African Journal of Biotechnology Vol. 9 (24), pp. 3494-3500, 14 June, 2010
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJBx09.057
ISSN 1684–5315 © 2010 Academic Journals

*Review*

# Current trends in chloroplast genome research

**Asifullah Khan, Ishtiaq A. Khan, Huma Asif and M. Kamran Azim***

H.E.J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan.

Accepted 21 December, 2009

**Chloroplast is an important cellular organelle of autotrophs which has an independent, circular, double-stranded DNA molecule termed as chloroplast genome. The chloroplast DNA (cpDNA) contains essential genes for its maintenance and operation. Several components of the photosystems and proteins involved in biosynthetic pathways are also encoded by the chloroplast genome. Exploring the genetic repository of this organelle is vital due to its conserved nature, small size, persistent gene organization and promising ability for transgenic expression. Therefore, cpDNA sequence information has been instrumental in phylogenetic studies and molecular taxonomy of plants. Chloroplast genome sequencing efforts have being initiated with conventional cloning and chain-termination sequencing technologies. Dedicated databases such as CGDB and GOBASE among others have been established as more and more complete cpDNA sequences are being reported. Presently, elegant molecular biology techniques including shotgun sequencing, rolling circle amplification (RCA), Amplification, Sequencing and Annotation of Plasteome (ASAP) and Next generation sequencing are being used to accelerate data output. Owing to many fold increase in submission of cpDNA sequences in nucleotide databases, challenges of in-depth data analysis stimulated the emergence of devoted annotation, assembling and phylogenetic software. Recently, reported bioinformatics software for chloroplast genome studies comprise of DOGMA for annotation, SCAN-SE, ARAGON and PREP suit for RNA analyses and CG viewer for circular map construction/comparative analysis. Faster algorithms for gene-order based phylogenetic reconstruction and bootstrap analysis have attracted the attention of research community. Current trends in sequencing strategies and bioinformatics with reference to chloroplast genomes hold great potential to illuminate more hidden corners of this ancient cell organelle.**

**Key words:** Organelle genomics, sequencing strategies, sequence analysis tools, plant molecular biology.

## INTRODUCTION

Chloroplast is an essential cellular organelle found in photosynthetic algae and plant cells (Xiong, 2009; Sugiura, 2003). According to evolutionary perception, chloroplast evolved by endosymbiosis between non-photosynthetic protists and photosynthetic cyanobacteria (Howe, 2003; Xiong, 2009; Raven, 2003). In addition to nuclear and mitochondrial genomes, photosynthetic organisms contain independent chloroplast genomic DNA (cpDNA). In higher plants, predominantly, the inheritance of chloroplast genomes occur through maternal or paternal parent but in some cases biparental inheritance has been reported (McKinnon, 2001; Hansen, 2007a). The cpDNA of higher plants is a double stranded, circular molecule, ranging in size of 120 - 160 kb (Odintsova and Yurina, 2006). Among angiosperms, this genome is highly conserved in size, structure and gene content (Olmstead and Palmer, 1994). Typical chloroplast DNA consists of large and small single copy regions (LSC, SSC respectively) separated by two duplicated inverted repeat regions (IRA and IRB) (Ravi, 2008) (Figure 1). Dinoflagellates, a group of algae have structurally diverse chloroplast genome, where genes are located on separate minicircular DNA rather than a single large circular molecule (Zhang, 1999; Howe, 2008). With the average size of 20-30 kb, the inverted repeat (IR) regions of cpDNA of land plants are highly conserved. The IR regions are responsible for the size variation among chloroplast genomes (Odintsova and Yurina, 2006; Ogihara et.

*Corresponding author. E-mail: kamran.azim@iccs.edu or mkamranazim@yahoo.co.uk. Tel: +9221-111222292 ext. 236. Fax: +9221-4819018.

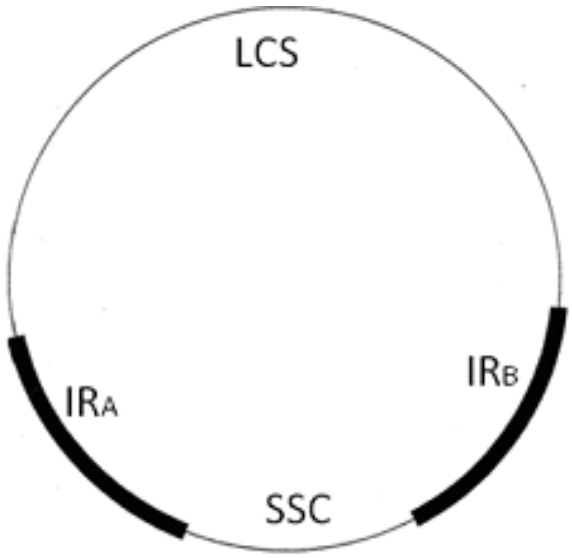


**Figure 1.** Generalized map of the chloroplast genome. IRA = inverted repeat region A; IRB = inverted repeat region B; LSC = large single copy region; SSC = small single copy region.

al., 2002). Moreover, the boundaries of IR regions with LSC and SSC are also important for expansion and contraction in genome and hence the gene content of this region varies greatly (Goulding et al., 1996, Plunkett and Downie, 2000).

Genome analysis of chloroplast DNA has revealed 60-200 open reading frames (ORFs) (Leister, 2003). According to their functions, chloroplast genes are divided into three groups which are the genes involved in photosynthesis, genes for the transcription/ translation system and genes related to photosynthetic metabolism (Odintsova and Yurina, 2006). Chloroplast genomes contain only 10% of the genes required for fully functional organelle whereas the remaining proteins in chloroplasts are encoded by nuclear genome (Jarvis and Robinson, 2004). Gene transfer from chloroplast genome to the nuclear genome has been reported resulting in reduction of size of cpDNA. Reduced size of dinoflagellates chloroplast genomes might have arisen on the basis of this concept (McFadden, 1999; Maliga, 2003; Howe, 2008).

Currently 170 chloroplast genomes from different species have been completely sequenced (NCBI Organelle Genome Resources; http://www.ncbi.nlm.nih.gov/genomes/). In the last few years, large increment in the GENBANK entries of chloroplast DNA sequences has been observed (Figure 2). It is noteworthy that > 70% known chloroplast genomic sequences has been submitted in the nucleotide databases during the last three years (Figure 2). This is primarily due to innovations in (a) chloroplast DNA sequencing and (b) bioinformatics tools for analysis of the sequencing data. In this review, new horizons being opened in this field of plant molecular biology by describing modern trends in chloroplast genome sequencing, data mining and analysis was summarised.

## CURRENT TRENDS IN cpDNA SEQUENCING

Recently, significant progress has been observed in chloroplast genome sequencing strategies. Traditionally, cp DNA is purified as an initial step for sequencing (Jansen, 2005). The purified cpDNA is then subjected to restriction endonucleases or random shearing followed by cloning of the resulting fragments in cloning vectors. The clones containing cpDNA fragments are then subjected to sanger-based sequencing. This approach has been used with some modification for sequencing of many chloroplast genomes. The sequencing of complete chloroplast genome of tobacco *(Nicotiana tobacum)* was reported using this strategy where overlapping restriction fragments of cpDNA were cloned in pHC79 vector for sequencing (Shinozaki, 1986). Similarly, *Zea mays* chloroplast genome was sequenced by cloning of overlapping cpDNA fragments into pUC19 and pKSII-Bluescript vectors (Maier, 1995). Sequencing of the Chinese wheat (*Triticum aestivum L.*) cpDNA was carried out by cloning the long restriction fragments in pBR322 and pGEM-T plasmids and then subcloning of the long cloned fragments into pBluescript or pUC vector for sequencing (Ogihara 2000). Korean Ginseng *(Panax schinseng Nees)* cpDNA was sequenced by cloning the *BamHI*, *SacI* and *ClaI* restriction fragments into pBluescript II vector followed by shotgun sequencing (Ki-Joong, 2004). In case of cucumber (*Cucumis sativus L.*) chloroplast genome, a fosmid library was initially generated by shearing the purified cpDNA into approximately 40 kb fragments and then the products of fosmid library was subsequently used to construct shotgun library to get complete genome sequence (Jin-Seog, 2005). Sequencing of *Solanum tuberosum* chloroplast genome was carried out by constructing shotgun library from PCR products of ~4 - 5 kb followed by random shearing into ~1 kb fragments and after end repairing these fragments, were cloned into pUC118 vector for sequencing (Chung, 2006).

Recently, sequencing of several chloroplast genomes has been reported with involvement of rolling circle amplification (RCA) step which improves the initial DNA template for sequencing (Jansen, 2005). During the sequencing of chloroplast genomes of *Gossypium hirsutum* (Lee, 2006) and *Citrus sinensis* (Baucher, 2006), purified cpDNA was initially subjected to RCA. The products of RCA were then digested by endonucleases and cloned for sequencing. During sequencing of *Huperzia lucidula and Welwitschia mirabilis* chloroplast genomes, RCA was employed followed by shotgun sequencing (Wolf, 2005; McCay, 2008).

Several researchers have utilized the approach of cloning long PCR products for sequencing purposes. Initially, different regions of chloroplast genome of size

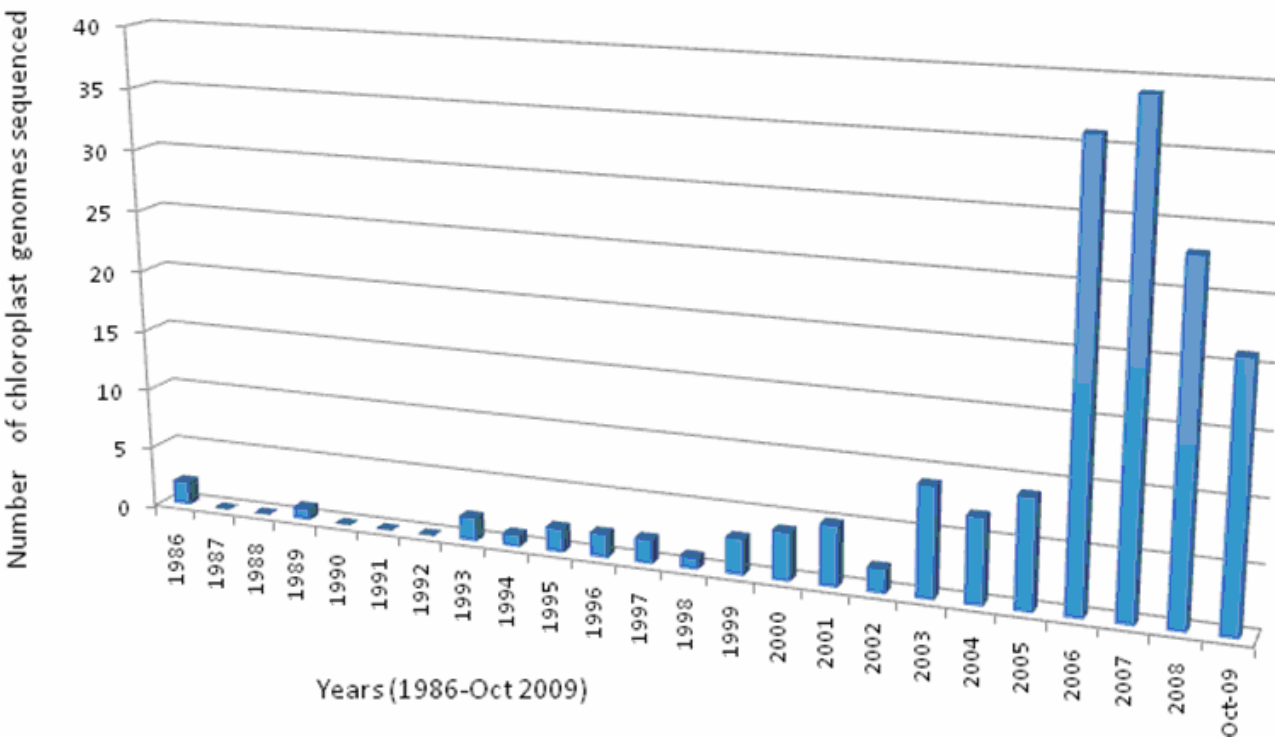


**Figure 2.** Completed chloroplast genome sequences submitted to Genbank during 1986 - October 2009.

range of 4 - 20 kb were amplified using total cellular DNA as template. For this purpose, consensus primer sets have been developed from alignment of known cpDNA sequences. The resultant long PCR products were then sheared into small fragments of 0.5 - 1.5 kb and cloned for sequencing purposes. Goremykin et al. (2003a, 2003b, 2005) have successfully adopted this strategy for chloroplast genome sequencing of several plants. Cattolico et al. (2008) sequenced the chloroplast genome of *Heterosigma akashiwo* with fosmid cloning approach, in which total genomic DNA was cloned in the form of large inserts called fosmid. The clones containing chloroplast genome fragments were then selected from the fosmid library by end sequencing of inserts and then the resultant sequences was compared with submitted sequences using BLAST searches (Cattolico et al. 2008). Although the fosmid library based approach seems to be quite valuable, the restriction digestion, cloning and subsequent chloroplast genome insert fosmid screening make it laborious and costly.

The availability of a large number of conserved cpDNA sequence information from diverse plant genera provided a basis for designing of cognate primer pairs for PCR amplification. An online database of PCR primers containing > 500 primer sequences has been developed to be employed in cpDNA sequencing (Heinze 2007) (http://bwf.ac.at/200/1859.html). Dhingra and Folta (2005) have introduced a new PCR based strategy called ASAP (Amplification, Sequencing and Annotation of Plasteome) for the sequencing of inverted repeat region (IR) of cp DNA. They designed 27 pairs of cognate primers, which resulted in a generation of 1-1.2 kb overlapping amplicons of IR region from 14 diverse genera using the total cellular DNA as template (Dhingra and Folta, 2005). This rapid, straightforward and inexpensive method has been used recently for cpDNA sequencing of several plant species (Masood, 2004; Chung, 2007; Mardanov, 2008; Logacheva, 2008). This includes the chloroplast genome sequencing of *Dendrocalamus latiflorus*, *Bambusa oldhamii* (Wu, 2009) and *Coix lacryma-joba* (Leseberg, 2009). Moreover, there is no need of purifying cpDNA, restriction digestion, cloning and colony screening. In the present review, the ASAP strategy for sequencing the IR regions of cpDNA of mango (*Mangifera indica*) and date palm (*Pheonix dactylifera*) (GenBank Accession numbers, *M. indica*, EF205595; *P. dactylifera*, FJ 212316) was adopted.

Rapid genome sequencing is one of the major challenges in contemporary genomic research and next generation sequencing technologies (NGSTs) have emerged to meet this (Church, 2005). Although, Sanger method-based DNA sequencing is the gold standard for large sequencing projects, it requires a large infrastructure and manpower including cloning of DNA into vectors, growth of host and purification of vectors. In the last few years, the NGSTs have been developed and adopted in many sequencing centers around the world (Shendure et. al., 2004). In the NGSTs, several methods developed in

parallel. There are two concepts at the core of NGST, which permitted its dramatic increase in throughput over the traditional Sanger sequencing. These are (1) DNA amplification without cloning and (2) DNA sequencing without chain termination. In NGSTs, each DNA fragment amplified by PCR independently followed by sequencing either by synthesis or by ligation (termed as sequencing-by-synthesis technology and multiplex polony sequencing protocol respectively) (Chan, 2005). This new technique performs sequencing in parallel using picotiter plates which results in a huge increase in sequence data compared to the contemporary Sanger sequencer. A next generation sequencer can easily generate a through put equivalent to that of more than 50 ABI 3730XL sequencers. Initially, this technology was adopted by large sequencing centers in North America and in Europe (Shendure et. al., 2004). Since 2007, even for smaller laboratories, sequencing using NGSTs started seeming more affordable. The NGSTs can be applied to a broad range of projects other than de novo sequencing such as functional genomics, metagenomics, comparative genomics, among others. (Mardis, 2007; Chan, 2005). With the advent of NGSTs, the chloroplast genomics entered into a new phase of sequencing. The next generation sequencing can easily be applied to cpDNA sequencing due to its conservation and availability of reference genome sequence data from many plant genera. Chloroplast genome sequencing of *Nandica domestica* and *Patanus occidentials* has been reported using NGSTs (Moore, 2006). Cronn, (2008) has sequenced cpDNA of eight plant species by multiplex tagging method through Solexa next generation technology. NGSTs coupled with multiplexing have dramatically reduced the time and cost to get the entire chloroplast genome sequences compared to 'conventional' approaches.

## CURRENT BIOINFORMATICS TOOLS FOR cpDNA SEQUENCE ANALYSES

While advancement in sequencing techniques are augmenting the rate of achieving cpDNA sequence data, the developments of bioinformatics tools for the analyses of resulting sequences are also important. Designing computer based tools are critical for functional and evolutionary perception of genome sequences. Genbank, EMBL and DDBJ, the primary nucleotide sequences databases, include sections for organelle genomes (http://www.ncbi.nlm.nih.gov/genomes/). The chloroplast genome database (CGDB: http://chloroplast.cbio.psu.edu/) and GOBASE (http://gobase.bcm.umontreal.ca/) are specialized chloroplast genetic repositories. These databases contain complete genome sequences, sequence alignments, PCR primer sequences database, gene families and other information related to complete chloroplast genomes (Cui, 2006; Heinze, 2007; Emmet, 2008). A database of 32 completely sequenced plastid DNA sequences has facilitated the annotation of chloroplast genomes (Kunisawa, 2004).

Dual Organelle genome Annotator (DOGMA) is a widely used web based annotation tool for mitochondrial and chloroplast genome analysis (Wyman, 2004). DOGMA can be used for rRNA, tRNA and protein genes detection. Glimer 2.0 (Delcher, 1999) and PFAM (Finn, 2006) have been developed for detection of open reading frames and conserved protein motifs in cpDNA sequences, respectively (Cattolico et al., 2008). Artemis and Artemis comparision tool (ACT) (Caver, 2005) originally developed for viewing and annotating of microbial genome has been applied for analysis of dignoflagellate cpDNA (Barbrook, 2006).

Chloroplast genome encodes tRNA, rRNA and in some species even tmRNA (Barbrook, 2006). Several expert tools for accurate annotation and analysis of different classes of RNA genes have been developed. SCAN-SE (Lowe and Eddy, 1997) is an interactive program for accurate identification of tRNA genes within DNA sequences and is being used in chloroplast genome analysis (Cattolico et al., 2008). Comparative RNA web (CRW) is an online database for annotation of 16S, 23S and 5S rRNA, tRNA and group-I and group-II introns. It is also used for phylogenetic studies on the basis of structurally diverse models of RNA types in this collection (Cannonel, 2002). ARAGORN is a heuristic algorithm developed for searching concurrent tRNA and tmRNA genes within a given sequence (Laslett, 2004) and has been applied to cpDNA sequences (Laslett, 2004; Lei, 2009). Another web service TFAM predicts the functional aspects of tRNAs on the basis of their sequences and anticodons (Taquist, 2007).

The remarkable post-transcriptional phenomenon of RNA editing has been observed in chloroplast genome of land plants (Tillich, 2006; Blanc and Davidson, 2003). The knowledge of RNA editing is important for accurate annotation of proteins. A list of RNA edits found in chloroplast genome is available in chloroplast genome database (Cui, 2006). CURE-Chloroplast is a useful tool developed for prediction of RNA editing in chloroplast genes. Cytidine to Uracil (C-to-U) are commonly found RNA edits observed in chloroplast mRNA transcripts and CURE-Chloroplast accurately predicts C-to-U RNA editing sites (Du, 2009). The PREP-Cp tool of PREP suit (Predictive RNA editors for plants) performs RNA editing prediction in plant chloroplast genes (Mower, 2009).

Comparative chloroplast genomics provides basis for phylogenetic, genotyping, genome mapping and functional genomics analysis in autotrophs. BLAST2 has been widely used for genome-level comparisons. However, it has limitation during comparison of sequences containing mononucleotide repeats greater than six nucleotides. Moreover, it also fails to show insertion deletion events (indel) greater than ten nucleotides (Raubenson, 2007). MULAN bioinformatics tool (Ovcharenko, 2005) overcame this issue by Indels analysis

up to 20 bp or more over the entire genome (Raubenson, 2007). CoreGene is another web resource to identify and catalogue set of core genes from two to five small genome sequences that is, mitochondria, chloroplast and viruses (Zafar, 2002). GeneOrder 3.0 software performs the identification of genes re-arrangements in small genomes. It is an interactive tool that can be utilized for comparative analysis of chloroplast genomes (Celamkoti, 2004). The combine use of CoreGenes and GeneOrder would be beneficial for understanding correlation and evolutionary distances in comparative chloroplast genomics (Keil, 2004). Reputer is a commonly used tool for repeat analysis (Kurtz, 2001). The direct and inverted repeats of chloroplast genome can be identified using Comparative Repeat Analysis program (http://bugmaster.jgi-psf.org/repeats/) which basically uses REPuter along with some additional features as well. The simple sequence repeat (SSR) can be characterized by using simple sequence repeat (SSR) Extractor utility (http://www.aridolan.com/ssr/ssr.aspx) (Raubenson, 2007).

A number of computational tools have developed for both visualization and comparison of circular genomes, including chloroplast. With the help of these tools, one can view and get information about the complete genome in a single glance. The program CGView (Stothard, 2005) has been applied for creating graphical maps of circular cpDNA (Cattolico et al., 2008). In addition to generation of circular maps, recently developed CGView Server (Grant, 2008) can compare up to three chloroplast genomes. The circular gene map can also be drawn using Genome Vx tool (Conant, 2008).

Availability of sufficient sequence data, predominantly, maternal inheritance and small size made chloroplast genome are important for phylogenetic reconstruction and evolutionary analysis. Efficient bioinformatics tools have been developed to utilize chloroplast sequence information for phylogenetic studies. GRAPPA and MGR performs phylogenetic tree construction from gene order changes (Moret, 2001, Yue, 2008a). DCM-GRAPPA is an advanced version of GRAPPA for high accuracy phylogenetic reconstruction of large number of organelles genomes (Tang, 2003). GRAPPA-IR is an extension in GRAPPA (Yue, 2008b) which specifically handles inverted repeat regions of cpDNA for phylogenetic purposes, as the boundaries of inverted repeat regions are hot spot for deletion and gene duplication (Plunkett and Downie, 2000). GRAPPA-TP is another extension of GRAPPA for phylogenetic reconstruction on the basis of transposable elements (Yue, 2008b). Recently, a fast algorithm for constructing phylogenetic tree on the basis of reversal and transposition types of genome rearrangements was developed. This new algorithm was checked for unichromosomal genome including chloroplast and found to be more elegant than DCM-GRAPPA and MGR (Bader 2008). Genome BLAST Distance Phylogeney (GBDP) (Henz, 2005) has also been successfully applied to compute accurate phylogenetic trees from all available completely sequenced plastid genomes (Auch, 2006). Bootstrapping analysis evaluates the strength of support for nodes on phylogenetic trees and it is widely used for tree estimation (Efron, 1996; Holmes, 2003). Quick Tree program produces bootstrapped neighbor-joining trees (Howe, 2002; Maruyama, 2009). Composition Vector Tree (CVTree) web server performs bootstrap test for phylogenetic tree construction, which is important for inferring evolutionary relatedness of complete microbial proteomes (Qi, 2004). Recently, a rapid bootstrap (RBS) heuristic algorithm was developed in RAxML (Randomized Axelerated Maximum Likelihood), which is comparatively faster than standard bootstrapping algorithms (Alexandros, 2008). Bayesian interference (BI) method based phylogenetic analyses are mostly performed using MrBayes (Huelsenbeck and Ronquist, 2001) while for Maximum parsimony (MP) and Maximum likelihood (ML) studies PAUP and Phylip phylogenetic packages are used, respectively (Swofford, 2003). Besides these, GARLI (Zwickl, 2006), MEGA4 (Tamura, 2007) and RAxML (Alexandros, 2008) tools have also being adopted for phylogenetic studies in case of chloroplast genomics (Hansen 2007b, Wu 2009).

## REFERENCES


Alexandros S (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst. Biol. 57(5):758-771.

Auch AF (2006). Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. BMC Bioinformatics, 7: p. 350.

Baucher MG (2006). The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. BMC Plant Biol. 6: p. 21.

Barbrook AC (2006). Comparative analysis of dinoflagellate chloroplast genome reveals rRNA and tRNA genes. BMC Genomics, 7: 297.

Blanc V, Davidson NO (2003). C-to-U RNA editing: Mechanisms leading to genetic diversity. J. Biol. Chem. 278(3): 1395-1398.

Bader B (2008). A fast algorithm for the multiple genome rearrangement problem with weighted reversals and transpositions. BMC Bioinformatics, 9: 516.

Cannonel JJ (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron andother RNAs. BMC Bioinformatics, 3: 2.

Caver TJ (2005). ACT: the Artemis comparing tool. Genome Anal. 21(16).

Conant GC (2008). Genome Vx: simple web-based creation of editable circular chromosome maps. Bioinformatics, 24: 861-862.

Cui L (2006). ChloroplastDB: the Chloroplast Genome Database. Nucleic Acids Res. 34 (Database issue).

Celamkoti S (2004). GeneOrder 3.0: Software for comparing the order of genes in pairs of small bacterial `genomes. BMC Bioinformatics, 5: 52.

Church GM (2005). Genomes for all. Sci. Am. pp. 46-54.

Chan EY (2005). Advances in sequencing technology. Mutat. Res. 573: 13-40.

Cronn R (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing by synthesis technology. Nucleic Acids Res. 36(19).

Chung SM (2007). Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies differences between chilling-tolerant and susceptible cucumber lines. Genome, 50: 215-225.

Chung HJ (2006). The complete chloroplast genome sequences of *Solanum tubersoum* and comparative analysis with Solanaceae specie identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep. 25: 1369-1379.

Catholico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, Lybrand T, McKay J, Ong HC, Sims E, Rocap G (2008). Chloropalst genome sequencing analysis of Heterosigma akashiwo CCMP 452 (West Atlantic) and NIES 293 (West Pacific) strains. BMC Genomics, 9: 211.
Delcher AL (1999). Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27(23).
Dhingra A, Folta KM (2005). ASAP: Amplification, sequencing & annotation of plastomes. BMC Genomics, 6: 176.
Du P (2009). CURE-Chloroplast: A chloroplast C-to-U RNA editing predictor for seed plants. BMC Bioinformatics, 10: 135.
Emmet A (2008). GOBASE: an organelle genome database. Nucleic Acids Res. pp. 1-5.
Efron B (1996). Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA, p. 93.
Finn RD (2006). Pfam: clans, web tools and services. Nucleic Acid Res. 34(Database issue): D247-D251,
Grant JR (2008). The CG View Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36(Web Server issue).
Goremykin VV, Hirsch-Ernst KI, Wolfl S (2003a). Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that Amborella is not a basal angiosperm. Mol. Biol. Evol. 20(9): 1499-1505.
Goremykin VV,Hirsch-Ernst KI, Wolfl S, Hellwig FH (2003b).The chloroplast genome of the basal angiosperm *Calycanthus fertilis*-structural and phylogenetic analyses. Plant Syst. Evol. 242: 119-135.
Goremykin VV,Holland B, Hirsch-Ernst KI, Hellwig FH (2005). Analysis of *Acorus calamus* Chloroplast Genome and Its Phylogenetic Implications. Mol. Biol. Evol. 22(9): 1813-1822.
Goulding S, Olmstead R, Morden C, Wolfe K (1996). Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 252: 195-206.
Hansen AK (2007a). Paternal, Maternal and Biparenat inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implication for Phylogenetic studies. Am. J. Bot. 94(1): 42-46.
Hansen DR (2007b). Phylogenetic and evolutionary implication of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae) and *Illicium* (Schisandraceae). Mol. Phylogenet. Evol. 45: 547-563.
Heinze B (2007). A database of PCR primers for the chloroplast genomes of higher plants. Plant Methods, 3: p. 4.
Henz SR (2005). Whole genome-based prokaryotic phylogeney. Bioinformatics, 21: 2329-2335.
Holmes S (2003). Bootstrapping Phylogenetic Trees: Theory and Methods. Stat. Sci. 18(2): 241-255.
Howe CJ (2003). Evolution of the chloroplast genome. Phil. Trans. R. Soc. Lond. 358: 99-107.
Howe CJ (2008). The remarkable chloroplast genome of Dinoflagellates. J. Exp. Bot. pp. 1-11.
Howe K (2002). QuickTree: building huge Neighbour-Joining trees of protein sequences. Bioinformatics, 18: 1546-1547.
Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17: 754-755.
Jarvis P, Robinson C (2004). Mechanisms of protein import and routing in chloroplast. Curr. Biol. 14: 1064-1077.
Jansen RK (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol. 395: 348-384.
Jin-Seog K (2005). Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. Plant Cell Rep. 25(4):334-40.
Ki-Joong K (2004). Complete chloroplast genome sequences from Korean Ginseng (Panax schinseng Nees) and comparative Analysis of Sequence Evolution among 17 Vascular Plants. DNA Res. 11: 247-261.
Kunisawa K, Kunisawa T (2004). Gene order database of plastid genomes. Data Sci. J. p. 3.
Keil B (2004). Comparative analysis and relationships of six important crop species chloroplast genomes using whole genome web-based informatics tools. Afr. J. Biotechnol. 3(4): 210-214.
Kurtz S (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29: 4633-4642.
Lei G (2009). Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insight into evolutionary changes in fern chloroplast genomes. BMC Evol. Biol. 9: 130.
Laslett D (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32(1):
Lee SB (2006). Complete chloroplast genome sequencing of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. BMC Genomics, 7: 61.
Leseberg CH (2009). The complete chloroplast genome of Coix lacryma-jobi and a comparative molecular evolutionary analysis of plastomes in cereals. J. Mol. Evol. 69: 311-318.
Leister D (2003). Chloroplast research in the genomic age. Trends Genet. 19(1).
Logacheva MD (2008). Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. *ancestrale*-A wild ancestor of cultivated buckwheat. BMC Plant Biol. 8: 59.
Lowe TM, Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5).
Maliga P (2003). Mobile plastids genes. Nature, p. 422.
Maier RM (1995). Complete Sequence of the Maize Chloroplast Genome: Gene Content, Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing. J. Mol. Biol. 251: 614-628.
Masood SM (2004). The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene, 340: 133-139.
Mardanov AV (2008). Complete Sequence of the Duckweed (Lemna minor) Chloroplast Genome: Structural Organization and Phylogenetic Relationships to Other Angiosperms. J. Mol. Evol. 66: 555-564.
Mardis ER (2007). The impact of next-generation sequencing technology on genetics. Trends Genet. 24: 3.
McCay SR (2008). The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evolutionary Biol. 8: 130.
McFadden G (1999). Ever decreasing circles. Nature, p. 400.
McKinnon GE (2001). Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. Genome, 44: 831-835.
Moore MJ (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. 6: p. 17.
Maruyama S (2009). Cyanobacterial contribution to the genomes of the plastid-lacking protists. BMC Evol. Biol. 9: 197.
Moret BME (2001). New approaches for reconstructing phylogenies from gene order data. Bioinformatics, 17(Suppl.1): S165-S173.
Mower JP (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res. 37(Web Server issue).
Olmstead RG, Palmer JD (1994). Chloroplast DNA systematic: A review of methods and data analysis. Am. J. Bot. 81(9): 1205-1224.
Odintsova MS, Yurina NP (2006). Chloroplast Genomics of Land Plants and Algae, in Biotechnological Applications of Photosynthetic Proteins: Biochips, Biosensors and Biodevices. Springer US.
Ogihara Y, Isono K, Kojima T (2002). Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol. Genet. Genomics, 266: 740-746.
Ogihara Y (2000). Chinese Spring Wheat (*Triticum aestivum* L.) Chloroplast genome: Complete Sequence and Contig Clones. Plant Mol. Biol. Rep. 18: 243-253.
Ovcharenko I (2005). Mulan: Multiple-sequence local alignment and visualization for studying function and evolution. Genet. Res. 15(1): 184-194.
Plunkett G, Downie S (2000). Expansion and contraction of the cp inverted repeat in Apiaceae subfamily Apioideae. Syst. Bot. 25: 648-667,
Qi J (2004). CVTree: a phylogenetic tree reconstruction tool based on whole genomes. Nucleic Acids Res. vol. 32, Web Server issue.
Raven JA (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biol. 4: 209.
Ravi V (2008). An update on chloroplast genomes. Plant Syst. Evol. 271: 101-102.
Raubenson LA (2007). Comparative chloroplast genomics: analysis including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics, 8: 174.

Sugiura M (2003). History of Chloroplast genomics. Photosynthetic Res. 76: 371-377.
Shinozaki K (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5(9): 2043-2049.
Swofford DL (2003). PAUP Phylogenetic analysis using parsimony (and other methods), version 4. Sunderland Massachusetts, USA.
Shendure J, Mitra RD, Varma C, Church GM (2004). Advanced sequencing technologies: methods and goals. Nat. Rev. Genet. 5: 335-344.
Stothard P (2005). Circular genome visualization and exploration using CG view. Bioinformatics, 21: 537-539.
Tamura K (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.
Taquist H (2007). TFAM 1.0: an online tRNA function classifier. Nucleic Acids Res. 35(Web Server issue).
Tillich M (2006). The evolution of chloroplast RNA editing. Mol. Biol. Evol. 23(10): 1912-1921.
Tang J (2003). Scaling up accurate phylogenetic reconstruction from gene-order data. Bioinformatics, 1(1): 1-8.
Wyman S (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics, 20(17).
Wolf PG (2005). The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350: 117-128.
Wu FH (2009). Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol. pp. 1-10.
Xiong AS (2009). Gene duplication, transfer andevolution in the chloroplast genome. Biotechnol. Adv. 27: 340-347.
Yue F (2008a). Gene rearrangements analysis and ancestral order inference from chloroplast genome with inverted repeat. BMC Genomics, 9(Suppl.1): S25.
Yue F (2008b). Phylogenetic reconstruction from transpositions. BMC Genomics, 9 (suppl 2): S15.
Zafar N (2002). Core Genes: A computational tool for identifying and cataloging “Core” genes in a set of small genomes. BMC Bioinformatics, 3: 12.
Zhang Z (1999). Single gene circles in dinoflagellate chloroplast genomes. Nature, 400: 155-159.
Zwickl DJ (2006). GARLI (Genetic Algorithm for Rapid Likelihood Inference). Version 0.942 (www.bio.utexas.edu/faculty/antisense/garli/Garli.html).