SCIENCE SET
OPEN ACCESS PUBLISHERS

Research Article

Journal of Agricultural, Earth and Environmental Sciences

Study on Chloroplast Genome of Three Corylus Genera in Hebei, China

Yanfang Peng[a†], Yuping Zhong[b†], Weitao Mao[c], Lingzhi Shao[a] and Jie Li[a*]

[a]*School of Biology and Food Science, Hebei Normal University for Nationalities*
[b]*Qianjiang City ecological energy Bureau*
[c]*Qianjiang City animal husbandry technology promotion station*
[†]*These authors have contributed equally to this work*

[*]**Corresponding author:** Jie Li, School of Biology and Food Science, Hebei Normal University for Nationalities. Email: lijie@hbun.edu.cn

Submitted: 29 January 2024 **Accepted:** 02 February 2024 **Published:** 09 February 2024

Citation: *Yuping Zhong, Yanfang Peng, Lingzhi Shao and Jie Li (2024) Study on Chloroplast Genome of Three Corylus Genera in Hebei, China. J of Agri Earth & Environmental Sciences 3(1), 01-04.*

Abstract

Hazelnut is an important economic tree species, which is loved by people all over the world and is in short supply at present.Its cultivation remains substantially based on named selections from local, wild vegetation. The breeding of excellent varieties is very important in hazelnut industry. However, the research on its chloroplast genome is limited, and the data of interspecific and intraspecific variation are also lacking, which hinders the improvement of breeding. Chloroplast genome as the second largest genetic system in plants after nuclear genome, has an independent evolutionary route and can reveal different evolutionary events at the level of plastome. In this study, we sequenced the chloroplast genomes of three species of hazelnut in Hebei Province. The data offer significant genetic information for the identification of Corylus species, taxonomic and phylogenetic studies, and molecular breeding

Keywords: Hazelnut, Corylus, Chloroplast Genome, Illumina Sequencing, Phylogenetic Relationship

Hazelnut (Corylus spp.) is a crop tree of worldwide agronomic importance, which has special nutritional value. It is widely used in food industry, such as bread, chocolate candy, etc., and plays an important role in human nutrition and health.[1, 2]. Current cultivation depends almost entirely on European hazelnut (C. avellana) and is restricted to Mediterranean-like climates that allow consistent yields. The scale of global hazelnut production closely mirrors that of pistachio (Pistacia vera) at ≈1 million tons of annual in-shell production (Food and Agriculture Organization of the United Nations, 2017). Hazelnuts belong to the order Fagales, family Betulaceae, subfamily Coryloideae, genus Corylus L. and are native to the temperate zones of the northern hemisphere (Europe, Asia Minor, Asia, North America). The number of species within the genus Corylus varies between 9 and 25, depending on the taxonomic authority, with current revisions based on morphological, molecular and hybridization studies suggesting 13 major species assigned to 4 subsections [3]. There are 8 species and 2 varieties of Corylus plants originated in China, accounting for about half of the world's Corylus species. In China, cold-hardy hybrid cultivars from Corylus heterophylla and Corylus avellana were recently released and are planted in northeastern China[3, 4]. Even though it has a significant place in agriculture, a limited number of studies exists about C. avellana at the molecular level. In addition, most species of Corylus can hybridize with each other to achieve the chloroplast genomes (chloroplast genome) introgression and transfer interspecifically, therefor the phylogeny and species definition of Corylus are still not completed. Since the middle of the 19th century, scholars at home and abroad have begun to study the phylogeny of Corylus, but so far, no clear conclusions have been drawn, especially on the origin, differentiation history and geographical distribution pattern of the genus. The chloroplast genome, as the second largest genetic system in plants after nuclear genome, has an independent evolutionary route and can reveal different evolutionary events at the level of plastomes. The chloroplast genome is of great value in revealing the origin and evolution of species and the genetic relationship among different species [5].

The chloroplast genome of three Corylus species were sequenced in this study via a next-generation sequencing platform. We gained valuable information including the length and content of highly variable regions, indels, microsatellites, and single nucleotide polymorphisms (SNPs). Our results provide vital information for species identification and for understanding the phylogenetic relationship and evolutionary classification within the Corylus plants.

Materials and Methods

Sample Collection and DNA Extraction

Fresh leaves of three Corylus species (Corylus heterophylla,Corylus mandshurica,Corylus heterophylla x Corylus avellana) were obtained from Hebei Normal University for Nationalities

in Chengde City, Hebei Province, China.DNA extraction was performed according to a modified CTAB protocol [6]. Subsequently, DNA concentrations were measured using a NanoDrop spectrophotometer 2000 (Thermo Fisher Scientific, USA).

DNA Fragmentation, Library Preparation, and High-Throughput Sequencing

The extracted DNA was disrupted by ultrasound, and DNA fragments of 350bp were recovered by gel cutting. A 350bp library was constructed by NEBNext® library creation kit, and sequenced by HiseqXtenPE150 sequencing platform.

Sequencing Data Assembly

The raw data were quality controlled using Trimmomatic 0.39, Subsequently, the data after quality control are spliced from scratch using SPAdes [7, 8]. The spliced contigs used Blast program to screen the chloroplast genome contigs and the screened chloroplast genome contigs was reassembled using Sequencher 4.10 [9].

Annotation of the cp Genome

The reference sequence was based on the published chloroplast genome reference sequence of hazelnut Corylus mandshurica (GenBank sequence receiving number MF375334), and the gene annotation was performed by Plann program.Some genes with unsuccessful or misannotated annotations are added manually in Sequin software.

Analysis of Intra-Species Variation

In this study, the sequences of determination and splicing were aligned by MAFFTV7 software [10]. Informative sites and variable sites in the whole chloroplast genome sequences as well as in the inverted repeat (IR), large single-copy (LSC), and small single-copy (SSC) regions were counted using MEGA X for comparing and aligning the sequence matrices [11].

Phylogenetic Analysis

A maximum likelihood phylogeny built-in PhyloSuite was inferred using IQ-TREE under the K81u+I+G4+F model for 5000 ultrafast bootstraps [12-14]. Bayesian inference (BI) of the phylogenies was implemented using MrBayes with the best fit model according to BIC: GTR+F+I+G4. The Markov chain Monte Chain (MCMC) analysis was run for 10,000,000 generations. The trees were sampled every 1,000 generations, with the first 25% discarded as burn in. Finally, the average standard deviation of split frequencies < 0.01 was checked.

Results

Characterization of Chloroplast Genome

Similar to the chloroplast genome structure of most angiosperms, the chloroplast genome of Corylus is a typical tetrad structure, including two single copy regions (LSC and SSC) and two identical reverse repeat regions [15]. The cp genomes of the three Corylus species ranged in length from 159,830bp(C. heterophylla x C. avellana) to 159,844bp(C. mandshurica).The chloroplast genome sequence assemblies are doubled stranded, single, circular, and demonstrate classic cp quadripartite architecture, being composed of a pair of IRs (26,098–26,099 bp), divided by a LSC (88,850–88,865 bp) and an SSC (18,782–18,783 bp) region.The GC contents of the cp genomes of all three Corylus were the same (36.5%)(Table 1, Fig 1).The cp genomes of the three Corylus encode 131genes, including 112 unique genes, including 18 genes in the IR region.Among the 112 unique genes, there are 78 protein coding genes 30 tRNAgenes, and 4 rRNAgenes.15 genes have one intron and two genes have two introns (ycf3 and clpP).

Table 1: The basic cp Genomes Information of Three Corylus Species

species	LSC	LSC% GC	IR	IR% GC	SSC	SSC% GC	Total Length	%GC
Corylus heterophylla x Corylus avellana	88850	34.2%	26099	42.5%	18782	30.2%	159830	36.5%
Corylus heterophylla	88856	34.2%	26098	42.5%	18783	30.2%	159835	36.5%
Corylus mandshurica	88865	34.2%	26098	42.5%	18783	30.2%	159844	36.5%

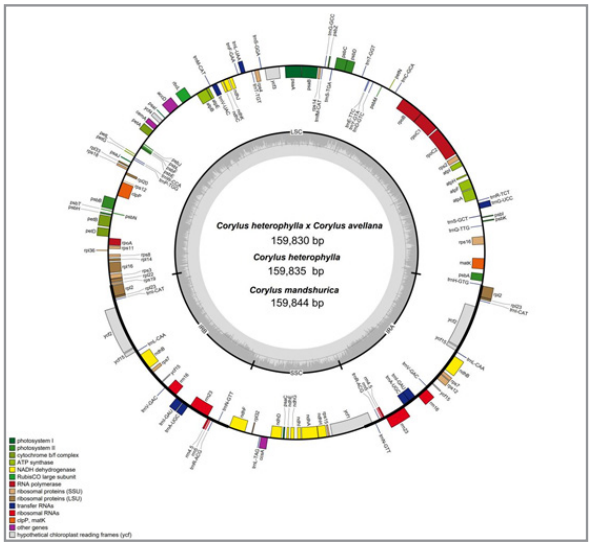

Figure 1: Circular Map of The Cp Genomes of Corylus Species.

Phylogenetic Analysis

At present, the classification of Betulaceae into Betuloideae and Coryloideae has been accepted by most taxonomists.The Betuloideae birch consists of only Alnus and Betula, while the Coryloideae consists of four genera: Corylus, Ostryopsis, Carpinus and Ostrya [16, 17]. In this study, 29 complete chloroplast genome sequences were used for phylogenetic analysis, including the chloroplast genome sequences of Alnus from the subfamily Betulinae.According to the results,Corylus heterophylla x Corylus avellana and Corylus mandshurica form a branch Fig 2.The data offer significant genetic information for the identification of Coryloideae species, taxonomic and phylogenetic studies, and molecular breeding.

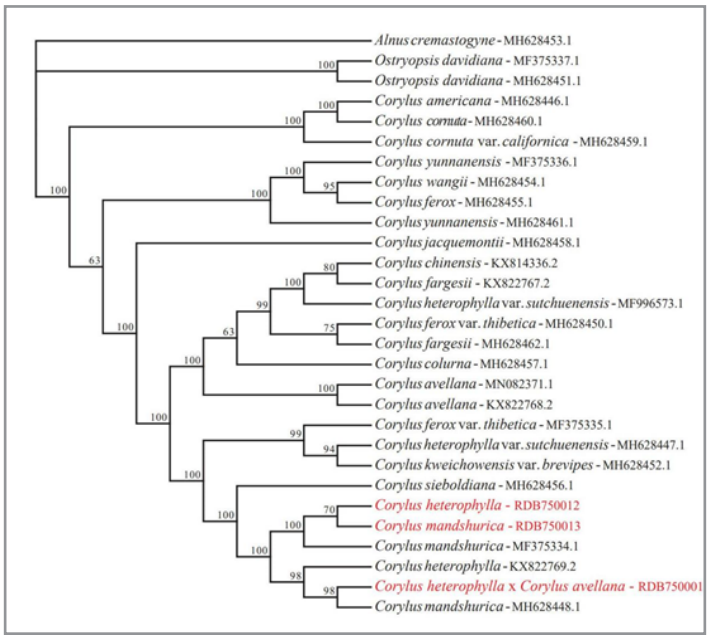

Figure 2: Phylogenetic Tree of Three Corylus Species

Conclusions

In this study, we reported the cp genome sequences of three Corylus species using Illumina paired-end sequencing.Compared with the published chloroplast genomes of Corylus, the chloroplast genomes of Corylus have the same typical circular tetrad structure as the chloroplast genomes of most higher plants. that is, it includes a pair of reverse complex regions and two separated large and small single copy regions.In addition, 131genes were annotated in the chloroplast genomes of three Corylus species. The functional classification, intron distribution and number of genes were the same, and there was no difference in GC content in different regions of the genome.

Our results provided insights into the characteristics of Corylus plants and the phylogenetic relationships between Betulaceae species.

Acknowledgements

Special thanks to Hebei Provincial Department of Science and Technology for the project funding support, thanks to Hebei Normal University for Nationalities for the research platform. (S&T Program of Hebei No. 20326807D).

References

1. Köksal Aİ, Artik N, Şimşek A, Güneş N (2006) Nutrient composition of hazelnut (Corylus avellana L.) varieties cultivated in Turkey. Food Chemistry 99: 509-515.
2. Wen-Han T, Li-Song L, Gui-Xi W (2012) Nutrient Composition Analysis of Hazelnut Kernel from Different Varieties. Food Science 33: 265-269.
3. Botta R, Molnar TJ, Erdogan V, Valentini N, Torello Marinoni D, et al. (2019) Hazelnut (Corylus spp.) Breeding. In: Al-Khayri JM, Jain SM, Johnson DV Springer International Publishing, Cham 157-219.
4. Gui-Xi W (2018) Progress in Cultivation and Utilization of Corylus L. Resources in China(I)——Corylus Germplasm Resources. Forest Research 31: 130-136.
5. Erdogan V, Mehlenbacher SA (2000) Interspecific Hybridization in Hazelnut (Corylus). Journal of the American Society for Horticultural Science American Society for Horticultural Science 125.
6. Aboul-Maaty NA, Oraby HA (2019) Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bulletin of the National Research Centre 43: 25.
7. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455-477.
9. Altschul SF, Madden TL, Schffer AA, Jinghui Z, Zheng Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25: 3389.
10. Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30: 772.

11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35: 1547-1549.
12. Zhang D, Gao F, Jakovli I, Zou H, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources 20: 348-355.
13. Lam-Tung N, Schmidt HA, Arndt VH, Quang MB (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology & Evolution 268-274.
14. Minh BQ, Nguyen MAT, Haeseler AV (2013) Ultrafast Approximation for Phylogenetic Bootstrap. Molecular Biology and Evolution 30: 1188-1195.
15. Jansen RK, Raubeson LA, Boore JL, Depamphilis CW, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in Enzymology 395:348.
16. Chen, Zhi-Duan, Manchester, Steven, R. (1999) Phylogeny and Evolution of the Betulaceae As Inferred from DNA Sequences, Morphology, and... American Journal of Botany 86: 1168.
17. Holstein N, Tamer SE, Weigend M (2018) The nutty world of hazel names – a critical taxonomic checklist of the genus Corylus (Betulaceae). European Journal of Taxonomy 409: 1-45.