BMC Genomics

BioMed Central

Research article

Open Access

The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms

Seung-Bum Lee[1], Charalambos Kaittanis[1], Robert K Jansen[2], Jessica B Hostetler[3], Luke J Tallon[3], Christopher D Town[3] and Henry Daniell*[1]

Address: [1]Dept. of Molecular Biology & Microbiology, University of Central Florida, Biomolecular Science, Building #20, Orlando, FL 32816–2364, USA, [2]Section of Integrative Biology and Institute of Cellular and Molecular Biology, Patterson Laboratories 141, University of Texas, Austin, TX 78712, USA and [3]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

Email: Seung-Bum Lee - sbumlee@mail.ucf.edu; Charalambos Kaittanis - daniell@mail.ucf.edu; Robert K Jansen - jansen@mail.utexas.edu; Jessica B Hostetler - cdtown@tigr.ORG; Luke J Tallon - cdtown@tigr.ORG; Christopher D Town - cdtown@tigr.ORG; Henry Daniell* - daniell@mail.ucf.edu

* Corresponding author

Published: 23 March 2006

BMC Genomics 2006, **7**:61 doi:10.1186/1471-2164-7-61

Received: 21 November 2005
Accepted: 23 March 2006

This article is available from: http://www.biomedcentral.com/1471-2164/7/61

© 2006 Lee et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Cotton (*Gossypium hirsutum*) is the most important fiber crop grown in 90 countries. In 2004–2005, US farmers planted 79% of the 5.7-million hectares of nuclear transgenic cotton. Unfortunately, genetically modified cotton has the potential to hybridize with other cultivated and wild relatives, resulting in geographical restrictions to cultivation. However, chloroplast genetic engineering offers the possibility of containment because of maternal inheritance of transgenes. The complete chloroplast genome of cotton provides essential information required for genetic engineering. In addition, the sequence data were used to assess phylogenetic relationships among the major clades of rosids using cotton and 25 other completely sequenced angiosperm chloroplast genomes.

Results: The complete cotton chloroplast genome is 160,301 bp in length, with 112 unique genes and 19 duplicated genes within the IR, containing a total of 131 genes. There are four ribosomal RNAs, 30 distinct tRNA genes and 17 intron-containing genes. The gene order in cotton is identical to that of tobacco but lacks *rpl22* and *infA*. There are 30 direct and 24 inverted repeats 30 bp or longer with a sequence identity ≥ 90%. Most of the direct repeats are within intergenic spacer regions, introns and a 72 bp-long direct repeat is within the *psaA* and *psaB* genes. Comparison of protein coding sequences with expressed sequence tags (ESTs) revealed nucleotide substitutions resulting in amino acid changes in *ndhC, rpl23, rpl20, rps3* and *clpP*. Phylogenetic analysis of a data set including 61 protein-coding genes using both maximum likelihood and maximum parsimony were performed for 28 taxa, including cotton and five other angiosperm chloroplast genomes that were not included in any previous phylogenies.

Conclusion: Cotton chloroplast genome lacks *rpl22* and *infA* and contains a number of dispersed direct and inverted repeats. RNA editing resulted in amino acid changes with significant impact on their hydropathy. Phylogenetic analysis provides strong support for the position of cotton in the Malvales in the eurosids II clade sister to *Arabidopsis* in the Brassicales. Furthermore, there is strong support for the placement of the Myrtales sister to the eurosid I clade, although expanded taxon sampling is needed to further test this relationship.

Background

The chloroplast is the site of photosynthesis, where light energy in photons is converted into chemical bond energy, via redox reactions, including inorganic carbon fixation at Calvin's cycle, finally yielding energy-rich carbohydrate molecules. Therefore, apart from the antennae, photosystem I and II complexes, which are found in the thylakoid membrane, the chloroplast contains the entire enzymatic machinery for carbohydrate biosynthesis in the stroma. Anabolic pathways such as protein, fatty acid, vitamin, and pigment biosynthesis take place in the chloroplast as well, indicating the organelle's ability to synthesize complex molecules. The chloroplast genome maintains a highly conserved organization [1,2] with most land plant genomes 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) [3]. The recent surge of interest in sequencing chloroplast genomes has provided a plethora of information on the organization and evolution of these genomes and new data for reconstructing phylogenetic relationships [2].

Chloroplast genetic engineering offers numerous advantages, including a high-level of transgene expression [4], multi-gene engineering in a single transformation event [4-7], transgene containment via maternal inheritance [8-11] or cytoplasmic male sterility [12], lack of gene silencing [4,13], position effect due to site specific transgene integration [14], and pleiotropic effects due to sub-cellular compartmentalization of transgene products [13,15,16]. Apart from expressing therapeutic agents, biopolymers, or transgenes to confer agronomic traits, plastid genetic engineering has been used to study plastid biogenesis and function, revealing mechanisms of plastid DNA replication origins, intron maturases, translation elements and proteolysis, import of proteins and several other processes [18]. Despite the potential of chloroplast genetic engineering, this technology has only recently been extended to the major crops, including soybean [19], carrot [20] and cotton [21], via somatic embryogenesis, achieving transgene expression in non-green plastids [22]. All other previous studies focused on direct organogenesis by bombardment of leaves containing mature green chloroplasts [22]. Lack of complete chloroplast genome sequences to provide 100% homologous species-specific chloroplast transformation vectors, containing suitable selectable markers and endogenous regulatory elements, is one of the major limitations to extend this concept to other useful crops [22,23].

The need for sequencing the cotton plastome is obvious, when considering its annual retail value of about $120 billion, making it America's most value-added crop. This is justified by the fact that cotton is the single most important textile fiber grown in 90 countries; the US accounts for 21% of the total world fiber production. In 2004–2005, US farmers planted 79% of the 5.7-million hectares of nuclear transgenic cotton. Upland cotton, *Gossypium hirsutum*, has the potential to hybridize with *G. tomentosum*, feral populations of *G. hirsutum*, and *G. hirsutum*/*G. barbadense* [21]. Therefore, geographical restrictions in planting genetically modified cotton are in place because of reports of pollen dispersal from transgenic cotton plants [25]. Chloroplast genetic engineering could minimize transgene escape because of maternal inheritance of transgenes [8-11]. In addition, other failsafe mechanisms, including cytoplasmic male sterility could be employed to contain transgenes [12].

The examination of phylogenetic relationships among angiosperms has received considerable attention during the past decade [reviewed in [26]]. Although there is considerable consensus about the circumscription and relationships among many of the major clades, most molecular phylogenetic analyses have examined numerous taxa but have relied on only a few gene sequences. Completely sequenced chloroplast genomes provide a rich source of nucleotide sequence data that can be used to address phylogenetic questions. Several recent studies have attempted to use completely sequenced genomes to resolve the identification of the basal lineages of flowering plants [27-29]. Use of many or all of the genes from the chloroplast genome provides many more characters for phylogeny reconstruction in comparison with previous studies that have relied on only a few genes. However, the limited number of available whole chloroplast genome sequences can result in misleading estimates of relationship [27,30]. This problem can be overcome as more complete chloroplast genome sequences become available.

In this article, we present the complete sequence of the chloroplast genome of upland cotton, *Gossypium hirsutum*. One goal of this paper is to examine gene content and gene order, and determine the distribution and location of repeated sequences. Secondly, the RNA editing sites in the cotton chloroplast genome are identified and examined, by comparing the DNA sequences with available expressed sequence tag (EST) sequences, because RNA editing plays a major role in several lineages of plants [31,32]. Lastly, protein-coding sequences from 61 genes are used to estimate phylogenetic relationships of cotton with 25 other angiosperms.

Results

Size, gene content, order and organization of the cotton chloroplast genome

Cotton complete chloroplast genome is 160,301 bp in length (Fig. 1), and includes a pair of inverted repeats 25,608 bp long, separated by a small and a large single

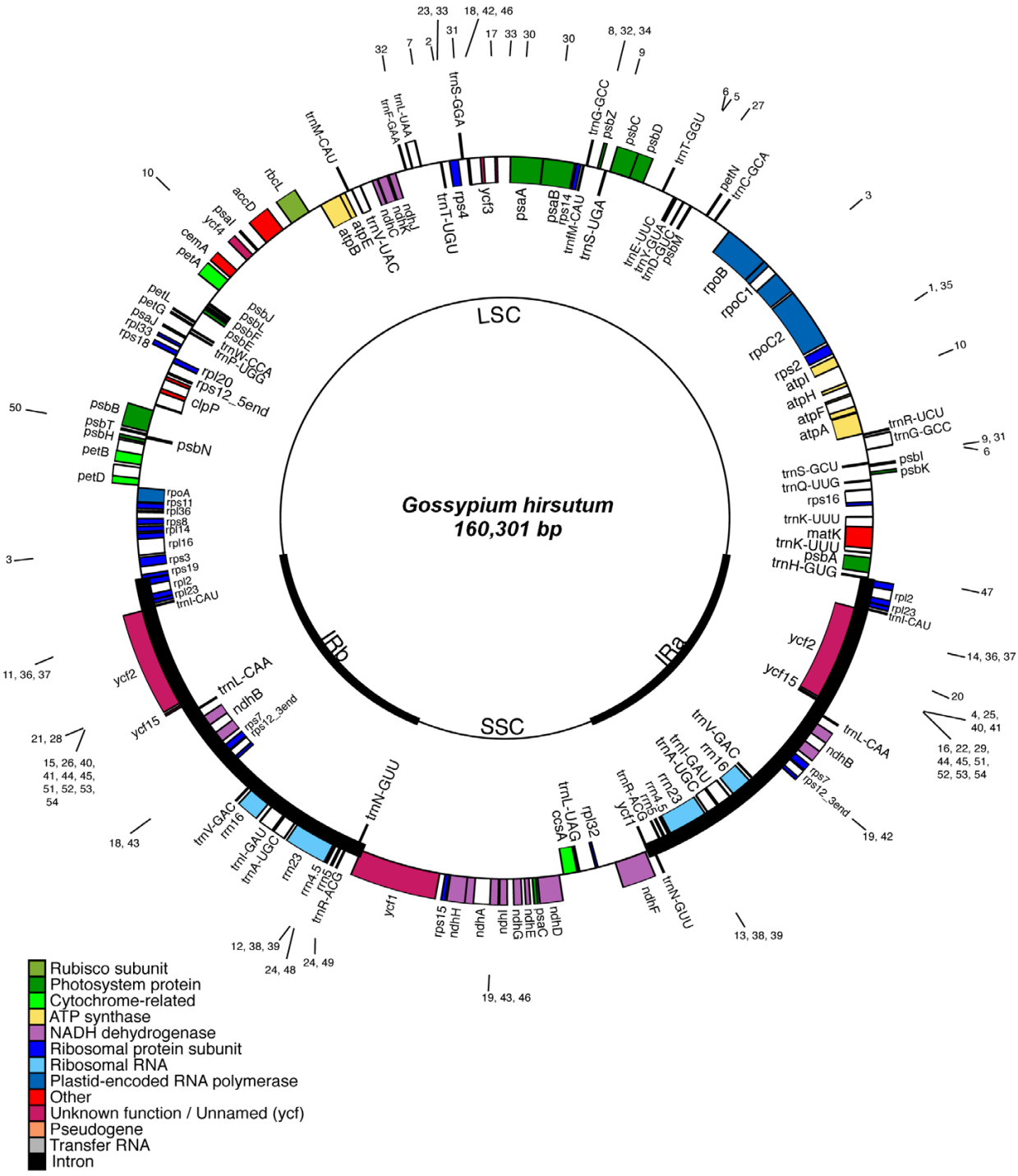

Figure 1
Gene map of the *Gossypium hirsutum* chloroplast genome. The thick lines indicate the extent of the inverted repeats (IRa and IRb), which separate the genome into small (SSC) and large (LSC) single copy regions. Genes on the outside of the map are transcribed in the clockwise direction and genes on the inside of the map are transcribed in the counterclockwise direction. Numbered lines around the map indicate the location of repeated sequences found in the cotton genome (see Table 1 for details). The SSC region is in the reverse orientation relative to tobacco [80]. This does not reflect any differences in gene order for cotton but simply reflects the well-known phenomenon that the SSC exists in two orientations in chloroplast genomes [84].

copy region of 20,269 bp and 88,816 bp, respectively. There are 112 unique genes within the cotton chloroplast genome and 19 of these are duplicated in the IR, giving a total of 131 genes (Fig. 1). Furthermore, there are four ribosomal and 30 distinct tRNA genes; seven of the tRNA genes and all rRNA genes are duplicated within the IR. There are 17 intron-containing genes, 15 of which contain one intron, whereas the remaining two have two introns. The gene order in the cotton plastid genome is identical to that of tobacco, but cotton lacks the *rpl22* and *infA* genes. Overall, genomic content is 37.25% GC and 62.75% AT, where 56.46% of the genome corresponds to protein coding genes and 43.54% to non-coding regions, including introns and intergenic spacers.

Repeat structure

Repeat analysis identified 30 direct and 24 inverted repeats 30 bp or longer with a sequence identity of at least 90% (Fig. 2 and Table 1). Twenty-three direct and 15 inverted repeats are 30 to 40 bp long, and the longest direct repeat is 72 bp. Most of the direct repeats are within intergenic spacer regions, intron sequences and *ycf2*, an essential hypothetical chloroplast gene [33]. Interestingly, a 72 bp-long direct repeat was found in the *psaA* and *psaB* genes, whereas a 34-bp forward repeat was within the *rrn23* gene, and a shorter, 32 bp-long direct repeat was identified in two serine transfer-RNA(*trnS*) genes that recognize different codons; *trnS-GCU* and *trnS-UGA*.

RNA editing

Comparison of the nucleotide sequences of protein coding genes and EST sequences retrieved from GenBank revealed that *rps16*, *rpl2*, *rpoC2*, *rps4* and *ycf1* have 100% sequence identity with their respective ESTs (data not shown). Eleven non-synonymous nucleotide substitutions, resulting in a total of nine amino acid changes, were identified within *ndhC*, *rpl23*, *rpl20*, *rps3* and *clpP* compared to respective ESTs, although their sequence identity was above 98% (Table 2). Surprisingly, there were no synonymous substitutions. All of the five aforementioned genes experienced one or two nucleotide substitutions, apart from the protease-encoding *clpP*, which had five variable sites. Lastly, in all but *rpl23*, the nucleotide substitutions had an impact on the hydropathy of the amino acid because they changed the amino acids from aliphatic to hydrophilic, and vice versa.

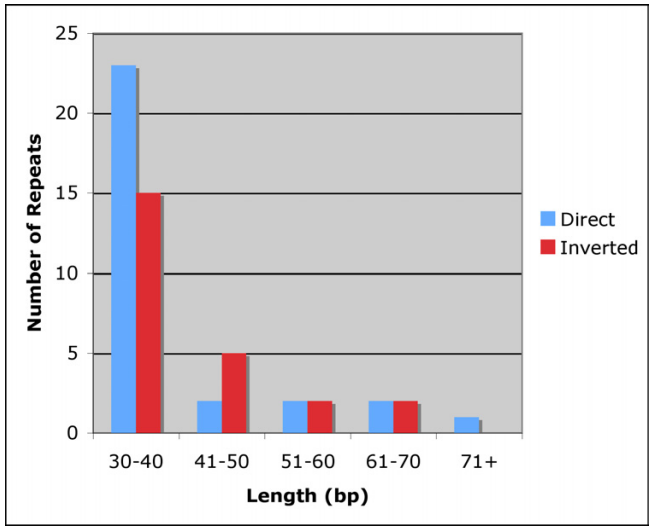

Figure 2
Histogram showing the number of repeated sequences ≥ 30 bp long with a sequence identity ≥ 90% in the cotton chloroplast genome.

Phylogenetic analysis

The data matrix for phylogenetic analyses included 61 protein-coding genes for 28 taxa (Table 3), including 26 angiosperms and two gymnosperm outgroups (*Pinus* and *Ginkgo*). The data set comprised 45,573 nucleotide positions but when the gaps were excluded there were 39,624 characters. Maximum Parsimony (MP) analyses resulted in a single, fully resolved tree with a length of 49,957, a consistency index of 0.46 (excluding uninformative characters) and a retention index of 0.62 (Fig. 3). Bootstrap analyses indicated that 24 of the 26 nodes were supported by values ≥ 95% with 19 of these with bootstrap values of 100%. Maximum Likelihood (ML) analysis resulted in a tree with a –lnL = 311251.33. The ML and MP trees had identical topologies so only the MP tree is shown in Figure 3.

Several major groups were supported within angiosperms and these groups are generally in agreement with recent classifications [26]. The most basal lineage was *Amborella* followed by the Nymphaeales. The next branch included *Calycanthus*, the sole representative of magnoliids in the data set. This was followed by a strongly supported clade of monocots, represented by members of three different orders (Acorales, Asparagales, and Poales). The monocots were then sister to the eudicots with the Ranunculales forming the earliest diverging eudicot clade. Within the core eudicots there were two major clades, one including the rosids and the second including the Caryophyllales sister to asterids. Within the rosid clade there were two major groups, the eurosids II and a group that included the Myrtales sister to the eurosids I. *Gossypium* in the Malvales was sister to *Arabidopsis* in the Brassicales.

Discussion

Implications for integration of transgenes

We have recently demonstrated stable transformation of the cotton plastid genome and maternal inheritance of transgenes via somatic embryogenesis [21]. In contrast to previous reports on integrating foreign genes in tomato

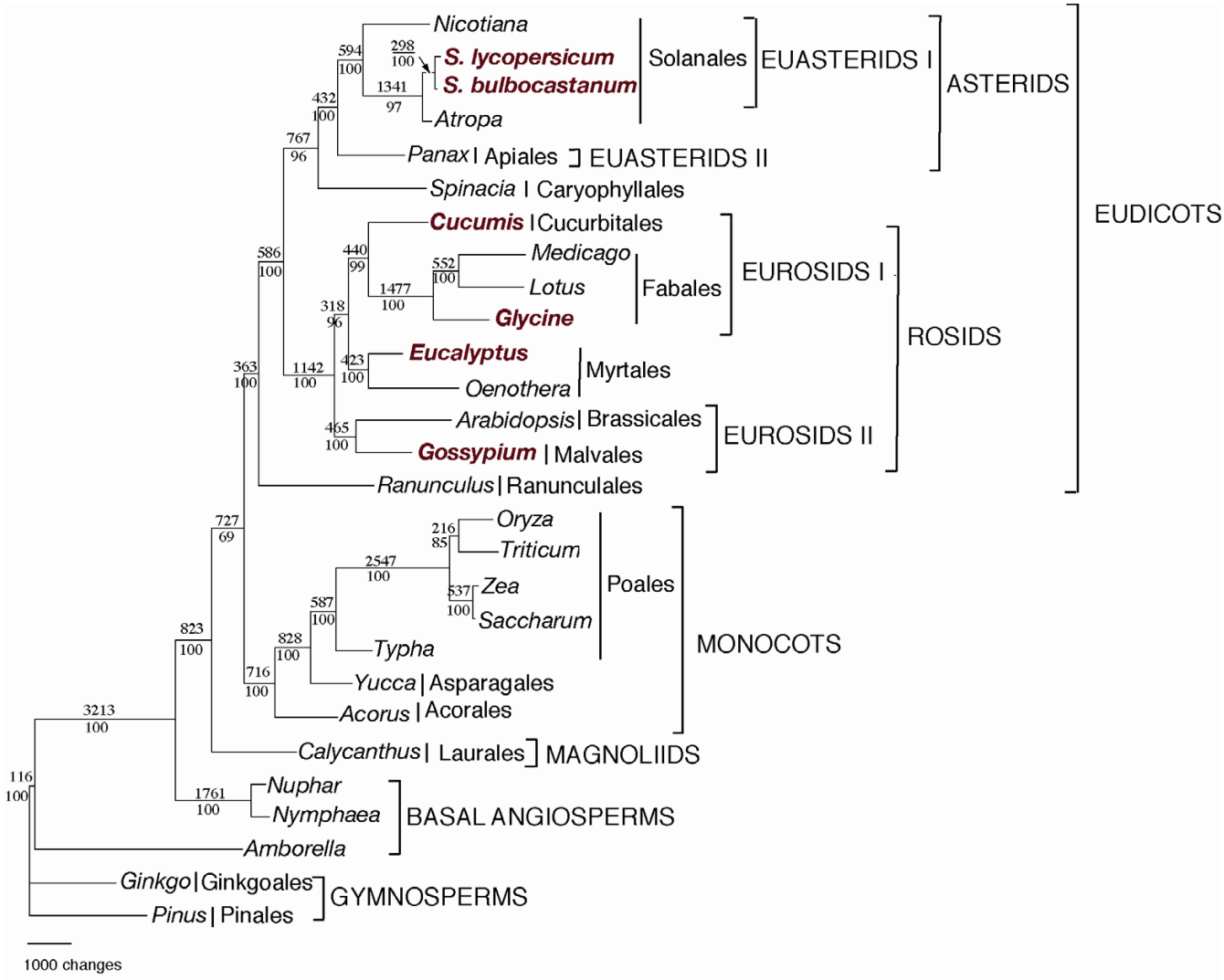

Figure 3
Parsimony tree based on 61 chloroplast protein-coding genes. The tree has a length of 49,957, a consistency index of 0.46 (excluding uninformative characters) and a retention index of 0.6. Numbers above node indicate number of changes along each branch and numbers below nodes are bootstrap support values. Taxa in red are those which have not appeared in any previous phylogenetic studies using 61 genes from complete chloroplast genome sequences. Ordinal and higher level group names follow APG II [85]. The maximum likelihood tree has the same topology but is not shown.

and potato chloroplast genomes using tobacco flanking sequences that do not have 100% sequence identity [24,34,35], the cotton plastid transformation vector was constructed using the PCR-amplified native cotton 16S/*trnI-trnA*/23S sequence. However, regulatory sequences used in the cotton plastid transformation were derived from tobacco or other heterologous sequences. With the availability of the entire cotton chloroplast genome sequence, it should now be possible to utilize endogenous regulatory sequences. Species-specific vectors should be effective for plastid transformation, especially in recalcitrant plants, because of transgene integration using flanking sequences with 100% sequence identity and endogenous promoters, 5' & 3'untranslated regions, thereby enhancing transcription and translation of transgenes. Also, the complete chloroplast genome provides the option of transgene integration into transcriptionally silent, active or read-through spacer regions for optimal transgene integration.

Thus far, transgenes conferring several useful agronomic traits, including insect [4,36,37], herbicide [8,38], and disease resistance [39], drought [13] and salt tolerance [20], phytoremediation [5], as well as cytoplasmic male

Table 1: Location of identified repeats in the cotton plastid genome. Table includes repeats at least 30 bp in size, with a sequence identity greater than or equal to 90%. IGS = Intergenic spacer. See Fig. 1 for location of repeats on the gene map.

Repeat Number	Size (bp)	Location
1	30	IGS
2	30	IGS
3	30	*rpoC1* intron, *rpl16* intron
4	30	*ycf2*
5	31	IGS
6	32	*psbI* (5 bp) – IGS, IGS
7	32	IGS
8	32	IGS
9	32	IGS (4 bp) – *trnS-GCU*, IGS (4 bp) – *trnS-UGA*
10	32	IGS
11	34	*ycf2*
12	34	IGS
13	34	*rrn23* exon
14	34	*ycf2*
15	34	*ycf2*
16	34	*ycf2*
17	35	*ycf3* intron
18	36	*ycf3* intron, IGS
19	38	*ndhA* intron, *rps12_3end* intron
20	38	*ycf2*
21	38	*ycf2*
22	38	*ycf2*
23	40	IGS
24	43	IGS
25	47	*ycf2*
26	52	*ycf2*
27	58	IGS
28	64	*ycf2*
29	64	*ycf2*
30	72	*psaA* exon, *psaB* exon
31	30	IGS (2 bp) – *trnS-GCU*, *trnS-GGA*
32	30	IGS
33	30	IGS
34	31	IGS
35	34	IGS
36	34	*ycf2*
37	34	*ycf2*
38	34	IGS
39	34	IGS
40	34	*ycf2*
41	34	*ycf2*
42	36	*ycf3* intron, IGS
43	38	IGS, *ndhA* intron
44	38	*ycf2*
45	38	*ycf2*
46	41	*ycf3* intron, *ndhA* intron
47	41	IGS
48	43	IGS
49	43	IGS
50	48	IGS
51	52	*ycf2*
52	52	*ycf2*
53	64	*ycf2*
54	64	*ycf2*

Table 2: Differences observed by comparison of cotton chloroplast genome sequences with EST sequences obtained by BLAST searches of GenBank.

Gene	Gene size (bp)	Sequence analyzed[a]	Number of variable sites	Variation type	Position(s)[b]	Amino acid change
clpP	591	228–537	5	A-G	523	M-A
				T-C	524	
				T-A	528	I-M
				T-G	531	G-S
				G-A	532	
ndhC	363	76–363	1	T-C	323	L-S
rpl20	354	1–354	2	A-G	263	K-R
				C-U	308	S-L
rpl23	282	85–282	1	C-U	89	S-L
rps3	657	274–657	2	T-G	275	L-R
				A-C	302	K-T

[a]Sequence analyzed coordinates based on the gene sequence, considering the first base of the initiation codon as bp 1. [b]Variable position is given in reference to the first base of the initiation codon of the gene sequence.

sterility [12], have been stably integrated and expressed, via the tobacco chloroplast genome. Using the chloroplast as a bioreactor, vaccine antigens [15,40-42], human therapeutic proteins [17,43-45], industrial enzymes [46] and

Table 3: Taxa included phylogenetic analyses with GenBank accession numbers and references. Taxa in bold are those which have not appeared in any previous phylogenetic studies using 61 genes from complete chloroplast genome sequences.

Taxon	GenBank Accession Numbers	Reference
Gymnosperms –Outgroups		
Pinus thunbergii	NC_001631	Wakasugi et al. 1994 [72]
Ginkgo biloba	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Basal Angiosperms		
Amborella trichopoda	NC_005086	Goremykin et al. 2003 [29]
Nuphar advena	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Nymphaea alba	NC_006050	Goremykin et al. 2004 [28]
Monocots		
Acorus americanus	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Oryza sativa	NC_001320	Hiratsuka et al. 1989 [73]
Saccharum officinarum	NC_006084	Asano et al. 2004 [74]
Triticum aestivum	NC_002762	Ikeo and Ogihara, unpublished
Typha latifolia	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Yucca schidigera	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Zea mays	NC_001666	Maier et al. 1995 [75]
Magnoliids		
Calycanthus floridus	NC_004993	Goremykin et al. 2003 [76]
Eudicots		
Arabidopsis thalliana	NC_000932	Sato et al. 1999 [77]
Atropa belladonna	NC_004561	Schmitz-Linneweber et al. 2002 [53]
Cucumis sativus	NC_007144	**Plader et al. unpublished**
Eucalyptus globulus	AY780259	**Steane 2005 [78]**
Glycine max	DQ317523	**Saski et al. 2005 [3]**
Gossypium hirsutum	DQ345959	**Current study**
Lotus corniculatus	NC_002694	Kato et al. 2000 [79]
Medicago truncatula	NC_003119	Lin et al., unpublished
Nicotiana tabacum	NC_001879	Shinozaki et al. 1986 [80]
Oenothera elata	NC_002693	Hupfer et al. 2000 [81]
Panax schinseng	NC_006290	Kim and Lee 2004 [82]
Ranunculus macranthus	DQ069337–DQ069702	Leebens-Mack et al 2005 [27]
Solanum lycopersicum	DQ347959	**Daniell et al. in press**
Solanum bulboscastanum	DQ347958	**Daniell et al. in press**
Spinacia oleracea	NC_002202	Schmitz-Linneweber et al. 2001 [83]

biomaterials [6,47,48] have been produced successfully in an environmental friendly way. Although many successful examples of plastid engineering in tobacco have set a solid foundation for various future applications, this technology has not been extended to many of the major crops, primarily due to the lack of complete chloroplast genome sequences and challenges in achieving homoplasmy in recalcitrant crops.

Evolutionary implications

Other than the IR, repeated sequences are generally considered to be uncommon in chloroplast genomes [1]. Furthermore, previous studies based on both filter hybridization and DNA sequencing have indicated that dispersed repeats are found more commonly in genomes that have experienced changes in genome organization [49,56], especially in highly rearranged algal genomes [51,52]. The most extensive examination of repeat structure in angiosperms was performed in legumes [3], which do have a single inversion and in some taxa a loss of one copy of the IR. These repeat analyses identified a substantial number highly conserved repeats ≥ 30 bp with a sequence identity of ≥ 90%. Many of these repeats were located in intergenic spacer regions and introns, with several located in the coding regions of *psaA*, *psaB*, and *ycf2*. Our examination of repeats in the cotton chloroplast genome (Table 1, Fig. 2) identified similar numbers of repeats as in legumes [3], and these are also located mostly in intergenic spacer regions and introns. Repeats in coding regions of cotton are located in the same genes as in legumes. Overall, it appears that dispersed repeats are very common in angiosperm chloroplast genomes, even in genomes that have not experienced rearrangements. Future comparative studies are needed to determine the functional and evolutionary role these repeats may play in chloroplast genomes.

DNA and EST sequence comparisons identified many nucleotide substitutions resulting in amino acid changes. Based on previous studies of *Atropa* [53] and tobacco [54], posttranscriptional RNA editing events result predominantly in C-to-U edits. However, analysis of the cotton genome and EST sequences indicates that only two of the eleven differences were C-to-U changes, suggesting that most of these changes are not mRNA edits but may simply represent intra-species polymorphisms. Evolutionary loss of RNA editing sites has been previously observed and could possibly be due to a decrease in the effect of RNA-editing enzymes [31]. Additionally, conversions other than C-to-U in cotton, as well as other crops, suggest that chloroplast genomes may be accumulating considerable amounts of nucleotide substitutions, where some genes might accrue more alterations than others, such as the *petL* and *ndh* genes that have a high frequency of RNA editing [55]. Therefore, despite the plastome's high conservation, variations occur post-transcriptionally, promoting translational efficiency due to transcript-protein complex binding and/or changes in the chloroplast microenvironment, like redox potential or light intensity [56,57].

The phylogeny based on 61 protein-coding genes for 28 angiosperms is congruent with relationships suggested in previous studies [summarized in [26]]. There is strong support for the monophyly all of the major clades of angiosperms, including monocots, eudicots, rosids, asterids, eurosids I, eurosids II, asterids I and asterid II. Our phylogenetic analyses have greatly expanded the taxon sampling of entire genomes because we included six genomes (in bold in Table 1 and Fig. 3) that have not been included in recently published phylogenies based on complete chloroplast genomes [27-29,58]. The sampling is particularly expanded in the rosids with four of the six genomes from this clade. Thus, we will focus our discussion of the phylogenetic implications of this expanded analysis on this group.

The rosid clade is very large and includes nearly 140 families representing almost one third of all angiosperms. The most recent phylogenies of this group [summarized in chapter 8 in [26]] indicate that there are seven major clades whose relationships still remain unresolved. Representatives of three of these major clades are included in our analyses, eurosids I, eurosids II, and Myrtales. The position of the Myrtales has been especially controversial with no clear resolution of the relationship of this order to other members of the rosids. Our 61 gene chloroplast phylogeny (Fig. 3) provides strong support for a sister relationship of the Myrtales with the eurosid I clade. A three-gene phylogeny of 560 angiosperms is congruent with our results [59], although support was very weak. However, a sister relationship between eurosids I and Myrtales is in conflict with two other recent phylogenies based on two chloroplast genes (*atpB*, *rbcL*), which placed the Myrtales sister to the eurosid II clade with weak support [60,61]. Although our results clearly favor a closer relationship of Myrtales to the eurosid I clade, expanded sampling of complete chloroplast genome sequences of rosids is needed to resolve this issue, especially since limited taxon sampling can lead to erroneous tree topologies [27,30].

Our chloroplast phylogeny (Fig. 3) also supports the sister relationship between the orders Cucurbitales and Fabales, two of the four nitrogen fixing clades of eurosids I. Furthermore, the position of cotton, a member of the order Malvales, as sister to *Arabidopsis* in the Brassicales, is in agreement with recently phylogenies of the eurosid II clade [26].

Conclusion

Our complete sequence of the cotton chloroplast genome provides the needed information for expanding chloroplast genetic engineering to this important crop plant. Although genome organization of cotton is very similar to other unrearranged angiosperm chloroplast genomes, identification of disperse repeats and potential RNA editing sites provides new insights into the evolution of this genome. Finally, phylogenetic analyses of sequences of 61 protein-coding genes for 26 angiosperms suggests that the order Myrtales is sister to the eurosid I clade but denser sampling is needed to test this result rigorously.

Methods

DNA isolation and amplification

Gossypium hirsutum plants cv. Coker310FR were grown from seedlings in soil pots, until they were 1 m tall. Prior to DNA extraction, the plants were placed in the dark for two days to reduce the chloroplast starch levels. After that, 10 g of young leaf tissue was collected for cpDNA isolation based on the sucrose step gradient centrifugation method by Sandbrink *et al* [62]. Isolation was followed by whole chloroplast genome Rolling Circle Amplification (RCA), using the Repli-g RCA kit (Qiagen, Inc.) following the methods outlined in [63]. After incubation at 30°C for 16 hr, the reaction was terminated with 10-minute incubation at 65°C. Digestion of the RCA product with *BstXI*, *EcoRI* and *HindIII* allowed verification of successful RCA plastome amplification, as well as assessment of its quality, prior to DNA sequencing.

DNA sequencing and genome assembly

DNA was sheared by nebulization, size fractionated to 4–6 kb, linker ligated and cloned into pHOS2, a TIGR medium copy vector. A total of 1619 good reads with an average length of 812 bases was generated during the random (1396 reads) and closure (223 reads) phases of sequencing. Sequences were assembled using TIGR assembler [64] and scaffolded using Bambus [65]. Sequence finishing included directed PCR to span gaps, directed primer walking on clones and transposon mediated sequencing of full clones to cover the entire genome and complete regions of low coverage and manual editing of sequences to resolve inconsistencies.

Gene annotation

The cotton genome was annotated using DOGMA [Dual Organellar GenoMe Annotator, [66]], after uploading a FASTA-formatted file of the complete plastid genome to the program's server. BLASTX and BLASTN searches, against a custom database of previously published plastid genomes, identified cotton's putative protein-coding genes, and tRNAs or rRNAs. For genes with low sequence identity, manual annotation was performed, after identifying the position of the start and stop codons, as well as the translated amino acid sequence, using the plastid/bacterial genetic code.

Examination of repeat structure

REPuter [67] was used to locate and count the direct (forward) and inverted (palindromic) repeats within the cotton chloroplast genome. For repeat identification, the following constraints were used: (i) minimum repeat size of 30 bp, and (ii) 90% or greater sequence identity, based on Hamming distance equal to 3 bp [3]. Manual verification of the identified repeats was performed in EditSeq, while performing intragenomic blast search of the identified repeat sequence.

Variation between coding sequences and cDNAs

Each of the gene sequences from the cotton chloroplast genome was used to perform a BLAST search of expressed sequence tags (ESTs) from GenBank. The retrieved *Gossypium hirsutum* ESTs were aligned with the corresponding annotated gene using ClustalX [68], followed by screening for nucleotide and amino acid changes using Megalign and its' plastid/bacterial genetic code. Because of variation in the length between an EST and the related gene, the length of the analyzed sequence was recorded.

Phylogenetic analysis

The 61 genes included in the analyses of Goremykin et al. [28,29] and Leebens-Mack et al. [27] were extracted from our new chloroplast genome sequences of cotton using the organellar genome annotation program DOGMA. [66]. The same set of 61 genes was extracted from chloroplast genome sequences of five other recently sequenced angiosperm chloroplast genomes, including tomato, potato, soybean, cucumber, and *Eucalyptus* (see Table 3 for complete list of genomes examined). In general, alignment of the DNA sequences was straightforward and simply involved removing gaps included in the data set because of the elimination of non-seed plants and adding the 61 genes for the new angiosperms to the aligned data matrix from Leebens-Mack et al. [27]. In some cases, small in frame insertions or deletions were required for correct alignment. For two genes, *ccsA* and *matK*, the DNA sequences were more divergent, requiring alignment using ClustalX [68] followed by manual adjustments.

Phylogenetic analyses using maximum parsimony (MP) and maximum likelihood (ML) were performed using PAUP* version 4.10 [69]. All phylogenetic analyses excluded gap regions. All MP searches were heuristic with 100 random addition replicates and TBR branch swapping with the Multrees option. The Hasegawa-Kishino-Yano (HKY; [70]) model of molecular evolution was used in ML analyses of the nucleotide sequences. ML analyses used TBR branch swapping with the Multrees option and one random addition replicate. Non-parametric bootstrap

analyses [71] were performed for MP analyses with 1000 replicates with TBR branch swapping, one random addition replicate, and the Multrees option and for ML analyses with 100 replicates with NNI branch swapping, one random addition replicate, and the Multrees option.

Abbreviations

cpDNA, chloroplast DNA; IR inverted repeat; SSC, small single copy; LSC, large single copy, bp, base pair; ycf, hypothetical chloroplast reading frame; rrn, ribosomal RNA; MP, maximum parsimony; ML, maximum likelihood; EST, expressed sequence tag; cDNA, complementary DNA.

Authors' contributions

SBL isolated chloroplasts, performed RCA amplification of cpDNA, genome annotation, analysis and submission of data to the GenBank; CK performed the repeat analyses, comparisons of DNA and EST sequences, assisted with extraction & alignment of DNA sequences for phylogenetic analyses and wrote a few sections of the first draft; JBH, LJT and CDT performed DNA sequencing and genome assembly; RKJ assisted with extracting and aligning DNA sequences, performed phylogenetic analyses, and wrote the phylogenetic portions of the manuscript; HD conceived and designed this study, interpreted data, wrote and revised several versions of this manuscript. All authors read and approved the final manuscript.

Acknowledgements

Investigations reported in this article were supported in part by grants from USDA 3611-21000-017-00D to Henry Daniell and from NSF DEB 0120709 to Robert K. Jansen.

References

1. Palmer JD: **Plastid chromosomes: structure and evolution.** In *The Molecular Biology of Plastids* Edited by: Bogorad L, Vasil K. San Diego: Academic Press; 1991:5-53.
2. Raubeson LA, Jansen RK: **Chloroplast genomes of plants.** In *Diversity and Evolution of Plants-Genotypic and Phenotypic Variation in Higher Plants* Edited by: Henry H. Wallingford: CABI Publishing; 2005:45-68.
3. Saski C, Lee S, Daniell H, Wood T, Tomkins J, Kim H-G, Jansen RK: **Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes.** *Plt Mol Biol* 2005, **59:**309-322.
4. 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.
5. Ruiz ON, Hussein H, Terry N, Daniell H: **Phytoremediation of organomercurial compounds via chloroplast genetic engineering.** *Plt Phys* 2003, **32:**1344-1352.
6. Lossl A, Eibl C, Harloff HJ, Jung C, Koop HU: **Polyester synthesis in transplastomic tobacco (*Nicotiana tabacum* L.): significant contents of polyhydroxybutyrate are associated with growth reduction.** *Plt Cell Rep* 2003, **21:**891-899.
7. Quesada-Vargas T, Ruiz ON, Daniell H: **Characterization of heterologous multigene operons in transgenic chloroplasts: transcription, processing, translation.** *Plt Physiol* 2005, **138:**1746-1762.
8. Daniell H, Datta R, Varma S, Gray S, Lee SB: **Containment of herbicide resistance through genetic engineering of the chloroplast genome.** *Nat Biotechnol* 1998, **16:**345-348.
9. Scott SE, Wilkenson MJ: **Low probability of chloroplast movement from oilseed rape (*Brassica napus*) into wild *Brassica rapa*.** *Nat Biotechnol* 1999, **17:**390-392.
10. Daniell H: **Molecular strategies for gene containment in transgenic crops.** *Nat Biotechnol* 2002, **20:**581-586.
11. Hagemann R: **The Sexual Inheritance of Plant Organelles.** In *Molecular Biology and Biotechnology of Plant Organelles* Edited by: Daniell H, Chase C. Dordrecht, The Netherlands: Springer Publishers; 2004:93-113.
12. Ruiz ON, Daniell H: **Engineering cytoplasmic male sterility via the chloroplast genome.** *Plt Physiol* 2005, **138:**1232-1246.
13. Lee SB, Kwon HB, Kwon SJ, Park SC, Jeong MJ, Han SE, Daniell H: **Accumulation of trehalose within transgenic chloroplasts confers drought tolerance.** *Mol Breed* 2003, **11:**1-13.
14. Daniell H, Khan M, Allison L: **Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology.** *Trends Plt Sci* 2002, **7:**84-91.
15. Daniell H, Lee SB, Panchal T, Wiebe PO: **Expression of cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts.** *J Mol Biol* 2001, **311:**1001-1009.
16. Leelavathi S, Reddy VS: **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.
17. Daniell H, Carmona-Sanchez O, Burns BB: **Chloroplast-derived vaccine antibodies, biopharmaceuticals, and edible vaccines in transgenic plants engineered via the chloroplast genome.** In *Molecular Farming Volume Chapter 8*. Edited by: Schillberg S. Germany: Wiley-VCH Verlag; 2004:113-133.
18. Daniell H, Cohill PR, Kumar S, Dufourmantel N: **Chloroplast Genetic Engineering.** In *Molecular Biology and Biotechnology of Plant Organelles* Edited by: Daniell H, Chase CD. Netherlands: Springer Publishers; 2004:443-490.
19. Dufourmantel N, Pelissier B, Garçon F, Peltier JM, Tissot G: **Generation of fertile transplastomic soybean.** *Plt Mol Biol* 2004, **55:**479-89.
20. Kumar S, Dhingra A, Daniell H: **Plastid expressed *betaine aldehyde dehydrogenase* gene in carrot cultured cells, roots and leaves confers enhanced salt tolerance.** *Plt Physiol* 2004, **136:**2843-2854.
21. Kumar S, Dhingra A, Daniell H: **Stable transformation of the cotton plastid genome and maternal inheritance of transgenes.** *Plt Mol Biol* 2004, **56:**203-216.
22. Daniell H, Kumar S, Duformantel N: **Breakthrough in chloroplast genetic engineering of agronomically important crops.** *Trends Biotechnol* 2005, **23:**238-245.
23. Maier RM, Schmitz-Linneweber C: **Plastid genomes.** In *Molecular Biology and Biotechnology of Plant Organelles* Edited by: Daniell H, Chase CD. Netherlands: Springer publishers; 2004:115-150.
24. Ruf S, Hermann M, Berger I, Carrer H, Bock R: **Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit.** *Nat Biotechnol* 2001, **19:**870-875.
25. Llewellyn D, Fitt G: **Pollen dispersal from two field trials of transgenic cotton in the Namoi valley, Australia.** *Mol Breeding* 1996, **2:**157-166.
26. Soltis DE, Soltis PS, Endress PK, Chase MW: *Phylogeny and evolution of Angiosperms* Sunderland Massachusetts: Sinauer Associates Inc; 2005.
27. 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.
28. Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH: **The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm.** *Mol Biol Evol* 2004, **21:**1445-1454.
29. Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH: **Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm.** *Mol Biol Evol* 2003, **20:**1499-1505.
30. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu Y-Q, Chase MW, Farris JS, Stefanoviæ S, Rice DW, Palmer JD, Soltis PS: **Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics.** *Trends Plant Sci* 2004, **9:**477-483.

31. Wolf PG, Rowe CA, Hasebe M: **High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*.** *Gene* 2004, **339:**89-97.
32. Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K: **RNA editing in hornwort chloroplasts makes more than half the genes functional.** *Nucl Acids Res* 2003, **31:**2417-2423.
33. Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R: **The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes.** *Plt J* 2000, **22:**97-104.
34. Sidorov VA, Kasten D, Pang SZ, Hajdukiewicz PT, Staub JM, Nehra NS: **Technical advance: stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker.** *Plant J* 1999, **19:**209-216.
35. Nguyen TT, Nugent G, Cardi T, Dix PJ: **Generation of homoplasmic plastid transformants of a commercial cultivar of potato (*Solanum tuberosum* L.).** *Plt Sci* 2005, **168:**1495-1500.
36. McBride KE, Svab Z, Schaaf DJ, Hogan PS, Stalker DM, Maliga P: **Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco.** *BioTechn* 1995, **13:**362-365.
37. 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.** *Proc Natl Acad Sci USA* 1999, **96:**1840-1845.
38. Iamtham S, Day A: **Removal of antibiotic resistance genes from transgenic tobacco plastids.** *Nat Biotechnol* 2000, **18:**1172-1176.
39. DeGray G, Rajasekaran K, Smith F, Sanford J, Daniell H: **Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi.** *Plant Physiology* 2001, **127:**852-862.
40. 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.
41. Koya V, Moayeri M, Leppla SH, Daniell H: **Plant based vaccine: mice immunized with chloroplast-derived anthrax protective antigen survive anthrax lethal toxin challenge.** *Infection and Immunity* 2005, **73:**8266-8274.
42. 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.
43. Staub JM, Garcia B, Graves J, Hajdukiewicz PTJ, Hunter P, Nehra N: **High-yield production of a human therapeutic protein in tobacco chloroplasts.** *Nat Biotechnol* 2000, **18:**333-338.
44. Fernandez-San Millan A, Mingo-Castel A, Miller M, Daniell H: **A chloroplast transgenic approach to hyper express and purify human serum albumin, a protein highly susceptible to proteolytic degradation.** *Plt Biotechn J* 2003, **1:**71-79.
45. Grevich JJ, Daniell H: **Chloroplast genetic engineering: Recent advances and future perspectives.** *Crit Rev Plt Sci* 2005, **24:**83-108.
46. 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.
47. Guda C, Lee SB, Daniell H: **Stable expression of biodegradable protein based polymer in tobacco chloroplasts.** *Plt Cell Rep* 2000, **19:**257-262.
48. Vitanen PV, Devine AL, Kahn S, Deuel DL, Van-Dyk DE, Daniell H: **Metabolic engineering of the chloroplast genome using the *E. coli ubiC* gene reveals that corismate is a readily abundant precursor for 4-hydroxybenzoic acid synthesis in plants.** *Plt Phys* 2004, **136:**4048-4060.
49. 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.
50. Milligan BG, Hampton JN, Palmer JD: **Dispersed repeats and structural reorganization in subclover chloroplast DNA.** *Mol Biol Evol* 1989, **6:**355-368.
51. Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB: **The *Chlamydomonas reinhardtii* plastid chromosome: Islands of genes in a sea of repeats.** *The Plant Cell* 2002, **14:**1-22.
52. Pombert J-F, Otis C, Lemieux C, Turmel M: **The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages.** *Mol Biol Evol* 2005, **22:**1903-1918.
53. Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM: **The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation.** *Mol Biol Evol* 2002, **19:**1602-1612.
54. Hirose T, Kusumegi T, Tsudzuki T, Sugiura M: **RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity.** *Mol Gen Genet* 1999, **262:**462-467.
55. Fiebig A, Stegemann S, Bock R: **Rapid evolution of RNA editing sites in a small non-essential plastid gene.** *Nucl Acids Res* 2004, **32:**3615-3622.
56. Monde RA, Schusterc G, Stern DB: **Processing and degradation of chloroplast mRNA.** *Biochimie* 2000, **82:**573-582.
57. Rochaix JD: **Posttranscriptional control of chloroplast gene expression. From RNA to photosynthetic complex.** *Plt Phys* 2001, **125:**142-144.
58. Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, Cheng C-H, Lin C-Y, Liu S-M, Chang C-C, Chaw S-M: **The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications.** *Mol Biol Evol* in press.
59. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WJ, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS: **Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences.** *Bot J Linn Soc* 2000, **133:**381-461.
60. Savolainen V, Chase MW, Morton CW, Soltis DE, Bayer C, Fay MF, de Bruijn A, Sullivan S, Qiu Y-L: **Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences.** *Syst Biol* 2000, **49:**306-362.
61. Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Lledo MD, Pintaud J-C, Powell M, Sheahan MC, Soltis DE, Soltis PS, Weston P, Whitten MW, Wurdack J, Chase MW: **Phylogeny of eudicots: a nearly complete familial analysis based on *rbcL* gene sequences.** *Kew Bull* 2000, **55:**257-309.
62. Sandbrink JM, Vellekoop P, Vanham R, Vanbrederode J: **A method for evolutionary studies on RFLP of chloroplast DNA, applicable to a range of plant species.** *Biochem Syst Ecol* 1989, **17:**45-49.
63. 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.** *Methods in Enzymology* 2005, **395:**348-384.
64. Sutton GG, White O, Adams MD, Kerlavage AR: **TIGRAssembler: A new tool for assembling large shotgun sequencing projects.** *Gen Sci Techn* 1995, **1:**9-19.
65. Pop M, Kosack DS, Salzberg SL: **Hierarchical scaffolding with bambus.** *Gen Res* 2004, **14:**149-159.
66. Wyman SK, Boore JL, Jansen RK: **Automatic annotation of organellar genomes with DOGMA.** *Bioinformatics* 2004, **20:**3252-3255.
67. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: **REPuter: the manifold applications of repeat analysis on a genomic scale.** *Nucl Acids Res* 2001, **29:**4633-4642.
68. Higgins DG, Thompson JD, Gibson TJ: **Using CLUSTAL for multiple sequence alignments.** *Meth Enzy* 1996, **266:**383-402.
69. Swofford DL: *PAUP*: Phylogenetic analysis using parsimony (*and other methods), ver. 4.0* Sunderland MA: Sinauer Associates; 2003.
70. Hasegawa M, Kishino H, Yano T: **Dating of the human-ape splitting by a molecular clock of mitochondrial DNA.** *J Mol Evol* 1985, **22:**160-174.
71. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39:**783-791.
72. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M: **Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*.** *Proc Natl Acad Sci USA* 1994, **91:**9794-9798.
73. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M: **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.
74. Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K: **Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes.** *DNA Res* 2004, **11:**93-99.
75. Maier RM, Neckermann K, Igloi GL, Kossel H: **Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing.** *J Mol Biol* 1995, **251:**614-628.
76. Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig FH: **The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis* – structural and phylogenetic analyses.** *Plt Syst Evol* 2003, **242:**119-135.
77. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S: **Complete structure of the chloroplast genome of *Arabidopsis thaliana*.** *DNA Res* 1999, **6:**283-290.
78. Steane DA: **Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae).** *DNA Res* 2005, **12:**215-220.
79. Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S: **Complete structure of the chloroplast genome of a legume, *Lotus japonicus*.** *DNA Res* 2000, **7:**323-330.
80. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: **The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression.** *EMBO J* 1986, **5:**2043-2049.
81. Hupfer H, Swaitek M, Hornung S, Herrmann RG, Maier RM, Chiu WL, Sears B: **Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenthera* plastomes.** *Mol Gen Genet* 2000, **263:**581-585.
82. Kim K-J, Lee H-L: **Complete chloroplast genome sequence from Korean Ginseng (*Panax schiseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants.** *DNA Res* 2004, **11:**247-261.
83. Schmitz-Linneweber C, Maier RM, Alcaraz JP, Cottet A, Herrmann RG, Mache R: **The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization.** *Plt Mol Biol* 2001, **45:**307-315.
84. Palmer JD: **Chloroplast DNA exists in two orientations.** *Nature* **301:**92-93.
85. APG II: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II.** *Bot J Linn Soc* 2003, **141:**399-436.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral