



## Original article

# Insights into plastome of *Fagonia indica* Burm.f. (Zygophyllaceae): organization, annotation and phylogeny



Mohamed S. Elshikh<sup>a</sup>, Mohammad Ajmal Ali<sup>a,\*</sup>, Fahad Al-Hemaid<sup>a</sup>, Soo Yong Kim<sup>b</sup>, Meena Elangbam<sup>c</sup>, Arun Bahadur Gurung<sup>d,\*</sup>, Prasanjit Mukherjee<sup>e</sup>, Mohamed El-Zaidy<sup>a</sup>, Joongku Lee<sup>f</sup>

<sup>a</sup> Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

<sup>b</sup> International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea

<sup>c</sup> Genetics Laboratory, Centre of Advanced Studies in Life Sciences, Manipur University, Canchipur 795 003, India

<sup>d</sup> Department of Basic Sciences and Social Sciences, North-Eastern Hill University, Shillong-793022, Meghalaya, India

<sup>e</sup> Department of Botany, Kumar Kalidas Memorial College, Pakur-816107, Jharkhand, India

<sup>f</sup> Department of Environment and Forest Resources, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

## ARTICLE INFO

## Article history:

Received 21 September 2021

Revised 3 November 2021

Accepted 4 November 2021

Available online 12 November 2021

## Keywords:

*Fagonia indica*  
Zygophyllaceae  
NGS  
Plastome  
Fabids

## ABSTRACT

The enhanced understanding of chloroplast genomics would facilitate various biotechnology applications; however, the chloroplast (cp) genome / plastome characteristics of plants like *Fagonia indica* Burm.f. (family Zygophyllaceae), which have the capability to grow in extremely hot sand desert, have been rarely understood. The *de novo* genome sequence of *F. indica* using the Illumina high-throughput sequencing technology determined 128,379 bp long cp genome, encode 115 unique coding genes. The present study added the evidence of the loss of a copy of the IR in the cp genome of the taxa capable to grow in the hot sand desert. The maximum likelihood analysis revealed two distinct sub-clades i.e. Krameriaceae and Zygophyllaceae of the order Zygophyllales, nested within fabids.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The cp (chloroplast) genome / plastome encodes several key proteins involved in the photosynthesis and in other metabolic processes important for the interactions of plants with the environment as well as defense against invading pathogens (Bobik and Burch-Smith, 2015). The availability of organelle or even whole genome sequence data in different databases repositories are gradually increasing because of the advancement of massively parallel next-generation DNA sequencing platforms and development of bioinformatics resources during last two decades; as a result, the characterization of over 5000 chloroplast (cp) genome sequences / plastome until September 2021 available in the

GenBank, have revolutionized the application of plastome genomics (Ali et al., 2020), genetic engineering to enhance plant agronomic traits (Cosa et al., 2001; Ruf et al., 2001; Dufourmant et al., 2004, 2005; Liu et al., 2007; Zhou et al., 2008; Singh et al., 2010; Lee et al., 2011; Jin et al., 2011, 2012, 2015; Zhang et al., 2015), synthesis of enzymes and biomaterials (Jin et al., 2011; Viitanen et al., 2004; Verma et al., 2010, enhancing nutrition (Shintani et al., 1998; Schneider, 2005; Apel et al., 2009; Jin et al., 2014), biopharmaceuticals (Grabowski et al., 2006; El Kaoutari et al., 2013; Kwon et al., 2013; Shenoy et al., 2014; Kohli et al., 2014; Holtz et al., 2015; Kwon et al., 2015), biomedical products (Daniell et al., 2016), and in understanding the genetic diversity, and phylogeny (Daniell et al., 2016; Brozynska et al., 2016; Jansen et al., 2007; Moore et al., 2010; Elshikh et al., 2020).

*Fagonia indica* Burm.f. (Family: Zygophyllaceae, Order: Zygophyllales, Clade: Fabids) (APG IV, 2016) is a densely to sparsely branched thorny herb approximately 60 × 100 cm in height and width respectively (Fig. 1), possess anticancer activity (Lam et al., 2014), is widely distributed in Asian and African deserts (El Hadidi, 1985; Basto, 2002; Beier et al., 2004; Beier, 2005) to the inner zone of Empty Quarter (-hottest sand desert) (Mandavil, 1986). A thorough survey of published reports revealed that the cp genomes of plants, like *F. indica*, which have the capability to

\* Corresponding authors.

E-mail addresses: [alimohammad@ksu.edu.sa](mailto:alimohammad@ksu.edu.sa) (M. Ajmal Ali), [arunbgurung@gmail.com](mailto:arunbgurung@gmail.com) (A. Bahadur Gurung).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

<https://doi.org/10.1016/j.sjbs.2021.11.011>

1319-562X/© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



**Fig. 1.** A twig of *Fagonia indica* Burm.f. (family Zygophyllaceae) in the flowering stage.

grow in extremely hot sand desert, have been rarely characterized. The present report deals the complete cp genome sequence of *F. indica*, and discuss its genome organization including gene content

and repeat features, phylogeny, and compare with the representative plants of major habitats to detect similarity and variations.

**2. Materials and methods**

**2.1. Plant material, DNA extraction and de novo genome sequencing**

The fresh leaves of *F. indica* were collected from the desert of Riyadh region, Saudi Arabia. The genomic DNA extracted using the Qiagen DNeasy Kit (Qiagen, Hilden, Germany) was subsequently used to construct short-insert libraries according to the manufacturer’s manual (Illumina, Inc., San Diego, USA), and sequenced as a single-end run of 51 bp using the DNA Illumina sequencing platform (Quail et al., 2012).

**2.2. The cp genome assembly and annotation**

The sequence raw reads were filtered using fastqc to obtain the high-quality clean sequence data by removing adaptor sequences. The high-quality filtered reads were then assembled using spades (Bankevich et al., 2012). The assembled cp genome was annotated using default parameters (Tillich et al., 2017; Stothard, 2000) of GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>), the NCBI GenBank sequence of *Larrea tridentata* (DC.) Coville (family Zygophyllaceae) (GenBank accession number NC\_028023.1) was used as a reference for annotation and for further comparison with the closely related *L. tridentata* using the mVISTA program (

**Fig. 2.** Gene map of the chloroplast (cp) genome of *Fagonia indica*.



genome.lbl.gov/vista/mvista/submit.shtml) in Shuffle-LAGAN mode (Brudno et al., 2003).

### 2.3. Repeat structure and small inversion

The tandem repeats were analyzed using the ‘Tandem Repeat Finder’ (<https://tandem.bu.edu/trf/trf.html>) (Benson 1999), (Timme et al., 2007), and REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) (Castro et al., 2013) was used to identify and locate disperse repeats including the direct (forward) and inverted (palindrome) repeats. The tandem repeats less than 15 bp in length and the REPuter redundant results were removed manually, and then the candidate small inversions (SIs) were identified when the repeats’ distance was less than 50 bp (Yang et al., 2010), and the likely secondary structures of the SIs were

evaluated using MFOLD (version 3.2) (<http://unafold.rna.albany.edu/?q=mfold>). The potential microsatellite regions were tracked by looking for five or more repeats of the nucleotides A and T using MISA (<http://misaweb.ipk-gatersleben.de/>) (Beier et al., 2017).

### 2.4. Phylogenetic analysis

A total number of 48 chloroplast genes (Supplementary Table S1) present in the cp genomes belonging to 49 taxa (Supplementary Table S2) from 49 different orders and the three outgroup sequences belonging to Gymnosperm clade were retrieved from the GenBank, and aligned using ClustalX (Thompson et al., 1997). The maximum likelihood (ML) analyses was performed using MEGAX software (Kumar et al., 2018).

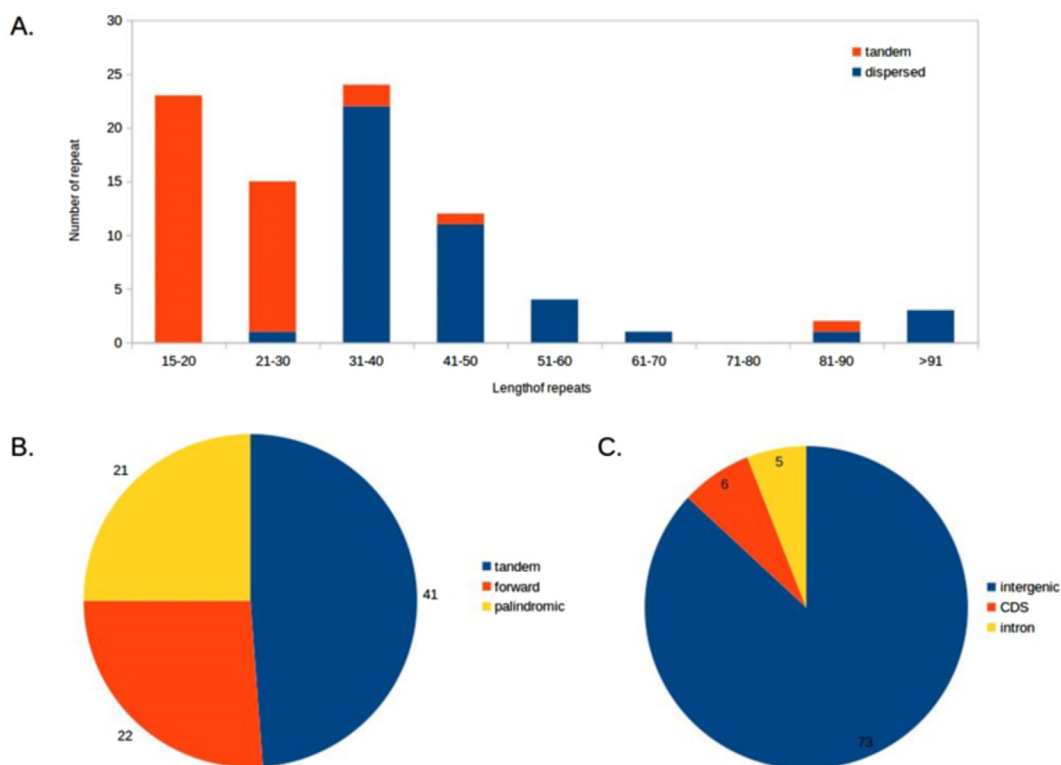


Fig. 5. Repeat structure analysis in the chloroplast (cp) genome. The cutoff value was 15 bp for a tandem repeat and 30 bp for a dispersed repeat. (A) frequency of repeats by length, (B) repeat type, (C) location distribution of repeats. (CDS: coding sequence).

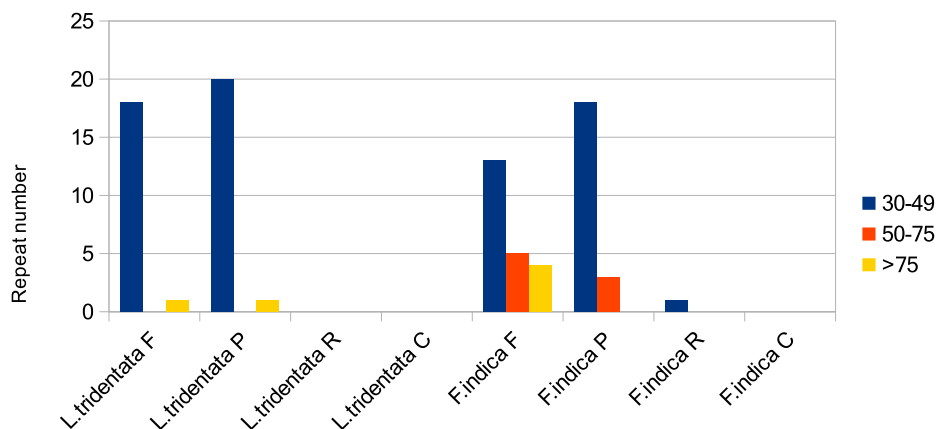


Fig. 6. Repeat structures in *Larrea tridentata* and *Fagonia indica* (Zygophyllaceae).

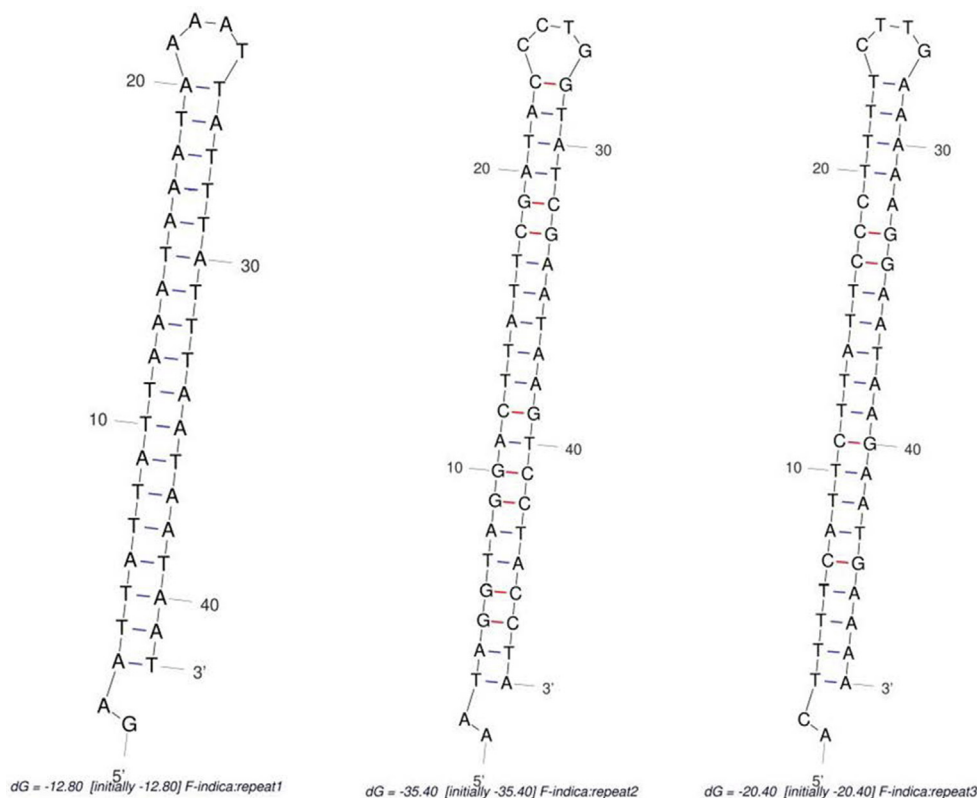


Fig. 7. Folded stem-loop structures in the three small inversions (Sis) of *Fagonia indica*.

**Table 2**  
Simple sequence repeat (SSR) loci in the chloroplast (cp) genome of *Fagonia indica*.

Position	Repeat	Repeat length of consensus	Locus	Region
285	A	10	<i>trnH-GUG-psbA</i>	intergenic
2380	T	11	<i>matK</i>	CDS
17,585	A	14	<i>trnS-GGA-ycf3</i>	intergenic
17,960	A	10	<i>ycf3</i>	intron
18,892	A	12	<i>ycf3</i>	intron
26,911	T	12	<i>psbZ-trnS-UGA</i>	intergenic
31,068	AT	6	<i>psbD-trnT-GGU</i>	intergenic
31,686	T	12	<i>trnT-GGU-trnE-UUC</i>	intergenic
32,350	A	11	<i>trnY-GUA-trnD-GUC</i>	intergenic
36,011	A	10	<i>trnC-GCA-rpoB</i>	intergenic
36,458	A	10	<i>trnC-GCA-rpoB</i>	intergenic
40,452	A	10	<i>rpoC1</i>	intron
44,927	A	13	<i>rpoC2</i>	CDS
46,844	A	11	<i>rpoC2</i>	CDS
49,092	A	11	<i>atpI-atpH</i>	intergenic
52,145	T	10	<i>atpF-atpA</i>	intergenic
58,939	T	11	<i>accD-psal</i>	intergenic
59,528	T	10	<i>accD-psal</i>	intergenic
60,008	A	10	<i>psal-ycf4</i>	intergenic
64,326	AT	6	<i>psbJ-psbL</i>	intergenic
66,729	A	11	<i>trnP-UGG-psaJ</i>	intergenic
68,792	T	11	<i>rps18-rpl20</i>	intergenic
69,919	TA	6	<i>rpl20-clpP</i>	intergenic
70,918	T	10	<i>clpP</i>	intron
76,067	T	11	<i>petB-petD</i>	intergenic
81,388	A	13	<i>rpl14-rpl16</i>	intergenic
110,755	T	10	<i>ycf1</i>	CDS
110,875	A	10	<i>ycf1</i>	CDS
111,474	A	10	<i>ycf1</i>	CDS
113,104	A	11	<i>ycf1</i>	CDS
114,482	TA	10	<i>ycf1-rps15</i>	intergenic
114,945	A	11	<i>rps15</i>	CDS
119,431	AT	6	<i>ndhI-ndhG</i>	intergenic
123,288	TA	7	<i>ndhD-ccsA</i>	intergenic
125,155	T	11	<i>rpl32</i>	CDS
125,883	T	10	<i>rpl32-ndhF</i>	intergenic
127,343	A	10	<i>ndhF</i>	CDS



### 3. Results

#### 3.1. Content and organization of plastome

The mapping of the assembled cp genome resulted into circular molecule (Fig. 2) with a total number of 115 unique genes [representing 1,28,379 base pair (bp) nucleotides (nt)] which includes 80 CDS (represents 80,200 bp nucleotides coding for 42,793 codons), 31 tRNA genes and four rRNA genes. The assembled cp genome sequence was submitted to the NCBI (GenBank accession number MN521457). The cp genome size of *F. indica* was approximately 128 kb, which was smaller than that of *L. tridentata* (Fig. 3, Table 1).

sents 1,28,379 base pair (bp) nucleotides (nt)] which includes 80 CDS (represents 80,200 bp nucleotides coding for 42,793 codons), 31 tRNA genes and four rRNA genes. The assembled cp genome sequence was submitted to the NCBI (GenBank accession number MN521457). The cp genome size of *F. indica* was approximately 128 kb, which was smaller than that of *L. tridentata* (Fig. 3, Table 1).

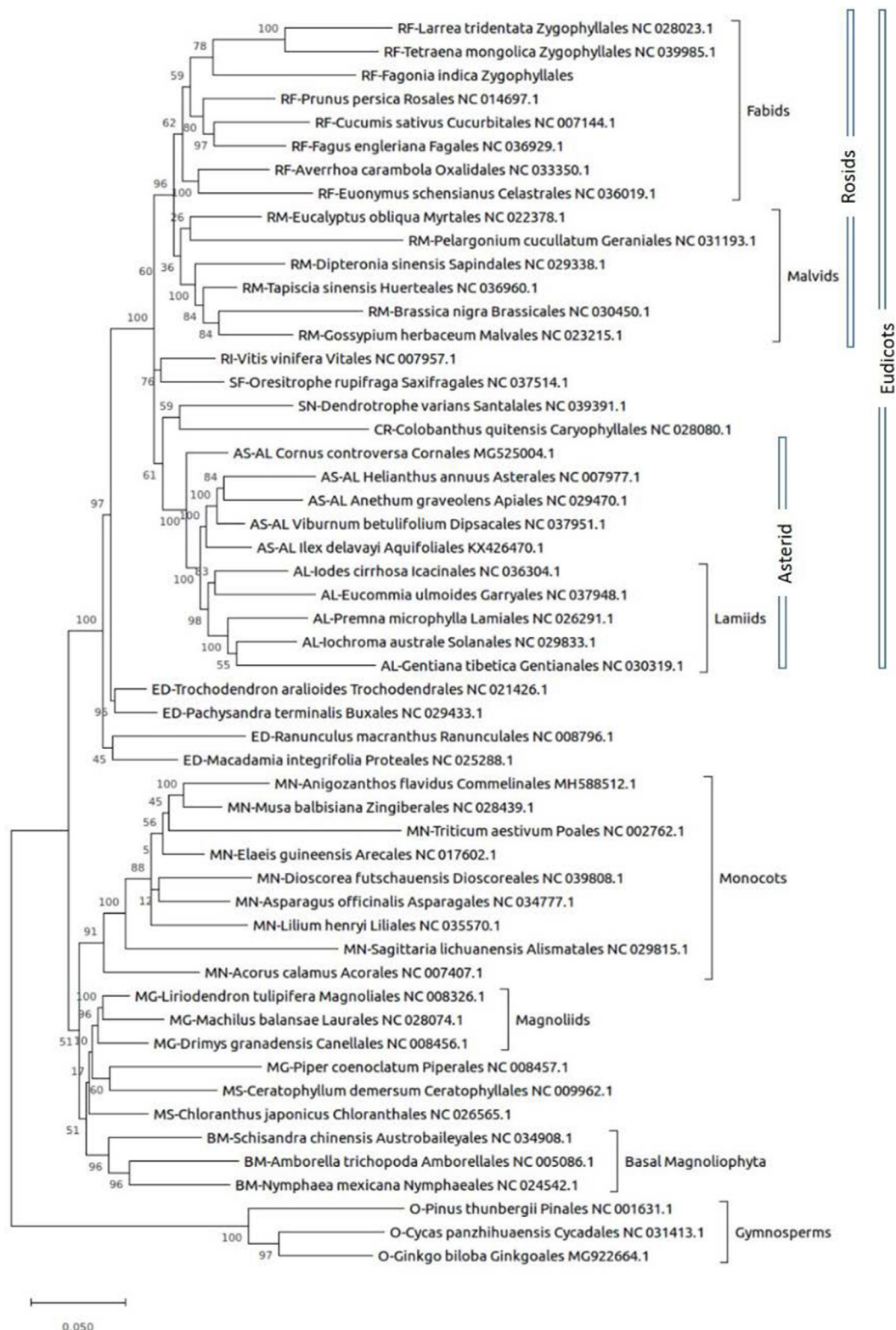


Fig. 8. Maximum likelihood (ML) phylogenetic tree inferred from 48 protein-coding chloroplast (cp) genes from 53 plant taxa.

The coding regions were less divergent than the non-coding regions (Fig. 4).

A total of 13 genes, including seven protein-coding genes and six tRNA genes, contained one or two introns (Supplementary Table S3). Among the intron-containing genes, *trnK-UUU* had the largest intron (2511 bp) that includes the *matK* gene, and *trnL-UAA* had the smallest intron (551 bp). The *ycf3* gene had two introns of 722 and 758 bp. The sequence analysis indicates that 58.49%, 6.87%, and 3.53% of the genome sequence encode proteins, tRNAs, and rRNAs, respectively, whereas 41.50% of the genome sequence is a non-coding sequence filled with introns, intergenic spacers, and pseudogenes. Based on the sequences of protein-coding and tRNA genes within the cp genome, Phe (0.05%) and Arg (0.0038%) were the most and least used amino acids, respectively (Supplementary Table S4).

The tandem and dispersed repeats were analyzed in the cp genome of *F. indica*. Forty-one tandem repeats were identified, of which 23 were 15–20 bp, 14 were 21–30 bp, two were 31–40 bp, one was 41–50 bp, and another one was 81–90 bp in size. Similarly, 43 dispersed repeats were identified, of which one was 21–30 bp, 22 were 31–40 bp, 11 were 41–50 bp, four were 51–60 bp, one was 61–70 bp, another one was 81–90 bp, and three were more than 91 bp in size. In total, 84 repeats were identified, of which 87% were in the intergenic spacer regions, 6% in introns, and 7% in the CDS regions, respectively (Fig. 5). The repeat structures in other members of Zygothylaceae (*L. tridentata*) were also analyzed using REPuter (Fig. 6). The forward and inverted repeats were common in *L. tridentata* and *F. indica*. In addition, in the same Zygothylaceae family, different repeat structures were found between *F. indica* and *L. tridentata*. Of the two Zygothylaceae cp genomes studied, *F. indica* contained the highest total number of repeats that were 75 bp or greater in length and SIs ranging from 11 to 24 bp in size. The folded stem-loop structures of the three SIs of *F. indica* are shown in Fig. 7.

Within the cp genome of *F. indica*, 37 different SSR loci were repeated more than five times (Table 2). Of these, 31 loci were homopolymers and six were di-polymers. All homopolymeric loci contained multiple A or T nucleotides, whereas all di-polymeric loci contained multiple AT or TA nucleotides. These SSR loci contribute to the A-T richness of the cp genome of *F. indica*.

The present maximum likelihood (ML) bootstrap analysis revealed two major clades—monocots and eudicots. In the eudicots clade, *F. indica* clades with *L. tridentata* and *T. mongolica* (family Zygothylaceae, order Zygothylales) nested within the clade fabids. The maximum likelihood tree (MLT) also revealed two distinct clades of Krameriaceae and Zygothylaceae (Fig. 8).

#### 4. Discussion

In the present study, the mapping of the assembled cp genome was found similar to the angiosperm (Raubeson et al., 2007), except for the loss of one copy of the IR as similar to majority of papilionoid (Doyle et al., 1996; Kato et al., 2000; Sasaki et al., 2005; Guo et al., 2007). The *rps16* gene was found in the cp genomes of most angiosperms, including the representatives of the early-branching lineages (Goremykin et al., 2003; Raubeson et al., 2007; Hansen et al., 2007); however, it was not found in the *F. indica*.

*F. indica* had a single copy of inverted repeat resulted into the inverted gene order compared to its taxonomically close relative *L. tridentata*. The lengths of the cp genomes of angiosperms remain variable primarily because of nucleotide substitutions, gene/intron losses, and expansion and contraction of the inverted repeat IR region (Jansen et al., 2007). It was noted that the coding regions were less divergent than the non-coding regions (Fig. 4); however,

further analysis showed that *clpP* and *accD* were the most divergent coding regions (Supplementary Table S5). Photosynthesis is the ultimate source of biomass production (Beadle and Long, 1985). The PAR (photosynthetically active radiation) intensity is an important factor that determines the rate of photosynthesis (Wimalasekera, 2019). The intensity of light varies in different major habitats (Warrant and Johnsen, 2013). The comparative cp genome analysis of *F. indica* as a representative from hot sand desert with the representatives of flowering plants occurring in different major habitats further supports the conservative pattern of the cp genome and suggests that the genes contained in the cp genome might not have roles solely in organism yield, rarity, or abundance and biomass, and in encountering stress (Elshikh et al., 2020).

The knowledge of phylogeny is used in almost every branch of biology (Yang and Rannala, 2012) including taxonomy (Philippe et al., 2005; APG IV, 2016), evolution (Edwards, 2009; Soltis et al., 2019) and comparative biology (Eisen, 1998; Mäser et al., 2001; Kellis et al., 2003; Pedersen et al., 2006; Lindblad et al., 2011), medicine (Marra et al., 2003; Grenfell et al., 2004; Salipante and Horwitz, 2006), and genomics (Paten et al., 2008; Green et al., 2010; Gronau et al., 2011; Li and Durbin, 2011; Ma, 2011). Moreover, the family Zygothylaceae has previously been treated as being related either to Geraniaceae (Geraniales) or to Sapindales/Rutales or Linales/Malpighiales (Sheahan and Chase, 1996). Secondly, the phylogenetic relationships of the two sister families, e.g., Zygothylaceae and Krameriaceae (Soltis et al., 1998; Savolainen et al., 2000; Wang et al., 2009; Tao et al., 2018) under the order Zygothylales, have often been controversial APG IV, 2016. The wood anatomy supports the separation of Krameriaceae from the Zygothylaceae (Carlquist, 2005). Granot and Grafi (2014) argued, based on epigenetic studies, that the placement of the families Krameriaceae and Zygothylaceae under the order Zygothylales should be re-examined. The present maximum likelihood (ML) bootstrap analysis revealed two major clades—monocots and eudicots. In the eudicots clade, *F. indica* clades with *L. tridentata* and *T. mongolica* (family Zygothylaceae, order Zygothylales) nested within the clade fabids. The maximum likelihood tree (MLT) also revealed two distinct clades of Krameriaceae and Zygothylaceae.

#### 5. Conclusions

The analyses of *de novo* genome sequence of *F. indica* (family Zygothylaceae) have added the evidence of the loss of a copy of the IR in the cp genome of the taxa capable to grow in the hot sand desert. The maximum likelihood analysis revealed two distinct sub-clades i.e. Krameriaceae and Zygothylaceae of the order Zygothylales.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The authors would like to extend their sincere appreciation to the Researchers Supporting Project number (RSP-2021/306), King Saud University, Riyadh, Saudi Arabia. This research was supported by the grant from the KRIBB Initiative Program of the Republic of Korea.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sjbs.2021.11.011>.

## References

- Ali, M.A., Kim, S.Y., Pan, T.K., Choi, S., Gurung, A.B., Farah, M.A., Alhemaïd, F., Alanazi, K.M., Elangbam, M., Lee, J., 2020. Complete chloroplast genome sequencing of *Saraca asoca*: cp genome characterization, comparison with Detarioideae and molecular authentication. *SJBS* (comm.)
- Apel, W., Bock, R., 2009. Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin a conversion. *Plant Physiol.* 151 (1), 59–66.
- APG, I.V., 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20.
- Bankevich, Anton, Nurk, Sergey, Antipov, Dmitry, Gurevich, Alexey A., Dvorkin, Mikhail, Kulikov, Alexander S., Lesin, Valery M., Nikolenko, Sergey I., Pham, Son, Prjibelski, Andrey D., Pyshkin, Alexey V., Sirotkin, Alexander V., Vyahhi, Nikolay, Tesler, Glenn, Alekseyev, Max A., Pevzner, Pavel A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (5), 455–477.
- Basto, M.F.P., 2002. *Zygophyllaceae*. in: Paiva, J., Gomes, I., Gomes, S. (Eds.), *Flora de Cabo Verde* 63. Centro de Botanica do Instituto de Investigação Científica Tropical, Lisboa-Praia, pp. 8–15.
- Beadle, C.L., Long, S.P., 1985. Photosynthesis—is it limiting to biomass production? *Biomass.* 8 (2), 119–168.
- Beier, Björn-Axel, 2005. A revision of the desert shrub *Fagonia* (Zygophyllaceae). *Systemat. Biodiver.* 3 (3), 221–263.
- Beier, B.-A., Nylander, J.A.A., Chase, M.W., Thulin, M., 2004. Phylogenetic relationships and biogeography of the desert plant genus *Fagonia* (Zygophyllaceae), inferred by parsimony and Bayesian model averaging. *Mol. Phylogenet. Evol.* 33 (1), 91–108.
- Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nuc. Acids Res.* 27 (2), 573–580.
- Bobik, K., Burch-Smith, T.M., 2015. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* 6, 781.
- Brozynska, Marta, Furtado, Agnelo, Henry, Robert J., 2016. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14 (4), 1070–1085.
- Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., Batzoglou, S., 2003. Global alignment: finding rearrangements during alignment. *Bioinformatics* 19S1, i54–i62.
- Carlquist, S., 2005. Wood anatomy of Krameriaceae with comparisons with Zygophyllaceae: phyletic, ecology and systematics. *Bot. J. Linn. Soc.* 149, 257–270.
- Castro, I., Pinto-Carnide, O., Ortiz, I.M., Martín, J.P., 2013. Chloroplast genome diversity in Portuguese grapevine (*Vitis vinifera* L.) cultivars. *Mol. Biotechnol.* 54, 528–540.
- Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134.
- Cosa, Brandy De, Moar, William, Lee, Seung-Bum, Miller, Michael, Daniell, Henry, 2001. Overexpression of the Bt cry 2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* 19 (1), 71–74.
- Doyle, Jeff J., Doyle, Jane L., Ballenger, J.A., Palmer, J.D., 1996. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol. Phylogenet. Evol.* 5 (2), 429–438.
- Dufourmantel, Nathalie, Pelissier, Bernard, Garçon, Frederic, Peltier, Gilles, Ferullo, Jean-Marc, Tissot, Ghislaine, 2004. Generation of fertile transplastomic soybean. *Plant Mol. Biol.* 55 (4), 479–489.
- Dufourmantel, N., Tissot, G., Goutorbe, F., Garçon, F., Muhr, C., Jansens, S., et al., 2005. Generation and analysis of soybean plastid transformants expressing *Bacillus thuringiensis* Cry1Ab protoxin. *Plant Mol. Biol.* 58, 659–668.
- Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? *Evol. Int. J. Org. Evol.* 63, 1–19.
- Eisen, J.A., 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8, 163–167.
- El Hadidi, N.M., 1985. *Zygophyllaceae*. In: Polhill, R.M. (Ed.), *Flora of Tropical East Africa*. Balkema, Rotterdam, p. 7.
- Kaoutari, Abdessamad El, Armougom, Fabrice, Gordon, Jeffrey L., Raoult, Didier, Henrissat, Bernard, 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* 11 (7), 497–504.
- Elshikh, Mohamed Soliman, Kim, Soo-Yong, Ali, Mohammad Ajmal, Al-Hemaid, Fahad, Chen, Shen-Ming, Choi, Sangho, Rahman, Mohammad Oliur, Elangbam, Meena, Lee, Joongku, 2020. Comparative analysis of cp genome of *Fagonia indica* growing in desert and its implications in pattern of similarity and variations. *Saudi J. Biol. Sci.* 27 (1), 229–232.
- Goremykin, V.V., Hirsch-Ernst, K.L., Wölfel, S., Hellwig, F.H., 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* 20, 1499–1505.
- Grabowski, Henry, Cockburn, Iain, Long, Genia, 2006. The market for follow-on biologics: how will it evolve? *Health Aff.* 25 (5), 1291–1301.
- Granot, Gila, Grafi, Gideon, 2014. Epigenetic information can reveal phylogenetic relationships within Zygophyllales. *Plant Syst. Evol.* 300 (8), 1819–1824.
- Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., et al., 2010. A draft sequence of the Neandertal genome. *Science* 328 (5979), 710–722.
- Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A., et al., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303 (5656), 327–332.
- Gronau, Ilan, Hubisz, Melissa J., Gulko, Brad, Danko, Charles G., Siepel, Adam, 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* 43 (10), 1031–1034.
- Guo, Xianwu, Castillo-Ramírez, Santiago, González, Víctor, Bustos, Patricia, Luis Fernández-Vázquez, José, Santamaría, Rosa, Arellano, Jesús, Cevallos, Miguel A, Dávila, Guillermo, 2007. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genom.* 8 (1), 228. <https://doi.org/10.1186/1471-2164-8-228>.
- Hansen, Debra R., Dastidar, Sayantani G., Cai, Zhengqiu, Penafior, Cynthia, Kuehl, Jennifer V., Boore, Jeffrey L., Jansen, Robert K., 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol. Phylogenet. Evol.* 45 (2), 547–563.
- Holtz, Barry R., Berquist, Brian R., Bennett, Lindsay D., Kommineni, Vally J.M., Muniguntti, Ranjith K., White, Earl L., Wilkerson, Don C., Wong, Kah-Yat I., Ly, Lan H., Marcel, Sylvain, 2015. Commercial-scale biopharmaceuticals manufacturing facility for plant-made pharmaceuticals. *Plant Biotechnol. J.* 13 (8), 1180–1190.
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., et al., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA.* 104, 19369–19374.
- Jin, Shuangxia, Daniell, Henry, 2014. Expression of  $\delta$ -tocopherol methyltransferase in chloroplasts results in massive proliferation of the inner envelope membrane and decreases susceptibility to salt and metal-induced oxidative stresses by reducing reactive oxygen species. *Plant Biotechnol. J.* 12 (9), 1274–1285.
- Jin, Shuangxia, Kanagaraj, Anderson, Verma, Dheeraj, Lange, Theo, Daniell, Henry, 2011. Release of hormones from conjugates: chloroplast expression of  $\beta$ -glucosidase results in elevated phytohormone levels associated with significant increase in biomass and protection from aphids or whiteflies conferred by sucrose esters. *Plant Physiol.* 155 (1), 222–235.
- Jin, Shuangxia, Singh, Nameirakpam D., Li, Lebin, Zhang, Xianlong, Daniell, Henry, 2015. Engineered chloroplast dsRNA silences cytochrome p450 monooxygenase, V-ATPase and chitin synthase genes in the insect gut and disrupts *Helicoverpa armigera* larval development and pupation. *Plant Biotechnol. J.* 13 (3), 435–446.
- Jin, S., Zhang, X., Daniell, H., 2012. *Pinellia ternata* agglutinin expression in chloroplasts confers broad spectrum resistance against aphid, whitefly, lepidopteran insects, bacterial and viral pathogens. *Plant Biotechnol. J.* 10, 313–327.
- Kato, T., Kaneko, T., Sato, S., Nakamura, Y., Tabata, S., 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res.* 7, 323–330.
- Kellis, Manolis, Patterson, Nick, Endrizzi, Matthew, Birren, Bruce, Lander, Eric S., 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423 (6937), 241–254.
- Kohli, Neha, Westerveld, Donevan R, Ayache, Alexandra C, Verma, Amrisha, Shil, Pollob, Prasad, Tuhina, Zhu, Ping, Chan, Sic L, Li, Qiuqiong, Daniell, Henry, 2014. Oral delivery of bioencapsulated proteins across blood-brain and blood-retinal barriers. *Mol. Ther.* 22 (3), 535–546.
- Kumar, S., Stecher, G., Li, M., Nkayac, C., Tamara, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
- Kwon, K.-C., Daniell, H., 2015. Low-cost oral delivery of protein drugs bioencapsulated in plant cells. *Plant Biotechnol. J.* 13, 1017.
- Kwon, Kwang-Chul, Nityanandam, Ramya, New, James S., Daniell, Henry, 2013. Oral delivery of bioencapsulated extendin-4 expressed in chloroplasts lowers blood glucose level in mice and stimulates insulin secretion in beta-TC 6 cells. *Plant Biotechnol. J.* 11 (1), 77–86.
- Lam, M., Wolff, K., Griffiths, H., Carmichael, A., 2014. Correction: an aqueous extract of *Fagonia cretica* induces DNA damage, cell cycle arrest and apoptosis in breast cancer cells via FOXO3a and p53 expression. *PLoS One* 9.
- Lee, S.-B., Li, B., Jin, S., Daniell, H., 2011. Expression and characterization of antimicrobial peptides Retrocyclin-101 and Protegrin-1 in chloroplasts to control viral and bacterial infections. *Plant Biotechnol. J.* 9, 100–115.
- Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., et al., 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482.
- Liu, C.-W., Lin, C.-C., Chen, J.J.W., Tseng, M.-J., 2007. Stable chloroplast transformation in cabbage (*Brassica oleracea* L. var. *capitata* L.) by particle bombardment. *Plant Cell Rep.* 26, 1733–1744.
- Ma, Jian, 2011. Reconstructing the history of large-scale genomic changes: biological questions and computational challenges. *J. Comput. Biol.* 18 (7), 879–893.



- Mandaville, James P., 1986. Plant life in the Rub' al-Khali (the Empty Quarter), southcentral Arabia. *Proc. R. Soc. Edinb.* 89, 147–157.
- Marra, M.A., Jones, S.J.M., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S.N., et al., 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300 (5624), 1399–1404.
- Mäser, P., Thomine, S., Schroeder, J.I., Ward, J.M., Hirschi, K., Sze, H., et al., 2001. Phylogenetic relationships within cation transporter families of *Arabidopsis*. *Plant Physiol.* 126, 1646–1667.
- Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. USA.* 107, 4623–4628.
- Paton, B., Herrero, J., Fitzgerald, S., Beal, K., Flicek, P., Holmes, I., Birney, E., 2008. Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* 18 (11), 1829–1843.
- Pedersen, Jakob Skou, Bejerano, Gill, Siepel, Adam, Rosenbloom, Kate, Lindblad-Toh, Kerstin, Lander, Eric S., Kent, Jim, Miller, Webb, Haussler, David, Durbin, Richard, 2006. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol.* 2 (4).
- Philippe, Hervé, Delsuc, Frédéric, Brinkmann, Henner, Lartillot, Nicolas, 2005. *Phylogenomics. Annu. Rev. Ecol. Evol. Syst.* 36 (1), 541–562.
- Quail, Michael, Smith, Miriam E, Coupland, Paul, Otto, Thomas D, Harris, Simon R, Connor, Thomas R, Bertoni, Anna, Swerdlow, Harold P, Gu, Yong, 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 13 (1), 341. <https://doi.org/10.1186/1471-2164-13-341>.
- Raubeson, Linda A, Peery, Rhiannon, Chumley, Timothy W, Dziubek, Chris, Fourcade, H Matthew, Boore, Jeffrey L, Jansen, Robert K, 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 8 (1), 174. <https://doi.org/10.1186/1471-2164-8-174>.
- Ruf, Stephanie, Hermann, Marita, Berger, Irving J., Carrer, Helaine, Bock, Ralph, 2001. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* 19 (9), 870–875.
- Salipante, S.J., Horwitz, M.S., 2006. Phylogenetic fate mapping. *Proc. Natl. Acad. Sci.* 103, 5448–5453.
- Saski, Christopher, Lee, Seung-Bum, Daniell, Henry, Wood, Todd C., Tomkins, Jeffrey, Kim, Hyi-Gyung, Jansen, Robert K., 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* 59 (2), 309–322.
- Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., et al., 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences. *Syst. Biol.* 49, 306–362.
- Schneider, C., 2005. Chemistry and biology of vitamin E. *Mol. Nutr. Food Res.* 49, 7–30.
- Sheahan, M.C., Chase, M.W., 1996. A phylogenetic analysis of Zygophyllaceae R. Br. based on morphological, anatomical and *rbc L* DNA sequence data. *Bot. J. Linn. Soc.* 122, 279–300.
- Shenoy, V., Kwon, K.-C., Rathinasabapathy, A., Lin, S., Jin, G., Song, C., et al., 2014. Oral delivery of Angiotensin-converting enzyme 2 and Angiotensin-(1-7) bioencapsulated in plant cells attenuates pulmonary hypertension. *Hypertension.* 64, 1248–1259.
- Shintani, David, DellaPenna, Dean, 1998. Elevating the vitamin E content of plants through metabolic engineering. *Science* 282 (5396), 2098–2100.
- Singh, A.K., Verma, S.S., Bansal, K.C., 2010. Plastid transformation in eggplant (*Solanum melongena* L.). *Transgenic Res.* 19, 113–119.
- Soltis, D.E., Soltis, P.S., Mort, M.E., Chase, M.W., Savolainen, V., Hoot, S.B., et al., 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Syst. Biol.* 47, 32–42.
- Soltis, P.S., Folk, R.A., Soltis, D.E., 2019. Darwin review: angiosperm phylogeny and evolutionary radiations. *Proc. R. Soc. B.* 286, 20190099.
- Stothard, Paul, 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28 (6), 1102–1104.
- Tao, Zhi-Bin, Wortley, Alexandra H., Lu, Lu, Li, De-Zhu, Wang, Hong, Blackmore, Stephen, 2018. Blackmore, S., Evolution of Angiosperm Pollen. 6. The Celastrales, Oxalidales, and Malpighiales (Com) Clade and Zygophyllales. *Ann. Missouri Bot. Gard.* 103 (3), 393–442.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nuc. Acids Res.* 25, 4876–4882.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, Timme, R.E., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* 94, 302–312.
- Verma, D., Kanagaraj, A., Jin, S.X., Singh, N.D., Kolattukudy, P.E., Daniell, H., 2010. Chloroplast-derived enzyme cocktails hydrolyse lignocellulosic biomass and release fermentable sugars. *Plant Biotechnol. J.* 8, 332–350.
- Viiitanen, P.V., Devine, A.L., Khan, M.S., Deuel, D.L., Van Dyk, D.E., Daniell, H., 2004. Metabolic engineering of the chloroplast genome using the *Echerichia coli* *ubiC* gene reveals that chorismate is a readily abundant plant precursor for phydroxybenzoic acid biosynthesis. *Plant Physiol.* 136, 4048–4060.
- Wang, H., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., et al., 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci.* 106, 3853–3858.
- Warrant, Eric J., Johnsen, Sönke, 2013. Vision and the light environment. *Curr. Biol.* 23 (22), R990–R994.
- Wimalasekera, R., 2019. Effect of light intensity on photosynthesis. in: *Photosynthesis, Productivity and Environmental Stress*. John Wiley & Sons, Ltd, pp. 65–73.
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al., 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 5.
- Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13, 303–314.
- Zhang, Jiang, Khan, Sher Afzal, Hasse, Claudia, Ruf, Stephanie, Heckel, David G., Bock, Ralph, 2015. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science* 347 (6225), 991–994.
- Zhou, F., Badillo-Corona, J.A., Karcher, D., Gonzalez-Rabade, N., Piepenburg, K., Borchers, A.-M.I., et al., 2008. High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. *Plant Biotechnol. J.* 6, 897–913.