

RESEARCH ARTICLE

The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis

Xiaohong Yao¹, Ping Tang¹, Zuozhou Li¹, Dawei Li¹, Yifei Liu², Hongwen Huang^{1,2*}

1 Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, The Chinese Academy of Sciences, Wuhan 430074, Hubei, China, **2** Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou, 510650, Guangdong, China

* huanghw@scbg.ac.cn



OPEN ACCESS

Citation: Yao X, Tang P, Li Z, Li D, Liu Y, Huang H (2015) The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. PLoS ONE 10(6): e0129347. doi:10.1371/journal.pone.0129347

Academic Editor: Zhong-Hua Chen, University of Western Sydney, AUSTRALIA

Received: February 2, 2015

Accepted: May 7, 2015

Published: June 5, 2015

Copyright: © 2015 Yao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The sequences of four chloroplast genomes were deposited in GenBank (accession numbers KP297242 through KP297245).

Funding: This work was partly supported by the Natural Scientific Foundation of China (31370251) and Fund of Science and Technology Service, Chinese Academy of Sciences (KFJ-EW-STS-076).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Actinidia chinensis is an important economic plant belonging to the basal lineage of the asterids. Availability of a complete *Actinidia* chloroplast genome sequence is crucial to understanding phylogenetic relationships among major lineages of angiosperms and facilitates kiwifruit genetic improvement. We report here the complete nucleotide sequences of the chloroplast genomes for *Actinidia chinensis* and *A. chinensis* var *deliciosa* obtained through de novo assembly of Illumina paired-end reads produced by total DNA sequencing. The total genome size ranges from 155,446 to 157,557 bp, with an inverted repeat (IR) of 24,013 to 24,391 bp, a large single copy region (LSC) of 87,984 to 88,337 bp and a small single copy region (SSC) of 20,332 to 20,336 bp. The genome encodes 113 different genes, including 79 unique protein-coding genes, 30 tRNA genes and 4 ribosomal RNA genes, with 16 duplicated in the inverted repeats, and a tRNA gene (*trnM-CAU*) duplicated once in the LSC region. Comparisons of IR boundaries among four asterid species showed that IR/LSC borders were extended into the 5' portion of the *psbA* gene and IR contraction occurred in *Actinidia*. The *clap* gene has been lost from the chloroplast genome in *Actinidia*, and may have been transferred to the nucleus during chloroplast evolution. Twenty-seven polymorphic simple sequence repeat (SSR) loci were identified in the *Actinidia* chloroplast genome. Maximum parsimony analyses of a 72-gene, 16 taxa angiosperm dataset strongly support the placement of Actinidiaceae in Ericales within the basal asterids.

Introduction

In plants, chloroplasts (cp) are key organelles for photosynthesis and are crucial in the biosynthesis of starch, fatty acids, pigments and amino acids [1]. Typically, cp genomes in angiosperms are highly conserved and have circular genomes ranging from 115 to 165 kb in length and consisting of a large-single-copy region (LSC; 80–90 kb) and a small-single-copy region (SSC; 16–27 kb), separated by an inverted repeat (IR) [2,3]. In contrast to nuclear and mitochondrial

genomes, chloroplast genomes are largely conserved in gene content, organization and structure [3]. However, mutations, duplications, losses and rearrangements of genes have been observed in several angiosperm lineages [4].

Complete plastid genome sequences of vascular plants were first reported in tobacco [5]. Advances in next-generation sequencing technologies have enabled the rapid acquisition of whole cp genome sequences at low cost. At present, over 600 cp genome sequences are currently deposited at the National Center for Biotechnology Information (NCBI) including all of the major lineages of the plant kingdom. To date, only three complete cp genomes have been sequenced for representatives of the basal asterids: *Ardisia polysticta* (Myrsinaceae) [6], *Camellia sinensis* (Theaceae) [7], and *Vaccinium macrocarpon* (Ericaceae) [8]. More cp genome sequences from additional taxa of the basal asterids are needed for understanding phylogenetic relationships among angiosperms [7].

Kiwifruit is an important fruit tree in the Actinidiaceae family of the asterids. This family includes 3 genera, i.e., *Actinidia*, *Clematoclethra* and *Saurauia* [9]. *Actinidia*, comprising 74 species, is widely distributed in Eastern Asia from just south of the Equator in the tropics to cold temperate regions as far north as 50° latitude [10]. *Actinidia chinensis* and *A. chinensis* var *deliciosa*, from which most commercial kiwifruit varieties have been developed, have received considerable attention over the last forty years [10,11].

As an important crop plant grown for fresh fruit, kiwifruit is cultivated in 14 countries. The world kiwifruit acreage was 166,000 hectares and annual kiwifruit production was around 2 million metric tons in 2012 [12]. Bacterial canker of kiwifruit, caused by *Pseudomonas syringae* pv. *actinidiae* (Psa), is currently the major cause of losses in kiwifruit production worldwide [13]. As an environmentally friendly approach, genetic engineering would be very useful to develop a method of disease resistance to Psa. Compared to nuclear genome engineering, chloroplast genetic engineering often gives high expression levels [14,15], which holds great promise for breeding kiwifruit cultivars with disease resistance. However, the lack of complete chloroplast genome sequences is one of the major limitations of extending this technology to most crops [15]. Hence, complete chloroplast genome sequences of kiwifruit are needed.

For the last three decades, numerous phylogenetic studies using chloroplast DNA sequence data have contributed to our understanding of the evolutionary relationships within the genus [16,17]. However, the interspecies relationships within the genus *Actinidia* remain largely controversial [10]. With the goal of developing molecular markers for further understanding phylogenetic relationships in *Actinidia* and chloroplast genome evolution in asterids, here we report the complete chloroplast genome sequences of *A. chinensis* and *A. chinensis* var *deliciosa*, the first cp genomes for the Actinidiaceae, obtained through de novo assembly of Illumina paired-end reads and produced by total DNA sequencing. Information on the complete chloroplast genome sequence will be instrumental for genetic engineering and breeding programs.

Materials and Methods

Plant materials

The genus *Actinidia* contains extensive natural ploidy variation with species known at four ploidy levels (diploid through octoploid). Four individuals, representing three ploidy levels of the *Actinidia chinensis* complex, including *A. chinensis* var. *chinensis* (2×) (accession number AC011), *A. chinensis* var. *chinensis* (4×) (accession number AC017), *A. chinensis* var *deliciosa* (4×) (accession number AD006) and *A. chinensis* var *deliciosa* (6×) (accession number AD019), were received from the National *Actinidia* Germplasm Repository of China (Wuhan,

China). Approximately 100 g young leaf tissue per individual was harvested for extraction of genomic DNA.

DNA sequencing and genome assembly

Total DNA was extracted from young leaves with the DNeasy Plant Mini Kit (Qiagen, CA, USA). DNA was sheared by nebulization with compressed nitrogen gas, yielding fragments of 300 bp in length, and fragmentation quality was checked on a Bioanalyzer 2100 (Agilent Technologies). Paired-end libraries were prepared with the Mate Pair Library Preparation Kit (Illumina, San Diego, California, USA) in accordance with the manufacturer's instructions. Genomic DNA was sequenced on a single lane with multiplexing on HiSeq2000 flow cell lanes (Illumina Inc.).

For each species, the raw reads were assembled into non-redundant contigs with Velvet1.2.07 [18], a de novo sequence assembly software package, with $k = 30$ and scaffolding contigs having a minimum length of 100 bp. All contigs were then mapped against the reference cp genome in *Camellia sinensis* [7] with BLAST (<http://blast.ncbi.nlm.nih.gov/>) similarity searches against the NCBI nr database by using the default search parameters. To identify the chloroplast contigs, all the returned contigs were blasted to the reference genomes. Primer walking and additional Sanger sequencing were then used to fill the gaps between the seven to twelve large contigs and to verify the junctions between the single-copy and the IRs regions (S1 Table).

Gene annotation

Initial gene annotation of the four chloroplast genomes was performed with Dual Organellar GenoMe Annotator (DOGMA; [19]). The tRNA genes were predicted with ARAGORN [20] and tRNAscan-SE [21]. The circular gene maps were drawn by the OrganellarGenomeDRAW tool (OGDRAW), followed by manual modification [22]. Comparison of *Actinidia* cp genome structures was performed by the mVISTA program in Shuffle-LAGAN mode [23] by using the annotation of *Camellia sinensis* as the reference.

Repeat structure

Size and location of both direct (forward) and inverted (palindromic) repeats in the *Actinidia* cp genome were determined by running REPuter [24] at a minimal repeat size ≥ 30 bp with a Hamming distance of 3.

Microsatellite (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) detection was performed by using MISA [25] with thresholds of nine repeat units for mononucleotide SSRs, five repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs.

Phylogenetic analysis

We sampled fifteen species representing the asteroid lineage of angiosperm to reconstruct a phylogeny of asterids with cp genes (S2 Table). The 72 protein-coding genes shared by the chloroplast genomes of 15 asterid members (S2 Table) were used for phylogenetic analysis. *Arabidopsis thaliana* was used as the outgroup. The sequences were aligned with MUSCLE [26] with the default settings and concatenated into a single alignment of 55,370 characters. Optimal trees were inferred with maximum parsimony (MP) and Maximum likelihood (ML) as implemented in PAUP* version 4.0b10 [27]. For all analyses, characters were equally weighted, gap regions were excluded in the phylogenetic analyses, and multistate characters were

treated as uncertainties. Prior to the ML analyses, the Akaike information criterion (AIC) was employed to determine the best model and parameters settings with the program Modeltest 3.07 [28]. Analyses were performed with the following options implemented: heuristic search mode used 1000 random-addition sequence replicates holding 20 trees at each step, tree bisection-reconnection (TBR) branch-swapping, MULTrees in effect, and steepest descent off. Bootstraps analyses were employed with heuristic searches and 1000 replicates. Because the results from ML analyses were congruent with the results of the MP analyses, they are not presented here.

Results and Discussion

Genome sequencing and assembly

Illumina paired-end (300 bp) sequencing produced 310.9, 252.8, 274.9 and 286.8 Mb of data for *A. chinensis* (2×), *A. chinensis* (4×), *A. chinensis* var *deliciosa* (4×) and *A. chinensis* var *deliciosa* (6×), respectively. The N50 of contigs ranged from 9898 bp to 22,416 bp. A total of twelve, eight, fifteen and seven contigs with lengths of 131,930, 131,859, 130,725 and 131,916 bp were found to be derived from cpDNAs for *A. chinensis* (2×), *A. chinensis* (4×), *A. chinensis* var *deliciosa* (4×) and *A. chinensis* var *deliciosa* (6×), respectively. Four junction regions between IRs and SSC/LSC in each cp genome were confirmed by PCR amplifications and Sanger sequencing. The sequences of four chloroplast genomes were deposited in GenBank (accession numbers KP297242 through KP297245).

Organization of chloroplast genome

The nucleotide sequences of the four cp genomes range from 156,346 bp in *A. chinensis* (2×) to 157,375 bp in *A. chinensis* var *deliciosa* (6×) (Table 1). All four cp genomes share the typical quadripartite cp structure, with a pair of inverted repeats (IRs 24,013–24,391 bp) each separated by a SSC (20,332–20,336 bp) and a LSC (87,984–88337 bp) (Table 1). This structure agrees well with that based on restriction mapping of [29], although the total lengths slightly differ.

When duplicated genes are counted only once, the *Actinidia* cp genomes contain 113 unique genes (unique ORFs were not taken into account) including 79 protein-coding genes, 30 tRNA genes and four rRNA coding genes (Fig 1 and Table 1). Four protein coding, four rRNA and eight tRNA genes are duplicated and located in the IR regions. Fourteen of the protein-coding genes and eight of the tRNA genes contain introns, 16 of which contain a single intron, whereas one (*ycf3*) has two introns (Table 2). The gene *rps12* is trans-spliced; the 5' end

Table 1. Summary of chloroplast genome characteristics in *Actinidia*.

	<i>A. chinensis</i> (2×)	<i>A. chinensis</i> (4×)	<i>A. chinensis</i> var <i>deliciosa</i> (4×)	<i>A. chinensis</i> var <i>deliciosa</i> (6×)
Size (bp)	156346	156810	156741	157375
LSC length (bp)	87984	88337	88267	88261
SSC length (bp)	20336	20333	20332	20332
IR length (bp)	24013	24070	24071	24391
Number of genes*	113	113	113	113
Protein-coding genes*	79	79	79	79
tRNA genes*	30	30	30	30
rRNA genes	4	4	4	4
GC content (%)	37.2	37.2	37.2	37.2

* numbers refer to the number of different genes

doi:10.1371/journal.pone.0129347.t001

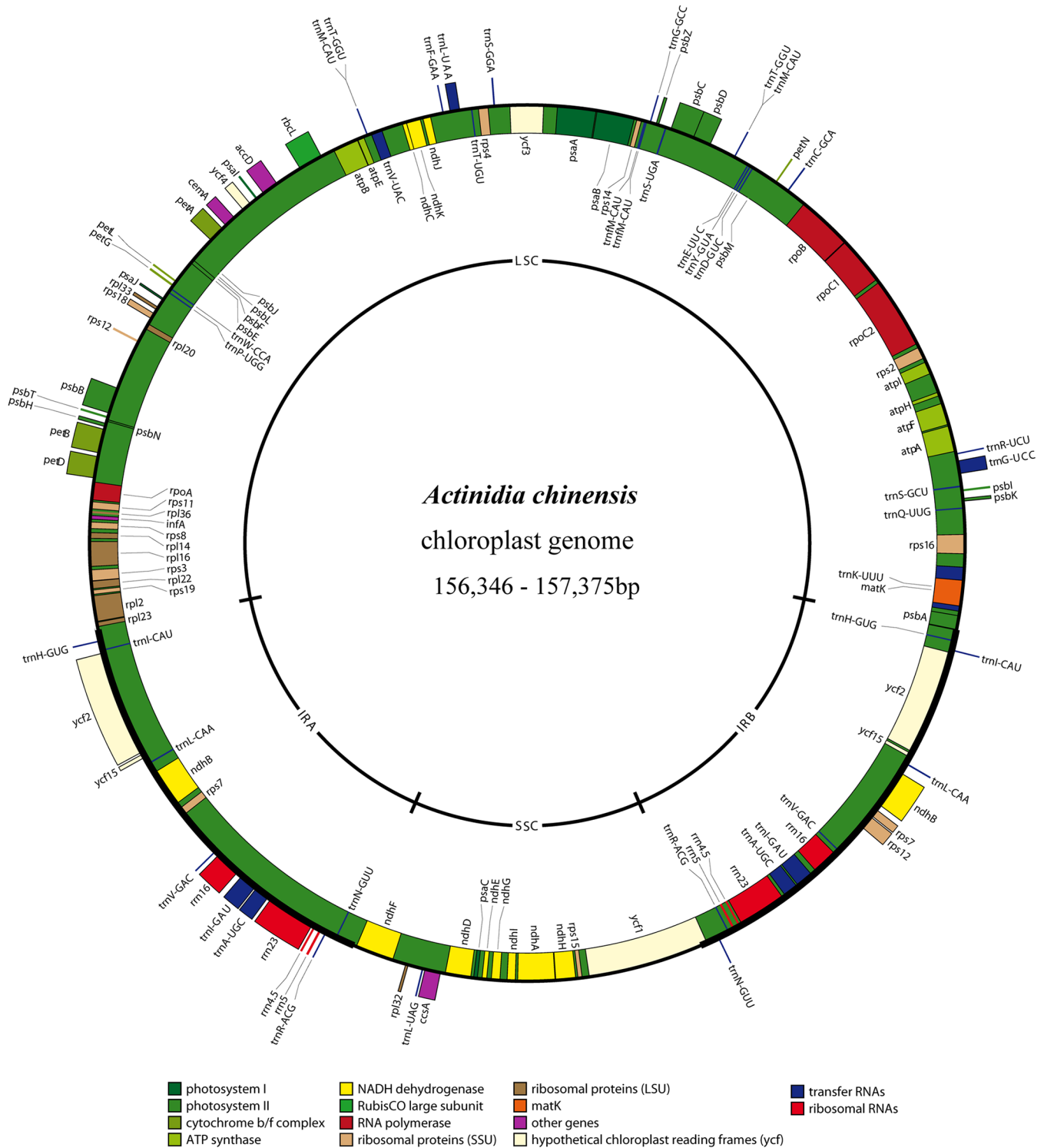


Fig 1. Gene maps of the *Actinidia* plastid genome. Genes shown on the outside of the map are transcribed clockwise, while genes on the inside are transcribed counter-clockwise. Genes belonging to different functional groups are color-coded.

doi:10.1371/journal.pone.0129347.g001

Table 2. List of genes encoded by the *Actinidia chinensis* plastome.

Category	Group of gene	Name of gene
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ, ycf3^a, ycf4</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	NADH oxidoreductase	<i>ndhA^b, ndhB^{b,c}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b/f complex	<i>petA, petB^b, petD^b, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF^b, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcL</i>
Self-replication	Large subunit ribosomal proteins	<i>rpl2^b, rpl14, rpl16^b, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	Small subunit ribosomal proteins	<i>rps2, rps3, rps4, rps7^c, rps8, rps11, rps12^a, rps14, rps15, rps16^b, rps18, rps19</i>
	RNA polymerase subunits	<i>rpoA, rpoB, rpoC1^b, rpoC2</i>
	Ribosomal RNAs	<i>rm16^c, rm23^c, rm4.5^c, rm5^c</i>
	Transfer RNAs	<i>trnA-UGC^{b,c}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC^b, trnH-GUG^c, trnI-CAU^c, trnI-GAU^{b,c}, trnK-UUU^b, trnL-CAA^c, trnL-UAA^b, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU^c, trnP-UGG, trnQ-UUG, trnR-ACG^c, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^c, trnV-UAC^b, trnW-CCA, trnY-GUA</i>
	Other genes	Translational initiation factor Maturase K Envelope membrane protein Subunit of acetyl-CoA-carboxylase c-type cytochrome synthesis gene Proteins of unknown function
Putative pseudogenes		<i>ycf15^c</i>

^aGene with two introns.

^bGene with one intron.

^cGenes located in the inverted repeats.

doi:10.1371/journal.pone.0129347.t002

exon is located in the LSC region and the 3' exon and intron are duplicated and located in the IR regions. Gene content is typical for the plastomes of the dicotyledonous angiosperms, being most similar to that of *Camellia sinensis* (Theaceae, [7]) as compared with those land plants sequenced. Most protein-coding genes have the standard AUG as the initiator codon [3], but *ndhD* has an initiator codon of ACG. The *trnfM-CAU* genes are duplicated in the LSC of the *Actinidia* chloroplast genome and separated by 14 bp with the same orientation. tRNA gene duplications have also been reported for black pine [30] and the green algae [31]. The overall GC content of the *Actinidia* chloroplast genome is 37.2%. The GC content of the *Actinidia* cp genome is close to that of *Ardisia polysticta* (37.07%) and other asterids [6].

The overall sequence identity of the four *Actinidia* cp genomes was plotted with mVISTA with *Camellia sinensis* as a reference (Fig 2). The IRs show lower sequence divergence than that

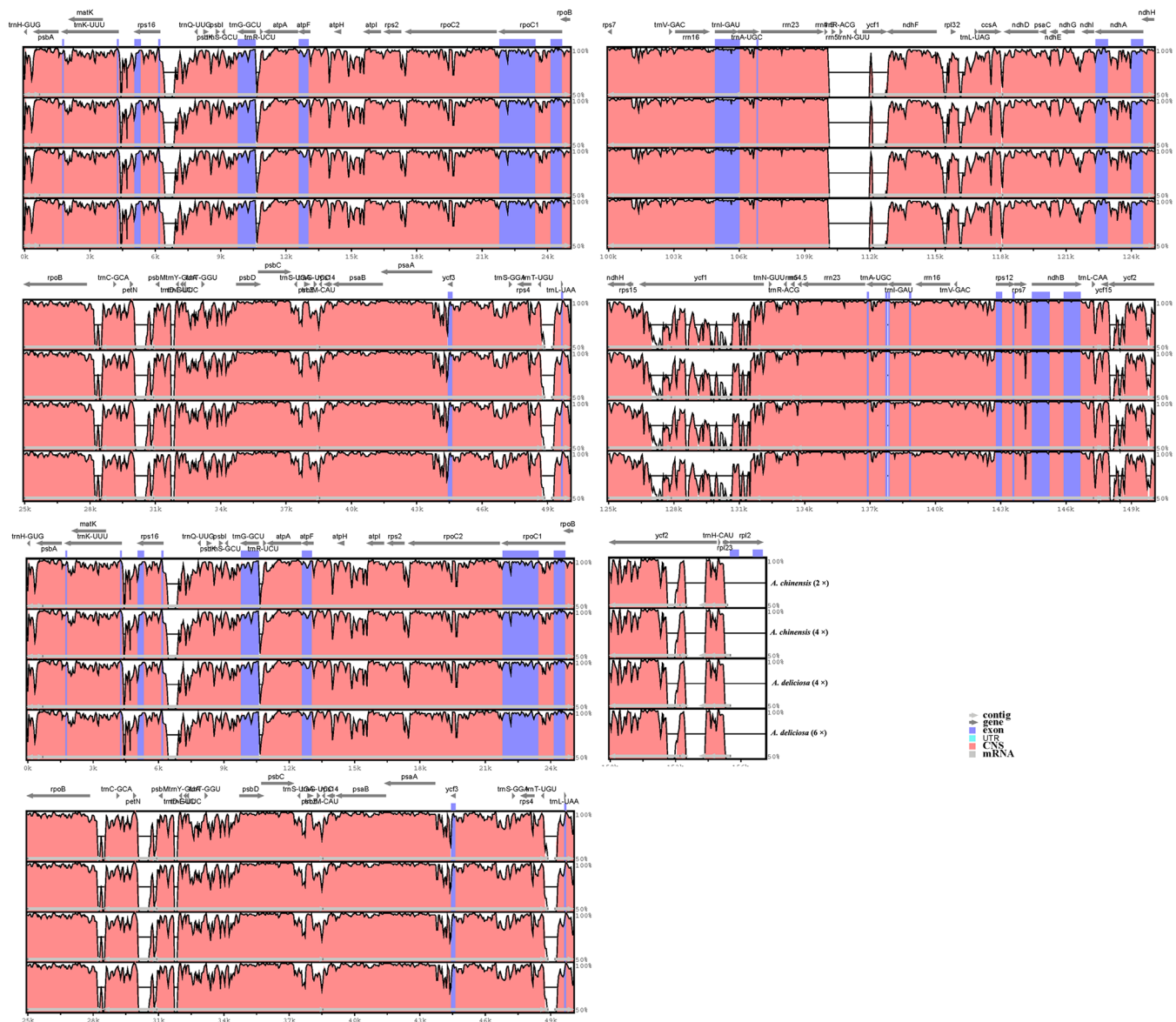


Fig 2. Sequence identity plots between four sequenced chloroplast genomes, with *Camellia sinensis* as a reference. The vertical scale indicates the identity percentage (50% to 100%). The horizontal axis corresponds to the coordinates within the chloroplast genome. Annotated genes are displayed along the top.

doi:10.1371/journal.pone.0129347.g002

in the single-copy regions, possibly due to copy correction between IR sequences by gene conversion [32]. As expected, non-coding regions exhibit a higher divergence than coding regions, and the most divergent regions among the four cp genomes are localized in the intergenic spacers. Intergenic regions with high degrees of divergence included *rps16-trnQ*, *petN-psbM*, *trnT-trnL*, *ndhD-ccsA*, *ndhI-ndhG*, *psbI-trnS*, *rrn5-ndhF*, and *trnH-psbA*. Therefore, developing universal primers for these intergenic regions could aid in assessing phylogenetic relationships among *Actinidia* species.

Repeat structure

Mononucleotide microsatellite length polymorphisms have been used as markers in cp genomes for understanding cp evolutionary history due to their high rates of variability [33]. In our MISA analyses, 65 SSRs with a length of at least 10 bp in the *Actinidia* plastid genomes were detected, of which 61 are mononucleotide repeats, two are dinucleotide repeats, and two are trinucleotide repeats. No pentanucleotides or hexanucleotides were found. Sixty homopolymer loci are composed of A/T repeats, whereas only one is composed of C repeats. All of the dinucleotides are composed of multiple copies of AT/TA repeats. SSR loci are mostly present in noncoding regions. Of the 65 loci, 27 were polymorphic (Table 3). Hence, the plastid genome sequences of *Actinida* will be useful in developing lineage-specific cpSSR markers, which are widely used in population-genetic and evolutionary studies of plant.

REPuter [24] identified a total of 34 pairs of repeats (30 bp or longer) with a sequence identity greater than 90% in the *Actinidia* cp genome, of which 29 are forward and 5 are inverted repeats (Table 4). Searches for shorter and/or more divergent repeats would likely identify many additional repeated sequences. The repeats range from 30 to 79 bp in length, and are repeated from two to nine times. Most of the repeated sequences occur in regions of noncoding DNA, whereas some were found in protein-coding regions (e.g. *ycf2*). A substantial number of repeated sequences identified in chloroplast genomes, especially in intergenic spacer regions, have been reported in a number of angiosperm lineages, including other asterids (e.g. [7]).

The availability of plastid genome sequences should provide valuable information for plastid genetic engineering [14,15]. Plastid transformation is based on homologous recombination between the vector and plastid genome sequences. Thus, knowledge of the nucleotide sequence of chloroplast genome would be helpful to identify the optimal intergenic spacers for transgene integration and to develop species-specific cp transformation vectors. In the present study, the direct or palindromic repeat sites found in the LSC region of the *Actinidia* cp genome represent potential site-specific recombination sites that could be used in the development of a kiwifruit-specific chloroplast vector.

ClpP lost in the *Actinida* cp genome

Gene loss and gene transfer to the nucleus is a common feature of cp genomes [34,35]. For instance, the *rpl22* gene of Fagaceae [36] and Fabaceae [37], the *infA* in rosids [38], the *rpl32* gene in *Populus* and Salicaceae [39,40], and *accD* in *Trifolium* [41] have been transferred to the nuclear genome. The *clpP* gene, which encodes the Proteolytic subunit of Clp-protease with over 200 amino acids, is widely distributed among the cp genome of various land plant species [42]. The *clpP* gene encoded in the *Camellia sinensis* chloroplast genome are absent in the *Actinida* cp genome [7]. The close relationship of these two plant families indicates that this difference reflects a relatively recent event, either gene loss or functional transfer to the nucleus.

Among the angiosperm plastids investigated thus far, *clpP* is also partially or completely missing from the plastid genome of *Passiflora* (Passifloraceae, [43]), *Trachelium caeruleum* (Fabaceae, [44]), *Scaevola* (Goodeniaceae) [42] (Wicke et al., 2011), and *Vaccinium macrocarpon*

Table 3. Polymorphic microsatellites in the chloroplast genome of *Actinidia*.

Unit	SSR type	Location
A	(A)n	<i>trnK-rps16</i> , <i>rps16-trnQ</i> , <i>trnS-trnR</i> , <i>atpF</i> intron, <i>atpH-atpI</i> , <i>psbZ-trnG</i> , <i>ycf3-trnS</i> , <i>trnT-trnL</i> , <i>ycf4-cemA</i> , <i>petA-psbJ</i> , <i>psbE-petL</i>
T	(T)n	<i>rps16-trnQ</i> , <i>trnT-trnL</i> , <i>ndhC-trnV</i> , <i>accD-psaI</i> , <i>psaI-ycf4</i> , <i>rps8-rpl14</i> , <i>rpl16</i> intron

doi:10.1371/journal.pone.0129347.t003

Table 4. Repeated sequences in the *Actinidia* chloroplast genomes.

No.	Size (bp)	Type	Repeat number	Location	Region
1	79	F	2	<i>trnM-CAU</i>	LSC
2	76	F	2	IGS (<i>rps12/psbT</i>)	LSC
3	76	F	2	IGS (<i>rps12/psbT</i>)	LSC
4	68	F	2	IGS (<i>rbcL/accD</i>)	LSC
5	57	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
6	56	F	2	IGS (<i>rps12/psbT</i>)	LSC
7	54	F	2	IGS (<i>rbcL/accD</i>)	LSC
8	49	F	3	IGS (<i>trnH-GUG/trnI-CAU</i>)	IRa,b
9	48	P	1	CDS (<i>rps18</i>)	LSC
10	47	P	1	IGS (<i>petB/petD</i>)	LSC
11	47	F	2	IGS (<i>rps12/psbT</i>)	LSC
12	44	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
13	42	F	4	IGS (<i>rps12/trnV-GAC</i>), <i>ndhA</i> intron	IRa,b
14	41	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
15	40	F	9	IGS (<i>trnN-GUU/ycf1</i>)	IRa,b
16	39	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
17	38	F	3	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
18	37	F	2	IGS (<i>trnH-GUG/trnI-CAU</i>)	IRa,b
19	37	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
20	36	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
21	35	F	2	CDS (<i>ycf2</i>)	IRa,b
22	35	F	2	IGS (<i>trnN-GUU/ ycf15</i>)	IRa,b
23	34	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
24	34	F	2	CDS (<i>rps18</i>)	LSC
25	34	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
26	33	F	2	CDS (<i>ycf2</i>)	IRa,b
27	33	P	1	IGS (<i>accD/psaI</i>)	LSC
28	32	F	2	IGS (<i>trnH-GUG/trnI-CAU</i>)	IRa,b
29	32	F	2	IGS(<i>rrn4.5/rrn5</i>),	IRa,b
30	32	F	2	IGS (<i>rps15/trnN-GUU</i>)	IRa,b
31	31	F	2	IGS (<i>rps12/psbT</i>)	LSC
32	30	F	2	CDS (<i>ycf2</i>)	IRa,b
33	30	P	1	<i>trnS-GCU</i>	LSC
34	30	P	1	CDS (<i>rps18</i>)	LSC

F: Forward; P: Inverted; IGS: Intergenic spacer; CDS: protein-coding regions.

doi:10.1371/journal.pone.0129347.t004

(Ericaceae, [8]). The availability of the sequenced *A. chinese* nuclear genome gave us opportunity to search for these missing genes by using the *C. sinensis* chloroplast-encoded proteins. Fourteen gene fragments encode the ATP-dependent *clp*-protease proteolytic subunit were annotated in the kiwifruit genome (<http://bioinfo.bti.cornell.edu/cgi-bin/kiwi/home.cgi>). In this study, two fragments (135 and 198 bp) from kiwifruit genome CDS sequences were found to have 77% and 89% similarity to *clpP* encoded in the cpDNA of *C. sinensis*, which suggests *clpP* has probably been replaced by a copy of the same gene that was transferred to the nucleus during chloroplast evolution, but this needs to be further explored.

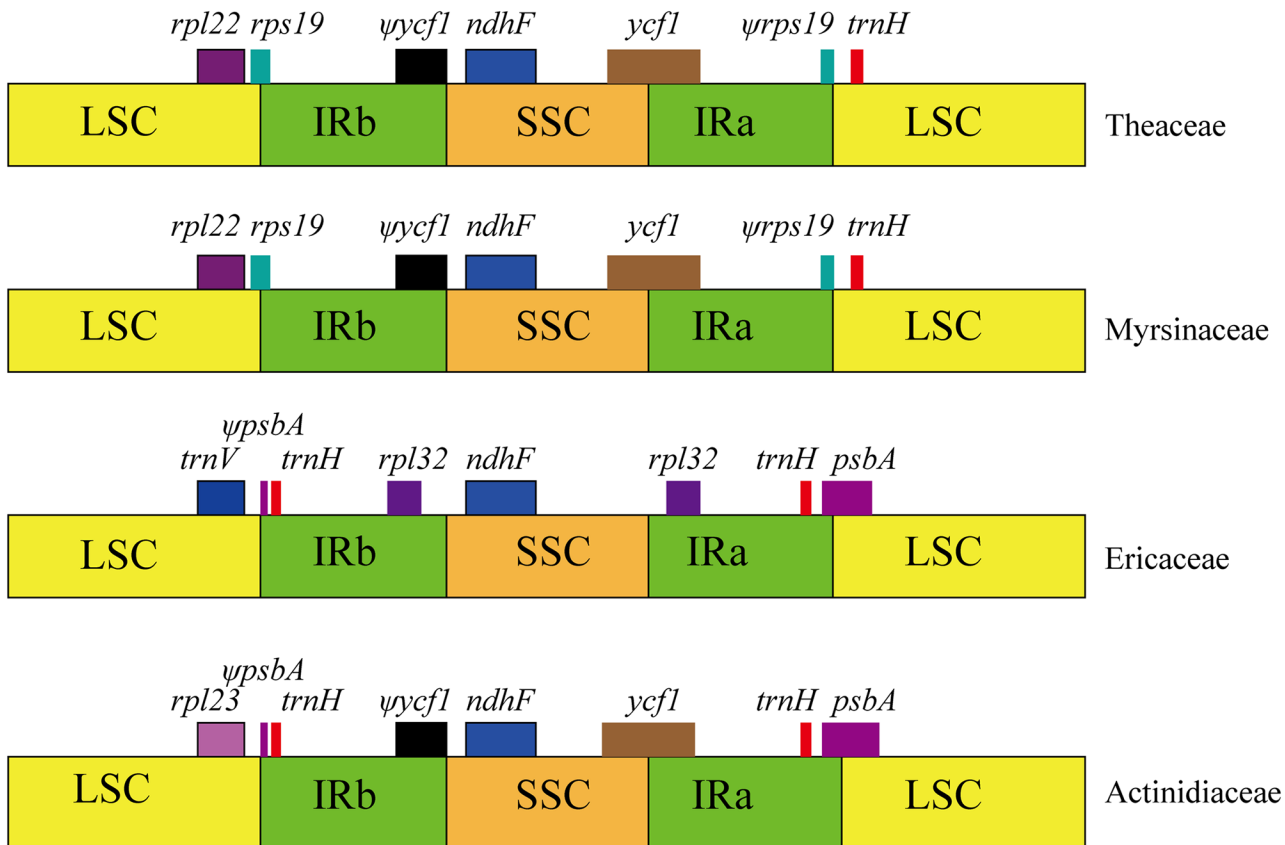


Fig 3. Comparison of the border positions of LSC, SSC, and IR regions in four basal asterid species. Boxes above the main line represent the genes at the IR/SC borders. The pseudogenes at the borders are shown by ψ.

doi:10.1371/journal.pone.0129347.g003

IR contraction

The IR boundaries were compared among four families in the asterids, including *Actinidia* (Actinidiaceae), *Vaccinium macrocarpon* (Ericaceae) [8], *Ardisia polysticta* (Myrsinaceae), [6] and *Camellia sinensis* (Theaceae) [7]. The cpDNA of *Actinidia* is collinear with the previously published plastomes of *Ardisia polysticta* and *Camellia sinensis* in gene order and overall homology. Fig 3 shows the detailed comparison of the IR/single copy (SC) boundaries between four representative members of basal asterids (*Camellia*, *Ardisia*, *Vaccinium*, and *Actinidia*). In the angiosperms, the downstream sequences of IRb/SSC are conserved, with the *ndhF* gene adjacent to it. In the *Actinidia*, *Camellia* and *Ardisia* cp genomes, the IR expands into the *ycf1* gene and inserts the *ycf1* pseudogene at the IRb/SSC border, whereas *ycf1* is lost in *Vaccinium*. For *Camellia* and *Ardisia*, the IRa/LSC junction is found within *rps19* and pseudogenes of *rps19* are located at the IRa/LSC boundary. However, in *Actinidia* the *rps19* gene does not extend into the IR region, and thus the *rps19* pseudogene is not observed. The IR extends into the *psbA* gene and inserts a short *psbA* pseudogene at the IRb/LSC border, which is similar to that of *Vaccinium*.

Compared to other species, the length of the IRs in the kiwifruit cp genome is fairly low; for example, in Theaceae, the size of the IR regions range from 26,025 bp to 26,057bp [7]; and in the basal asterids, such as Ericaceae, the length of the IRs are 34,232 bp [8]. The IR sequence in *Actinidia* is about 2 kb smaller than that in *Ardisia* cpDNA [6]. Hence, there has also been contraction of the IR in *Actinidia* at the IR/SSC boundary relative to the IRs in *Camellia* and

Ardisia. However, size variation by contractions within IR sequences contributes little to overall size variation between cp genomes of different taxa in asterids.

Phylogenetic analysis

To gain an insight into the position of the *Actinidia* within the asterids, we generated data sets of 72 protein-coding genes from the completely sequenced chloroplast genomes of 15 asterids and one outgroup (S2 Table). The 72 genes contain 8,671 potentially parsimony-informative and 9,342 uninformative nucleotide characters. Parsimony analyses yielded a single most-parsimonious tree (MPTs) with a length of 35,711, a consistency index of 0.66 and a retention index of 0.44 (Fig 4). The MP bootstrap tree resolved 16 nodes, of which 12 had strong bootstrap support (BS) of 99–100%, one had moderate support of 63%, and three had weak support at 5–30%. The results strongly support the position of the Actinidiaceae within Ericales, the basal position of Ericales within asterids, and the subdivision of euasterids into euasterids I (Gentianales, Lamiales, Solanales) and euasterids II (Apiales, Asterales). The phylogenetic trees in this study also indicates a close relationship between the Actinidiaceae and the Ericaceae, with high bootstrap support (100%). These results agree with data confirmed by phylogenetic methods based on 3 coding and 3 non-coding chloroplast DNA markers [45] and 78 orthologous chloroplast genes [6]. However, interrelationships among six families (Sarraceniaceae, Actinidiaceae, Roridulaceae, Clethraceae, Cyrillaceae, and Ericaceae) within the ericoid group based on these markers remain unresolved. The sequenced cp genomes provide large amounts of genetic information for improving resolution in phylogenetic studies. Thus, expanded taxon sampling will be required to acquire the accurate relationship in asterids.

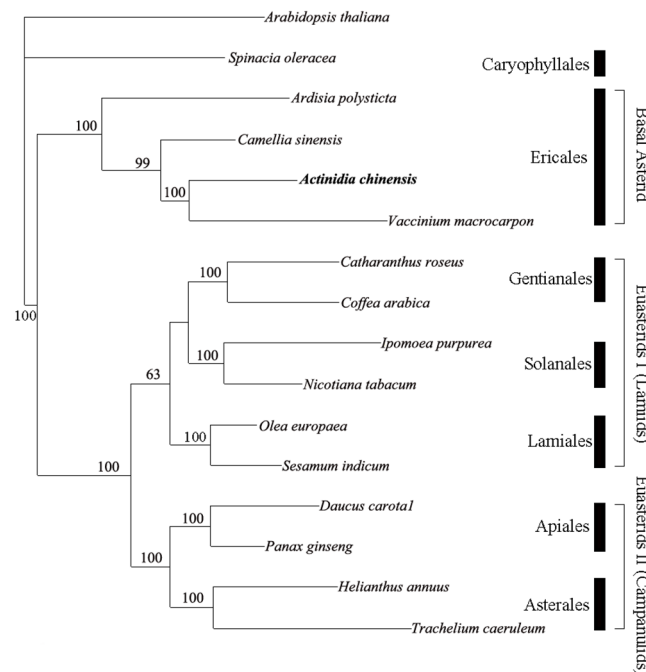


Fig 4. Phylogenetic position of *Actinidia* as inferred by MP analyses of 72 protein-coding genes. The MP tree has a length of 35,711, with a consistency index of 0.66 and a retention index of 0.44. Numbers above the lines indicate the maximum parsimony bootstrap value > 50% for each clade. The position of *Actinidia* is shown in boldface.

doi:10.1371/journal.pone.0129347.g004

Conclusions

The complete chloroplast sequence of *Actinidia* of the basal asterid was obtained with Illumina sequencing technology and Sanger sequencing. This is the first chloroplast genome sequenced in the Actinidiaceae family. The chloroplast genome of *Actinidia* has a very similar size and organization to those of other sequenced angiosperms. However, IR contraction was observed in *Actinidia* by comparing its cp genome with those of others in the asterids. Furthermore, we found that the *clpP* gene was lost in the *Actinidia* chloroplast genomes and may have been transferred to the nucleus during chloroplast evolution. The abundant and variable cpSSR loci identified in *Actinidia* will be useful in characterizing the population genetics of *Actinidia*. The phylogenetic relationships yielded by a dataset of sequences of 72 shared protein-coding genes of *Actinidia* and 14 other asterids genomes strongly support the placement of Actinidiaceae in Ericales within the basal asterids. Our data presented in this paper should provide important information to facilitate plastid genetic engineering of kiwifruit. For instance, the complete DNA sequence of chloroplast genomes of *Actinidia* is helpful to identify the optimal intergenic spacers for transgene integration and to develop kiwifruit-specific chloroplast vector.

Supporting Information

S1 Table. Primers used for gap closure and assembly.

(DOCX)

S2 Table. Accession numbers of plastome sequences of asterids included in phylogenetic analyses.

(DOCX)

Acknowledgments

We wish to thank Peter W. Fritsch and Lei Gao for their valuable comments on the manuscript. This work was partly supported by the Natural Scientific Foundation of China (31370251) and Fund of Science and Technology Service, Chinese Academy of Sciences (KFJ-EW-ST5-076).

Author Contributions

Conceived and designed the experiments: XHY ZZZL HWH. Performed the experiments: PT XHY. Analyzed the data: XHY ZZZL. Contributed reagents/materials/analysis tools: XHY YFL DWL. Wrote the paper: XHY HWH.

References

1. Neuhaus HE, Emes MJ (2000) Nonphotosynthetic Metabolism in Plastids. *Annu Rev Plant Physiol Plant Mol Biol* 51: 111–140. PMID: [15012188](#)
2. Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, editor. *Molecular Biology of Plastids*. San Diego, CA: Academic Press. pp. 5–53.
3. Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ, editor. *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*. Cambridge, MA: CABI. pp. 45–68.
4. Lee HL, Jansen RK, Chumley TW, Kim KJ (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol* 24: 1161–1180. PMID: [17329229](#)
5. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, et al. (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2043–2049. PMID: [16453699](#)

6. Ku C, Hu JM, Kuo CH (2013) Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. PLoS ONE 8(4): e62548. doi: [10.1371/journal.pone.0062548](https://doi.org/10.1371/journal.pone.0062548) PMID: [23638113](https://pubmed.ncbi.nlm.nih.gov/23638113/)
7. Yang JB, Yang SX, Li HT, Yang J, Li DZ (2013) Comparative chloroplast genomes of *Camellia* Species. PLoS ONE 8: e73053. doi: [10.1371/journal.pone.0073053](https://doi.org/10.1371/journal.pone.0073053) PMID: [24009730](https://pubmed.ncbi.nlm.nih.gov/24009730/)
8. Fajardo D, Senalik D, Ames M, Zhu HY, Steffan SA, Harbut R, et al (2013) Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. Tree Genet Genomes 9: 489–498.
9. Li JQ, Li XW, Soejarto DD (2007) Actinidiaceae. In: Wu ZY, Raven PH, Hong DY, editors. Flora of China. Beijing: Science Press and Missouri: Missouri Botanical Garden. pp. 334–360.
10. Huang H, Zhong C, Jiang Z, Li XW, Yao XH, Li DW, et al. (2013) *Actinidia*: taxonomy, resources, domestication, cultivation. Beijing, Science Press.
11. Huang H, Ferguson AR (2007) Genetic resources of kiwifruit: domestication and breeding. Hortic Rev 33: 1–121.
12. Belrose, Inc. (2013) World Kiwifruit Review. 2013, editor. Pullman, WA: Belrose, Inc.
13. Scortichini M, Marcelletti S, Ferrante P, Petriccione M, Firrao G (2012) *Pseudomonas syringae* pv. *actinidiae*: a re-emerging multi-faceted, pandemic pathogen. Molecular Plant Pathol 13: 631–640. doi: [10.1111/j.1364-3703.2012.00788.x](https://doi.org/10.1111/j.1364-3703.2012.00788.x) PMID: [22353258](https://pubmed.ncbi.nlm.nih.gov/22353258/)
14. Daniell H, Khan MS, Allison L (2002) Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Sci 7: 84–91. PMID: [11832280](https://pubmed.ncbi.nlm.nih.gov/11832280/)
15. Maliga P (2004) Plastid transformation in higher plants. Annu Rev Plant Biol 55: 289–313. PMID: [15377222](https://pubmed.ncbi.nlm.nih.gov/15377222/)
16. Li JQ, Huang H, Sang T (2002) Molecular phylogeny and infrageneric classification of *Actinidia* (Actinidiaceae). Syst Bot 27: 408–415.
17. Chat J, Jauregui B, Petit RJ, Nadot S (2004) Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. Am J Bot 91: 736–747. doi: [10.3732/ajb.91.5.736](https://doi.org/10.3732/ajb.91.5.736) PMID: [21653428](https://pubmed.ncbi.nlm.nih.gov/21653428/)
18. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829. doi: [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107) PMID: [18349386](https://pubmed.ncbi.nlm.nih.gov/18349386/)
19. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255. PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
20. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32: 11–16. PMID: [14704338](https://pubmed.ncbi.nlm.nih.gov/14704338/)
21. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964. PMID: [9023104](https://pubmed.ncbi.nlm.nih.gov/9023104/)
22. Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52: 267–274. PMID: [17957369](https://pubmed.ncbi.nlm.nih.gov/17957369/)
23. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32: 273–279.
24. Kurtz S, Choudhuri JV, Ohlebusch E, Scheleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29: 4366–4642.
25. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor and Appl Genet 106: 411–422. PMID: [12589540](https://pubmed.ncbi.nlm.nih.gov/12589540/)
26. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 1–19. PMID: [14706121](https://pubmed.ncbi.nlm.nih.gov/14706121/)
27. Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods), Version 4b 10. Sunderland, Massachusetts: Sinauer.
28. Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818. PMID: [9918953](https://pubmed.ncbi.nlm.nih.gov/9918953/)
29. Hudson KR, Gardner RC (1988) Organisation of the chloroplast genome of kiwifruit (*Actinidia deliciosa*). Curr Genet 13:339–342.
30. Tsudzuki J, Ito S, Tsudzuki T, Wakasugi, Sugiura M (1994) A new gene encoding tRNA (Pro) (GGG) is present in the chloroplast genome of black pine: a compilation of 32 tRNA genes from black pine chloroplasts. Curr Genet 26: 153–158. PMID: [8001170](https://pubmed.ncbi.nlm.nih.gov/8001170/)

31. Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, Tsudzuki J, et al. (1997) Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc Natl Acad Sci USA* 94: 5967–5972. PMID: [9159184](#)
32. Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* 46: 85–94. PMID: [16553897](#)
33. Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci USA* 92: 7759–7763. PMID: [7644491](#)
34. Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci USA* 100: 8828–8833. PMID: [12817081](#)
35. Kleine T., Maier U.G. and Leister D. (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* 60: 115–138. doi: [10.1146/annurev.arplant.043008.092119](#) PMID: [19014347](#)
36. Jansen RK, Saski C, Lee S-B, Hansen AK, Daniell H (2011) 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* 28: 835–847. doi: [10.1093/molbev/msq261](#) PMID: [20935065](#)
37. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J* 10: 3073–3078. PMID: [1915281](#)
38. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13: 645–658. PMID: [11251102](#)
39. Cusack BP, Wolfe KH (2007) When gene marriages don't work: divorce by subfunctionalization. *Trends Genet* 23: 270–272. PMID: [17418444](#)
40. Ueda M, Fujimoto M, Arimura SI, Murata J, Tsutsumi N, Kadowaki KI (2007) Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene* 402: 51–56. PMID: [17728076](#)
41. Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, et al. (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res* 20: 1700–1710. doi: [10.1101/gr.111955.110](#) PMID: [20978141](#)
42. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76: 273–97. doi: [10.1007/s11103-011-9762-4](#) PMID: [21424877](#)
43. Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104: 19369–19374. PMID: [18048330](#)
44. Haberle R, Fourcade HM, Boore J, Jansen R (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol* 66: 350–361. doi: [10.1007/s00239-008-9086-4](#) PMID: [18330485](#)
45. Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, et al. (2002) Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol Phylogenet Evol* 24: 274–301. PMID: [12144762](#)