

RESEARCH ARTICLE

The Chloroplast Genome of *Elaeagnus macrophylla* and *trnH* Duplication Event in Elaeagnaceae

Kyoung Su Choi, Ogyeong Son, SeonJoo Park*

Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongsangbuk-do, 712–749, Korea

* sjpark01@ynu.ac.kr



OPEN ACCESS

Citation: Choi KS, Son O, Park S (2015) The Chloroplast Genome of *Elaeagnus macrophylla* and *trnH* Duplication Event in Elaeagnaceae. PLoS ONE 10(9): e0138727. doi:10.1371/journal.pone.0138727

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

Received: June 24, 2015

Accepted: September 2, 2015

Published: September 22, 2015

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

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by the Ministry of Environment in Korea (Daegu Regional Environmental Office, 2014).

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

Abstract

Elaeagnaceae, which harbor nitrogen-fixing actinomycetes, is a plant family of the Rosales and sister to Rhamnaceae, Barbeyaceae and Dirachmaceae. The results of previous molecular studies have not strongly supported the families of Elaeagnaceae, Rhamnaceae, Barbeyaceae and Dirachmaceae. However, chloroplast genome studies provide valuable phylogenetic information; therefore, we determined the chloroplast genome of *Elaeagnus macrophylla* and compared it to that of Rosales such as IR junction and *infA* gene. The chloroplast genome of *Elaeagnus macrophylla* is 152,224 bp in length and the *infA* gene of *E. macrophylla* was pseudogenized. Phylogenetic analyses based on 79 genes in 30 species revealed that *Elaeagnus* was closely related to *Morus*. Comparison of the IR junction in six other rosids revealed that the *trnH* gene contained the LSC region, whereas *E. macrophylla* contained a *trnH* gene duplication in the IR region. Comparison of the LSC/IRb (J_{LB}) and the IRa/LSC (J_{LA}) regions of Elaeagnaceae (*Elaeagnus* and *Shepherdia*) and Rhamnaceae (*Rhamnus*) showed that *trnH* gene duplication only occurred in the Elaeagnaceae. The complete chloroplast genome of *Elaeagnus macrophylla* provides unique characteristics in rosids. The *infA* gene has been lost or transferred to the nucleus in rosids, while *E. macrophylla* lost the *infA* gene. Evaluation of the chloroplast genome of *Elaeagnus* revealed *trnH* gene duplication for the first time in rosids. The availability of *Elaeagnus* cp genomes provides valuable information describing the relationship of Elaeagnaceae, Barbeyaceae and Dirachmaceae, IR junction that will be valuable to future systematics studies.

Introduction

Rosales consists of approximately 7,700 species distributed into about 260 genera and nine families, Rosaceae, Ulmaceae, Cannabaceae, Moraceae, Urticaceae, Rhamnaceae, Barbeyaceae, Dirachmaceae and Elaeagnaceae. Elaeagnaceae has been placed near Barbeyaceae, Dirachmaceae and Rhamnaceae [1]. Molecular analyses of Rosales has shown that relationships among Ulmaceae, Cannabaceae, Moraceae and Urticaceae were strongly supported [2]. However, phylogenetic relationships among Barbeyaceae, Dirachmaceae, Rhamnaceae and Elaeagnaceae were weakly supported and not certain [2–4].

Elaeagnus L. belong to Elaeagnaceae, a small family that also contains *Hippophae* L and *Shepherdia* Nutt [5,6]. *Elaeagnus* consists of approximately 60 species distributed in Asia, Australia, southern Europe and North America [7]. *Elaeagnus macrophylla* is a popular ornamental plant valued for its aesthetic qualities and sweetly scented flowers. *E. macrophylla* is native to Eastern Asia.

The plant chloroplast (cp) genome consists of large inverted repeats (IRa, IRb) separated by a large single-copy (LSC) region and a small single-copy (SSC) region [8,9]. Approximately 100–120 genes are located in the cp genome, which is highly conserved [10]. However, some species in Asteraceae have been shown to have inversions [11], rearrangements have been observed in *Pelargonium* [12], and gene loss and IR variations have been found in early-divergent eudicots [13, 14].

Recent studies of the IR region have enabled its use as an important marker describing relationships among plants. The IR region of most angiosperms ranges from 20 kb to 28 kb [12]. Previous analyses have shown various expansions or contractions of IR in some plants, such as 25 kb in *Cycas* [15], 114 bp in *Cryptomeria* [16] and 76 kb in *Pelargonium* [12]. Plunkett and Downie [17] reported IR expansion/contraction in Apioideae. IRa and IRb contain four junctions, J_{LA} (LSC/IRa border), J_{SA} (SSC/IRa border), J_{SB} (LSC/IRb border) and J_{LB} (SSC/IRb border). Most angiosperm plant IRb and IRa contain *rps19-rpl2* and *rps19-trnH*, respectively [18], while most monocot IRb and IRa contain *rps19-trnH* and *trnH-rps19-psbA*, respectively [19].

Previous studies have analyzed the complete chloroplast genome sequences of rosids and identified unique features such as inversion and gene transfer in their plastids [20]. Fagaceae of rosids showed changes in gene order in response to 51 kb inversions in *Glycine* and loss of the IR region in *Medicago* [20]. Genes of some rosid plastomes have been transferred to the nucleus [21,22]. For example, the *infA* gene (*Gossypium* [23], *Arabidopsis* [24], *Oenothera* [25] and *Lotus* [26]) and *rpl22* gene (*Castanea*, *Quercus* and *Passiflora* [22]) were transferred to the nucleus.

Here, we report the complete sequence of the chloroplast genome of *E. macrophylla* in Elaeagnaceae for the first time. In this review, we provide comparative analyses of the chloroplast genome of rosids species such as the *infA* gene, *rpl22* gene and IR junction. Specifically, we describe the structure of the chloroplast genome, IR junction characteristics and gene contents, which will better resolve phylogenetic relationships among rosids and Rosales.

Materials and Methods

Ethics, plant samples, sequencing, mapping and analysis

This research was approved by the Ministry of Environment in Korea (Daegu Regional Environmental Office). *Elaeagnus macrophylla* is not endangered or protected species. *Elaeagnus macrophylla* leaves were obtained from Dokdo Island (Korean Government), Korea. Total DNA was extracted using a DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) and quantified with a HiGen™ Gel & PCR Purification System (Biofact Inc., Daejeon, Korea). Genomic DNA was sequenced using an Illumina Miseq sequencer (Illumina Inc., San Diego, CA). A total of 4,284,888 pair-end sequence reads of 300 bp were generated from the sequencing library with a median insert size of 500 bp, after which genome coverage was estimated using the CLC Genomics Workbench, v. 7.0.4 (CLC Bio, Aarhus, Denmark).

The complete chloroplast genome sequence was annotated using a Dual Organellar Genome Annotator [DOGMA] [27]. All of the identified tRNA genes were further verified using the corresponding structures predicted by tRNAscan-SE [28]. A circle cp genome map was drawn using the OGDRAW program [29]. Geneious v.6.1.7 [30] was employed to compare the cp genome of *E. macrophylla*, *Morus indica* and *Prunus kansuensis*.

Phylogenetic analyses

A total of 79 genes sequences from 30 species (S1 Table) were aligned using MAFFT [31]. Phylogenetic analysis was conducted based on the maximum likelihood (ML) using the GTR+R+I model in RAxML v. 7.2.6 [32] and 1,000 bootstrap replicates. ML analysis resulted in a single tree with $-\ln L = 485,367.343$.

PCR amplification and comparative analysis of IR junctions

Six species of Elaeagnaceae (*Elaeagnus* and *Shepherdia*) and six species of Rhamnaceae (*Rhamnus*) were evaluated (S2 Table) using the following primers specific for the LSC/IRb junction and IRa/LSC junction designed with Primer3 [33]: 1) *rps19-rpl2*: Forward, CGCTCGGGACC AAGTTACTA; Reverse, GGGTTATCCTGCACTTGGAA 2) *rpl2-psbA*: Forward, ATGTTGG GGTGAACCAGAAA; Reverse, GCTGCTTGGCCTGTAGTAGG. Total DNA was extracted as described by Allan et al. [34] and then subjected to PCR amplification. PCR cocktail (25 μ l) consisted of 250ng genomic DNA, 1X DiastarTM Taq DNA buffer, 0.2 mM of each dNTP, 10 pM of each primer and 0.025 U of DiastarTM Taq DNA polymerase (SolGent Co., Korea). The amplification conditions were as follows: initial denaturation at 95°C for 2 min, followed by 35 cycle of 95°C for 20sec, 56°C for 40sec, and 72°C for 50sec, with a final extension at 72°C for 5min, after which samples were held at 8°C. Amplification products were purified using a HiGenTM Gel & PCR Purification System (Biofact Inc., Daejeon, Korea). Nucleotide sequences of the *rps19-rpl2* region and *rps19-psbA* regions were aligned with Geneious v. 6.1.7 [30].

Results

Comparison of the chloroplast genome of *Elaeagnus macrophylla* to those of other rosids

The cp genome sequence of *E. macrophylla* was submitted to GenBank and assigned accession number KP211788. The cp genome contains 152,224 bp, the LSC has 82,136 bp, the SSC has 18,278 bp and the IR has 25,905 bp (Fig 1). We identified 113 unique genes in *E. macrophylla*, 80 protein coding genes, 29 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes. The genome consisted of 73.4% coding genes (111,792 bp), including 60.5% protein-coding genes (92,119 bp), 7% tRNA genes (10,625 bp) and 5.9% rRNA genes (9,048 bp). Additionally, 18 genes encoded introns among unique genes of *E. macrophylla*, among which 12 are protein-coding genes and six are tRNA genes. Three protein coding genes include two introns (*clpP*, *ycf3* and *rps12*), and the overall A+T content of *E. macrophylla* is 63.9%. The A+T percentages are higher in the SSC region (69.4%) than the LSC (65%) and IR regions (57.3%).

Previous studies, *E. macrophylla* belong to Rosales [2] and complete chloroplast genome of Rosales studied in *Morus indica* (NC_008359) and *Prunus kansuensis* [22]. Therefore, the genome features of *E. macrophylla* were compared to *M. indica* and *P. kansuensis* (Table 1). The total size of the chloroplast genome was longer in *P. kansuensis* (157,790 bp) than *E. macrophylla* (152,224 bp) and *M. indica* (158,484 bp). The length of the LSC region (82,136 bp to 87,386 bp) differed significantly from the SSC (18,278bp to 19,745 bp) and IR regions (26,387 bp to 26,678 bp). The average AT content of the Rosales cp genome is 63%, with the highest being observed in *M. indica* (63.63%).

Genes of *infA* and *rpl22* in rosids

The functional gene sequences of *infA* and *rpl22* are highly variable in rosids. The *infA* gene of rosids differs from that of most asterids (*Helianthus*, *Guizotia*, *Lactuca* and *Jacobaea*), monocots (*Dioscorea*), magnoliids (*Drimys*) and chloranthales (*Chloranthus*) owing to the presence

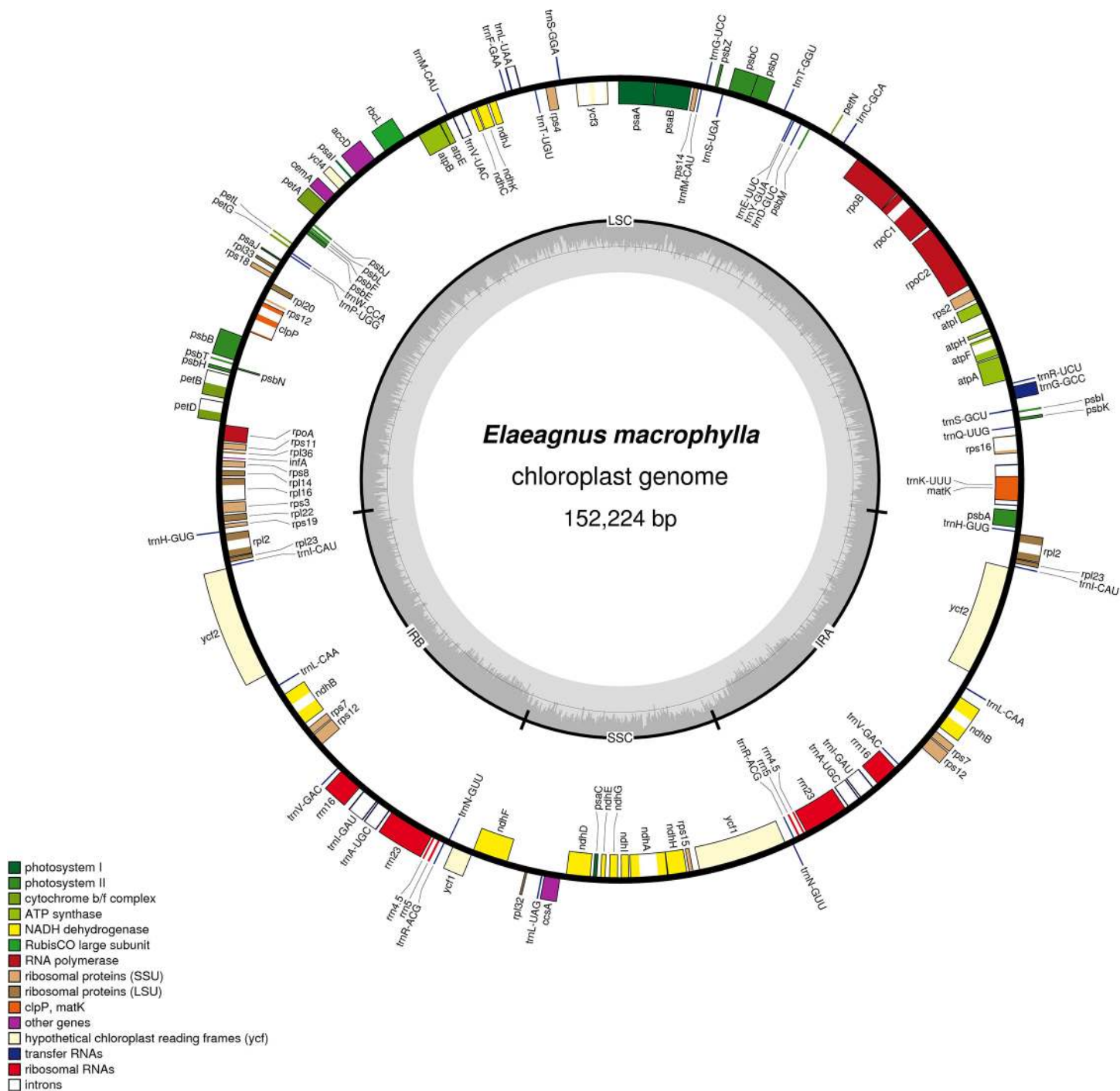


Fig 1. Complete chloroplast genome of *Elaeagnus macrophylla*. Genes drawn inside the circle are transcribed clockwise, while outside genes are counterclockwise. The gray plot in the inner circle corresponds to the GC content.

doi:10.1371/journal.pone.0138727.g001

of the pseudogene, *infA*, in *Elaeagnus* and other rosids. However, other plants such as *Dioscorea*, *Helianthus*, *Guizotia*, *Lactuca*, *Jacobaea*, *Chloranthus* and *Drimys* encode homologous sequences (Fig 2).

Table 1. General features of *Elaeagnus macrophylla* and comparison to those of Rosales.

Genome features	<i>E. macrophylla</i>	<i>Morus indica</i>	<i>Prunus kansuensis</i>
Total length (bp)	152,224 bp	158,484 bp	157,790 bp
LSC length	82,136 bp	87,386 bp	85,811 bp
SSC length	18,278 bp	19,742 bp	19,151 bp
IR length	25,905 bp	25,678 bp	26,387 bp
A+T content	63.9%	63.63%	63.2%
Number of genes	129	129	129
Number of gene duplicated in IR ^a	17	16	16

Prunus kansuensis has re-annotation in this study.

^a *rps12* is not included in this number; only genes completely duplicated are included.

doi:10.1371/journal.pone.0138727.t001

The *rpl22* gene of *Arabidopsis*, *Glycine* and *Lotus* showed an internal stop codon. However, the *rpl22* gene of the *rpl22* gene of other plants consists of the start codon (methionine) to stop codon (data not shown). Nevertheless, the size of the *rpl22* gene in another 18 species ranged from 252 bp in *Cucumis* to 552 bp in *Guizotia*, while it was 423 bp in *Elaeagnus*.

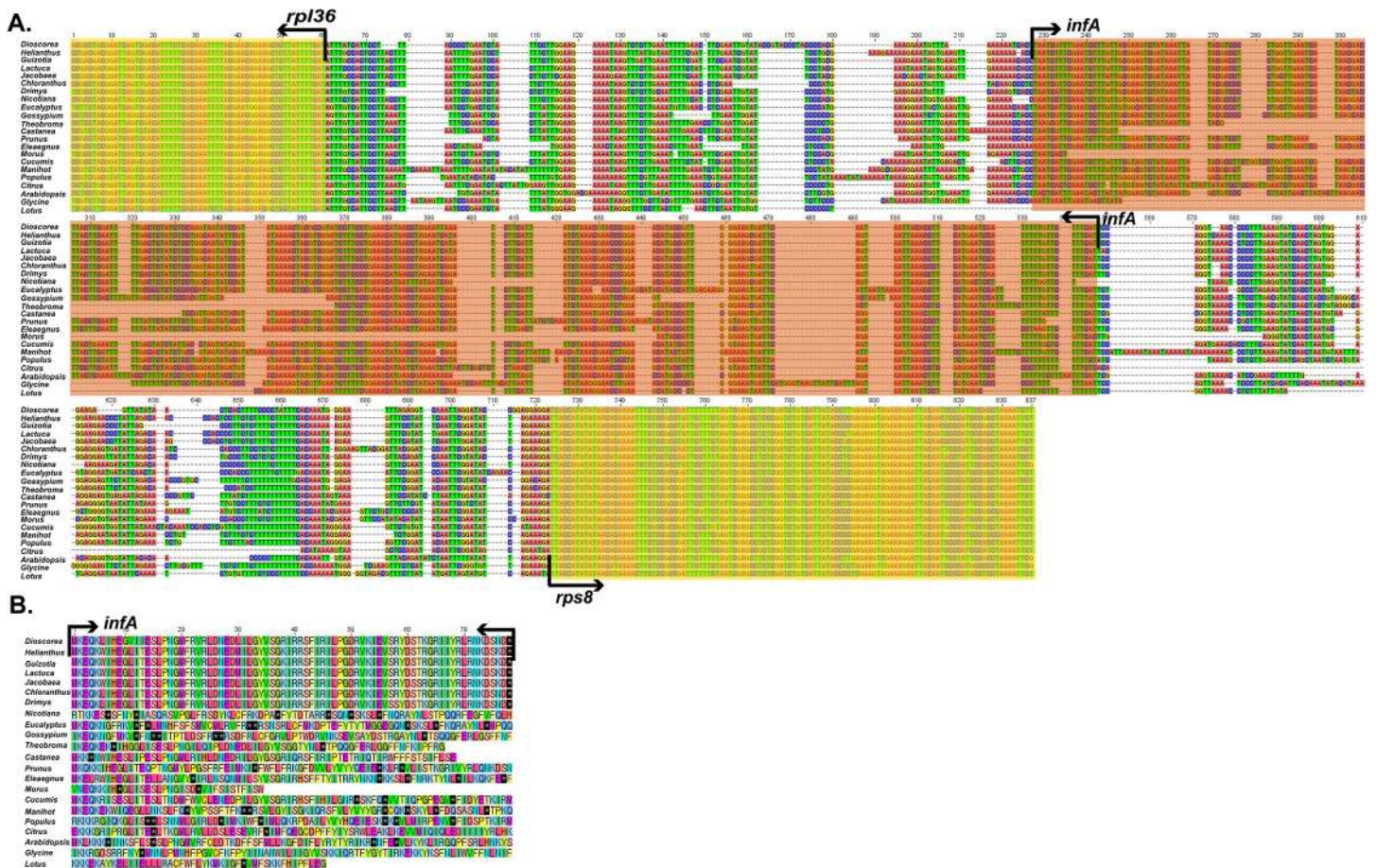


Fig 2. *rpl36-rps8* region sequence alignment of 22 species in angiosperms. A. The *rpl36-rps8* region aligned nucleotide sequence. B. Amino acid sequence alignment of *infA* (red box in Fig 2A). *Dioscorea* in Monocots; *Helianthus*, *Guizotia*, *Lactuca*, *Jacobaea*, and *Nicotiana* in Astrids; *Chloranthus* in Chloranthales; *Drymopsis* in Magnoliids; *Eucalyptus*, *Gossypium*, *Theobroma*, *Castanea*, *Prunus*, *Morus*, *Cucumis*, *Manihot*, *Populus*, *Citrus*, *Arabidopsis*, *Glycine*, and *Lotus* in Rosids.

doi:10.1371/journal.pone.0138727.g002

Comparison of IR region in Rosids

We compared the IR region of seven species (*Elaeagnus*, *Morus*, *Prunus*, *Oenothera*, *Manihot*, *Castanea* and *Theobroma*) of rosids (Fig 3). In *Prunus*, *Oenothera*, *Manihot* and *Theobroma*,

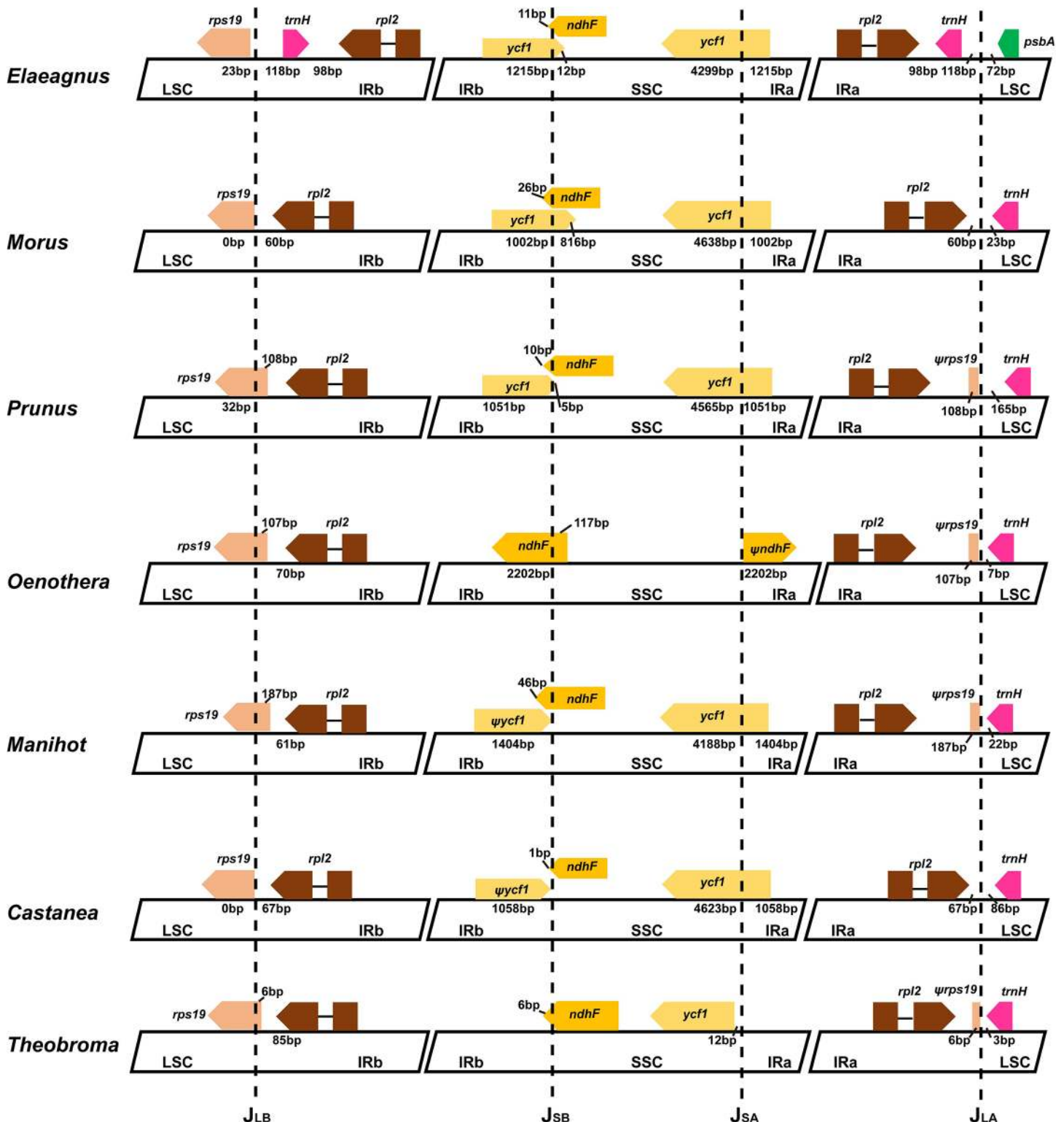


Fig 3. Comparison of four junctions (LSC/IRb, IRb/SSC, SSC/IRa and IRa/LSC) among eight rosid genomes.

doi:10.1371/journal.pone.0138727.g003

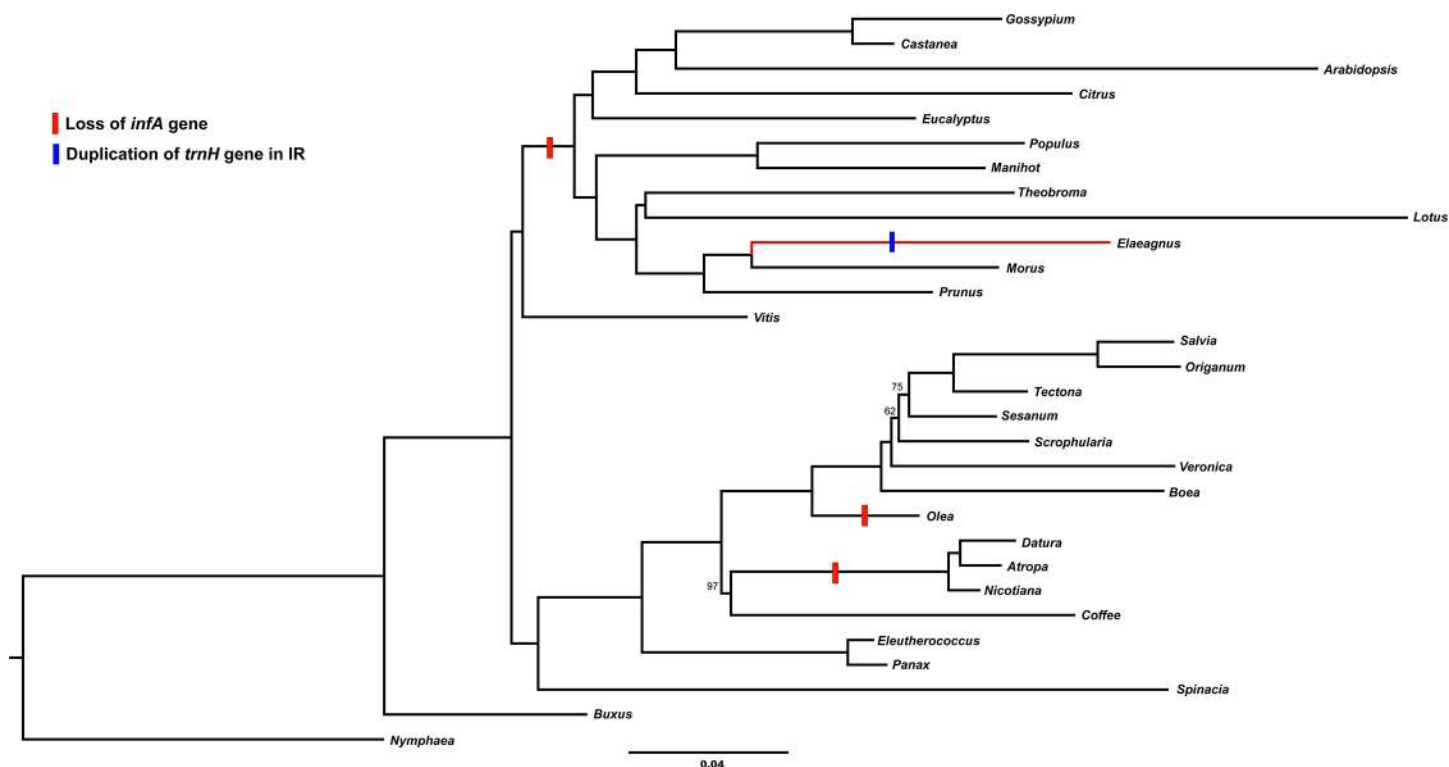


Fig 4. Maximum likelihood tree based on 79 coding genes. The maximum likelihood phylogram has an ML value of $-\ln L = 48,5367.343$. The red box indicated loss of the *infA* gene and the blue box indicated duplication of the *trnH* in the IR region.

doi:10.1371/journal.pone.0138727.g004

J_{LB} occur within the *rps19* gene, resulting in partial duplication of this gene in IRa at J_{LA} (108 bp, 107 bp, 187 bp and 96 bp, respectively). Nevertheless, *rps19* is not duplicated in *Morus* and *Castanea*. In contrast, differing gene arrangement such as complete duplication of *trnH* was observed in the LSC/IRb and LSC/IRa border regions of *Elaeagnus*.

The *ycf1* gene is duplicated in the IRb/SSC (J_{SB}) and SSC/IRa (J_{SA}) borders of rosids. This gene duplication varies from 1,002 bp (*Morus*) to 1,404 bp (*Manihot*). In *Oenothera*, 2022 bp of *ndhF* were duplicated in the IR region. Conversely, the *ndhF* and *ycf1* genes of *Theobroma* are not duplicated in the IR.

Phylogenetic analyses of *Elaeagnus* and Rosids

Phylogenetic analysis was conducted using a gene data matrix based on 79 genes from 30 species with 75,370 bp aligned nucleotides (Fig 4). Rosids and asterids form two well supported monophyletic sister groups with strong support (100% bootstrap values). Rosids are a well-defined group with two strongly supported clades: Fabidae (*Prunus*, *Morus*, *Elaeagnus*, *Lotus*, *Theobroma*, *Manihot* and *Populus*); and Malvidae (*Gossypium*, *Castanea*, *Arabidopsis*, *Citrus* and *Eucalyptus*). The results of the present study confirmed that the genus *Elaeagnus* belongs to Fabidae and forms a sister relationship with *Morus* with 100% bootstrap values.

The *infA* gene has been lost from many angiosperms in land plants, and Millen et al. [21] suggested functional replacement of a nucleus copy. Our results indicate that the *infA* gene has been lost from rosids (including *Elaeagnus*). We also found that *trnH* duplication of the IR region was only present in *Elaeagnus*.

Comparison of the *trnH* gene in the IR region between Rhamnaceae and Elaeagnaceae

A previous study reported that Elaeagnaceae is closely related to Rhamnaceae, Dirachmaceae and Barbeyaceae [2,14,35]. In the present study, the chloroplast genome data revealed that the gene order in the LSC/IRb region of *E. macrophylla* continued to *rps19*, *trnH* and *rpl2*, while the IRa/LSC region continued to *rpl2*, *trnH* and *psbA*. Therefore, we compared the LSC/IRb (J_{LB}) and the IRa/LSC (J_{LA}) regions of Elaeagnaceae (*Elaeagnus* and *Shephedia*) and Rhamnaceae (*Rhamnus*). Fourteen species of Elaeagnaceae (*Elaeagnus* and *Shephedia*) and Rhamnaceae (*Rhamnus*) does experiments and aligned the sequences of *rps19-rpl2* (J_{LB} region) and *rpl2-psbA* (J_{LA}) regions.

The *rps19-rpl2* region of Elaeagnaceae differed from that of Rhamnaceae (Fig 5). The *rps19* (Fig 5A) and *rpl2* (Fig 5C) regions of Elaeagnaceae were highly similar to those of Rhamnaceae, whereas the areas surrounding the *trnH* and *trnH* gene differed greatly between these families (Fig 5B and 5D). Additionally, the *rpl2-psbA* region (Fig 6) between Elaeagnaceae and Rhamnaceae could be distinguished by the $\psi rps19$ gene (Fig 6E). The *rpl2*, *trnH*, and *psbA* genes are conserved in Elaeagnaceae and Rhamnaceae, whereas Elaeagnaceae has long gaps among coding genes and the *trnH* gene (*rpl2-trnH* and *trnH-psbA*).

Discussion

Rosids comprise the largest clade among eudicots (20 orders and 140 families), and include plants that form nitrogen-fixing symbioses (Elaeagnaceae, Rhamnaceae, Rosaceae, and Ulmaceae) [36], as well as many important crops (Fabaceae (legumne) and Rosaceae (fruit crops)). Accordingly, rosids plants have been very well studied, and almost the entire chloroplast genome is known [14, 20, 36–39].

Rosid gene contents

Previous studies revealed that the *infA* and *rpl22* genes and *atpF* intron have been lost or subjected to psuedogenation in rosids. Millen et al. [21] and Jansen et al. [22] found that the chloroplast genes, *infA* and *rpl22*, are transferred to the nucleus in rosids. However, the intron was lost from the *atpF* gene of Cassava (*Manihot esculenta*) [38]. Moreover, the *infA* gene has been independently lost multiple times from angiosperms and most rosids [22, 26]. Phylogenetic

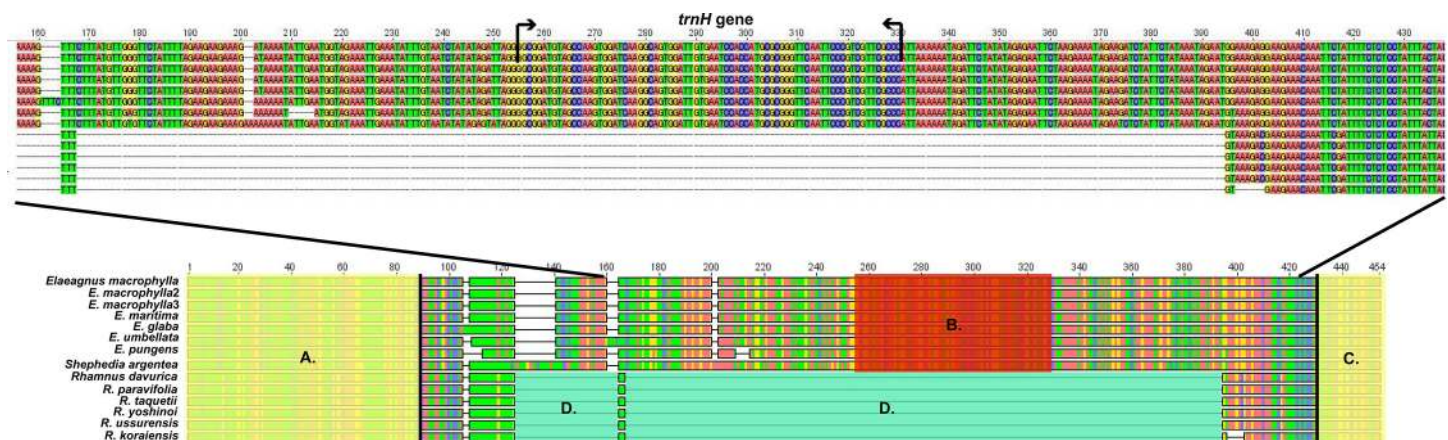


Fig 5. Comparison of the *rps19-rpl2* region sequences between Elaeagnaceae (*Elaeagnus* and *Shephedia*) and Rhamnaceae (*Rhamnus*). A: partial *rps19* gene B: *trnH* gene C: partial *rpl2* gene D: gaps.

doi:10.1371/journal.pone.0138727.g005

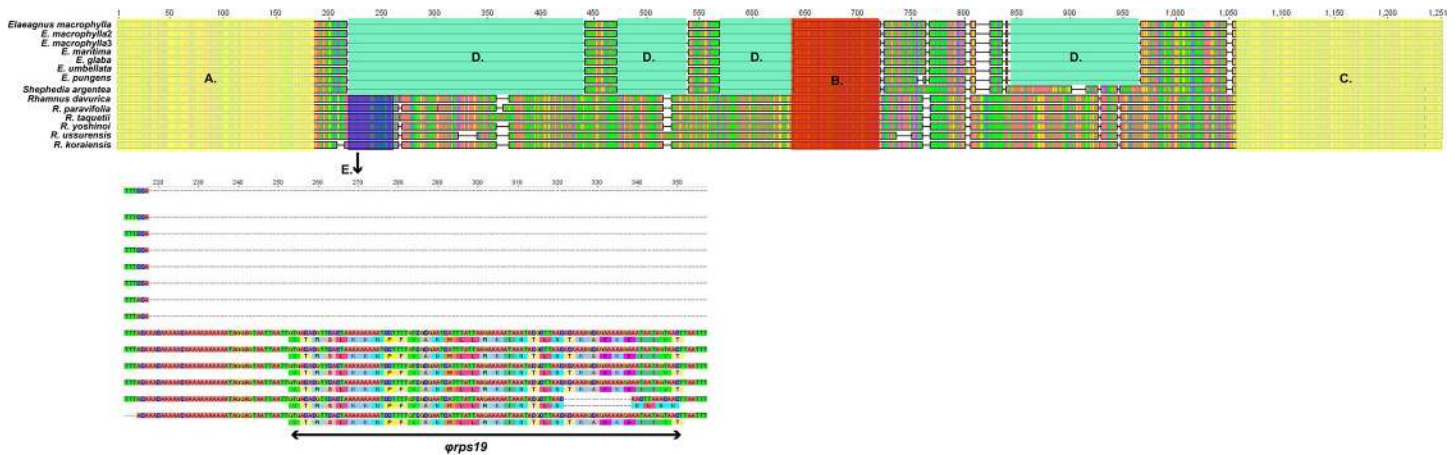


Fig 6. Comparison of *rpl2-psbA* region sequences between Elaeagnaceae (*Elaeagnus* and *Shepherdia*) and Rhamnaceae (*Rhamnus*). A: partial *rpl2* gene, B: *trnH* gene, C: partial *psbA* gene, D: gaps, E: *ψrps19* gene. Below Fig indicates the nucleotide sequence and translation of *ψrps19* (annotation by DOGMA).

doi:10.1371/journal.pone.0138727.g006

studies placed *Elaeagnus* sister to *Morus* in the Rosales clade [2, 26], and complete chloroplast genome analysis of *Morus* did not reveal an *infA* gene [40]. The *rpl22* gene has been lost from Fabaceae (*Glycine* and *Medicago*) and Fagaceae (*Castanea* and *Quercus*), and these plants have been independently transferred to the nucleus [22].

Our results also revealed the putative loss or formation of a pseudogene of the *infA* gene in *E. macrophylla* (Fig 2B). Moreover, the loss of the *infA* gene of 12 rosids was observed in this study (Fig 4). Su et al. [41] showed that *Quercus*, *Francoa* and *Cucumis* contain intact *infA* genes; however, no *infA* gene was observed in *Cucumis* in the present study (Fig 4).

The *rpl22* gene of *E. macrophylla* is intact, while it was lost from *Arabidopsis*, *Glycine*, *Lotus* and *Castanea*, and present in varying lengths in 18 other species.

Special event in the IR region of *Elaeagnus*

The chloroplast genome of land plants is highly conserved structurally, and the junction of large inverted repeats (IRs) is not essential to chloroplast genome function [18]. Because of black pine, *Conopholis* and *Phelipanche* of Orobanchaceae and *Erodium* was not present the IR region [42–44]. However, the IR region is a variable site on the chloroplast genome with useful features [17,45,46].

The gene arrangement of the IR region in most eudicots is different from that of monocots. The gene arrangement of basal plants and monocots in J_{LB} (LSC/IRb region) are *rpl23*, *rpl2*, *trnH*, *rps19* and *rpl22*, while gene arrangement in J_{LA} (IRa/LSC region) is *rpl23*, *rpl2*, *trnH*, *rps19*, and *psbA* [22, 41]. Thus, the *trnH* gene contains two IR regions. However, most eudicots do not undergo *trnH* gene duplication in the IR region. Nevertheless, the IR region border of most eudicots, including rosids plants, contains the *rps19* or *ψrps19* gene [14,20,22–25,38].

As shown in Fig 3, the LSC-IRb junction of the *Elaeagnus* species shows insertion of the *trnH* gene, whereas the other rosids species do not contain the *trnH* gene. Comparison of the LSC-IRb region of closely related species of Rhamnaceae revealed an approximately 600 bp gap after the *rps19* gene (Fig 5). In contrast, the IRa-LSC region contains 600bp gaps in Elaeagnaceae species. The *trnH* gene of Elaeagnaceae and Rhamnaceae is the same length, but Elaeagnaceae does not include the *rps19* gene (Fig 6). Consequently, Elaeagnaceae and Rhamnaceae have different gene contents and arrangements in the IR region.

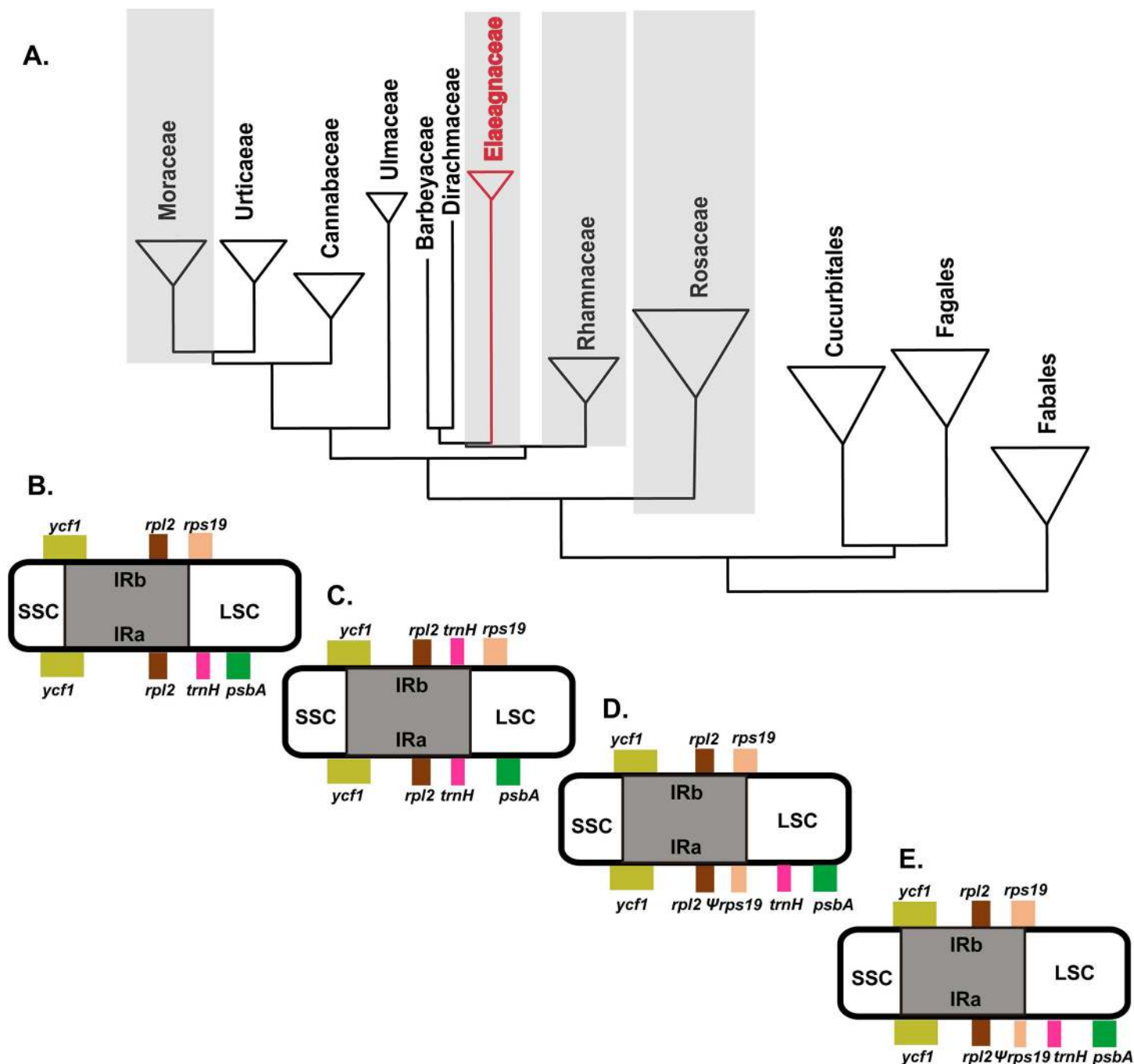


Fig 7. Duplication of *trnH* gene in Elaeagnaceae. A: Previous phylogenetic tree of Rosales (Zhang et al, [2]), B: Four junctions (LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC) of *Morus* in Moraceae, C: Four junctions of *Elaeagnus* in Elaeagnaceae, D: Four junctions of *Rhamnus* in Rhamnaceae, E: Four junctions of *Prunus* in Rosaceae.

doi:10.1371/journal.pone.0138727.g007

Comparisons of J_{LB} and J_{LA} in Rosales revealed that the *rps19* gene is not duplicated in *Morus*, whereas, *Prunus* contains a 108 bp duplication of the *rps19* gene. The gene *ycf1* is duplicated from 1,002 bp in *Morus* to 1,051 bp in *Prunus*. However, the *trnH* gene is duplicated in the J_{LB} (*rps19* is not duplicated) and J_{LA} border, and 1,215 bp of *ycf1* is duplicated in *Elaeagnus*. Hence, the IR length of *Elaeagnus* was longest in *Morus* and *Prunus*.

Wang et al. [19] has suggested two possible mechanisms of the evolution of IR expansions in Monocots. Wang et al. [19], double-strand break (DSB) events occurred within the IRb, after which the free 3' end of the broken strand was repaired against the homologous sequence in IRa. The repaired sequence then extends over the original IR-LSC junction, reaching the area downstream of *trnH*, resulting in duplication of the *trnH* gene in the newly repaired IRb. Similarly, the IR region extends in Elaeagnaceae.

Duplication of the *trnH* gene in Elaeagnaceae

Our data analyses confirmed IR evolution in Rosales (Fig 7A). The incomplete *rps19* gene of *Prunus* in Rosaceae (Fig 7E) and *Rhamnus* in Rhamnaceae (Fig 7D) was duplicated in the IR region. Conversely, *Morus* in Moraceae did not contain a duplicated *rps19* gene in the IR region (Fig 7B). Only the Elaeagnaceae was duplicated in the *trnH* gene (Fig 7C). The *trnH* gene duplication is a useful marker in Rosales, such as Dirachmaceae, Barbeyaceae and Elaeagnaceae. In a previous study, Richardson et al. [3] suggested a sister relationship between Rhamnaceae, Dirachmaceae and Barbeyaceae. In contrast, Zhang et al. [2] suggested a sister relationship among Elaeagnaceae, Dirachmaceae and Barbeyaceae, but this was not well supported in the Elaeagnaceae clade. Consequently, analyses of *trnH* duplication in the LSC/IRb junction and the IRa/LSC junction from different Moraceae and Elaeagnaceae would be of great value in systematics studies.

Conclusions

Here, we present the complete chloroplast genome of *Elaeagnus macrophylla* and compare it to that of rosids. The *infA* gene has been lost from the chloroplast genome or transferred to the nucleus in angiosperms [21]. Most rosids, including *E. macrophylla*, show loss of the *infA* gene. The chloroplast genome consists of a LSC (Large Single Copy), SSC (Small Single Copy) and two IR (Inverted Repeat) regions. The IR region is between 20 and 30 kb in length in angiosperms, and clearly differs among closely related species. The IR region of *E. macrophylla* differs owing to *trnH* gene duplication. Phylogenetic analysis strongly supports a monophyletic group of Rosales (*Elaeagnus*, *Morus* and *Prunus*). Previous studies did not clearly support Elaeagnaceae, Rhamnaceae, Dirachmaceae and Barbeyaceae in the molecular phylogenetic tree. In the present study, comparison of *trnH* gene duplication in two closely related families, Elaeagnaceae and Rhamnaceae, showed that no duplication occurred in Rhamnaceae, but that it occurred in Elaeagnaceae. Consequently, *trnH* gene duplication in Elaeagnaceae offers information that will be useful for systematics and elucidation of the relationship between Elaeagnaceae, Dirachmaceae and Barbeyaceae.

Supporting Information

S1 Table. Phylogenetic study taxa and Genebank accession numbers of references.
(DOCX)

S2 Table. IR junction analysis taxa and accession numbers.
(DOCX)

Acknowledgments

We thank Dr. Seongjun Park and Dr. Gurusamy Raman for valuable comments, JH Choi for sampling assistance. This research was supported by the Ministry of Environment in Korea (Daegu Regional Environmental Office, 2014).

Author Contributions

Conceived and designed the experiments: SP KSC. Performed the experiments: KSC OS. Analyzed the data: KSC. Wrote the paper: KSC.

References

1. Angiosperm Phylogeny Group (APG). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGIII. Bot J Linn Soc. 2009; 161: 105–121.
2. Zhang S-D, Soltis DE, Yang Y, Li D-Z, Yi T-S. Multi-gene analysis provides a well-supported phylogeny of Rosales. Mol phylogenet Evol. 2011; 60: 21–28. doi: [10.1016/j.ympev.2011.04.008](https://doi.org/10.1016/j.ympev.2011.04.008) PMID: [21540119](https://pubmed.ncbi.nlm.nih.gov/21540119/)
3. Richardson JE, Fay MF, Cronk QCB, Bowman D, Chase MA. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* plastid DNA sequences. Am J Bot. 2000; 87(9): 1309–1324. PMID: [10991902](https://pubmed.ncbi.nlm.nih.gov/10991902/)
4. Soltis DE, Soltis PS, Chase MW, Markt ME, Albach DC, Zanis M et al. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequence. Bot J Linn Soc. 2000; 133: 381–461.
5. Mabberly DJ. The plant-book. Cambridge University Press; 1998.
6. Sun M, Lin Q. A revision of *Elaeagnus* L. (Elaeagnaceae) in mainland China. J Syst Evol. 2010; 48(5): 356–390.
7. Heywood VH, Brummitt RK, Culham A, Seberg O. Flowering plant families of the world. Kew Publishing. London UK; 2007. pp. 135–136.
8. Palmer JD. Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. 1990; 6: 115–120. PMID: [2132730](https://pubmed.ncbi.nlm.nih.gov/2132730/)
9. Palmer JD. Plastid chromosomes. structure and evolution. In cell culture and somatic cell genetics in plants. Vol. 7A, The molecular biology of plastids, Vasil I.K., and Bograd L., eds. (San Diego: Academic Press); 1991. pp. 5–53.
10. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011; 76: 273–297. doi: [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4) PMID: [21424877](https://pubmed.ncbi.nlm.nih.gov/21424877/)
11. Jansen RK, Palmer JD. A chloroplast DNA inversion markers and ancient evolutionary split in the sunflower family (Asteraceae). Proc Natl Acad Sci USA. 1987; 84: 5818–5822. PMID: [16593871](https://pubmed.ncbi.nlm.nih.gov/16593871/)
12. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL et al. The chloroplast genome sequence of *Pelargonium X hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 2006; 23(11): 2175–2190. PMID: [16916942](https://pubmed.ncbi.nlm.nih.gov/16916942/)
13. Raubeson LA, Peery R, Chumley TW, Dzibek C, Fourcade HM, Boore JL et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics. 2007; 8: 174. PMID: [17573971](https://pubmed.ncbi.nlm.nih.gov/17573971/)
14. Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene. 2013; 528: 120–131. doi: [10.1016/j.gene.2013.07.037](https://doi.org/10.1016/j.gene.2013.07.037) PMID: [23900198](https://pubmed.ncbi.nlm.nih.gov/23900198/)
15. Wu CS, Wang YN, Liu SM. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights in to cpDNA evolution and phylogeny of extant seed plants. Mol Biol Evol. 2007; 24:1366–1379. PMID: [17383970](https://pubmed.ncbi.nlm.nih.gov/17383970/)
16. Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics diversified genomic structure of coniferous species. BMC Plant Biol. 2008; 8: 70. doi: [10.1186/1471-2229-8-70](https://doi.org/10.1186/1471-2229-8-70) PMID: [18570682](https://pubmed.ncbi.nlm.nih.gov/18570682/)
17. Plunkett GM, Dawnie SR. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst Bot. 2000; 25(4): 648–667.
18. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet. 1996; 252: 195–206. PMID: [8804393](https://pubmed.ncbi.nlm.nih.gov/8804393/)
19. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 2008; 8: 36. doi: [10.1186/1471-2148-8-36](https://doi.org/10.1186/1471-2148-8-36) PMID: [18237435](https://pubmed.ncbi.nlm.nih.gov/18237435/)
20. Saski C, Lee L-B, Daniell H, Wood TC, Tomkins J, Kim H-G et al. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Mol Biol. 2005; 59:309–322. PMID: [16247559](https://pubmed.ncbi.nlm.nih.gov/16247559/)

21. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*. 2001; 13: 645–658. PMID: [11251102](#)
22. Jasnen RK, Saski C, Hansen AK, Daniell H. Complete plastid genome sequences of three Rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol Biol Evol*. 2011; 28(1): 835–847. doi: [10.1093/molbev/msq261](#) PMID: [20935065](#)
23. Ibrahim RIH, Azuma J-I, Sakamoto M. Complete nucleotide sequence of the Cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants. *Genes Genet Syst*. 2006; 81: 311–321. PMID: [17159292](#)
24. Sato S, Makamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res*. 1999; 6: 283–290. PMID: [10574454](#)
25. Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, Chiu W-L et al. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenthera* plastomes. *Mol Gen Genet*. 2006; 263: 581–585.
26. Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *PNAS*. 2007; 104(49): 19369–19374. PMID: [18048330](#)
27. Wyman SK, Jansen RK, Boore HL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20: 3252–3255. PMID: [15180927](#)
28. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005; 33(2): W686–689.
29. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2009; 25(11): 1451–1452.
30. Biomatters: Geneious v.6.1.7 [<http://www.geneious.com/>]
31. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002; 30: 3059–3066. PMID: [12136088](#)
32. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-server. *Syst Biol*. 2008; 75: 758–771.
33. Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res*. 2012; 40(15): e115. PMID: [22730293](#)
34. Allan GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols*. 2006; 1: 2320–5. PMID: [17406474](#)
35. Soltis DE, Gitzendanner MA, Soltis PS. A 567-taxon data set for angiosperms: the challenges posed by bayesian analyses of large data sets. *Int J Plan Sci*. 2007; 168: 137–157.
36. Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM et al. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc Natl Acad Sci USA*. 1995; 92: 2647–2651.
37. Endress PK, Friis EM. Rosids-Reproductive structures, fossil and extant, and their bearing on deep relationships: Introduction. *Pl Syst Evol*. 2006; 260: 83–85.
38. Daniell H, Wurdack KJ, Kanagaraj A, Lee S-B, Saski C, Jansen RK. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of group II intron. *Theor Appl Genet*. 2008; 116: 723–737. doi: [10.1007/s00122-007-0706-y](#) PMID: [18214421](#)
39. Jung S, Cho I, Sosinski B, Abbott A, Main D. Comparative genomic sequence analysis of strawberry and other rosids reveals significant microsynteny. *BMC Research Notes*. 2010; 3: 168. doi: [10.1186/1756-0500-3-168](#) PMID: [20565715](#)
40. Ravi V, Khurana JP, Tyagi AK, Khurana P. The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet Genomes*. 2006; 3: 49–59.
41. Su H-J, Hongenhou SA, Al-Sadi AM, Kuo C-H. Complete chloroplast genome sequence of Omani Lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *PLoS ONE*. 2014; 9: 11.
42. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thumbergii*. *Proc Natl Acad Sci USA*. 1994; 91: 9794–9798. PMID: [7937893](#)
43. Blazier JC, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. 2011; 76: 263–272. doi: [10.1007/s11103-011-9753-5](#) PMID: [21327834](#)

44. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wuckett NJ, Zhang Y, Renner SS et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell*. 2013; 25: 3711–3725. doi: [10.1105/tpc.113.113373](https://doi.org/10.1105/tpc.113.113373) PMID: [24143802](https://pubmed.ncbi.nlm.nih.gov/24143802/)
45. Hansen DR, Dastidar SG, Penafior C, Kuehl JV, Boore JL, Janse RK. Phylogenetic and evolutionary implications of complete chloroplast genome sequences for four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Discorea* (Dioscoreaceae), and *Illium* (Schisandraceae). *Mol Phylogen Evol*. 2007; 45: 547–563.
46. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE*. 2013; 8(2): e57607. doi: [10.1371/journal.pone.0057607](https://doi.org/10.1371/journal.pone.0057607) PMID: [23460883](https://pubmed.ncbi.nlm.nih.gov/23460883/)