INNSPUB

Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 15, No. 1, p. 94-100, 2019
http://www.innspub.net

RESEARCH PAPER OPEN ACCESS

Chloroplast genome: An important tool for inferring phylogenetic relationship

Samaila S. Yaradua[*1,2], Dhafer A. Alzahrani[1]

[1]*Department of Biology, King Abdulaziz University, Jeddah, Saudi Arabia*

[2]*Centre for Biodiversity and Conservation, Department of Biology, Umaru Musa Yaradua University, Katsina, Nigeria*

Article published on July 30, 2019

Key words: Chloroplast genome, Phylogenomics, Taxonomic problems.

Abstract

Chloroplast genome is an organelle characterized with definite DNA (cpDNA), whose primary function is to perform photosynthesis. The genome is reported to be derived from cyanobacterium in the process of endosymbiosis and co-evolution. The cp genome characteristic differs among different taxa, its size ranges from 135, 000 to 165, 000 base pairs (bp) with a pair of inverted repeats (IRa and IRb) separated by small single copy (SSC) and large single copy (LSC). Here, we discuss the importance and needs to utilize the chloroplast genome in phylogeny. Considering the taxonomic problem in various taxonomic levels, several studies have proven the viability of the cp genome in resolving taxonomic problems.

***Corresponding Author:** Samaila S. Yaradua ✉ dryaradua@gmail.com

Introduction

According to a theory known as modern evolutionary theory, all life on earth is believed to evolve from a common ancestor, and this this phenomenon explains that all living things on earth are related to one another.

Characters from morphological features that can be observed looking at the phenotype of living organism, as well as the genetic makeup of the living organism known as molecular characteristics, are used to infer phylogenetic relationships. Recently, only the molecular characteristics are being used in phylogenetic analysis because they play an important role in showing relationship among taxa in every taxonomic hierarchy.

The organization and function of the DNA sequences as well as their evolution make them vital tools for inferring evolutionary relationships. Due to advancement in technology, sequencing of DNA became accessible since last decade, and its cost became cheap (http://www.genome.gov/sequencingcosts/). With this effect, generating a different kind of sequences is cheaply for a taxonomist to infer phylogenetic relationship among taxa of interest. As reported by (Eisen and Fraser 2003; Lemmon and Lemmon 2013), researchers can now investigate phylogenetic relationship using comparative genomic analysis called phylogenomics rather than the conventional approach which involve using one or few genes in inferring relationships, because sometimes phylogenetic tree from different genes become incongruent. Genome content and its organization, as well as comparison of DNA sequences are widely used to reconstruct phylogenetic relationships to solve taxonomic problems.

Organellar genome is a good tool to infer phylogenetic relationship because changes that normally occur in genomesuch as deletions and insertion of nucleotides in introns, changes in the gene order are very rare in the organellar genome. Changes that may happen in the complete genome such as duplications of gene and variation in the genetic code are very vital tools to be used as molecular markers for every taxonomic hierarchy (Rokas and Holland, 2000). Result of phylogenetic relationship tress using one or few regions (i.e. conventional approach) sometimes are incongruent (Teichmannand Mitchison 1999) this is as a result of the different evolutionary event that individual gene went through. Additionally, one or few genes may not contain enough phylogenetic informative variation; this result to weakly resolved phylogenetic trees. Contrary, phylogenomics provide and show valid evolutionary relationships and resolve complex phylogenetic relationships that cannot be determined using conventional approach (Delsuc *et al.*, 2005). Several experimental studies have revealed the strength of phylogenomics to determine complicated phylogenetic relationships. For instance, phylogenomic analysis with plastome sequences as reported by Henriquez *et al.*, (2014) was able to determined strongly supported phylogenies of Araceae. Also, the study conducted by Ma *et al.*, (2014) and Pyron *et al.*, (2014) which was done by phylogenomic analyses to solve phylogenetic relationship problems at the genus and species level in the temperate and woody bamboo snakes respectively.

Chloroplast phylogenomics

In the plant cell, apart from the nuclear genome the cell contains two additionally genomes namely chloroplast (the plastid) and mitochondrion genomes. As reported by Dong *et al.*, (2012) the chloroplast genome differs from the other two genomes in many ways, for instance, the chloroplast genome not often show evidence of intra or inter molecular recombination, therefore the organization and content are highly conserved. Due to the mentioned characteristics, the plastid became a vital tool to study phylogenetic relationships. Beside the plastome genome sequence, sequences from nuclear genome such as internal transcribed spacer (ITS) are also combined with sequences of the chloroplast genome in phylogenetic studies. Chloroplast DNA has been known to reveal information on molecular variation for inferring phylogenetic relationships. Previous molecular phylogenetic studies using chloroplast DNA sequences were based on the comparison of restriction site polymorphism and gene order changes

at a wide range of taxonomic levels (Olmstead and Palmer 1994; Jansen *et al*., 1998).

A Brief overview of chloroplast evolution and structure

Chloroplast is the most distinguishing features that differentiate plant cell from other cells; the organelle is semi-autonomous and was derived through endosymbiosis from cyanobacteria over a billion year ago (Timmis *et al*., 2005). The chloroplast genome organization and content are well conserved in both gymnosperms and angiosperms. The genes present in the genome can be classified and grouped based on their functions (Kim and Lee 2004; Yi and Kim 2012; Li *et al*., 2013; Huang *et al*., 2014; Zhang *et al*., 2014). The first class or group comprised of genes that function in photosynthetic activities which involves photosystem I and II. Second group is genes that are involved in central dogma (replication, transcription, and translation) such genes includes transfer RNA, ribosomal RNA, ribosomal subunit genes and RNA polymerase genes (Mullet 1988).Conserved open reading frames (ORFs) which include protein-codinggenes like maturaseK, (*matK*), hypothetical chloroplast open reading frame (*ycfs*) and chloroplast envelope membrane protein (*cemA*) made up the third class.

The structure of chloroplast in flowering plant (angiosperms) is circular in shape (Fig. 1) and its size ranges from 122 to 225 kilobases (kb) as reported by (Wu *et al*., 2010; Wicke *et al*., 2011). The configuration of the genome is quadripartite comprising of large single copy (LSC) and small single copy (SSC) separated by a pair of inverted repeat of the same length (Fig. 1) (Saski *et al*., 2005; Yang *et al*., 2013; Yang *et al*., 2014). In most of the species, the size of the Inverted Repeat is within 20-30 kb with an exception in *Pelargonium hortotum,* which has about 76 kb (Palmer *et al*., 1987; Chumley *et al*., 2006).The positions of boundaries between the single copy regions and inverted repeats also varies among species as shown in many researches, for example *rps19* gene in some species is located in the IR while in others is located in single copy repeat specifically the largest single copy, the two inverted repeat are exact reverse complement duplicates; therefore both of them, have the same gene content as well as their arrangement. The IRs evolved slower than the single copy region and they function as stabilizing region of the chloroplast genome (Perry and Wolfe 2002). Because of the existence of the inverted repeat, the chloroplast genome contained signatures of evolutionary history compared with nuclear and mitochondrial genome this is attributed to the low mutation rate in the genome. Gene mutation does occur in the inverted repeat among species with only one complement of the inverted repeat, in this case, the same substitution rate is comparable to that in the single copy region as reported (Ravi *et al*., 2008).

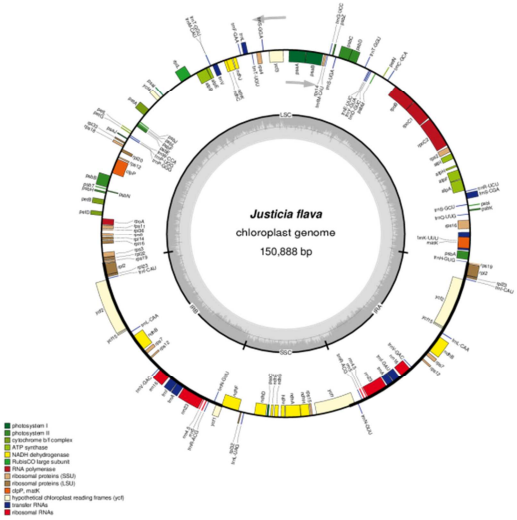

Fig. 1. Gene map of the *J. flava* chloroplast genome. Genes outside the circles are transcribed in counter clockwise direction and those inside in clockwise direction. Known functional genes are indicated in the colored bar. The GC and AT content are denotes by the dark grey and light grey colour in the inner circle respectively. LSC indicates large single copy; SSC, indicates small single copy and IR, indicates inverted repeat.

Chloroplast DNA sequences as molecular markers and their utility in phylogenomics

Phylogeny is the evolutionary history of species which is presented in a phylogenetic tree by comparing homologous characters. These characters are shared between organisms that are descended from a

common ancestor. The characters consist of morphological features, genome organization, biochemical pathways, chemical compounds, genes and the arrangement of nucleotides or amino acids (Delsuc *et al*., 2005). The evolutionary relationship is measured by the differences between the homologous sequences present in different species. Though some of the homologous characters might evolve differently as a result of their biological nature which is attributed by differences in the rate of evolution as well as mutational saturation. Therefore not all of them are a good tool to be used in inferring phylogenetic relationship (Gribaldo and Philippe 2002).A good tool to be used as a maker in phylogenetic studies should provide enough informative sites that is to say the rate of substitution should be optimum, not high to allow comparison. Galtier and Gouy (1995) reported that markers that show true ancestry must be well conserved. An additional important characteristic of good markers is that they must be acquired only through inheritance not by other forms such as horizontal gene transfer or from one organism to another.

DNA sequences from chloroplast genome have been used as a primary tool to infer evolutionary relationships in several phylogenetics studies of the plant. This is due to their evolution and conservesnature which enables them to preserve phylogenetic signals that show evolutionary relationships. In addition, evolutionary processes like pseudogene formation, rearrangements in genome and gene duplication do occur in the chloroplast genome but are not very common as in the case of other genomes like the nuclear genome (Palmer, 1985). The conserved nature of the chloroplast organellar genome allows designing of primers targeting region conserved well beyond species boundaries, and amplification of molecular makers. In spite of the low evolutionary rate in the plastome compared to its counterpart the mitochondrial and nuclear genome, its small size and high frequency in the plant tissue such as the leaf make it easier to sequence the genome.

Recently, breakthroughs in sequencing technology have provided a means of rapid sequencing of the complete plastid genome, thereby making it possible to use the plastid genome in the phylogenetic analysis (phylogenomics). This system is now used widely as a common means of species identification (Wu *et al*., 2010; Nock *et al*., 2011), systematic studies and in phylogenetic analysis of plants (Jansen *et al*., 2007; Moore *et al*., 2007). Using the complete chloroplast genome (phylogenomics) provide better statistical support and result in indicating a valid relationship with minimized sample errors that occur when using one or few genes (Blair *et al*., 2002; Wolf *et al*., 2004). Though, it doesn't mean that well supported phylogenetic tree from complete chloroplast genome is always correctbecause systematic errors occur with the increase in the data set (Philippe *et al*., 2005; Jeffroy *et al*., 2006; Brinkmann and Philippe 2008). Violations of the model used are the factors responsible for systematic error. If there is a model violation, the phylogenetic tree will have errors which compute with the genuine phylogenetic signal, thereby showing a phylogenetic relationship that is not correct (Delsuc *et al*., 2005). There are different kinds of model violation as reported by Rodríguez-Ezpeleta *et al*., (2007), they include site heterogenous amino/acid or nucleotide sequence replacement, site interdependent evolution, cross-site rate variation and compositional heterogeneity among others. Long branch attraction is an example of systematic error that occurs in most phylogenetic analyses, as reported in (Felsenstein, 1978) it occurs when the result of a phylogenetic tree shows taxa that evolve faster than other species are closely related.

Strongly supported artificial nodes are as a result of the presence of systematic errors which mainly occur due to the increase in the number of data. Identification of systematic error in phylogenetic tree from a single gene is easy and can be achieved by observing incongruence in the phylogenetic tree from two different genes. This cannot be applied in phylogenomics because the genome contains all the genes. Data partitioning strategy and different tree reconstruction model are some of the approaches suggested to detect systematic errors in phylogenomics. The advantage of using

complete chloroplast genome is that any changes in the genome can be easily observed due to its conserved nature, this feature can be useful in showing phylogenetic relationship among taxa in different taxonomic level (Jansen *et al.*, 2005; Philippe *et al.*, 2005; Jansen *et al.*, 2007).

Conclusion

A complete chloroplast genome is a promising tool in inferring phylogenetic relationship to solve various taxonomic problems at different taxonomic level. Genes and simple sequence repeat in the genome are also important tools for identification and classification of species as well as in genetic diversity studies. More studies and sequencing of cp genome of unexplored taxa are required to have a better understanding about phylogenetic relationship among taxa.

References

Blair JE, Ikeo K, Gojobori T, Hedges SB. 2002. The evolutionary position of nematodes. BMC Evolutionary Biology **2,** 7.

Brinkmann H, Philippe H. 2008. Animal phylogeny and large-scale sequencing: progress and pitfalls. Journal of Systematic and Evolution **46,** 274-286.

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. 2006. The complete chloroplast genome sequence of *Pelargonium* x hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Molecular Biology and Evolution **23,** 2175-2190.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nature Review Genetics **6,** 361-375.

Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One **7,** e35071.

Eisen JA, Fraser CM. 2003. Phylogenomics: intersection of evolution and genomics. Science **300,** 1706-1707.

Felsenstein J. 1978. Cases in which Parsimony or Compatibility Methods will be Positively Misleading. Systematic Biology **27,** 401-410.

Galtier N, Gouy M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proceeding of the National Academy of Science of the United States of America **92,** 11317-11321.

Gribaldo S, Philippe H. 2002. Ancient Phylogenetic Relationships. Theoretical Population Biology **61,** 391-408.

Henriquez CL, Arias T, Pires JC, Croat TB, Schaal BA. 2014. Phylogenomics of the plant family Araceae. Molecular Phylogenetic and Evolution **75,** 91-102.

Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology **14,** 151.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceeding of the National Academy of Science of the United States of America **104,** 19369-19374.

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in Enzymology **395,** 348-384.

Jansen RK, Wee JL, Millie D. 1998. Molecular systematics of plants II. In: Soltis DE, Soltis PS, Doyle JJ, editors. Springer US. p. 87-100.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence. Trends in Genetics **22,** 225-231.

Kim KJ, Lee HL. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panaxschinseng nees*) and comparative analysis of sequence evolution among 17 vascular plants. DNA Research **11,** 247-261.

Lemmon EM, Lemmon AR. 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. Annual Review of Ecology Evolution and Systematics **44,** 99-121.

Li R, Ma PF, Wen J, Yi TS. 2013. Complete Sequencing of Five Araliaceae Chloroplast Genomes and the Phylogenetic Implications. PLoS One **8,** 1-15.

Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe

Moore MJ, Bel CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceeding of the National Academy of Science of the United States of America **104,** 19363-19368.

Mullet JE. 1988. Chloroplast Development and Gene Expression **1967,** 475-502.

Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnology Journal **9,** 328-333.

Olmstead RG, Palmer JD. 1994. Chloroplast DNA systematics: A review of methods and data analysis. American Journal of Botany **81,** 1205-1224.

Palmer JD, Nugent JM, Herbon LA. 1987. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proceeding of the National Academy of Science of the United States of America **84,** 769-773.

Palmer JD. 1985. Chloroplast DNA and molecular phylogeny. Bio Essays **2,** 263-267.

Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. Journal of Molecular Evolution **55,** 501-508.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. Phylogenomics. Annual Review in Ecology, Evolution and Systematic **36,** 541-562.

Pyron RA, Hendry CR, Chou VM, Lemmon EM, Lemmon AR, Burbrink FT. 2014. Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). Molecular Phylogenetic and Evolution **81,** 221-231.

Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. Plant Systematic and Evolution **271,** 101-122.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Systematic Biology **56,** 389-399.

Rokas A, Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. Trends in Ecology and Evolution **15,** 454-459.

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK. 2005. Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. Plant Molecular Biology **59,** 309-322.

Teichmann SA, Mitchison G. 1999. Is there a phylogenetic signal in prokaryote proteins? Journal of Molecular Evolution **49,** 98-107.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Review Genetics **5,** 123, 135.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Molecular Biology **76,** 273-297.

Wolf YI, Rogozin IB, Koonin EV. 2004. Coelomata and not ecdysozoa: Evidence from genome-wide phylogenetic analysis. Genome Research **14,** 29-36.

Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H, Duvall MR, Lin CS. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biology **10,** 68.

Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Molecular Ecology Resources **5,** 1024-1031.

Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evolutionary Biology **13,** 84.

Yi DK, Kim KJ. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS One 7. Arundinarieae (poaceae). Systematic Biology **63,** 933-950.

Zhang Y, Ma J, Yang B, Li R, Zhu W, Sun L, Tian J, Zhang L. 2014. The complete chloroplast genome sequence of *Taxuschinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene **540,** 201-209.