JOURNAL OF CURRENT RESEARCH IN SCIENCE (ISSN 2322-5009)
CODEN (USA): JCRSDJ
2014, Vol. 2, No. 6, pp: 838-850
Available at www.jcrs010.com

ORIGINAL ARTICLE

CHLOROPLAST GENOME STUDY, NEW TOOL IN PLANT BIOTECHNOLOGY; *GOSSYPIUM* SPP. AS A MODEL CROP

Farshid TALAT[1,2*] , Kunbo WANG[2]
1. West Azerbaijan Agricultural and Natural Resources Research Center, AREEO, Urmia, Iran
2. Cotton Research Institute, Chinese Academy of Agricultural Sciences / Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang 455000, Henan, China
** Corresponding authors email: farshid.talat@gmail.com*

ABSTRACT: This is totally true that if there is one feature that distinguishes plant from animal life on our earth, it is not plants being primarily sessile, as a few animals also share this trait, rather, it is the reliance of plants on solar energy to generate molecules with energy-rich bonds, the fuel that will be used by almost the entire biosphere (including plants themselves) to build other organized molecules and drive the rest of the processes that we know as life. Chloroplasts are the sites of this wonderful process. Chloroplast research have significant advantage of genomics and genome sequencing, and a new picture is emerging of how the chloroplast functions and communicates with other cellular compartments. As a world's leading textile crop and a model system for studies of many biological processes, genomics research of cottons has advanced rapidly in the past few years. *Gossypium* contains 5 tetraploid (AD_1 to AD_5, 2n = 4×) and 47 diploid species (designated A through G, plus K, 2n = 2×), but the origin and evolution of allotetraploid*Gossypium* has remained controversial.

Key words: cp DNA, Cotton, Genome sequencing, *Gossypium*

INTRODUCTION

Questions concerning the evolution of organelles have been a key force driving studies of organelle molecular biology (Daniell et al., 2004b). It is widely accepted that the first plastids originated from an endosymbiotic event between a photosynthetic bacterium (cyanobacteria) and a non-photosynthetic host (Howe et al., 2003). The green lineage among the descendants of this first photosynthetic eukaryote (there was a separate red lineage), eventually colonized the planet outside the oceans, around 450 million years ago (Willis et al., 2002, Lopez-Juez and Pyke 2005). The engulfed cyanobacteria changed in to organelles as chloroplast in which small degrees of genetic autonomy as well as a large degree of biochemistry were retained, but losing some of their original functions (Davis et al., 2008, Lopez-Juez and Pyke 2005). They needed to synthesize and accumulate their required proteins within and in their surrounding cytoplasm, locate them to their correct destination, divide and propagate (Lopez-Juez and Pyke 2005). The ability of chloroplast to accomplish photosynthesis determined the development of plants throughout the land and its need to adapt to environmental signals, such as light or the availability of raw materials (Lopez-Juez and Pyke 2005). The chloroplasts were also developed into a variety of derivatives (Figure 1), including other plastid types including etioplasts, eliaplasts, amyloplasts and proplastids, to carry out essential or specialized functions other than photosynthesis in other cells, (Waters et al., 2004).

Chromoplasts are responsible for pigment synthesis and storage. Elaioplasts specialize in the lipids storage and amyloplasts store starch through the polymerization of glucose. Etioplasts are chloroplasts that have not been exposed to light and are usually found in plants grown in the dark. If a plant is kept out of light for several days, its normal chloroplasts will actually convert into etioplasts. Proplastids are the progenitor of all plastid types. Therefore the chloroplasts and its derivatives came under the control of developmental signals and affected the cells harboring them, or become influenced by the same environmental cues, to insure their function remained possible under a variety of conditions (Rodermel 2001, Lopez-Juez and Pyke 2005).

Molecular research over the past three decades have revealed many prokaryotic features in the modern-day plant organelles, including some aspects of organelle division, genome organization and coding content, transcription, translation, RNA processing, and protein turn-over (Gray 2004). The confirmation of the basic

endosymbiosis hypothesis (has raised many questions as to how evolution has shaped the modern day chloroplasts. It is still under debate whether there was a monophyletic (single) or paraphyletic (multiple) origin event for the plastid genome (Palmer 2003, Gray 2004). Complete genome sequences from the chloroplast to diverse taxa will lead to resolve this debate and generate additional information in support of the evolutionary relationships among the land plants.

CHLOROPLASTS AND OTHER PLASTID TYPES

Chloroplasts are the most noticeable feature of green cells in leaves and, excluding the vacuole, probably constitute the largest percentage of space within mesophyll cells (Lopez-Juez and Pyke 2005). Plastids conduct multiple functions and carry out a variety of critical biochemical processes other than photosynthesissuch as starch synthesis, nitrogen metabolism, sulfate reduction, fatty acid synthesis, DNA, and RNA synthesis (Zeltz et al. 1993). Plastid DNA (ptDNA) are attached to membranein clusters called plastid nucleoids and each particular type of plastid carries identical copies of the DNA, s (Kobayashi et al., 2002, Sato et al., 1993, Sato et al., 2001, Maliga 2004,Kuroiwa 1991, Maliga 2004). The number of plastids and ptDNA is highly variable depending on the cell type (Bendich 1987, Maliga 2004). In tobacco, the meristematic cells contain 10-14 proplastids, each containing 1-2 nucleoids per organelle, whereas leaf cells may contain 100 chloroplasts, with 10-14 nucleoids each, giving as much as 10,000 copies of the ptDNA per cell (Bendich 1987, Maliga 2004). The chloroplast genome organization is highly conserved (Palmer 1991, Raubeson et l., 2005) and composed of a single circular chromosome with a quadripartite structure that includes two copies of an inverted repeat (IR) that separate the large and small single copy regions (LSC and SSC) (Figure 2). The size of this circular genome varies from 35 to 217 kb but, the majority of plastid genomes from photosynthetic organisms are between 115-165 kb (Jansen et al. 2005). Compared to the nuclear and mitochondrial genomes, the plastid genome is quite conserved across taxa (Maier et al., 2004). However, due to comparisons of whole chloroplast genome sequence, differences in the general architecture (tobacco and Arabidopsis) have been reported (Hiratsuka et al., 1989, Doyle et al. 1992, Palmer and Stein 1986) and can mainly be attributed to evolutionary expansion/contraction or loss of the inverted repeat, genome rearrangements, dispersed repeats, and indels (Hiratsuka et al. 1989, Doyle et al. 1992, Palmer and Stein 1986, Maier et al., 2004). Since the inverted repeat is present in several algae, it seems likely that it is an ancient feature which has been later lost in individual branches during evolution (Palmer 1991). Characteristically, the IR-region contains a complete rRNA operon. Duplicated rRNA operons are also observed in cyanobacterial genomes which argue for a selective pressure to increase rRNA gene number (Palmer 1991). Speculatively, the IR-organization may play a direct role in maintaining the conserved structure of the chloroplast genome and also indirectly conserving genes encoded by the IR, as these genes characteristically have lower rates of nucleotide substitutions than those encoded in single copy regions (Curtis et al., 1984, Wolfe et al., 1987).

GENE TRANSFER

It has been noted that cyanobacterial genes for processes no longer needed inside the host are not found in plant cells now (e.g., motility-related genes) (Maier et al., 2004). The plastid genome is small (100-200 genes) in compare with the typical cyanobacterium genome composed of 3,000-4,000 genes (Maier et al., 2004). At first glance, it seems that many of the cyanobacterial genes have been disappeared, while, it became apparent that the plastid's proteome, despite its small genome, contained 1,000 to 5,000 proteins which is comparable in size to a cyanobacterial proteome (Martin et al., 1998, Rujan et al., 2001). Detailed analysis of homologies between modern plastid and nuclear genomes revealed substantial amounts of plastid-derived DNA in the nucleus (Maier et al 2004). This has been observed in Spinach (Cheung et al., 1989), various chenopod species (Ayliffe et al., 1988), potato (du Jardin 1990), tomato, tobacco (Ayliffe et al., 1992), rice, and *Arabidopsis* (Kebeish, 2007). These findings provided the stage to further study gene transfer to the nucleus. This information can provide invaluable phylogenetic markers such as the *rpl22* loss to the nucleus in the legumes (Gantt et al., 1991) that was discovered by chloroplast comparative genomics in analyzing whole genome sequences.

WHY DO PLASTIDS HAVE GENOMES?

The chloroplast offers a particularly unfriendly environment for DNA. The chemistry of

photosynthesis generates high concentrations of various oxygen species that are highly mutagenic (Allen et al., 1996). Whatever the selective pressures are that have reduced the plastid genome to its current size are unknown. The question still open is why this was not driven to completion. There are several hypothesis to address this question. First, it has been argued that several of the organelle encoded proteins are highly hydrophobic and hence would not easily cross the plastid envelope when translated in the cytoplasm (von Heijne 1986; Palmer 1997). A previous described argument suggests the highly hydrophobic light harvesting chlorophyll proteins are universally nuclear-encoded and the hydrophilic large subunit (*rbcL)* of RuBisCO, with few exceptions, is plastid-encoded (Maier et al., 2004).Additionally, other explanations for the maintenance of the plastid chromosome are that plastid proteins could be toxic in the cytosol (Martin et al., 1998). It has also been proposed that as gene transfer is an ongoing process, the last remnants of the plastid chromosome will eventually disappear over time (Herrmann 1997). The genes that appear to have remained are categorized as; rubisco subunit, photosystem proteins, cytochrome-related, ATP synthase, NADH dehydrogenase, ribosomal protein subunits, ribosomal RNAs, plastid encoded RNA polymerase, and open reading frames with unknown function.

PHYLOGENETIC UTILITY OF CHLOROPLAST GENOMES

Most previous molecular phylogenetic studies of flowering plants have relied on one to several genes from the chloroplast, mitochondria, and/or nuclear genomes, though most of these analyses were based on chloroplast markers (RFLP and SSR) (Jansen et al., 2006). During the past few years there has been a rapid increase in the number of studies using complete genes and intergenic regions from completely sequenced chloroplast genomes for estimating phylogenetic relationships among angiosperms (Goremykin et al., 2003a, b, 2004, 2005, Leebens-Mack et al., 2005, Chang et al., 2006, Lee et al., 2006a, Jansen et al., 2006, Ruhlman et al., 2006, Bausher et al., 2006, Cai et al., 2006). These studies have resolved a number of issues regarding relationships among the major clades, including the identification of either Amborella alone or Amborella + Nymphaeales as the sister group to all other angiosperms, these studies also lend strong support for the monophyly of magnoliids, monocots, and eudicots, the position of magnoliids as sister to a clade that includes both monocots and eudicots, the placement of vitaceae as the earliest diverging lineage of rosids, and the sister group relationship between caryophyllales and Asterids. However, some issues remain unresolved, including the monophyly of the eurosid I clade and relationships among the major clades of rosids (Jansen et al., 2006; Soltis et al., 2005). Completely sequenced chloroplast genomes provide a rich source of data that can be used to address phylogenetic questions at deep nodes in the angiosperm tree (Jansen et al., 2006; Goremykin et al., 2003a, b, 2004, 2005, Leebens-Mack et al., 2005, Chang et al. 2006, Lee et al., 2006a, Bausher et al., 2006, Cai et al., 2006). The use of DNA sequences from all of the shared chloroplast genes provides many more characters for phylogeny reconstruction compared to previous studies that have relied on only one or a few genes to address the same questions (Jansen et al., 2006). However, the whole genome approach can result in misleading estimates of relationships because of limited taxon sampling (Jansen et al., 2006, Leebens-Mack et al., 2005, Soltis et al., 2004, Stefanovic et al., 2004, Martin et al., 2005) and the use of incorrect models of sequence evolution in concatenated datasets (Jansen et al., 2006; Goremykin et al., 2005, Lockhart et al., 2005). Thus, there is a growing interest in expanding the taxon sampling of complete chloroplast genome sequences and developing new evolutionary models for phylogenetic analysis of chloroplast sequences (Jansen et al., 2006) to overcome these concerns. To date, there are more than 292 chloroplast genome sequences available.

CHLOROPLAST MOLECULAR MARKERS

Since the first report on chloroplast DNA variation based on restriction patterns (Vedel et al., 1976), there has been increasing interest in chloroplast genomic sequence for the purposes of population genetics and phylogenetic studies (McCauley 1995; Morand-Prieur 2002). The use of chloroplast DNA (cpDNA) restriction fragment length polymorphisms (RFLP) as genetic markers in interspecific hybridization showed that most angiosperm species display maternal inheritance of the chloroplast genome (Reboud et al., 1993, Morand-Prieur 2002). It has been recently noted that there is little intraspecific variation among angiosperm chloroplast DNA (Morand-Prieur

2002) and that the highest frequency of mutations is found in the noncoding regions (Palmer 1992). It has been recently discovered that chloroplast simple sequence repeats are highly useful markers for size variations that are easy to analyze by using PCR and polyacrylamide gel electrophoresis (Powell et al., 1995, Morand-Prieur 2002). The complete tobacco chloroplast genome sequence has been mined for simple sequence repeats that resulted in high levels of intra and interspecific diversity among solanaceous species (Powell et al., 1995, Provan et al., 1999, Bryan et al., 1999) the presence of which indicates the necessity for whole genome chloroplast sequence to develop polymorphic markers to reveal diversity at the intra- and interspecific level.

PLASTIDS AND BIOTECHNOLOGY

Plastid transformation involves transforming one or a few chloroplast DNA copies, followed by gradually diluting plastids carrying nontransformed copies on a selective medium (Maliga 2004). The most common integration site in chloroplast transformation is the transcriptionally active intergenic spacer region between *trnI/trnA.* This region is located in the inverted repeat near one of the two origins of replication. The plastid transformation approach has been shown to have a number of advantages, most notably with regard to its high transgene expression levels (De Cosa et al., 2001), capacity for multi-gene engineering in a single transformation event (De Cosa et al., 2001, Lossl et al., 2003, Ruiz et al., 2003, Quesada-Vargas et al., 2005), and ability to accomplish transgene containment via maternal inheritance (Daniell 2002). Moreover, chloroplasts appear to be an ideal compartment for the accumulation of certain proteins, or their biosynthetic products, which would be harmful if accumulated in the cytoplasm (Daniell et al., 2001, Lee et al., 2003, Leelavathi et al., 2003, Ruiz et al., 2005). In addition, gene silencing has not been observed in association with this technique (et al., 2001, Lee et al. 2003, Dhingra et al., 2004). Because of these advantages, the chloroplast genome has been engineered to confer several useful agronomic traits, including herbicide resistance (Daniell et al., 1998), insect resistance (McBride et al., 1995, Kota et al., 1999), disease resistance (DeGray et al., 2001), drought tolerance (Lee et al., 2003), salt tolerance (Kumar et al., 2004a), and phytoremediation (Ruiz et al., 2003). The chloroplast genome has also been utilized in the field of molecular pharming, for the expression of biomaterials, human therapeutic proteins, and vaccines for use in humans or other animals (Guda et al., 2000, Staub et al., 2000, Fernandez-San Milan et al., 2003, Leelavathi et al., 2003, Molina et al., 2004, Viitanen et al., 2004, Watson et al., 2004, Koya et al., 2005, Grevich et al., 2005, Daniell et al., 2005b, Kamarajugadda et al., 2006). Lack of complete chloroplast genome sequences is still one of the major limitations to extend this technology to useful crops. Chloroplast genome sequences are necessary for identification of spacer regions for integration of transgenes at optimal sites via homologous recombination, as well as endogenous regulatory sequences for optimal expression of transgenes (Maier et al., 2004, Daniell et al., 2005b). In land plants, about 40-50% of each chloroplast genome contains non-coding spacer and regulatory regions (Jansen et al., 2005). Identity between vector sequences and target sequence is necessary (DeCosa et al., 2001, Daniell et al., 2004b, Daniell et al., 2005b, Dhingra et al., 2004, Lee et al., 2006b), as transformation vectors with homologous sequence from another species have not yielded high frequency transformations so far even in tobacco, in which plastid transformation is highly efficient (Daniell et al., 2004b, Degray et al., 2001). Therefore, further genome sequencing projects of crop plant plastid chromosomes is one of the more pressing needs in this field to identify intergenic sequences as well as endogenous regulatory elements (Daniell et al., 2004b). Our knowledge of the organization and evolution of chloroplast genomes has been expanding rapidly because of the large numbers of completely sequenced genomes published in the past decade. The use of information from whole chloroplast genome sequence has added to our understanding of chloroplast biology, the origins and relationships of land plants, and allowed development of useful traits to aid in worldwide needs. Many crop nuclear genomes have been mapped and/or partially sequenced, but there is limited or no information about their chloroplast genomes.

FUTURE TRENDS IN CHLOROPLAST RESEARCH

ADOPTING EXPERIMENTAL TOOLS FROM OTHER FIELDS

Many research fields exist in plant and animal studiesand serve as sources of advanced methods for chloroplast research. In particular, the area of redox regulation of chloroplast worksisemerging as a hot topic in this field of research, making it necessary to establish redox markers and in vivo

sensors in chloroplast sciences (Dietz, 2008; Ute et al., 2011).

POST-TRANSLATIONAL MODIFICATIONS

While reversible phosphorylation of thylakoid proteins is a well characterized post-translational modification in chloroplasts (Pesaresi et al., 2009), during the last decade, other post-translationalproteins has been described. Protein S-nitrosylation has emerged as the most important mechanism for transduction of the bioactivity of nitric oxide.Also several chloroplast proteins have been described to become S-nitrosylated. Glutathionylation is a more recently described redox post translational modification representing the major form of S-thiolation in cells in which a mixed disulfide between a free thiol on a protein and a molecule of glutathione is formed (Rouhier et al., 2008).

PROSPECTS OF TRANSCRIPTOME ANALYSES

Using of next generation sequencing (NGS) approaches may provide a cheaper alternative to hybridization-based microarray platformstowards transcript quantification (Leister et al., 2008). Biological inducible systems enable reconstruction of effects on nuclear and plastid gene expressions with high temporal resolution (Pesaresi et al., 2007) andpromises to extend the power of transcriptomics.

TOWARDS THE CHLOROPLAST INTERACTOME

One of the most important approaches to elucidate protein interactions is the fractionation of native protein complexes using one or several molecular biology methods such as one- or multidimensional electrophoreses, chromatography, or density-gradient centrifugations. In further, matrix assisted laser desorption ionization time-of-flight (MALDI–TOF) tandem mass spectrometry (MS–MS) method can be used to detect the identity of isolated complexes(Liu et al., 2008). Although this approach is straightforward, it has been utilized in only a relatively small number of studies of plant protein complexes so far and should be more widely explored.

SYSTEMS BIOLOGY

Combining the analysis of multiple datasets generated by transcriptomics and other 'omics' will be essential for the in silico reconstruction of regulatory networks (Nacu et al., 2007). Eventually, a chloroplast development or functional state-driven nuclear gene expression network would be constructed based on genome-wide identification of transcription factors via yeast one-hybrid assays. Additionally, their direct and indirect target genes will be recognized using chromatin immuno-precipitation with an antibody against such transcription factors followed by a NGS approach and identification of genes whose expression levels in the loss-of-function mutant background are affected (Jung, 2010).

TAILORED CHLOROPLASTS

Transformation capacity of chloroplasts by homologous transformation and drive high-level expression of transgenes in chloroplasts, coupled with their maternal mode of inheritance in most species of interest, make chloroplasts a prime target for biotechnological improvement of crop plants. However, commercial varieties harboring transgenic chloroplasts have not been generated yet (Bock, 2010; Ute et al., 2011). One promising target for modifying chloroplast functions is photorespiration.

RECENT ADVANCES IN COTTON GENOMICS

Cottons (*Gossypiumspp*.) belong to the genus *Gossypium* of the family *Malvaceae*. *Gossypium* consists of 45–50 species (appendix 1), with 40–45 being diploids (2n = 26) and 5 being allotetraploids (2n = 52). The species are grouped into eight genome groups, designated A through G and K, on the basis of chromosome pairing affinities (Zhang et al., 2007). At the tetraploid level, there are five species, designated $(AD)_1$ through $(AD)_5$ for their genome constitutions. Phylogenetic analyses clustered the diploid species of *Gossypium* into two major lineages, including the 13 D-genome species lineage and the 30~32 A-, B-, E-, F-, C-, G-, and K-genome species lineage, and the polyploid species into one lineage, that is, the 5 AD-genome species lineage (Figure 3). Of the Gossypium species, four are cultivated in agriculture, including two allotetraploids (*G. hirsutum* and *G. barbadense*) and two diploids (*G. herbaceum* and *G. arboreum*). *Gossypiumhirsutum*, also known as Upland cotton, Long Staple Cotton, or Mexican Cotton, produces over 90% of the world's cotton; *G. barbadense*, also known as Sea Island Cotton, Extra Long Staple Cotton, American Pima, or Egyptian Cotton, contributes 8% of the world's cotton; and *G. herbaceum*, also known as

Levant Cotton, and *G. arboreum*, also known as Tree Cotton, together provide 2% of the world's cotton. Cottons are not only a world's leading textile fiber and oilseed crop, but also a crop that is of significance for fuel energy and bio engergy production. Although cottons are native to tropics and subtropics naturally, including the Americas, Africa and Asia, they are cultivated in nearly 100 countries. China, India, USA and Pakistan are the top four cotton growing countries, accounting for approximately 2/3of the world's cotton. According to the Food and Agriculture Organization (FAO) of the United Nations (http://www.fao.org), the cotton planting area reached about 34.1 million hectares and the total world's cotton production had a record of about 26.3 million metric tons in 2012/2013. Cotton products include fibers and seeds that have a variety of uses. Cotton fibers sustain one of the world's largest industries, the textile industry, for wearing apparel, home furnishings, and medical supplies, whereas cottonseeds are widely used for food oil, animal feeds, and industrial materials (such as soap). Cottonseed oil is ranked fifth in production and consumption volume among all vegetable oils in the past decades, accounting for 8% of the world's vegetable oil consumption (Zhang et al. 2007 and Lee SB et al. 2006a).

Moreover, nearly a billion barrels of petroleum worldwide are used in every year to synthesize artificial "synthetic" fibers. Further improvement of cotton fibers in yield and quality will replace or significantly reduce the consumption of fossil oil for synthetic fiber production, thus being saved for energy production. Finally, cottonseed oil, the main by-product of cotton fiber production, could be potentially used as biofuel. In addition to their economic importance, cottons are an excellent model system for several important biological studies, including plant genome size evolution, plant polyploidization and single-celled biological processes. The genomes of angiosperm plants vary over 1000 folds in size, ranging from 100 to >100,000 Mb/1C (haploid). It has long been recognized that polyploidy is a common, prominent, ongoing, and dynamic process of genome organization, function diversification, and evolution in angiosperms. The genomes of most angiosperms are thought to have incurred one or more polyploidization events during evolution. Studies have demonstrated that genome doubling has also been significant in the evolutionary history of all vertebrates and in many other eukaryotes. It is estimated that about 70% of the flowering plant species are polyploids. For instance, of the world-leading field, forage, horticultural, and environmental crops, many are contributed by polyploid species, such as cotton, wheat, soybean, potatoes, canola, sugarcane, Brassica, oats, peanut, tobacco, rose, coffee, and banana. Therefore, studies of both genome size evolution and polyploidization have long attracted the interests of scientists in different disciplines. Nevertheless, much remains to be learned. Examples include impacts of polyploidization on genome size, genome organization, gene duplication and function, and gene family evolution; the role of transposable elements in structural and regulatory gene evolution and gene functions; and mechanisms and functional significance of rapid genome changes.

Cottons have several advantages over other polyploid complexes for plant genome size and polyploidization studies. First, the genome sizes of 37 of the 45~50 *Gossypium* species, including all eight genomes and polyploidy species, have been determined and shown to vary extremely significantly (Figure 3). At the diploid level, the genome sizes vary by three folds, ranging from 885 Mb/1C in the D-genome species to 2,572 Mb/1C in the K-genome species. Within each lineage, the genome sizes vary most in the A+F+B+E+C+G+K lineage, ranging from 1,311 to 2,778 Mb/1C with a difference of 1,467Mb (110.2%); second in the D-genome lineage, ranging from 841 to 934 Mb/1C with a difference of 93Mb (10.5%); and least in the polyploidy lineage, ranging from 2,347 to 2,489 Mb/1C with a difference of 142Mb (5.9%). Variations were also observed within a species. For instance, within *G. hirsutum*, the variation (n = 5) was from 2,347 to 2,489 Mb/1C, differing by 142Mb (5.9%) while within *G. arboreum*, the variation (n = 5) was from 1,677 to 1,746 Mb/1C, differing by 69Mb (4.0%). Second, the evolutionary history of the allotetraploid species of *Gossypium* has been established (Figure 3), especially for the two cultivated AD-genome cottons, *G. hirsutum* and *G. barbadense*, and their closely related diploid progenitors, *G. herbaceum* (A_1), *G. arboreum* (A_2), *G. raimondii* (D_5), and G. gossypioides (D_6).The A-genome species are African-Asian in origin, whereas the D-genome species are endemic to the New World subtropics, primarily Mexico. Following the transoceanic dispersal of an A-genome taxon to the New World, hybridization between the immigrant A-genome taxon and a local D-genome taxon led to the origin and evolution of the New World allopolyploids (AD-genome). Subsequent to the polyploidization

event, the allopolyploids radiated into three sublineages, among which included are the world's commercially most important species, *G. hirsutum* and *G. barbadense*. Studies showed that the A subgenome of the AD-genome-cultivated cottons is the most closely related to the genome of the extant diploid *G. herbaceum* (A_1); the D subgenome of the AD-genome-cultivated cottons is the most closely related to the genome of the extant diploid, *G. raimondii*(D_5) or *G. gossypioides* (D_6) ; and the cytoplasm of the AD genome-cultivated cottons is the most closely related to that of the extant diploids *G. herbaceum* (A_1) and *G. arboreum* (A_2). Sequence analysis and paleontological record suggest that the A-genome and the D-genome groups diverged from a common ancestor 5–10 million years ago, and that the two diverged diploid genomes became reunited in a common nucleus to form the polyploid cottons, via allopolyploidization, in the mid-Pleistocene, or 1-2 million years ago. Finally, as in the wheat polyploid complex, cottons have a long history of research at the cytological level. A wealth of cytogenetic stocks has been developed, including artificially synthesized AD-genome polyploids between the A-genome and D-genome diploid species as well as individual chromosome addition and substitution lines. These cytogenetic stocks are unique and valuable not only for cotton genetics research, but also for deciphering the ramifications of polyploidization on genome organization, function, and evolution. Cotton fiber is an excellent single-celled model system for studies of many single-celled biological processes, particularly cell expansion and cellulose biosynthesis. Cotton fibers are unicellular, unbranched, simple trichomes that differentiate from the protoderm of developing seeds. There are probably over one-half million quasi-synchronously elongating fibers in each boll or ovary. Although all plant cells extend to some degree during development and differentiation, cotton fibers can reach up to 5.0 cm in length in some genotypes, being among the longest cells. Therefore, they offer a unique opportunity to study cell expansion at the single cell level. Cellulose is a major component of the cell walls of all higher plants, constituting perhaps the largest component of plant biomass, with an estimated annual world production of 100 million metric tons. The fiber cell wall of cottons consists of >90% cellulose. Therefore, cotton fiber cells have long been used as a model system to study cellulose biosynthesis that is the basis for biomass-based bioenergy production (Zhang et al., 2007). Chloroplast research takes significant advantage of genomics and genome sequencing, and a new picture is emerging of how the chloroplast functions and communicates with other cellular compartments. In terms of evolution, it is now known that only a fraction of the many proteins of cyanobacterial origin were rerouted to higher plant plastids. Reverse genetics and novel mutant screens are providing a growing catalogue of chloroplast protein–function relationships, and the characterization of plastid-to-nucleus signaling mutants reveals cell–organelle interactions. Recent advances in transcriptomics and proteomics of the chloroplast make this organelle one of the best understood of all plant cell compartments. The need for sequencing the cotton plastome is obvious, when considering its annual retail value in the world cotton producing countries making it those country's most value-added crop. Chloroplast genetic engineering could minimize transgene escape because of maternal inheritance of transgenes. In addition, other failsafe mechanisms, including cytoplasmic male sterility could be employed to contain transgenes.

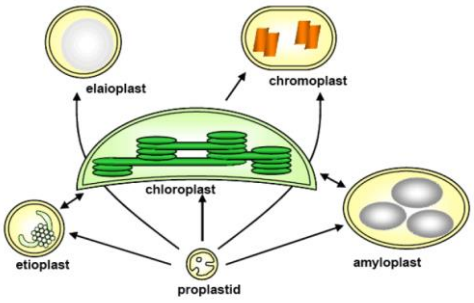

Figure. 1 Diversity of plastid types and their interconversions. Chloroplasts occupy the center of the figure to signify their evolutionary role as ancestors of all other plastid types (taken from Lopes-Juez and Pyke 2005)

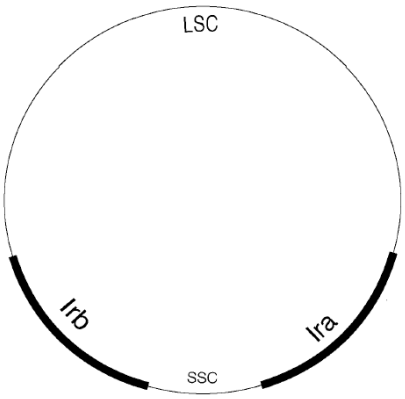

Figure. 2 Typical organization of a plastid chromosome in its circular monomeric form. Large and small single copy regions (LSC, SSC) are separated by the inverted repeats Ira and IRb (Jansen et al., 2005).

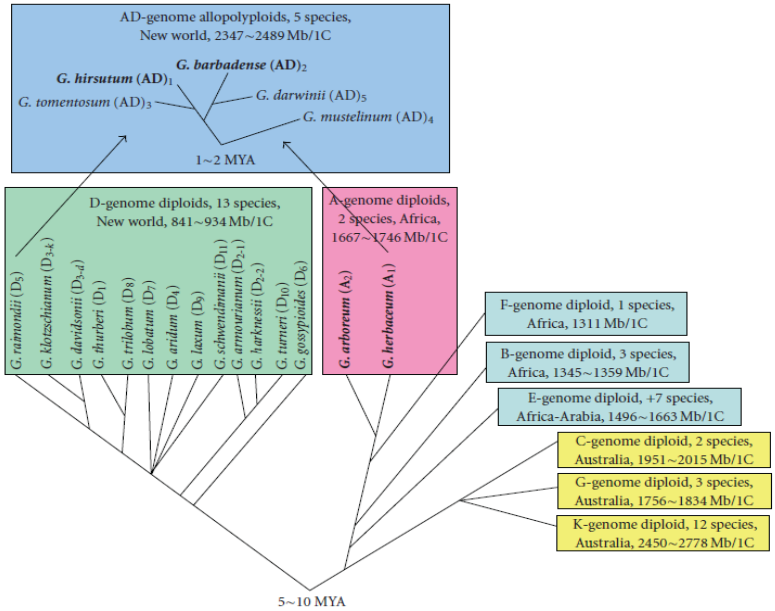

Figure. 3: Phylogeny and evolution of *Gossypium* species. The phylogenetic data is from Wendel and Cronn, the genome sizes are from Hendrix and Stewart, and genomic designations follow Endrizzi et al. and Percival. The species in bold face are cultivated. MYA: million years ago (Zhang et al., 2007).

REFERENCES

Allen JF, Raven JA .Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. J Mol Evol 1996: 42:482-492.

Ayliffe MA, Timmis JN, Steele SN. Homologies to chloroplast DNA in the nuclear DNA of a number of Chenopod species. Theor Appl Genet 1988: 75:282-285.

Ayliffe MA, Timmis JN. Plastid DNA sequence homologies in the tobacco nuclear genome. Mol. Gen. Genet 1992: 236:105-112.

Bausher MG, Singh ND, Lee S-B, Jansen RK , Daniell H. The complete chloroplast genome sequence of *Citrus sinensis*(L.) Osbeckvar 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. BMC Plant Biology 2006: 6:21-29.

Bendich AJ. Why do chloroplasts and mitochondria contain so many copies of their genome? Bioessays 1987: 6:279-282.

Bock R, Warzecha H. Solar-powered factories for new vaccines and antibiotics. Trends Biotechnol 2010: 28:246–252.

Bryan GJ, McNicoll J, Ramsey G, Meyer RC, De Jong WS. Polymorphic simple sequences repeat markers in chloroplast genomes of Solanaceous plants. Theor Appl Genet 1999:99:859-867.

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson J, dePamphilis CW , Jansen RK. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: Implications for the phylogeny of magnoliids. BMC Evol Biol 2006: 6:77.

Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC , Chaw SM. The chloroplast genome of *Phalaenopsis Aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol 2006:23:279-291.

Cheung WY, Steele Scott N. A contigous sequence in spinach nuclear DNA is homologous to three separated sequences in chloroplast DNA. Theor Appl Genet 1989: 77:625-633.

Cosner ME, Jansen RK, Palmer JD, Downie SR. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): Multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr Genet 1997:31: 419–429

Curtis SE, Clegg MT .Molecular evolution of chloroplast DNA sequences. Mol Biol Evol 1984: 1:291-301.

Daniell H, Cohill P, Kumar S, Dufourmantel N , Dubald M . Chloroplast genetic engineering; In Daniell H, Chase C (eds) molecular biology and biotechnology of plant organelles. SpringerPublishers, Netherlands 2004b: pp 423-468.

Daniell H, Datta R, Varma S, Gray S, Lee SB. Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnoly 1998: 16:345-348.

Daniell H, Dhingra A .Multigene engineering: dawn of an exciting new era in biotechnology. Curr Opin Biotechnol 2002: 13:136-141.

Daniell H, Khan M, Allison L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Sci 2002: 7:84-91.

Daniell H, Kumar S, Duformantel N. Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol 2005b: 23(5):238–245.

Daniell H, Lee SB, Pahchal T, Wiebe P. Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. J Mol Biol 2001: 311:1001-1009.

Daniell H. Molecular strategies for gene containment in transgenic crops. Nat Biotechnol 2002: 20: 581–586.

Davis N, Biddlecom N. On the relationship between GC content and the number of predicted microRNA binding sites by MicroInspector. Comput Biol Chem 2008: 32(3): 222-6.

DeCosa B, Moar W, Lee SB, Miller M, Daniell H. Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. Nat Biotechnol 2001: 9:71–74.

DeGray G, Rajasekaran K, Smith F, Saford J, Daniell H. Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi. Plant Physiol 2001: 127:852-862.

Dhingra A, Portis A Jr, Daniell H. Enhanced translation of a chloroplast-expressed *rbcS*gene restores small subunit levels and photosynthesis in nuclear *rbcS*antisense plants. Proc Natl Acad Sci USA 2004: 101:6315-6320.

Dietz KJ. Redox signal integration: from stimulus to networks and genes. Physiol Plant 2008: 133: 459–468.

Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc Natl Acad Sci, USA 1992: 89:7723-7726.

Du Jardin P. Homologies to plastid DNA in the nuclear and mitochondrial genomes of potato. Theor Appl Genet 1990: 79:807-812.

Endrizzi JE, Turcotte EL, Kohel RJ. Genetics, cytology, and evolution of *Gossypium*. Advances in Genetics 1985: 23: 271–375.

Fernandez-San M A, Mingeo-Castel AM, Miller M, Daniell H . A chloroplast transgenic approach to hyper-express and purify human serum albumin, a protein highly susceptible to proteolytic degradation. Plant Biotechnol J 2003: 1:71-79.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and

involved the gain of an intron. EMBO J 1991: 18:2621-2630.
Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH. Analysis of the *Amborellatrichopoda*chloroplast genome sequence suggests that *Amborella*is not a basal angiosperm. Mol Biol Evol 2003a: 20:1499-1505.
Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH. The chloroplast genome of the "basal" angiosperm *Calycanthusfertilis*- structural and phylogenetic analyses. Plant Syst Evol 2003b: 242:119-135.
Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH. The chloroplast genome of *Nymphaeaalba*: whole-genome analyses and the problem of identifying the most basal angiosperm. Mol Biol Evol 2004: 21:1445-1454.
Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH. Analysis of *Acoruscalamus*chloroplast genome and its phylogenetic implications. Mol Biol Evol 2005: 22:1813-1822.
Gray Michael W. Plastid genomes. In: Molecular Biology and Biotechnology of Plant Organelles. 2004: pp 115-150, Daniell, H. and Chase, C.D., Eds., Springer publishers, Netherlands.
Grevich JJ, Daniell H. Chloroplast genetic engineering: Recent advances and future perspectives. Crit Rev Plant Sci 2005: 24:83-108.
Guda C, Lee SB, Daniell H. Stable expression of biodegradable protein based polymer in tobacco chloroplasts. Plant Cell Rep 2000: 19:257-262.
Herrmann RG .Eukaryotism, towards a new interpretation. In H. E. Schenk, K. W. Jeon, N. E. Muller, W. Schwemmler, K. W. Jeon (Eds), Eukaryotism and Symbiosis 1997:pp 73-118, Springer: Berlin.
Hiratsuka J, Shimada H, Whittier R. The complete sequence of the *rice* (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 1989: 217:185–194.
Howe CJ, Barbrook AC. Evolution of the chloroplast genome. Philos Trans R SocLondBBiol Sci 2003:358(1429): 99-106.
Jansen RK, Kaittanis C, Saski C, Lee SB, Tompkins J, Alverson AJ, Daniell H. Phylogenetic analyses of *Vitis*(Vitaceae) based on complete chloroplast genome sequences: Effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol 2006: 6:32-42.
Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J , Cui L . Methods for obtaining and analyzing chloroplast genome sequences. Meth Enzy mol 2005: 395:348-384.
Jung HS, Chory J. Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? Plant Physiol 2010: 152:453–459.
Kamarajugadda S, Daniell H .Choroplast derived anthrax and other vaccine antigens: their immunogenic and immunoprotective properties. Expert Rev Vaccines 2006: 5:839-849.
Kebeish R.Chloroplasticphotorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. Nat. Biotechnol 2007:25:593–599.
Kobayashi T, Takahara M, Miyagishima S, Kuroiwa H, Sasaki N. Detection and localization of chloroplast encoded HU-like protein that organizes chloroplast nucleoids. Plant Cell 2002: 14:1579-1589.
Kota M, Daniel H, Varma S, Garczynski SF, Gould F, William MJ. Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. ProcNatlAcad, Sci USA 1999: 96: 1840–1845.
Koya V, Moayeri M, Leppla SH, Daniell H. Plant based vaccine: mice immunized with chloroplast-derived anthrax protective antigen survive anthrax lethal toxin challenge. Infect Immun 2005: 73:8266-8274.
Kumar S, Dhingra A, Daniell H. Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots and leaves confers enhanced salt tolerance. Plant Physiol 2004a: 136:2843-2854.
Kuroiwa T. The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei. Int Rev Cytol 1991: 128:1-62.
Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD ,Daniell H. The complete chloroplast genome sequence of *Gossypiumhirsutum*: organization and

phylogenetic relationships to other angiosperms. BMC Genomics 2006a: 7:61-72.

Lee SB, Kwon H, Kwon S. Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. Mol Breed 2003: 11:1-13.

Lee SM, Kang K, Chung H, Yoo SH, Xu XM, B. Lee SB, Cheong JJ, Daniell , Kim M. Plastid transformation in the monocotyledonous cereal crop, rice (*Oryza sativa*) and transmission of transgenes to their progeny. Mol Cells 2006b: 21:401-410.

Leebens-Mack J, Raubeson LA, Cui L, Kuehl J, Fourcade M, Chumley T, Boore JL, Jansen RK , dePamphilis CW. Identifying the basal angiosperms in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol 2005: 22:1948-1963.

Leelavathi S and Reddy V. Chloroplast expression of His-tagged GUS fusions: a general strategy to overproduce and purify foreign proteins using transplastomic plants as bioreactors. Mol Breed 2003:11:49-58.

Leelavathi S, Gupta N, Maiti S, Ghosh A, Reddy VS. Overproduction of an alkali- and thermo-stable xylanase in tobacco chloroplasts and efficient recovery of the enzyme. Mol Breed 2003: 11:59-67.

Leister D and Kleine T. Towards a comprehensive catalog of chloroplast proteins and their interactions. Cell Res 2008: 18:1081–1083.

Liu X, Yang WC, Gao Q, Regnier F. Toward chromatographic analysis of interacting protein networks. J. Chromatogr. A 2008: 1178:24–32.

Lockhart PJ, Penny D. The place of *Amborella* within the radiation of angiosperms. Trends Plant Sci 2005: 10:201–202.

Lopez-Juez E, Pyke KA. Plastids unleashed: their development and their integration in plant development. Int J Dev Biol 2005: 49:557-577.

Lossl A, Eibl C, Harloff HJ, Jung C, Koop HU. Polyester synthesis in transplastomic tobacco (*Nicotianatabacum*L.): significant contents of polyhydroxybutyrate are associated with growth reduction. Plant Cell Rep 2003: 21:891-899.

Maier RM, Schmitz-Linneweber. Plastid genomes. In: Molecular biology and biotechnology of plant organelles 2004:pp 115-150, Daniell H, Chase CD, Eds, Springer publishers, Netherlands

Maliga Pal. Plastid transformation in higher plants. Annu Rev Plant Biol 2004: 55: 289-313.

Martin W, Deusch O, Stawski N, Grunheit N, Goremykin V. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. Trends Plant Sci 2005: 10:203–209.

Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol 1998: 118:9-17.

McBride K, Svab Z, Schaaf D, Hogan P, Stalker D, Maliga P. Amplification of a chimeric Bacillus gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. Biotech 1995: 13:362-365.

McCauley DE. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution 1995: 10:198-202.

Molina A, Herva-Stubbs S, Daniell H, Mingo-Castel AM, Veramendi J. High yield expression of a viral peptide animal vaccine in transgenic tobacco chloroplasts. Plt Biotechnol J 2004: 2:141–153.

Morand-Prieur ME, Vedel F, Raquin C, Brached S, Sihachakr D, Frascaria-Lacoste N. Maternal inheritance of a chloroplast microsatllite marker in controlled hybrids between *Fraxinus excelsior* and *Fraxinusangustifolia*. Mol Ecol 2002: 11:613-617.

Nacu S, Critchley-Thorne R, Lee P, Holmes S. Gene expression network analysis and applications to immunology. Bioinformatics 2007: 23:850–858.

Palmer JD, Stein DB. Conservation of chloroplast genome structure among vascular plants. Curr Genet 1986: 10:823-833.

Palmer JD. Organelle genomes: going, going, gone! Science 1997: 275:790-791.

Palmer JD. Plastid chromosomes: structure and evolution. In RG Hermann, Ed, The molecular biology of plastids. Cell culture and somatic cell genetics of plants, vol. 7A, Springer-Verlag, Vienna 1991: pp 5–53.

Palmer JD. The symbiotic birth and spread of plastids: how many times and who done it? J Phycol 2003: 39:4-11.

Pesaresi P, Schneider A, Kleine T, Leister D. Interorganellar communication. Curr Opin Plant Biol 2007: 10:600–606.

Pesaresi P. Arabidopsis *STN7* kinase provides a link between short- and long-term

photosynthetic acclimation. Plant Cell 2009: 21:2402–2423.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. Polymorphic simple sequence repeat regions in chloroplast genomes: application to the population genetics of pines. PNAS 1995: 92:7759-7763.

Provan J, Soranzo N, Wilson N, Goldstein D, Powell W. A low mutation rate for chloroplast microsatellites. Genetics 1999: 153:943-947.

Quesada-Vargas T, Ruiz ON, Daniell H. Characterization of heterologous multigene operons in transgenic chloroplasts: transcription, processing, translation. Plant Physiol 2005: 128:1746-1762.

Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry R (Ed) Diversity and evolution of plants-genotypic and phenotypic variation in higher plants. CABI Publishing, Wallingford 2005: pp 45–68.

Reboud X, Zeyll C. Organelle inheritance in plants. Heredity 1993: 72:132-140.

Rodermel S. Pathways of plastid-to-nucleus signaling. Trends Plant Sci 2001: 6:471-478.

Rouhier N, Lemaire SD, Jacquot JP. The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. Annu Rev Plant Biol 2008: 59:143–166.

Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD , Daniell D. Complete plastid genome sequence of *Daucuscarota*: Implications for biotechnology and phylogeny of angiosperms. BMC Genomics 2006: 7:224.

Ruiz O, Hussein S, Terry N, Daniell H. Phytoremediation of organomercurial compounds via chloroplast genetic engineering. Plant Physiol 2003: 132:1344-1352.

Ruiz ON, Daniell H. Engineering cytoplasmic male sterility via the chloroplast genome by expression of ketothiolase. Plant Physiol 2005: 138:1232-1246.

Ruiz ON, Hussein H, Terry N, Daniell H. Phytoremediation of organomercurial compounds via chloroplast genetic engineering. Plt Phys 2003: 32:1344–1352.

Rujan T, Martin W. How many genes in *Arabidopsis* come from cyanobacteria? An estimate from 386 protein phylogenies. Trends Genet 2001: 17:113-120.

Sato N, Albrieux C, Joyard J, Douce R , Kuroiwa T. Detection and characterization of a plastid DNA-binding protein which may anchor plastid nucleoids. EMBO J 1993: 12:555-561.

Sato N, Ohta N. DNA binding specificity and dimerization of the DNA binding domain of the PEND protein in the chloroplast envelope membrane. Nucleic Acids Res 2001: 29:2244-2250.

Soltis DE, Soltis PS, Endress PK, Chase MW. Phylogeny and evolution of Angiosperms. Sunderland Massachusetts: Sinauer Associates Inc 2005: pp: 150-172.

Staub JM, Garcia B, Graves J. High yield production of a human therapeutic protein in tobacco chloroplasts. Nat Biotechnol 2000: 18:333-338.

Stefanovic S, Rice DW, Palmer JD. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella*or monocots? BMC Evol Biol 2004: 4:35-48.

Ute A, Pesaresib P, Pribila M, Hertlea A , Leister D. Update on chloroplast Research: new tools, new topics, and new trends. Molecular Plant 2011: 4:1–16.

Vedel F, Quetier F, Bayen M. Specific cleavage of chloroplast DN and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. Theor. Appl Genet 1976: 263:440-442.

Viitanen PV, Devine AL, Kahn S, Deuel DL, Van-Dyk DE, Daniell H. Metabolic engineering of the chloroplast genome using the *E.coliubiC*gene reveals that corismate is a readily abundant precursor for 4-hydroxybenzoic acid synthesis in plants. Plt Phys 2004: 136:4048–4060.

Von Heijne G. Why mitochondria need a genome. FEBS Lett 1986: 198:1-4.

Waters MT, Fray RG, Pyke KA .Stromule formation is dependent upon plastid size, plastid differentiation status and the density of plastids within the cell. Plant J 2004: 39:655-667.

Watson J, Koya V, Leppla SH, Daniell H. Expression of *Bacillus anthracis*protective antigen in transgenic chloroplasts of tobacco, a non-food/feed crop. Vaccine 2004: 22: 4374–4384.

Willis KJ, Macelwain JC. The Evolution of Plants Oxford Univ Press 2002: Oxford.

Wolfe KH, Li WH, Sharp PM. Rates of nucleotide subsistution vary greatly among plant mitochondrial, chloroplast, and nuclear

DNAs. Proc Natl Acad Sci USA 1987: 84:9054-9058.
Zeltz P, Hess WR, Neckermann K, Borner T, Kossel H. Editing of the chloroplast *rpoB*transcript is independent of chloroplast translation and shows different patterns in barley and maize. EMBO J 1993: 12:4291-4296.
Zhang HB, Li Y, Wang B, Chee PW. Recent advances in cotton genomics. Inter J of Pl Genomics 2007: 2008:1-20.