Plant Kit (Qiagen, Hilden, Germany) was used for the subsequent extraction, according to the manufacturer's protocol. DNA concentration of a representative sub sample was photometrically determined using the NanoDrop 2000 (Thermo Fischer Scientific, Waltham, USA) spectral photometer.

Amplification and scoring

Primer testing was performed in a TProfessional thermocycler (Biometra, Jena, Germany) with a touchdown protocol, consisting of 15 minutes of initial denaturation (95 °C), followed by ten cycles of 60s of denaturation (94 °C), 60s annealing (60 °C; Δ-1 °C per cycle), and 60s extension (72 °C). Subsequently another 25 cycles were performed, using the same temperatures and times, but maintaining 50 °C in the annealing step. The reaction was finalized by 20 minutes of extension (72 °C). Each PCR mix (15 µl) contained about 5 ng of sample DNA, 1.05 pmol Tris–HCl and 0.263 pmol $(NH_4)_2SO_4$, 37.5 pmol $MgCl_2$, 2.5 pmol of each dNTP, and 1 U HOT FIREPol® Taq polymerase from Solis BioDyne (Tartu, Estonia). Polymorphic primer pairs were multiplexed with up to four primer pairs. Each sample volume contained 2.5 pmol M13 Primer (fluorescent dye labeled), 0.5 pmol forward and 1.25 pmol reverse primer of each tested primer pair in the respective multiplex. Various multiplex attempts were used during initial primer tests. However, there is no recommended multiplex composition and any subset of the introduced primer pairs could be successfully multiplexed, as long as expected allele size ranges are separated by at least 10 base pairs.

The fragment sizes of the PCR products were determined using an ABI Genetic Analyzer 3130xl (Applied Biosystems, Foster City, CA, United States) and GeneScan™ 500 ROX™ size standard (Thermo Fischer Scientific, Waltham, USA), as well as the accompanying Genemapper software v3.7 (Applied Biosystems, Foster City, CA, United States).

Initial tests for successful amplification, and fragment length polymorphisms were assessed using a subsample of eight *J. nigra*, and *J. regia* samples, each. Subsequently, the polymorphic markers were tested in the enlarged sample set mentioned above for an estimation of their degree of polymorphism and delimitation between fragment sizes. Furthermore, we repeated the final marker test, performing an independent PCR, as well as allele scoring to provide evidence for reproducible results using our *Juglans* cpSSR set.

Analysis of the genetic structure

GenAlEx 6.5 software (Peakall and Smouse, 2006) was used to count observed alleles within the sample set and to observe the allele frequencies, as well as general genetic chloroplast structure within the sample set. Genetic structure was visualized using the STRUCTURE 2.3.4 software (Pritchard et al., 2000). Results were independently repeated ten times with 10,000 burn-ins and 100,000 MCMC iterations, each using a standard admixture model. The summary of the results from all repetitions was visualized using CLUMPAK (Kopelman et al., 2015) and summarized using the STRUCTURE-SELECTOR browser application (Li and Liu, 2018). The number of genetic sub clusters was estimated following the method of Evanno et al. (2005). Haplotype distribution was visualized using the statistical programming language R (R core team, 2013).

Table 1
Origin and number of different sample populations of five *Juglans* species included in the test sample set. Number of populations per country of origin is displayed in brackets.

Species	Origin of the plant material	Populations	Sample size per population
J. nigra	USA (4)	4	4, each
J. regia	India (1), Hungary (1), Czech Republic (1), Germany (3), Switzerland (1), Canada (1), and Liechtenstein (1)	9	4, each
J. mandshurica	China (1)	1	12
J. cinerea	USA (3)	3	6 + 6 + 4
J. × *intermedia*	France (3), and Germany (1)	4	4, each

Results

Within the tested marker set of 21 cpSSRs, 14 were polymorphic in the reduced, initial test set and were further investigated. Three of those polymorphic markers (Jcp03, Jcp07, and MJcp05) exhibited a more or less continuous transition between fragment sizes of different alleles in the large test set and where thus discarded from the final marker set. The results of the remaining eleven *Juglans* cpSSRs (Table 1) were successfully reproduced in an independent investigation of the same DNA samples.

The introduced cpSSR marker set was able to separate different Juglans species, such as *J. regia* (Figure 1, S1). Within single species, genetic chloroplast structure was observed. The first four individuals of the *J. regia* sample set, for example, belong to a population from Dachigam (India) and three of them were assigned to a distinct genetic cluster (pink), while the remaining *J. regia* samples were assigned to the blue genetic cluster. Moreover, the individuals of the *J.* × *intermedia* samples were well distinguishable by three genetic clusters in the STRUCTURE analysis, which mostly match the four different varieties included in our test sample set (four adjacent samples, each).

Chloroplast haplotype variation (Figure 2, Table S2) ranged between one and four different haplotypes within our test sample set. While *J. nigra* and *J. mandshurica* samples exhibited no haplotype variation, we observed two different haplotypes within the *J. regia* and *J. cinerea* samples, respectively. The majority of the *J. regia* sub sample consisted of haplotype 3, complemented by about 10 % haplotype 2. *J. cinerea* samples were divided in two sections of haplotype 5 and 6, with about 40 % and 60 %, each. Within the *J.* × *intermedia* samples, we observed four different haplotypes of which about the half of the samples exhibited haplotype 1, which was common in the *J. nigra* sub sample. The other half of the *J.* × *intermedia* sub sample consisted of haplotypes 7, 8, and 9.

Discussion

The newly developed chloroplast SSR marker set for *Juglans* species revealed clear separation among different *Juglans* species. In addition, we observed genetic variation within the chloroplast genomes of some *Juglans* species. To our knowledge, this was previously restricted to SNP based studies (Bai et al., 2010, Zhao et al., 2018) or whole chloroplast genome sequencing (Dong et al., 2017). Even though we observed no genetic chloroplast variation within *Juglans* species *J. nigra*, and *J. mandshurica*, one has to consider the limitations of our sample set. The samples included in our test set originated from a very limited number of populations and there is no information available about the geographic distance between distinct populations. All included *J. mandshurica* samples, for example, originate from the exact same population and low genetic differentiation within the chloroplast genomes is likely. In general, our sample set is not representative for the entire genetic diversity of any of the included species. Therefore, we expect genetic variation at chloroplast SSRs to be discovered in future *Juglans* investigations using the introduced marker set. We can confirm the observation of Ebrahimi et al. (2019), who inferred a slow evolution rate in *Juglans* chloroplast genomes. Due to the expected low genetic variation in organelle genomes (e.g., Provan et al., 2001; Souframanien and Reddy, 2015), and the maternal inheritance of chloroplast genomes in *Juglans*, missing genetic variation in single species within our sample test set is barely surprising. Nevertheless, we were able to detect chloroplast polymorphisms within three of the five studied *Juglans* species, including the important timber and nut tree species *J. regia*.

J. × *intermedia* exhibited higher chloroplast haplotype variation when compared with *J. nigra*. *J.* × *intermedia* is a *Juglans* hybrid, which is usually harvested from *J. nigra* trees, which were fertilized by *J. regia* pollen. Due to the maternal inheritance of chloroplast genomes in *Juglans*, one would expect a similar haplotype distribution in *J. nigra* and *J.* × *intermedia*. In fact, the most prominent haplotype in *J.* × *intermedia* was haplotype 1, which was observed in all *J. nigra* samples within our test set. This could indicate a much higher haplotype differentiation in *J. nigra*, which was, by chance, not represented in our sample set. Since our investigated plant material contained four different *J.* × *intermedia* breeds, the high genetic differentiation in the hybrids might be a result of the four different mother linages. To achieve consistent seed material, *Juglans* breeders select distinct clones as mother trees. Contrarily, the *J. nigra* plant material was harvested from four different locations, but from populations, with unknown genetic

Table 2
Primer pair information about the introduced chloroplast SSR marker set for *Juglans* species.

Primer	Primer Sequence (5'-3')	Repeat Motif	Fragment Length	Tm(°C)	GC Content (%)	Observed Alleles
Jcp02 F	TCTTCAAGTACCAARGAC	$A_{(10)}T_{(10)}$	260	52.56	44	3 R, C
Jcp02 R	TCTTTATTCTCCACCCCA			53.16	44	
Jcp04 F	GGTACCTTCGGTTTGTTT	$A_{(11)}$	251	54.49	44	4 R, C, I
Jcp04 R	ACTCCCTTCTTTGTTTATCC			54.49	40	
Jcp05 F	TCATCAAACCGTCACCCA	$T_{(17)}$	476	65.70	50	3 C
Jcp05 R	GAACAACAAAGAGCGAGTAA			65.00	40	
Jcp06 F	AGGACTCTTCTGACCAAA	$T_{(13)}$	314	52.04	44	5 C, I
Jcp06 R	GAAGTCCACACTCAGCAA			55.71	50	
Jcp08 F	AAAACCCCAGCCATACCA	$A_{(8)}$	432	58.53	50	3 R, C
Jcp08 R	AGATCYACTGACCGTTACAA			58.47	45	
Jcp12 F	CCCTCTATTTTGGTGCCTTT	$T_{(16)}$	346	56.68	50	4 C, I
Jcp12 R	CCACTTCGTATGATTCCCT			56.87	42	
Jcp14 F	GGGGGGAAAACAAAAAAG	$T_{(12)}$	479	54.98	44	4 R, C, I
Jcp14 R	AAGAGGTTGATAGCGAGA			54.66	44	
Jcp15 F	CAGCGGAGGACCAAGAAA	$T_{(13)}$	332	58.78	56	3 C, I
Jcp15 R	ACCAACCCGAGACATAAACAAA			58.61	41	
MJcp01 F	ATTAAACCCGAAACTCCC	$T_{(9)}[N_{(20)}]^*A_{(12)}$	402	53.60	44	4 C, I
MJcp01 R	CTTCTTCCTTCCCCTTCT			55.02	50	
MJcp02 F	TGGAGGGGGTATAAATGG	Multiple indels	251	54.29	50	3 R, I
MJcp02 R	TGCGCGGATCAAGATAAT			55.25	44	
MJcp03 F	TTTCTCTCCCCACACTTC	$T_{(8)}A_{(12)}$	265	53.89	50	5 C, I
MJcp03 R	CACCCAAACAATCATTGC			53.43	44	

*5'-CAATTGGAAATTTCCAATTG-3'; N polymorphic in *J. nigra*; R polymorphic in *J. regia*; M polymorphic in *J. mandshurica*; C polymorphic in *J. cinerea*; I polymorphic in *J. × intermedia*

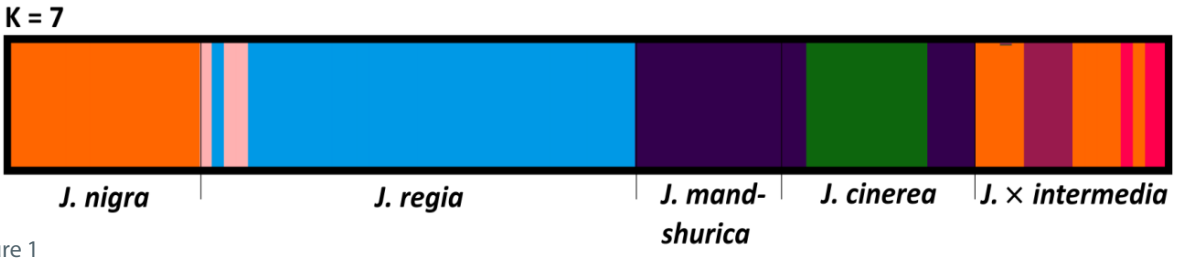


Figure 1
Genetic chloroplast variation within and amongst five different *Juglans* species characterized using the introduced cpSSR markers.

variation. It is possible, that, by sampling effects, the haplotype variation in the selected *J. × intermedia* mother trees was higher than in the randomly sampled *J. nigra* individuals in our sample set.

The single base pair motif in our markers can lead to errors in the clear discrimination of alleles. Therefore, we double checked the results in an independent repetition of all tested samples and discarded error-prone markers (Jcp03, Jcp07, and MJcp05) from the final marker set. We were able to independently reproduce the results for all of the remaining eleven markers for all investigated species. For the entire data of more than 1000 single values, only three exceeded a deviation of 0.3 base pairs. The deviation in fragment sizes between the two repetitions was generally low with a mean deviation of only 0.068 base pairs. Hence, the introduced cpSSR marker set contains a reliable toolkit despite the mononucleotide repeat motifs and might be of particular interest for future research in the economically important hardwood *Juglans* tree species. Four of the introduced markers (Jcp02, MJcp01, MJcp02, and MJcp03) contain more than one SSR motif, which were either directly adjacent or too close to each other to design suitable primers in between. In our test set, we observed comparable numbers of different alleles for the mentioned multi motif cpSSRs (Mean = 3.75), when compared to the single motif SSRs (Mean = 3.86), though.

Altogether, the introduced cpSSR marker set harbors the potential to detect variation within the chloroplast genomes of single *Juglans* species, while offering well reproducible results. Hence, it can contribute to future studies by providing a labor and cost-effective method for *Juglans* haplotype investigations.

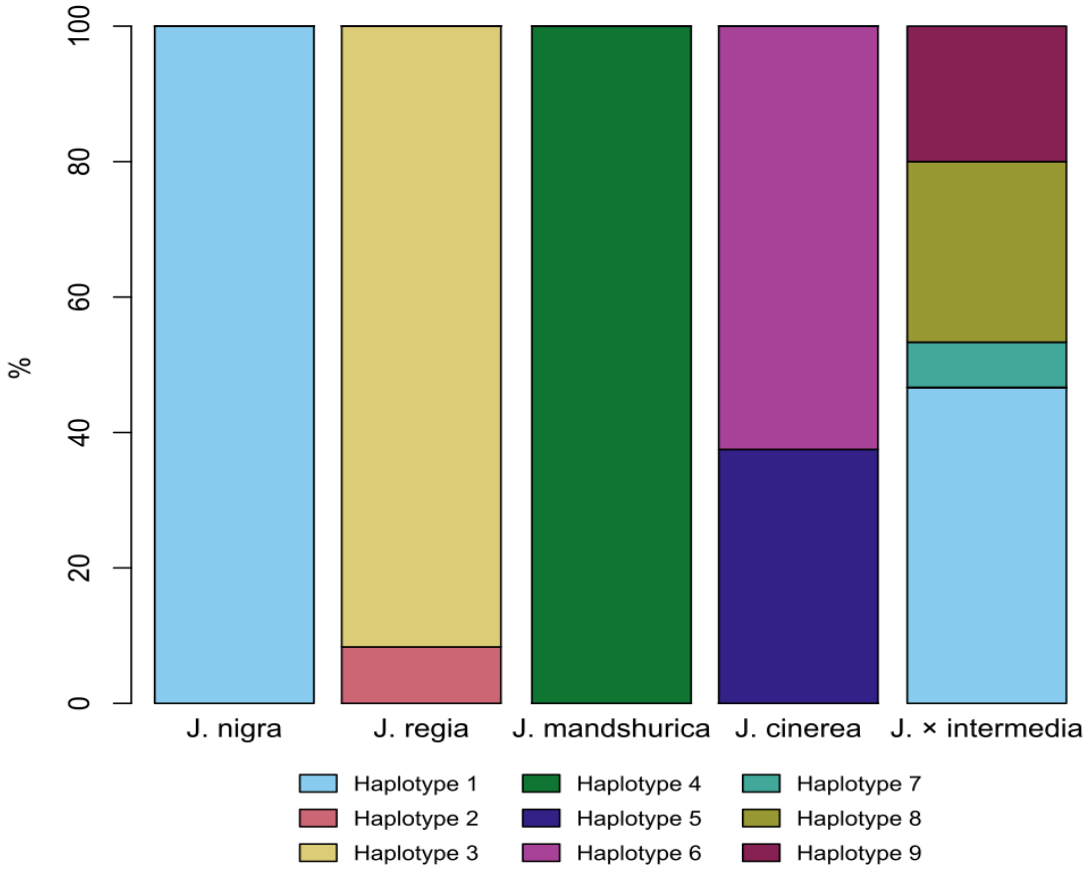


Figure 2
Chloroplast haplotype composition within and amongst five different *Juglans* species characterized using the introduced cpSSR markers.

Acknowledgements

The authors are grateful for the permission for the sample collection in the Juglans variety collection by Wolfgang Hertel (IG-Nuss) and Bernhard Mettendorf for his supporting background information about the sampled *Juglans* species and guidance in the collection stand.

Funding

This research was funded by the German Federal Ministry of Food and Agriculture (BMEL) represented by the Fachagentur Nachwachsende Rohstoffe e. V. (FNR) grant number FKZ 2222NR044B.

References

Bai WN, Liao WJ, Zhang DY (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytologist, 188(3), 892-901. https://doi.org/10.1111/j.1469-8137.2010.03407.x

Bernard A, Lheureux F, Dirlewanger E (2018) Walnut: past and future of genetic improvement. Tree genetics & genomes, 14(1), 1. https://doi.org/10.1007/s11295-017-1214-0

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of Non-Templated Nucleotide Addition by Taq DNA Polymerase: Primer Modifications that Facilitate Genotyping. BioTechniques, 20(6), 1004–1010. https://doi.org/10.2144/96206st01

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie CA, McGranahan G (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. Journal of the American Society for Horticultural Science, 130(3), 348-354. https://doi.org/10.21273/jashs.130.3.348

Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, Jin X, Suo Z (2017) Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. Front Plant Sci. Jun 30; 8:1148. https://doi.org/10.3389/fpls.2017.01148

Ebrahimi A, Mathur S, Lawson SS, LaBonte NR, Lorch A, Coggeshall MV, Woeste KE (*2019)* Microsatellite borders and micro-sequence conservation in *Juglans*. Scientific Reports, 9(1), 3748. https://doi.org/10.1038/s41598-019-39793-z

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Götz J, Gailing O (2022) A novel set of polymorphic chloroplast microsatellite markers for northern red oak (*Q. rubra* L.). Plant Genetic Resources, 20(2), 174-177. https://doi.org/10.1017/S1479262122000156

Götz J, Leinemann L, Gailing O, Hardtke A, Caré O (2024) Development of a highly polymorphic chloroplast SSR set in *Abies grandis* with transferability to other conifer species – A promising toolkit for gene flow investigations. Ecology and Evolution. https://doi.org/10.1002/ece3.11593

Hu Y, Woeste KE, Zhao P (2017) Completion of the chloroplast genomes of five Chinese Juglans and their contribution to chloroplast phylogeny. Frontiers in Plant Science, 7, 231924. https://doi.org/10.3389/fpls.2016.01955

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 15:1179–1191. https://doi.org/10.1111/1755-0998.12387

Li YL, Liu JX (2018) StructureSelector: A web based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources, 18:176–177. https://doi.org/10.1111/1755-0998.12719

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peng S, Yang G, Liu C, Yu Z, Zhai M (2017) The complete chloroplast genome of the *Juglans regia* (Juglandales: Julandaceae). Mitochondrial DNA Part A, 28(3), 407-408. https://doi.org/10.3109/19401736.2015.1127367

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology & Evolution, 16(3), 142-147. https://doi.org/10.1016/S0169-5347(00)02097-8

R Core Team R. (2013) R: A language and environment for statistical computing.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology, 18(2), 233-234. https://doi.org/10.1038/72708

Souframanien J, Reddy KS (2015) De novo assembly, characterization of immature seed transcriptome and development of genic-SSR markers in black gram [*Vigna mungo* (L.) Hepper]. PLoS One, 10(6), e0128748. https://doi.org/10.1371/journal.pone.0128748

Zhang R, Zhu A, Wang X, Yu J, Zhang H, Gao J, Cheng Y, Deng X (2010) Development of *Juglans regia* SSR markers by data mining of the EST database. Plant Molecular Biology Reporter, 28, 646-653. https://doi.org/10.1007/s11105-010-0192-2

Zhao P, Zhou HJ, Potter D, Hu YH, Feng XJ, Dang M, Feng L, Zulfigar S, Liu W-Z, Zhao G-F,Woeste K (2018) Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). Molecular Phylogenetics and Evolution, 126, 250-265. https://doi.org/10.1016/j.ympev.2018.04.014