Chloroplast Genome Evolution in Actinidiaceae: clpP Loss, Heterogenous Divergence and Phylogenomic Practice

Wen-Cai Wang2, Si-Yun Chen3, Xian-Zhi Zhang1*

1 College of Forestry, Northwest A&F University, Yangling, Shaanxi, China, 2 School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, United Kingdom, 3 Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

* zhangxianzhi@nwsuaf.edu.cn

Abstract

Actinidiaceae is a well-known economically important plant family in asterids. To elucidate the chloroplast (cp) genome evolution within this family, here we present complete genomes of three species from two sister genera (Clematoclethra and Actinidia) in the Actinidiaceae via genome skimming technique. Comparative analyses revealed that the genome structure and content were rather conservative in three cp genomes in spite of different inheritance pattern, i.e. paternal in Actinidia and maternal in Clematoclethra. The clpP gene was lacked in all the three sequenced cp genomes examined here indicating that the clpP gene loss is likely a conspicuous synapomorphic characteristic during the cp genome evolution of Actinidiaceae. Comprehensive sequence comparisons in Actinidiaceae cp genomes uncovered that there were apparently heterogenous divergence patterns among the cpDNA regions, suggesting a preferred data-partitioned analysis for cp phylogenomics. Twenty non-coding cpDNA loci with fast evolutionary rates are further identified as potential molecular markers for systematics studies of Actinidiaceae. Moreover, the cp phylogenomic analyses including 31 angiosperm plastomes strongly supported the monophyly of Actinidia, being sister to Clematoclethra in Actinidiaceae which locates in the basal asterids, Ericales.

Introduction

Actinidiaceae as an economically important plant family includes the well-known kiwifruit species, i.e. Actinidia chinensis Planch. and A. chinensis var. delicosa H.L.Li that can be eaten as fresh fruit or made into juice, jelly and dried fruit. There are c. 357 species included in three genera, i.e. Actinidia Lindl., Clematoclethra Maxim. and Saurauia Willd., in the kiwifruit family [1]. This family is widely distributed in Asia and the Americas displaying highly diverse morphological traits. In Clematoclethra species have bisexual flowers in one individual i.e. hermaphrodites, while in Actinidia species are dioecious or functionally dioecious [1–3]. The chloroplast (cp) genomes of Clematoclethra and Actinidia descend differently, that is with maternal and paternal transmission in Clematoclethra and Actinidia respectively [4]. However, currently little is known about the cp genomes evolution in these two closely related lineages.
The size of cp genomes in flowering plants ranges from 19 [5] to 218 kb [6], typically consisting of a pair of inverted repeats (IRs), a large single-copy (LSC) region and a small single-copy (SSC) region with the former one separating the latter two apart [7]. In spite of large degree conservation of organization and structure of the cp genomes in comparison to nuclear and mitochondrial genomes, gene losses and/or additions within cp genomes have been observed in several angiosperm lineages, e.g. orchids [8], Campanulaceae [9] and Rafflesiaaceae [10]. Additionally, gene transfer between plastome, chondrome and nucleus has also been revealed in plants [11–13]. It has been proven that cp genome structural variations are accompanied by speciation over time, which likely provides evolutionary information [14]. Whether there are common structure variations in the cp genomes alongside the evolution of Actinidiaceae is so far little known.

Rates of nucleotide change between non-coding (intergenic regions and introns) and coding (exons of protein-coding genes) sequences can vary tremendously within a given taxonomic group [15–16]. As lack of phylogenetic information is one of the major problems leading to incompletely resolved phylogenetic trees [17], thus to generate maximum phylogenetic signal, the non-coding sequences should be especially considered, in particular when inferring phylogenies at lower taxonomic levels e.g. at generic level [18–19]. By far the phylogeny of Actinidiaceae, particularly the genus Actinidia, remains largely elusive and controversial [20–21]. Therefore, it is valuable to develop global non-coding molecular markers from Actinidiaceae plastomes to resolve phylogenetic relationships within this family (e.g. the genus Actinidia).

Genome skimming technique is so far one of the simplest and the most economical methodologies to obtain high-copy sequences [22–23], for instance plastome sequences, and to conduct intergeneric or family-wide phylogenomic analyses [24–26]. It is now convenient to obtain complete cp genome sequences and mine markers for plant systematics via genome skimming [27–28]. To date, there are few cp genomes that have been sequenced from basal asterids lineages [29–32]. As one of them, Actinidiaceae has only one species (A. chinensis) been sequenced the cp genomes [33]. Undoubtedly, more sequenced cp genomes from other species of Actinidiaceae will not only improve our understanding of the cp genome evolution in the kiwifruit family, but will also aid resolving the phylogenetic relationships of the Actinidiaceae, even of the asterids [34]. In this study, we generated and characterized the complete cp genomes of one representative of Clematochlethra (C. lanosa) and two species of Actinidia (A. polygama and A. tetramera) from Actinidiaceae using genome skimming approach. By applying comparative genomic analyses we aimed to 1) test whether the cp genomes of Clematochlethra and Actinidia evolved congruently in spite of their contrasting genetic patterns; 2) identify the most variable cp genome-wide markers for subsequent evolutionary studies of the kiwifruit family; and 3) uncover the potentiality of cp phylogenomic practice at varying taxonomic levels in asterids.

**Materials and Methods**

**Plant materials and DNA sequencing**

Fresh leaves of Actinidia polygama (ZZX15021), Actinidia tetramera (ZZX15014) and Clematochlethra lanosa (ZZX15007) were collected from Huoditang forestry farm of Qinling Mountains in Shaanxi Province, China, in July 2015. Field studies were permitted and supported by Northwest A&F University to whom the Huoditang forestry farm belongs as an education and research experimental base. The voucher specimens of these three species were all deposited at the Trees Herbarium of Northwest A&F University.

Total genomic DNA was extracted from fresh leaf tissues of single individual per species using CTAB method [35]. Paired-end libraries (100 bp read length, insert size c. 500 bp) for
each species were constructed from fragmented genomic DNA following standard Illumina protocols (Illumina, California, USA). DNAs from different species were indexed by barcodes and pooled together. Prepared libraries for all three species were finally sequenced in one lane on the Illumina HiSeq 2000 platform in Beijing Genomics Institute (BGI) in Shenzhen, China.

**Genome assembly, annotation and repeat analysis**

Raw reads of each species were quality-filtered and de novo assembled using CLC Genomics Workbench v7.5 software (CLC Bio, Aarhus, Denmark). Contigs with length < 300 bp and sequence coverage < 50 were discarded [36]. The remaining contigs were analyzed by a BLAST search (http://blast.ncbi.nlm.nih.gov/) against the cp genome of *A. chinensis* [33] which was used as a reference in this study. Aligned contigs with ≥ 90% similarity and query coverage were designated as cp genomic contigs and ordered according to the reference. Small gaps were filled using trimmed Illumina reads, as conducted in Bock et al. [37] and Zhang et al. [26].

Initial cp genome annotation was performed using DOGMA software [38]. Start/stop codons and intron/exon boundaries were adjusted manually by comparing to the available *A. chinensis* cp genome [33]. tRNA genes were further conformed using tRNAscan-SE 1.21 [39]. OGDRAW program [40] was applied to graphically display the physical map of the circular cp genomes.

Tandem and palindromic repeats were determined by applying program REPuter [41]. Minimal tandem repeat size was set to 10 bp, intervening size between palindromic repeats constrained to a maximal length of 2 kb. Simple sequence repeats (SSRs) were detected using MISA perl script (http://pgrc.ipk-gatersleben.de/misa/). Minimum repeat unit was defined as 10 for mono-, 6 for di-, and 5 for tri-, tetra-, penta-, and hexanucleotide SSRs [42].

**Genome comparison and divergent hotspot identification**

Chloroplast genome sequences of two additional Actinidiaceae species, i.e. *A. chinensis* (NC_026690) and *A. chinensis* var. *deliciosa* (NC_026691), were downloaded from GenBank. Together with our own sequenced cp genomes of *A. polygama*, *A. tetramera* and *C. lanosa*, we aligned these five cp genomes using MAFFT program [43] and manually adjusted where necessary. Full alignments with annotations were visualized using VISTA viewer [44] by referring to *A. chinensis* [33]. Multiple alignments of LSC, SSC, IRa (representing the two identical IRs), coding and non-coding regions in the matrix of five cp genomes were extracted respectively. Each region was compared to tally numbers of constant and variable alignment positions in MEGA v5.2 [45]. The number of nucleotide substitutions, in terms of single nucleotide polymorphisms (SNPs), divided by the total aligned length in each partition resulted in the percent variability for each region compared [19].

Five Actinidiaceae cp genomes were analysed to identify rapidly evolving molecular markers which can be used in subsequent phylogenetic studies. Both the coding and non-coding DNA fragments > 200 bp in each cp genome were extracted separately by applying “Extract Sequences” option of DOGMA [38]. Then the homologous loci were aligned individually using MUSCLE [46] implemented in Geneious v9.0 [47] with default settings. Manual adjustments were made where necessary. The proportion of mutational events for each coding and non-coding loci was calculated as follows: the proportion of variation = (NS/L) × 100, where NS = the number of nucleotide substitutions, L = the aligned sequence length.

**Phylogenomic analyses**

Thirty-one complete cp genomes representing eight major lineages of the superasterids were included for phylogenomic analyses (S1 Table). *Arabidopsis thaliana* (L.) Heynh. and *Vitis*
vitis L. from superrosids were defined as outgroups according to previous studies [34, 48]. Given concerns about the likely effects of missing data on phylogenetic inferences [49], we analyzed the following two kinds of dataset accordingly: first, 79 plastid protein-coding unique genes of Actinidaeae (S2 Table) were used, while that absent in other taxa were treated as missing data (“data-incomplete”); second, analyses based merely on the 56 common protein-coding genes (S3 Table) shared by all the representatives (S1 Table) were conducted (“data-complete”). Each gene was aligned individually using MUSCLE [46] in Geneious v9.0 [47], and then concatenated as a supermatrix for phylogenetic analyses. Gaps were not coded in both datasets.

Phylogenetic analyses of the concatenated datasets were carried out by maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). MP analyses were implemented in PAUP* v4.0b10 [50]. Parsimony heuristic tree searches were performed with 1000 random addition sequence replicates, TBR branch swapping, MulTrees option in effect. Branch support (MPBS) was evaluated by 1000 bootstrap replicates [51].

ML and BI trees were inferred using RAxML v.8.2.8 [52] and MrBayes 3.2.6 [53] respectively in the CIPRES Science Gateway v3.3.3 [54], with concatenated supermatrix partitioned by genes. RAxML searches relied on the GTR + CAT model of nucleotide substitution, and 1,000 fast bootstrap ML reps were performed to assess the relative degree of support (MLBS) for internal nodes. In Bayesian analysis, two runs with four chains were run for 20,000,000 generations under the model GTR + G, with a sampling every 1,000 generations till convergence (the average standard deviation of split frequencies was less than 0.01). The first 25% of trees were discarded as burn-in, and the remaining trees were used to construct majority-rule consensus tree and estimate posterior probabilities (PP).

Results

Genome assembly and features

About 20 million reads were generated via genome skimming for each species. There were 516,840, 332,220, and 401,858 paired-end reads that were aligned for *A. polygama*, *A. tetramera* and *C. lanosa*, respectively (Table 1). Based on de novo and reference-assisted assembly, complete cp genomes of the three species were obtained with high genome coverage (> 200×) (Table 1). The sequences of three cp genomes were deposited in GenBank (accession numbers: KX345297-KX345299).

Table 1. Statistic summary of the chloroplast genome sequencing, assembly and features.

|                  | *Actinidia polygama* | *Actinidia tetramera* | *Clematoclethra lanosa* |
|------------------|----------------------|-----------------------|-------------------------|
| Total paired-end reads | 20,224,568           | 20,188,340            | 20,200,620              |
| Aligned paired-end reads | 516,840              | 332,220               | 401,858                 |
| Mean coverage    | 330.1                | 210.8                 | 256.4                   |
| Size (bp)        | 156,583              | 157,659               | 156,731                 |
| LSC length (bp)  | 88,568               | 89,572                | 89,032                  |
| SSC length (bp)  | 20,397               | 20,497                | 20,399                  |
| IRs length (bp)  | 23,809               | 23,795                | 23,650                  |
| Number of genes  | 131                  | 131                   | 131                     |
| Protein-coding genes | 83                   | 83                    | 83                      |
| Structure RNAs   | 48                   | 48                    | 48                      |
| GC content (%)   | 37.2                 | 37.1                  | 37.1                    |
| Coding regions (%) | 49.4                | 49.3                  | 50.0                    |

doi:10.1371/journal.pone.0162324.t001
Eventually determined cp genome size of *A. polygama*, *A. tetramer* and *C. lanosa* was 156,583, 157,659 and 156,731 bp respectively (Table 1). All three complete cp genomes exhibited the typical quadripartite structure of land plants, consisting of a pair of IRs (23,650–23,809 bp) separated by one LSC (88,568–89,572 bp) and one SSC (20,397–20,497 bp) region (Table 1). An identical set of 131 genes with the same linear order were encoded in three cp genomes. Here we reported the cp genome map of *C. lanosa* as the representative (Fig 1).

![Gene map of Clematoclethra lanosa representing the chloroplast genome structure of three Actinidiaceae species.](image-url)
genes are unique, of which 79 are protein-coding genes (Fig 1). The clpP gene was undetectable in both Clematoclethra and Actinidia cp genomes (Fig 1, S2 Table). Among the 11 intron-containing genes (atpF, ndhA, ndhB, petB, petD, rpl2, rpl16, rpoC1, rps12, rps16, ycf3), only rps12 and ycf3 contained two intervening introns, while others contained one intron each. The boundaries between IRs and single copy regions were conserved in Clematoclethra and Actinidia cp genomes, with ycf1 and psbA extending into the IRb (Fig 1).

Three classes of repeats (tandem, palindromic and SSRs) were investigated in our sequenced cp genomes. Tandem repeats that ranged from 10 to 20 bp were the most abundant, that had their sizes between 20 and 30 bp were secondarily abundant, while that with lengths greater than 30 bp were much less in all three species (Fig 2A). Palindromic repeats were much more in the cp genome of C. lanosa (99) than that in two Actinidia species (34 and 41 respectively) (Fig 2B). Analysis of occurrence and type of SSRs indicated that four categories of SSRs, i.e. mono-, di-, tri-, and hexa-nucleotide were detected. Mononucleotide repeats were the most dominant, while other types were far less in the cp genomes of three species (Fig 2C). Not all the SSRs mentioned above occurred in all the species, that is, there were three types in A. polygama (mono-, di- and tri-nucleotide) and A. tetramera (mono-, di- and hexa-nucleotide), but only two types in C. lanosa (mono- and tri-nucleotide) (Fig 2C).

**Sequence divergence estimation and molecular marker identification**

Sequence divergences among the five Actinidiaceae species (A. chinensis, A. chinensis var. deliciosa, A. polygama, A. tetramera and C. lanosa) were visually quantified by “percent identity plots” and estimated by the percentage of nucleotide substitutions (SNPs) (Fig 3). Global alignments further confirmed that the five cp genomes of Actinidiaceae were perfectly syntenic (Fig 3A). Across the whole cp genomes, we found that sequence variations were not uniform but highly heterogenous among different regions (Fig 3A). The estimation of the percentage of SNPs in LSC, SSC and IRs (here in terms of IRa) regions suggested that sequences of the IRs diverged at slower rates (1.21%) than that of their adjacent single-copy regions, LSC (2.94%) and SSC (3.92%) (Fig 3B). Coding regions (2.30%) were more conserved than non-coding regions (3.81%) as expected (Fig 3B).
Distribution patterns of variable characters (uninformative + informative SNPs) differed largely in coding and non-coding loci (> 200 bp) in these five Actinidiaceae cp genomes (Fig 4). It showed that the proportion of variability in non-coding loci ranged from 0.00% to 9.04% with mean value as 3.36% which was 1.6 times greater than that in the coding loci (2.01% on average) (Fig 4). Variations in most (57 out of 65) of the coding DNA fragments were less than 3% (Fig 4A). Only three genes i.e. rpl20, accD and ycf1 had their variations exceeded 5%. In contrast, there were 32 non-coding loci showed relative high degree of divergence (above 3.6%, Fig 4B). Considering the relatively high percentage of variations and the convenience of universal primers developing for PCR and sequencing experiments, we chose 20 highly variable intergenic loci as potential molecular markers for subsequent phylogenetic studies (Table 2). Percentage of variations of these 20 non-coding loci all exceeded 3.7%, among which 16 had a percentage of variable characters (VCs) greater than 4% (Table 2).
Phylogenomic reconstruction

The final length of the aligned matrix for the “data-incomplete” and “data-complete” data sets contained 91,222 and 58,876 unambiguously aligned nucleotide characters respectively. Because these two data sets generated congruent phylogenetic trees, we here only reported the results based on the 56 common protein-coding plastid genes (“data-complete” data set) as shown in Fig 5. Asterids was highly supported as monophyletic (100/100/1.00) and the sister relationship of lamiids and campanulids was well supported (100/100/1.00). Corales was resolved as the basal most group in the asterids (100/100/1.00), and the Ericales diverged subsequently (100/100/1.00). Caryophyllales was revealed as the sister to asterids in the superasterids clade (100/100/1.00). The sister relationship of Ericales and core asterids (lamiids + campanulids) also obtained high support values (100/100/1.00).

The Actinidiaceae were recovered as a monophyly in Ericales with high statistic supports (100/100/1.00). The phylogenomic tree indicated a branching order of Actinidiaceae, Ericaceae, Primulaceae and Theaceae as (((Actinidiaceae, Ericaceae), Theaceae), Primulaceae) in Ericales, each node getting high support values (100/100/1.00). The sampled four Actinia taxa clustered in a clade with high statistic supports (100/100/1.00), being sister to Clematoclethra. A. tetramer a diverged firstly (95/97/1.00), followed by A. polygama (100/100/1.00) in the Actinidia lineage. As expected, A. chinensis and A. chinensis var. deliciosa were closely related (100/100/1.00).

Discussion

The conserved chloroplast genomes of Clematoclethra and Actinidia

Selective pressure leads to divergent traits during plant evolution, which may be tracked in chloroplast genomes, for instance, the non-photosynthetic lifestyle of certain plant groups such as Epipogium aphyllum and Epipogium roseum [5], Orobanche [13], and Petrosavia stellaris [55] tends to be associated with the enormously reduced cp genome size. Complete sequencing and comparative analyses here revealed that the Clematoclethra species (C. lanosa) bears a highly conserved cp genome in terms of its architecture and linear sequence order with either the two newly sequenced Actinidia (A. polygama and A. tetramer a) in this study or the published A. chinensis cp genome [33]. Chloroplast genome size of C. lanosa (156.7 kb) is similar to that of A. polygama (156.6 kb) and A. tetramer a (157.7 kb) (Table 1), showing a typical cp genome size range (115 to 165 kb) of angiosperms [56]. Although the different inheriting patterns of the Actinidia and Clematoclethra cp genomes, i.e. paternal passing in Actinidia and maternal in Clematoclethra [4], no significant structural variation was observed (Figs 1 and 3A). The gene content, gene order, GC content, size of IRs regions, and junctions of single copy and IR regions (Table 1, Fig 1) all indicate highly syntenic nature of the cp genomes in the two closely related genera within Actinidiaceae.

clpP gene was completely lost in the sequenced cp genomes of Actinidia [this study, 33] and also was absent in Clematoclethra (C. lanosa). The lack of certain plastid genes is not unusual since previous studies illustrated that several cp genes have been lost in certain groups, such like, the clpP gene in Primula poissonii [31] and Camellia oleifera [32], the accD gene in Trifolium [57], the rpl32 gene in Populus [58], and the infA gene in rosids [59]. The clpP gene loss detected in both Clematoclethra and Actinidia is most likely a conspicuous common feature during their cp genomes evolution, suggesting that the clpP gene loss event might happen in their common ancestor. More species from the kiwifruit family, especially representatives of the third genus Saurauia, will be sequenced to determine whether clpP gene loss is synapomorphy of the Actinidiaceae.
Occurrences and types of tandem repeats are quite similar with approximately the same levels in each cp genomes (Fig 2A). However, the palindromic repeats in the cp genome of *C. lanosa* were more than that in two *Actinidia* species (Fig 2B). Different abundance of palindromic repeats in the cp genomes of *Clematoclethra* and *Actinidia* would likely provide additional evolutionary information since presence and abundance of repetitive sequences in cp or nuclear genomes may contain phylogenetic signal [16, 60]. Mononucleotide SSRs are abundant in all the three cp genomes (Fig 2C), which can be used as potential cpSSR markers in studies on population genetics of kiwifruit species [33].

**Heterogenous divergence patterns in chloroplast genomes and molecular markers development**

Alignment of our three cp genomes and two other published cp genomes of Actinidiaceae (*A. chinensis* and *A. chinensis* var. *deliciosa*) [33] uncovered obvious heterogenous sequence...
divergence within the Actinidiaceae cp genomes (Fig 3A). The single-copy regions (LSC and SSC) show a higher sequence divergence than that in the inverted repeat regions (in terms of IRa) (Fig 3B). This is possible because the gene conversion of IRs constrained the nucleotide variations of inverted repeat sequences [61], which results in the lower genetic divergence of IRs. The coding regions are more conserved than non-coding regions (Fig 3B), likely resulting from the strong selective pressure on the function constraint of coding sequences [62]. Different regions of cp genomes display varied evolutionary rates reflecting that data-partitioned cp phylogenomic analysis such as partitioning cp genomes data by LSC, SSC and IRs [16], by coding and non-doing regions [63], or by genes [27–28], may be more reasonable in the Actinidiaceae cp phylogenomic practice.

Sequence variation analyses of the coding and non-coding loci (longer than 200 bp) across the five examined cp genomes indicate that each locus shows independent/distinct evolutionary patterns (Fig 4) as revealed in Poaceae [64], mimosoid legume [65] and Silene [66]. In general, coding genes would have their sequence variations (percentage of variable characters) lower than 4% but some exceptions are observed here, for instance, accD (in the LSC region), rpl20 (in the LSC region) and ycf1 (in the SSC region). These exceptions display more SNPs than all other coding loci particularly with rpl20 as one which varies the most quickly. There have been some observations of accelerated variation rate of the plastid genes, such as psb in Poaceae [67], rps in Saxifragales [68], and rps and clpP genes in Caryophyllales [69]. Nevertheless, those speeded divergence observations were thought as results of abnormal DNA replication, repair or recombination [65, 70]. It is likely that the heterogenous divergences among plastid genes in Actinidiaceae probably are driven by similar mechanism as in the above described groups.
Chloroplast DNA fragments such like matK, psbC-trnS, rbcL, and trnL-trnF have been used most commonly in Actinidia phylogenetic studies [20–21, 71], but the resulted trees have not been resolved clearly. The difficulty mainly lies in the lack of phylogenetic information due to the low sequence variation of commonly used genetic markers [21]. Therefore, we need to develop rapidly evolving genetic markers for the phylogenetic study of Actinidiaceae, particularly of the genus Actinidia, which will shed light on our understanding of the evolutionary history of this valuable economical plant group. The average proportion of variability in non-coding loci is higher (3.36%) than that in the coding loci (2.01%) (Fig 4). In addition, it is convenient to design universal PCR and Sanger sequencing primers for non-coding DNA fragments with length from c. 300 bp to 1,500 bp by using the adjacent conserved coding sequences [18–19]. We thus choose 20 highly variable non-coding cpDNA loci as potential molecular markers (Table 2) for subsequent phylogenetic studies of the kiwifruit family. Some of these loci such as trnT(UGU)-trnL(UAA) and rpl32-trnL(UAG) has been used in phylogenetic studies of intractable plant groups, e.g. in woody bamboos (Poaceae, Bambusoideae) with relatively robust solutions [72]. These 20 cp genome-wide intergenic loci will probably provide sufficient phylogenetic information for the reconstruction of the phylogeny of Actinidiaceae.

Chloroplast phylogenomic practice

The presence of missing data is not unusual in phylogenomic analyses, though its impact on phylogenetic evaluation has been widely debated [73–76]. Using our sequenced three cp genomes via genome skimming together with 28 other publically available plastomes data (S1 Table), we reconstructed the phylogeny of Ericales and its related groups. The “data-incomplete” and “data-complete” data sets give rise to a congruent topology of the 31 sampled taxa.
with high level of confidence (Fig 5, from the “data-complete” data set), which indicates that the relatively small missing data (23 out of 79 plastid genes) probably have no ill impact on the phylogenetic inferences.

The relationships derived from our phylogenomic analyses among the major lineages of asterids, i.e. Cornales, Ericales and core asterids (lamkiads + campanulids), are highly supported as stated previously [34, 46, 77]. The branching orders of species within lamkiads and campanulids are consistent with recent studies [78–79] as ((Solanales, Lamiales), ((Dipsacales, Apiales), Asterales)). Our analyses also recover the phylogenetic relationships of four sampled families in Ericales as (((Actinidiaceae, Ericaceae), Theaceae), Primulaceae), in agreement with the result of Yao et al. [33]. The Actinidiaceae is a monophyly with the sampled four *Actinidia* taxa clustering in a clade as sister to *Clematoclethra* (Fig 5). The sister relationship of *A. chinensis* and *A. chinensis* var. *deliciosa* is consistent with the results of previous studies [20, 71]. Moreover, phylogenetic placements of *A. polygama* and *A. tetramera* are also resolved with high supports within *Actinidia*.

To date, the genome skimming method has been thought as an economical and advanced technique with great potential not only in the plastomes but also for the mitochondrial and nuclear genomes studies [80–83]. Chloroplast genome has been broadly employed as a routine practice in the plant systematics because of its essentially recombination-free and high degree of identical copies per cell [84]. Applying genome skimming on the expanded taxon samples will facilitate the phylogenomic study of the large asterids clade in angiosperm.

**Conclusions**

In summary, in the present study we generated cp genomes of three species from the family Actinidaceae, including one species of *Clematoclethra* and two species of *Actinidia*, using genome skimming approach. These finished cp genomes will facilitate the development of plastid genetic engineering for these valuable fruit species and provide additional information about the evolutionary history of cp genomes in the whole kiwifruit family. The cp genomes of *Clematoclethra* and *Actinidia* are highly conserved in the structure and content in spite of their contrasting genetic processes of plastomes. The clpP gene loss likely provides a prominent synapomorphic characteristic of Actinidaceae. Obviously heterogenous divergence patterns in different regions of the Actinidaceae cp genomes are revealed, according to which 20 rapidly evolving intergenic loci are identified as potential cp genome-wide markers for subsequent phylogenetic studies of this family. The Actinidaceae and its subordinate genera are both strongly supported as monophyletic located in the Ericales in basal asterids. The phylogenetic relationships within the large asterids clade, even within the enigmatical genus *Actinidia*, have been robustly resolved based on the cp genomes data, suggesting the bright probability of cp phylogenomics for tackling difficult phylogenies in flowering plants.

**Supporting Information**

S1 Table. Taxa and GenBank accession numbers included in the phylogenomic analyses.  
(DOCX)

S2 Table. List of 79 unique plastid genes of Actinidaceae included in the “data-incomplete” data set of phylogenomic analyses.  
(DOCX)

S3 Table. List of 56 common unique plastid genes included in the “data-complete” data set of phylogenomic analyses.  
(DOCX)
Acknowledgments

We would like to thank Zheng-Shan He and Hong-Tao Li for the help in the data analyses. We are grateful to Zhi-Rong Zhang and Jun-Bo Yang for the help in the laboratory experiments.

Author Contributions

Conceptualization: W-CW X-ZZ.
Formal analysis: W-CW S-YC X-ZZ.
Funding acquisition: X-ZZ.
Investigation: W-CW.
Project administration: X-ZZ.
Resources: S-YC X-ZZ.
Software: S-YC.
Supervision: X-ZZ.
Writing – original draft: W-CW X-ZZ.
Writing – review & editing: W-CW X-ZZ.

References

1. Li JQ, Li XW, Soejarto DD. Actinidiaceae. In: Wu ZY, Raven PH, Hong DY, editors. Flora of China. Beijing: Science Press and Missouri: Missouri Botanical Garden. pp. 334–360. 2007.
2. Ming R, Bendahmane A, Renner SS. Sex chromosomes in land plants. Annual Review of Plant Biology. 2011; 62:485–514. doi: 10.1146/annurev-arplant-042110-103914 PMID: 21526970
3. Renner SS. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. American Journal of Botany. 2014; 101:1588–1596. doi: 10.3732/ajb.1400196 PMID: 25326608
4. Chat J, Chalak L, Petit R. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. Theoretical and Applied Genetics. 1999; 99:314–322.
5. Schelkunov MI, Shratinikova VY, Nuraliev MS, Selosse MA, Penin AA, Logacheva MD. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids Epipogium aphylum and Epipogium roseum. Genome Biology and Evolution. 2015; 7:1179–1191. doi: 10.1093/gbe/evv019 PMID: 25635040
6. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of Pelargonium × hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Molecular Biology and Evolution. 2006; 23:2175–2190. PMID: 16916942
7. Wicke S, Schneeweiss GM, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology. 2011; 76:273–297. doi: 10.1007/s11103-011-9762-y PMID: 21424877
8. Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, et al. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Molecular Biology and Evolution. 2014; 31:3095–3112. doi: 10.1093/molbev/msu252 PMID: 25172958
9. Knox EB. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. Proceedings of the National Academy of Sciences. 2014; 111:11097–11102.
10. Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, et al. Possible loss of the chloroplast genome in the parasitic flowering plant Rafflesia lagascae (Rafflesiaceae). Molecular Biology and Evolution. 2014; 31:793–803. doi: 10.1093/molbev/msu051 PMID: 24458431
11. Jansen RK, Sasaki C, Lee S-B, 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
25. Weitemier K, Straub SC, Fishbein M, Liston A. Intragenomic polymorphisms among high-copy loci: a genome-wide study of nuclear ribosomal DNA in the grape family (Vitaceae). Peer J. 2015; 3:e718. doi: 10.7717/peerj.718 PMID: 25653903

26. Zhang N, Wen J, Zimmer EA. Congruent deep relationships in the grape family (Vitaceae) based on sequences of chloroplast genomes and mitochondrial genes via genome skimming. PLoS One. 2015; 10:e0144701. doi:10.1371/journal.pone.0144701 PMID: 26658830

27. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytologist. 2016; 209:855–70. doi: 10.1111/nph.13617 PMID: 26350789

28. Givnish TJ, Zuluaga A, Marques I, Lam VK, Gomez MS, Iles WJ, et al. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. Cladistics. 2016; 32:1–25.

29. Fajardo D, Senalik D, Ames M, Zhu H, Steffen SA, Harbut R, et al. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. Tree Genetics & Genomes. 2013; 9:489–98.

30. Ku C, Hu J-M, Kuo C-H. Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. PLoS One. 2013; 8:e62548. doi: 10.1371/journal.pone.0062548 PMID: 23638113

31. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Molecular Ecology Resources. 2014; 14:1024–31. doi: 10.1111/1755-0998.12251 PMID: 24620934
32. Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z. Comparative chloroplast genomes of Camellia species. PLoS One. 2013; 8:e73053. doi: 10.1371/journal.pone.0073053 PMID: 24009730
33. 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. doi: 10.1371/journal.pone.0129347 PMID: 26046631
34. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society. 2016; 0:1–20.
35. Doyle JJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin. 1987; 19:11–5.
36. Wu C-S, Chaw S-M. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Genome Biology and Evolution. 2015; 7:2000–2009. doi: 10.1093/gbe/evv125 PMID: 26116919
37. Bock DG, Kane NC, Ebert DP, Rieseberg LH. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytologist. 2014; 201:1021–1030. doi: 10.1111/nph.12560 PMID: 24245977
38. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004; 20:3252–3255. PMID: 15180927
39. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research. 2005; 33:686–689.
40. Lohse M, Drechsel O, Kahliu S, Bock R. OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research. 2013; gkt289.
41. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research. 2001; 29:4633–42. PMID: 11713313
42. Zhao H, Yang L, Peng Z, Sun H, Yue X, Lou Y, et al. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus Phyllostachys. Scientific Reports. 2015; 5:5018–5027.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 2013; 30:772–80. doi: 10.1093/molbev/mst010 PMID: 23329690
44. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Research. 2004; 32:273–279.
45. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 2011; 28:2731–2739. doi: 10.1093/molbev/msr121 PMID: 21546353
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32:1792–1797. PMID: 15034147
47. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28:1647–1649. doi: 10.1093/bioinformatics/bts199 PMID: 22543367
48. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al. Angiosperm phylogeny: 17 genes, 640 taxa. American Journal of Botany. 2011; 98:704–730. doi: 10.3732/ajb.1000404 PMID: 21613169
49. Wiens JJ, Morrill MC. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Systematic Biology. 2011; 60:1–13.
50. Swofford DL. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10. 2003.
51. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985; 39:783–791.
52. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 0.1–2.
53. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 2012; 61:539–542. doi: 10.1093/sysbio/sys029 PMID: 22357727
54. Miller MA, Pfeiffer W, Schwartz T, editors. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010; 2010: IEEE.
Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion.

60. Dodsworth S, Chase MW, Kelley LJ, Leitch IJ, Macas J, Novak P, Piednoel M, Weiss-Schneeweiss H, Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of
61. many parallel losses of
62. Jinping SS, Burke SV, Duvall MR. Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. Plant Systematics and Evolution. 2014; 300:1421–1436.
63. Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Molecular Phylogenetics and Evolution. 2013; 72:82–89. doi: 10.1093/molphy/mvs006 PMID: 24373909
64. Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. Molecular Biology and Evolution. 2002; 19:2084–2091. PMID: 12446800
65. Jones SS, Burke SV, Duvall MR. Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. Plant Systematics and Evolution. 2014; 300:1421–1436.
66. Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Scientific Reports. 2015; 5:16958–16972. doi: 10.1038/srep16958 PMID: 26592928
67. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus Silene. Genome Biology and Evolution. 2012; 4:294–306. doi: 10.1093/gbe/evs006 PMID: 22474429
68. Matsuo Y, Yamazaki Y, Ogihara Y, Tsunewaki K. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. Molecular Biology and Evolution. 2002; 19:2084–2091. PMID: 12446800
69. Dong W, Xu C, Cheng T, Zhou S. Complete chloroplast genome of Sedum sarmentosum and chloroplast genome evolution in Saxifragales. PLoS One. 2013; 8:e77965. doi: 10.1371/journal.pone.0077965 PMID: 24205047
70. Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. A recurring syndrome of accelerated plastid evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Molecular Phylogenetics and Evolution. 2014; 72:82–89. doi: 10.1016/j.mpev.2013.12.004 PMID: 24373909
71. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proceedings of the National Academy of Sciences. 2008; 105:18424–18429.
72. Li J, Huang H, Sang T. Molecular phylogeny and infrageneric classification of Actinidia (Actinidiaceae). Systematic Botany. 2002; 27:408–415.
73. Zhang X-Z, Zeng C-X, Ma P-F, Haevermans T, Zhang Y-X, Zhang L-N, et al. Multi-locus plastid phylogenetic biogeography supports the Asian hypothesis of the temperate woody bamboos (Poaceae: Bambusoideae). Molecular Phylogenetics and Evolution. 2016; 96:118–129. doi: 10.1016/j.mpev.2015.11.025 PMID: 26723898
74. Streicher JW, Schulte JA, Wiens JJ. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. Systematic Biology. 2015; 65:128–145. doi: 10.1093/sysbio/syv058 PMID: 26300450
75. Huang H, Knowles LL. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic Biology. 2014; 65:357–365. doi: 10.1093/sysbio/syu046 PMID: 24996413
75. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology. 2009; 58:130–145. doi: 10.1093/sysbio/syp017 PMID: 20525573
76. Jiang W, Chen S-Y, Wang H, Li D-Z, Wiens JJ. Should genes with missing data be excluded from phylogenetic analyses? Molecular Phylogenetics and Evolution. 2014; 80:308–18. doi: 10.1016/j.ympev.2014.08.006 PMID: 25124098
77. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms-infering the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evolutionary Biology. 2014; 14:23–49. doi: 10.1186/1471-2148-14-23 PMID: 24533922
78. Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences. 2015; 112:12764–12769.
79. Zanne AE, Tank DC, Cornell WK, Eastman JM, Smith SA, FitzJohn RG, et al. Three keys to the radiation of angiosperms into freezing environments. Nature. 2014; 506:89–92. doi: 10.1038/nature12872 PMID: 24362564
80. Dodsworth S. Genome skimming for next-generation biodiversity analysis. Trends in Plant Science. 2015; 20:525–527. doi: 10.1016/j.tplants.2015.06.012 PMID: 26205170
81. Wicke S, Schneeweiss GM. Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. Hörandl, E & Appelhans, MS (eds), Next-Generation Sequencing in Plant Systematics. 2015; Chapter 1. doi: 10.14630/000002
82. Hao W, Fan S, Hua W, Wang H. Effective extraction and assembly methods for simultaneously obtaining plastid and mitochondrial genomes. PLoS ONE. 2014; 9:e108291. doi: 10.1371/journal.pone.0108291 PMID: 25251391
83. Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, et al. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences. 2014; 2:1400042–1400048.
84. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biology. 2016; 17:134–162. doi: 10.1186/s13059-016-1004-2 PMID: 27339192