

Complete Chloroplast Genome of *Sedum sarmentosum* and Chloroplast Genome Evolution in Saxifragales

Wenpan Dong¹, Chao Xu^{1,2}, Tao Cheng^{1,2}, Shiliang Zhou^{1*}

1 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, **2** Graduate University of Chinese Academy of Sciences, Beijing, China

Abstract

Comparative chloroplast genome analyses are mostly carried out at lower taxonomic levels, such as the family and genus levels. At higher taxonomic levels, chloroplast genomes are generally used to reconstruct phylogenies. However, little attention has been paid to chloroplast genome evolution within orders. Here, we present the chloroplast genome of *Sedum sarmentosum* and take advantage of several available (or elucidated) chloroplast genomes to examine the evolution of chloroplast genomes in Saxifragales. The chloroplast genome of *S. sarmentosum* is 150,448 bp long and includes 82,212 bp of a large single-copy (LSC) region, 16,670 bp of a small single-copy (SSC) region, and a pair of 25,783 bp sequences of inverted repeats (IRs). The genome contains 131 unique genes, 18 of which are duplicated within the IRs. Based on a comparative analysis of chloroplast genomes from four representative Saxifragales families, we observed two gene losses and two pseudogenes in *Paeonia obovata*, and the loss of an intron was detected in the *rps16* gene of *Penthorum chinense*. Comparisons among the 72 common protein-coding genes confirmed that the chloroplast genomes of *S. sarmentosum* and *Paeonia obovata* exhibit accelerated sequence evolution. Furthermore, a strong correlation was observed between the rates of genome evolution and genome size. The detected genome size variations are predominantly caused by the length of intergenic spacers, rather than losses of genes and introns, gene pseudogenization or IR expansion or contraction. The genome sizes of these species are negatively correlated with nucleotide substitution rates. Species with shorter duration of the life cycle tend to exhibit shorter chloroplast genomes than those with longer life cycles.

Citation: Dong W, Xu C, Cheng T, Zhou S (2013) Complete Chloroplast Genome of *Sedum sarmentosum* and Chloroplast Genome Evolution in Saxifragales. PLoS ONE 8(10): e77965. doi:10.1371/journal.pone.0077965

Editor: Nicolas Salamin, University of Lausanne, Switzerland

Received: April 7, 2013; **Accepted:** September 9, 2013; **Published:** October 18, 2013

Copyright: © 2013 Dong et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was partly supported by the National Natural Science Foundation of China (30930010) and the Ministry of Science and Technology (2012BAC01B05-6 and 2011FY120200-6). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: slzhou@ibcas.ac.cn

Introduction

Chloroplasts are one of the main distinctive characteristics of plant cells. The major function of chloroplasts is to perform photosynthesis [1]. Typically, the size of chloroplast genomes in higher plants ranges from 120 to 160 kb, and a pair of inverted repeats (IRs) divides the genome into a large single copy (LSC) region and a small single copy (SSC) region. Most chloroplast genomes contain 110–130 distinct genes; the majority of these genes (approximately 79) encode proteins, which are mostly involved in photosynthesis, while the remainder of the genes encode transfer RNAs (approximately 30) or ribosomal RNAs (4) [2].

Since the first chloroplast genome from tobacco (*Nicotiana tabacum*) was published [3], more than 200 complete chloroplast genomes from protists, thallophytic, bryophytic, and vascular plants have been made available in GenBank. Although the chloroplast genomes of vascular plants are highly

conserved in their basic structures, comparative genomic studies have revealed occasional structural changes, such as inversions, gene or intron losses, and rearrangements among plant lineages. The most notable examples of gene loss were found in the parasitic plants *Cuscuta* [4,5], *Epifagus* [6], and *Rhizanthella* [7], which have lost some or all of their photosynthetic ability. Loss of chloroplast genes is rare in photosynthetic species but can occur if a gene has been transferred to the nuclear genome or functionally replaced by a nuclear gene [8]. For instance, the *rpl22* gene of Fagaceae and Passifloraceae [9], the *infA* gene of Brassicaceae [10], and the *rpl32* gene of *Populus* [11] have transferred to the nuclear genome. Only 18 genes found in angiosperm chloroplast genomes contain introns, and most of them are quite conserved. However, the introns of the *rpoC1*, *rpl2*, and *atpF* genes have been independently lost from the chloroplast genomes of some angiosperm lineages [10,12–15]. Extensions or contractions of IR regions that cause variations in genome

size, together with gene losses and nucleotide insertions/deletions (indels), are frequently observed within intergenic spacers [16].

The nucleotide substitution rate of chloroplast genes is lower than that of nuclear genes but higher than that of mitochondrial genes [17]. “The overall relative rate of synonymous substitutions of mitochondrial, chloroplast, and nuclear genes in all seed plants is 1:3:10” [18]. However, the rate of chloroplast genome evolution appears to be taxon and gene dependent. For example, the substitution rates observed in the chloroplast genomes of gnetophytes are significantly higher than in other gymnosperms [19,20]; the Poaceae have experienced accelerated chloroplast genome rearrangements and nucleotide substitutions compared to other monocots [14]; and the genes encoding ribosomal proteins, RNA polymerase, and ATPase in Geraniaceae undergo nucleotide substitutions more rapidly than photosynthetic genes [21].

Considering its small size, simple structure and conserved gene content, the chloroplast genome has become an ideal model for evolutionary and comparative genomic researches. Comparative studies of chloroplast genomes have mostly been focused on a target species such as *Panicum virgatum* [22]; genera such as *Oenothera* [23,24]; or families such as Solanaceae [25,26], Poaceae [27,28], Pinaceae [29,30], and Asteraceae [31]. At higher taxonomic levels, information on chloroplast genomes is useful not only for phylogenetic studies [10,32,33] but also for understanding the genome evolution underlying gene and intron losses, genome size variations, and nucleotide substitutions. For this purpose, Saxifragales is an ideal group, in which four completely sequenced and one nearly completely sequenced chloroplast genomes are available, representing five major lineages in the order, i.e., the woody lineage, Haloragaceae + Penthoraceae, Crassulaceae, the Saxifragaceae alliance and the fence-riding Paeoniaceae [34].

As defined in the APG III system (2009), Saxifragales includes 15 families and is divided into six major lineages [35]. The Saxifragales are morphologically diverse, including herbs, shrubs, and large trees [34]. Saxifragales represents one of the early diversified orders of rosids. It was estimated that the order has diverged from eurosids for 89.1–97.6 million years and all major lineages had diverged one another in Cretaceous [36]. It would be interesting to know what has happened to their chloroplast genomes after such a long time of evolution. In Saxifragales the complete chloroplast genome of *Liquidambar formosana* (Hamamelidaceae) and the protein-coding genes of *Heuchera sanguinea* (Saxifragaceae) have been reported [32], and the genomes of *Paeonia obovata* (Paeoniaceae) and *Penthorum chinense* (Penthoraceae) have also been determined (They will be published soon in another paper). If a chloroplast genome of Crassulaceae is available, we will be able to probe into the chloroplast genome evolution in the order. Thus, we chose to determine the chloroplast genome of *Sedum sarmentosum* (Crassulaceae) which is a frequently observed herbal ornamental plant with some medicinal values.

Here, we first report the complete sequence of the chloroplast genome of *S. sarmentosum*. Then, we present the results of comparative analyses of four representative

chloroplast genomes of Saxifragales species, using the genome of the closely related but basal species *Vitis vinifera* as a reference. Special emphases were placed on changes in genome structure, the variations of nucleotide substitution rates and genome sizes, and the associations between them.

Materials and Methods

DNA extraction and sequencing

Genomic DNA was extracted from fresh young leaves of an *S. sarmentosum* plant found in the Beijing Botanical Garden (Institute of Botany, Chinese Academy of Sciences, Beijing, China) using the mCTAB method [37]. The genome was sequenced following Dong et al. [35]. Fifty-five specific primers (Table S1) were used to bridge gaps in the chloroplast genome.

Chloroplast genome annotation

Genome annotation was accomplished using the Dual Organellar Genome Annotator (DOGMA) [38] to annotate the genes encoding proteins, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs). All of the identified tRNA genes were further verified using the corresponding structures predicted by tRNAscan-SE 1.21 [39].

Comparative chloroplast genomic analysis

The mVISTA program was employed in Shuffle-LAGAN mode [40] to compare the complete chloroplast genomes of *S. sarmentosum* and three other species (*Liquidambar formosana*, *Penthorum chinense*, and *Paeonia obovata*). The chloroplast genome of *Vitis vinifera* was used as a reference. To assess the variability of the coding regions of the four chloroplast genomes together with that of *Heuchera sanguinea* (Saxifragaceae) [32], the nucleotide sequences of all protein-coding genes were aligned to those of the reference genome using ClustalX [41] and adjusted manually using Se-AL 2.0 [42].

Estimation of substitution rates

The relative rates of sequence divergence in the five Saxifragales species and the reference were analyzed using the PAML v4.4 package [43]. The program yn00 was employed to estimate dN, dS, and dN/dS under the F3x4 substitution matrix using the Nei–Gojobori method. Genes with the same functions were grouped following previous studies [44–46]. Analyses were carried out on (1) concatenated common protein-coding genes, except for lost genes or pseudogenes from any species; 2) datasets corresponding to the same functions, i.e., for *atp*, *pet*, *ndh*, *psa*, *psb*, *rpl*, *rpo*, and *rps*; and 3) datasets corresponding to singular genes, i.e., for *cemA*, *matK*, *ccsA*, *clpP*, *rbcL*, and *ycf1*. Tajima’s relative rate test implemented in MEGA 5 was used to compare evolutionary rates among the lineages [47,48]. Kruskal-Wallis and Spearman’s rank correlation tests were conducted using the R software package (<http://www.r-project.org>)

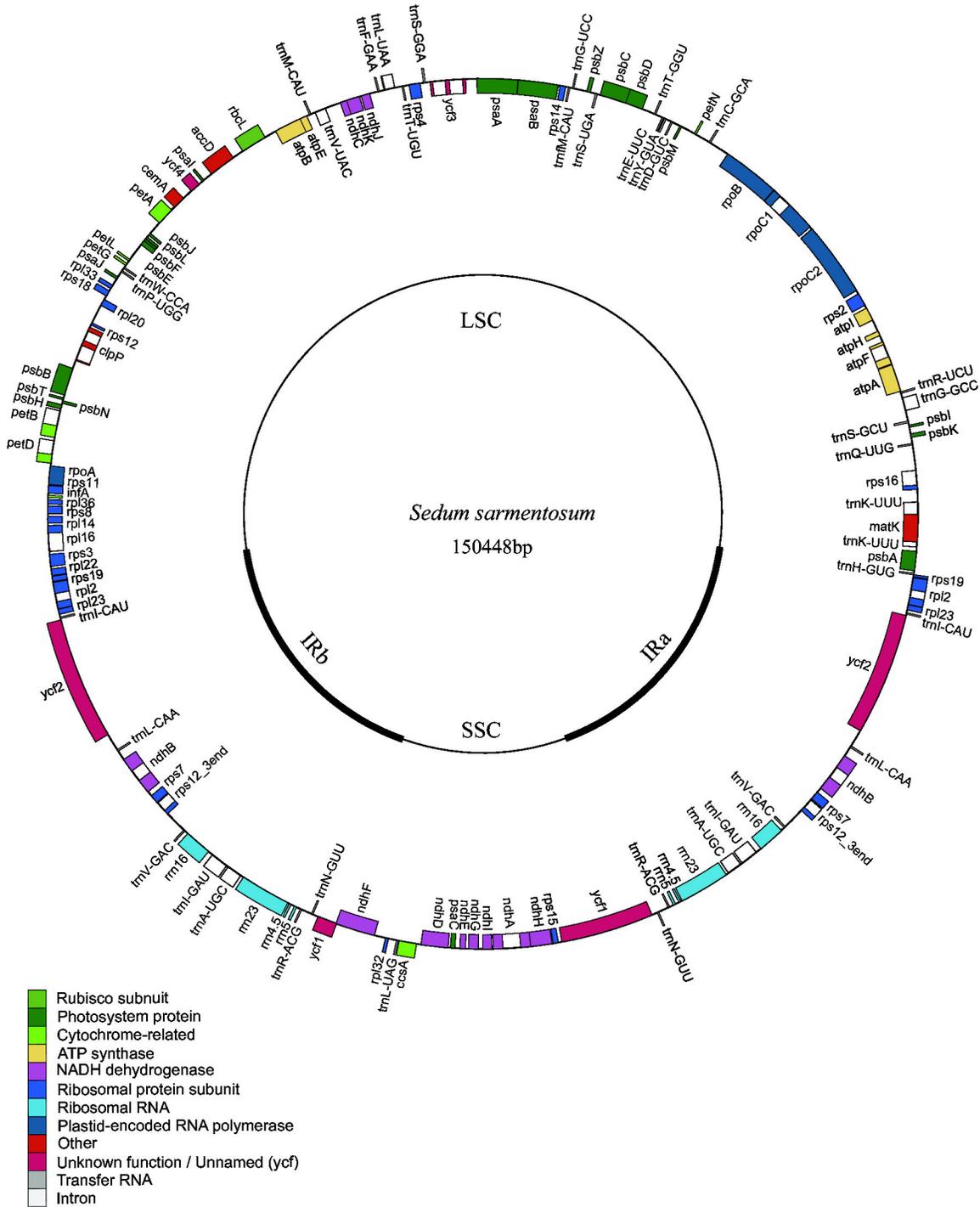


Figure 1. Chloroplast genome map of *Sedum sarmentosum*. The genes inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are shown in different colors. The thick lines indicate the extent of the inverted repeats (IRa and IRb) that separate the genomes into small single-copy (SSC) and large single-copy (LSC) regions.

doi: 10.1371/journal.pone.0077965.g001

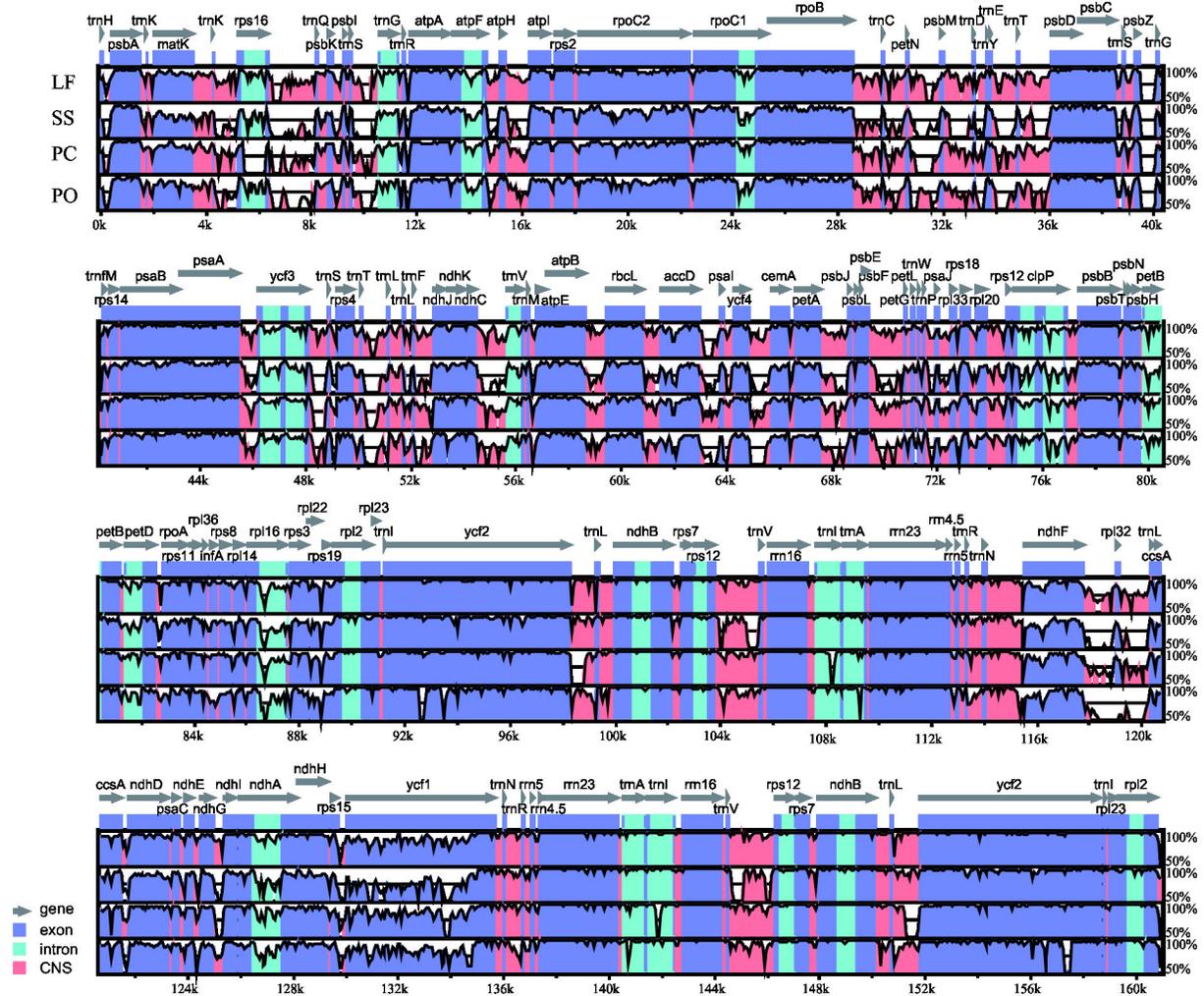


Figure 2. Identity plot comparing the chloroplast genomes of four Saxifragales species using *Vitis vinifera* as a reference sequence. The vertical scale indicates the percentage of identity, ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genome regions are color coded as protein-coding, rRNA, tRNA, intron, and conserved non-coding sequences (CNS). Abbreviations - LF: *Liquidambar formosana*; SS: *Sedum sarmentosum*; PC: *Penthorum chinense*; and PO: *Paeonia obovata*.

doi: 10.1371/journal.pone.0077965.g002

Results and Discussion

Genome content and organization in *Sedum sarmentosum*

The complete chloroplast genome of *S. sarmentosum* (JX427551) is 150,448 bp in size and exhibits a typical circular structure including a pair of IRs (25,783 bp each) that separate the genome into two single-copy regions (LSC 82,212 bp; SSC 16,670 bp; Figure 1). Coding regions (91,260 bp), including protein-coding genes (79,413 bp), tRNA genes (2,801 bp), and rRNA genes (9,046 bp), account for 60.66% of the genome, while noncoding regions (59,188 bp), including introns (17,750 bp) and intergenic spacers (41,438 bp), account for the

remaining 39.34% of the genome. The overall A+T content of the whole genome is 62.24% (Table 1).

There are a total of 113 genes in the genome, including 79 protein-coding genes, 30 tRNA genes, and 4 ribosomal RNA genes (Figure 1 and Table S2). Eighteen genes contain introns (one class I intron, *trnL^{UUA}*, and 17 class II introns), and three of these genes, *clpP*, *rps12*, and *ycf3*, exhibit two introns. The 5'-end exon of the *rps12* gene is located in the LSC region, and the intron and 3'-end exon of the gene are situated in the IR region.

Genome organization of Saxifragales

The organization of the chloroplast genome is rather conserved within Saxifragales (Figure 2). Neither translocations

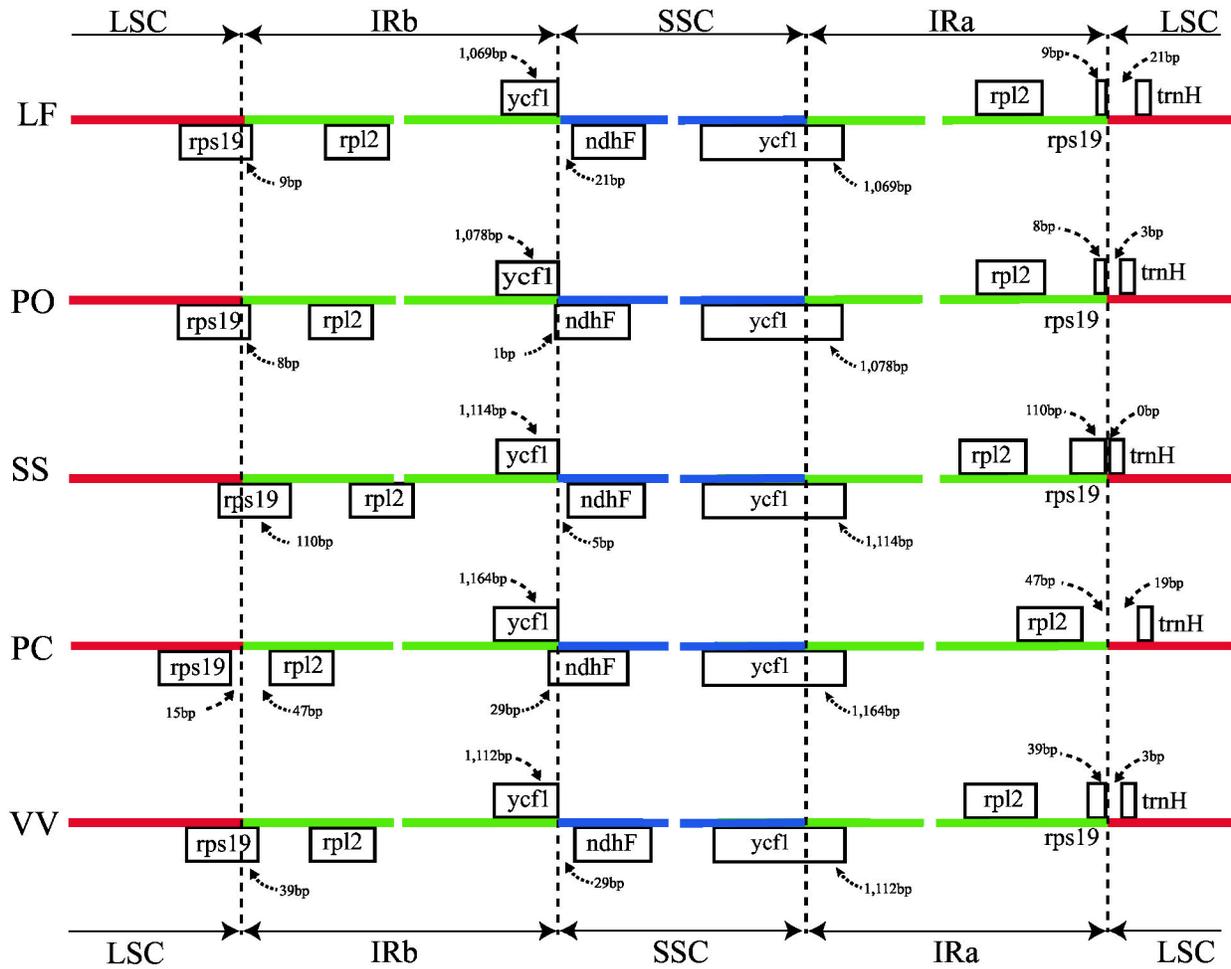


Figure 3. Comparison of junction positions between single copy and IR regions among four Saxifragales genomes and *Vitis vinifera*. Abbreviations - LF: *Liquidambar formosana*; PO: *Paeonia obovata*; SS: *Sedum sarmentosum*; PC: *Penthorum chinense*; and VV: *Vitis vinifera*.

doi: 10.1371/journal.pone.0077965.g003

nor inversions were detected in the four Saxifragales genomes. Similar to other angiosperms, the IR region is more conserved in these species than the LSC and SSC regions. Differences were observed in terms of genome size, gene losses, intron losses, the pseudogenization of protein-coding genes, and IR expansion and contraction.

Genome size. In terms of the chloroplast genome size observed among the examined Saxifragales species, *S. sarmentosum* exhibits the smallest genome. The genome of *L. formosana* (160,410 bp) is approximately 10 kbp larger than that of *S. sarmentosum*, 6.2 kbp larger than that of *Penthorum chinense*, and 2.7 kbp larger than that of *Paeonia obovata*, though it is 0.5 kbp smaller than that of *V. vinifera*, an out-group species. The detected sequence length variation is mainly attributable to the difference in the length of the noncoding region (Table 1). The *S. sarmentosum* genome contains the smallest noncoding region among the four analyzed chloroplast genomes.

Gene loss. Two genes, *infA* and *rpl32*, have been lost from the chloroplast genome of *Paeonia obovata*. The only case of gene loss observed in the sampled Saxifragales species was for *infA* which functions as a translation initiation factor that assists in the assembly of the translation initiation complex [49]. Loss of *infA* appears to have independently occurred multiple times during the evolution of land plants, and this gene is also possibly transferred to the nucleus [50]. Therefore, its loss in *Paeonia obovata* does not represent a unique phenomenon. The *rpl32* gene is one of the nine genes encoding the large ribosomal protein subunit, and loss of this gene would cause functional problems. In *Populus*, the *rpl32* gene has been transferred to the nucleus [11]. In *Paeonia obovata*, whether the *rpl32* gene has been transferred to the nucleus remains to be investigated.

Intron loss. The *rps16* gene in the chloroplast genome of *Penthorum chinense* has lost its only intron, which is a phenomenon that is also observed in *Trachelium* [51] and

Table 1. Characteristics of four Saxifragales and an outgroup (*Vitis vinifera*) chloroplast genomes.

Features	<i>Sedum sarmentosum</i>	<i>Paeonia obovata</i>	<i>Penthorum chinense</i>	<i>Liquidambar formosana</i>	<i>Vitis vinifera</i>
Genome size	150448	152736	156686	160410	160928
Length of LSC	82212	84399	86735	88945	89147
Length of SSC	16670	17031	20399	18917	19065
Length of IR	25783	25653	25776	26274	26358
Coding size	91260	89941	91756	91840	90878
Intron size ¹	17750	17461	16380	17216	17977
Spacer size	41438	45334	48550	51354	52073
AT content (%)	62.24%	61.57%	62.73%	62.05%	62.20%
Total number of genes	131	127	131	131	131
Protein-coding genes	79	75	79	79	79
Duplicated genes	18	18	18	18	18
tRNA genes	30	30	30	30	30
rRNA genes	4	4	4	4	4
Genes with introns	18	18	17	18	18
Pseudogenes	1	3	1	1	1

¹ The *trnK* intron is not considered.

doi: 10.1371/journal.pone.0077965.t001

Table 2. Substitution rates of 72 protein-coding genes in five Saxifragales chloroplast genomes.

Taxa	Nonsynonymous (dN)	Synonymous (dS)	dN/dS
<i>Liquidambar formosana</i>	0.0269±0.0007	0.1234±0.0032	0.2178
<i>Heuchera sanguinea</i>	0.0299±0.0008	0.1480±0.0035	0.2019
<i>Penthorum chinense</i>	0.0314±0.0008	0.1590±0.0037	0.1975
<i>Paeonia obovata</i>	0.0422±0.0009	0.2358±0.0048	0.1788
<i>Sedum sarmentosum</i>	0.0609±0.0011	0.2931±0.0056	0.2077

Vitis vinifera was used as an outgroup. Data are presented as the means ± standard error.

doi: 10.1371/journal.pone.0077965.t002

Pelargonium [52]. However, the genome of *Penthorum chinense* is conventional, whereas those of *Trachelium* and *Pelargonium* have been extensively restructured. The intron loss observed for *rps16* is therefore unusual in normal angiospermous chloroplast genomes.

The *rpl2* intron loss has been reported in Saxifragaceae genera, including *Saxifraga* and *Heuchera* [13]. This phenomenon was confirmed in *Heuchera sanguinea* (GenBank accession number: GQ998409; HQ664603) but rejected in *Heuchera micrantha* (GenBank accession number: EF207446) and *Saxifraga stolonifera* (GenBank accession number: EF207457) based on a re-examination of the sequences. The *rpl2* intron is present in all of the Saxifragales genomes examined in this study, suggesting that intron loss in the *rpl2* gene is occasional.

Gene pseudogenization. The *rpl22* and *rps18* are pseudogenes only in Paeoniaceae, but *ycf15* has pseudogenized in all families of Saxifragales. The *rpl22* sequence of *Paeonia obovata* contains an insertion of a C residue at position 236, which causes a reading frame shift and internal stop codons (Figure S1). The *rps18* gene of *Paeonia obovata* exhibits a mutation at the 58th codon and a deletion of a single nucleotide at the 86th codon (Figure S2). The *rps18* gene encodes a small ribosomal protein and has only been found lost or pseudogenized in the chloroplast genomes of some parasitic plants [49].

The *ycf15* gene, which displays a small open reading frame (ORF), is located immediately downstream of the *ycf2* gene. The *ycf15* gene of tobacco is potentially functional [3], but the validity of *ycf15* as a protein-coding gene is questionable [53–55]. Expression studies in spinach have suggested that *ycf15* may be transcribed, but not spliced [56]. The *ycf15* is certainly a pseudogene in Saxifragales. There is a deletion of ~400 bp in *Penthorum chinense* (Figure S3) and an inversion of 15 bp in *S. sarmentosum*. Although the extra sequences (compared to *Nicotiana*) found in *S. sarmentosum* and the other three species were ignored, the *ycf15* genes of these species exhibit premature stop codons. Because small inversions and microstructural changes mostly occur in introns and intergenic spacers [57], the extra sequence of *ycf15* is more appropriately considered an intron of a pseudogene.

IR expansion and contraction. The expansion and contraction of the border regions between the two IR regions and the single-copy regions have contributed to genome size variations among plant lineages [58]. Therefore, we compared the exact IR border positions and their adjacent genes among the four Saxifragales chloroplast genomes (Figure 3). The portion of *ycf1* located in the IR region varies from 1069 bp to 1164 bp. The *ndhF* gene shares some nucleotides with the *ycf1* pseudogene (1 bp in *Paeonia obovata* and 29 bp in *Penthorum chinense*) or is separated from *ycf1* by spacers (5 bp in *S. sarmentosum*, 26 bp in *L. formosana* and 29 bp in *V. vinifera*).

The IRa/LSC border is generally located upstream of the *trnH^{GUG}* gene. The *trnH^{GUG}* gene is separated from the *rps19* pseudogene or the *rpl2* gene by a spacer except in *S. sarmentosum* which does not contain a spacer (Figure 3). However, in *Penthorum chinense*, the *rps19* gene does not extend into the IR region, and thus, the *rps19* pseudogene is not observed. Although there are expansions or contractions of IR regions observed among the sampled representatives of Saxifragales, they contribute little to the overall size differences in the chloroplast genomes.

Genome evolution

Nonsynonymous (dN) and synonymous (dS) substitutions and their ratio ($\omega = dN/dS$) are indicators of the rates of evolution and natural selection [59]. Relative to the reference species *V. vinifera*, these parameters (Table 2) were compared among the protein-coding chloroplast genes of the five representative species of Saxifragales to investigate genome evolution. The dN and dS of *S. sarmentosum* are 2.27 fold and 2.3 fold larger than those of *L. formosana*, respectively,

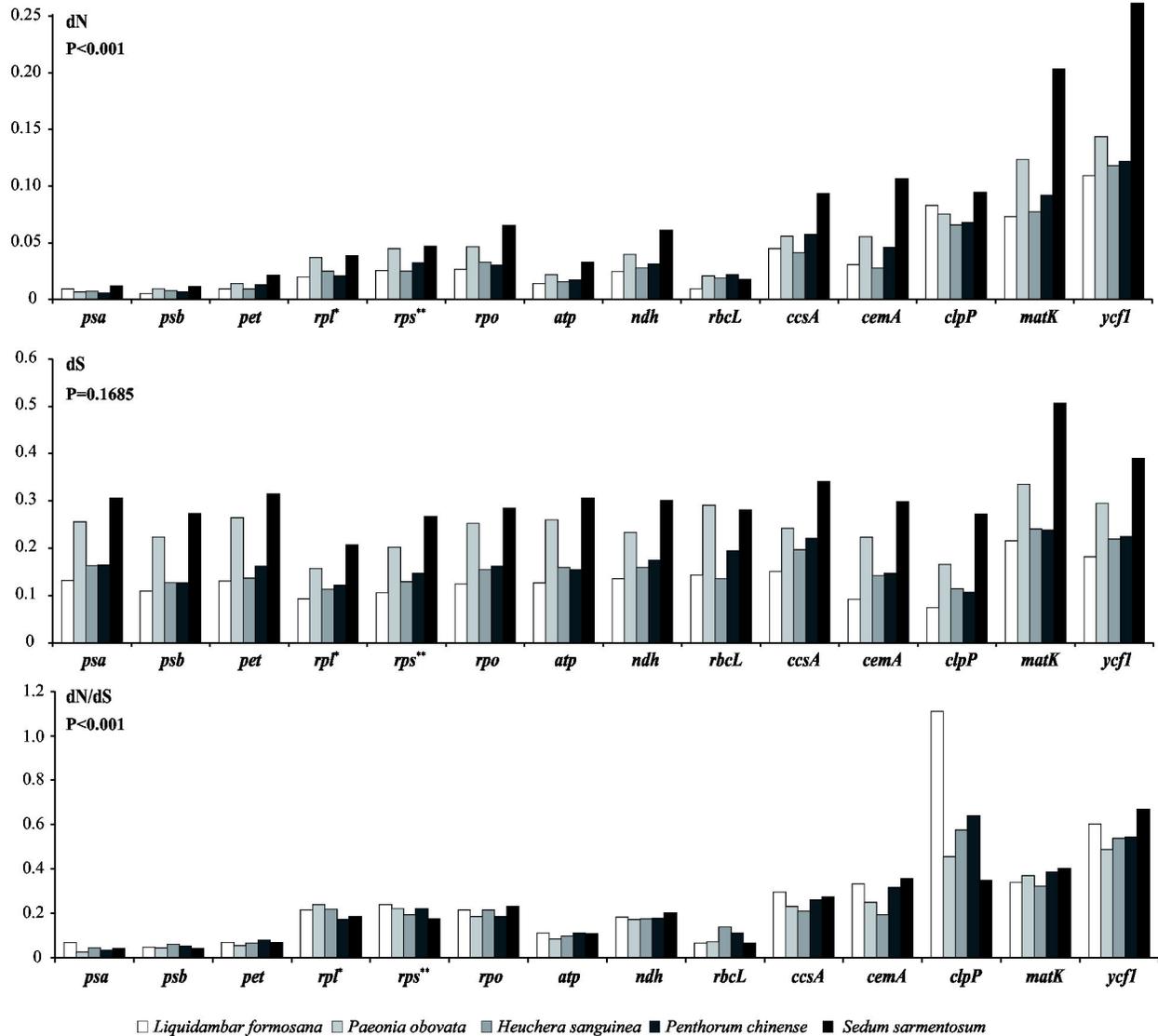


Figure 4. Nonsynonymous substitution (dN), synonymous substitution (dS), and dN/dS values for individual Saxifragales genes and groups of genes. *Without the *rpl22* and *rpl32* genes; **without the *rps18* gene.

doi: 10.1371/journal.pone.0077965.g004

whereas the values of *Paeonia obovata* are accordingly 1.56 fold and 1.88 fold higher than those of *L. formosana*. The ω values of these Saxifragales species are significantly smaller than 1, suggesting the existence of purifying selection on the chloroplast protein-coding genes of Saxifragales species. Both the dN and dS values also consistently indicated that *Paeonia obovata* and *S. sarmentosum* have evolved significantly more rapidly than the other three species (χ^2 -test, all $P < 0.001$).

Association of dN with gene functions. Variations in evolutionary rates can be related to genome structure and the function of genes [10,44]. In Saxifragales species, the observed genome structures are quite similar, without any remarkable restructuring being detected. In comparison with the out-group *V. vinifera*, the dS shows similar values (Kruskal–

Wallis tests, $P = 0.1685$), whereas the dN values differ significantly ($P < 0.001$) among gene groups sorted according to gene functions (Figure 4). The *psa*, *psb*, and *pet* genes exhibit the lowest dN values, while the *ycf1* gene presents the highest dN values. The *matK*, *clpP*, *ccsA*, and *cemA* genes have evolved more rapidly than the other gene groups in Saxifragales. Apart from individual genes, *rpo* exorted the highest mean dN value, followed by *ndh*, *rps*, and *rpl*. Moreover, a strong positive correlation (Spearman's rank correlation, $r_s = 0.94$, $P < 0.001$) was observed between dN and dS among genes of ribosomal gene groups (*rpl*, *rpo* and *rps*), but no significant correlation was found between these two values for any other gene groups.

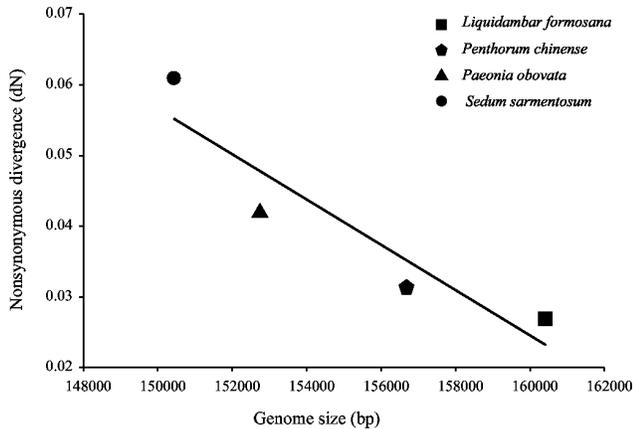


Figure 5. Negative correlation between nonsynonymous substitution (dN) values and chloroplast genome size. The nonsynonymous substitution values are based on an analysis of all protein-coding genes, except for *rp132*, *infA*, *rp122*, and *rps18*.

doi: 10.1371/journal.pone.0077965.g005

Accelerated evolution of the *S. sarmentosum* and *Paeonia obovata* genomes. Tajima's relative rate tests of dN strongly ($P < 0.001$) suggest an acceleration of nucleotide substitution rates in the genomes of *S. sarmentosum* and *Paeonia obovata*. These two genomes are similar to others in terms of their genome structure and number of genes and appear to be under the same purifying selection, but they differ from other species in terms of their genome size, life histories and the systematic positions of the families to which the species belong.

There are obvious differences in genome size among the four species of Saxifragales. These genome size variations are mainly due to length differences in spacers rather than differences in coding genes or introns (Table 1). Interestingly, the sizes of these genomes are negatively correlated with the observed substitution rates (Figure 5). There are strong positive correlations between the numbers of substitutions, repeats and indels in *Cephalotaxus* [60] and Araceae [61], and repeats may play a more important role in sequence divergence [62]. The higher substitution rates found in the genomes of *Paeonia obovata* and *S. sarmentosum* imply that their genomes are more relaxed to changes, giving rise to the loss of some "non-essential" sequences and reducing genome sizes. The reduction of genome size results in saving energy in a cell and promoting the efficiency of replication with lower costs [63-65]. According to this hypothesis, small genomes are more likely to be derived.

Species with short generation times usually evolve more quickly and exhibit higher substitution rates [66-68]. *Paeonia obovata* and *S. sarmentosum* happen to be perennial herbs of short life histories among Saxifragales species, and the species with the lowest substitution rate, *L. formosana*, is a large tree. Therefore, the length of life cycles of these species would have played a role in the evolution of their chloroplast genomes.

Studies on the *rbcl* gene have revealed a positive correlation between molecular evolution and species diversity in angiosperms [67]. In Saxifragales weak positive correlation ($r=0.64$) was observed between the dN values and the species diversity of the five lineages [37]. The lineage represented by *S. sarmentosum* is the most species-rich one which includes ca 1,370 species [69] and has the highest substitution rate. The woody lineage to which *L. formosana* belongs has much few species (ca. 107) and has the lowest dN value. However, there are only 32 species in Paeniaceae [70] but the dN value of *Paeonia obovata* genome is the second largest. Although such a general tendency may hold true, the species diversity may not be a good indicator of genome evolution rate for specific taxa.

Conclusion

The chloroplast genomes of Saxifragales species have undergone evolution at the gene level, rather than the genome level, as no significant structural changes are observed among their genomes. However, the examined genomes differ in size, and the detected genome size variations are predominantly due to the length of intergenic spacers, rather than losses of genes and introns, gene pseudogenization or IR expansion or contraction. The genome sizes of these species appear to be negatively correlated with their nucleotide substitution rates. Species with short life histories tend to exhibit smaller genome sizes and higher nucleotide substitution rates. As every species displays its own unique evolutionary history, it is difficult to draw a conclusion without exceptions. It is clear that genome evolution is taxon dependent.

Supporting Information

Table S1. *Sedum sarmentosum*-specific primers used to sequence the complete chloroplast genome. (XLS)

Table S2. A list of genes found in the *Sedum sarmentosum* chloroplast genome. Intron-containing genes are marked by asterisks (*). (XLS)

Figure S1. Alignment of the *rp122* region in five Saxifragales species. Codons highlighted in red represent stop codons and codons highlighted in green represent unformed triplet codons. The numbers indicate the positions of nucleotides. (PDF)

Figure S2. Alignment of the *rps18* region in five Saxifragales species. Codons highlighted in red represent stop codons and codons highlighted in green represent unformed triplet codons. The numbers indicate the positions of codons. (PDF)

Figure S3. Alignment of the *ycf15* region in five Saxifragales species, *Nicotiana*, and *Vitis*. The uninterrupted form of *Nicotiana* was used as a reference. Codons highlighted in red represent stop codons and codons highlighted in green represent unformed triplet codons. The nucleotides in red and blue indicate an inversion of the sequence. (PDF)

Acknowledgements

We are grateful to Shu-Miaw Chaw for her revision of the manuscript and suggestions, and to Jianhua Xue and Jing Yu for their help in the genome sequencing.

References

- Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Physiol Plant Mol Biol* 51: 111-140. doi: 10.1146/annurev.arplant.51.1.111. PubMed: 15012188.
- Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW et al. (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*. Academic Press. pp. 348-384.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2043-2049. PubMed: 16453699.
- Funk HT, Berg S, Krupinska K, Maier UG, Krause K (2007) Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol* 7: 45. doi:10.1186/1471-2229-7-45. PubMed: 17714582.
- McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW (2007) Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol* 7: 57. doi:10.1186/1471-2229-7-57. PubMed: 17956636.
- Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A* 89: 10648-10652. doi:10.1073/pnas.89.22.10648. PubMed: 1332054.
- Delannoy E, Fujii S, des Francs-Small Colas C, Brundrett M, Small I (2011) Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28: 2077-2086. doi:10.1093/molbev/msr028. PubMed: 21289370.
- Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M et al. (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res* 20: 1700-1710. doi:10.1101/gr.111955.110. PubMed: 20978141.
- Jansen RK, Saski C, Lee SB, 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. PubMed: 20935065.
- Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW 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 U S A* 104: 19369-19374. doi:10.1073/pnas.0709121104. PubMed: 18048330.
- Ueda M, Fujimoto M, Arimura S, Murata J, Tsutsumi N et al. (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. doi:10.1016/j.gene.2007.07.019. PubMed: 17728076.
- Downie SR, Llanas E, KatzDownie DS (1996) Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNA's. *Syst Bot* 21: 135-151. doi:10.2307/2419744.
- Downie SR, Olmstead RG, Zurawski G, Soltis DE, Soltis PS et al. (1991) Six Independent Losses of the Chloroplast DNA *rpl2* Intron in Dicotyledons: Molecular and Phylogenetic Implications. *Evolution* 45: 1245-1259. doi:10.2307/2409731.
- Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK (2010) Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J Mol Evol* 70: 149-166. doi:10.1007/s00239-009-9317-3. PubMed: 20091301.
- Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saski C et al. (2008) 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* 116: 723-737. doi:10.1007/s00122-007-0706-y. PubMed: 18214421.
- Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. *Plant Syst Evol* 271: 101-122. doi:10.1007/s00606-007-0608-0.
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A* 84: 9054-9058. doi:10.1073/pnas.84.24.9054. PubMed: 3480529.
- Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol* 49: 827-831. doi:10.1016/j.ympev.2008.09.009. PubMed: 18838124.
- McCoy SR, Kuehl JV, Boore JL, Raubeson LA (2008) The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol Biol* 8: 130. doi:10.1186/1471-2148-8-130. PubMed: 18452621.
- Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol* 52: 115-124. doi:10.1016/j.ympev.2008.12.026. PubMed: 19166950.
- Guisinger MM, Kuehl JNV, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc Natl Acad Sci U S A* 105: 18424-18429. doi:10.1073/pnas.0806759105. PubMed: 19011103.
- Young HA, Lanzatella CL, Sarath G, Tobias CM (2011) Chloroplast genome variation in upland and lowland switchgrass. *PLOS ONE* 6: e23980. doi:10.1371/journal.pone.0023980. PubMed: 21887356.
- Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K et al. (2008) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. *Mol Biol Evol* 25: 2019-2030. doi:10.1093/molbev/msn149. PubMed: 18614526.
- Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K et al. (2008) The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic Acids Res* 36: 2366-2378. doi:10.1093/nar/gkn081. PubMed: 18299283.
- Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T et al. (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor Appl Genet* 112: 1503-1518. doi:10.1007/s00122-006-0254-x. PubMed: 16575560.
- Jo YD, Park J, Kim J, Song W, Hur CG et al. (2011) 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* 30: 217-229. doi:10.1007/s00299-010-0929-2. PubMed: 20978766.
- Diekmann K, Hodkinson TR, Wolfe KH, van den Bekerom R, Dix PJ et al. (2009) Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* p. L.). *DNA Research* 16: 165-176

Author Contributions

Conceived and designed the experiments: WD SZ. Performed the experiments: WD CX. Analyzed the data: WD TC. Wrote the manuscript: WD SZ.

28. Cahoon AB, Sharpe RM, Mysayphonh C, Thompson EJ, Ward AD et al. (2010) The complete chloroplast genome of tall fescue (*Lolium arundinaceum*; Poaceae) and comparison of whole plastomes from the family Poaceae. *Am J Bot* 97: 49-58. doi:10.3732/ajb.0900008. PubMed: 21622366.
29. Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM (2011) Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biol Evolution* 3: 309-319. doi:10.1093/gbe/evr026. PubMed: 21402866.
30. Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol Evolution* 2: 504-517. doi:10.1093/gbe/evq036. PubMed: 20651328.
31. Doorduyn L, Gravendeel B, Lammers Y, Ariyurek Y, Chin AWT et al. (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res* 18: 93-105. doi:10.1093/dnares/dsr002. PubMed: 21444340.
32. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A* 107: 4623-4628. doi:10.1073/pnas.0907801107. PubMed: 20176954.
33. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci U S A* 104: 19363-19368. doi:10.1073/pnas.0708072104. PubMed: 18048334.
34. Jian SG, Soltis PS, Gitzendanner MA, Moore MJ, Li R et al. (2008) Resolving an ancient, rapid radiation in Saxifragales. *Syst Biol* 57: 38-57. doi:10.1080/10635150801888871. PubMed: 18275001.
35. Dong W, Xu C, Cheng T, Lin K, Zhou S (2013) Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol Evolution* 5: 989-997. doi:10.1093/gbe/evt063. PubMed: 23595020.
36. Xue JH, Dong WP, Cheng T, Zhou SL (2012) Nelumbonaceae: Systematic position and species diversification revealed by the complete chloroplast genome. *J Syst Evolution* 50: 477-487. doi:10.1111/j.1759-6831.2012.00224.x.
37. Li J, Wang S, Jing Y, Wang L, Zhou S (2013) A modified CTAB protocol for plant DNA extraction. *Chin Bulletin Bot* 48: 72-78.
38. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255. doi:10.1093/bioinformatics/bth352. PubMed: 15180927.
39. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33: W686-W689. doi:10.1093/nar/gki366. PubMed: 15980563.
40. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273-W279. doi:10.1093/nar/gkh053. PubMed: 15215394.
41. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882. doi:10.1093/nar/25.24.4876. PubMed: 9396791.
42. Rambaut A (1996) Se-AI: sequence alignment editor. version 2.0. Oxford: University of Oxford, Department of Zoology.
43. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.
44. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH et al. (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol* 23: 279-291. PubMed: 16207935.
45. Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K (2002) Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Mol Biol Evol* 19: 2084-2091. doi:10.1093/oxfordjournals.molbev.a004033. PubMed: 12446800.
46. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol Evolution* 4: 294-306. doi:10.1093/gbe/evs006. PubMed: 22247429.
47. Tajima F (1993) Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis. *Genetics* 135: 599-607. PubMed: 8244016.
48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
49. 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: 273-297. doi:10.1007/s11103-011-9762-4. PubMed: 21424877.
50. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT 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. doi:10.1105/tpc.13.3.645. PubMed: 11251102.
51. Haberle RC, Fourcade HM, Boore JL, Jansen RK (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. PubMed: 18330485.
52. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ et al. (2006) 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* 23: 2175-2190. doi:10.1093/molbev/msl089. PubMed: 16916942.
53. Goremykin V, Hirsch-Ernst K, W Iff S, Hellwig F (2003) The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Plant Syst Evol* 242: 119-135. doi:10.1007/s00606-003-0056-4.
54. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM et al. (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174. doi:10.1186/1471-2164-8-174. PubMed: 17573971.
55. Tangphatsornruang S, Uthapaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T et al. (2011) Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475: 104-112. doi:10.1016/j.gene.2011.01.002. PubMed: 21241787.
56. Schmitz-Linneweber C, Maier RM, Alcaraz JP, Cottet A, Herrmann RG et al. (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Mol Biol* 45: 307-315. doi:10.1023/A:1006478403810. PubMed: 11292076.
57. Kim KJ, Lee HL (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol Cells* 19: 104-113. PubMed: 15750347.
58. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM et al. (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 8: 36. doi:10.1186/1471-2148-8-36. PubMed: 18237435.
59. Yang ZH, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17: 32-43. doi:10.1093/oxfordjournals.molbev.a026236. PubMed: 10666704.
60. Yi X, Gao L, Wang B, Su YJ, Wang T (2013) The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol Evolution* 5: 688-698. doi:10.1093/gbe/evt042. PubMed: 23538991.
61. Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD et al. (2012) Mutational dynamics of aroid chloroplast genomes. *Genome Biol Evolution* 4: 1316-1323. doi:10.1093/gbe/evs110. PubMed: 23204304.
62. McDonald MJ, Wang WC, Huang HD, Leu JY (2011) Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLOS Biol* 9: e1000622. PubMed: 21697975.
63. Comeron JM (2001) What controls the length of noncoding DNA? *Curr Opin Genet Dev* 11: 652-659. doi:10.1016/S0959-437X(00)00249-5. PubMed: 11682309.
64. Andersson SGE, Kurland CG (1998) Reductive evolution of resident genomes. *Trends Microbiol* 6: 263-268. doi:10.1016/S0966-842X(98)01312-2. PubMed: 9717214.
65. Selsosse MA, Albert BR, Godelle B (2001) Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evol* 16: 135-141. doi:10.1016/S0169-5347(00)02084-X. PubMed: 11179577.
66. Bousquet J, Strauss SH, Doerksen AH, Price RA (1992) Extensive variation in evolutionary rate of *rbcl* gene-sequences among seed plants. *Proc Natl Acad Sci U S A* 89: 7844-7848. doi:10.1073/pnas.89.16.7844. PubMed: 1502205.
67. Barraclough TG, Harvey PH, Nee S (1996) Rate of *rbcl* gene sequence evolution and species diversification in flowering plants (angiosperms). *Proc R Soc Lond B-Biol Sci* 263: 589-591. doi:10.1098/rspb.1996.0088.

68. Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science* 322: 86-89. doi:10.1126/science.1163197. PubMed: 18832643.
69. Mort ME, Soltis DE, Soltis PS, Francisco-Ortega J, Santos-Guerra A (2002) Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Syst Bot* 27: 271-288.
70. Hong DY (2010) *Peonies of the world: Taxonomy and phytogeography*. UK: Royal Botanic Gardens.