



8-31-2006

## Complete Plastid Genome Sequence of *Daucus Carota*: Implications for Biotechnology and Phylogeny of Angiosperms

Tracey Ruhlman

Seung-Bum Lee

Robert K. Jansen

Jessica B. Hostetler

Luke J. Tallon

*See next page for additional authors*

Follow this and additional works at: [https://repository.upenn.edu/dental\\_papers](https://repository.upenn.edu/dental_papers)

 Part of the [Dentistry Commons](#)

---

### Recommended Citation

Ruhlman, T., Lee, S., Jansen, R. K., Hostetler, J. B., Tallon, L. J., Town, C. D., & Daniell, H. (2006). Complete Plastid Genome Sequence of *Daucus Carota*: Implications for Biotechnology and Phylogeny of Angiosperms. *BMC Genomics*, 7 (222), <http://dx.doi.org/10.1186/1471-2164-7-222>

At the time of publication, author Henry Daniell was affiliated with the University of Central Florida. Currently, he is a faculty member at the School of Dental Medicine at the University of Pennsylvania

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/dental\\_papers/302](https://repository.upenn.edu/dental_papers/302)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Complete Plastid Genome Sequence of *Daucus Carota*: Implications for Biotechnology and Phylogeny of Angiosperms

## Abstract

### Background

Carrot (*Daucus carota*) is a major food crop in the US and worldwide. Its capacity for storage and its lifecycle as a biennial make it an attractive species for the introduction of foreign genes, especially for oral delivery of vaccines and other therapeutic proteins. Until recently efforts to express recombinant proteins in carrot have had limited success in terms of protein accumulation in the edible tap roots. Plastid genetic engineering offers the potential to overcome this limitation, as demonstrated by the accumulation of BADH in chromoplasts of carrot taproots to confer exceedingly high levels of salt resistance. The complete plastid genome of carrot provides essential information required for genetic engineering. Additionally, the sequence data add to the rapidly growing database of plastid genomes for assessing phylogenetic relationships among angiosperms.

### Results

The complete carrot plastid genome is 155,911 bp in length, with 115 unique genes and 21 duplicated genes within the IR. There are four ribosomal RNAs, 30 distinct tRNA genes and 18 intron-containing genes. Repeat analysis reveals 12 direct and 2 inverted repeats  $\geq 30$  bp with a sequence identity  $\geq 90\%$ . Phylogenetic analysis of nucleotide sequences for 61 protein-coding genes using both maximum parsimony (MP) and maximum likelihood (ML) were performed for 29 angiosperms. Phylogenies from both methods provide strong support for the monophyly of several major angiosperm clades, including monocots, eudicots, rosids, asterids, eurosids II, euasterids I, and euasterids II.

### Conclusion

The carrot plastid genome contains a number of dispersed direct and inverted repeats scattered throughout coding and non-coding regions. This is the first sequenced plastid genome of the family Apiaceae and only the second published genome sequence of the species-rich euasterid II clade. Both MP and ML trees provide very strong support (100% bootstrap) for the sister relationship of *Daucus* with *Panax* in the euasterid II clade. These results provide the best taxon sampling of complete chloroplast genomes and the strongest support yet for the sister relationship of Caryophyllales to the asterids. The availability of the complete plastid genome sequence should facilitate improved transformation efficiency and foreign gene expression in carrot through utilization of endogenous flanking sequences and regulatory elements.

### Disciplines

Dentistry

### Comments

At the time of publication, author Henry Daniell was affiliated with the University of Central Florida. Currently, he is a faculty member at the School of Dental Medicine at the University of Pennsylvania

### Author(s)

Tracey Ruhlman, Seung-Bum Lee, Robert K. Jansen, Jessica B. Hostetler, Luke J. Tallon, Christopher D. Town, and Henry Daniell

Research article

Open Access

## Complete plastid genome sequence of *Daucus carota*: Implications for biotechnology and phylogeny of angiosperms

Tracey Ruhlman<sup>1</sup>, Seung-Bum Lee<sup>1</sup>, Robert K Jansen<sup>2</sup>, Jessica B Hostetler<sup>3</sup>, Luke J Tallon<sup>3</sup>, Christopher D Town<sup>3</sup> and Henry Daniell\*<sup>1</sup>

Address: <sup>1</sup>Dept. of Molecular Biology & Microbiology, University of Central Florida, Biomolecular Science, Building #20, Room 336, Orlando, FL 32816-2364, USA, <sup>2</sup>Section of Integrative Biology and Institute of Cellular and Molecular Biology, Patterson Laboratories 141, University of Texas, Austin, TX 78712, USA and <sup>3</sup>The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

Email: Tracey Ruhlman - daniell@mail.ucf.edu; Seung-Bum Lee - daniell@mail.ucf.edu; Robert K Jansen - jansen@mail.utexas.edu; Jessica B Hostetler - daniell@mail.ucf.edu; Luke J Tallon - daniell@mail.ucf.edu; Christopher D Town - daniell@mail.ucf.edu; Henry Daniell\* - daniell@mail.ucf.edu

\* Corresponding author

Published: 31 August 2006

Received: 19 May 2006

BMC Genomics 2006, 7:222 doi:10.1186/1471-2164-7-222

Accepted: 31 August 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/222>

© 2006 Ruhlman 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:** Carrot (*Daucus carota*) is a major food crop in the US and worldwide. Its capacity for storage and its lifecycle as a biennial make it an attractive species for the introduction of foreign genes, especially for oral delivery of vaccines and other therapeutic proteins. Until recently efforts to express recombinant proteins in carrot have had limited success in terms of protein accumulation in the edible tap roots. Plastid genetic engineering offers the potential to overcome this limitation, as demonstrated by the accumulation of BADH in chromoplasts of carrot taproots to confer exceedingly high levels of salt resistance. The complete plastid genome of carrot provides essential information required for genetic engineering. Additionally, the sequence data add to the rapidly growing database of plastid genomes for assessing phylogenetic relationships among angiosperms.

**Results:** The complete carrot plastid genome is 155,911 bp in length, with 115 unique genes and 21 duplicated genes within the IR. There are four ribosomal RNAs, 30 distinct tRNA genes and 18 intron-containing genes. Repeat analysis reveals 12 direct and 2 inverted repeats  $\geq 30$  bp with a sequence identity  $\geq 90\%$ . Phylogenetic analysis of nucleotide sequences for 61 protein-coding genes using both maximum parsimony (MP) and maximum likelihood (ML) were performed for 29 angiosperms. Phylogenies from both methods provide strong support for the monophyly of several major angiosperm clades, including monocots, eudicots, rosids, asterids, eurosids II, euasterids I, and euasterids II.

**Conclusion:** The carrot plastid genome contains a number of dispersed direct and inverted repeats scattered throughout coding and non-coding regions. This is the first sequenced plastid genome of the family Apiaceae and only the second published genome sequence of the species-rich euasterid II clade. Both MP and ML trees provide very strong support (100% bootstrap) for the sister relationship of *Daucus* with *Panax* in the euasterid II clade. These results provide the best taxon sampling of complete chloroplast genomes and the strongest support yet for the sister relationship of Caryophyllales to the asterids. The availability of the complete plastid genome sequence should facilitate improved transformation efficiency and foreign gene expression in carrot through utilization of endogenous flanking sequences and regulatory elements.

## Background

The plastid is a nearly autonomous organelle because it contains the biochemical machinery necessary to replicate and transcribe its own genome and carry out protein synthesis. Within angiosperms the plastid genome includes approximately 120 to 130 genes and usually ranges in size from 120 to 170 kilobases (kb) [1-3]. Of the estimated 3000 or so distinct proteins found in the higher plant plastid [4,5] only a small fraction are encoded by the plastid genome [6]. The bulk of the plastid proteome is nuclear encoded, translated on cytosolic ribosomes and subsequently translocated across the plastid envelopes [7].

The circular plastid genome is divided into four regions: large single copy (LSC), small single copy (SSC) and the inverted repeat (IR) which is present in exact duplicate separated by the two single copy regions. Restriction site analysis indicates that the molecule exists in two orientations present in equimolar proportions within a single plant [8]. The circular molecule undergoes interconversion into a dumbbell-shaped conformation that is facilitated by the IR. Concerted evolution within the IR [9,10] suggests intramolecular recombination between the repeats may be occurring.

The advantages of plastid transformation for bioengineering are several-fold and include the integration of multiple genes in a single transformation event [11-13], lack of gene silencing [14-16], position effect due to site-specific transgene integration [17], and minimization of pleiotropic effects due to compartmentalization of recombinant proteins [15,18,19]. The presence of many copies of the plastid genome within the many plastids in each cell contributes to high levels of foreign protein expression [14]. Plastid genetic engineering could minimize transgene escape because of maternal inheritance of transgenes [17,20-24] and the possibility of employing cytoplasmic male sterility to contain transgenes [25].

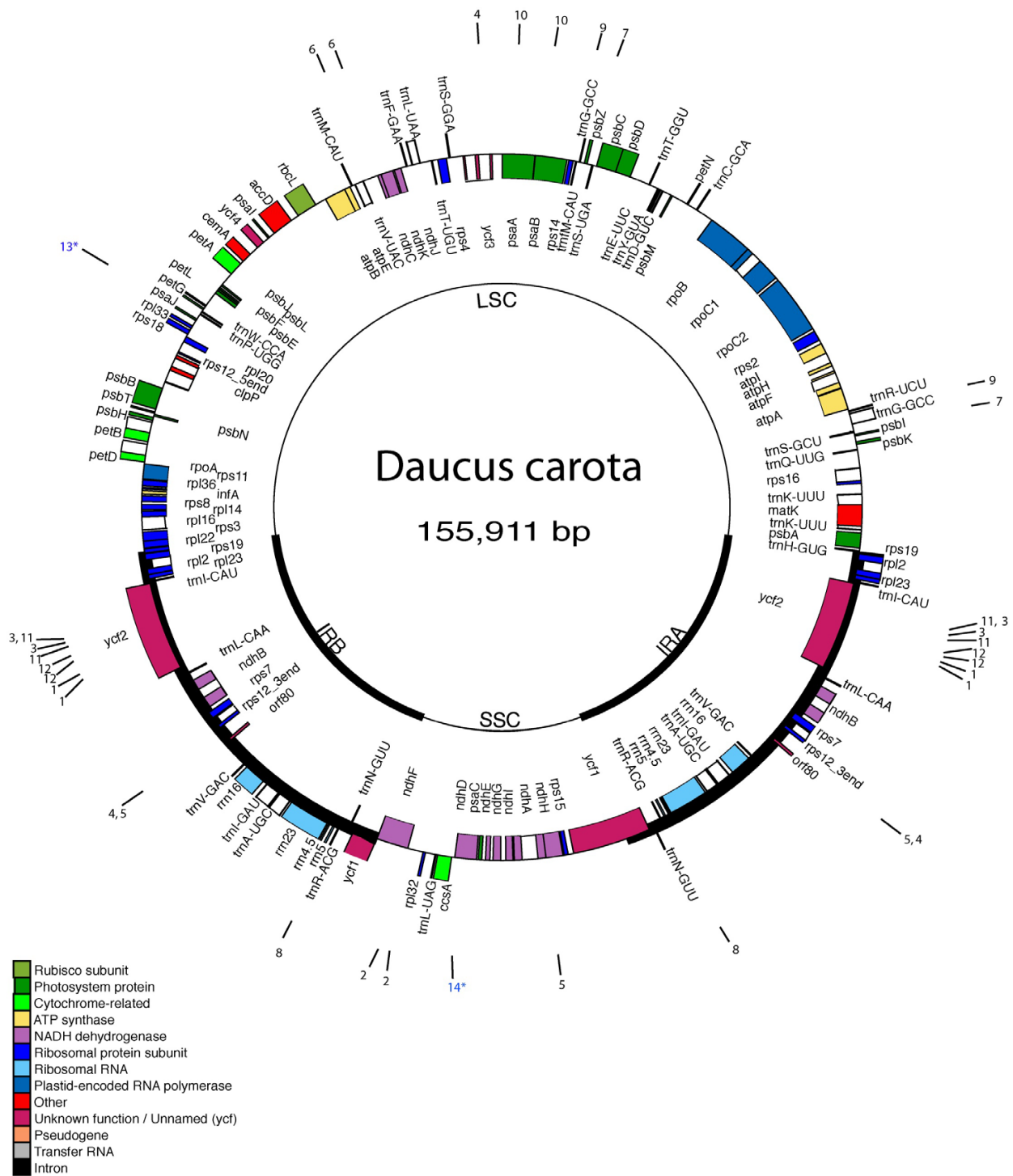
The ability to transform the plastid genome of higher plants has facilitated the accumulation of foreign proteins previously found to be recalcitrant in plant expression systems [26]. Until very recently breakthroughs in this regard have been limited to *Nicotiana tabacum* (tobacco) where plastid transformation has become routine [27,28]. As the field progresses to encompass the expression of vaccine antigens and other therapeutic proteins via the plastid genome [29], there is a growing interest in developing a crop system for oral delivery of these recombinant products. *Daucus carota* (carrot) has been proposed as an ideal candidate for this application for several reasons. Cultivated carrot is a biennial, the reproductive structures are not present until the second year, yet the root crop is suitable for harvest in the first year [30]. This feature further

ensures the ability to contain foreign genes in the field by eliminating the possibility of dispersal by pollen and seed. In terms of storage, the root may be realistically maintained up to six months without any processing under typical commercial conditions [31]. With an average annual value of over 70 million dollars, carrot ranks in the top ten among commercial vegetable crops in the United States adding to the interest in biotechnological improvement of this species [31].

Transformation of the carrot plastid genome has been accomplished, and expression of betaine aldehyde dehydrogenase (BADH) from spinach in carrot plastids was found to confer salt tolerance up to 400 mM NaCl [32]. In this case native carrot plastid sequences flanking the integration site were amplified by PCR from primers derived from the tobacco genome, due to the scarcity of carrot plastid DNA sequences in the public databases. Despite the potential of plastid genetic engineering, this technology has only recently been extended to a few major crops, including soybean [33], carrot [32] and cotton [34], via somatic embryogenesis, achieving transgene expression initially via non-green plastids [28]. Most previous studies focused on direct organogenesis by bombardment of leaves containing mature green plastids [28].

Although overall gene content and order are highly conserved among land plants, this same conservation is not observed in non-coding sequences such as introns and intergenic spacers (IGS), which along with the untranslated regions of genes (UTRs), comprise about 50 % of the plastome [35-38]. Genes for input traits such as insect [14,39] and herbicide resistance [23], salt [32] and drought tolerance [15] and pathogen resistance [40] as well as output traits such as the production of therapeutic proteins [19,41-43] are targeted to IGS regions to avoid disruption of endogenous genes. Integration of foreign sequences is dependent on homologous recombination between the transformation vector and the plastid genome. It is possible to achieve integration without 100% sequence identity between the vector and plastid genome sequence but recombination and hence transformation efficiency is impaired when sequences are divergent [37,44]. Additionally, evaluations of UTRs from a variety of species indicate the need to employ species-specific regulatory elements, such as promoters and translation sequences, to elevate the level of foreign protein expression [45,46].

Completely sequenced plastid genomes also provide a valuable source of phylogenetic data for resolving relationships among angiosperms [35,47-50]. The use of DNA sequences from shared plastid genes provides many more characters for phylogeny reconstruction relative to previous molecular phylogenies based on one to several



**Figure 1**

Map of the *Daucus carota* plastid 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 ticks around the map indicate the location of repeated sequences found in the carrot genome; black = direct, blue = palindrome; \* indicates that repeated sequence begins at the same position (see Table 2 for details).

**Table 1: Intron-containing genes found in the carrot plastome**

Protein coding	Transfer RNA
<i>rps16</i>	<i>trn-Lys</i> (UUU)
<i>atpF</i>	<i>trn-Gly</i> (GCC)
<i>rpoC1</i>	<i>trn-Leu</i> (UAA)
<i>yf3+</i>	<i>trn-Val</i> (UAC)
<i>clpP+</i>	<i>trn-Ala</i> (UGC)*
<i>petB</i>	<i>trn-Ile</i> (GAU)*
<i>petD</i>	
<i>rpl16</i>	
<i>rpl2*</i>	
<i>ndhB*</i>	
<i>rps12*+</i>	
<i>ndhA</i>	

\* genes are located in the IR; + genes contain two introns

plastid genes. However, the use of complete plastid genome sequences is constrained because of limited taxon sampling, a phenomenon that can often lead to incorrect tree topologies [e.g., [35,49,51-54]]. Thus, there is an increased need to expand taxon sampling of complete plastid genomes to overcome this problem. Currently there are 35 published plastid genome sequences of angiosperms [37,55]. Some major lineages have multiple genome sequences available, especially basal angiosperms, monocots, and rosids, whereas other major clades are represented by only one or two taxa. The euasterid II clade represents one lineage that is undersampled. This group, comprising four major subclades with approximately 35 families and 32,000 species [56], has only one published genome sequence from *Panax* [57,58].

In this paper, we report on complete plastid genome sequence of *Daucus carota*, the first sequenced member of the family Apiaceae. We describe the organization of this genome and we present a phylogenetic analysis of *Daucus* and 29 other angiosperm plastid genomes based on 61 shared protein-coding genes. This is only the second published plastid genome sequence of the species-rich euasterid II clade. The complete plastid genome sequence of *Daucus* also provides valuable information for the application of plastid genetic engineering to this economically important crop plant [46].

**Results**

**Size, gene content, order and organization of the carrot plastid genome**

The complete *Daucus carota* plastid genome is 155,911 base pairs (bp) in length (Fig. 1). The inverted repeat is 27,051 bp and the two copies are separated by two single copy regions; the large single copy region is 84,242 bp long and the small single copy region is 17,567 bp. There are a total of 136 predicted coding regions, 115 of which are unique and 21 are duplicated in the IR. On LSC/IRb boundary, the IR extends into *rps19*, resulting in the duplication of a portion of this gene. There are 81 unique protein-coding genes, 10 of which are duplicated in the IR. Also in the IR region is the ribosomal operon, which includes all 4 rRNA genes as well as tRNA-Ile and tRNA-Ala. There are five additional tRNAs within the IR resulting in a total of 37 tRNA genes, 30 of which are unique. There are 18 genes containing introns, with 15 of these with only a single intron (Table 1). Non-coding sequences, including IGS regions and introns, comprise

**Table 2: Repeats identified in the carrot plastid genome**

Repeat #	Length	Location 1	Location 2
Direct			
1*	70	<i>ycf2</i> (IRa,b)	<i>ycf2</i> (IRa,b)
2	42	<i>ndhF:rpl32</i>	<i>ndhF:rpl32</i>
3*	34	<i>ycf2</i> (IRa,b)	<i>ycf2</i> (IRa,b)
4*	42	<i>ycf3</i> intron 2	<i>rps12_3end:orf80</i> (IRa,b)
5*	41	<i>rps12_3end:orf80</i> (IRa,b)	<i>ndhA</i> intron
6	34	<i>ndhC:trnV-UAC</i>	<i>ndhC:trnV-UAC</i>
7	33	<i>trnS-GCU</i>	<i>trnS-UGA</i>
8*	31	<i>rrn5:trnR-ACG</i> (IRa,b)	<i>trnR-ACG:rrn5</i> (IRa,b)
9	30	<i>trnG-GCC</i>	<i>trnG-GCC</i>
10	30	<i>psaB</i>	<i>psaA</i>
11*	30	<i>ycf2</i> (IRa,b)	<i>ycf2</i> (IRa,b)
12*	30	<i>ycf2</i> (IRa,b)	<i>ycf2</i> (IRa,b)
Palindrome			
13	30	<i>psaj:rpl33</i>	<i>psaj:rpl33</i>
14	30	<i>ccsA:ndhD</i>	<i>ccsA:ndhD</i>

Repeats at least 30 bp in size, with a sequence identity greater or equal than 90% are included.

Repeats marked with \* are present in the IR and are therefore duplicated.

See Fig. 1 for location of repeats on the gene map.

**Table 3: Taxa included in phylogenetic analyses with GenBank accession numbers and references**

Taxon	GenBank Accession Numbers	Reference
Gymnosperm Outgroups		
<i>Pinus thunbergii</i>	<a href="#">NC 001631</a>	Wakasugi <i>et al.</i> 1994 [95]
<i>Ginkgo biloba</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
Basal Angiosperms		
<i>Amborella trichopoda</i>	<a href="#">NC 005086</a>	Goremykin <i>et al.</i> 2003 [47]
<i>Nuphar advena</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
<i>Nymphaea alba</i>	<a href="#">NC 006050</a>	Goremykin <i>et al.</i> 2004 [48]
Magnoliids		
<i>Calycanthus floridus</i>	<a href="#">NC 004993</a>	Goremykin <i>et al.</i> 2003 [76]
Monocots		
<i>Acorus americanus</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
<i>Oryza sativa</i>	<a href="#">NC 001320</a>	Hiratsuka <i>et al.</i> 1989 [96]
<i>Saccharum officinarum</i>	<a href="#">NC 006084</a>	Asano <i>et al.</i> 2004 [97]
<i>Triticum aestivum</i>	<a href="#">NC 002762</a>	Ikeo and Oghihara, unpublished
<i>Typha latifolia</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
<i>Yucca schidigera</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
<i>Zea mays</i>	<a href="#">NC 001666</a>	Maier <i>et al.</i> 1995 [98]
Eudicots		
<i>Arabidopsis thaliana</i>	<a href="#">NC 000932</a>	Sato <i>et al.</i> 1999 [99]
<i>Atropa belladonna</i>	<a href="#">NC 004561</a>	Schmitz-Linneweber <i>et al.</i> 2002 [100]
<i>Citrus sinensis</i>	<a href="#">DQ864733</a>	Bausher <i>et al.</i> unpublished
<i>Cucumis sativus</i>	<a href="#">NC 007144</a>	Plader <i>et al.</i> unpublished
<i>Daucus carota</i>	<a href="#">DQ898156</a>	Current study
<i>Eucalyptus globulus</i>	<a href="#">AY780259</a>	Steane 2005 [101]
<i>Glycine max</i>	<a href="#">NC 007942</a>	Saski <i>et al.</i> 2005 [78]
<i>Gossypium hirsutum</i>	<a href="#">NC 007944</a>	Lee <i>et al.</i> [36]
<i>Lotus corniculatus</i>	<a href="#">NC 002694</a>	Kato <i>et al.</i> 2000 [75]
<i>Medicago truncatula</i>	<a href="#">NC 003119</a>	Lin <i>et al.</i> , unpublished
<i>Nicotiana tabacum</i>	<a href="#">NC 001879</a>	Shinozaki <i>et al.</i> 1986 [102]
<i>Oenothera elata</i>	<a href="#">NC 002693</a>	Hupfer <i>et al.</i> 2000 [77]
<i>Panax schinseng</i>	<a href="#">NC 006290</a>	Kim and Lee 2004 [58]
<i>Ranunculus macranthus</i>	<a href="#">DQ069337–DQ069702</a>	Leebens-Mack <i>et al.</i> 2005 [49]
<i>Solanum lycopersicum</i>	<a href="#">DQ347959</a>	Daniell <i>et al.</i> [37]
<i>Solanum bulbocastanum</i>	<a href="#">NC 007943</a>	Daniell <i>et al.</i> [37]
<i>Spinacia oleracea</i>	<a href="#">NC 002202</a>	Schmitz-Linneweber <i>et al.</i> 2001 [103]
<i>Vitis vinifera</i>	<a href="#">NC 007957</a>	Jansen <i>et al.</i> 2006 [35]

43.61 % of the carrot plastome. The overall nucleotide composition is 62.34 % AT and 37.66 %.

### Repeat analysis

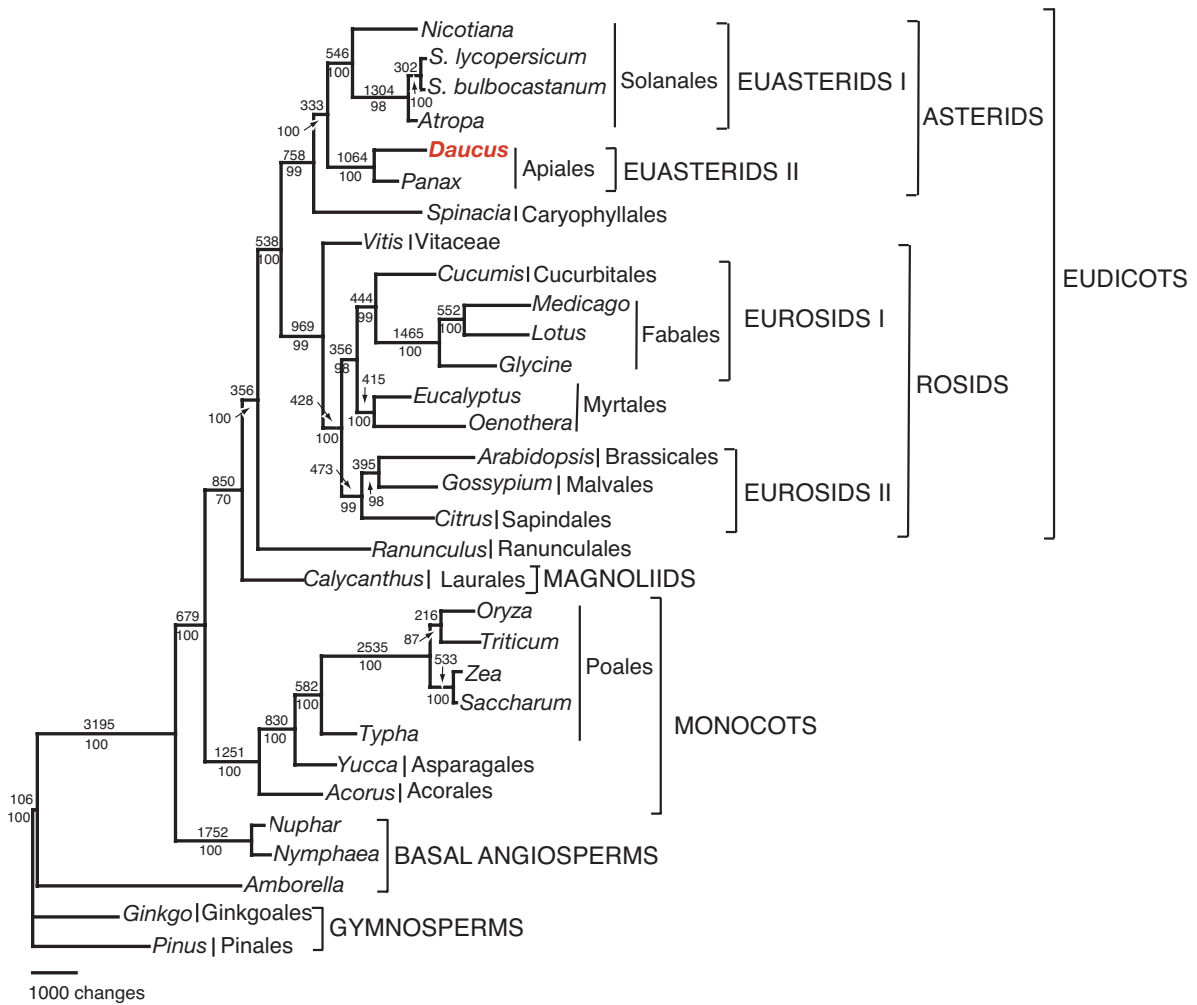
Repeat analysis identified 12 direct repeats and 2 palindromes of  $\geq 30$  bp with a sequence identity of  $\geq 90$  % (Hamming distance of 3). Repeated sequences were found in IGS regions, introns and within coding sequence (Table 2). There are 4 direct repeats in *ycf2*, with repeated sequences ranging up to 70 bp in length.

### Phylogenetic analysis

Our phylogenetic data set included 61 protein-coding genes for 31 taxa (Table 3), including 29 angiosperms and two gymnosperm outgroups (*Pinus* and *Ginkgo*). The data set comprised 45,582 nucleotide positions but when the gaps were excluded to avoid regions with ambiguous

alignment due to length variation there were 39,490 characters.

Maximum Parsimony (MP) analyses resulted in a single, fully resolved tree with a length of 54,140, a consistency index of 0.44 (excluding uninformative characters) and a retention index of 0.60 (Fig. 2). Bootstrap analyses indicated that 26 of the 28 nodes were supported by values  $\geq 95$ % and 19 of these had a bootstrap value of 100%. Maximum likelihood (ML) analysis resulted in a single tree with  $-\ln L = 312205.340$ . ML bootstrap values also were also high, with values of  $\geq 95$ % for 24 of the 28 nodes and 22 nodes with 100% bootstrap support. The ML and MP trees had very similar topologies but differed in three places. (1) The MP tree placed *Amborella* as the sister group to all other angiosperms, whereas the ML tree placed *Amborella* sister to the Nymphaeales, and together



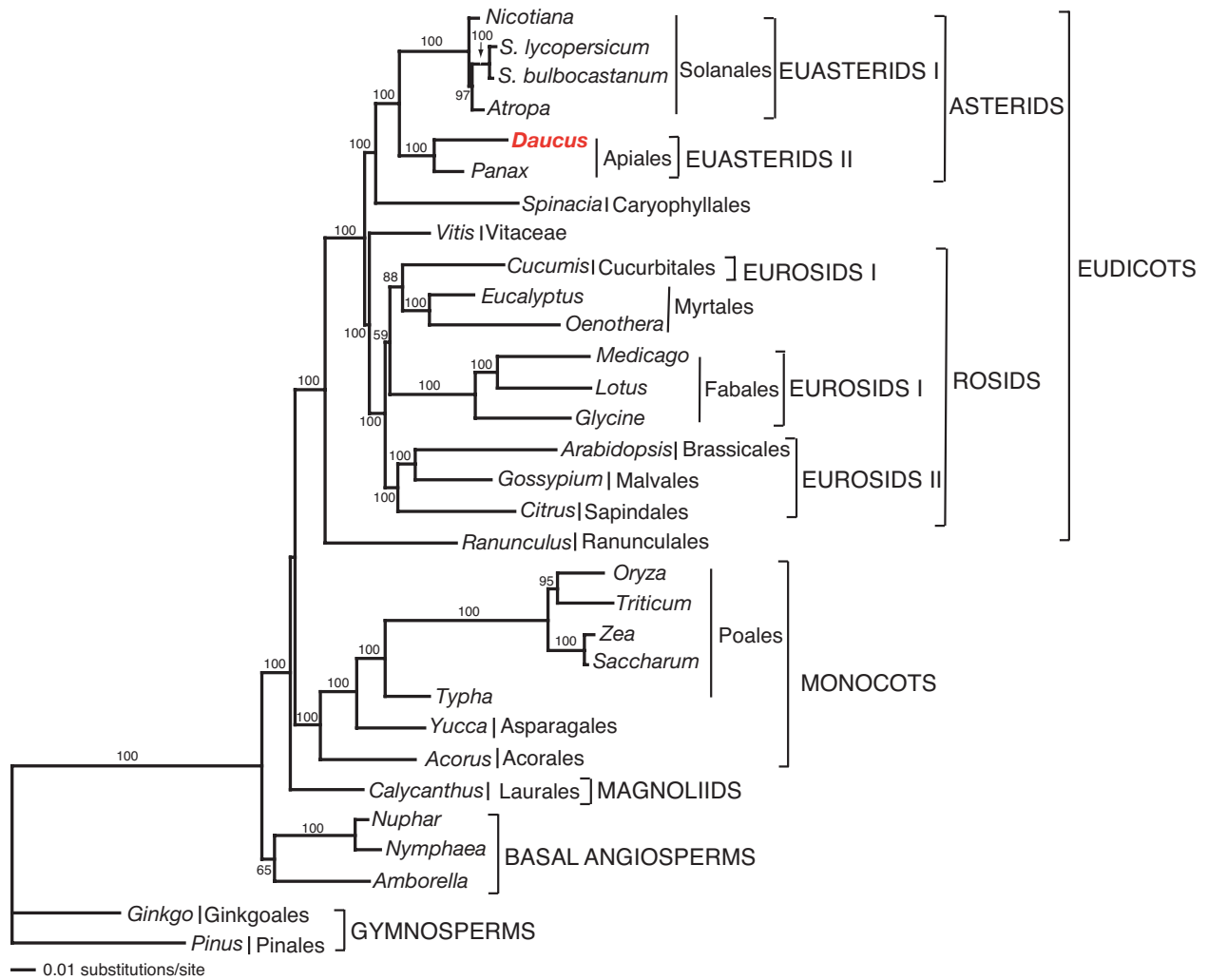
**Figure 2**

Phylogenetic tree of 31-taxon data set based on 61 plastid protein-coding genes using maximum parsimony. The tree has a length of 54,140, a consistency index of 0.44 (excluding uninformative characters) and a retention index of 0.60. Numbers above node indicate number of changes along each branch and numbers below nodes are bootstrap support values. Ordinal and higher level group names follow APG II [104]. Taxon in red is *Daucus*, the new genome reported in this paper.

this group formed the sister group of all other angiosperms. Support for the relationships of basal angiosperms in the MP tree is strong (100%) but only moderate in the ML tree (65%). (2) The MP tree placed *Calycanthus* sister to the eudicots, whereas the ML tree positioned *Calycanthus* as sister to a large clade that included both monocots and eudicots. Support for the different placements of *Calycanthus* was weak in both MP and ML analyses. (3) Relationships among the rosids differed, especially the position of *Cucumis* and the monophyly of eurosids I. The MP tree (Fig. 2) provides strong support for the monophyly of the eurosid I clade because

*Cucumis* is sister to the three legume taxa. In contrast, the ML tree (Fig. 3) places *Cucumis* sister to the two examined taxa of Myrtales, and support for this relationship is not as strong (88% bootstrap value). These three differences were detected in recent phylogenies based on complete plastid genome sequences of basal angiosperms [49] and rosids [35]. The remaining angiosperms formed two major clades, one including monocots and a second including the eudicots (Figs. 2, 3). Monophyly of the monocots was strongly supported (100% bootstrap value for both MP and ML). Ranunculales were strongly supported as sister to the remaining eudicots. There were two





**Figure 3**

Phylogenetic tree of 31-taxon data set based on 61 plastid protein-coding genes using maximum likelihood. The tree has a ML value of  $-\ln L = 312205.340$ . Numbers at nodes are bootstrap support values  $\geq 50\%$ . Ordinal and higher level group names follow APG II [104]. Taxon in red is *Daucus*, the new genome reported in this paper.

major clades of core eudicots, one including the rosids and the second including the Caryophyllales + asterids. Both MP and ML trees provide very strong support (100% bootstrap) for the sister relationship of *Daucus* with *Panax* in the euasterid II clade.

**Discussion**

**Implications for plastid genetic engineering**

An important agricultural crop worldwide, carrot has long attracted attention from the research community. It was the first crop species in which somatic embryogenesis was demonstrated [59]. This ability, to regenerate entire plants from cell or tissue cultures, has helped to maintain inter-

est in carrot as our technology has advanced to include the improvement of agronomic species via genetic manipulation. The carrot nucleus has been the recipient of foreign genes to confer input characteristics such as pathogen resistance [60] and herbicide resistance [61]. Recently an extensive analysis of four *Agrobacterium rhizogenes* strains and twelve *Daucus carota* genotypes examined the utility of green fluorescent protein (GFP) as a selectable marker for nuclear transformants [62] demonstrating continued interest in this system for expression of foreign proteins.

Most interesting has been the exploration of carrot as an ideal platform for the production of proteins of signifi-

cance to pharmacology. The small isoform of human glutamic acid decarboxylase (GAD65) has been identified as a major autoantigen contributing to the onset of insulin-dependent diabetes mellitus (IDDM) [63]. Expression of the GAD65 cDNA in transgenic carrot and tobacco resulted in an immunoreactive product which retained appropriate enzymatic function. Unfortunately levels of expression were quite low for both tobacco leaves and carrot taproots, on the order of 0.040 % and 0.012 % of total soluble protein (TSP), respectively [64]. A heterologous version of human GAD65 having the N-terminus substituted with GAD67 from rat was expressed in tobacco and was able to achieve stable accumulation of functional immunoreactive product up to 0.19% of the TSP. Although oral administration of disease associated autoantigens such as GAD65 can lead to the induction of tolerance in the murine model, dosage on the order of milligrams per week per mouse are required [65]. Expression levels and stable accumulation will have to be improved by orders of magnitude to make oral dosage truly feasible.

Vaccine antigens have been expressed in a number of plant species [66-69], including carrot. Transformation experiments have introduced a hemagglutinin glycoprotein [70] and a novel chimeric polyepitope antigen [71], both for neutralizing immunization against measles virus into the carrot nuclear genome. Extracted proteins demonstrated immunogenicity, raising antibody titers in sera of injected mice. However, quantitative data on protein accumulation in transgenic plants is lacking, especially for taproots. An estimate based on ELISA of 2% of membrane fraction in crude membrane preparations from carrot leaf is offered, but no mention is made of protein content in the root [70]. It is noteworthy that these extracts were homogenized with Freund's adjuvant (1:1) prior to initial injection and at each boost. Freund's adjuvant is employed to enhance antibody formation suggesting that plant extracts from nuclear transformants were insufficient to induce the desired immune response.

The need for an alternative expression system to obtain high levels of protein accumulation in carrot roots becomes apparent if efficacious oral delivery is to be accomplished. Furthermore a system for oral delivery of antigens should ideally include the adjuvant, limiting further the need for post harvest processing. Transformation of the plant plastid has demonstrated the capacity to produce substantial quantities of functional and immunoreactive proteins [reviewed in [29]].

Stable transformation of the *Daucus carota* plastid genome via somatic embryogenesis has been demonstrated recently [32]. Integration of foreign genes in these experiments was accomplished through the use of flanking

sequences that were PCR-amplified from the native carrot plastid ribosomal operon, whereas the regulatory sequences used to facilitate expression of the transgenes were derived from tobacco and bacteriophage T7. When assayed for BADH enzyme function, roots of carrot plastid transformants showed activity up to 74.8 % of leaf tissue. In root tissues, plastome copy number is generally about 5 % of the level in mature leaves. The notably high activity is probably due to the elevated concentration of root chromoplasts in carrot, the plastid type responsible for the orange coloration. With the availability of the entire carrot plastid genome sequence, it will now be possible to incorporate native translation regulatory sequences into transformation constructs to further enhance foreign protein accumulation in carrot plastids. Additionally, detailed knowledge of this genome will allow the identification of optimal intergenic spacer regions for the integration of transgenes.

Receptor-mediated translocation of antigens and other pharmaceutical proteins across the intestinal mucosa offers the potential to make plant-produced, orally delivered vaccines and therapeutics a reality. The toxin of *Vibrio cholerae* (CT) is recognized as one of the most potent mucosal adjuvants. The holotoxin is composed of the A subunit, responsible for toxicity, and the non-toxic homopentameric B subunit (CTB), which facilitates entry into epithelial cells of the intestine by binding the GM1 receptor followed by endocytosis. Recombinant forms excluding the A subunit are rendered non-toxic, and when fused to another antigen, the B subunit cannot only carry this antigen across the intestine, but also strongly potentiate the antigen's immunogenicity [72,73]. Recently a fusion construct of CTB and GFP expressed in transplastomic tobacco demonstrated the efficacy of CTB to deliver foreign proteins to the circulatory system of mice, which were fed pulverized leaf tissue from plastid transformants. Between the two protein sequences investigators included the cleavage site for the ubiquitous protease furin to facilitate the intracellular cleavage of GFP. Quantitative ELISA revealed accumulation of CTB-GFP in transgenic plants ranging from 19.09 to 21.3% of TSP. Following oral administration of CTB-GFP expressing leaf material to mice, fluorescence microscopy and immunohistochemical analyses confirmed the presence of GFP in the mouse intestinal mucosa, liver and spleen while CTB remained in the intestinal cells [74]. Remarkable levels of protein accumulation coupled to a receptor-mediated oral delivery mechanism offers realistic hope for the possibility of plant-derived, orally delivered therapeutic proteins.

#### **Genome organization and evolution**

The *Daucus* genome with two copies of an IR separating the SSC and LSC regions is identical in architecture to most sequenced angiosperm plastid genomes [2]. The size

of the genome at 155,911 bp is also within the known range for angiosperms, which generally vary from 150,519 [75] to 162,686 bp [47] for taxa that have both copies of the IR. The size of the *Daucus* IR at 27,051 bp is at the upper end of the size range of other sequenced genomes, which vary from 23,302 (*Calycanthus*) [76] to 27,807 (*Oenothera*) bp [77]. Gene content and order of the *Daucus* plastid genome are identical to *Panax* [58], the only other published euasterid II genome.

A number of recent comparisons of plastid genomes of angiosperms have identified dispersed direct and inverted repeats [35-37,78]. The carrot genome contains similar numbers and sizes of repeats to these other angiosperms (Table 2, Fig. 1). In most cases, these repeats are located in intergenic spacer regions and in introns but several also occur in tRNAs and protein-coding genes. Examination of repeats in highly rearranged algal and angiosperm genomes have demonstrated a correlation between both the number and location of the repeats and the propensity for rearrangements [79,80]. The role of dispersed repeats in unrearranged plastid genomes remains unknown.

#### Phylogenetic implications

The phylogenies based on 61 protein-coding plastid genes for 29 angiosperms (Figs. 2, 3) are largely congruent with relationships suggested by previous studies based single and multiple genes [56] and a number of recent phylogenies based on complete plastid genome sequences [35,36,48-50,76,81]. There is strong support for the monophyly the major clades of angiosperms, including monocots, eudicots, rosids, asterids, eurosids II, asterids I and asterid II. The three areas of incongruence between the MP and ML trees regarding relationships of basal angiosperms, *Calycanthus*, and eurosids I were identified previously [35,49], and are likely due to limited taxon sampling and misspecification of model parameters in large concatenated, multigene data sets.

The position of Caryophyllales within angiosperms has been controversial in the past. Previous molecular phylogenies clearly indicated that this order is a member of the eudicot clade [56], but relationships of Caryophyllales to other major eudicot lineages remains uncertain. The order has been suggested to be allied to rosids, asterids, or simply as an unresolved major eudicot clade sister to the Dilleniaceae [82]. Two recent phylogenies based on 61 shared plastid gene sequences provided support for a sister relationship between the Caryophyllales and asterids [35,36], however, only a single representative of the euasterid II clade was included. The addition of the *Daucus* plastid genome to this data set increases the level of support for a sister relationship of asterids and Caryophyllales (Figs. 2, 3).

Finally, the multiple plastid gene phylogenies also provide strong support for the monophyly of the euasterid II clade (Figs. 2, 3). This result is not surprising given that the Araliaceae (*Panax*) and Apiaceae (*Daucus*) have been considered sister families for a long time based on both morphological and molecular data [56]. Expanded taxon sampling of the other three major clades of euasterids II is needed to further test the monophyly and relationships of this large, diverse angiosperm clade.

#### Conclusion

This is the first sequenced plastid genome of the family Apiaceae and only the second published genome sequence of the species-rich euasterid II clade. Both MP and ML trees provide very strong support (100% bootstrap) for the sister relationship of *Daucus* with *Panax* in the euasterid II clade. These results provide the best taxon sampling of complete chloroplast genomes and the strongest support yet for the sister relationship of Caryophyllales to the asterids.

The availability of the complete plastid genome sequence should facilitate improved transformation efficiency and foreign gene expression in carrot through utilization of endogenous flanking sequences and regulatory elements. The ability to express high levels of foreign protein, particularly those of clinical interest, makes plant plastids an attractive target for biotechnology. As a biennial crop which is amenable to relatively long term storage, carrot taproots may provide a feasible platform for the production of pharmaceutical proteins for oral delivery. The availability of the complete plastid genome sequence should facilitate improved transformation efficiency and foreign gene expression in carrot through utilization of endogenous flanking sequences and regulatory elements.

#### Methods

##### DNA isolation and amplification

*Daucus carota* L. cv half long plants were purchased fresh from a local market with leaves intact. Leaf tissue (10 g) was collected for plastid isolation based on the sucrose step gradient centrifugation method of Palmer [83]. Isolation was followed by whole plastid genome Rolling Circle Amplification (RCA) using the Repli-g RCA kit (Qiagen, Inc.) following the methods outlined in [84]. 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 amplification of the plastome, as well as assessment of its quality prior to genome 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 1231 high quality reads

with an average length of 808 bases was generated during the random (1126 reads) and closure (105 reads) phases of sequencing. Sequences were assembled using TIGR assembler [85] and scaffolded using Bambus [86]. Sequence finishing included directed PCR to span gaps and directed primer walking of clones to cover the entire genome and to complete regions of low depth of coverage.

#### **Annotation and analysis of repeat structure**

The *Daucus carota* plastid genome was annotated using DOGMA [87], which performs BLASTX searches against a custom database of previously published plastid genomes to identify putative coding sequences. The user submits a FASTA-formatted input file of the complete plastid genome sequence for analysis. DOGMA identified putative start and stop codons, which must then be confirmed by the user for each putative protein-coding gene. Identification of intron and exon boundaries and tRNAs and rRNAs must also be confirmed. The fully annotated plastid genome of *Daucus carota* was submitted to NCBI GenBank with the following accession number [GenBank: [DQ898156](#)].

Analysis of repeat structure was carried out using Comparative Repeat Analysis (CRA) [88]. The settings for identifying direct and inverted (palindromic) repeats included a size range between 30–5000 bp and a Hamming distance of 3 (limiting hits to sequence identity of  $\geq 90\%$ ).

#### **Phylogenetic analysis**

The 61 genes included in the analyses of Goremykin *et al.* [47], Leebens-Mack *et al.* [49], Lee *et al.* [36], and Jansen *et al.* [35] were extracted from the plastid genome sequence of *Daucus* using the DOGMA [87]. The same set of 61 genes was extracted from plastid genome sequences of 30 other sequenced plastid genomes (see Table 3 for complete list of genomes examined). All 61 protein-coding genes of the 31 taxa were translated into amino acid sequences, aligned using MUSCLE [89] followed by manual adjustments, and then nucleotide sequences of these genes were aligned by constraining them to the aligned amino acid sequences. A Nexus file with character sets for phylogenetic analyses was generated after nucleotide sequence alignment was completed. The complete nucleotide alignment is available online at Chloroplast Genome Database [90].

Phylogenetic analyses using maximum parsimony (MP) and maximum likelihood (ML) were performed with PAUP\* version 4.10b10 [91]. Phylogenetic analyses excluded gap regions to avoid alignment ambiguities in regions with variation in sequence lengths. All MP searches included 100 random addition replicates and TBR branch swapping with the Multrees option. Model-

test 3.7 [92] was used to determine the most appropriate model of DNA sequence evolution for the combined 61-gene dataset. Hierarchical likelihood ratio tests and the Akaike information criterion were used to assess which of the 56 models best fit the data, which was determined to be GTR + I +  $\Gamma$  by both criteria. For ML analyses we performed an initial parsimony search with 100 random addition sequence replicates and TBR branch swapping, which resulted in a single tree. Model parameters were optimized onto the parsimony tree. We fixed these parameters and performed a ML analysis with three random addition sequence replicates and TBR branch swapping. The resulting ML tree was used to re-optimize model parameters, which then were fixed for another ML search with three random addition sequence replicates and TBR branch swapping. This successive approximation procedure was repeated until the same tree topology and model parameters were recovered in multiple, consecutive iterations. This tree was accepted as the final ML tree (Fig. 3). Successive approximation has been shown to perform as well as full-optimization analyses for a number of empirical and simulated datasets [93]. Non-parametric bootstrap analyses [94] were performed for MP analyses with 1000 replicates with TBR branch swapping, 1 random addition replicate, and the Multrees option and for ML analyses with 100 replicates with NNI branch swapping, 1 random addition replicate, and the Multrees option.

#### **Abbreviations**

cpDNA, plastid DNA; IR inverted repeat; SSC, small single copy; LSC, large single copy, bp, base pair; plastome, plastid genome; ycf, hypothetical chloroplast reading frame; rrn, ribosomal RNA; MP, maximum parsimony; ML, maximum likelihood.

#### **Authors' contributions**

SBL isolated plastids, performed RCA amplification of cpDNA, genome annotation, analysis and submission of data to the GenBank; TR performed the repeat analyses, drew the genome map and wrote some 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 and NIH R01 GM 63879 to Henry Daniell and from NSF DEB 0120709 to Robert K. Jansen. We thank Zhenggui Cai for assistance with alignment of amino acid and nucleotide sequences.

## References

- 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.
- 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.
- Sugiura M: **The chloroplast genome.** *Plant Mol Biol* 1992, **19**:149.
- Colas des Francs-Small C, Szurek B, Small I: **Proteomics, bioinformatics and genomics applied to plant organelles.** In *Molecular Biology and Biotechnology of Plant Organelles: Chloroplast and Mitochondria* Edited by: Daniell H, Chase C. Dordrecht, The Netherlands: Springer; 2004:179.
- Richly E, Leister D: **NUMTs in sequenced eukaryotic genomes.** *Mol Biol Evol* 2004, **21**:1081.
- Shimada H, Sugiura M: **Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes.** *Nucl Acids Res* 1991, **19**:983.
- Zerges W: **Translation in chloroplasts.** *Biochimie* 2000, **82**:583.
- Palmer JD: **Chloroplast DNA exists in two orientations.** *Nature* 1983, **301**:92-93.
- Kolodner R, Tewari KK, Warner RC: **Physical studies on the size and structure of the covalently closed circular chloroplast DNA from higher plants.** *Biochim Biophys Acta* 1976, **447**:144.
- Kolodner RD, Tewari KK: **Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism.** *Nature* 1975, **256**:708.
- Quesada-Vargas T, Ruiz ON, Daniell H: **Characterization of heterologous multigene operons in transgenic chloroplasts: transcription, processing, and translation.** *Plant Physiol* 2005, **138**:1746-1762.
- Ruiz ON, Hussein HS, Terry N, Daniell H: **Phytoremediation of organomercurial compounds via chloroplast genetic engineering.** *Plant Physiol* 2003, **132**:1344.
- 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.** *Plant Cell Rep* 2003, **21**:891.
- De Cosa 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, **19**:71-74.
- Lee SB, Kwon HB, Kwon SJ, Park SC, Jeong MJ, Han SE, Byun MO, Daniell H: **Accumulation of trehalose within transgenic chloroplasts confers drought tolerance.** *Mol Breed* 2003, **11**:1.
- Dhingra A, Portis AR 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.
- Daniell H, Khan MS, Allison L: **Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology.** *Trends Plant Sci* 2002, **7**:84.
- 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.
- Daniell H, Lee SB, Panchal T, Wiebe PO: **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.
- Birky CW Jr: **The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models.** *Annu Rev Genet* 2001, **35**:125.
- Hagemann R: **The Sexual Inheritance of Plant Organelles.** In *Molecular Biology and Biotechnology of Plant Organelles: Chloroplast and Mitochondria* Edited by: Daniell H, Chase C. Dordrecht, The Netherlands: Springer; 2004:93.
- Daniell H: **Molecular strategies for gene containment in transgenic crops.** *Nat Biotechnol* 2002, **20**:581-586.
- 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.
- Scott SE, Wilkinson MJ: **Low probability of chloroplast movement from oilseed rape (*Brassica napus*) into wild *Brassica rapa*.** *Nat Biotechnol* 1999, **17**:390-392.
- Ruiz ON, Daniell H: **Engineering cytoplasmic male sterility via the chloroplast genome by expression of  $\beta$ -ketothiolase.** *Plant Physiol* 2005, **138**:1232-1246.
- Bogorad L: **Engineering chloroplasts: an alternative site for foreign genes, proteins, reactions and products.** *Trends Biotechnol* 2000, **18**:257-263.
- Grevich JJ, Daniell H: **Chloroplast genetic engineering: recent advances and future perspectives.** *Crit Rev Plant Sci* 2005, **24**:83-107.
- Daniell H, Kumar S, Dufourmantel N: **Breakthrough in chloroplast genetic engineering of agronomically important crops.** *Trends Biotechnol* 2005, **23**:238.
- Daniell H, Chebolu S, Kumar S, Singleton M, Falconer R: **Chloroplast-derived vaccine antigens and other therapeutic proteins.** *Vaccine* 2005, **23**:1779-1783.
- Yan WK, Hunt LA: **Reanalysis of vernalization data of wheat and carrot.** *Ann Bot* 1999, **84**:615-619.
- Handbook of Energy Crops** [[http://www.hort.purdue.edu/newcrop/duke\\_energy/Daucus\\_carota.html](http://www.hort.purdue.edu/newcrop/duke_energy/Daucus_carota.html)]
- 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* 2004, **136**:2843.
- Dufourmantel N, Pelissier B, Garcon F, Peltier G, Ferullo JM, Tissot G: **Generation of fertile transplastomic soybean.** *Plant Mol Biol* 2004, **55**:479-489.
- Kumar S, Dhingra A, Daniell H: **Stable transformation of the cotton plastid genome and maternal inheritance of transgenes.** *Plant Mol Biol* 2004, **56**:203-216.
- Jansen RK, Kaitanis C, Lee SB, Saski C, Tomkins 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.
- Lee SB, Kaitanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H: **The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms.** *BMC Genomics* 2006, **7**:61.
- Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK: **Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes.** *Theor Appl Genet* 2006, **112**:1503-1518.
- Kim JS, Jung JD, Lee JA, Park HW, Oh KH, Jeong WJ, Choi DW, Liu JR, Cho KY: **Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome.** *Plant Cell Rep* 2006, **25**:334-340.
- Kota M, Daniell H, Varma S, Garczynski SF, Gould F, Moar WJ: **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.
- 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 Physiol* 2001, **127**:852-862.
- Daniell H, Streatfield SJ, Wycoff K: **Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants.** *Trends Plant Sci* 2001, **6**:219-226.
- 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.** *Plant Biotechnol J* 2003, **1**:71.
- 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.
- Zubko MK, Zubko EI, Zuilen Kv, Meyer P, Day A: **Stable transformation of petunia plastids.** *Trans Res* 2004, **13**:523.
- Kramzar LM, Mueller T, Erickson B, Higgs DC: **Regulatory sequences of orthologous *petD* chloroplast mRNAs are highly specific among *Chlamydomonas* species.** *Plant Mol Biol* 2006, **60**:405-422.
- Daniell H, Ruiz ON, Dhingra A: **Chloroplast genetic engineering to improve agronomic traits.** *Meth Mol Biol* 2005, **286**:111.
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH: **Analysis of the *Amborella trichopoda* chloroplast genome sequence sug-**

- gests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 2003, **20**:1499-1505.
48. Goremykin VV, Hirsch-Ernst KI, Wolff 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.
  49. Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW: **Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone.** *Mol Biol Evol* 2005, **22**:1948-1963.
  50. 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.
  51. Soltis DE, Soltis PS: ***Amborella* not a "basal angiosperm"? Not so fast.** *Am J Bot* 2004, **91**:997-1001.
  52. Stefanovic S, Rice DW, Palmer JD: **Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots?** *BMC Evol Biol* 2004, **4**:35.
  53. 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.
  54. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, Chase MW, Farris JS, Stefanovic 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.
  55. [<http://megasun.bch.umontreal.ca/ogmp/projects/other/cp.list.html>].
  56. Soltis DE, Soltis PS, Endress PK, Chase MW: *Phylogeny and Evolution of Angiosperms* Sunderland, MA: Sinauer Associates Inc; 2005.
  57. Kim KJ, Lee HL: **Widespread occurrence of small inversions in the chloroplast genomes of land plants.** *Mol Cells* 2005, **19**:104-113.
  58. Kim KJ, Lee HL: **Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants.** *DNA Res* 2004, **11**:247-261.
  59. Steward FC, Mapes M, Mears K: **Growth and organized development of cultured cells: II. Organization in cultures grown from freely suspended cells.** *Am J Bot* 1958, **45**:705-708.
  60. Takaichi M, Oeda K: **Transgenic carrots with enhanced resistance against two major pathogens, *Erysiphe heraclei* and *Alternaria dauci*.** *Plant Sci* 2000, **153**:135-144.
  61. Aviv D, Amsellem Z, Gressel J: **Transformation of carrots with mutant acetolactate synthase for *Orobanche* (broomrape) control.** *Pest Manag Sci* 2002, **58**:1187-1193.
  62. Baranski R, Klocke E, Schumann G: **Green fluorescent protein as an efficient selection marker for *Agrobacterium rhizogenes* mediated carrot transformation.** *Plant Cell Rep* 2005:1-8.
  63. Sanjeevi CB, Falorni A, Robertson J, Lernmark A: **Glutamic acid decarboxylase (GAD) in insulin-dependent diabetes mellitus.** *Diabet Nutr Metab* 1996, **9**:167-182.
  64. Porceddu A, Falorni A, Ferradini N, Cosentino A, Calcinaro F, Faleri C, Cresti M, Lorenzetti F, Brunetti P, Pezzotti M: **Transgenic plants expressing human glutamic acid decarboxylase (GAD65), a major autoantigen in insulin-dependent diabetes mellitus.** *Mol Breed* 1999, **5**:553.
  65. Arakawa T, Yu J, Chong DK, Hough J, Engen PC, Langridge WH: **A plant-based cholera toxin B subunit-insulin fusion protein protects against the development of autoimmune diabetes.** *Nat Biotechnol* 1998, **16**:934-938.
  66. Ma JK, Barros E, Bock R, Christou P, Dale PJ, Dix PJ, Fischer R, Irwin J, Mahoney R, Pezzotti M, Schillberg S, Sparrow P, Stoger E, Twyman RM: **Molecular farming for new drugs and vaccines. Current perspectives on the production of pharmaceuticals in transgenic plants.** *EMBO Rep* 2005, **6**:593-599.
  67. Ma JK, Chikwamba R, Sparrow P, Fischer R, Mahoney R, Twyman RM: **Plant-derived pharmaceuticals – the road forward.** *Trends Plant Sci* 2005, **10**:580-585.
  68. Chargelegue D, Drake PM, Obregon P, Prada A, Fairweather N, Ma JK: **Highly immunogenic and protective recombinant vaccine candidate expressed in transgenic plants.** *Infect Immun* 2005, **73**:5915-5922.
  69. Ma JK, Drake PM, Chargelegue D, Obregon P, Prada A: **Antibody processing and engineering in plants, and new strategies for vaccine production.** *Vaccine* 2005, **23**:1814-1818.
  70. Marquet-Blouin E, Bouche FB, Steinmetz A, Muller CP: **Neutralizing immunogenicity of transgenic carrot (*Daucus carota* L.)-derived measles virus hemagglutinin.** *Plant Mol Biol* 2003, **51**:459.
  71. Bouche FB, Marquet-Blouin E, Yanagi Y, Steinmetz A, Muller CP: **Neutralising immunogenicity of a polyepitope antigen expressed in a transgenic food plant: a novel antigen to protect against measles.** *Vaccine* 2003, **21**:2065.
  72. Holmgren J, Adamsson J, Anjuere F, Clemens J, Czerkinsky C, Eriksson K, Flach CF, George-Chandy A, Harandi AM, Lebens M, Lehner T, Lindblad M, Nygren E, Raghavan S, Sanchez J, Stanford M, Sun JB, Svennerholm AM, Tengvall S: **Mucosal adjuvants and anti-infection and anti-immunopathology vaccines based on cholera toxin, cholera toxin B subunit and CpG DNA.** *Immunol Lett* 2005, **97**:181-188.
  73. Holmgren J, Czerkinsky C: **Mucosal immunity and vaccines.** *Nat Med* 2005, **11**:S45-53.
  74. Limaye A, Koya V, Samsam M, Daniell H: **Receptor-mediated oral delivery of a bioencapsulated green fluorescent protein expressed in transgenic chloroplasts into the mouse circulatory system.** *FASEB J* 2006, **20**:959-961.
  75. 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.
  76. Goremykin V, Hirsch-Ernst KI, Wolff S, Hellwig FH: **The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis* – structural and phylogenetic analyses.** *Plant Syst Evol* 2003, **242**:199-135.
  77. Hupfer H, Swiatek 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 *euoenothera* plastomes.** *Mol Gen Genet* 2000, **263**:581-585.
  78. Sasaki C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK: **Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes.** *Plant Mol Biol* 2005, **59**:309-322.
  79. Pombert JF, Otis C, Lemieux C, Turmel M: **The chloroplast genome sequence of the green alga *Pseudoclonium akinetum* (Ulvoophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages.** *Mol Biol Evol* 2005, **22**:1903-1918.
  80. Chumley TW, Mower JP, Boore JL, Fourcade HM, Caille PJ, Jansen RK: **The complete chloroplast genome sequence of *Pelargonium x hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants.** *Mol Biol Evol* 2006 in press.
  81. Goremykin VV, Hirsch-Ernst KI, Hellwig FH: **Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications.** *Mol Biol Evol* 2005, **22**:1813-1822.
  82. Soltis DESP, 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.
  83. Palmer JD: **Isolation and structural analysis of chloroplast DNA.** In *Methods in Enzymology Volume 118*. Academic Press; 1986:167-186.
  84. Jansen RK, 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.** In *Methods in Enzymology Volume 395*. Academic Press; 2005:348-384.
  85. Sutton GG, Adams MD, Kerlavage AR: **TIGRAssembler: A new tool for assembling large shotgun sequencing projects.** *Gen Sci Techn* 1995, **1**:9-19.
  86. Pop M, Kosack DS, Salzberg SL: **Hierarchical scaffolding with *Bambus*.** *Genome Res* 2004, **14**:149-159.
  87. Wyman SK, Jansen RK, Boore JL: **Automatic annotation of organellar genomes with DOGMA.** *Bioinformatics* 2004, **20**:3252-3255.
  88. [<http://bugmaster.igi-psf.org/repeats/>].

89. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5**:113.
90. [<http://chloroplast.cbio.psu.edu>].
91. Swofford D: **PAUP\*: Phylogenetic analysis using parsimony (\*and other methods), ver. 4.0.** 2003.
92. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
93. Sullivan JA, Abzo Z, Joyce P, Swofford DL: **Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation.** *Mol Biol Evol* 2005, **22**:1386-1392.
94. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
95. 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 U S A* 1994, **91**:9794-9798.
96. 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.
97. 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.
98. Maier RM, Neckerkmann 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.
99. 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.
100. 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.
101. Steane DA: **Complete nucleotide sequence of the chloroplast genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae).** *DNA Res* 2005, **12**:215-220.
102. 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.
103. 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.** *Plant Mol Biol* 2001, **45**:307.
104. **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](http://www.biomedcentral.com/info/publishing_adv.asp)

