

REVIEW

Open Access



# Chloroplast genomes: diversity, evolution, and applications in genetic engineering

Henry Daniell<sup>1\*</sup>, Choun-Sea Lin<sup>2</sup>, Ming Yu<sup>1</sup> and Wan-Jung Chang<sup>2</sup>

## Abstract

Chloroplasts play a crucial role in sustaining life on earth. The availability of over 800 sequenced chloroplast genomes from a variety of land plants has enhanced our understanding of chloroplast biology, intracellular gene transfer, conservation, diversity, and the genetic basis by which chloroplast transgenes can be engineered to enhance plant agronomic traits or to produce high-value agricultural or biomedical products. In this review, we discuss the impact of chloroplast genome sequences on understanding the origins of economically important cultivated species and changes that have taken place during domestication. We also discuss the potential biotechnological applications of chloroplast genomes.

## Introduction

Chloroplasts are active metabolic centers that sustain life on earth by converting solar energy to carbohydrates through the process of photosynthesis and oxygen release. Although photosynthesis is often recognized as the key function of plastids, they also play vital roles in other aspects of plant physiology and development, including the synthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins and a plethora of metabolites, and the assimilation of sulfur and nitrogen. Metabolites that are synthesized in chloroplasts are important for plant interactions with their environment (responses to heat, drought, salt, light, and so on) and their defense against invading pathogens. So, chloroplasts serve as metabolic centers in cellular reactions to signals and respond via retrograde signaling [1, 2]. The chloroplast genome encodes many key proteins that are involved in photosynthesis and other metabolic processes.

The advent of high-throughput sequencing technologies has facilitated rapid progress in the field of chloroplast genetics and genomics. Since the first chloroplast genome, from tobacco (*Nicotiana tabacum*), was sequenced in 1986 [3], over 800 complete chloroplast genome sequences have been made available in the National Center for Biotechnology Information (NCBI) organelle genome database, including 300 from crop and tree genomes. Insights gained from complete chloroplast genome sequences have enhanced our understanding of plant biology and diversity; chloroplast genomes have made significant contributions to phylogenetic studies of several plant families and to resolving evolutionary relationships within phylogenetic clades. In addition, chloroplast genome sequences have revealed considerable variation within and between plant species in terms of both sequence and structural variation. This information has been especially valuable for our understanding of the climatic adaptation of economically important crops, facilitating the breeding of closely related species and the identification and conservation of valuable traits [4, 5]. Improved understanding of variation among chloroplast genomes has also allowed the identification of specific examples of chloroplast gene transfer to plant nuclear or mitochondrial genomes, which has shed new light on the relationship between these three genomes in plants.

In addition to improving our understanding of plant biology and evolution, chloroplast genomics research has important translational applications, such as conferring protection against biotic or abiotic stress and the development of vaccines and biopharmaceuticals in edible crop plants. Indeed, the first commercial-scale production of a human blood protein in a Current Good Manufacturing Processes (cGMP) facility was published recently [6]. The lack of conservation of intergenic spacer regions, even among chloroplast genomes of closely related plant species, and the species specificity of regulatory sequences have facilitated the development of highly efficient transformation vectors for the integration and expression of foreign genes in chloroplasts. Because

\* Correspondence: hdaniell@upenn.edu

<sup>1</sup>Department of Biochemistry, School of Dental Medicine, University of Pennsylvania, South 40th St, Philadelphia, PA 19104-6030, USA  
Full list of author information is available at the end of the article

the published literature is rarely cross-referenced, this review highlights the impact of chloroplast genomes on various biotechnology applications. In addition to our enhanced understanding of chloroplast biology, we discuss in depth the roles of chloroplast genome sequences in improving our understanding of intracellular gene transfer, conservation, diversity, and the genetic basis by which chloroplast transgenes are engineered to enhance plant agronomic traits or to produce high-value agricultural or biomedical products. In addition, we discuss the impact of chloroplast genome sequences on increasing our understanding of the origins of economically important cultivated species and changes that occurred during domestication.

### Advances in chloroplast genome sequencing technology

One of the important factors in the rapid advancement of the chloroplast genomics field is improvement in sequencing technologies. In studies conducted before the availability of high-throughput methods, isolated chloroplasts were used for the amplification of the entire chloroplast genome by rolling circle amplification [7–12]. An alternative strategy is to screen bacterial artificial chromosome (BAC) or fosmid libraries using chloroplast genome sequences as probes [13–20]; however, these methods are subject to many challenges, including difficulty in constructing good-quality BAC or fosmid libraries, large numbers of PCR reactions, and the possibility of contamination from other organellar DNA [21–32]. The PCR approach is also difficult to apply to species that have no relatives whose chloroplast genomes have been sequenced or those with highly rearranged chloroplast genomes.

The development of next-generation sequencing (NGS) methods provided scientists with faster and cheaper methods to sequence chloroplast genomes. Moore and colleagues [33] first reported using NGS to determine chloroplast genome sequences, in *Nandina* and *Platanus*. Although multiple NGS platforms are available for chloroplast genome sequencing [34], Illumina is currently the major NGS platform used for chloroplast genomes [21, 32, 35, 36] because it allows the use of rolling circle amplification products [35, 37]. Investigators can then use bioinformatics platforms to perform de novo assembly without the need for reference genome sequences; from these assemblies it is possible to identify consensus chloroplast genome sequences [32]. A third-generation sequencer, the PacBio system which uses single molecule real-time (SMRT) sequencing, is now widely used in chloroplast genome sequencing [38–43]. Its advantage is long read lengths [44], which facilitate de novo genome assembly, particularly in the four chloroplast junctions between the inverted repeat (IR) and single-copy regions.

The low accuracy (~85 % of the raw data) of the long reads produced by the PacBio platform [45] can be corrected by combining the latest chemistry with a hierarchical genome assembly process algorithm; accuracy rates as high as 99.999 % can be achieved after such post-error corrections [46]. Accuracy can also be increased using Illumina short reads [42]. In a study of *Potentilla micrantha*, sequencing with the Illumina platform produced seven contigs covering only 90.59 % of the chloroplast genome; by contrast, using the PacBio platform with error correction, the entire genome was successfully assembled in a single contig [39].

### Chloroplast genome structure

The chloroplast genomes of land plants have highly conserved structures and organization of content; they comprise a single circular molecule with a quadripartite structure that includes two copies of an IR region that separate large and small single-copy (LSC and SSC) regions (Fig. 1a, b). The chloroplast genome includes 120–130 genes, primarily participating in photosynthesis, transcription, and translation. Recent studies have identified considerable diversity within non-coding intergenic spacer regions, which often include important regulatory sequences [13]. Despite the overall conservation in structure, chloroplast genome size varies between species, ranging from 107 kb (*Cathaya argyrophylla*) to 218 kb (*Pelargonium*), and is independent of nuclear genome size (Table 1). Certain lineages of land-plant chloroplast genomes also show significant structural rearrangements, with evidence of the loss of IR regions or entire gene families. Furthermore, there is also evidence for the existence of linear chloroplast genomes, as illustrated in Fig. 1b. The percentage of each form within the cell varies in different reports [47, 48].

Like the genes, the introns in land-plant chloroplast genomes are generally conserved, but the loss of introns within protein-coding genes has been reported in several plant species [49], including barley (*Hordeum vulgare*) [8], bamboo (*Bambusa* sp.) [28], cassava (*Manihot esculenta*) [20], and chickpea (*Cicer arietinum*) [7]. The proteins encoded by genes in which intron loss is known to occur have diverse functions; they include an ATP synthase (atpF), a Clp protease (clpP), an RNA polymerase (rpoC2), and ribosomal proteins (rpl2, rps12, and rps16) [49]. The majority of reported intron losses have been observed in specific plant groups or species, although some examples of intron loss (such as that in *clpP*) occur in diverse plant species, including monocots (Poaceae), eudicots (Onagraceae and Oleaceae) and gymnosperms (*Pinus*) [49].

### Diversity of chloroplast genome sequences

At higher taxonomic levels (family level), protein-coding regions and conserved sequences of the chloroplast



**Table 1** Alphabetical list of crop and tree species that have complete annotated chloroplast genome sequences

Species	Common name	Accession	Genome size (bp)	Uses	Reference(s)
Crops					
<i>Acorus gramineus</i>	Sweet flag	NC_026299	152849	Ornamental	[169]
<i>Agrostis stolonifera</i>	Creeping bent grass	NC_008591	136584	Forage	[8]
<i>Allium cepa</i>	Onion	NC_024813	153538	Vegetable	[170]
<i>Ananas comosus</i>	Pineapple	NC_026220	159636	Fruit	[171]
<i>Anthriscus cerefolium</i>	Chervil	NC_015113	154719	Medicinal	[172]
<i>Artemisia frigida</i>	Fringed sagewort	NC_020607	151076	Medicinal	[173]
<i>Atropa belladonna</i>	Belladonna	NC_004561	156687	Medicinal	[174]
<i>Brassica napus</i> (2)*	Canola	NC_016734	152860	Oil	[175]
<i>Calanthe triplicata</i>	Black orchid	NC_024544	158759	Flower	[176]
<i>Cannabis sativa</i> (2)	Marijuana	NC_027223	153854	Fiber	[177]
<i>Capsicum annuum</i> (2)	Pepper	NC_018552	156781	Vegetable	[178]
<i>Carica papaya</i>	Papaya	NC_010323	160100	Fruit	[179]
<i>Catharanthus roseus</i>	Madagascar periwinkle	NC_021423	154950	Flower	[180]
<i>Cenchrus americanus</i>	Pearl millet	NC_024171	140718	Cereals	[181]
<i>Cicer arietinum</i>	Chickpea	NC_011163	125319	Vegetable	[7]
<i>Coix lacryma-jobi</i>	Job's tears	NC_013273	140745	Cereals	[29]
<i>Colocasia esculenta</i>	Taro	NC_016753	162424	Vegetable	[182]
<i>Cucumis sativus</i> (3)	Cucumber	NC_007144	155293	Vegetable	[183]
<i>Curcuma roscoeana</i>	Jewel of Burma	NC_022928	159512	Medicinal	[184]
<i>Cymbidium tortisepalum</i> (5)	Cymbidium orchid	NC_021431	155627	Flower	[55]
<i>Cypripedium formosanum</i> (3)	Formosa's lady's slipper	NC_026772	178131	Flower	[32]
<i>Daucus carota</i>	Carrot	NC_008325	155911	Vegetable	[9]
<i>Dendrobium catenatum</i>	Dendrobium orchid	NC_024019	152221	Flower	[56]
<i>Dieffenbachia seguine</i>	Dumbcane	NC_027272	163699	Ornamental	[185]
<i>Digitaria exilis</i>	White fonio	NC_024176	140908	Cereals	[181]
<i>Echinochloa oryzicola</i>	Late barnyard grass	NC_024643	139891	Cereals	[186]
<i>Ephedra equisetina</i>	Ma Huang	NC_011954	109518	Medicinal	[187]
<i>Erycina pusilla</i>	Mini orchid	NC_018114	143164	Flower	[36]
<i>Fagopyrum esculentum</i> (2)	Common buckwheat	NC_010776	159599	Cereals	[188]
<i>Festuca arundinacea</i> (4)	Kentucky fescue	NC_011713	136048	Forage	[189]
<i>Fragaria vesca</i> (6)	Wild strawberry	NC_015206	155691	Fruit	[190]
<i>Glycine max</i> (9)	Soybean	NC_007942	152218	Oil	[18]
<i>Glycyrrhiza glabra</i>	Common liquorice	NC_024038	127943	Medicinal	[74]
<i>Gossypium barbadense</i> (22)	Sea island cotton	NC_008641	160317	Fiber	[69]
<i>Guizotia abyssinica</i>	Ramtila	NC_010601	151762	Bird seed	[191]
<i>Helianthus annuus</i> (9)	Common sunflower	NC_007977	151104	Oil	[192]
<i>Heliconia collinsiana</i>	Platanillo	NC_020362	161907	Ornamental	[193]
<i>Hordeum vulgare</i>	Barley	NC_008590	136462	Cereals	[8]
<i>Hyoscyamus niger</i>	Henbane	NC_024261	155720	Medicinal	[194]
<i>Ipomoea batatas</i>	Sweet potato	NC_026703	161303	Vegetable	[195]
<i>Ipomoea purpurea</i>	Common morning glory	NC_009808	162046	Ornamental	[196]
<i>Lactuca sativa</i>	Lettuce	NC_007578	152765	Vegetable	[197]
<i>Lilium superbum</i>	Turk's-cap lily	NC_026787	152069	Flower	[198]

**Table 1** Alphabetical list of crop and tree species that have complete annotated chloroplast genome sequences (*Continued*)

<i>Lolium multiflorum</i> (2)	Ryegrass	NC_019651	135175	Lawn	[199]
<i>Lotus japonicus</i>	Birdsfoot trefoil	NC_002694	150519	Forage	[200]
<i>Manihot esculenta</i>	Cassava	EU117376	161453	Starch crop	[20]
<i>Masdevallia picturata</i> (2)	Masdevallia orchid	NC_026777	157423	Flower	[32]
<i>Musa textilis</i>	Banana	NC_022926	161347	Fruit	[184]
<i>Nicotiana tabacum</i> (4)	Tobacco	Z00044	155943	Tobacco	[3]
<i>Nuphar advena</i>	Spatterdock	NC_008788	160866	Medicinal	[201]
<i>Nymphaea alba</i> (2)	White water-lily	NC_006050	159930	Flower	[24]
<i>Oncidium hybrid</i>	Oncidium	NC_014056	146484	Flower	[54]
<i>Oryza sativa</i> (6)	Rice	X15901	134525	Cereals	[202]
<i>Panax ginseng</i> (2)	Ginseng	NC_006290	156318	Medicinal	[203]
<i>Panicum virgatum</i>	Switchgrass	NC_015990	139619	Biofuel	[204]
<i>Paphiopedilum armeniacum</i> (2)	Slipper orchid	NC_026779	162682	Flower	[32]
<i>Parthenium argentatum</i>	Guayule	NC_013553	152803	Biofuel	[205]
<i>Pelargonium</i> (2)	Geranium	NC_008454	217942	Flower	[206]
<i>Phalaenopsis hybrid</i> (3)	Phalaenopsis orchid	NC_007499	148964	Flower	[51]
<i>Phaseolus vulgaris</i>	Kidney bean	NC_009259	150285	Bean	[78]
<i>Pisum sativum</i>	Pea	NC_014057	122169	Vegetable	[76]
<i>Raphanus sativus</i>	Radish	NC_024469	153368	Vegetable	[207]
<i>Ravenala madagascariensis</i>	Traveller's tree	NC_022927	166170	Ornamental	[184]
<i>Ricinus communis</i>	Castor bean	NC_016736	163161	Oil	[208]
<i>Saccharum hybrid</i> (2)	Sugarcane	NC_005878	141182	Sugar	[209]
<i>Salvia miltiorrhiza</i>	Redroot sage	NC_020431	151328	Medicinal	[210]
<i>Secale cereale</i>	Rye	NC_021761	114843	Cereals	[64]
<i>Sesamum indicum</i>	Sesame	NC_016433	153324	Oil	[211]
<i>Solanum lycopersicum</i> (11)	Tomato	NC_007898	155461	Vegetable	[13]
<i>Solanum tuberosum</i>	Potato	DQ231562	155312	Starch crop	[212]
<i>Sorghum bicolor</i> (2)	Sorghum	NC_008602	140754	Cereals	[8]
<i>Spinacia oleracea</i>	Spinach	NC_002202	150725	Vegetable	[213]
<i>Trifolium grandiflorum</i> (8)	Large-flower hop clover	NC_024034	125628	Forage	[74]
<i>Triticum aestivum</i> (6)	Bread wheat	NC_002762	134545	Cereals	[63]
<i>Vanilla planifolia</i>	Vanilla	NC_026778	148011	Fruit	[32]
<i>Vigna radiata</i> (3)	Mung bean	NC_013843	151271	Bean	[79]
<i>Zea mays</i>	Maize	NC_001666	140384	Cereals	[62]
<i>Zingiber spectabile</i>	True ginger	NC_020363	155890	Ornamental	[193]
Trees and perennial plants					
<i>Abies koreana</i>	Fir	NC_026892	121373	Wood	[214]
<i>Actinidia chinensis</i> (2)	Kiwifruit	NC_026690	156346	Fruit	[215]
<i>Amentotaxus formosana</i>	Taiwan catkin yew	NC_024945	136430	Timber	[216]
<i>Araucaria heterophylla</i>	Norfolk island araucaria	NC_026450	146723	Timber	[217]
<i>Bambusa multiplex</i> (4)	Golden goddess bamboo	NC_024668	139394	Ornamental	[91]
<i>Bambusa oldhamii</i>	Green bamboo	NC_012927	139350	Vegetable	[28]
<i>Berberis bealei</i>	Beale's mahonia	NC_022457	164792	Ornamental	[218]
<i>Bismarckia nobilis</i>	Bismarck palm	NC_020366	158210	Ornamental	[193]
<i>Buxus microphylla</i>	Japanese box	NC_009599	159010	Ornamental	[219]



**Table 1** Alphabetical list of crop and tree species that have complete annotated chloroplast genome sequences (*Continued*)

<i>Calocedrus formosana</i>	Taiwan incense-cedar	NC_023121	127311	Timber	[220]
<i>Calycanthus floridus</i>	Carolina-allspice	NC_004993	153337	Medicinal	[23]
<i>Camellia oleifera</i> (13)	Tea oil plant	NC_023084	156971	Oil	[221]
<i>Camellia reticulata</i>	To-tsubaki	NC_024663	156971	Flower	[222]
<i>Carludovica palmata</i>	Toquilla palm	NC_026786	158545	Fiber	[198]
<i>Castanea mollissima</i>	Chestnut	NC_014674	160799	Fruit	[14]
<i>Cathaya argyrophylla</i>	Cathaya	NC_014589	107122	Timber	[223]
<i>Cedrus deodara</i>	Cedar	NC_014575	119299	Timber	[223]
<i>Cephalotaxus wilsoniana</i> (2)	Wilson plum yew	NC_016063	136196	Timber	[224]
<i>Chrysobalanus icaco</i>	Coco plum	NC_024061	162775	Fruit	[225]
<i>Citrus sinensis</i> (2)	Orange	NC_008334	160129	Fruit	[12]
<i>Cocos nucifera</i>	Coconut	NC_022417	154731	Oil	[226]
<i>Coffea arabica</i>	Coffee	NC_008535	155189	Beverage	[10]
<i>Corymbia gummifera</i> (4)	Red bloodwood	NC_022407	160713	Timber	[227]
<i>Corynocarpus laevigata</i>	Karaka nut	NC_014807	159202	Fruit	[37]
<i>Cryptomeria japonica</i>	Sugi	NC_010548	131810	Timber	[228]
<i>Dendrocalamus latiflorus</i>	Sweet giant bamboo	NC_013088	139394	Vegetable	[28]
<i>Elaeis guineensis</i>	African oil palm	NC_017602	156973	Oil	[229]
<i>Eucalyptus globulus</i> (32)	Eucalyptus	NC_008115	160286	Timber	[230]
<i>Hevea brasiliensis</i>	Rubber tree	NC_015308	161191	Rubber	[231]
<i>Jasminum nudiflorum</i>	Winter jasmine	NC_008407	165121	Ornamental	[232]
<i>Jatropha curcas</i>	Barbados nut	NC_012224	163856	Biofuel	[233]
<i>Juniperus bermudiana</i> (4)	Bermuda juniper	NC_024021	127659	Timber	[234]
<i>Larix decidua</i>	European larch	NC_016058	122474	Timber	[224]
<i>Licania sprucei</i> (3)	Licania	NC_024065	162228	Ornamental	[225]
<i>Liquidambar formosana</i>	Chinese sweetgum	NC_023092	160410	Timber	[30]
<i>Liriodendron tulipifera</i>	Tulip tree	NC_008326	159886	Timber	[235]
<i>Metasequoia glyptostroboides</i>	Dawn redwood	NC_027423	131887	Timber	[236]
<i>Millettia pinnata</i>	Indian beech	NC_016708	152968	Ornamental	[81]
<i>Morus indica</i> (3)	White mulberry	NC_008359	158484	White mulberry	[237]
<i>Nageia nagi</i>	Asian bayberry	NC_023120	133722	Timber	[220]
<i>Nandina domestica</i>	Heavenly bamboo	NC_008336	156599	Ornamental	[33]
<i>Nerium oleander</i>	Oleander	NC_025656	154903	Ornamental	[238]
<i>Olea europaea</i> (5)	Olive	NC_015604	155862	Oil	[239]
<i>Phoenix dactylifera</i>	Date palm	NC_013991	158462	Fruit	[240]
<i>Phyllostachys edulis</i> (4)	Moso bamboo	NC_015817	139679	Timber	[89]
<i>Picea sitchensis</i> (3)	Sitka spruce	NC_011152	120176	Timber	[35]
<i>Pinus taiwanensis</i> (12)	Taiwan red pine	NC_027415	119741	Timber	[241]
<i>Platanus occidentalis</i>	American sycamore	NC_008335	161791	Ornamental	[33]
<i>Podocarpus lambertii</i> (3)	Podocarpus	NC_023805	133734	Ornamental	[242]
<i>Populus alba</i>	White poplar	NC_008235	156505	Timber	[243]
<i>Prinsepia utilis</i>	Himalayan cherry	NC_021455	156328	Ornamental	[244]
<i>Prunus persica</i> (6)	Peach	NC_014697	157790	Fruit	[14]
<i>Pseudophoenix vinifera</i>	Florida cherry palm	NC_020364	157829	Ornamental	[193]
<i>Pseudotsuga sinensis</i>	Chinese douglas	NC_016064	122513	Timber	[224]

**Table 1** Alphabetical list of crop and tree species that have complete annotated chloroplast genome sequences (*Continued*)

<i>Pyrus pyrifolia</i> (2)	Chinese pear	NC_015996	159922	Fruit	[245]
<i>Quercus rubra</i> (4)	Oak	NC_020152	161304	Timber	[246]
<i>Sapindus mukorossi</i>	Soapberries	NC_025554	160481	Medicinal	[247]
<i>Taiwania cryptomerioides</i> (2)	Taiwania	NC_016065	132588	Timber	[224]
<i>Theobroma cacao</i>	Cacao tree	HQ336404	160604	Beverage	[14]
<i>Vaccinium macrocarpon</i>	Large cranberry	NC_019616	176045	Fruit	[248]
<i>Vitis vinifera</i>	Wine grape	NC_007957	160928	Fruit	[19]
<i>Wollemia nobilis</i>	Wollemia	NC_027235	145630	Timber	[249]

\*The number of species in the same genus as the listed species that have sequenced and annotated chloroplast genomes is shown in parentheses

genome can be used for phylogenetic analysis and domestication studies [49]. Earlier phylogenetic analyses utilized partial chloroplast DNA sequences. The use of variable regions or multiple DNA fragments dramatically enhanced the utility of these analyses but there is insufficient information in these sequences to provide the high-resolution necessary to differentiate closely related taxa, particularly some within-species taxa whose taxonomic relationships are unclear. Complete chloroplast genome sequences are valuable for deciphering phylogenetic relationships between closely related taxa and for improving our understanding of the evolution of plant species.

In this section, we discuss several examples of comparisons of chloroplast genomes, within and between crop species, that have provided unique insight into evolutionary relationships among taxa. We also discuss the origin and geographic distribution of economically important species, as well as their adaptations to different climatic conditions and the use of genome information in their breeding and conservation.

A key application of the chloroplast genome in agriculture is the identification of commercial cultivars and the determination of their purity. DNA barcodes derived from the chloroplast genome can be used to identify varieties and in the conservation of breeding resources. Success in breeding is determined by genetic compatibility and chloroplast genomes serve as a valuable tool for identifying plants that are likely to be closely related and, therefore, genetically compatible. Understanding the genetic relationships between cultivated crops and their wild relatives informs efforts to introduce specific advantageous traits into cultivated crops. In the section below, we discuss how chloroplast genomes have been used to elucidate the evolutionary relationships and domestication history of a few major crops and how this informs breeding programs.

### Breeding

The Orchidaceae is a large family that encompasses about 6–11 % of all angiosperms [50] and is important

in floriculture. Many commercially important orchid species belong to the subfamily Epidendroideae and chloroplast genomes of several species from this subfamily have been sequenced [51–58]. Because it is easy to perform inter-generic crossing in orchids and because the record of breeding is sometimes incomplete, it is often difficult to validate the parental origin of commercially important varieties [54]. Corrected parental information is important for breeding and variety identification. In an investigation of the Oncidiinae, a subtribe within the Epidendroideae, PCR products derived from eight conserved regions in 15 commercial varieties resolved their phylogenetic relationship at the species level [54] and helped to resolve putative errors in parental origin. Parental records had indicated that *Odontoglossum* ‘Violetta von Holm’, *Odontoglossum* ‘Margarete Holm’ and *Odontocidium* ‘Golden Gate’ are derived from the same female parent (*Odontoglossum bickoniense*) but phylogenetic analyses of ‘Violetta von Holm’ did not correlate with those of ‘Golden Gate’ or ‘Margarete Holm’ [54]. A possible reason for inconsistencies between the chloroplast DNA-based phylogenetic tree and the parental record is chloroplast capture. Chloroplast capture is the introgression of chloroplasts from one species into another after intrageneric and intergeneric hybridization [59]. Although chloroplast genomes provide useful information for phylogenetic analyses involving closely related taxa, chloroplast capture by hybridization may distort phylogenetic relationships if captured chloroplast genomes or genes included therein are used [60]. The use of both nuclear and chloroplast genomes can provide more complete phylogenies [4, 61].

### Phylogenetic studies

There are several published chloroplast genomes from cereals, including those from sorghum (*Sorghum bicolor*), barley [8], maize (*Zea mays*) [62], wheat (*Triticum aestivum*) [63], rye (*Secale cereale*) [64], and rice (*Oryza sativa*) [65]. Rice is one of the world’s most important crops and is the primary carbohydrate source for the global human population (<http://www.ers.usda.gov/topics/crops/rice.aspx>). The *Oryza* species are classified into ten

genome types, including six diploids (AA, BB, CC, EE, FF, and GG) and four allotetraploids (BBCC, CCDD, HHJJ, and HHKK). Attempts to clarify the evolutionary relationships between cultivated rice and its wild relatives remain contentious and inconclusive [4]. For example, there are two wild species that have an AA genome in Australia, *Oryza meridionalis* (annual) and *Oryza rufipogon* (perennial). *Oryza sativa* was domesticated from Asian *O. rufipogon* 10,000 years ago [65]. Nevertheless, analysis of complete Australian and Asian wild rice chloroplast genomes indicated that Australian *O. rufipogon* chloroplast genomes are more similar to those of Australian *O. meridionalis* than to those of Asian *O. rufipogon* [65–67]. Using 19 chloroplast genomes of *Oryza* AA genome species, a robust phylogenetic tree was established, which will aid in improving rice crops and in conservation strategies [4, 5].

Cotton is the most important textile fiber crop and the first cotton (*Gossypium hirsutum*) chloroplast genome was published in 2006 [11]. The diploid *Gossypium* species comprise eight genome groups (A to G and K genomes). *Gossypium hirsutum* (upland cotton), the most widely planted cotton species in the world, is an allotetraploid of the ancestral A and D genome species [68]. Chloroplast genome sequences are available for 22 *Gossypium* species and these can be used to glean information about the evolution and domestication of this crop [11, 68, 69] (Table 1). Simple sequence repeat primers were used to investigate 41 species of *Gossypium*, including all eight genome groups and allotetraploid species [70]. The results indicated that two modern A-genome species, *Gossypium herbaceum* and *Gossypium arboreum*, were not cytoplasmic donors of tetraploid (AD) species; instead, the AD genome species originated from an extinct ancestor species of the modern A genome [68, 70].

### Domestication

Information on chloroplast genomes is useful for understanding the domestication of several crops, particularly legumes [71]. The chloroplast genome structure of legumes is very interesting; it contains multiple rearrangements, including large inverted segments and loss of inverted repeats [72]. An example is a 51-kb inversion that was first identified in the soybean (*Glycine max*) chloroplast genome sequence [18] before being reported in most members of the subfamily Papilionoideae [7, 73–77]. A 78-kb reversion was subsequently confirmed in *Phaseolus* and *Vigna* chloroplast genomes [78, 79]. More recently, 36-kb [80] and 5.6-kb [81] inversions inside the 51-kb inversion were identified. There are many important genes within these inverted regions but no gene is disturbed and plant survival and performance are not affected. These unique characteristics are not only very

useful in phylogenetic studies [82] but also provide important information for chloroplast transformation in legumes. Chloroplast structure is also important for the design of primers needed in the amplification of sequences for further domestication and phylogenetic analysis.

*Citrus* is one of the most commercially important fruit genera. In 2006, the first *Citrus* chloroplast genome, that of sweet orange (*Citrus × sinensis*), was published [12] and this served as a reference genome for subsequent publications [83, 84]. Phylogenetic analysis of 34 chloroplast genomes of *Citrus* (28) and *Citrus*-related genera (6) indicated that citrus fruits have the same common ancestor [84, 85]. In four genes (*matK*, *ndhF*, *ycf1*, and *ccsA*), single-nucleotide variations and insertion/deletion frequencies were clearly higher than average and showed that these genes have been positively selected. The *matK* gene encodes a maturase that is involved in splicing type II introns and the *matK* sequence is often used in phylogenetic and evolutionary studies [84]. Positive selection of *matK* is observed not only in citrus but is common in several other plant species. In fact, more than 30 plant groups have been shown to undergo positive selection of *matK* genes, indicating that the gene is subject to a number of different ecological selective pressures [86]. The *ndhF* gene encodes a subunit of the chloroplast NAD(P)H dehydrogenase (NDH) complex. Chloroplast NDH monomers are sensitive to high light stress, suggesting that the *ndh* genes may also be involved in stress acclimation [87]. These studies indicated that *matK* and *ndhF* show positive selection in Australian species, potentially contributing to their adaptation to a hot, dry climate [84, 85].

Bamboo is an economically and ecologically important forest plant in Asia [88]. Bamboo grows quickly and new culms are regenerated from the rhizome after harvesting, making it a sustainable and ecologically and environmentally friendly crop. The first two bamboo chloroplast genomes have been published [28] and many more bamboo chloroplast genomes are now available [88–93]. Bamboo has a long juvenility and it is difficult to obtain flowers for taxonomic studies; consequently the taxonomic relationships of bamboo have proven challenging to unravel on the basis of traditional reproductive organ morphology. Furthermore, the extremely low rate of sequence divergence meant that the taxonomic and phylogenetic relationships of temperate woody bamboos at lower taxonomic levels proved difficult to resolve [88]. These relationships were eventually resolved with high-resolution phylogenetic trees using 25 bamboo chloroplast genomes [93]. In addition to woody bamboos, chloroplast genomes have also been published for herbaceous bamboo [88, 92]. An interesting phenomenon identified in herbaceous bamboo chloroplast genomes is that of gene transfer from the mitochondrial genome to



the chloroplast genome. This was an unusual observation, as the chloroplast genome is thought to be nearly immune to the transfer of DNA from nuclear and mitochondrial genomes [88, 92, 94]. A possible reason for this recalcitrance to DNA transfer is the lack of an efficient DNA uptake system [94]. Prior to its observation in herbaceous bamboo, this phenomenon was only observed in two eudicot chloroplast genomes [94] and in monocots [88, 92].

### Transfer of chloroplast genes to nuclear or mitochondrial genomes

There are three distinct genomes in plant cells: nuclear, mitochondrial, and plastid. Mitochondria are believed to have evolved from a single endosymbiotic event by the uptake of a proteobacterium, whereas chloroplasts evolved from endosymbiosis of a cyanobacterium, after which there was a massive transfer of genes from the chloroplast to the nucleus [95]. There are distinct translation systems in these organelles: nuclear-encoded genes are translated in the cytosol and the protein products are then transported to the locations in which they function, including chloroplasts [96], whereas chloroplast-encoded proteins are directly synthesized within the chloroplast. Multi-subunit functional protein complexes that are involved in photosynthesis or protein synthesis are also assembled within chloroplasts.

Gene content, number, and structure are conserved in the chloroplast genome sequences of most autotrophic land plants [97, 98] but some protein-encoding genes are absent in specific species [49]. The loss of genes such as *infA*, *rpl22*, and *ndh* from the chloroplast genome and their intracellular transfer to the nuclear or mitochondrial genomes provide valuable information for phylogenetic analyses and evolutionary studies. It is very easy to identify the chloroplast origin of genes in plant mitochondrial or nuclear genomes [99, 100] by intracellular gene transfer [32], but this could also lead to erroneous phylogenetic relationships when short sequences are used instead of complete chloroplast genome sequences.

The chloroplast *translation initiation factor 1* (*infA*) is a homolog of the essential gene *infA* in *Escherichia coli* [101, 102]. This gene initiates translation in collaboration with two nuclear-encoded initiation factors to mediate interactions between mRNA, ribosomes, and initiator tRNA-Met [102]. Many parallel losses of chloroplast-encoded *infA* have occurred during angiosperm evolution [102] (Fig. 2). Nuclear-encoded *infA* genes have been identified in *Arabidopsis thaliana*, soybean, tomato (*Solanum lycopersicum*), and ice plant (*Mesembryanthemum crystallinum*) [102]. Protein sequences of nuclear-encoded *infA* in these four species contain chloroplast transit peptides. Studies using soybean

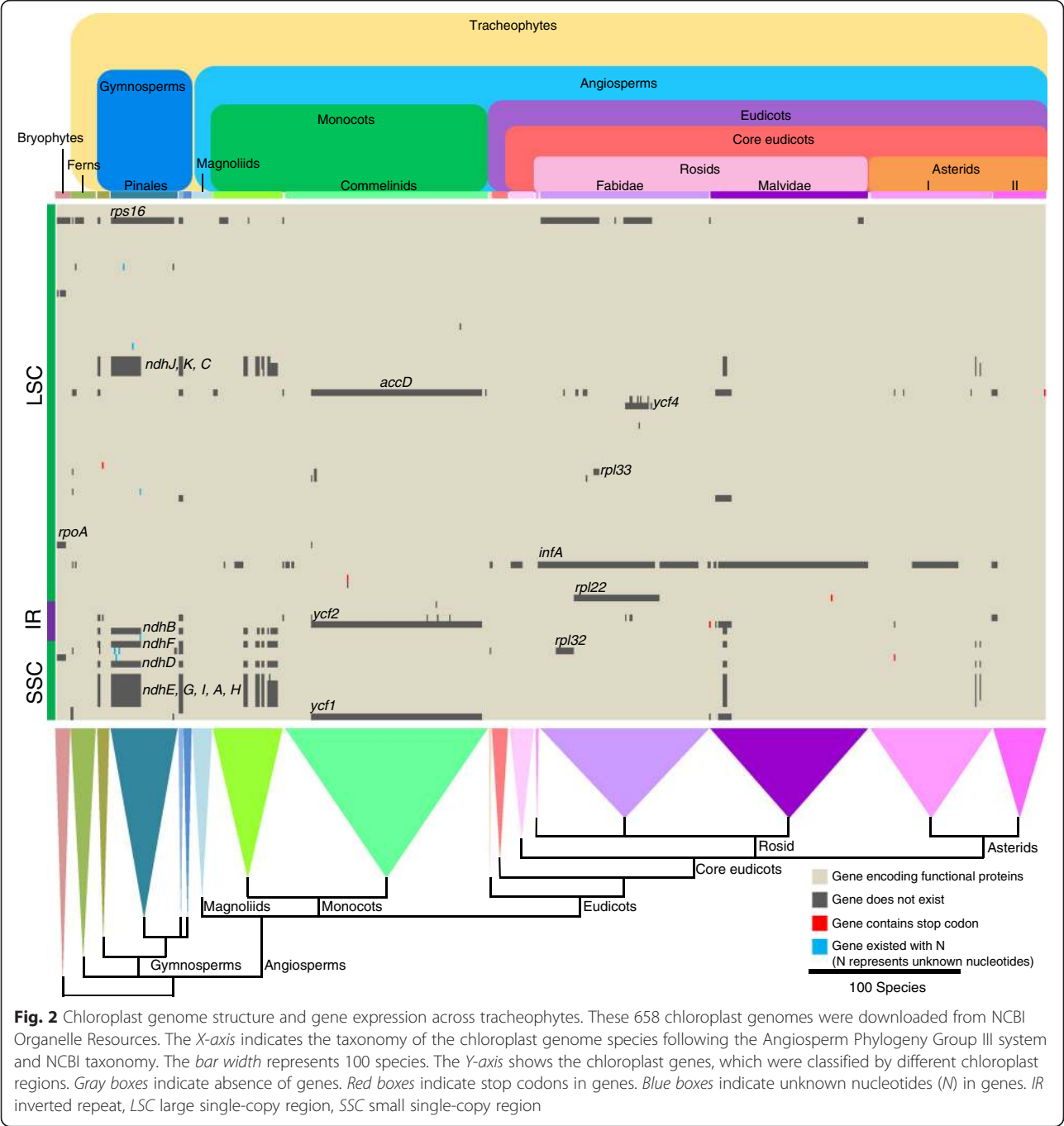
and *A. thaliana* *infA*-GFP proteins have shown that nuclear-encoded *infA* genes are translated in the cytosol and transported into chloroplasts [102]. Many more chloroplast-encoded *infA* deletions have been identified recently (Fig. 2).

There are 57 chloroplast genomes in 26 genera in which the essential gene *rpl22* is reported to have been deleted from the chloroplast and transferred to the nuclear genome (Fig. 2) [14, 103]. Nuclear-encoded *rpl22* contains a transit peptide that is predicted to deliver this protein from the cytosol to chloroplasts. These peptides are diverse, suggesting that there were two independent *rpl22* transfers in the Fabaceae and the Fagaceae [14]. Similar transfer to the nucleus has also been observed for *rpl32* deletion from chloroplast genomes [104–106].

Eleven chloroplast genes encode *ndh* subunits, which are involved in photosynthesis. The *ndh* proteins assemble into the photosystem I complex to mediate cyclic electron transport in chloroplasts [107, 108] and facilitate chlororespiration [109]. Some autotrophic plants lack functional *ndh* genes in their chloroplast genomes [36, 51, 54, 55, 110–115] (Fig. 2). Unlike the single gene losses described previously, the entire family of *ndh* genes has been deleted in these plants. Seven orchid chloroplast genomes indicated at least three independent *ndh* deletions [32]. Some orchid *ndh* DNA fragments were identified in the mitochondrial genome but the complete *ndh* genes required to translate putative functional protein complexes are absent [32]. In the nuclear genome of Norway spruce, only non-functional plastid *ndh* gene fragments are present [116]. Normal photosynthesis is observed in these *ndh*-deleted species [32, 117]. Furthermore, *ndh*-deleted transformants are autotrophic and produce carbohydrates through photosynthesis [107, 118–121].

Many more chloroplast-gene deletions have been observed, including deletions of *accD*, *ycf1*, *ycf2*, *ycf4*, *psaI*, *rpoA*, *rpl20*, *rpl23*, *rpl33*, and *rps16*; many unique gene deletions have been identified in only one or a few species (*psbJ*, *rps2*, *rps14*, and *rps19*) (Fig. 2). The functions of these genes, phenotypes of their knock-out mutants, and evidence for their transfer are summarized in Additional file 1. Most essential genes that have been lost from chloroplast genomes have been transferred to the nucleus to maintain the plant's photosynthetic capacity, with the exception of *ycf1* and *ycf2*.

In summary, chloroplast genome sequences are most valuable for understanding plant evolution and phylogeny. Databases of not only plant genomes but also plant transcriptomes will be useful in investigating deletion events or the transfer of chloroplast genes to other organellar genomes to complement such deletions.



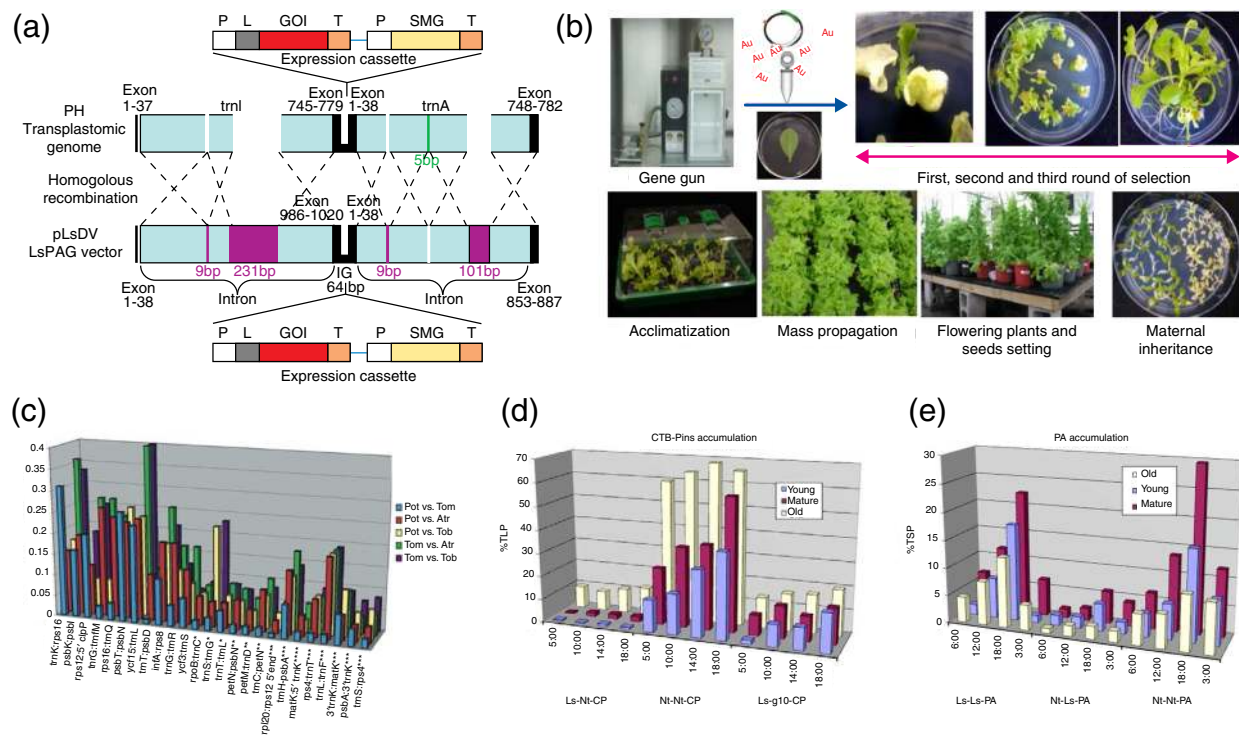
### Advances in chloroplast genome engineering

In the past century, desirable agronomic traits, including yield enhancement and resistance to pathogens or abiotic stress, were achieved by breeding cultivated crops with their wild relatives. As explained above, chloroplast genome sequences are very useful in the identification of closely related, breeding-compatible plant species. With the advent of modern biotechnology, desirable traits from unrelated species can now be readily introduced

into commercial cultivars. Such genetically modified crops have revolutionized agriculture in the past two decades, dramatically reducing the use of chemical pesticides and herbicides while enhancing yield. For most commercial cultivars, herbicide- or insect-resistance genes are introduced into the nuclear genome. There are, however, a few limitations for nuclear transgenic plants, including low levels of expression (<1 % total soluble protein (TSP)) and potential escape of transgenes via pollen.

Engineering the introduction of foreign genes into the chloroplast genome addresses both of these concerns. Just two copies of transgenes are typically introduced into the nuclear genome, whereas up to 10,000 transgene copies have been engineered into the chloroplast genome of each plant cell, resulting in extremely high levels of foreign gene expression (>70 % TSP) [122]. Most importantly, chloroplast genomes are maternally inherited in most cultivated crops, minimizing or eliminating transgene escape via pollen [123].

The basic process of chloroplast engineering is explained in Fig. 3a, b. Chloroplast genome engineering is accomplished by integrating foreign genes into intergenic spacer regions without disrupting the native chloroplast genes (Fig. 3a). Two chloroplast genes are used as flanking sequences to facilitate integration of transgene cassettes. Transgene cassettes include a selectable marker gene and gene(s) of interest, both regulated by chloroplast gene promoters and untranslated regions (UTRs; Fig. 3a). Chloroplast genome sequences are essential to build transgene cassettes because they provide



**Fig. 3** Basic process of chloroplast genetic engineering, diversity in intergenic spacer regions, and impact of transgene integration (endogenous versus heterologous genome sequences). **a** Complexity of heterologous sequence integration into intergenic spacer regions between lettuce and tobacco. The schematic diagram represents recombination between the tobacco transplastomic genome and the lettuce transformation vector [128]. Purple bars represent unique lettuce intron sequence; the green bar represents unique tobacco intron sequence; black bars are exon regions; blue regions are looped out sequence. The expression cassette comprises: promoters (P), leader sequence (L), gene of interest (GOI), terminators (T), and selectable marker gene (SMG). IG intergenic spacer region. **b** Basic process of chloroplast genetic engineering. Gene delivery is performed by bombardment with gold microparticles coated with chloroplast vectors, followed by three rounds of selection to achieve homoplasmy. After confirmation of transgene integration, plants are grown in the greenhouse to increase biomass. Chloroplast transgenes are maternally inherited without Mendelian segregation of introduced traits. **c** Comparison of 21 of the most variable intergenic spacer regions among Solanaceae chloroplast genomes. *Atr* *Atropa*, *Pot* potato, *Tob* tobacco, *Tom* tomato. \*Tier 1, \*\*tier 2, and \*\*\*tier 3 regions reported in the paper by Shaw et al. [250]. Plotted values were converted from percentage identity to sequence divergence on a scale from 0 to 1 as shown on the Y-axis; these values demonstrate a wide range of sequence divergence in different regions. Nucleotide sequences were determined by a bridging shotgun method and genome annotation was performed using the Dual Organellar GenoMe Annotator [13]. **d, e** Decrease in the expression of transgenes regulated by heterologous *psbA* promoters and untranslated regions (UTRs) engineered via tobacco chloroplast genomes. When the lettuce (*La*) *psbA* regulatory region was used in tobacco (*Na*) chloroplasts or vice versa, transgene expression is dramatically reduced. **d** Accumulation of a cholera toxin B subunit (CTB) and proinsulin (Pins) fusion protein (CP) was quantified by densitometry and **e** anthrax protective antigen (PA) accumulation was estimated by enzyme-linked immunosorbent assay (ELISA). Total leaf protein (TLP) or total soluble protein (TSP) data are presented as a function of light exposure and developmental stage. The order of young, mature, and old is different in **d** and **e** because of the accumulation of more CTB-Pins in older leaves and PA in mature leaves [128]. Young (top five), mature (fully grown), and old (bottom three) leaves were fully expanded and were cut from plants grown in the greenhouse for 8–10 weeks

both flanking and regulatory sequences. Transgene cassettes that are inserted into bacterial plasmids are called chloroplast vectors and they are bombarded into plant cells using gold particles and a gene gun (Fig. 3b). Because of the presence of chloroplast DNA in the nuclear or mitochondrial genome, transgene cassettes may integrate via homologous or non-homologous recombination events; but any transgenes that are integrated within the nuclear or mitochondrial genome will not be expressed because chloroplast regulatory sequences are not functional in other genomes. If such integration occurs, the transgenes could be easily identified by evaluation of their integration site and eliminated [124].

One of the challenges of creating chloroplast transgenic (transplastomic) plants is the elimination of all untransformed copies (>10,000 per cell) of the native chloroplast genome and replacing them with transformed genomes that contain integrated transgene cassettes. The absence of the native chloroplast genome and the presence of only the modified genomes is referred to as the homoplasmic state, which is typically achieved after two or three rounds of selection (Fig. 3b). The most effective selectable marker used is the *aadA* gene, which confers resistance to streptomycin and spectinomycin. These antibiotics bind specifically to chloroplast ribosomes and disrupt protein synthesis without interfering with any other cellular process. Efforts to transform the chloroplast genome of cereal crops have been mostly unsuccessful. This could be due to the instability of chloroplast DNA in the mature leaves of cereals [47] or to a requirement for better selectable markers [125].

Table 2 provides the first global, comprehensive summary of the power of chloroplast genetic engineering, utilizing valuable information generated by the sequencing of chloroplast genomes described in previous sections. This table includes the most complete list of chloroplast genomes that have been engineered for enhanced agronomic traits or the production of different bio-products, including biopolymers, industrial enzymes, biopharmaceuticals, and vaccines. Within Table 2, transgenes are grouped according to their functions and are organized according to their site of integration. The efficiency of transgene expression is also included in Table 2, providing important information about the regulatory sequences used to express the transgenes.

### Impact of sequence diversity in the chloroplast genome on transgene integration

Figure 3a shows examples of transplastomic genomes that have been transformed with either an endogenous or a heterologous flanking sequence. Every single nucleotide change in the heterologous sequence was subsequently edited out and corrected to achieve 100 % homology to the native sequence within the intergenic

spacer region (Fig. 3a). The repetitive editing process significantly reduces the efficiency of transgene integration when using heterologous flanking sequences. This challenge is made even more difficult by inadequate conservation of intergenic spacer regions, even within the same family. Figure 3c shows comparisons of 21 of the most variable intergenic spacer regions; only four of the >150 spacer regions, including the *trnI/trnA* spacer region, are conserved among members of the Solanaceae. Among grass chloroplast genomes, not a single intergenic spacer region is conserved [8]. This necessitates construction of species-specific chloroplast vectors using endogenous flanking sequences and underscores the need to sequence the chloroplast genomes of economically important crop species.

### Ideal sites in the chloroplast genome for transgene integration

The selection of a suitable intergenic spacer region from among more than 100 sites found in each chloroplast genome is a major concern. Statements on the lack of positional effects in the transplastomic literature are common and are used to contrast chloroplast genetic engineering with nuclear transgene integration, which is often associated with profound differences in the expression of transgenes dependent on their site of integration. Evidence shows, however, that there are also positional effects within the chloroplast genome (Table 2). IR regions are found in duplicate in most chloroplast genomes; therefore, transgenes should be inserted within the IR region instead of the SSC or LSC regions because this should double the copy number of transgenes. Integration of a transgene cassette into one copy of the IR facilitates integration into the other copy, thereby enhancing selection pressure to achieve homoplasmy through this copy correction mechanism, a characteristic feature of the chloroplast genome [126–128]. Therefore, the site of integration plays a crucial role in transgene expression level and in enhancing homoplasmy under selection by antibiotics. Most importantly, in all sequenced chloroplast genomes within a single plant species, the DNA sequence in one copy of the IR is identical to that in the other copy, without any exception (Table 1).

An early controversy in the chloroplast genetic engineering field was the suitability of transcriptionally silent spacer regions, where native genes (for example, *rbcL/accD*) are located on opposite strands of the chloroplast genome, or transcriptionally active spacer regions, where native genes (for example, *trnA/trnI*) are located within operons on the same strand. After a herbicide resistance gene was introduced into the transcriptionally active spacer region for the first time [129], most subsequent studies preferentially used this site of integration (Table 2). The integration of transgenes into the transcriptionally active

**Table 2** Engineering the chloroplast genome for biotechnology applications

Site of integration	Transgenes	Regulatory sequences	Efficiency of expression	Engineered traits or products	Reference(s)
Insect or pathogen tolerance					
trnI/trnA	<i>Bgl-1</i>	5'psbA/3'psbA	>160-fold enzyme	Resistance against whitefly and aphid	[141]
trnI/trnA	<i>Pta</i>	5'psbA/3'psbA	7.1–9.2 % TSP	Broad-spectrum resistance against aphid, whitefly, Lepidopteran insects, bacterial and viral pathogens	[142]
trnI/trnA	<i>PelB1, PelD2</i>	5'psbA/3'psbA	~2.42 units mg <sup>-1</sup> FW	Resistance against <i>Erwinia</i> soft rot	[150]
trnI/trnA	<i>RC1011, PG12</i>	5'psbA/3'	17–38 % TSP	Resistance to <i>Erwinia</i> soft rot and tobacco mosaic virus	[140]
trnI/trnA	<i>cpo</i>	Prrn/psbA/psbA	NR	Resistance to fungal pathogens in vitro ( <i>Aspergillus flavus</i> , <i>Fusarium verticillioides</i> , and <i>Verticillium dahliae</i> ) and in planta ( <i>Alternaria alternata</i> )	[251]
trnI/trnA	<i>Bt cry2Aa2</i> operon	Prrn/ggagg/psbA	45.3 % TSP	100 % mortality of cotton bollworm, beet armyworm; cuboidal Bt crystals formation	[137]
trnI/trnA	<i>Bt cry9Aa2</i>	Prrn/ggagg/rbcL	~10 % of TSP	Resistance to <i>Phthorimaea operculella</i>	[252]
trnI/trnA	<i>msi-99</i>	Prrn/ggagg/psbA	21–43 % TSP	Resistance to in planta challenge of <i>Pseudomonas syringae</i> , <i>Aspergillus flavus</i> , <i>Fusarium moniliforme</i> , <i>Verticillium dahlia</i> , and <i>Colletotrichum destructivum</i>	[253]
trnI/trnA	<i>sporamin1, CeCPI2, and chitinase2</i>	Prrn/TpsbA	0.85–1 % TSP	Resistance against <i>Spodoptera litura</i> and <i>Spodoptera exigua</i> leaf spot, as well as soft rot diseases	[254]
trnI/trnA	<i>MSI-99</i>	Prrn/Trps16	89.75 µg g <sup>-1</sup> FW	Resistance against rice blast fungus	[255]
trnV/rps12/7	<i>cry1A(c)</i>	Prrn/rbcL/rps16	3–5 % of TSP	Resistance to larvae of <i>Heliothis virescens</i> , <i>Helicoverpa zea</i> , and <i>Spodoptera exigua</i>	[256]
trnV/rps12/7	<i>cry1Ab</i>	Prrn/T7gene10/rbcL	NR	Resistance to caterpillar of <i>Anticarsia gemmatilis</i>	[145]
rbcL/accD	<i>cry2Aa2</i>	Prrn/ggagg/psbA	2–3 % of TSP	Resistance to <i>Heliothis virescens</i> , <i>Helicoverpa zea</i> , and <i>Spodoptera exigua</i>	[257]
Abiotic stress tolerance					
trnI/trnA	<i>tps1</i>	Prrn/ggagg/psbA	>169-fold transcript	Drought tolerance: growth in 6 % polyethylene glycol and rehydration after 24 days of drought	[258]
trnI/trnA	<i>merA/merB</i>	Prrn/ggagg/psbA	NR	Phytoremediation: high level tolerance to the organomercurial compounds, up to 400 µM phenylmercuric acetate	[259]
trnI/trnA	<i>badh</i>	Prrn/T7 g10/rps16	93–101 µM g <sup>-1</sup> FW	Salt tolerance: carrot plants survived up to 400 mM NaCl	[135]
trnI/trnA	<i>γ-TMT</i>	Prrn/T7g 10/TpsbA	>7.7 % TSP	Increased salt and heavy metal tolerance, enhanced accumulation of α-tocopherol in seeds	[153]
trnI/trnA	<i>mt1</i>	Prrn/T7 g10/Trps16	NR	Phytoremediation: resistant to mercury, up to 20 µM	[260]
trnV/rps12/7	<i>b-bar1</i>	Prrn/TrbcL	>7 % TSP	Resistance to the herbicide phosphinothricin	[261]
trnV/rps7/12	<i>EPSPS</i>	Prrn/Trps16	>10 % TSP	Resistance to the herbicide glyphosate	[262]
rbcL/accD	<i>EPSPS/aroA</i>	Prrn/ggagg/psbA	NR	Resistance to glyphosate (>5 mM)	[129]



**Table 2** Engineering the chloroplast genome for biotechnology applications (*Continued*)

rbcl/accD	<i>mALS</i>	PpsbA/TpsbA	NR	Tolerant to pyrimidinylcarboxylate, imidazolinon, and sulfonylurea/pyrimidinylcarboxylate herbicides	[263]
rbcl/accD	<i>Bar</i>	Prrn/rbcl/psbA	NR	Herbicide resistance: up to 25 µg ml <sup>-1</sup> glufosinate	[264]
rbcl/rbcl	<i>Hppd</i>	psbA/psbA/3'rbcl	5 % TSP	Resistance to herbicide	[265]
rbcl/accD	<i>panD</i>	Prrn/rbcl 3'	>4-fold β-alanine	Tolerance to high-temperature stress	[266]
trnfM/trnG	<i>lycopene β-cyclase</i>	atpl/rps16	0.28 mg g <sup>-1</sup> DW	Herbicide resistance and triggers conversion of lycopene	[133]
prs14/trnG	<i>HTP, TCY, TMT</i>	Prrn/T7 g10/Trbcl	NR	Increase in vitamin E in fruit; cold-stress tolerance	[267]
Other agronomic traits					
trnI/trnA	<i>phaA</i>	Prrn/psbA/psbA	14.71β-ketothiolase mg <sup>-1</sup> FW	Engineered cytoplasmic male sterility	[268]
trnI/trnA	<i>RbcS</i>	T7g10 or psbA	>150-fold RbcS transcript	Restoration of RuBisCO activity in <i>rbcS</i> mutants	[136]
rbcl/accD	<i>TC, γ -TMT</i>	PpsbA/Trsp16	3 nmol h <sup>-1</sup> mg <sup>-1</sup> FW	Vitamin E accumulation in tobacco and lettuce	[269]
rbcl/accD	<i>CrtZ, CrtW</i>	Prrn/Trps16	NR	Accumulation of astaxanthin fatty acid esters in lettuce	[270]
trnV/orf708	<i>BicA</i>	psbA/psbA/psbA	~0.1 % TSP	CO <sub>2</sub> capture within leaf chloroplasts	[271]
trnV/3'rps12	<i>Trx f, Trx m</i>	prrn T7G10/rps12	NR	Starch synthesis/chloroplast redox regulation	[272]
trnfM/trnG	<i>CV-N</i>	Prrn/T7g10/TatpA	~0.3 % TSP	Increased mRNA stability and protein stability with the expression of CV-N in chloroplasts	[273]
trnI/trnA	<i>Bgl-1</i>	5'psbA/3'psbA	44.4 units Bgl1 g <sup>-1</sup> FW	β-Glucosidase increased enzyme cocktail efficiently to release sugar from paper, citrus peel, and wood	[141]
trnI/trnA	<i>ubiC</i>	5'psbA/3'psbA	25 % DW	250-fold higher pHBA polymer accumulation than nuclear transgenic lines	[149]
trnI/trnA	<i>man 1</i>	5'psbA/3'psbA	25 units g <sup>-1</sup> FW	Mannanase increased enzyme cocktail released sugar from paper, citrus peel, and wood	[274]
trnI/trnA	<i>cutinase or swollenin</i>	5'PsbA/3'PsbA	47.7 % reduction of MGDG and DGDG in cutinase and 68.5 % in swollenin	Swollenin enlarged and irreversibly unwound cotton fiber; cutinase showed esterase and lipase activity; used in enzyme cocktails	[275]
trnI/trnA	<i>bgl1</i>	5'psbA/3'psbA	14 units mg <sup>-1</sup> FW	Enzyme cocktails produced glucose from filter paper, pine wood, or citrus peel	[150]
	<i>swo1</i>		NR		
	<i>xyn2</i>		421 units mg <sup>-1</sup> FW		
	<i>Acetyl sytan esterase</i>		NR		
	<i>celD</i>		493 units mg <sup>-1</sup> FW		
	<i>celO</i>		442 units mg <sup>-1</sup> FW		
	<i>Lipase</i>		NR		
trnI/trnA	<i>Cutinase</i>	Prrn/PpsbA	15 units mg <sup>-1</sup> FW	Artemisinic acid for several isoprenoid products	[276]
	<i>PMK, MVK, MDD, AACT, HMGS, HMGRt; IPP, FPP, ADS, CYP71AV1, AACPR</i>		0.1 mg g <sup>-1</sup> FW		
trnI/trnA	<i>Cel6A, Cel6B</i>	Prrn/rbcl/rbcl	2–4 % TSP	Hydrolyzed crystalline cellulose	[277]
trnfM/trnG	<i>bgl1C, cel6B, cel9A, xeg74</i>	Prrn/T7g10/Trbcl	5– 40 % TSP	Cell wall-degrading enzyme activity	[278]

**Table 2** Engineering the chloroplast genome for biotechnology applications (*Continued*)

rbcl/accD	<i>phbC, phbA, phbB</i>	Prrn/rbcl 3'	0.16 % DW	Polyhydroxybutyrate (PHB) accumulation in leaves	[279]
rbcl/accD	<i>crtZ, crtW</i>	Prrn/Trps16	>0.5 % DW	Astaxanthin accumulation	[280]
trnV/rps7	<i>EGPh</i>	psbA/psbA/Trps16	25 % TSP	Chloroplast-derived $\beta$ -1,4-endoglucanase (EGPh) was recovered from dry leaves and digested carboxymethyl cellulose (CMC) substrate	[281]
trnI/trn A	<i>EX4</i>	PpsbA/TpsbA	14.3 % TSP	CTB-EX4 showed increased insulin secretion similar to the commercial injectable EX4 in pancreatic $\beta$ -cells and in mice fed with cells expressing EX4 in chloroplasts	[160]
trnI/trn A	<i>MBP</i>	PpsbA/TpsbA	2 % TSP	Amyloid loads were reduced in ex vivo studies in human Alzheimer's brain and in vivo in Alzheimer's mice fed with bio-encapsulated CTB-MBP. Abeta was also reduced in retinae and loss of retinal ganglion cells was prevented	[162]
trnI/trn A	<i>FVIII</i>	PpsbA/TpsbA	370 mg g <sup>-1</sup> FW	Feeding of the HC/C2 antigen mixture substantially suppressed T-helper cell responses and inhibitor formation against FVIII in hemophilia A mice	[282]
trnI/trn A	<i>HSA</i>	PpsbA/TpsbA	26 % TSP	In vitro chaperone activity of Trx m and Trx f	[283]
trnI/trn A	<i>EDA</i>	PpsbA/TpsbA	2.0 % TSP	The vaccine adjuvant EDA from fibronectin retains its proinflammatory properties when expressed in tobacco chloroplasts	[284]
trnI/trn A	<i>Proinsulin</i>	PpsbA/TpsbA	47 % TSP in tobacco, 53 % TLP in lettuce	Oral delivery of proinsulin in plant cells lowered glucose levels comparably to injectable commercial insulin	[285]
trnI/trn A	<i>HSA</i>	psbA/psbA/psbA	~11 % TSP	First report of human blood protein in chloroplasts; function not evaluated	[286]
trnI/trn A	<i>IGF</i>	psbA/psbA/psbA	32.7 % TSP	Promoted growth of cultured HU-3 cells in a dose-dependent manner	[287]
trnI/trnA	<i>FIX</i>	PpsbA/TpsbA	1 mg g <sup>-1</sup> DW (0.56 % TLP)	Oral delivery of CTB-FIX lettuce cells suppressed inhibitor formation against FIX in hemophilia B mice	[6]
trnI/trnA	<i>FIX</i>	Ppsba/TpsbA	3.8 % TSP; 0.4 mg g <sup>-1</sup> FW	Tolerance induction via complex immune regulation, involving tolerogenic dendritic and T-cell subsets	[288]
trnI/trnA	<i>GAA</i>	Ppsba/TpsbA	5.7 mg g <sup>-1</sup> DW	Reduced toxic antibody responses in enzyme replacement therapy in Pompe mice	[289]
trnI/trnA	<i>ACE2</i> <i>Ang-(1-7)</i>	PpsbA/TpsbA	CTB-ACE2: 2.14 % TLP CTB-Ang1-7: 8.7 % TLP	Oral delivery of ACE2 and Ang (1-7) significantly improved cardiopulmonary structure and functions, decreased the elevated right ventricular systolic blood pressure and improved pulmonary blood flow in animals with induced pulmonary hypertension	[161]
trnI/trn A	<i>BACE</i>	Prrn/TpsbA	2.0 % TSP	Immunogenic response against the BACE antigen in mice	[290]
trnI/trn A	<i>IFNa2b</i>	Prrn/TpsbA	3 mg g <sup>-1</sup> FW	Protected cells against VSV CPE and HIV; increased MHC I antibody on splenocytes and total number of natural killer cells and protected mice from a highly metastatic lung tumor	[291]

**Table 2** Engineering the chloroplast genome for biotechnology applications (*Continued*)

trnI/trn A	<i>CTB-pins</i>	Prrn/T7g10/TpsbA and rps16	16 % TSP in tobacco, 72 % TLP in lettuce	CTB-proinsulin-fed non-obese diabetic mice significantly decreased inflammation (insulinitis); insulin-producing $\beta$ cells in pancreatic islets were highly protected, increased in insulin production with lower blood or urine glucose levels; increased expression of immunosuppressive cytokines	[128, 292]
rbcl/accD	<i>IFN-<math>\gamma</math></i>	PpsbA/TpsbA	6 % TSP	Protection of human lung carcinoma cells against infection by encephalomyocarditis virus	[293]
rbcl/accD	<i>hTrx</i>	PpsbA/Trps16	1 % TSP	Protected mouse from hydrogen peroxide	[294]
rbcl/accD	<i>A1AT</i>	PpsbA/TrbcL	2 % TSP	Binds to porcine pancreatic elastase	[295]
rbcl/accD	<i>TGF<math>\beta</math>3</i>	Prrn/T7g10/psbC	12 % TLP	Inhibits mink lung epithelial cell proliferation	[296]
trnV/3'rps12	<i>hCT-1</i>	Prrn/G10L/Trps16	5 % TSP	Biologically active on human hepatocarcinoma cell line	[297]
trnV/rps7/12	<i>hST</i>	PpsbA or Prrn/G10L/Trps16	0.2–7.0 % TSP	Promotes growth of Nb2 cells in a dose-dependent manner	[298]
trnM/trnG	<i>pal, cpl-1</i>	Prrn/T7g10/TpsbA	~30 % TSP	Bacteriolytic activity and kills <i>Streptococcus pneumoniae</i> , the causative agent of pneumonia	[299]
trnI/trn A	<i>ESAT-6</i>	5'psbA/3'psbA	~7.5 % TSP	Hemolysis of red blood cells and GM1 binding	[165]
trnI/trn A	<i>AMA1</i>	5'psbA/3'psbA	7.3 % TSP in tobacco, 13.2 % TSP in lettuce	Long-term immunity against cholera challenge; inhibition of malarial parasite; protection correlated with IgA and IgG1	[164]
trnI/trn A	<i>MSP1</i>	5'psbA/3'psbA	10.1 % TSP in tobacco, 6.1 % TSP in lettuce		
trnI/trn A	<i>2 L21</i>	5'psbA/3'psbA	6.0 % TSP		
trnI/trn A	<i>Pag</i>	5'psb/3'psbA	~29.6 % TSP	Macrophage lysis assay, systemic immune response, toxin neutralization assay, mice survived (100 %) challenge with lethal doses of anthrax toxin	[300, 301]
trnI/trn A	<i>L1</i>	PpsbA/TpsbA	20–26 % TSP	Induced systemic immune response and produced neutralizing antibodies in mice	[302]
trnI/trnA	<i>RA4</i>	PpsbA/T psbA	0.2 % TLP	Oral administration elicited both mucosal and systemic Th1/Th2 responses to reduce <i>Toxoplasma</i> parasite load	[303]
trnI/trnA	<i>rFaeG</i>	PpsbA/TrbcL	>1 % DW	Transplastomic plants expressing the rFaeG protein could possibly be used for delivery of an oral vaccine against porcine F4+ ETEC infections	[304]
trnI/trn A	<i>F1-V</i>	Prrn/TpsbA	14.8 % TSP	Orally immunized mice heavily challenged with plague ( <i>Yersinia pestis</i> ) were protected better than those given IP injections	[305]
trnI/trn A	<i>CTB-2 L21</i>	PpsbA/TpsbA	31.1 % TSP	Immunogenic in mice following IP or oral administration	[306]
trnI/trnA	<i>VP8*</i>	psbA/psbA/Trps16	600 $\mu$ g g <sup>-1</sup> FW	Induced strong immune response and virus neutralization	[307]
trnI/trn A	<i>CtxB</i>	Prrn/ggagg/TpsbA	4.1 % TSP	Efficient GM1 ganglioside-binding	[308]
trnI/trn A	<i>LTB</i>	Prrn/ggagg/TpsbA	2.5 % TSP	GM1 ganglioside-binding assay	[309]
trnI/trn A	<i>LecA</i>	Prrn/T7g10/TpsbA	7 % TSP	Systemic immune response in mice	[310]

**Table 2** Engineering the chloroplast genome for biotechnology applications (*Continued*)

trnI/trnA	<i>BACE</i>	Prrn/TpsbA	2.0 % TSP	Immunogenic response against the BACE antigen in mice	[290]
rbcl/accD	<i>OspA</i> , <i>OspA-T</i>	PpsbA/TpsbA	1–10 % TSP	Systemic immune response and protection against <i>Borrelia burgdorferi</i> (Lyme disease)	[311]
trnN/trnR	<i>LTB</i>	Prrn/T7g10/TrbcL	2.3 % TSP	GM1 ganglioside-binding assay; oral immunization partially protected mice from cholera toxin challenge	[312]
trnN/trnR	<i>DPT</i>	Prrn/T7g10/TrbcL	0.8 % TSP	Immunogenic in orally inoculated mice with freeze-dried chloroplast-derived multi-epitope DPT protein	[313]
trnN/trnR	<i>C4V3</i>	Prrn/T7g10/TrbcL	~15 $\mu\text{g mg}^{-1}$ DW	Plant-derived C4V3 has elicited both systemic and mucosal antibody responses in mice, as well as CD4+ T cell proliferation responses	[314]
trnN/trnR	<i>L1</i>	Prrn/TrbcL	>2 % of TSP	Proper folding and display of conformational epitopes for L1 in the fusion protein by antigen capture ELISA	[315]
trnM/trnG	<i>p24</i>	Prrn/T7g10/TrbcL	~4 % TSP	Induced strong CD4+ and CD8+ T-cell responses in mice	[316]
trnG/trnM	<i>HEV E2</i>	Prrn/psbA/TpsbA	1.09 $\text{ng } \mu\text{g}^{-1}$ TSP	Immune response in mice against hepatitis E virus	[317]
trnH/trnK	<i>CSFV E2</i>	Prrn/TpsbA	1–2 % TSP	Immune response in mice against swine fever	[318]
rrn16/rps12/7	<i>TetC</i>	Prrn/T7 g10/TrbcL atpB/TrbcL	10–25 % TSP	Mice developed systemic immune response and survived the tetanus toxin challenge	[319]
rrn16/trnI	<i>E7</i>	PpsbA/Trps	3–8 % TSP	Several therapeutic HPV-specific E7-based vaccine formulations have been tested in animal models and some have advanced into clinical trials	[320]

**Abbreviations:** *Ang* (1–7) Angiotensin (1–7), *BACE* human b-site APP cleaving enzyme, *Bgl*  $\beta$ -glucosidase, *CPE* carbapenemase-producing Enterobacteriaceae, *CTB* cholera toxin B subunit, *DGDG* digalactosyldiacylglycerol, *DPT* diphtheria, pertussis, tetanus, *DW* dry weight, *EDA* extra domain A-fibronectin, *ELISA* enzyme-linked immunosorbent assay, *ETEC* enterotoxigenic *Escherichia coli*, *EX4* exendin-4, *FVIII* coagulation factor VIII, *FW* fresh weight, *HPV* human papilloma virus, *IP* intraperitoneal, *MBP* myelin basic protein, *MGDG* monogalactosyldiacylglycerol, *NR* not recorded, *RbcS* small subunit of RuBisCO, *RuBisCO* ribulose-1,5-bisphosphate carboxylase/oxygenase, *TLP* total leaf protein, *TSP* total soluble protein, *VSV* vesicular stomatitis virus

spacer region (trnI/trnA) has led to 25-fold higher expression of transgenes compared with the transcriptionally silent spacer region (rbcl/accD) [130], possibly due to the presence of multiple promoters (heterologous and endogenous) that enhance transcription. Introns present within *trnI/trnA* genes (used as flanking sequences) also provide efficient processing of native or foreign transcripts. The *trnA* gene intron includes a chloroplast origin of replication and produces more copies of the template (chloroplast vectors) for integration of the transgene cassette [131]. In fact, among 114 transgenes in different plant species in Table 2, 71 are integrated at the *trnA/trnI* site of the chloroplast genome, confirming the unique advantages of this site [127, 129, 130].

### Role of chloroplast genome regulatory sequences in transgene expression

In addition to the site of integration, regulatory sequences located upstream (promoter, 5' UTR) and

downstream (3' UTR) of transgenes play a major role in determining their expression level. The *psbA* regulatory region, first used almost 25 years ago [131], still appears to be the best option for use in an expression cassette, as the *psbA* gene encodes the most highly translated protein in the chloroplast [132] and it can also mediate light-induced activation of translation [128]. Indeed, almost all highly expressed transgenes (>70 % TSP, >25 % dry weight) utilize the *psbA* regulatory region; among 114 transgenes expressed via the chloroplast genome, 84 use the *psbA* regulatory sequence (Table 2). Other endogenous regulatory sequences that are used include *rbcl* and *atpA*, which result in lower transgene expression levels than the *psbA* promoter/5' UTR.

Using regulatory regions from photosynthetic genes has the advantage of light regulation, making them ideal for transgene expression in photosynthetic organs (leaves; Fig. 3d, e). However, when the lettuce *psbA* regulatory region was used in tobacco chloroplasts or vice

versa, transgene expression was dramatically reduced (Fig. 3d, e) [128]. Nucleotide differences within the *psbA* 5' UTR between tobacco and lettuce (*Lactuca sativa*) resulted in changes that decreased the interaction of RNA-binding proteins and produced variation in the size of the stem, bulge, and terminal loop of the UTR [128]. In addition, most regulatory proteins (including sigma factors that bind to the promoter region) are nuclear encoded and transported to chloroplasts. This underscores a caveat associated with using regulatory sequences for transgene expression: the need to make species-specific chloroplast vectors to accommodate highly specific regulatory region-binding proteins.

Heterologous regulatory sequences are necessary for transgene expression that is independent of cellular control, especially in non-photosynthetic organs such as fruits and edible roots, where chloroplast protein synthesis is poor [133]. A heterologous UTR (T7 *gene10*) was first evaluated for expression in leaves [127, 134] and

was subsequently tested in non-green tissues. When the expression of *BETAINE ALDEHYDE DEHYDROGENASE* (*BADH*) was regulated by the T7 *gene10* UTR in carrot (*Daucus carota*) plants, 75 % of the expression level in leaves was observed in non-green edible roots, conferring the highest level of salt tolerance (400 mM NaCl) found in the published literature (Fig. 4i, j) [135]. Although T7 *gene10* has been successfully used to engineer salt tolerance in non-green tissues, its expression level is not as high as that of the *psbA* regulatory sequence in leaves [136]. The only other heterologous UTR that expressed transgenes at high levels is that from the *Bacillus thuringiensis* (*Bt*) operon [137]. Use of this operon produced the highest level of insecticidal toxin protein (52 % TLP) ever reported in the published literature [137]. These high levels of toxin accumulation in chloroplasts could result from the combination of high-level expression and protein stability; the Bt protein formed cuboidal crystals within chloroplasts (Fig. 4e) due to co-expression of a chaperone



**Fig. 4** Engineering the chloroplast genome to confer biotic/abiotic stress tolerance or expression of high-value products. **a–d** Industrial production of blood clotting factor IX (FIX) bioencapsulated in lettuce plants in a hydroponic cGMP facility. **a** Biomass production of FIX-expressing plants. **b–d** Steps in capsule preparation. After harvesting and lyophilization of fresh leaves, freeze-dried FIX-accumulating leaves were powdered and prepared as capsules [6]. **e–g** Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to the formation of the Bt insecticidal crystal protein. In bioassays with the *Helicoverpa zea*, **f** eating the transplastomic leaf kills the caterpillar, while **g** the control leaf is consumed by the growing caterpillar [137]. **h** Ultrastructure of the chloroplast envelope membrane of transplastomic  $\gamma$ -tocopherol methyltransferase ( $\gamma$ -TMT) tobacco plants shows the formation of multiple layers of inner envelope membranes as the result of  $\gamma$ -TMT overexpression [153]. **i, j** Expression of *BETAINE ALDEHYDE DEHYDROGENASE* (*BADH*) in carrot plants. **i** Transgenic carrot plants thrived in soil irrigated with 400 mM sodium chloride, whereas untransformed carrot plants showed retarded growth in the presence of salt. **j** Carrot roots from transplastomic plants [135]. **k** Phenotypes of tomato fruits from transplastomic tomato plants expressing lycopene  $\beta$ -cyclase transgenes compared with wild-type plants. Fruits were harvested at different ripening stages. Orange color of ripe fruits indicates efficient conversion of red lycopene into orange  $\beta$ -carotene (provitamin A) [154]



that facilitates folding. When fed, transplastomic leaves, cotton bollworm (*Helicoverpa* sp.) were killed with a single bite of leaf and insects that had 40,000-fold increased resistance to Bt were also killed (Fig. 4f, g). Nevertheless, expression of this transgene in tomato fruit is very poor [133, 138, 139] and further research is needed to enhance transgene expression in fruits.

## Engineering the chloroplast genomes for biotechnology applications

### Conferring stress tolerance

In the past decade, chloroplast genetic engineering has focused primarily on the overexpression of target genes with the potential to enhance biotic stress tolerance, which is very important for plant protection and yield enhancement. Yield loss due to insect pests can be very serious in many countries. In addition to cotton bollworm resistance conferred by hyper-expression of Bt protein in chloroplasts [137], there are many other striking recent examples of improved biotic stress tolerance. Retrocyclin-101 and Protegrin-1 protect against *Erwinia* soft rot and tobacco mosaic virus (TMV), which result in yield loss in several cultivated crops [140]. Whitefly and aphid resistance has been accomplished by expressing  $\beta$ -glucosidase [141], which releases insecticidal sugar esters from hormone conjugates. Multiple resistances against aphids, whiteflies, lepidopteran insects, and bacterial and viral pathogens were achieved by expressing the *Pinellia ternata agglutinin* (PTA) gene in the chloroplast genome [142]. More than 40 transgenes have been stably integrated into and expressed within the chloroplast genome, conferring important agronomic traits, including insect resistance in edible crops cabbage (*Brassica oleracea*) [143], soybean [144, 145], and eggplant (*Solanum melongena*) [146].

More recently, scientists have begun to explore new strategies to downregulate specific target genes. One such approach is to express double-stranded RNAs (dsRNAs) within the chloroplast genome and to use RNA interference (RNAi) to confer the desired agronomic traits, mainly resistance to insects that cause severe yield loss. This strategy has been demonstrated by expressing long or short dsRNAs that activate RNAi and disrupt target genes in insects, providing efficient protection against insects without the need for chemical pesticides. One such example is the suppression of three essential proteins required for insect survival—lepidopteran chitin synthase (Chi), cytochrome P450 monooxygenase (P450), and V-ATPase—using dsRNAs in the tobacco chloroplast system [147]. Each dsRNA was expressed independently in chloroplasts and leaves were fed to insects. The transcription level of target genes in *Helicoverpa* insects decreased to almost undetectable levels in the midgut, resulting in a significant

reduction in the net weight of larvae and in pupation rate [147]. Transplastomic potato plants producing  $\beta$ -actin-targeting long dsRNA were lethal to Colorado potato beetle (*Leptinotarsa decemlineata*) larvae, providing yet another crop protection mechanism [148].

### Synthesis of enzymes and biomaterials

In addition to improved resistance against both biotic and abiotic stress, the chloroplast genome has been engineered to produce useful enzymes, biomaterials, and biofuels, or even to enhance biomass. The first report of metabolic engineering using chloroplast genomes produced the highest level of the poly(p-hydroxybenzoic acid (pHBA) polymer (25 % dry weight) in normal healthy plants despite the diversion of a major metabolic intermediate [149]. The first use of plant-derived enzyme cocktails for the production of fermentable sugars from lignocellulosic biomass was accomplished recently [150]. Unlike the single biofuel enzymes previously expressed in chloroplasts, nine different genes from bacteria or fungi were expressed in *E. coli* or tobacco chloroplasts using a new technique that enabled the insertion of fungal genes with several introns, eliminating the need to prepare cDNA libraries. Industrial fermentation systems are currently limited by high cost and low production capacity; chloroplast-derived enzyme cocktails offer several striking advantages, including significantly reduced cost, improved stability of chloroplast-derived enzymes, and no need for enzyme purification. Interestingly, expression of  $\beta$ -glucosidase released hormones from conjugates, resulting in elevated phytohormone levels and increased biomass [141], an unexpected outcome of enzyme expression.

### Enhancing nutrition

Seed oils, such as those from soybean, rapeseed (*Brassica napus*), and maize, are the major dietary source of vitamin E. They have very low  $\alpha$ -tocopherol content but relatively high levels of  $\gamma$ -tocopherol. Only a few seed oils, such as sunflower (*Helianthus annuus*) seed oil, contain high levels of  $\alpha$ -tocopherol, an important precursor of vitamin E [151].  $\gamma$ -Tocopherol is the biosynthetic precursor of  $\alpha$ -tocopherol, suggesting that the  $\alpha$ -tocopherol biosynthetic pathway catalyzed by  $\gamma$ -tocopherol methyl transferase ( $\gamma$ -TMT) is the rate-limiting step [152]. Engineering of the  $\gamma$ -tmt gene into the chloroplast genome resulted in the formation of multiple layers of the inner chloroplast envelope (Fig. 4h) due to  $\gamma$ -TMT overexpression, with around tenfold higher conversion of  $\gamma$ -tocopherol to  $\alpha$ -tocopherol in seeds [153]. Likewise, introducing *lycopene  $\beta$ -cyclase* genes into the tomato plastid genome increased the conversion of lycopene into provitamin A ( $\beta$ -carotene), with obvious phenotypic changes (Fig. 4k) [154].

### Biopharmaceuticals

At present, protein drugs are extremely expensive; for example, >90 % of the global population cannot afford insulin, a drug needed to treat the global diabetes epidemic. The high cost of protein drugs is due to their production in prohibitively expensive fermentation systems (which cost more than \$450–700 million to build depending on their capacity [155, 156]), prohibitively expensive purification from host proteins, the need for refrigerated storage and transport, and the short shelf-life of the final product. Protein drugs made by plant chloroplasts overcome most of these challenges because they do not require expensive fermentation systems and are produced in federal drug administration (FDA)-approved hydroponic greenhouses (Fig. 4a) [157]. Lettuce leaves expressing protein drugs are lyophilized and stored indefinitely at ambient temperature without losing their efficacy (Fig. 4b–d) [6]. The plant cell wall protects protein drugs from acids and enzymes in the stomach because human enzymes do not digest plant cell wall glycans. Human gut microbes, however, have evolved to break down every glycosidic bond in the plant cell wall and therefore release the protein drug into the gut lumen, directing its delivery to the blood or immune system [158, 159].

Oral delivery of several human therapeutic proteins expressed in chloroplasts is highly efficacious in the treatment of several human diseases, including diabetes, cardiovascular disease, pulmonary hypertension, and Alzheimer's disease. Most proteins were expressed in tobacco chloroplasts for initial evaluation and were subsequently expressed in lettuce chloroplasts for advancing them to the clinic. Oral delivery of exendin-4, which modulates the secretion of insulin in a glucose-dependent manner, lowered glucose in diabetic animals by stimulating the production of insulin in a manner similar to that of the injectable drug [160]. Oral delivery of angiotensin-converting enzyme 2 (ACE2) and angiotensin (Ang) (1–7) significantly improved cardiopulmonary structure and function, decreased elevated right ventricular systolic blood pressure, and improved pulmonary blood flow in animals with induced pulmonary hypertension [161]. Oral delivery of plant cells expressing ACE2 and Ang (1–7) also reduced endotoxin-induced uveitis (EIU) and dramatically decreased cellular infiltration and retinal vasculitis, as well as damage and folding in experimental autoimmune uveoretinitis [158]. It is also possible to orally deliver protein drugs across the blood–brain barrier to the Alzheimer's brain to remove plaques [162].

The first industrial-scale production of human blood clotting factor in a cGMP facility was reported recently [6] (Fig. 4a–d). In a 1000 ft<sup>2</sup> hydroponic cGMP facility, it is possible to produce up to 30,000 doses for a 20-kg

pediatric patient. Clotting factor made in lettuce was stable for up to 2 years when lyophilized cells were stored at ambient temperature, completely eliminating the need for the cold chain. This enables the first commercial development of an oral drug and addresses the extremely expensive purification, cold storage and transportation, and short shelf-life of current protein drugs. Oral delivery of a broad dose range was effective in the prevention of antibody formation after injection of clotting factor IX (FIX), further facilitating human clinical studies.

### Vaccines against infectious diseases

The current iteration of vaccines, using attenuated bacteria or viruses, offer protection against major infectious diseases but they also present major challenges. For example, the oral polio vaccine that is used around the globe has caused severe polio resulting from mutations and recombination with other viruses [163]. In addition, all current vaccines require cold storage and transportation, making distribution in developing countries a major challenge. Many of these challenges can be overcome by using chloroplasts.

One successful chloroplast-derived vaccine conferred dual immunity against cholera and malaria in animal studies [164]. Cholera is a major disease causing high mortality, with the only licensed vaccine being not only expensive but also limited in its duration of protection. No vaccine is currently available for malaria. The cholera toxin-B subunit (CTB) of *Vibrio cholerae* was fused to the malarial vaccine antigen apical membrane antigen-1 (AMA1) and merozoite surface protein-1 (MSP1) and expressed in lettuce or tobacco chloroplasts. While no suitable models exist to test human malaria, a cholera toxin challenge using mice immunized with chloroplast-expressed CTB was highly effective and provided the longest duration of protection in the published literature [164]. These early results show that chloroplasts are ideal for producing low-cost booster vaccines against several infectious diseases [165] for which the global population has been primed previously (Table 2), but lack of an oral priming strategy is still a major limitation in this field.

### Moving forward

It is amazing that the chloroplast genome can express >120 foreign genes from different organisms, including bacteria, viruses, fungi, animals, and humans. The insertion of commercially useful traits, including herbicide and insect resistance, into soybean resulted in high-level expression and superior transgene containment, with no antibiotic selectable markers; but even so, these lines were not developed commercially. Nevertheless, recurring concerns about insect resistance

against biopesticides have resulted in new USDA requirements on planting *Bt* corn [122], which may eventually require utilization of the transplastomic approach to confer agronomic traits. The nuclear transgenic approach is inadequate to develop products when higher-level transgene expression is a requirement. Thus, chloroplast transformation has a unique advantage in advancing the field of molecular farming for the production of vaccines, biopharmaceuticals, or other bio-products.

Although products with high-level expression have now advanced to the clinic or are in commercial development, a better understanding of chloroplast translation is required to improve several other gene products. The availability of chloroplast genome sequences should help in the development of codon optimization programs using highly expressed chloroplast genes, but among the ~3000 cultivated crops, sequenced chloroplast genomes are available for crops from fewer than 70 genera. Major funding agencies have not supported crop chloroplast genome sequence projects because of the misconception that all chloroplast genomes are similar, as evidenced by the publication of fewer than ten crop chloroplast genome sequences between 1986 and 2004. This review illustrates the importance of sequencing more crop chloroplast genomes for various biotechnology applications. Furthermore, new selectable markers are needed to transform the chloroplast genomes of cereals, which has been elusive for the past two decades.

Chloroplast genome sequences will be valuable assets in herbal medicine. Most medicinal plants are rare species and very little information is available to confirm their identity. DNA barcodes derived from chloroplast genomes will be useful for identifying varieties and resources; this concept is also valuable in the identification of the origin of cultivated crops and their close relatives to enhance breeding or transfer of useful traits. Molecular techniques to sequence the genomes of single chloroplasts could help to eliminate chloroplast-like sequences that are present in the mitochondrial or nuclear genome. The ability to sequence chloroplast genomes using minimal leaf materials could help us to understand variations in different segments of a variegated leaf in horticultural crops. Further, determining complete chloroplast genome sequences from fossils or recently extinct plants could shed more light on chloroplast genome evolution; help us to understand these species' inadequate fitness to cope with environmental changes; and help us to build new phylogenetic trees. The technology for isolating DNA from fossils is already available [166–168]. All of these goals can be accomplished with less expensive and more accurate genome sequences, utilizing longer read sequencing technology and new bioinformatics tools.

## Additional file

**Additional file 1: Table S1.** The chloroplast genes which are absent in specific species, their knock out phenotypes and transfer to nuclear genomes. (DOCX 23 kb)

## Abbreviations

ACE2, Angiotensin-converting enzyme 2; BAC, bacterial artificial chromosome; *Bt*, *Bacillus thuringiensis*; CTB, Cholera toxin B subunit; cGMP, Current Good Manufacturing Processes; CTB, Cholera toxin B subunit; dsRNA, double-stranded RNA; FIX, clotting factor IX; *infA*, translation initiation factor 1; IR, inverted repeat; LSC, large single copy; *ndh*, NAD(P)H dehydrogenase; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; RNAi, RNA interference; SSC, small single copy;  $\gamma$ -TMT,  $\gamma$ -Tocopherol methyltransferase; TSP, total soluble protein; UTR, untranslated region; Ang (1–7), Angiotensin (1–7)

## Acknowledgements

We acknowledge valuable contributions from editors of *Genome Biology* (Drs Dominique Morneau and Ripudaman Bains) to enhance the flow and presentation to non-specialists and reviewers for their critical comments.

## Funding

Research from the Daniell Laboratory included in this review was supported by the Bill and Melinda Gates Foundation (OPP1031406), the National Institutes of Health (NIH; R01 HL107904, HL109442, GM 63879, EY 024564), Novo Nordisk, Bayer, and Department of Energy ARPA-E grants to Henry Daniell.

## Authors' contributions

HD and CSL wrote this review. MY assembled Table 2 and Figs. 3 and 4 with guidance from HD. WJC assembled Table 1 and Figs. 1 and 2 with guidance from CSL. All authors read and approved the final manuscript.

## Competing interests

Henry Daniell, as a pioneer in the field of chloroplast genetic engineering, has several patents in this field but has no financial conflict of interest to declare. A complete list of published patents is available at the Google Scholar weblink: <http://scholar.google.com/citations?user=7sow4jwAAAAJ&hl=en>

## Author details

<sup>1</sup>Department of Biochemistry, School of Dental Medicine, University of Pennsylvania, South 40th St, Philadelphia, PA 19104-6030, USA. <sup>2</sup>Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan.

Published online: 23 June 2016

## References

- Bobik K, Burch-Smith TM. Chloroplast signaling within, between and beyond cells. *Front Plant Sci.* 2015;6:781.
- Daniell H, Chan H-T, Pasoreck EK. Vaccination through chloroplast genetics: affordable protein drugs for the prevention and treatment of inherited or infectious human diseases. *Annu Rev Genet.* 2016. In Press.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 1986;5:2043–9.
- Wambugu P, Brozynska M, Furtado A, Waters D, Henry R. Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci Rep.* 2015;5:13957.
- Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol J.* 2016;14:1070–85.
- Su J, Zhu L, Sherman A, Wang X, Lin S, Kamesh A, et al. Low cost industrial production of coagulation factor IX bioencapsulated in lettuce cells for oral tolerance induction in hemophilia B. *Biomaterials.* 2015;70:84–93.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol Phylogenet Evol.* 2008;48:1204–17.
- Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, et al. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet.* 2007;115:571–90.

9. Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics*. 2006;7:222.
10. Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnol J*. 2007;5:339–53.
11. Lee SB, Kaittani 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.
12. Bausher MG, Singh ND, Lee SB, Jansen RK, Daniell H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol*. 2006;6:21.
13. Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, et al. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor Appl Genet*. 2006;112:1503–18.
14. Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol Biol Evol*. 2011;28:835–47.
15. Wolfe KH, Morden CW, Palmer JD. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A*. 1992;89:10648–52.
16. Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ. The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res Notes*. 2008;1:61.
17. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res*. 1999;6:283–90.
18. Saski 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;59:309–22.
19. Jansen RK, Kaittani C, Saski C, Lee SB, 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.
20. Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saski C, Jansen RK. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor Appl Genet*. 2008;116:723–37.
21. Wu J, Liu B, Cheng F, Ramchiary N, Choi SR, Lim YP, Wang X-W. Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. *Front Plant Sci*. 2012;3:234.
22. Grivet D, Heinze B, Vendramin G, Petit R. Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. *Mol Ecol Notes*. 2001;1:345–9.
23. Goremykin V, Hirsch-Ernst K, Wölfl S, Hellwig F. The chloroplast genome of the 'basal' angiosperm *Elaeagnus argentea*—structural and phylogenetic analyses. *Plant Syst Evol*. 2003;242:119–35.
24. Goremykin V, Hirsch-Ernst K, Wölfl 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;21:1445–54.
25. Goremykin V, Hirsch-Ernst K, Wölfl S, Hellwig FH. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol*. 2003;20:1499–505.
26. Dhingra A, Foltz KM. ASAP: amplification, sequencing & annotation of plastomes. *BMC Genomics*. 2005;6:176.
27. Heinze B. A database of PCR primers for the chloroplast genomes of higher plants. *Plant Methods*. 2007;3:4.
28. Wu FH, Kan DP, Lee SB, Daniell H, Lee YW, Lin CC, et al. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiol*. 2009;29:847–56.
29. Leseberg CH, Duvall MR. The complete chloroplast genome of *Coix lacryma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *J Mol Evol*. 2009;69:311–8.
30. Dong W, Xu C, Cheng T, Lin K, Zhou S. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales. *Genome Biol Evol*. 2013;5:989–97.
31. Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, Skyabin KG. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J Mol Evol*. 2008;66:555–64.
32. Lin CS, Chen JJ, Huang YT, Chan MT, Daniell H, Chang WJ, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci Rep*. 2015;5:9040.
33. Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Foltz KM, Soltis DE. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol*. 2006;6:17.
34. Wang W, Messing J. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS One*. 2011;6:e24670.
35. Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res*. 2008;36:e122.
36. Pan IC, Liao DC, Wu FH, Daniell H, Singh ND, Chang C, et al. Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical orchid breeding. *PLoS One*. 2012;7:e34738.
37. Atherton RA, McComish BJ, Shepherd LD, Berry LA, Albert NW, Lockhart PJ. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. *Plant Methods*. 2010;6:22.
38. Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, et al. Organellar genomes of white spruce (*Picea glauca*): assembly and annotation. *Genome Biol Evol*. 2016;8:29–41.
39. Ferrarini M, Moretto M, Ward JA, Šurbanovski N, Stevanović V, Giongo L, et al. An evaluation of the PacBio RS platform for sequencing and *de novo* assembly of a chloroplast genome. *BMC Genomics*. 2013;14:670.
40. Wu Z, Gui S, Quan Z, Pan L, Wang S, Ke W, et al. A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. *BMC Plant Biol*. 2014;14:289.
41. Li Q, Li Y, Song J, Xu H, Xu J, Zhu Y, et al. High-accuracy *de novo* assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. *New Phytol*. 2014;204:1041–9.
42. Redwan R, Saidin A, Kumar S. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biol*. 2015;15:196.
43. Chen X, Li Q, Li Y, Qian J, Han J. Chloroplast genome of *Aconitum barbatum* var. *puberulum* (Ranunculaceae) derived from CCS reads using the PacBio RS platform. *Frontiers Plant Sci*. 2015;6:42.
44. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323:133–8.
45. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*. 2012;7:e47768.
46. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:563–9.
47. Oldenburg DJ, Bendich AJ. DNA maintenance in plastids and mitochondria of plants. *Frontiers Plant Sci*. 2015;6:883.
48. Oldenburg DJ, Bendich AJ. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Curr Genet*. 2016;62:431–42.
49. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A*. 2007;104:19369–74.
50. Pillion Y, Chase MW. Taxonomic exaggeration and its effects on orchid conservation. *Conserv Biol*. 2007;21:263–5.
51. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol*. 2006;23:279–91.
52. Jheng CF, Chen TC, Lin JY, Chen TC, Wu WL, Chang CC. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci*. 2012;190:62–73.
53. Kim GB, Kwon Y, Yu HJ, Lim KB, Seo JH, Mun JH. The complete chloroplast genome of *Phalaenopsis* 'Tiny Star'. *Mitochondrial DNA*. 2016;27:1300–2.
54. Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H, et al. Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of



- molecular markers for identification and breeding in *Oncidiinae*. *BMC Plant Biol.* 2010;10:68.
55. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol.* 2013;13:84.
  56. Luo J, Hou BW, Niu ZT, Liu W, Xue QY, Ding XY. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One.* 2014;9, e99016.
  57. Li J, Chen C, Wang ZZ. The complete chloroplast genome of the *Dendrobium strongylanthum* (Orchidaceae: Epidendroideae). *Mitochondrial DNA.* 2016;27:3048–9.
  58. da Rocha Perini V, Leles B, Furtado C, Prosdociimi F. Complete chloroplast genome of the orchid *Cattleya crispata* (Orchidaceae: Laeliinae), a Neotropical rupicolous species. *Mitochondrial DNA.* 2015. doi:10.3109/19401736.2014.1003850.
  59. Tsitrone A, Kirkpatrick M, Levin DA. A model for chloroplast capture. *Evolution.* 2003;57:1776–82.
  60. Soltis DE, Kuzoff RK. Discordance between nuclear and chloroplast phylogenies in the heuchera group (Saxifragaceae). *Evolution.* 1995;49:727–42.
  61. Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol Trends Plants.* 1991;5:65–84.
  62. 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;251:614–28.
  63. Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, et al. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Mol Genet Genomics.* 2002;266:740–6.
  64. Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B. Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PLoS One.* 2014;9, e85761.
  65. Brozynska M, Omar ES, Furtado A, Crayn D, Simon B, Ishikawa R, Henry RJ. Chloroplast genome of novel rice germplasm identified in northern Australia. *Tropical Plant Biol.* 2014;7:111–20.
  66. Waters DL, Nock CJ, Ishikawa R, Rice N, Henry RJ. Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecol Evol.* 2012;2:211–7.
  67. Sotowa M, Ootsuka K, Kobayashi Y, Hao Y, Tanaka K, Ichitani K, et al. Molecular relationships between Australian annual wild rice, *Oryza meridionalis*, and two related perennial forms. *Rice.* 2013;6:26.
  68. Xu Q, Xiong G, Li P, He F, Huang Y, Wang K, et al. Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: origin and evolution of allotetraploids. *PLoS One.* 2012;7, e37128.
  69. Ibrahim RI, Azuma J, Sakamoto M. Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among nine dicot plants. *Genes Genet Syst.* 2006;81:311–21.
  70. Li P, Li Z, Liu H, Hua J. Cytoplasmic diversity of the cotton genus as revealed by chloroplast microsatellite markers. *Genet Resour Crop Evol.* 2014;61:107–19.
  71. Pickersgill B, Deboucq DG. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theor Appl Genet.* 2005;110:432–44.
  72. Palmer JD, Osorio B, Thompson WF. Evolutionary significance of inversions in legume chloroplast DNAs. *Curr Genet.* 1988;14:65–74.
  73. Sherman-Broyles S, Bombarely A, Grimwood J, Schmutz J, Doyle J. Complete plastome sequences from *Glycine syndetika* and six additional perennial wild relatives of soybean. G3 (Bethesda). 2014;4:2023–33.
  74. Sabir J, Schwarz E, Ellison N, Zhang J, Baeshen NA, Mutwakil M, et al. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnol J.* 2014;12:743–54.
  75. Cai Z, Guisinger M, Kim H-G, Ruck E, Blazier JC, McMurtry V, et al. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol.* 2008;67:696–704.
  76. Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, et al. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 2010;20:1700–10.
  77. Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, Vasiliev GV. Nuclear-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastid candidate genes encoding acetyl-CoA carboxylase subunits. *PLoS One.* 2015;10:e0119835.
  78. Guo X, Castillo-Ramírez S, González V, Bustos P, Fernández-Vázquez JL, Santamaría RI, et al. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics.* 2007;8:228.
  79. Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaisaisanwong P, Yoocha T, Jomchai N, Tragoonrun S. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Res.* 2010;17:11–22.
  80. Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, et al. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Ann Bot.* 2014;113:1197–210.
  81. Kazakoff SH, Imelfort M, Edwards D, Koehorst J, Biswas B, Batley J, et al. Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feedstock tree *Pongamia pinnata*. *PLoS One.* 2012;7, e51687.
  82. Schwarz EN, Ruhlman TA, Sabir JS, Hajrah NH, Alharbi NS, Al-Malki AL, et al. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J Syst Evol.* 2015;53:458–68.
  83. Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. Complete chloroplast genome sequence of Omani Lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. *PLoS One.* 2014;9, e113049.
  84. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol.* 2015;32:2015–35.
  85. Caspermeyer J. Most comprehensive study to date reveals evolutionary history of *Citrus*. *Mol Biol Evol.* 2015;32:2217–8.
  86. Chen SL, Xiao PG. Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *J Plant Res.* 2010;123:241–7.
  87. Peng L, Yamamoto H, Shikanai T. Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *Biochim Biophys Acta.* 1807;2011:945–53.
  88. Ma PF, Zhang YX, Guo ZH, Li DZ. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Sci Rep.* 2015;5:11608.
  89. Zhang YJ, Ma PF, Li DZ. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS One.* 2011;6, e20596.
  90. Burke SV, Grennan CP, Duvall MR. Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of Bambusoideae. *Am J Bot.* 2012;99:1951–61.
  91. Gao J, Li K, Gao LZ. The complete chloroplast genome sequence of the *Bambusa multiplex* (Poaceae: Bambusoideae). *Mitochondrial DNA.* 2014;18:1–3.
  92. Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evol Biol.* 2015;15:50.
  93. Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst Biol.* 2014;63:933–50.
  94. Smith DR. Mitochondrion-to-plastid DNA, transfer: it happens. *New Phytol.* 2014;202:736–8.
  95. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 2004;5:123–35.
  96. Li H, Chiu CC. Protein transport into chloroplasts. *Annu Rev Plant Biol.* 2010; 61:157–80.
  97. Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol Biol Evol.* 2014;31:3095–112.
  98. Tiller N, Bock R. The translational apparatus of plastids and its role in plant development. *Mol Plant.* 2014;7:1105–20.
  99. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 2010;6, e1000834.
  100. Smith DR, Crosby K, Lee RW. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biol Evol.* 2011;3:365–71.



101. Cummings HS, Hershey JWB. Translation initiation-factor *IF1* is essential for cell viability in *Escherichia coli*. *J Bacteriol*. 1994;176:198–205.
102. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*. 2001;13:645–58.
103. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J*. 1991;10:3073–8.
104. Park S, Jansen RK, Park S. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictrioideae. *BMC Plant Biol*. 2015;15:40.
105. Ueda M, Fujimoto M, Arimura S, Murata J, Tsutsumi N, Kadowaki K. Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene*. 2007;402:51–6.
106. Cusack BP, Wolfe KH. When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet*. 2007;23:270–2.
107. Ueda M, Kuniyoshi T, Yamamoto H, Sugimoto K, Ishizaki K, Kohchi T, et al. Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in *Marchantia polymorpha*. *Plant J*. 2012;72:683–93.
108. Munekage Y, Hashimoto M, Miyake C, Tomizawa K-I, Endo T, Tasaka M, Shikanai T. Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*. 2004;429:579–82.
109. Peltier G, Courmac L. Chlororespiration. *Annu Rev Plant Biol*. 2002;53:523–50.
110. 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;91:9794–8.
111. Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JL, Eguiarte LE, et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *Am J Bot*. 2015;102:1115–27.
112. McCoy SR, Kuehl JV, Boore JL, Raubeson LA. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol Biol*. 2008;8:130.
113. Braukmann TWA, Kuzmina M, Stefanović S. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr Genet*. 2009;55:323–37.
114. Blazier J, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. 2011;76:263–72.
115. Weng M-L, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Mol Biol Evol*. 2014;31:645–59.
116. Ranade SS, García-Gil MR, Rosselló JA. Non-functional plastid *ndh* gene fragments are present in the nuclear genome of Norway spruce (*Picea abies* L. Karsch): insights from in silico analysis of nuclear and organellar genomes. *Mol Genet Genomics*. 2016;291:935–41.
117. Ruhlman TA, Chang WJ, Chen JJ, Huang YT, Chan MT, Zhang J, et al. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. *BMC Plant Biol*. 2015;15:100.
118. Takabayashi A, Endo T, Shikanai T, Sato F. Post-illumination reduction of the plastoquinone pool in chloroplast transformants in which chloroplastic NAD(P)H dehydrogenase was inactivated. *Biosci Biotechnol Biochem*. 2002;66:2107–11.
119. Burrows PA, Sazanov LA, Svab Z, Maliga P, Nixon PJ. Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid *ndh* genes. *EMBO J*. 1998;17:868–76.
120. Kofer W, Koop H-U, Wanner G, Steinmüller K. Mutagenesis of the genes encoding subunits A, C, H, I, J and K of the plastid NAD(P)H-plastoquinone-oxidoreductase in tobacco by polyethylene glycol-mediated plastome transformation. *Mol Gen Genet*. 1998;258:166–73.
121. Shikanai T, Endo T, Hashimoto T, Yamada Y, Asada K, Yokota A. Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc Natl Acad Sci U S A*. 1998;95:9705–9.
122. Jin S, Daniell H. The engineered chloroplast genome just got smarter. *Trends Plant Sci*. 2015;20:622–40.
123. Daniell H. Transgene containment by maternal inheritance: effective or elusive? *Proc Natl Acad Sci U S A*. 2007;104:6879–80.
124. Dheeraj V, Daniell H. Chloroplast vector systems for biotechnology applications. *Plant Physiol*. 2007;145:1129–43.
125. Dunne A, Maple-Grødem J, Gargano D, Haslam RP, Napier JA, Chua NH, et al. Modifying fatty acid profiles through a new cytokinin-based plastid transformation system. *Plant J*. 2014;80:1131–8.
126. Kavanagh TA, Thanh ND, Lao NT, McGrath N, Peter SO, Horváth EM, et al. Homeologous plastid DNA transformation in tobacco is mediated by multiple recombination events. *Genetics*. 1999;152:1111–22.
127. Guda C, Lee SB, Daniell H. Stable expression of a biodegradable protein-based polymer in tobacco chloroplasts. *Plant Cell Rep*. 2000;19:257–62.
128. Ruhlman T, Verma D, Samson N, Daniell H. The role of heterologous chloroplast sequence elements in transgene integration and expression. *Plant Physiol*. 2010;152:2088–104.
129. Daniell H, Datta R, Varma S, Gray S, Lee SB. Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat Biotechnol*. 1998;16:345–8.
130. Krichevsky A, Meyers B, Vainstein A, Maliga P, Citovsky V. Autoluminescent plants. *PLoS One*. 2010;5, e15461.
131. Daniell H, Vivekananda J, Nielsen B, Ye G, Tewari K, Sanford J. Transient foreign gene expression in chloroplasts of cultured tobacco cells after biolistic delivery of chloroplast vectors. *Proc Natl Acad Sci U S A*. 1990;87:88–92.
132. Klein RR, Mullet JE. Control of gene expression during higher plant chloroplast biogenesis. Protein synthesis and transcript levels of *psbA*, *psaA-psaB*, and *rbcL* in dark-grown and illuminated barley seedlings. *J Biol Chem*. 1987;262:4341–8.
133. Wurbs D, Ruf S, Bock R. Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *Plant J*. 2007;49:276–88.
134. McBride KE, Schaaf DJ, Daley M, Stalker DM. Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proc Natl Acad Sci U S A*. 1994;91:7301–5.
135. 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;136:2843–54.
136. Dhingra A, Portis Jr AR, 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 U S A*. 2004;101:6315–20.
137. 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;19:71–4.
138. Ruf S, Hermann M, Berger IJ, Carrer H, Bock R. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat Biotechnol*. 2001;19:870–5.
139. Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, et al. High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. *Plant Biotechnol J*. 2008;6:897–913.
140. Lee SB, Li B, Jin S, Daniell H. Expression and characterization of antimicrobial peptides Retrocyclin-101 and Protegrin-1 in chloroplasts to control viral and bacterial infections. *Plant Biotechnol J*. 2011;9:100–15.
141. Jin S, Kanagaraj A, Verma D, Lange T, Daniell H. Release of hormones from conjugates: chloroplast expression of beta-glucosidase results in elevated phytohormone levels associated with significant increase in biomass and protection from aphids or whiteflies conferred by sucrose esters. *Plant Physiol*. 2011;155:222–35.
142. Jin S, Zhang X, Daniell H. *Pinellia ternata* agglutinin expression in chloroplasts confers broad spectrum resistance against aphid, whitefly, Lepidopteran insects, bacterial and viral pathogens. *Plant Biotechnol J*. 2012;10:313–27.
143. Liu CW, Lin CC, Chen JJ, Tseng MJ. Stable chloroplast transformation in cabbage (*Brassica oleracea* L. var. *capitata* L.) by particle bombardment. *Plant Cell Rep*. 2007;26:1733–44.
144. Dufourmantel N, Pellissier B, Garçon F, Peltier G, Ferullo JM, Tissot G. Generation of fertile transplastomic soybean. *Plant Mol Biol*. 2004;55:479–89.
145. Dufourmantel N, Tissot G, Goutorbe F, Garçon F, Muhr C, Jansens S, et al. Generation and analysis of soybean plastid transformants expressing *Bacillus thuringiensis* Cry1Ab protoxin. *Plant Mol Biol*. 2005;58:659–68.
146. Singh AK, Verma SS, Bansal KC. Plastid transformation in eggplant (*Solanum melongena* L.). *Transgenic Res*. 2010;19:113–9.
147. Jin S, Singh ND, Li L, Zhang X, Daniell H. Engineered chloroplast dsRNA silences cytochrome p450 monooxygenase, V-ATPase and chitin synthase genes in the insect gut and disrupts *Helicoverpa armigera* larval development and pupation. *Plant Biotechnol J*. 2015;13:435–46.

148. Zhang J, Khan SA, Hasse C, Ruf S, Heckel DG, Bock R. Pest control. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science*. 2015;347:991–4.
149. Viitanen PV, Devine AL, Khan MS, Deuel DL, Van Dyk DE, Daniell H. Metabolic engineering of the chloroplast genome using the *Escherichia coli* *ubiC* gene reveals that chorismate is a readily abundant plant precursor for p-hydroxybenzoic acid biosynthesis. *Plant Physiol*. 2004;136:4048–60.
150. Verma D, Kanagaraj A, Jin SX, Singh ND, Kolattukudy PE, Daniell H. Chloroplast-derived enzyme cocktails hydrolyse lignocellulosic biomass and release fermentable sugars. *Plant Biotechnol J*. 2010;8:332–50.
151. Schneider C. Chemistry and biology of vitamin E. *Mol Nutr Food Res*. 2005;49:7–30.
152. Shintani D, DellaPenna D. Elevating the vitamin E content of plants through metabolic engineering. *Science*. 1998;282:2098–100.
153. Jin S, Daniell H. Expression of gamma-tocopherol methyltransferase in chloroplasts results in massive proliferation of the inner envelope membrane and decreases susceptibility to salt and metal-induced oxidative stresses by reducing reactive oxygen species. *Plant Biotechnol J*. 2014;12:1274–85.
154. Apel W, Bock R. Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. *Plant Physiol*. 2009;151:59–66.
155. Grabowski H, Cockburn I, Long G. The market for follow-on biologics: how will it evolve? *Health Aff*. 2006;25:1291–301.
156. Spök A, Karner S, Stein AJ, Rodríguez-Cerezo E. Plant molecular farming. Opportunities and challenges. *JRC Scientific and Technical Reports* 2008. <http://ftp.jrc.es/EURdoc/JRC43873.pdf>. Accessed 17 May 2016.
157. Holtz BR, Berquist BR, Bennett LD, Kommineni VJ, Munigunt RK, White EL, et al. Commercial-scale biotherapeutics manufacturing facility for plant-made pharmaceuticals. *Plant Biotechnol J*. 2015;13:1180–90.
158. Kwon KC, Daniell H. Low-cost oral delivery of protein drugs bioencapsulated in plant cells. *Plant Biotechnol J*. 2015;13:1017–22.
159. El Kaoutari A, Armougom F, Gordon JL, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol*. 2013;11:497–504.
160. Kwon KC, Nityanandam R, New JS, Daniell H. Oral delivery of bioencapsulated exendin-4 expressed in chloroplasts lowers blood glucose level in mice and stimulates insulin secretion in beta-TC6 cells. *Plant Biotechnol J*. 2013;11:77–86.
161. Shenoy V, Kwon KC, Rathinasabapathy A, Lin SN, Jin GY, Song CJ, et al. Oral delivery of angiotensin-converting enzyme 2 and angiotensin-(1–7) bioencapsulated in plant cells attenuates pulmonary hypertension. *Hypertension*. 2014;64:1248–59.
162. Kohl N, Westerveld DR, Ayache AC, Verma A, Shil P, Prasad T, et al. Oral delivery of bioencapsulated proteins across blood–brain and blood-retinal barriers. *Mol Ther*. 2014;22:535–46.
163. Chan HT, Daniell H. Plant-made oral vaccines against human infectious diseases—are we there yet? *Plant Biotechnol J*. 2015;13:1056–70.
164. Davoodi-Semiromi A, Schreiber M, Nalapalli S, Verma D, Singh ND, Banks RK, et al. Chloroplast-derived vaccine antigens confer dual immunity against cholera and malaria by oral or injectable delivery. *Plant Biotechnol J*. 2010;8:223–42.
165. Lakshmi PS, Verma D, Yang X, Lloyd B, Daniell H. Low cost tuberculosis vaccine antigens in capsules: expression in chloroplasts, bio-encapsulation, stability and functional evaluation in vitro. *PLoS One*. 2013;8, e54708.
166. Hagelberg E, Hofreiter M, Keyser C. Ancient DNA: the first three decades. *Philos Trans R Soc Lond B Biol Sci*. 2015;370:20130371.
167. Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, et al. Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature*. 1990;344:656–8.
168. Rosselló JA. The never-ending story of geologically ancient DNA: was the model plant *Arabidopsis* the source of Miocene Dominican amber? *Biol J Linn Soc*. 2014;111:234–40.
169. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol*. 2016;209:1747–56.
170. von Kohn C, Kielkowska A, Havey MJ. Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasts of onion. *Genome*. 2013;56:737–42.
171. Nashima K, Terakami S, Nishitani C, Kunihiya M, Shoda M, Takeuchi M, et al. Complete chloroplast genome sequence of pineapple (*Ananas comosus*). *Tree Genet Genomes*. 2015;11:60.
172. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Botany*. 2015;40:336–51.
173. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, et al. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One*. 2013;8, e57533.
174. 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;19:1602–12.
175. Hu Z-Y, Hua W, Huang S-M, Wang H-Z. Complete chloroplast genome sequence of rapeseed (*Brassica napus* L.) and its evolutionary implications. *Genet Resour Crop Evol*. 2011;58:875–87.
176. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*. 2014;14:1024–31.
177. Oh H, Seo B, Lee S, Ahn D-H, Jo E, Park J-K, Min G-S. Two complete chloroplast genome sequences of *Cannabis sativa* varieties. *Mitochondrial DNA*. 2016;27:2835–7.
178. Jo YD, Park J, Kim J, Song W, Hur CG, Lee YH, Kang BC. Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Rep*. 2011;30:217–29.
179. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*. 2008;452:991–6.
180. Ku C, Chung WC, Chen LL, Kuo CH. The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. *PLoS One*. 2013;8, e68518.
181. Mariac C, Scarcelli N, Pouzadou J, Barnaud A, Billot C, Faye A, et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol Ecol Resour*. 2014;14:1103–13.
182. Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ. Mutational dynamics of aroid chloroplast genomes. *Genome Biol Evol*. 2012;4:1316–23.
183. Plader W, Yukawa Y, Sugiura M, Malepszy S. The complete structure of the cucumber (*Cucumis sativus* L.) chloroplast genome: its composition and comparative analysis. *Cell Mol Biol Lett*. 2007;12:584–94.
184. Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical ginger (Zingiberales)? *Ann Bot*. 2014;113:119–33.
185. Wang B, Han L, Chen C, Wang Z. The complete chloroplast genome sequence of *Dieffenbachia seguine* (Araceae). *Mitochondrial DNA*. 2016;27:2913–4.
186. Ye CY, Lin Z, Li G, Wang YY, Qiu J, Fu F, Zhang H, et al. *Echinochloa* chloroplast genomes: insights into the evolution and taxonomic identification of two weedy species. *PLoS One*. 2014;9, e113657.
187. Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol*. 2009;52:115–24.
188. Logacheva MD, Samigullin TH, Dhirga A, Penin AA. Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. ancestrale—a wild ancestor of cultivated buckwheat. *BMC Plant Biol*. 2008;8:59.
189. Cahoon AB, Sharpe RM, Mysayphonh C, Thompson EJ, Ward AD, Lin A. The complete chloroplast genome of tall fescue (*Lolium arundinaceum*; Poaceae) and comparison of whole plastomes from the family Poaceae. *Am J Bot*. 2010;97:49–58.
190. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet*. 2011;43:109–16.
191. Dempewolf H, Kane NC, Ostevik KL, Geleta M, Barker MS, Lai Z, et al. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol Ecol Resour*. 2010;10:1048–58.
192. Timme RE, Kuehl JV, Boore JL, Jansen RK. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of

- divergent regions and categorization of shared repeats. *Am J Bot.* 2007;94:302–12.
193. Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics.* 2013;29:65–87.
  194. Sanchez-Puerta MV, Abbona CC. The chloroplast genome of *Hyoscyamus niger* and a phylogenetic study of the tribe Hyoscyameae (Solanaceae). *PLoS One.* 2014;9, e98353.
  195. Yan L, Lai X, Li X, Wei C, Tan X, Zhang Y. Analyses of the complete genome and gene expression of chloroplast of sweet potato (*Ipomoea batata*). *PLoS One.* 2015;10, e0124083.
  196. McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 2007;7:57.
  197. Kanamoto H, Yamashita A, Asao H, Okumura S, Takase H, Hattori M, et al. Efficient and stable transformation of *Lactuca sativa* L. cv. Cisco (lettuce) plastids. *Transgenic Res.* 2006;15:205–17.
  198. Mennes CB, Lam VK, Rudall PJ, Lyon SP, Graham SW, Smets EF, Merckx VS. Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). *J Biogeogr.* 2015;42:1123–36.
  199. Hand ML, Spangenberg GC, Forster JW, Cogan NO. Plastome sequence determination and comparative analysis for members of the *Lolium-Festuca* grass species complex. G3 (Bethesda). 2013;3:607–16.
  200. Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res.* 2000;7:323–30.
  201. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007;8:174.
  202. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, et al. 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;217:185–94.
  203. 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;11:247–61.
  204. Young HA, Lanzatella CL, Sarath G, Tobias CM. Chloroplast genome variation in upland and lowland switchgrass. *PLoS One.* 2011;6, e23980.
  205. Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. *BMC Plant Biol.* 2009;9:131.
  206. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, 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;23:2175–90.
  207. Jeong YM, Chung WH, Mun JH, Kim N, Yu HJ. De novo assembly and characterization of the complete chloroplast genome of radish (*Raphanus sativus* L.). *Gene.* 2014;551:39–48.
  208. Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, et al. Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLoS One.* 2011;6, e21743.
  209. Calsa Junior T, Carraro DM, Benatti MR, Barbosa AC, Kitajima JP, Carrer H. Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Curr Genet.* 2004;46:366–73.
  210. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One.* 2013;8, e57607.
  211. Yi DK, Kim KJ. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS One.* 2012;7, e35872.
  212. Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, et al. The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Rep.* 2006;25:1369–79.
  213. 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;45:307–15.
  214. Yi DK, Yang JC, So S, Joo M, Kim DK, Shin CH, et al. The complete plastid genome sequence of *Abies koreana* (Pinaceae: Abietoideae). *Mitochondrial DNA.* 2016;27:2351–3.
  215. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS One.* 2015;10, e0129347.
  216. Hsu CY, Wu CS, Chaw SM. Ancient nuclear plastid DNA in the yew family (Taxaceae). *Genome Biol Evol.* 2014;6:2111–21.
  217. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour.* 2015;15:1067–78.
  218. Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene.* 2013;528:120–31.
  219. Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, Jansen RK. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol Phylogenet Evol.* 2007;45:547–63.
  220. Wu CS, Chaw SM. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J.* 2014;12:344–53.
  221. Shi C, Liu Y, Huang H, Xia EH, Zhang HB, Gao LZ. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One.* 2013;8, e59620.
  222. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol.* 2014;14:151.
  223. Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol Evol.* 2010;2:504–17.
  224. Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol.* 2011;3:1284–95.
  225. Male PJ, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, et al. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol Ecol Resour.* 2014;14:966–75.
  226. Huang YY, Matzke AJ, Matzke M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One.* 2013;8, e74736.
  227. Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, et al. Chloroplast genome analysis of Australian eucalypts—*Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Mol Phylogenet Evol.* 2013;69:704–16.
  228. Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol.* 2008;8:70.
  229. Uthapaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoocha T, Jomchai N, et al. Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene.* 2012;500:172–80.
  230. Steane D, Jones R, Vaillancourt R. A set of chloroplast microsatellite primers for *Eucalyptus* (Myrtaceae). *Mol Ecol Notes.* 2005;5:538–41.
  231. Tangphatsornruang S, Uthapaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T, Jomchai N, Tragoonrungs S. Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene.* 2011;475:104–12.
  232. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol.* 2007;24:1161–80.
  233. Asif MH, Mantri SS, Sharma A, Srivastava A, Trivedi I, Gupta P, et al. Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet Genomes.* 2010;6:941–52.
  234. Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, et al. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biol Evol.* 2014;6:580–90.

235. Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, de Pamphilis CW, et al. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol Biol*. 2006;6:77.
236. Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y, et al. The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Front Plant Sci*. 2015;6:447.
237. Ravi V, Khurana JP, Tyagi AK, Khurana P. The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet Genomes*. 2006;3:49–59.
238. Straub SC, Moore MJ, Soltis PS, Soltis DE, Liston A, Livshultz T. Phylogenetic signal detection from an ancient rapid radiation: effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Mol Phylogenet Evol*. 2014;80:169–85.
239. Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol*. 2011;11:80.
240. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One*. 2010;5, e12762.
241. Fang MF, Wang YJ, Zu YM, Dong WL, Wang RN, Deng TT, Li ZH. The complete chloroplast genome of the Taiwan red pine *Pinus taiwanensis* (Pinaceae). *Mitochondrial DNA*. 2016;27:2732–3.
242. Vieira Ldo N, Faoro H, Rogalski M, Fraga HP, Cardoso RL, de Souza EM, et al. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. *PLoS One*. 2014;9, e90618.
243. Okumura S, Sawada M, Park YW, Hayashi T, Shimamura M, Takase H, Tomizawa K. Transformation of poplar (*Populus alba*) plastids and expression of foreign proteins in tree chloroplasts. *Transgenic Res*. 2006;15:637–46.
244. Wang S, Shi C, Gao LZ. Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. *PLoS One*. 2013;8, e73946.
245. Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, Katayama H. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. *Tree Genet Genomes*. 2012;8:841–54.
246. Alexander LW, Woeste KE. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. *Tree Genet Genomes*. 2014;10:803–12.
247. Yang B, Li M, Ma J, Fu Z, Xu X, Chen Q, et al. The complete chloroplast genome sequence of *Sapindus mukorossi*. *Mitochondrial DNA*. 2016;27:1825–6.
248. Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, et al. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genet Genomes*. 2013;9:489–98.
249. Yap JY, Rohner T, Greenfield A, Van Der Merwe M, McPherson H, Glenn W, et al. Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): structure and evolution. *PLoS One*. 2015;10, e0128126.
250. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot*. 2005;92:142–66.
251. Ruhlman TA, Rajasekaran K, Cary JW. Expression of chloroperoxidase from *Pseudomonas pyrocinia* in tobacco plastids for fungal resistance. *Plant Sci*. 2014;228:98–106.
252. Chakrabarti SK, Lutz KA, Lertwiriawong B, Svab Z, Maliga P. Expression of the *cry9Aa2 B.t.* gene in tobacco chloroplasts confers resistance to potato tuber moth. *Transgenic Res*. 2006;15:481–8.
253. 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;127:852–62.
254. Chen PJ, Senthilkumar R, Jane WN, He Y, Tian Z, Yeh KW. Transplastomic *Nicotiana benthamiana* plants expressing multiple defence genes encoding protease inhibitors and chitinase display broad-spectrum resistance against insects, pathogens and abiotic stresses. *Plant Biotechnol J*. 2014;12:503–15.
255. Wang YP, Wei ZY, Zhang YY, Lin CJ, Zhong XF, Wang YL, et al. Chloroplast-expressed MSI-99 in tobacco improves disease resistance and displays inhibitory effect against rice blast fungus. *Int J Mol Sci*. 2015;16:4628–41.
256. 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. *Biotechnology (N Y)*. 1995;13:362–5.
257. 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 U S A*. 1999;96:1840–5.
258. Lee SB, Kwon HB, Kwon SJ, Park SC, Jeong MJ, Han SE, et al. Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. *Mol Breeding*. 2003;11:1–13.
259. Ruiz ON, Hussein HS, Terry N, Daniell H. Phytoremediation of organomercurial compounds via chloroplast genetic engineering. *Plant Physiol*. 2003;132:1344–52.
260. Ruiz ON, Alvarez D, Torres C, Roman L, Daniell H. Metallothionein expression in chloroplasts enhances mercury accumulation and phytoremediation capability. *Plant Biotechnol J*. 2011;9:609–17.
261. Lutz KA, Knapp JE, Maliga P. Expression of bar in the plastid genome confers herbicide resistance. *Plant Physiol*. 2001;125:1585–90.
262. Ye GN, Hajdukiewicz PT, Broyles D, Rodriguez D, Xu CW, Nehra N, Staub JM. Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. *Plant J*. 2001;25:261–70.
263. Shimizu M, Goto M, Hanai M, Shimizu T, Izawa N, Kanamoto H, et al. Selectable tolerance to herbicides by mutated acetolactate synthase genes integrated into the chloroplast genome of tobacco. *Plant Physiol*. 2008;147:1976–83.
264. Iamtham S, Day A. Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat Biotechnol*. 2000;18:1172–6.
265. Dufourmantel N, Dubald M, Matringe M, Canard H, Garçon F, Job C, et al. Generation and characterization of soybean and marker-free tobacco plastid transformants over-expressing a bacterial 4-hydroxyphenylpyruvate dioxygenase which provides strong herbicide tolerance. *Plant Biotechnol J*. 2007;5:118–33.
266. Fouad WM, Altpeter F. Transplastomic expression of bacterial L-aspartate-alpha-decarboxylase enhances photosynthesis and biomass production in response to high temperature stress. *Transgenic Res*. 2009;18:707–18.
267. Lu Y, Rijzaani H, Karcher D, Ruf S, Bock R. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Natl Acad Sci U S A*. 2013;110:E623–32.
268. Ruiz ON, Daniell H. Engineering cytoplasmic male sterility via the chloroplast genome by expression of beta-ketothiolase. *Plant Physiol*. 2005;138:1232–46.
269. Yabuta Y, Tanaka H, Yoshimura S, Suzuki A, Tamoi M, Maruta T, Shigeoka S. Improvement of vitamin E quality and quantity in tobacco and lettuce by chloroplast genetic engineering. *Transgenic Res*. 2013;22:391–402.
270. Harada H, Maoka T, Osawa A, Hattan J, Kanamoto H, Shindo K, et al. Construction of transplastomic lettuce (*Lactuca sativa*) dominantly producing astaxanthin fatty acid esters and detailed chemical analysis of generated carotenoids. *Transgenic Res*. 2014;23:303–15.
271. Pengelly JJ, Forster B, von Caemmerer S, Badger MR, Price GD, Whitney SM. Transplastomic integration of a cyanobacterial bicarbonate transporter into tobacco chloroplasts. *J Exp Bot*. 2014;65:3071–80.
272. Sanz-Barrio R, Corral-Martinez P, Ancin M, Seguí-Simarro JM, Farran I. Overexpression of plastidial thioredoxin f leads to enhanced starch accumulation in tobacco leaves. *Plant Biotechnol J*. 2013;11:618–27.
273. Elghabi Z, Karcher D, Zhou F, Ruf S, Bock R. Optimization of the expression of the HIV fusion inhibitor cyanovirin-N from the tobacco plastid genome. *Plant Biotechnol J*. 2011;9:599–608.
274. Agrawal P, Verma D, Daniell H. Expression of *Trichoderma reesei* beta-mannanase in tobacco chloroplasts and its utilization in lignocellulosic woody biomass hydrolysis. *PLoS One*. 2011;6, e29302.
275. Verma D, Jin S, Kanagaraj A, Singh ND, Daniel J, Kolattukudy PE, et al. Expression of fungal cutinase and swollenin in tobacco chloroplasts reveals novel enzyme functions and/or substrates. *PLoS One*. 2013;8, e57187.
276. Saxena B, Subramanian M, Malhotra K, Bhavesh NS, Potlakyala SD, Kumar S. Metabolic engineering of chloroplasts for artemisinic acid biosynthesis and impact on plant growth. *J Biosci*. 2014;39:33–41.
277. Yu LX, Gray BN, Rutzke CJ, Walker LP, Wilson DB, Hanson MR. Expression of thermostable microbial cellulases in the chloroplasts of nicotine-free tobacco. *J Biotechnol*. 2007;131:362–9.
278. Petersen K, Bock R. High-level expression of a suite of thermostable cell wall-degrading enzymes from the chloroplast genome. *Plant Mol Biol*. 2011;76:311–21.



279. Arai Y, Shikanai T, Doi Y, Yoshida S, Yamaguchi I, Nakashita H. Production of polyhydroxybutyrate by polycistronic expression of bacterial genes in tobacco plastid. *Plant Cell Physiol.* 2004;45:1176–84.
280. Hasunuma T, Miyazawa SI, Yoshimura S, Shinzaki Y, Tomizawa KI, Shindo K, et al. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *Plant J.* 2008;55:857–68.
281. Nakahira Y, Ishikawa K, Tanaka K, Tozawa Y, Shiina T. Overproduction of hyperthermostable beta-1,4-endoglucanase from the archaeon *Pyrococcus horikoshii* by tobacco chloroplast engineering. *Biosci Biotechnol Biochem.* 2013;77:2140–3.
282. Sherman A, Su J, Lin S, Wang X, Herzog RW, Daniell H. Suppression of inhibitor formation against FVIII in a murine model of hemophilia A by oral delivery of antigens bioencapsulated in plant cells. *Blood.* 2014;124:1659–68.
283. Sanz-Barrio R, Fernandez-San Millan A, Corral-Martinez P, Segui-Simarro JM, Farran I. Tobacco plastidial thioredoxins as modulators of recombinant protein production in transgenic chloroplasts. *Plant Biotechnol J.* 2011;9:639–50.
284. Farran I, McCarthy-Suarez I, Rio-Manterola F, Mansilla C, Lasarte JJ, Mingo-Castel AM. The vaccine adjuvant extra domain A from fibronectin retains its proinflammatory properties when expressed in tobacco chloroplasts. *Planta.* 2010;231:977–90.
285. Boyhan D, Daniell H. Low-cost production of proinsulin in tobacco and lettuce chloroplasts for injectable or oral delivery of functional insulin and C-peptide. *Plant Biotechnol J.* 2011;9:585–98.
286. 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;1:71–9.
287. Daniell H, Ruiz G, Denes B, Sandberg L, Langridge W. Optimization of codon composition and regulatory elements for expression of human insulin like growth factor-1 in transgenic chloroplasts and evaluation of structural identity and function. *BMC Biotechnol.* 2009;9:33.
288. Wang X, Su J, Sherman A, Rogers GL, Liao G, Hoffman BE, et al. Plant-based oral tolerance to hemophilia therapy employs a complex immune regulatory response including LAP<sup>+</sup>CD4<sup>+</sup> T cells. *Blood.* 2015;125:2418–27.
289. Su J, Sherman A, Doerfler PA, Byrne BJ, Herzog RW, Daniell H. Oral delivery of Acid Alpha Glucosidase epitopes expressed in plant chloroplasts suppresses antibody formation in treatment of Pompe mice. *Plant Biotechnol J.* 2015;13:1023–32.
290. Youm JW, Jeon JH, Kim H, Min SR, Kim MS, Joung H, et al. High-level expression of a human beta-site APP cleaving enzyme in transgenic tobacco chloroplasts and its immunogenicity in mice. *Transgenic Res.* 2010;19:1099–108.
291. Arlen PA, Falconer R, Cherukumilli S, Cole A, Cole AM, Oishi KK, Daniell H. Field production and functional evaluation of chloroplast-derived interferon-alpha2b. *Plant Biotechnol J.* 2007;5:511–25.
292. Ruhlman T, Ahangari R, Devine A, Samsam M, Daniell H. Expression of cholera toxin B-proinsulin fusion protein in lettuce and tobacco chloroplasts—oral administration protects against development of insulinitis in non-obese diabetic mice. *Plant Biotechnol J.* 2007;5:495–510.
293. 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 Breeding.* 2003;11:49–58.
294. Lim S, Ashida H, Watanabe R, Inai K, Kim YS, Mukougawa K, et al. Production of biologically active human thioredoxin 1 protein in lettuce chloroplasts. *Plant Mol Biol.* 2011;76:335–44.
295. Nadai M, Bally J, Vitel M, Job C, Tissot G, Botterman J, Dubald M. High-level expression of active human alpha1-antitrypsin in transgenic tobacco chloroplasts. *Transgenic Res.* 2009;18:173–83.
296. Gisby MF, Mellors P, Madesis P, Ellin M, Lavery H, O'Kane S, et al. A synthetic gene increases TGF beta 3 accumulation by 75-fold in tobacco chloroplasts enabling rapid purification and folding into a biologically active molecule. *Plant Biotechnol J.* 2011;9:618–28.
297. Farran I, Rio-Manterola F, Iniguez M, Garate S, Prieto J, Mingo-Castel AM. High-density seedling expression system for the production of bioactive human cardiotrophin-1, a potential therapeutic cytokine, in transgenic tobacco chloroplasts. *Plant Biotechnol J.* 2008;6:516–27.
298. Staub JM, Garcia B, Graves J, Hajdukiewicz PT, Hunter P, Nehra N, et al. High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat Biotechnol.* 2000;18:333–8.
299. Oey M, Lohse M, Scharff LB, Kreikemeyer B, Bock R. Plastid production of protein antibiotics against pneumonia via a new strategy for high-level expression of antimicrobial proteins. *Proc Natl Acad Sci U S A.* 2009;106:6579–84.
300. 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;73:8266–74.
301. 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;22:4374–84.
302. Millan AFS, Ortigosa SM, Hervás-Stubbis S, Corral-Martinez P, Segui-Simarro JM, Gaetan J, et al. Human papillomavirus L1 protein expressed in tobacco chloroplasts self-assembles into virus-like particles that are highly immunogenic. *Plant Biotechnol J.* 2008;6:427–41.
303. Yacono MD, Farran I, Becher ML, Sander V, Sanchez VR, Martin V, et al. A chloroplast-derived *Toxoplasma gondii* GRA4 antigen used as an oral vaccine protects against toxoplasmosis in mice. *Plant Biotechnol J.* 2012;10:1136–44.
304. Kolotilin I, Kaldis A, Devriendt B, Joensuu J, Cox E, Menassa R. Production of a subunit vaccine candidate against porcine post-weaning diarrhea in high-biomass transplastomic tobacco. *Plos One.* 2012;7, e42405.
305. Arlen PA, Singleton M, Adamovicz JJ, Ding Y, Davoodi-Semiromi A, Daniell H. Effective plague vaccination via oral delivery of plant cells expressing F1-V antigens in chloroplasts. *Infect Immun.* 2008;76:3640–50.
306. Molina A, Hervás-Stubbis S, Daniell H, Mingo-Castel AM, Veramendi J. High-yield expression of a viral peptide animal vaccine in transgenic tobacco chloroplasts. *Plant Biotechnol J.* 2004;2:141–53.
307. Lentz EM, Mozgovoj MV, Bellido D, Santos MJD, Wigdorovitz A, Bravo-Almonacid FF. VP8\* antigen produced in tobacco transplastomic plants confers protection against bovine rotavirus infection in a suckling mouse model. *J Biotechnol.* 2011;156:100–7.
308. 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;311:1001–9.
309. Kang TJ, Loc NH, Jang MO, Jang YS, Kim YS, Seo JE, Yang MS. Expression of the B subunit of *E. coli* heat-labile enterotoxin in the chloroplasts of plants and its characterization. *Transgenic Res.* 2003;12:683–91.
310. Chebolu S, Daniell H. Stable expression of Gal/GalNAc lectin of *Entamoeba histolytica* in transgenic chloroplasts and immunogenicity in mice towards vaccine development for amoebiasis. *Plant Biotechnol J.* 2007;5:230–9.
311. Glenz K, Bouchon B, Stehle T, Wallich R, Simon MM, Warzecha H. Production of a recombinant bacterial lipoprotein in higher plant chloroplasts. *Nat Biotechnol.* 2006;24:76–7.
312. Rosales-Mendoza S, Alpuche-Solis AG, Soria-Guerra RE, Moreno-Fierros L, Martinez-Gonzalez L, Herrera-Diaz A, Korban SS. Expression of an *Escherichia coli* antigenic fusion protein comprising the heat labile toxin B subunit and the heat stable toxin, and its assembly as a functional oligomer in transplastomic tobacco plants. *Plant J.* 2009;57:45–54.
313. Soria-Guerra RE, Alpuche-Solis AG, Rosales-Mendoza S, Moreno-Fierros L, Bendik EM, Martinez-Gonzalez L, Korban SS. Expression of a multi-epitope DPT fusion protein in transplastomic tobacco plants retains both antigenicity and immunogenicity of all three components of the functional oligomer. *Planta.* 2009;229:1293–302.
314. Rubio-Infante N, Govea-Alonso DO, Alpuche-Solis AG, Garcia-Hernandez AL, Soria-Guerra RE, Paz-Maldonado LMT, et al. A chloroplast-derived C4V3 polypeptide from the human immunodeficiency virus (HIV) is orally immunogenic in mice. *Plant Mol Biol.* 2012;78:337–49.
315. Waheed MT, Thones N, Muller M, Hassan SW, Gottschamel J, Lössl E, et al. Plastid expression of a double-pentameric vaccine candidate containing human papillomavirus-16 L1 antigen fused with LTB as adjuvant: transplastomic plants show pleiotropic phenotypes. *Plant Biotechnol J.* 2011;9:651–60.
316. Gonzalez-Rabade N, McGowan EG, Zhou F, McCabe MS, Bock R, Dix PJ, et al. Immunogenicity of chloroplast-derived HIV-1 p24 and a p24-Nef fusion protein following subcutaneous and oral administration in mice. *Plant Biotechnol J.* 2011;9:629–38.
317. Zhou YX, Lee MYT, Ng JMH, Chye ML, Yip WK, Zee SY, Lam E. A truncated hepatitis E virus ORF2 protein expressed in tobacco plastids is immunogenic in mice. *World J Gastroenterol.* 2006;12:306–12.
318. Shao HB, He DM, Qian KX, Shen GF, Su ZL. The expression of classical swine fever virus structural protein E2 gene in tobacco chloroplasts for applying chloroplasts as bioreactors. *C R Biol.* 2008;331:179–84.



319. Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, et al. Expression of tetanus toxin Fragment C in tobacco chloroplasts. *Nucleic Acids Res.* 2003;31:1174–9.
320. Morgenfeld M, Lentz E, Segretin ME, Alfano EF, Bravo-Almonacid F. Translational fusion and redirection to thylakoid lumen as strategies to enhance accumulation of human papillomavirus E7 antigen in tobacco chloroplasts. *Mol Biotechnol.* 2014;56:1021–31.