Opinion   **Open Access**

Chloroplast Genome Sequencing: Some Reflections

Ibrar Ahmed*

Department of Biochemistry, Quaid-i-Azam University, Islamabad, 45320, Pakistan.

Opinion

Chloroplast genome in plants-also known as a plastome-is a short, circular, double stranded DNA molecule of 83-292 kb in size [1]. This haploid genome lacks recombination, and exhibits a uniparental inheritance which is predominantly maternal in angiosperms but paternal in gymnosperms [2]. Sequence information contained in chloroplast DNA loci has been widely used in plant systematics, elucidating relationships from intraspecific through deep divergence [3-7]. In some species, chloroplast genomes have also been transformed with exogenous genes encoding important metabolites, including genes for production of vaccines against human diseases [8-10]. Selected chloroplast DNA loci have been used as barcodes for identifying plant species [11,12], although the concept of a universal barcode has inherent limitations [4]. Whole chloroplast genome sequence has been suggested as a super-barcode [13], but this idea might need to be reconsidered for out-crossing plant species that are, or have been living in sympatry.

Plastome sequencing has traditionally been reliant on cloning of DNA fragments to form DNA libraries followed by long-range PCR, or using a large set of DNA primers to amplify and sequence overlapping DNA fragments. Plastome sequencing appears to have benefited a little of recent advances in high-throughput sequencing technologies, as the number of plastomes currently available in public repositories is far little than anticipated. Even some researchers who use high-throughput sequencing technologies fail to appreciate that the plastome, rich in AT nucleotides, contains a number of mononucleotide microsatellite (As / Ts). They have used sequencing platforms such as Roche 454 [14-17] that inherently have high rates of insertion-deletion (indel) errors at mononucleotide microsatellites. Often such studies need to further validate their sequencing results by using additional platforms [14], or to fill-in the gaps at mononucleotide microsatellites by amplifying missing sequences through conventional PCR and Sanger Sequencing [17], which requires additional time and money. Instead, using an alternate platform (e.g. Illumina's MiSeq) that produces short reads of exact length and avoids errors at mononucleotide microsatellites can give better results.

In a typical plastome sequencing run, about 5-10% of sequences belong to plastome; majority of the sequences come from nuclear genome [18]. The percentage of plastome reads can be increased by enriching chloroplasts before the organelle DNA extraction [19]. Alternatively, rolling circle amplification may be used to increase the proportion of the short reads, using oligonucleotide primers which preferentially amplify AT rich regions [20]. With increased proportion of chloroplast DNA, a better approach is to multiplex or combine several plastomes in a single run. Multiplexing can give ample sequence coverage depth while giving an output data file of suitable size (~1 GB per plastome sample), making downstream analyses easier on a desktop computer. If the size of output data file is bigger (usually 3-4 GB), high computation resources (such as a server) are required for subsequent analyses. Large output data files offer little extra advantage in terms of sequence quality and coverage depth.

References

1. . Palmer JD (1986) Isolation and structural analysis of chloroplast DNA. Methods Enzymol 118: 167-186.

2. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76: 273-297.

3. Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, et al. (2012) Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 4: 1316-1323.

4. Ahmed I, Biggs PJ, Matthews PJ, Naeem M, Lockhart PJ, et al. (2013) Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of Colocasia esculenta (L.) Schott (Araceae) and closely related taxa. Mol Ecol Resour 13: 929-937.

5. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci United States Am 99: 12246-12251.

6. Kugita M, Akira K, Yuhei Y, Yuko T, Tohoru M, et al. (2003) The complete nucleotide sequence of the hornwort (Anthoceros formosae) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res 31: 716-721.

7. Yamane K, Yasui Y, Ohnishi O (2003) Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, Fagopyrum cymosum (Polygonaceae). Am J Bot 90: 339-346.

8. Lössl AG, Waheed MT (2011) Chloroplast-derived vaccines against human diseases: achievements, challenges and scopes. Plant Biotechnol J 9: 527-539.

9. Waheed MT, Thones N, Muller M, Hassan SW, Gottschamel J, et al. (2011) Plastid expression of a double-pentameric vaccine candidate containing human papillomavirus-16 L1 antigen fused with LTB as adjuvant: transplastomic plants show pleiotropic phenotypes. Plant Biotechnol J 9: 651-660.

10. 10. Waheed MT, Thones N, Muller M, Hassan SW, Mona NR, et al. (2011) Transplastomic expression of a modified human papillomavirus L1 protein leading to the assembly of capsomeres in tobacco: a step towards cost-effective second-generation vaccines. Transgenic Res 20: 271-282.

11. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, et al. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3: e2802.

12. Peter MH, Laura LF, John L, Mehrdad H, Sujeevan R (2009) CBOL Plant Working Group A DNA barcode for land plants. Proc Natl Acad Sci U S A 106: 12794-12797.

13. 13. Li X, Yang Y, Robrt JH, Maurizio R, Yitao W, et al. (2015) Plant DNA barcoding: from gene to genome. Biol Rev 90: 157-166.

14. Zhang T, Zhang X, Hu S, Yu J (2011) An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. Plant Methods 7: 38.

15. Liu Y, Huo N, Dong L, Yi W, Zhang S, et al. (2013) Complete chloroplast genome sequences of Mongolia medicine Artemisia frigida and phylogenetic relationships with other plants. PLoS One 8: e57533.

16. Zhang H, Li C, Miao H, Xiong S (2013) Insights from the complete chloroplast genome into the evolution of Sesamum indicum L. PLoS One 8: e80508.

*Corresponding author: Ibrar Ahmed, Department of Biochemistry, Quaid-i-Azam University, Islamabad, 45320, Pakistan. Tel: +92 51 9064000; E-mail: iaqureshi_qau@yahoo.com

17. Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, et al. (2010) The chloroplast genome sequence of mungbean (Vigna radiata) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17: 11-22.

18. Jansen RK, Linda AR, Jeffrey LB, Claude W, Timothy W, et al. (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol 395: 348-384.

19. Shi C, Na Hu, Hui H, Ju G, Zhao Y, et al. (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS One 7: e31468.

20. Atherton RA, Bennet JM, Lara DS, Berry AL, Nick WA, et al. (2010) Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant Methods 6:22.