

RESEARCH

Open Access



Complete plastid genome of *Eriobotrya japonica* (Thunb.) Lindl and comparative analysis in Rosaceae

Liqun Shen¹, Qijie Guan¹, Awais Amin¹, Wei Zhu¹, Mengzhu Li¹, Ximin Li², Lin Zhang¹ and Jingkui Tian^{1*}

Abstract

Eriobotrya japonica (Thunb.) Lindl (loquat) is an evergreen Rosaceae fruit tree widely distributed in subtropical regions. Its leaves are considered as traditional Chinese medicine and are of high medical value especially for cough and emesis. Thus, we sequenced the complete plastid genome of *E. japonica* to better utilize this important species. The complete plastid genome of *E. japonica* is 159,137 bp in length, which contains a typical quadripartite structure with a pair of inverted repeats (IR, 26,326 bp) separated by large (LSC, 89,202 bp) and small (SSC, 19,283 bp) single-copy regions. The *E. japonica* plastid genome encodes 112 unique genes which consist of 78 protein-coding genes, 30 tRNA genes and 4 rRNA genes. Gene structure and content of *E. japonica* plastid genome are quite conserved and show similarity among Rosaceous species. Five large indels are unique to *E. japonica* in comparison with *Pyrus pyrifolia* and *Prunus persica*, which could be utilized as molecular markers. A total of 72 simple sequence repeats (SSRs) were detected and most of them are mononucleotide repeats composed of A or T, indicating a strong A or T bias for base composition. The Ka and Ks ratios of most genes are lower than 1, which suggests that most genes are under purifying selection. The phylogenetic analysis described the evolutionary relationship within Rosaceae and fully supported a close relationship between *E. japonica* and *P. pyrifolia*.

Keywords: *Eriobotrya*, Loquat, Chloroplast genome, Rosaceae, Gene evolution

Background

Chloroplast (cp), which is derived from free-living cyanobacteria through endosymbiosis (Keeling 2004), plays an essential role in photosynthesis and many biosynthetic activities such as biosynthesis of certain amino acids and fatty acids. Chloroplast contains its own genome which displays a typical quadripartite structure with two copies of inverted repeats separated by large single copy and small single copy (Nguyen et al. 2015). In general, plastomes of angiosperms range from 120 to 170 kb and mostly contain 100–120 different genes. The chloroplast genome is usually recognized as highly conserved in gene structure and content, especially in closely related groups. However, IR loss or expansion/contraction which

contributes to the variation in genome size can be easily found in some clades. Gymnosperms such as *Taxus chinensis var. mairei* (Zhang et al. 2014b) and *Cephalotaxus oliveri* (Yi et al. 2013) were identified to lose one copy of IR, leading to a decrease in genome size. Ma et al. (2013) found a significant IR expansion in *Mahonia bealei* in which 15 genes had an additional duplication in IR regions. Slight shift of IR/SC boundaries appears more commonly and may cause small changes in the extent of IR (Goulding et al. 1996). Consequently, structure differences of IR have been considered as important features to provide insights into gene evolution among species.

The plastid genome is good resource to provide sufficient information for phylogenetic analysis and DNA barcoding. Thanks to rapid development of next-generation sequencing, the number of whole plastid genome available is increasing constantly, which makes large-scale phylogenetic research based on plastid genomes possible (Jansen et al. 2007). In addition, comparative analysis

*Correspondence: tjtk@zju.edu.cn

¹ Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China

Full list of author information is available at the end of the article

among species can provide large amount of genetic information such as insertion or deletion (indel) and nucleotide substitutions, which can be utilized for diversity analysis and molecular markers (Cho et al. 2015).

There are over 100 genera containing approximately 3000 species in Rosaceae which is medium-sized, but of high economic importance (Potter et al. 2007). As the third most important economical family in temperate regions, many edible fruits like apple (*Malus*), loquat (*Eriobotrya*), pear (*Pyrus*) and peach (*Prunus*) as well as ornamental plants such as rose (*Rosa*) are included. However, till now, phylogenetic relationships within Rosaceae family have been uncertain. Traditional morphological studies divided Rosaceae into 4 subfamilies including Spiraeoideae, Maloideae, Rosoideae and Prunoideae according to the type of fruit (Kalkman 1988). Molecular studies however, put forward different opinions that Rosaceae was divided into three subfamilies: Dryadoideae, Rosoideae and Spiraeoideae (Potter et al. 2007). As the plastid genome can offer useful phylogenetic information, several plastomes of Rosaceae plants have been sequenced and reported, such as *Malus domestica* (http://www.rosaceae.org/projects/apple_genome), *Pyrus pyrifolia* (AP012207.1), *Prunus persica* (HQ336405) and *Fragaria virginiana* (NC_019602). Most of these studies focused on the comparative analysis and evolutionary studies within genera. Wang et al. (2013) reported the chloroplast genome of *Prinsepia utilis* and reconstructed phylogenetic relationships within Rosaceae but with low bootstrap values. So more closely related plastid genomes are needed to deal with intrafamilial classifications within Rosaceae.

Eriobotrya japonica (Thunb.) Lindl (loquat), which belongs to Rosaceae, is widely distributed in temperate and subtropical zones of Asian, European and American countries (Gisbert et al. 2009). *Eriobotrya japonica* originated in China and later was introduced to Mediterranean basin, Japan and Florida (Blasco et al. 2014). *Eriobotrya japonica* is an important economic fruit crop with high edible, medical as well as ornamental values. The fruit of *E. japonica* is greatly appreciated by consumers due to its good taste and high nutrition values. Its leaves, where pharmaceutically active compounds proved to be found (Sharpe 2010), have long been considered as one kind of traditional Chinese medicine. Due to natural multiplication and artificial cultivation for long time, genetic variation can be easily found within loquat which makes the identification and classification of loquat more difficult. Molecular markers such as RAPD, AFLP or SSRs have been rapidly developed these years and applied into loquat studies. Vilanova et al. (2001) applied RAPD technology into 33 loquat cultivars for identification from different regions. Soriano et al. (2005) first demonstrated

the usefulness of SSRs markers in genetic analysis and identification in *E. japonica*. Many researches utilized AFLP and SSRs markers to investigate genetic variation, diversity and identification among loquat accessions (He et al. 2011; Fukuda et al. 2013; Blasco et al. 2014), which provided information for diversity analysis and plant breeding program. However, few molecular markers based on plastomes have been developed since completed plastid genome of *E. japonica* is not available, lagging behind other species of Rosaceae.

In this study, we report the first complete plastid genome of *E. japonica* using next-generation sequencing method (Illumina HiSeq 2000) and conduct comparative analysis with other Rosaceous species, *P. pyrifolia* and *P. persica* in particular, which will not only help with species identification or germplasm selection but also provide insights into phylogenetic evolution of Rosaceae family.

Methods

Taxon sampling, DNA sequencing and genome assembly

Fresh leaves of *Eriobotrya japonica* were collected from the nursery garden in Zhejiang University to prepare for isolation of chloroplast DNA. The two-step Percoll gradient method (Aronsson and Jarvis 2002) was used to isolate purified *E. japonica* chloroplasts. The chloroplast DNA was then extracted and purified using the CTAB method (He 2011) with slight modifications. The whole chloroplast library was constructed by 5 µg purified cp DNA. The cp genome of *E. japonica* was then sequenced using Illumina HiSeq 2000 system, which generated in total 4,484,369 raw pair-end reads for this project. As many low-quality reads were included in the generated raw reads, SolexaQA v1.0 (Cox et al. 2010) was used to filter low-quality reads with the settings-h 27 and-l 60. Then 6050 contigs with an average length of 206 bp were generated by Soapdenovo v1.3 (Luo et al. 2012) with Kmer size equal to 59 bp. The similar strategy with Zhang et al. (2014b) was then used for the assembly. The complete plastid genome of *Pyrus pyrifolia* (AP012207) extracted from NCBI (<http://www.ncbi.nlm.nih.gov/>) was selected as a reference and Blastz (Schwartz et al. 2003) perl script was used to map all contigs to the reference in order to determine the order and direction of these contigs. The following steps was taken repeatedly to fill up the gaps: first, BLAT v.34 (Kent 2002) was used to map raw sequence reads to both ends of the assembled contigs; second, assembled contigs were extended by connecting up with overlapping reads which were best overlapped with the contigs (Zhang et al. 2014b). For the gaps remaining between contigs, we designed six pairs of primers for PCR amplification to close these gaps (Additional file 1: Table S1). Also, to avoid mistakes during our assembly,

another four primers were designed to confirm IR/SC boundaries (Additional file 1: Table S1). All PCR products were sequenced by conventional Sanger sequencing and the results were consistent with our assembly. Thus, the whole plastid genome of *E. japonica* was obtained.

Genome annotation, codon usage and comparative analysis

The *Eriobotrya japonica* plastid genome was annotated using the program Dual Organellar GenoMe Annotator (Wyman et al. 2004). Start and stop codons of protein-coding genes were then manually checked and adjusted if necessary, by comparing *E. japonica* with other Rosaceae plastomes. Any genes not found by DOGMA were verified missing by the use of blastn and blastx online searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Transfer RNA genes were identified and further confirmed with tRNAscan-SE 1.21 (Lowe and Eddy 1997) using the mito/chloroplast source setting. Then the whole record was deposited into Genbank with accession number KT633951. The physical map of the complete plastid genome was drawn by OGdraw v1.2 online tool (Lohse et al. 2007). Codon usage and relative synonymous codon usage (RSCU) analysis of all protein-coding genes were conducted using MEGA5 (Tamura et al. 2011). In order to perform identity analysis, seven Rosaceae cp genomes, *Fragaria virginiana* (NC_019602), *Pentactina rupicola* (JQ041763), *Prinsepia utilis* (NC_021455), *Prunus persica* (HQ336405), *Pyrus pyrifolia* (AP012207), and *Rosa odorata* (KF753637) were extracted from NCBI and were aligned with *E. japonica* respectively to compute pairwise identity using mVISTA program (Frazer et al. 2004). Large indels with no less than 40 bp in size were detected through pairwise alignment among *E. japonica*, *P. pyrifolia* and *P. persica*. For these unique large indels of *E. japonica*, five pairs of primers were also designed to avoid assembly errors (Additional file 2: Table S2).

Repeat structure identification

Reputer online program (Kurtz et al. 2001) was used to detect forward, palindrome, reverse and complementary repeat structures with structure size greater than 20 bp and hamming distance equal to 0 (100% identity).

Simple sequence repeats (SSRs) were detected within completed genomes of *Eriobotrya japonica* using MISA v1.0 (Thiel et al. 2003) with parameter settings of mononucleotide, dinucleotide, tri- or tetranucleotide and hexanucleotide repeats no less than 10, 12, 15 and 24 bases respectively.

Nucleotide substitution in coding regions

All 78 functional protein-coding genes were extracted from *Eriobotrya japonica*, *Pyrus pyrifolia* and *Prunus*

persica chloroplast genome. Each gene of *P. pyrifolia* and *P. persica* was aligned with that of *E. japonica* using clustalX v2.1 (Thompson et al. 1997). The alignment file was then put into Dnasp v5 (Librado and Rozas 2009) to calculate the synonymous (Ks) and nonsynonymous (Ka) substitution rates. Transition (Ts), transversion (Tv) and P-distance of each gene were calculated using MEGA5 (Tamura et al. 2011).

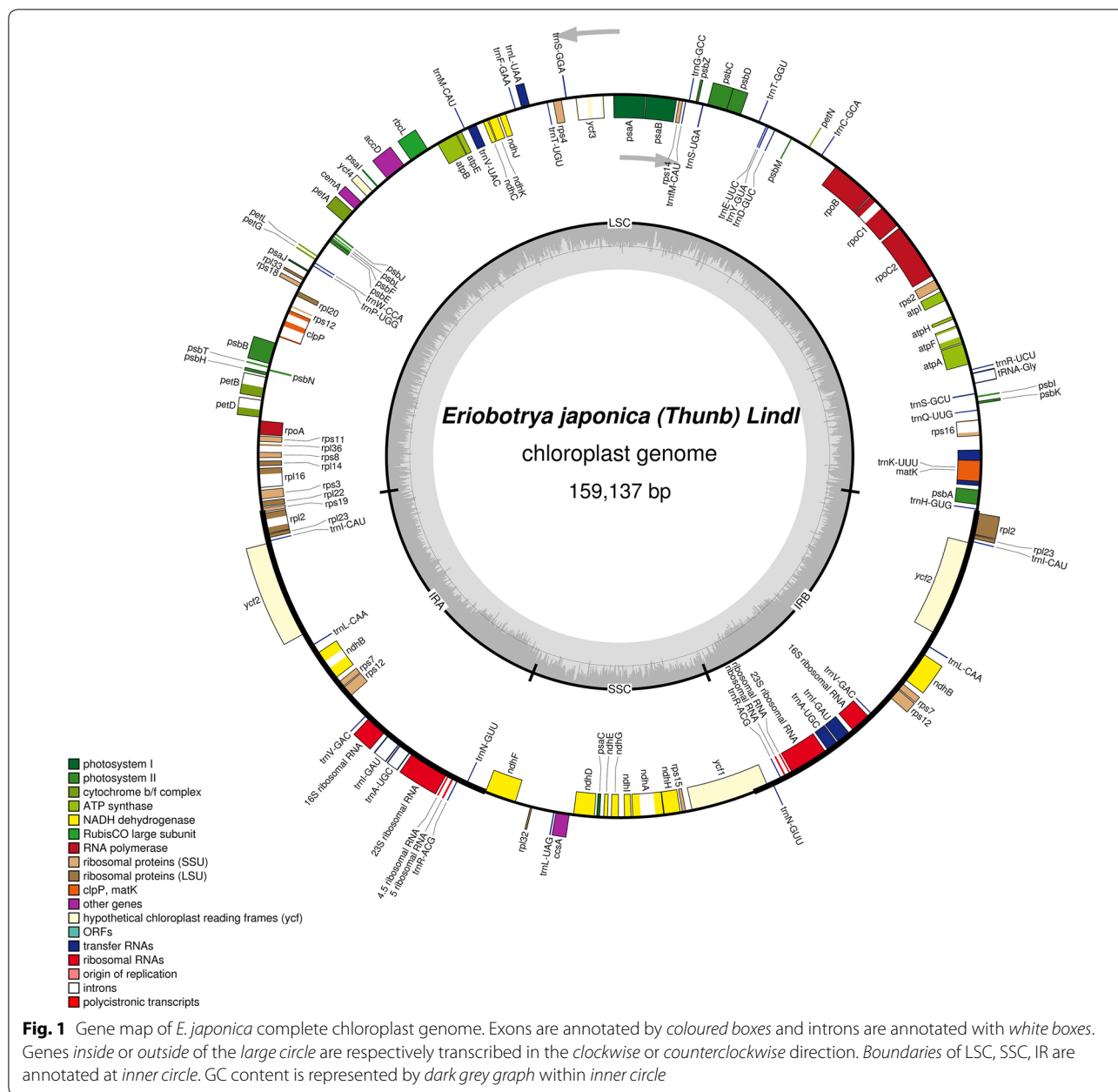
Phylogenetic analysis

All 78 protein-coding genes extracted from the plastomes of 7 Rosaceous plants (*Fragaria virginiana*, NC_019602; *Pentactina rupicola*, JQ041763; *Prinsepia utilis*, NC_021455; *Prunus persica*, HQ336405; *Pyrus pyrifolia*, AP012207; *Rosa odorata*, KF753637 and *Eriobotrya japonica*) and *Morus indica* (NC_008359, as an outgroup) were used to construct Maximum likelihood (ML) tree and Maximum parsimony (MP) tree. All 78 protein coding genes were concatenated and then aligned using MAFFT v7 (Katoh and Standley 2013). Gblock v0.91 (Talavera and Castresana 2007) was used to select the conserved regions, which resulted in a total length of 67,104 bases for MP and ML analysis. The aligned sequences of conserved regions were then tested in DAMBE v5.3.19 (Xia and Lemey 2009) for saturation of substitution. The results revealed that Iss (index of substitution saturation) < Iss.c (the critical Iss value), indicating no substantial saturation in these sequences (Xia et al. 2003). The General Time Reversible model of substitution, incorporating invariant sites and a gamma distribution (GTR + I+G), was selected as the best model by jmodeltest2.1.7 (Posada 2008). To construct the ML tree, RaxML7.0.4 (Stamatakis 2006) was used with 1000 non-parametric replications for bootstrap settings. The base frequencies estimated by RaxML was 0.305 (A), 0.174 (C), 0.203 (G), 0.317 (T) while rate matrix was 1.54, 3.07, 0.24, 0.77, 3.37, 1.00 for ac, ag, at, cg, ct, gt, respectively. PAUP*4.0b10 (Swofford 2003) was used to construct MP tree with the following options implemented: Accctran was used, gaps were treated as missing, heuristic search mode used 1000 replications for bootstrap support, tree bisection-reconnection (TBR) branch-swapping, Mul-Trees in effect, and steepest descent off.

Results

Genome features of *E. japonica*

The complete plastid genome of *Eriobotrya japonica* is 159,137 bp in length with a double-strand circle structure with a pair of IRs of 26,326 bp separated by a small single copy of 19,283 bp and a large single copy of 87,202 bp (Fig. 1). The plastid genome includes 112 functional genes in all, of which 78 are protein-coding genes, 30 are tRNA genes and 4 are rRNA genes (Table 1). Eighteen



genes duplicates in IR region including 6 protein-coding genes, 7 tRNA genes and 4 rRNA genes. Eighteen genes comprising six tRNA genes contain one or two introns. Both *ycf3* and *clpP* gene contain two introns while *rps12* gene is trans-spliced with a share 5' exon in LSC region and two 3' exons located in IR region. Three genes, *ycf1*, *rps19* and *infA* was found as non-functional genes in the *E. japonica* plastid genome. *ycf1* and *rps19*, located in IR/SC boundary regions, became truncated as incomplete duplications of the normal copy. *infA* gene was found in *R. odorata* while it was non-functional in *E. japonica* and

some other Rosaceae plastomes as several stop codons were identified in *infA* coding regions (Fig. 2).

Overall, 22,772 codons encoding 78 functional protein-coding genes were identified from *E. japonica* plastid genome and classified depending on codon usage (Table 2). There were 2388 codons encoding for leucine and 252 codons for cysteine, which denoted respectively the most and the least amino acids. The RSCU analysis showed a great A/T bias in the protein-coding genes, especially in the third position of the codon with G/C content at 25.9%. This pattern has been identified in former studies (Redwan et al.

Table 1 List of genes located in *E. japonica*

Group of genes	Gene names
Ribosomal RNA genes	<i>rrn16[#], rrn23[#], rrn4.5[#], rrn5[#]</i>
Transfer RNA genes	<i>trnA-UGC[#], trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC[*], trnG-UCC, trnH-GUG, trnI-CAU[#], trnI-GAU[#], trnK-UUU[*], trnL-CAA[#], trnL-UAA[*], trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU[#], trnP-UGG, trnQ-UUG, trnR-UCU, trnR-ACG[#], trnS-UGA, trnS-GCU, trnS-GGA, trnT-GGU, trnT-UGU, trnV-UAC[*], trnV-GAC[#], trnW-CCA, trnY-GUA</i>
Small subunit of ribosome	<i>rps2, rps3, rps4, rps7[#], rps8, rps11, rps12[*], rps14, rps15, rps16[*], rps18, rps19[#]</i>
Large subunit of ribosome	<i>rpl2[#], rpl14, rpl16[*], rpl20, rpl22, rpl23[#], rpl32, rpl33, rpl36</i>
DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1[*], rpoC2</i>
Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>
Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
Subunits of cytochrome	<i>petA, petB[*], petD[*], petG, petL, petN</i>
Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF[*], atpH, atpI</i>
ATP-dependent protease	<i>clpP[*]</i>
Large subunit of Rubisco	<i>rbcl</i>
Subunits of NADH	<i>ndhA[*], ndhB[#], ndhC, ndhD, ndhE, ndhF[#], ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Maturase	<i>matK</i>
Envelop membrane protein	<i>cemA</i>
Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
c-Type cytochrome synthesis gene	<i>ccsA</i>
Conserved open reading frames	<i>ycf1, ycf2[#], ycf3[*], ycf4</i>

* Gene with intron

Gene duplicated in IR

2015; Tangphatsornruang et al. 2011). However, whether this pattern relates to selection for translational efficiency or nucleotide compositional biases remains unclear.

Comparison with other plastid genomes in Rosaceae

Structure and content of *Eriobotrya japonica* plastid genome are conserved and share similar features with other Rosaceae plastid genomes (Table 3). The average length of six cp genomes is 157,619 bp. Among these seven Rosaceae plants, *Pyrus pyrifolia* is the largest in size at 159,922 bp while *Fragaria virginiana* is shortest at 155,621 bp. *Eriobotrya japonica* ranks following *P. pyrifolia* with 159,137 bp. *P. pyrifolia* contains the longest LSC and IR, which are 87,901 and 26,392 respectively. The SSC size of *E. japonica* is the largest, which is 46 bp longer than *P. pyrifolia*.

The multialignment of Rosaceae plastid genomes (Fig. 3) illustrated that the IR regions revealed higher identity in comparison with single copy (SC) regions. *Eriobotrya japonica* plastid genome is most similar to *P. pyrifolia* and most divergent from *Rosa odorata* and *Fragaria virginiana*. Another aspect is that coding regions show higher conservation than non-coding regions.

The GC content is similar among these species ranging from 36.7 to 37.2% as the GC content of seed plants usually ranges from 34 to 40% (Guisinger et al. 2011; Yap et al. 2015; Cai et al. 2006; Raubeson et al. 2007). *Rosa odorata* and *F. virginiana* contain the highest GC content (37.2%) while *P. pyrifolia* contains the lowest (36.6%). In

addition, IR regions contain the highest GC content followed by LSC and SSC regions.

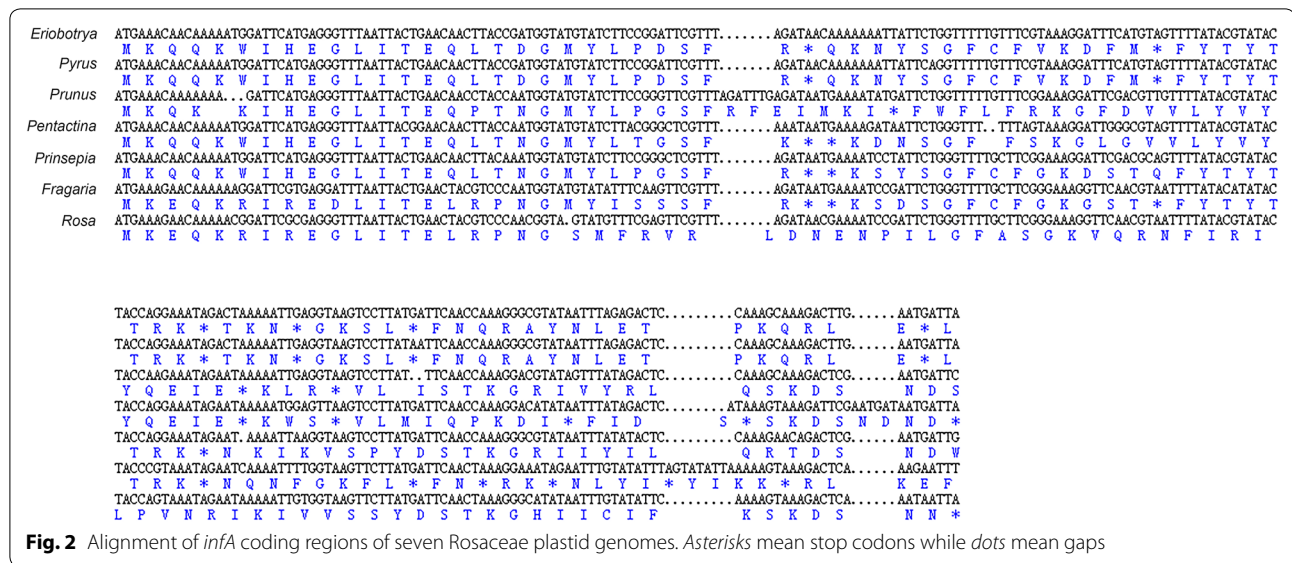
Repeat structure

A total of 42 repeat structures containing no less than 20 bp with 100% identity were found (Additional file 3: Table S3). The number of direct, palindrome, complement and reverse structures are 29, 8, 3, 2, respectively. Of all these structures, the longest one is 59 bp located between *rpl32* and *trnL-UAG*. Most repeat structures located in intergenic or intron regions while two repeat structures located in *ycf1* gene.

Seventy-two simple sequence repeats including 70 mononucleotides and 2 dinucleotides were identified in *E. japonica* plastid genome (Table 4). Among 70 mononucleotides, 26 A stretches, 42 T stretches and 2 C stretches were detected while no G stretch was found. Two dinucleotide repeats, are both composed of A and T, 1 AT stretch (6 repeat motifs) and 1 TA (8 repeat motifs) stretch. The size of all SSRs are between 10 and 20 bp in length. Only five genes, *atpB*, *rpoB*, *rpoC2*, *matK* and *ycf1* appear to harbor one or two SSRs while others are all located in intergenic or intron regions. The numbers of SSRs located in LSC, IR and SSC are 58, 4, and 10, respectively.

Nucleotide substitution and indels in *Eriobotrya japonica*, *Pyrus pyrifolia* and *Prunus persica*

In this research, 78 functional protein-coding genes were classified into eleven groups according to their function,



the Ka/Ks among *Pyrus pyrifolia*, *Prunus persica* and *Eriobotrya japonica* were computed. Although protein-coding regions were always conserved among closely related groups, Ka and Ks varied in different regions (Table 5) and different functions (Additional file 4: Table S4). Both Ks and Ka of all regions between *P. persica* and *E. japonica* were higher than those between *P. pyrifolia* and *E. japonica*. In two comparisons, Ka and Ks were highest in SSC regions and lowest in IR regions. When compared with *P. persica*, *psaC* revealed the highest Ks while *rps18* revealed the highest Ka and genes related with ATP synthase had the highest Ks value. When compared with *P. pyrifolia*, *rpl32* and *rpl36* revealed the highest Ks and Ka value respectively and genes of Photosystem I described the highest Ks value. The Ka/Ks values of all gene groups were less than one in both two comparisons (Fig. 4a) and only one gene *accD* contained the Ka/Ks value slightly greater than one when compared with *P. pyrifolia*.

A total of 675 and 5040 nucleotide substitutions were found in *E. japonica* plastid genome when compared with *P. pyrifolia* and *P. persica* respectively. The LSC region contained the most nucleotide substitutions as it's the largest area of the genome (Table 5). Transitions and transversions in coding regions are shown in Additional file 5: Table S5. Overall, transitions occurred more than transversions when compared with *P. persica* than with *P. pyrifolia*. Also, the Ts and Tv ratio varied among different gene groups (Fig. 4b). Genes related with Small subunit of ribosome and Photosystem II had a relatively higher ratio in both two comparisons. Specially, the line charts of two comparisons both illustrated similar Ts/Tv tendency, which indicated a Ts/Tv bias for different functional groups.

Large indels (>40 bp) were also identified among these three species (Table 6). Compared to *P. pyrifolia*, 11 indels

were detected in *E. japonica* plastid genome comprising 7 deletions and 4 insertions. The largest indel was a 417 bp deletion located in intergenic region between *trnR-UCU* and *atpA*. When compared to *P. persica*, 20 indels including 8 deletions and 12 insertions were identified. All large indels were distributed in non-coding regions of LSC and SSC, while none were located in IR regions.

IR expansion and contraction

The SC and IR boundaries of six Rosaceae plastid genomes were aligned in Fig. 5. *Eriobotrya japonica* contained almost the same IR/SC structure with *Pyrus pyrifolia*, *Prunus persica* and *Pentactina rupicola* in which IRb/SC boundaries lay respectively in coding regions of *rps19* and *ndhF*. *Rosa odorata* and *Fragaria virginiana* contained another structure with *rps19* gene and *ndhF* gene apart from LSC/IRb boundary, which led to the loss of *rps19* pseudogene at LSC/IRa boundary. *Prinsepia utilis* contained a LSC/IRb boundary with *rps19* involved and a IRb/SSC boundary apart from *ndhF*. The IRa/SSC boundary of all six species extended to *ycf1* coding regions with the range from 978 bp (*P. utilis*) to 1105 bp (*R. odorata*), which led to a nonfunctional *ycf1* gene in IRb. The IRa/LSC boundary revealed a large variation from 4 (*R. odorata*) to 104 bp (*P. pyrifolia*) as *trnH-GUG* gene located upstream of IRa/LSC boundary.

Phylogenetic analysis

All 78 protein-coding genes possessed by seven Rosaceae plastomes were extracted to perform ML and MP analysis using *Morus indica* as an outgroup.

Maximum likelihood analysis resulted in a single tree with—lnL of 150822.7449 using GTR + G + I substitution

Table 2 Codon usage and RSCU analysis of *E. japonica* cp genome

Amino acid	Codon	No.	RSCU	tRNA	Amino acid	Codon	No.	RSCU	tRNA
Phe	UUU	850	1.34		Tyr	UAU	691	1.61	
	UUC	420	0.66	<i>trnF-GAA</i>		UAC	166	0.39	<i>trnY-GUA</i>
Leu	UUA	814	2.05	<i>trnL-UAA</i>	TER	UAA	43	1.65	
	UUG	485	1.22	<i>trnL-CAA</i>		UAG	19	0.73	
	CUU	495	1.24		His	CAU	414	1.54	
	CUC	145	0.36			CAC	123	0.46	<i>trnH-GUG</i>
	CUA	305	0.77	<i>trnL-UAG</i>	Gln	CAA	634	1.55	<i>trnQ-UUG</i>
	CUG	144	0.36			CAG	182	0.45	
Ile	AUU	987	1.50		Asn	AAU	834	1.55	
	AUC	364	0.55	<i>trnI-GAU</i>		AAC	244	0.45	<i>trnN-GUU</i>
	AUA	622	0.95		Lys	AAA	904	1.53	<i>trnK-UUU</i>
Met	AUG	537	1.00	<i>trnM-CAU</i>		AAG	281	0.47	
Val	GUU	472	1.47		Asp	GAU	743	1.62	
	GUC	134	0.42	<i>trnV-GAC</i>		GAC	175	0.38	<i>trnD-GUC</i>
	GUA	504	1.57	<i>trnV-UAC</i>	Glu	GAA	897	1.50	<i>trnE-UUC</i>
	GUG	173	0.54			GAG	299	0.50	
Ser	UCU	485	1.71		Cys	UGU	190	1.51	
	UCC	259	0.92	<i>trnS-GGA</i>		UGC	62	0.49	<i>trnC-GCA</i>
	UCA	327	1.15	<i>trnS-UGA</i>	TER	UGA	16	0.62	
	UCG	155	0.55		Trp	UGG	396	1.00	<i>trnW-CCA</i>
Pro	CCU	363	1.56		Arg	CGU	301	1.33	<i>trnR-ACG</i>
	CCC	172	0.74			CGC	97	0.43	
	CCA	268	1.15	<i>trnP-UGG</i>		CGA	312	1.38	
	CCG	130	0.55			CGG	99	0.44	
		ACU	483	1.64		Ser	AGU	364	1.29
Thr	ACC	211	0.72	<i>trnT-GGU</i>	AGC		109	0.39	
	ACA	363	1.23	<i>trnT-UGU</i>	Arg	AGA	407	1.80	<i>trnS-GCU</i>
	ACG	123	0.41			AGG	138	0.61	
Ala	GCU	590	1.85		Gly	GGU	528	1.35	
	GCC	194	0.61			GGC	166	0.42	<i>trnG-GCC</i>
	GCA	349	1.10	<i>trnA-UGC</i>		GGA	620	1.58	<i>trnG-UCC</i>
	GCG	141	0.44			GGG	254	0.65	

Table 3 Summary of seven Rosaceae plastid genome features

	<i>E. japonica</i>	<i>P. pyrifolia</i>	<i>P. rupicola</i>	<i>P. persica</i>	<i>F. virginiana</i>	<i>P. utilis</i>	<i>R. odorata</i>
Accession	KT633951	NC_015996	NC_016921	NC_014697	NC_019602	NC_021455	KF753637
Length	159,137	159,922	156,612	157,790	<i>155,621</i>	156,328	156,634
LSC	87,202	87,901	84,970	85,968	85,586	85,239	85,767
SSC	19,283	19,237	18,941	19,060	<i>18,145</i>	18,485	18,761
IR	26,326	26,392	26,351	26,381	<i>25,945</i>	26,302	26,053
GC% overall	36.7	36.6	36.8	36.8	37.2	36.9	37.2
in LSC	34.5	34.3	34.6	34.6	35.1	34.7	35.1
in IR	42.7	42.7	42.6	42.6	42.8	42.7	42.7
in SSC	30.3	30.4	30.6	30.6	31.1	30.5	31.2

Number in bold italics indicates the largest value of line

Number in italics indicates the smallest value of line

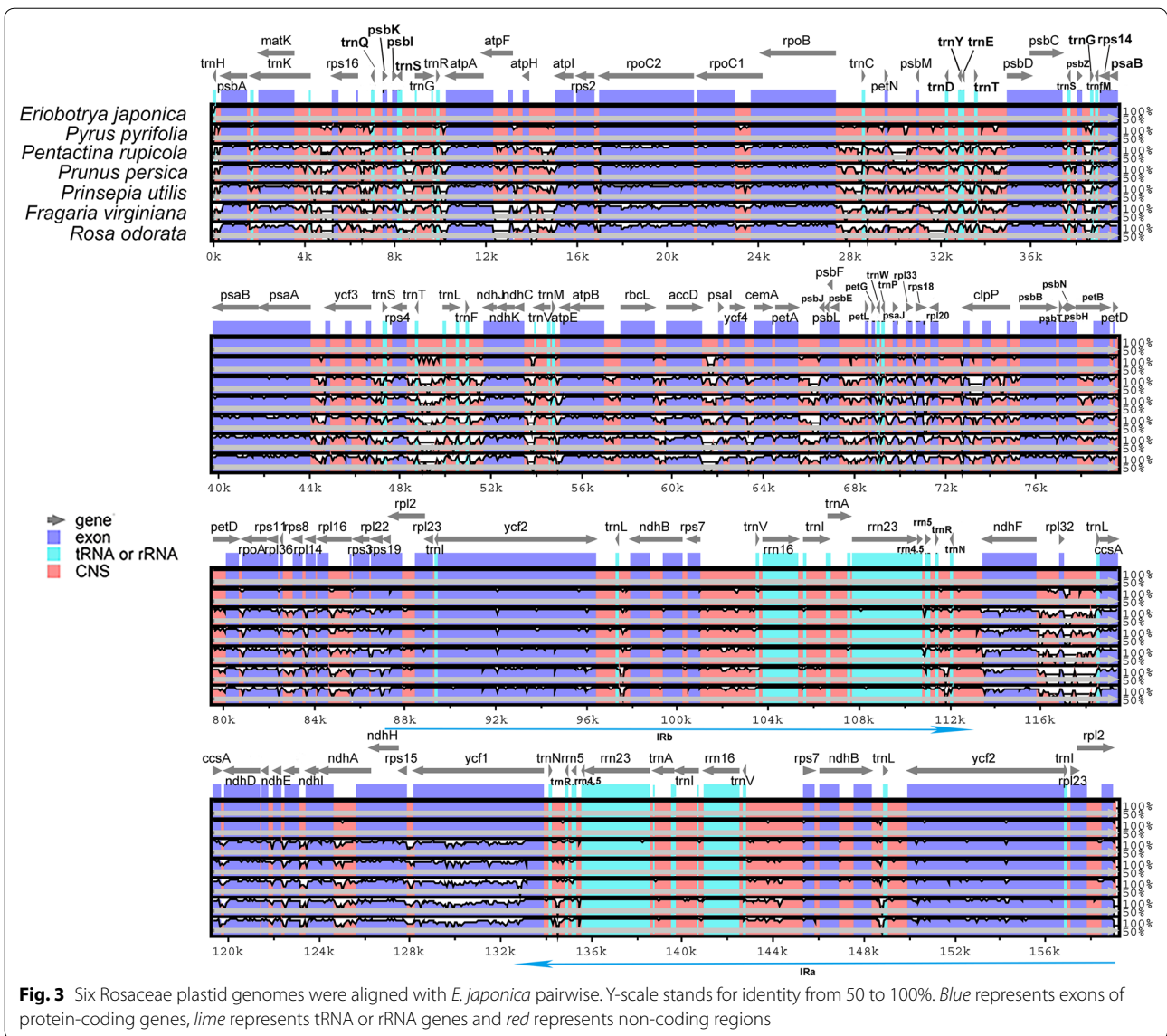


Fig. 3 Six Rosaceae plastid genomes were aligned with *E. japonica* pairwise. Y-scale stands for identity from 50 to 100%. Blue represents exons of protein-coding genes, lime represents tRNA or rRNA genes and red represents non-coding regions

model. Maximum parsimony analysis generated one most parsimonious tree with a length of 10,628, a consistency index of 0.8852, and a retention index of 0.7117. Our MP (Additional file 6: Fig. S1) and ML (Fig. 6) trees shared similar topology. Both of them strongly supported a clade of *Fragaria virginiana* and *Rosa odorata* as sister to the remaining sampled Rosaceae species. *Prunus persica* and *Prinsepia utilis* form a clade sister to *Pentactina rupicola*, *Pyrus pyrifolia*, and *Eriobotrya japonica*. *Pyrus pyrifolia* and *Eriobotrya japonica* are sister to each other with a 100% bootstrap value.

Discussion

We compared *Eriobotrya japonica* with available sequences of Rosaceae plastid genomes, indicating the conservation of *E. japonica* plastid genome with similar

structure and gene content. However, one significant difference among these species is that *Rosa odorata* encodes *infA* gene which is non-functional in other plastid genomes due to several stop codons in coding regions (Fig. 2). It's well recognized that plastid DNA transfers to nuclear DNA at a considerably high rate (Huang et al. 2003). However, until now, successful gene transfers to nuclear genome have been documented for only four genes in Rosids, *infA*, *rpl22*, *rpl32* and *rpoA* (Jansen et al. 2011). *infA* gene, which encodes for translation initiation factor1 was reported to transfer from plastid genome to nuclear genome for functional use several times especially in Rosids (Millen et al. 2001). So further studies could be focused on gene evolution of *infA* in Rosaceae.

Table 4 Distribution of SSRs (mononucleotide) loci in the *E. japonica* chloroplast genome

Size (bp)	Number and start position			
	A stretch	C stretch	T stretch	G stretch
10	9 (13897, 62517, 68030, 84710, 115751, 117249, 124942, 125608, 143104)	0	17 (174, 4752, 8304, 9472, 9942, 11831, 13310, 14438, 14463, 16746, 26698, 56906, 84155, 85388, 86446, 103167, 130945)	0
11	6 (16759, 44490, 46845, 51587, 68061, 158959)	0	6 (2753, 6523, 9110, 12905, 18997, 87311)	0
12	3 (187, 27779, 79320)	0	4 (1612, 51543, 67720, 85645)	0
13	2 (48460, 74027)	0	1 (71138)	0
14	1 (37903)	2 (25627, 116685)	6 (12551, 32565, 37948, 65963, 73345, 123131)	0
15	1 (6806)	0	3 (71854, 74083, 116600)	0
16	4 (69960, 80822, 131576)	0	2 (14900, 125047)	0
17	1 (7681)	0	1 (82414)	0
18	0	0	0	0
19	0	0	1 (59616)	0
20	0	0	1 (83566)	0
Total	26	2	42	0

Table 5 Synonymous rate, nonsynonymous rate, transition (Ts) and transversion (Tv) in LSC, IR and SSC regions among *E. japonica*, *P. pyrifolia* and *P. persica*

Region	<i>E. japonica</i> vs <i>P. pyrifolia</i>				<i>E. japonica</i> vs <i>P. persica</i>			
	Ks	Ka	Ts	Tv	Ks	Ka	Ts	Tv
LSC	0.0062	0.0014	209	313	0.0717	0.0091	1863	1706
IR	0.0035	0.00004	3	6	0.0110	0.0041	81	77
SSC	0.0091	0.0026	60	76	0.0942	0.0175	584	572
All	0.0053	0.0024	272	395	0.0694	0.0138	2528	2355
Ratio	Ka/Ks = 0.4528		Ts/Tv = 0.6858		Ka/Ks = 0.2126		Ts/Tv = 1.0964	

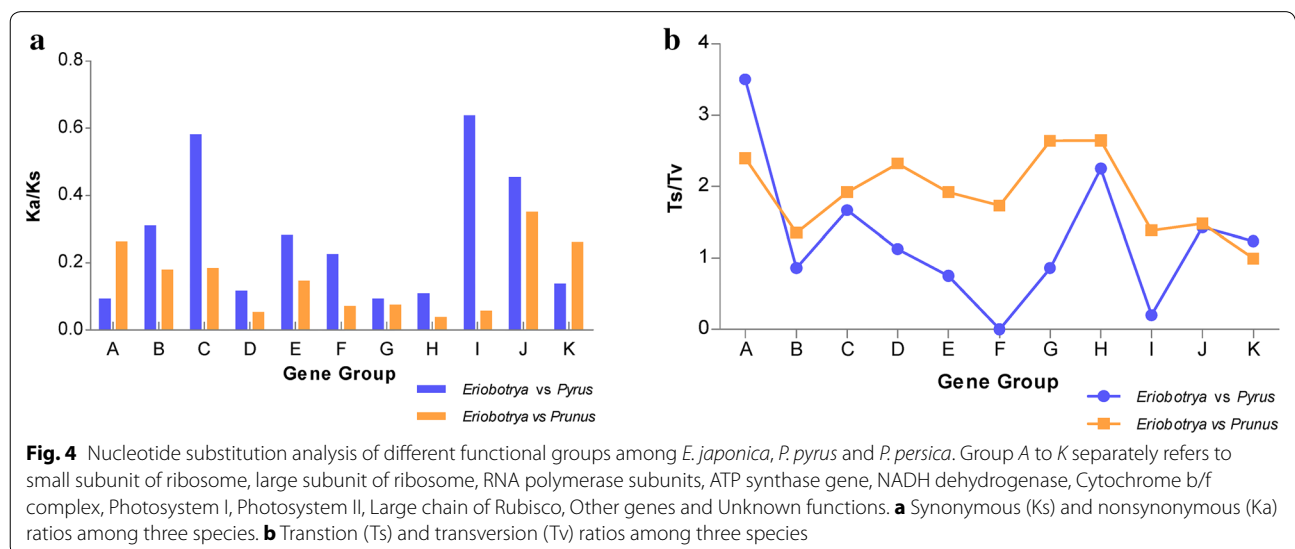


Table 6 Large indels identified among *E. japonica*, *P. pyrifolia* and *P. persica*

Type	<i>Eriobotrya</i> vs <i>Pyrus</i>			<i>Eriobotrya</i> vs <i>Prunus</i>		
	Location	Size (bp)	Repeat motifs	Location	Size (bp)	Repeat motifs
Deletion	<i>rps16-trnQ-UUG</i>	182	polyA	<i>rps16-trnQ-UUG</i>	151	
	<i>trnR-UCU-atpA</i>	417	AAT	<i>trnR-UCU-atpA</i>	40	
	<i>trnT-GGU-psbD</i>	54		<i>trnS-UGA-psbZ</i>	58	
	<i>psaA-ycf3</i>	52	polyA	<i>ndhC-trnV-UAC</i>	48	
	<i>ndhC-trnV-UAC</i>	56		<i>ndhC-trnV-UAC</i>	44	
	<i>ndhC-trnV-UAC</i>	48	polyT	<i>trnM-CAU-atpE</i>	45	TTTTG
	<i>rpl2-trnH-GUG</i>	50		<i>ccsA-ndhD</i>	50	AA
Insertion				<i>rpl2-trnH-GUG</i>	96	TA
	<i>petN-psbM</i>	79	TTCG	<i>rps16-trnQ-UUG</i>	138	
	<i>trnT-UGU-trnL-UAA</i>	42	CTCAAATATATGTTTATCAAT	<i>trnS-GCU-trnG-GCC</i>	151	
	<i>rpl32-trnL-UAG</i>	69		<i>rpoB-trnC-GCA</i>	47	
	<i>accD-psal</i>	129	AA	<i>rpoB-trnC-GCA</i>	124	
				<i>psbZ-trnG-GCC</i>	148	
				<i>rps4-trnT-UGU</i>	48	
				<i>trnT-UGU-trnL-UAA</i>	42	
				<i>trnT-UGU-trnL-UAA</i>	78	
				<i>rpl33-rps18</i>	61	TTTAT
				<i>rps18-rpl20</i>	146	
				<i>ndhF-rpl32</i>	191	AATTT
				<i>rpl32-trnL-UAG</i>	59	

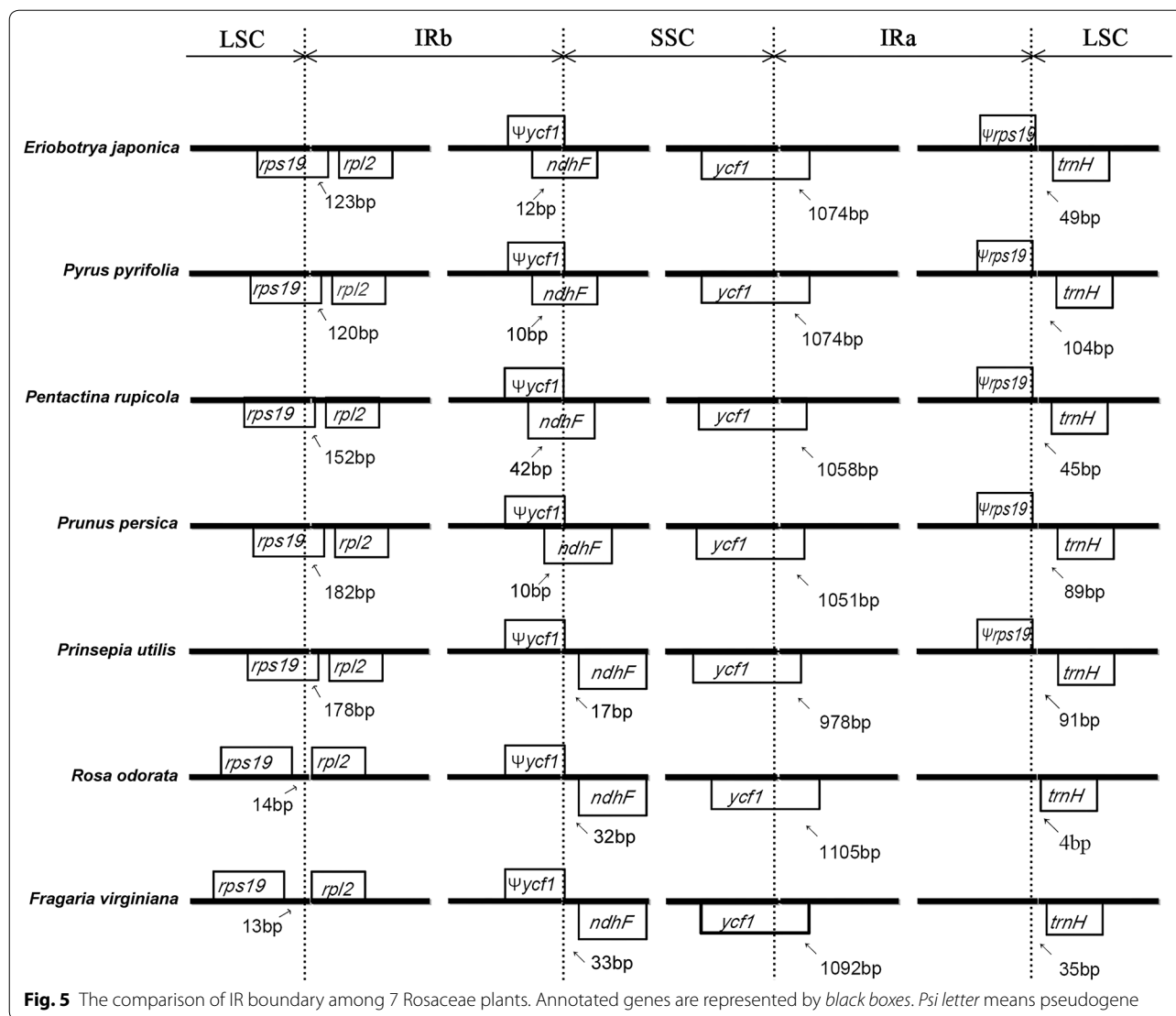
Simple sequence repeats (SSRs) have been widely used as molecular markers, which is useful for plant breeding and linkage map construction. However, due to the lack of plastid genomes in Rosaceae, few chloroplast SSRs have been utilized. In this study, a total of 72 SSRs have been found. SSRs are much more abundant in non-coding regions as they always contain more mutations than conserved coding regions (Bodin et al. 2013). Of all 72 SSRs, A and T stretches accounted for 97%, which was similar with the observation in *Pyrus pyrifolia* plastid genome where 31 A stretches, 34 T stretches, 2 C stretches and no G stretch were detected (Terakami et al. 2012). These results reveal that SSR has a strong AT bias which is consistent with many studies (Ma et al. 2013; Kuang et al. 2011; Do et al. 2013).

Large indels (>40 bp) were found through the comparison among *E. japonica*, *P. pyrifolia* and *P. persica*. The distribution of large indels revealed that IR regions as well as coding regions were more conserved than other regions as none of large indels located in IR and coding regions. Many of indels were found to be flanked by short repeated motifs or polyA or T structures (Table 6), suggesting that many large indels could be attributed to slipped-strand mispairing (Levinson and Gutman 1987; Terakami et al. 2012). Five large indels were unique in *E. japonica* plastid genome including 4 deletions and 1

insertion and were located in *rps16-trnQ-UUG*, *trnR-UCU-atpA*, *ndhC-trnV-UAC*, *rpl2-trnH-GUG* and *trnT-UGU-trnL-UAA*, respectively. *Pentactina rupicola* plastid genome was then checked for these indels and the results showed that the deletion at *ndhC-trnV-UAC* was found in *P. rupicola*, which suggested that the other four indels might be unique to *Eriobotrya* genus.

The Ka and Ks ratio has been used to denote the rate of divergence and methods of selection pressure. The value of less than, equal to or great than 1 have been considered as purifying, neutral or positive selection, respectively (Redwan et al. 2015). *Pyrus pyrifolia* revealed a lower divergence than *P. persica* from *E. japonica* as Ka and Ks between *E. japonica* and *P. pyrifolia* were much lower (Additional file 4: Table S4). Both Ka and Ks in IR regions were much lower than those in LSC and SSC (Table 5), which suggested that IR region evolved at a slower rate than LSC and SSC (Yi and Kim 2012). The Ka and Ks ratios of all gene groups were lower than 1, suggesting that a purifying selection might act on most *E. japonica* chloroplast genes.

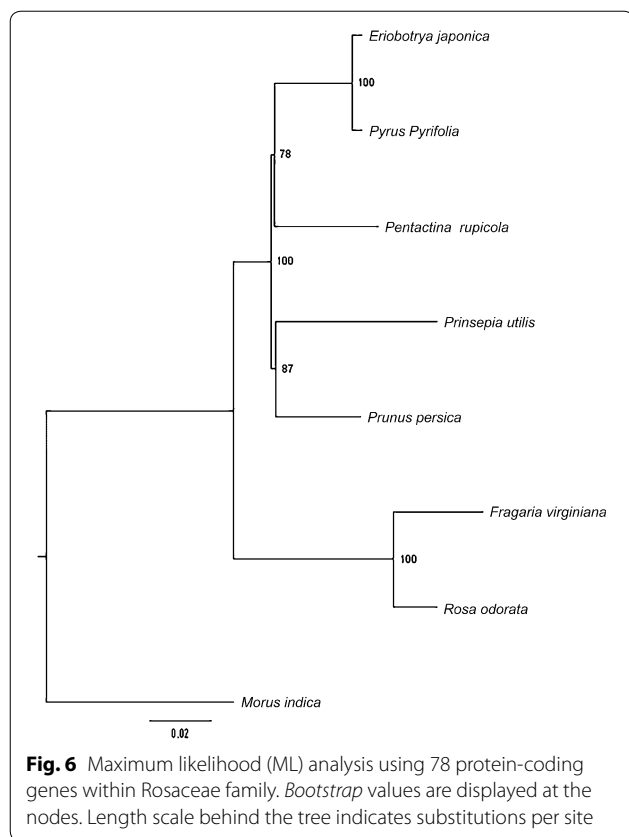
It has been documented that the higher Ts and Tv ratio denotes lower level of divergence and lower ratio denotes higher level of divergence (Yang and Yoder 1999; Dane et al. 2015). This might be a little incompatible with our results that Ts/Tv value is lower when compared with *P.*



pyrifolia since the closest relationship between *P. pyrifolia* and *E. japonica* (Table 5). From another perspective, P-distance (nucleotide substitution per site) value was found to be much smaller in *P. pyrifolia* than in *P. persica* compared to *E. japonica* (Additional file 5: Table S5). Thus, we can infer that the lower Ts and Tv ratio between *P. pyrifolia* and *E. japonica* may be caused by extremely low Ts rate, which would then be consistent with a lower divergence between *E. japonica* and *P. pyrifolia*. If both P-distance and Ts/Tv value are considered, *ycf1* with 0.068 P-distance and 0.906 Ts/Tv value are regarded as the most divergent gene of all, which could be utilized as DNA barcode for identification.

Phylogenetic analysis based on plastid genome has been rapidly developed as more and more researchers use single genes, multiple genes or different regions of

plastid genome to perform phylogenetic reconstruction (Wang et al. 2013; Walker et al. 2014; Zhang et al. 2014a). Rosaceae contains nearly a hundred genera with high economic values. Yet, taxa relationship among Rosaceae has not been investigated enough and is still vague. Su et al. (2014) performed maximum likelihood analysis within Rosids using 62 conserved chloroplast genes, which suggested but with low bootstrap that *P. persica* and *P. rupicola* formed a clade while *P. pyrifolia* and *P. utilis* formed another clade. However, another ML analysis was performed using 78 protein-coding genes (Wang et al. 2013) described that *P. persica* and *P. utilis* formed a clade while *P. pyrifolia* and *P. rupicola* formed another clade. As these two studies were incongruent, we included those four genera, along with *R. odorata*, *F. virginiana* and *E. japonica* in our phylogenetic analysis to clarify the internal



relationships within Rosaceae. Our phylogeny was similar to that of Wang et al. (2013), but with higher bootstrap values. This is also consistent with traditional studies that put *Eriobotrya* (loquat) and *Pyrus* (pear) into the subfamily Maloideae. Due to the lack of Rosaceae plastid genomes available online, in future study, more plastomes such as Spiraeae need to be involved to construct a larger-scale phylogeny within Rosaceae.

Conclusions

The complete plastid genome of *E. japonica* is 159,137 bp in length with a typical quadripartite structure. Comparison with other Rosaceae species revealed that *E. japonica* plastid genome is quite conserved in gene content and structure and showed the highest identity with *P. pyrifolia*. A total of 72 SSRs were found including two dinucleotide repeats and most of them were composed of A/T, which showed a strong A/T bias in base composition of SSRs. Four large indels appeared to be unique to *E. japonica* and could be utilized as markers for inter-generic identification. The K_a and K_s value among three species revealed differences among different genes and K_a/K_s was always less than 1, suggesting a purifying selection on these genes. Both MP and ML phylogenetic analysis shared similar topology and strongly supported

the closest relationship between *E. japonica* and *P. pyrifolia*.

Additional files

Additional file 1: Table S1. Primers designed for gaps and IR/SC boundaries.

Additional file 2: Table S2. Primers list for InDels validation of *E. japonica*.

Additional file 3: Table S3. Repeat structures found in *E. japonica* chloroplast genome.

Additional file 4: Table S4. Synonymous (K_s) and nonsynonymous (K_a) substitution rates among *E. japonica*, *P. pyrifolia* and *P. persica*.

Additional file 5: Table S5. Transitions (Ts), Transversions (Tv) and P-distance among *E. japonica*, *P. persica* and *P. pyrifolia*.

Additional file 6: Figure S1. Maximum parsimony (MP) analysis using 78 protein-coding genes within Rosaceae family. *Bootstrap* values are displayed at the nodes.

Authors' contributions

LZ, JT, ML and LS designed the experiments. LZ, QG, XL, ML and AA prepared samples and performed the experiments. LS and ML did the sequence assembly. WZ, LS, QG contributed to the analysis of data. LS, WZ and AA designed and drafted the manuscript. All authors read and approved the final manuscript.

Author details

¹ Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China. ² Changshu Qiushi Technology Co. Ltd., Changshu 215500, China.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 81274033), Science and Technology Support Program of Jiangsu Province (No. BE2014654), and Changshu Technology Support Program (No. CS201405).

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential competing interests.

Availability of data and materials

The dataset(s) supporting the conclusions of this article is(are) available in the NCBI repository, under Accession number KT633951.

Received: 21 May 2016 Accepted: 18 November 2016

Published online: 29 November 2016

References

- Aronsson H, Jarvis P (2002) A simple method for isolating import-competent Arabidopsis chloroplasts. *FEBS Lett* 529(2–3):215–220. doi:10.1016/S0014-5793(02)03342-2
- Blasco M, del Mar Naval M, Zuriaga E, Badenes ML (2014) Genetic variation and diversity among loquat accessions. *Tree Genet Genome* 10(5):1387–1398
- Bodin SS, Kim JS, Kim JH (2013) Complete chloroplast genome of *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae): comparative genomics and evaluation of universal primers for Liliales. *Plant Mol Biol Rep* 31(6):1407–1421. doi:10.1007/s11105-013-0616-x
- Cai ZQ, Penafior C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of Drimys, Liriodendron, and Piper: implications for the phylogenetic relationships of magnoliids. *BMC Evol Biol* 6:77. doi:10.1186/1471-2148-6-77

- Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ (2015) Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). PLoS ONE 10(5):0125332
- Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics. doi:10.1186/1471-2105-11-485
- Dane F, Wang Z, Goertzen L (2015) Analysis of the complete chloroplast genome of *Castanea pumila* var. *pumila*, the Allegheny chinkapin. Tree Genet Genomes 11(1):1–6
- Do HDK, Kim JS, Kim JH (2013) Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene 530(2):229–235. doi:10.1016/j.gene.2013.07.100
- 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/gkh458
- Fukuda S, Nishitani C, Hiehata N (2013) Genetic diversity of loquat accessions in Japan as assessed by SSR markers. J Jpn Soc Hortic Sci 82(2):131–137
- Gisbert AD, Romero C, Martínez-Calvo J, Leida C, Llácer G, Badenes ML (2009) Genetic diversity evaluation of a loquat (*Eriobotrya japonica* (Thunb) Lindl) germplasm collection by SSRs and S-allele fragments. Euphytica 168(1):121–134
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252(1–2):195–206. doi:10.1007/Bf02173220
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28(1):583–600. doi:10.1093/molbev/msq229
- He X (2011) DNA extraction of seven species plants of Melastomaceae using modified CTAB method. Chin J Trop Agric 10:020
- He Q, Li XW, Liang GL, Ji K, Guo QG, Yuan WM, Zhou GZ, Chen KS, van de Weg WE, Gao ZS (2011) Genetic diversity and identity of Chinese loquat cultivars/accessions (*Eriobotrya japonica*) using apple SSR markers. Plant Mol Biol Rep 29(1):197–208
- Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422(6927):72–76. doi:10.1038/nature01435
- Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA 104(49):19369–19374. doi:10.1073/pnas.0709121104
- Jansen RK, Sasaki 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(1):835–847. doi:10.1093/molbev/msq261
- Kalkman C (1988) The phylogeny of the rosaceae. Bot J Linn Soc 98(1):37–59. doi:10.1111/j.1095-8339.1988.tb01693.x
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. doi:10.1093/molbev/mst010
- Keeling PJ (2004) Diversity and evolutionary history of plastids and their hosts. Am J Bot 91(10):1481–1493. doi:10.3732/ajb.91.10.1481
- Kent WJ (2002) BLAT—The BLAST-like alignment tool. Genome Res 12(4):656–664. doi:10.1101/gr.229202
- Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome 54(8):663–673. doi:10.1139/G11-026
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29(22):4633–4642. doi:10.1093/nar/29.22.4633
- Levinson G, Gutman GA (1987) Slipped-strand mispairing—a major mechanism for DNA-sequence evolution. Mol Biol Evol 4(3):203–221
- Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452. doi:10.1093/bioinformatics/btp187
- Lohse M, Dreichsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52(5–6):267–274. doi:10.1007/s00294-007-0161-y
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964. doi:10.1093/nar/25.5.955
- Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Zhu XQ, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. doi:10.1186/2047-217x-1-18
- Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528(2):120–131
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13(3):645–658
- Nguyen PAT, Kim JS, Kim JH (2015) The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. Planta 242(1):223–237
- Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256. doi:10.1093/molbev/msn083
- Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS (2007) Phylogeny and classification of Rosaceae. Plant Syst Evol 266(1–2):5–43. doi:10.1007/s00606-007-0539-9
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genom 8(1):174
- Redwan R, Saidin A, Kumar S (2015) Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. BMC Plant Biol 15(1):1
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W (2003) Human-mouse alignments with BLASTZ. Genome Res 13(1):103–107. doi:10.1101/gr.809403
- Sharpe RH (2010) Loquat: botany and horticulture. Horticult Rev 23:233
- Soriano JM, Romero C, Vilanova S, Llacer G, Badenes ML (2005) Genetic diversity of loquat germplasm (*Eriobotrya japonica* (Thunb) Lindl) assessed by SSR markers. Genome 48(1):108–114. doi:10.1139/G04-101
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688–2690. doi:10.1093/bioinformatics/btl446
- Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH (2014) Complete chloroplast genome sequence of omani lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. PLoS ONE 9(11):113049
- Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other Methods). Version 4
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biol 56(4):564–577. doi:10.1080/10635150701472164
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739. doi:10.1093/molbev/msr121
- Tangphatsornruang S, Uthaisaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T, Jomchai N, Tragoonrun S (2011) Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. Gene 475(2):104–112. doi:10.1016/j.gene.2011.01.002
- Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, Katayama H (2012) Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. Tree Genet Genomes 8(4):841–854. doi:10.1007/s11295-012-0469-8
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106(3):411–422. doi:10.1007/s00122-002-1031-0

- 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(24):4876–4882
- Vilanova S, Badenes ML, Martinez-Calvo J, Llacer G (2001) Analysis of loquat germplasm (*Eriobotrya japonica* Lindl) by RAPD molecular markers. *Euphytica* 121(1):25–29. doi:10.1023/A:1012051207948
- Walker JF, Zanis MJ, Emery NC (2014) Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia Burkei* (Madieae, Asteraceae). *Am J Bot* 101(4):722–729. doi:10.3732/ajb.1400049
- Wang S, Shi C, Gao LZ (2013) Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of rosaceae chloroplast genomes. *PLoS ONE* 8(9):e73946. doi:10.1371/journal.pone.0073946
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255. doi:10.1093/bioinformatics/bth352
- Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. In: Salemi M (ed) *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, 2nd edn. Cambridge University Press, Cambridge, pp 615–630
- Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Mol Phylogenet Evol* 26(1):1–7
- Yang ZH, Yoder AD (1999) Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol* 48(3):274–283. doi:10.1007/Pl00006470
- Yap J-Y, Rohner T, Greenfield A, Van Der Merwe M, McPherson H, Glenn W, Kornfeld G, Marendy E, Pan AY, Wilton A (2015) Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): structure and evolution. *PLoS ONE* 10(6):e0128126
- Yi DK, Kim KJ (2012) Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* 7(7):e35872
- 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 Evol* 5(4):688–698. doi:10.1093/gbe/evt042
- Zhang Y, Li L, Yan TL, Liu Q (2014a) Complete chloroplast genome sequences of *Praxelis (Eupatorium catarium Veldkamp)*, an important invasive species. *Gene* 549(1):58–69. doi:10.1016/j.gene.2014.07.041
- Zhang YZ, Ma J, Yang BX, Li RY, Zhu W, Sun LL, Tian JK, Zhang L (2014b) The complete chloroplast genome sequence of *Taxus chinensis* var *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. *Gene* 540(2):201–209. doi:10.1016/j.gene.2014.02.037

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
