Saudi Journal of Biological Sciences 27 (2020) 229–232

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Comparative analysis of cp genome of *Fagonia indica* growing in desert and its implications in pattern of similarity and variations

Mohamed Soliman Elshikh [a], Soo-Yong Kim [b,*], Mohammad Ajmal Ali [a,*], Fahad Al-Hemaid [a], Shen-Ming Chen [c], Sangho Choi [b], Mohammad Oliur Rahman [d], Meena Elangbam [e], Joongku Lee [f,*]

[a] Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia
[b] International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea
[c] Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taipei 106, Taiwan, ROC
[d] Department of Botany, University of Dhaka, Dhaka 1000, Bangladesh
[e] Genetics Laboratory, Centre of Advanced Studies in Life Sciences, Manipur University, Canchipur 795 003, India
[f] Department of Environment and Forest Resources, Chungnam National University, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

ARTICLE INFO

Article history:
Received 2 July 2019
Revised 1 August 2019
Accepted 25 August 2019
Available online 31 August 2019

Keywords:
cp DNA
Fagonia indica
Zygophyllaceae
Photosynthesis
Habitats
Biomass
Crop

ABSTRACT

The chloroplasts genome encodes several key proteins that involves in the process of the photosynthesis and also in other metabolic processes important for growth and development, yield, biomass, and plant interactions with their environment. The present study aimed to sequencing of cp genome of *Fagonia indica* Burm.f (Zygophyllaceae), -a plant that occurs even in the hot desert condition of the inner zone of Rub′ al-Khali (the Empty Quarter) of south-central Arabia, and its comparative analyses with the representative of the sequence of the different categories [viz. (a) with the other member of the family Zygophyllaceae, and with the representatives from: (b) different clade of the angiosperms, (c) flowering plants occurs in different major habitats, (d) different groups of plants, (e) different group of plants having range of biomass, (f) C3 and C4 plants, and (g) the representative from very common, rare and major high yielding crop of the world] to unravel the genetic pattern of similarity and variations. The comparison of *F. indica* genome in different categories showed strong evidence and further support for the conservative pattern of chloroplast genome, the coding and non-coding region remains conserved even in phylogenetically distant eukaryotic clades, and might not have the sole roles in organism′s yield, rarity or abundance and biomass, and in encountering the stress. Nevertheless, the result could be useful for molecular phylogenetic and molecular ecological and molecular mechanism of photosynthesis.

1. Introduction

Fagonia indica Burm.f. [family Zygophyllaceae; included in the Fabids (APG IV, 2016)] is a thorny medicinal herb possess anticancer activity (Lam et al., 2014), growing widely in Asian and African deserts (Beier, 2005), and able to grow even in the hot desert condition of the inner zone of Rub' al-Khali (the Empty Quarter) of south-central Arabia where annual precipitation generally less than 35 mm or sometimes very less or no rain for several years (Mandavil, 1986). Since the enhanced understanding of chloroplast biology came with the recent advances in biotechnology and bioinformatics during last two decades (Shendure and Ji, 2008; Shendure et al., 2017) have impact on various biotechnological applications including genetic engineering to enhance plant agronomic traits (Shintani et al., 1998; Viitanen et al., 2004; Apel and Bock, 2009; Verma et al., 2010; Kwon et al., 2013, 2015; Jin et al., 2011, 2014; Holtz et al., 2015; Brozynska et al., 2016; Daniell et al., 2016). The chloroplasts genome encodes several key proteins that involves in the process of the photosynthesis and in other metabolic processes important for the growth and development, and plant interactions with their environment like responses to heat, drought, salt, light, etc., (Bobik and Burch-Smith, 2015). The dynamics of genetic diversity have fascinated to naturalists for centuries, the present study aimed to comparative analyses of cp

* Corresponding authors.
E-mail addresses: soodole@kribb.re.kr, sooyongkimkribb@gmail.com (S.-Y. Kim), ajmalpdrc@gmail.com, alimohammad@ksu.edu.sa (M.A. Ali), joongku@cnu.ac.kr (J. Lee).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2019.08.016
1319-562X/

genome to unravel the genetic pattern of similarity and variations in context with organism's yield, rarity or abundance and biomass, and in encountering the stress.

2. Materials and methods

2.1. cp genome sequencing

The leaf material of *F. indica* was collected, and stored at 4 °C for overnight. The modified Percoll gradient buffer method (Kim and Kim, 2013) was then applied to isolate the chloroplasts. The DNA was extracted from the isolated chloroplast using DNeasy Plant Mini Kit (Qiagen, Seoul, Korea). The Purified cp DNA was fragmented, and used to construct short-insert libraries following the manufacturer's manual (Illumina). The sequencing was performed at DNA Illumina sequencing platform.

2.2. Assembly, annotation and comparative analyses of the cp genome

The high quality reads were assembled using spades (Bankevich et al., 2012). The assembled genome was annotated using GeSeq (Tillich et al., 2017), the cp genome sequence of *Tibetia liangshanensis* (NC_036109.1) was used as reference. The comparative analysis of the cp genome was analyzed using the mVISTA program in Shuffle-LAGAN mode (Brudno et al., 2003). For the comparative cp genomic analyses, the cp genome sequence of *F. indica* was compared with the *Larrea tridentata* (-a member of the Order Zygophyllales; Family Zygophyllaceae; sub family Larreoideae; GenBank NC_028023.1), and also with the representative of the sequence of the different categories [viz. i. the representatives from different clade of the angiosperms including Rosids (Annexure I), ii. the representative of flowering plants occurs in different major habitats such as in hot/desert habitat, temperate rain forest, temperate/tropical and sub-tropical regions, cold desert, wetland/aquatic, sea grass/aquatic/marine habitat (Annexure II), iii. the representative of different groups of plants such as Algae, Bryophytes, Pteridophyte and Gymnosperm (Annexure III), iv. the representative from range of biomass such as unicellular plant, plant having thalloid body, small size herbaceous plant, living giant Gymnosperm from desert habitat, smallest flowering plants, very rapidly growing medium sized herbs, trees occurs in desert and tree with large canopy (Annexure IV), v. the representative from C3 and C4 plant (Annexure V), and vi. the representative from common, rare and major high yielding crop of the world (Annexure VI)] were plotted using the mVISTA program (Brudno et al., 2003) with the annotation of *F. indica* as reference (Fig. 1), and the percent identity plot for comparison were drawn.

Annexure I:- Apiales: *Anethum graveolens* (NC_029470), *Daucus carota* (NC_008325), *Panax schinseng* (NC_006290); Asterales: *Helianthus annuus* (NC_007977), *Lactuca sativa* (DQ383816), *Trachelium caeruleum* (NC_010442); Brassicales: *Arabidopsis thaliana* (NC_000932), *Brassica rapa* (DQ231548); Buxales: *Buxus microphylla* (EF380351); Caryophyllales: *Spinacia oleracea* (NC_002202); Cucurbitales: *Cucumis sativus* (NC_007144); Fabales: *Glycine* max (NC_007942), *Lotus corniculatus* (NC_002694), *Medicago truncatula* (NC_003119); Gentianales: *Coffea arabica* (NC_008535); Geraniales: *Pelargonium* × *hortorum* (NC_008454), Lamiales: *Jasminum nudiflorum* (NC_008407); Malphigiales: *Manihot esculenta* (EU117376), *Passiflora biflora* (NC_038120), *Populus trichocarpa* (NC_008235), Malvales: *Gossypium hirsutum* (NC_007944); Myrtales: *Eucalyptus globulus* (NC_008115), *Oenothera elata* (NC_002693); Proteales: *Platanus occidentalis* (NC_008335); Ranunculales: *Nandina domestica* (NC_008336), *Ranunculus macranthus* (NC_008796); Rosales: *Morus indica* (NC_008359); Sapindales: *Citrus sinensis* (NC_008334); Solanales: *Atropa belladonna* (NC_004561), *Ipomoea purpurea* (EU118126), *Nicotiana sylvestris* (NC_007500), *N. tabacum* (NC_001879), *N. tomentosus* (NC_007602), *Solanum bulbocastanum* (NC_007943), *S. lycopersicum* (DQ347959), *S. tuberosum* (NC_008096); Vitales: *Vitis vinifera* (NC_007957); Zygophyllales: *Larrea tridentate* (NC_028023).

Annexure II:- Hot/desert habitat: *Haloxylon persicum* (NC_027669); Temperate Rain forest: *Panax vietnamensis* (NC_033519); Temperate/tropical and sub-tropical regions: *Praxelis clematidea* (NC_023833); Cold desert: *Arabis alpine* (NC_023367); Wetland/Aquatic: *Nelumbo nucifera* (NC_025339); Sea grass/Aquatic/Marine habitat: *Zostera marina* (NC_036014).

Annexure III:- Algae: *Chlamydomonas reinhardtii* (BK000554); Bryophytes: *Marchantia paleacea* (NC_001319); Pteridophyte: *Adiantum capillus-veneris* (NC_004766); Gymnosperm: *Welwitschia mirabilis* (EU342371).

Annexure IV:- Unicellular plant: Algae- *Chlamydomonas reinhardtii* (BK000554); Thalloid plant body: Bryophytes- *Marchantia paleacea* (NC_001319); Small size herbaceous plant: Pteridophyte- *Adiantum capillus-veneris* (NC_004766); Living Gymnosperm from desert habitat: Gymnosperm- *Welwitschia mirabilis* (EU342371); Smallest flowering plants: Angiosperm- *Wolffia australiana* (NC_015899); Very rapidly growing medium sized herbs: Angiosperm- *Praxelis clematidea* (KF922320); Trees occurs in desert: Angiosperm- *Haloxylon persicum* (KF534479); Tree with large canopy: Angiosperm- *Ficus religiosa voucher* (NC_033979).

Annexure V:- C3 Plant: Poaceae- *Oryza sativa* (NC_008155), *Triticum aestivum* (NC_002762), Solanaceae- *Solanum tuberosum* (NC_008096), *Nicotiana tabacum* (Z00044), Fabaceae-*Glycine* max (NC_007942), Malvaceae- *Gossypium thurberi* (NC_015204), Fabaceae- *Arachis hypogaea* (NC_037358); C4 plant: Poaceae- *Aristida ternipe* (NC_037164), *Zea mays* (NC_001666), *Saccharum officinarum* (NC_035224), Amaranthaceae- *Bienertia sinuspersici* (KU726550), Euphorbiaceae- *Euphorbia esula* (NC_033910).

Annexure VI:- Rare: *Berchemiella wilsonii* var. *wilsoni* (KY926621); Major high yielding crop of the world: *Oryza sativa* (NC_008155), *Triticum aestivum* (NC_002762), *Zea mays* (NC_001666); Common: *Cynodon dactylon* (NC_034680).

3. Results and discussion

The high quality reads were *de novo* assembled, resulted into a contig of 128,379 bp with GC content of 34.02% which is consistent with the cp genome of *L. tridentata* (35.09%). The cp genome size of *F. indica* cp was approximately 7.5 kb, and smaller than *L. tridentata* cp genome. Interestingly, *F. indica* has single copy of inverted repeat resulting into the inverted gene order compared to its closest relative *L. tridentata.* The length of angiosperm cp genomes remains variable primarily due to expansion and contraction of the inverted repeat IR region and the single copy boundary regions. It is evident from the analysis that the coding region is less divergent than the non-coding region. However further analysis showed that *clpP* and *accD* were the most divergent coding regions.

With the increasing global population and demand for food, and the rising global temperatures and decreasing water resources, it is important to understand the genomic mechanisms of photosynthetic genes respond to abiotic and biotic stress which may lead to enhance the yield of the crops (Cushman 2001; Berry et al. 2011, 2013). The genome annotation resulted a total of 115 unique coding genes were annotated which includes 80 protein coding (represent 80200 bp nucleotides coding for 42,793 codons), 31 tRNA and 4 rRNA genes. The gene order in *F. indica* was similar to the angiosperm's gene order except for the loss of one copy of the IR and by the presence of a single, large inversion that reverses the order of the genes between rbcL and rps16, the similar cases have also been reported in some cp genome of legumes previously

Fig. 1. Percent identity plot for comparison of *Fagonia indica* chloroplast genome with the representative of plants (a) occurs in different major habitats, (b) of major eukaryotic clade [Algae to Angiosperm], (c) range of biomass [from unicellular to large canopy], (d) with major important crop [(Rice, Wheat and Maize), common plant (*Cynodon*) and rare plant (*Berchemiella wilsonii*].

(Doyle et al., 1996; Kato et al., 2000; Saski et al., 2005; Guo et al., 2007).

The comparative analyses of cp genome of the representatives from different clade of the angiosperms including Rosids with *F. indica* as reference revealed that coding region was more conserved than the non-coding region; however, *clpP* (Clp protease proteolytic subunit), *ycf1*, *ycf2* and *ycf4* were the most divergent coding region among all taxa included in the analyses. The comparative analyses of the representatives from different habitats with *F. indica* as reference also revealed that coding regions were more conserved than the non-coding regions, but here in this case the *ycf1* was the most divergent coding region among all taxa.

The percent identity plot of the representatives from different groups of plants and also the different group of plants having range of biomass with *F. indica* as reference revealed that coding regions were more conserved than the non-coding regions; however, it was interesting to note the considerable differences in between cp genome of *F. indica* and *C. reinhardtii.*

Further, the comparative analyses of cp genome of the representatives from C3 and C4 plant and the representative from common, rare and major high yielding crop of the world revealed that coding region was more conserved than the non-coding region as similar to the results of previous comparative analyses, and the *ycf1* and *ycf2* were the most divergent coding region. It was interesting to note that the loss of *accD* gene in rice, wheat, maize and *Cynodon* (the representative from common, major high yielding crop). The accD, ycf2 genes are exclusively transcribed by the nuclear-encoded plastid RNA polymerase (NEP) which plays a role in maintaining housekeeping functions in proplastids, and in plastid development from proplastids (Hajdukiewicz et al. 1997; Allison 2000; Swiatecka-Hagenbruch et al. 2008).

In summary, the comparison of *F. indica* genome in different categories showed strong evidence and further support for the conservative pattern of chloroplast genome, the coding and non-coding region remains conserved even in phylogenetically distant eukaryotic clades, and might not have the sole roles in organism's yield, rarity or abundance and biomass, and in encountering the stress. Further the present findings also support the anterograde and retrograde signaling which insures the highly coordinated expression of the photosynthetic genes (Jung and Chory, 2010).

Declaration of Competing Interest

The authors report no conflicts of interest in this work.

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding the work through the research group project

(RG-1439-84). This study was supported by the KRIBB Initiative Program of the Republic of Korea.

References

Allison, L.A., 2000. The role of sigma factors in plastid transcription. Biochimie 82, 537–548.

Apel, W., Bock, R., 2009. Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin a conversion. Plant Physiol. 151, 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, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A.S., Lesin, V., Nikolenko, S., Pham, S., Prjibelski, A., Pyshkin, A., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477.

Beier, B.-A., 2005. A revision of the desert shrub *Fagonia* (Zygophyllaceae). Systemat. Biodiver. 3 (3), 221–263.

Berry, J.O., Zielinski, A.M., Patel, M., 2011. Gene expression in mesophyll and bundle sheath cells of C4 plants. In: Raghavendra, A.S., Sage, R.F. (Eds.), C4 photosynthesis and related CO_2 concentrating mechanisms. Advances in photosynthesis and respiration, vol 32. Springer, Dordrecht, pp. 221–256.

Berry, J.O., Yerramsetty, P., Zielinski, A.M., Mure, C.M., 2013. Photosynthetic gene expression in higher plants. Photosynth. Res. 117 (1–3), 91–120.

Bobik, K., Burch-Smith, T.M., 2015. Chloroplast signaling within, between and beyond cells. Front. Plant Sci. 6, 781.

Brozynska, M., Furtado, A., Henry, R.J., 2016. Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnol. J.14, 1070–1085.

Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., Batzoglou, S., 2003. Glocal alignment: finding rearrangements during alignment. Bioinformatics 19S1, i54–i62.

Cushman, J., 2001. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. Plant Physiol. 127, 1439–1448.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17 (1), 134.

Doyle, J.J., Doyle, J.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.

Guo, X., Castillo-Ramírez, S., González, V., Bustos, P., Fernández-Vázquez, J.L., Santamaría, R.I., Arellano, J., Cevallos, M.A., Dávila, G., 2007. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. BMC Genom. 8, 228.

Hajdukiewicz, P.T.J., Allison, L.A., Maliga, P., 1997. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J. 16, 4041–4048.

Holtz, B.R., Berquist, B.R., Bennett, L.D., Kommineni, V.J., Munigunti, R.K., White, E.L., Wilkerson, D.C., Wong, K.Y., Ly, L.H., Marcel, S., 2015. Commercial-scale biotherapeutics manufacturing facility for plant-made pharmaceuticals. Plant Biotechnol. J. 13 (8), 1180–1190.

Jin, S., Daniell, H., 2014. Expression of gamma-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, 1274–1285.

Jin, S., Kanagaraj, A., Verma, D., Lange, T., Daniell, H., 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, 222–235.

Jung, H.-S., Chory, J., 2010. Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? Plant Physiol. 152, 453–459.

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 (6), 323–330.

Kim, J.S., Kim, J.H., 2013. Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two lilies (*Lilium longiflorum* and *Alstroemeria aurea*). PLoS One 8 (6), e68180.

Kwon, K.C., Daniell, H., 2015. Low-cost oral delivery of protein drugs bioencapsulated in plant cells. Plant Biotechnol. J. 13, 1017–1022.

Kwon, K.C., Nityanandam, R., New, J.S., Daniell, H., 2013. Oral delivery of bioencapsulated exendin-4 expressed in chloroplasts lowers blood glucose level in mice and stimulates insulin secretion in beta-TC6 cells. Plant Biotechnol. J. 11, 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 (7), e102655.

Mandavil, J.P., 1986. Plant life in the Rub' al-Khali (the Empty Quarter), south-central Arabia. Proc. R. Soc. Edinb. 89B, 147–157.

Saski, C., Lee, S.B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.G., Jansen, R.K., 2005. Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. Plant Mol. Biol. 59 (2), 309–322.

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat Biotechnol. 26, 1135–1145.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H., 2017. DNA sequencing at 40: past, present and future. Nature 550, 345–353.

Shintani, D., DellaPenna, D., 1998. Elevating the vitamin E content of plants through metabolic engineering. Science. 282, 2098–2100.

Swiatecka-Hagenbruch, M., Emanuel, C., Hedtke, B., Liere, K., Börner, T., 2008. Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. Nucl. Acid Res. 36, 785–792.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq – versatile and accurate annotation of organelle genomes. Nuc. Acids Res. 45, W6–W11.

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.

Viitanen, 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 p-hydroxybenzoic acid biosynthesis. Plant Physiol. 136, 4048–4060.