Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/8/20240434

Application of chloroplast genome analysis in plant taxonomy

Yanan Luo

School of Bangor, Central South University of Forestry and Technology, Changsha, Hunan, China, 410004

3097056840@qq.com

Abstract. Chloroplasts are important components of plants and photosynthetic organs, which have played an important role in the long history of biological evolution. Because of this, chloroplast genome research has gradually become the focus of plant research. With the further development of biotechnology, it has been found that the information of chloroplast genome structure and sequence is of great value in revealing the origin of species, evolutionary evolution and the relationship between different species. At the same time, chloroplast transformation technology, which has obvious advantages over nuclear transformation, has shown great potential in genetic improvement and the production of biological agents. Chloroplast genome structure and sequence analysis is the foundation of chloroplast transformation. Based on these important roles of chloroplasts, the application of chloroplast genome in plant classification, variation and distribution analysis was discussed in this paper. At the same time, the current analysis and research methods were summarized and briefly introduced according to different objectives. According to this study, ISSR marker analysis, DNA barcoding technology and high-throughput sequencing technology are still the main biogenic analysis methods for chloroplast gene analysis, and chloroplast transformation is the focus of future research.

Keywords: Chloroplast Genome, Evolutionary Analysis, Classification, Bioinformation, Gene Analysis.

1. Introduction

Chloroplasts, as a kind of endemic organelles with autonomous inheritance, are common in terrestrial plants, algae and some protists. They not only play an important role in energy conversion, but also provide research materials for the study of plant genetics and evolution [1, 2]. They are related to plant photosynthesis, growth and development, and are closely related to plant genetic variation and adaptability. A comprehensive understanding of the chloroplast genome and its role in biological evolution is of great significance to further study and give full play to the chloroplast function in the future. At the same time, a large number of proteins in chloroplasts come from the nuclear genome, which is only a semi-autonomous organelle. The in-depth study of chloroplast genome is helpful to understand the interaction between the nuclear genome and the chloroplast genome [3, 4]. Current studies on chloroplasts of angiosperms have shown that the chloroplast genome of angiosperms is highly conserved in terms of structure, gene content and organization. Therefore, the variation of cpDNA at the intraspecies level can better reflect the evolutionary degree and relationship between different related species. Chloroplast genomes have been widely used in phylogenetic and

DOI: 10.54254/2753-8818/8/20240434

evolutionary studies among different populations. As more and more plant chloroplast whole genomes are sequenced, more valuable information and more new markers with potentially high resolution can be selected for further exploration of phylogenetic relationships and intraspecific diversity in the future. In recent years, the progress of whole chloroplast genome sequencing has provided a new direction for molecular identification and plant classification [5]. With the continuous improvement of sequencing technology and upgrading of sequencing platform, a series of assembly and annotation software such as plasmid SPAdes and NOVOPlasty GetOrganelle have been developed and updated. This also provides new opportunities and challenges for chloroplast genome analysis [6-8]. Based on the important roles of chloroplasts, the application of chloroplast genome in plant classification, variation and distribution analysis was discussed in this paper. The importance of chloroplast genome research is demonstrated through the current cases of orchid plant classification, oak plant geographical distribution analysis, C. tinctorius variation type analysis and tobacco chloroplast gene transformation analysis. The significance of this paper is that it can help readers further understand the application of chloroplast genome analysis, and have a positive effect on the research and wide application of chloroplast genome transformation and variation analysis technology in the future.

2. Chloroplast gene sequencing

2.1. Superior of chloroplast gene sequencing

In recent years, the rapid development of whole chloroplast genome sequencing is mainly due to the superiority of chloroplast genome over nuclear genome in material processing and data analysis. The main advantages are: high expression rate and locatable integration of foreign genes; Can directly express functional genes from prokaryotes; High conversion efficiency; There is no carrier sequence, position effect and pleiotropy. Less gene silencing; Stable descendant material; Genetic pollution caused by pollen drift can be effectively controlled, as well as high environmental safety and other advantages. Studying the evolution of chloroplast genome can provide a more in-depth understanding of the relationship between different plants, the transcription, shear and expression rules of chloroplast genes, and the information exchange between the nuclear genome, etc., which can provide beneficial information for carrier design and improve the homogenization [9].

2.2. Chloroplast genome research area

The chloroplast genome generally has a closed-loop double-stranded deoxyribonucleic acid (DNA) structure, and the chloroplast genome structure of land plants is usually composed of a large single copy region (LSC), a small single copy region (SSC), and a large single copy region (LSC) and two inverted repeat (IR) regions [8]. In the long process of evolution, the structural order of these parts remains unchanged, and the differences between chloroplast genomes of different species are mainly manifested in the length and direction changes of IR regions [10]. In terms of length variation, the IR range of gymnosperms such as Pinus thunbergii has shrunk to only 495 bp, in some legumes, such as Pisum sativum and Medicago truncatula, the IR interval has completely disappeared. In contrast to this formation is an increase in the length of the IR region, for example, Pelargonium hortorum has increased its IR region to 76 kb during evolution, Euglena gracilis contains three tandem repeats in the same direction. There are also variations in the direction of the DNA.IR region. The IR chloroplast genome in Porphyra purpurea has a relatively independent genome and genetic sequence. Unlike the nuclear genome, which generally has complex repetitive sequences, its gene sequences are conserved and there are abundant variation sites in the spacer region, and the appropriate evolutionary rate can provide relatively reliable information for the study of different levels of plant relatives, phylogenetic relationships and genetic diversity. It is found that molecular barcodes based on cp base groups have excellent potential for species identification, and complete cp genome sequences can provide reliable barcodes for accurate identification of plant species and population levels. At the same time, comparing cp genome sequences provides an opportunity to find sequence variation and identify mutation hotspots, as well as detect gene deletions and duplicate events. Therefore, mutation hotspots

DOI: 10.54254/2753-8818/8/20240434

and SSRS obtained from chloroplast genome sequences can be used as effective molecular markers for species identification and population genetics. At present, most people use simple repeat sequence (ISSR) marker analysis to analyze cpDNA variation in different species, which is a relatively simple and available method [10, 11].

3. Current application of chloroplast genome analysis

3.1. Application to plant classification

Orchids, as the largest plant group, occupy a high position in plant classification. The application of chloroplast genome analysis has had a great impact on orchid plant classification. At present, the National Center for Biotechnology Information (NCBI) has uploaded complete chloroplast genome data for more than 1,000 orchid species. According to the current investigation, the proportion in chloroplast genome of Orchidaceae species is about 35.6%, and the differences in chloroplast genes are mainly differences in location and structure [12]. Such as relatively primitive taxa pocket orchid genus IR area significantly increased the size and number of genes but SSC area decreases greatly, some typical SSC genes, such as ycf1 reading frame (assuming the chloroplast genes), psaC (optical system I gene) and ndhD (NADH) REDOX enzyme gene has been transferred to the IR area. These characteristics helped confirm its position in taxonomy. At the same time, the codon preference, gene deletion, duplication and variation of plants are also different due to the joint action of selection and mutation [13]. In order to analyze these differences, more advanced methods include DNA barcoding analysis and chloroplast molecular marker analysis. DNA barcoding is a new technique for rapid and accurate identification of species by using one or several conserved fragments in the genome [14]. Chloroplast DNA sequences (such as matK, ycf5, rbcL, rpoC1, trnH-psbA, rpoB, accD and ndhJ) are widely used in plant DNA barcoding. Molecular marker technology has been widely used in genome mapping, genetic breeding, species identification and so on. Wen et al. developed nine pairs of polymorphic chloroplast microsatellite primers in their study. The polymorphic chloroplast microsatellite primers developed for Dendrobium officinale, an endangered plant, will be a useful tool for studying the genetic diversity, population genetic structure and evolution of Dendrobium officinale and establishing effective conservation strategies [15].

3.2. Application in plant distribution analysis

In addition to plant classification, plant geographic distribution analysis is also used to analyze chloroplast genome variation. The study of three quercus species from Crimean-Caucasus shows that the haplotypes of Crimean and Caucasian quercus have distinct structures, and the chloroplast haplotypes in this region belong to several different phylogenetic lines. The study was compared with other plant data that had been collected, and the effectiveness of various chloroplast fragments and markers was evaluated. In this study, the genotyping scheme was optimized, in which stepwise sequencing, chloroplast DNA microsatellite locus analysis (cpSSR), enzyme digestion analysis (PCR-RFLP), and sequencing were used together for cpDNA detection. This study shows that the analysis of chloroplast genome variation also has a high influence on the analysis of plant geographic distribution [16].

3.3. Application in plant variation analysis

In addition to classification and distribution, chloroplast genome analysis can also be used to analyze plant variation types. In landscape and agriculture, the variation analysis of plants can determine the family and genus of plants, which plays a great role in plant culture and application value exploration. For example, in the study of the Carthamus tinctorius (C. tinctorius) variant type, phylogenetic analysis showed that the 3 variant types of C. tinctorius and the Centaurea diffusa of Cynara form a monophyletic branch. This proved that the medicinal plant C. tinctorius (in Carthamus) belonged to Cynara of the Asteraceae family. In this study, high-throughput sequencing was used to determine the genomic DNA sequence, and the chloroplast genome of the plant was assembled and annotated. The

DOI: 10.54254/2753-8818/8/20240434

chloroplast genome sequence, IR boundary and spacer information loci of three flower colors of the medicinal plant C. tinctorius were analyzed, and the chloroplast genome sequence of the plant was analyzed phylogenically. It is proved that chloroplast variation can not only verify the position of plants in family phylogeny, but also provide a theoretical basis for the identification, development and utilization of plant germplasm resources [17].

3.4. Application in gene transformation analysis

The chloroplast genome has a large number of copies, each gene in the plant chloroplast copy number of up to 10,000, and genes in the IR region are doubled, the proportion of the target protein expressed in the soluble protein can be as high as 46.1%, so it is very suitable for genetic transformation. In addition, according to relevant studies, tobacco, as an important cash crop, has a higher gene expression rate of chloroplast genes than other crops. Due to the low production cost and low probability of gene contamination of chloroplast gene transformation, its related research has brought relatively good feedback in tobacco production and research. Because of this, chloroplast gene transformation as a new method of transgenic research, and related technologies have developed rapidly. According to recent studies, the relevant steps include selection of suitable explants, selection of vector insertion sites, vector construction, gene gun transformation (PEG method can also be used), screening, regeneration, and identification and analysis of transformed offspring [18-20].

4. Conclusion

This paper mainly discusses the application of chloroplast genome in plant classification, variation and distribution analysis, and summarizes the current analysis and research methods according to different objectives. Then, the importance of chloroplast genome research is demonstrated through the current cases of orchid plant classification, oak plant geographical distribution analysis, C. tinctorius variation type analysis and tobacco chloroplast gene transformation analysis. Through the analysis of the above research results, it can be concluded that ISSR marker analysis is still the main means of cpDNA analysis at present, in addition to DNA barcoding technology and high-throughput sequencing technology as the main technical support means to promote the process of research. At the same time, chloroplast genome transformation is a hot project under the background of gene editing technology. Moreover, the research shows that tobacco is likely to become the focus of chloroplast gene transformation research species. Although the chloroplast genome has been widely used, there is still a research gap, and the data in the database only accounts for a small part, which still needs further supplement. Of greater concern is the information on the chloroplast genome, which is still very limited and needs to be further enhanced. Further research will be carried out along the lines of expanding the chloroplast gene bank and the plant gene bank, while new analytical methods and priorities will continue to be developed.

References

[1] Svab Z, Maliga P. Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. Proc Natl Acad Sci USA, 2007, 104(17): 7003～7008

[2] Grevich J J, Daniell H. Chloroplast genetic engineering: recent advances and future perspectives. Critical Reviews in Plant Sciences, 2005, 24: 83～107

[3] Twyford A D, Ness R W. Strategies for complete plastid genome sequencing [J]. Mol Ecol Resour, 2017, 17(5): 858-868.

[4] Yu X Y, Zuo L H, Lu D D, et al. Comparative analysis of chloroplast genomes of five Robinia species: Genome comparative and evolution analysis [J]. Gene, 2019, 689: 141-151.

[5] Ni Lianghong, Zhao Zhili, Mima. Advances in chloroplast genome research of medicinal plants [J]. Chinese Materia Medica, 2015, 38(9): 1990-1994. (In Chinese)

[6] Antipov D, Hartwick N, Shen M, et al. plasmid SPAdes: Assembling plasmids from whole genome sequencing data [J]. Bioinformatics, 2016, 32(22): 3380-3387.

DOI: 10.54254/2753-8818/8/20240434

[7] Dierckxsens N, Mardulyn P, Smits G. NOVOP lasty: De novo assembly of organelle genomes from whole genome data [J]. Nucleic Acids Res, 2017, 45(4): e18.
[8] Jin J J, Yu W B, Yang J B, et al. Get Organelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes [J]. Genome Biol, 2020, 21(1): 241.
[9] Raubeson L A, Jansen R K. Chloroplast genomes of plants. In: Henry R J ed. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. UK:CABI publishing, 2005. 4568
[10] Amiryousefi A., Hyvönen J., and Poczai P., 2018, IRscope: an online program to visualize the junction sites of chloroplast genomes, Bioinformatics, 34(17): 3030-3031.
[11] CBOL Plant Working Group., 2009, A DNA barcode for land plants, Proceedings of the National Academy of Sciences of the United States of America, 106(31): 12794-12797.
[12] Jansen R K, Raubeson L A, Boore J L, et al. Methods for obtaining and analyzing whole chloroplast genome sequences［J］. Methods Enzymol, 2005, 395: 348-384.
[13] Wicke S, Schneeweiss G M, Depamphilis C W, et al. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function［J］. Plant Mol Biol, 2011, 76 (3-5): 273-297.
[14] Li Y X, Li Z H, Schuiteman A, et al. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes［J］. Mol Phylogenet Evol, 2019, https: //doi.org/10.1016/j. ympev.2019.106540.
[15] Liu H, Liu L, Wang Z, etc. Research progress of chloroplast genome in orchids [J]. Chinese Wild Plant Resources, 2019,42(07):73-79.
[16] Huang T Tang M, Chen X et al. Chloroplast genome characteristics and phylogeny of four Quercus species of Cyclobalanopsis [J]. Botany of Guangxi, 2019,43(04):741-754.
[17] Ding Y, Bi G, Hu S et al. Chloroplast genome characteristics and phylogenetic analysis of different flower color variation types of Safflower [J]. Chinese Herbal Medicine, 2019,54(01):262-271. (in Chinese)
[18] Daniell H, Khan M S, Allison L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Science, 2002, 7: 84～91
[19] Daniell H, Khan M S, Allison L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Science, 2002, 7: 84～91
[20] Svab Z, Maliga P. Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. Proc Natl Acad Sci USA, 2007, 104(17): 7003～7008