

Complete chloroplast genome sequencing of sago palm (*Metroxylon sagu* Rottb.): Molecular structures, comparative analysis and evolutionary significance

Leonard Whye Kit Lim, Hung Hui Chung, Hasnain Hussain



PII: S2452-0144(20)30076-5

DOI: <https://doi.org/10.1016/j.genrep.2020.100662>

Reference: GENREP 100662

To appear in: *Gene Reports*

Received date: 14 March 2020

Accepted date: 25 March 2020

Please cite this article as: L.W.K. Lim, H.H. Chung and H. Hussain, Complete chloroplast genome sequencing of sago palm (*Metroxylon sagu* Rottb.): Molecular structures, comparative analysis and evolutionary significance, *Gene Reports* (2018), <https://doi.org/10.1016/j.genrep.2020.100662>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Complete Chloroplast Genome Sequencing of Sago Palm (*Metroxylon sagu* Rottb.):  
Molecular Structures, Comparative Analysis and Evolutionary Significance**

Leonard Whye Kit Lim<sup>a\*</sup>, Hung Hui Chung<sup>a\*</sup>, & Hasnain Hussain<sup>b</sup>

<sup>a</sup> Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.

<sup>b</sup> Centre of Sago Research (CoSAR), Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.

\* *Corresponding authors: lwkl1993@gmail.com; hhchung@unimas.my*

Number of figures: 7

Number of tables: 5

Number of supplementary figures: 3

Number of supplementary tables: 3

**ABSTRACT**

The sago palm (*Metroxylon sagu* Rottb.) is an underappreciated essential food crop of the century and is deemed to be the ultimate solution to future food security issues globally. To grasp a better comprehension on the molecular basis of sago palm chloroplast, a complete characterization of the 157,300 bp chloroplast genome sequences (encompassing 85,257 bp large single copy region (LSC) and 17,533 bp small single copy region (SSC), sandwiched by a pair of 27,245 bp inverted repeats (IRs)) was performed. The genome houses 113 unique genes: 80 protein-coding genes, 29 tRNA genes and four ribosomal genes. The gene order, AT content and orientation of the *M. sagu* chloroplast genome closely resembles that of its congeneric species, *M. warburgii*. The distribution patterns of short and long repeats were also determined along with four highly variable regions, 216 mutation events (164 single nucleotide polymorphisms (SNPs) and 52 insertion/deletions (indels)) between the two *Metroxylon* genus members. The maximum likelihood phylogenetic tree had unravelled the close relationship of *M. sagu*, with its closest neighbour *M. warburgii*, together with other Calameae members like *Pigafetta elata*, *Calamus caryotoides*, and *Salacca ramosiana*. This study had dissected on the entire chloroplast genome of *M. sagu* and characterized the contents at the molecular level. It is anticipated that the yielded treasurable information will be channelled for species identification and evolutionary analysis within the *Metroxylon* genus in future.

**Keywords:** *Metroxylon sagu*, chloroplast genome, Areaceae, genomic structure, comparative analysis

## 1. Introduction

Chloroplasts are semiautonomous intracellular organelles responsible for photosynthesis and a wide range of housekeeping biochemical processes, which are found in plant cells as well as eukaryotic algae (Vothknecht & Westhoff, 2001). This organelle has its own genome originated from endosymbiotic event (McFadden, 2001; Gray, 2004). A chloroplast (cp) genome has highly conserved circular quadripartite structure and it is commonly made up of a pair of inverted repeats (IRs), a large single-copy region (LSC) as well as a small single-copy region (SSC). Generally, chloroplast genomes can span in the range of 120 to 210 kilo basepair (kb) in length, encoding 110 to 130 genes in total (Chumley et al., 2006; Ravi et al., 2007; Lin et al., 2010; Dong et al., 2013). The inverted repeats in plants normally covers the length of 20 to 30 kb (Palmer & Delwiche, 1998).

The chloroplast genome has been put into the research limelight in the recent years in the development of genetic markers for species identification, DNA barcoding, population genetics, phylogeography and phylogeny because it generally undergoes slower evolutionary rate compared to that of the nuclear genome and is usually inherited uniparentally (maternally in majority of the angiosperms) (Song et al., 2019). Most of the genes accommodated in the chloroplast genome are actively involved in vital biological processes such as transcription, translation and photosynthesis (Ahlert et al., 2003). The RNA editing is one major posttranslational modification mechanism occurring in the chloroplast that involves the editing of a single specific nucleotide at the level of transcript as part of the repair mechanism (Bock, 2000; Schmitz-Linneweber & Barkan, 2007). Usually, the editing involves the C- (Cytidine) to -U (Uridine) conversion, which causes non-silent mutations leading to emergence or elimination of initiation or termination codons (Uthaipaisanwong et al., 2012). Since the first deposition of chloroplast genome (liverwort, *Marchantia polymorpha*) into the GenBank back in the year 1986 (Ohyama et al., 1986), the chloroplast genome database grew exponentially with the emergence of next-generation-sequencing (NGS) technologies to a massive size of 13,982 as of May 2019.

The sago palm (*Metroxylon sagu* Rottb.) is a member of the Arecaceae (Palmae family) which encompasses more than 200 genera and this family contains an estimation of 2600 species. To date, there are a total of 38 complete chloroplast genomes from the Arecaceae family that have been deposited into the GenBank, including *M. warburgii* (natangura palm), *Pigafetta elata* (black wanga palm), *Mauritia flexuosa* (fan palm), *Eugeissona tristis* (bertam), *Elaeis guineensis* (oil palm), *Phoenix dactylifera* (date palm) and *Cocos nucifera* (coconut palm) (Barrett et al., 2016; Uthaipaisanwong et al., 2012; Yang et al., 2010; Huang et al., 2013). The sago palm is one important food crop member of the Arecaceae family because it has the capability to yield starch three- to four-fold to that of rice, wheat and corn as well as 17-fold to that of cassava (Karim et al., 2008). At its maximum starch producing growth stage (the “*Angau Muda*” stage), the starch production per trunk can reach 39% to 41% on dry weight basis (Lim, 1991; Pei-Lang et al., 2006), yielding 216 to 219 kg of starch (Hamanishi et al., 1999).

Chloroplasts play essential role in starch biosynthesis, especially in the leaves or specialized amyloplasts in tissues that stores starch of staple crop plants (Pfister & Zeeman, 2016). In sago palm, the starch-storing tissue is located at the trunk whereby the trunk can reach a maximum height of 10 to 15 metres tall and 35 to 75 centimetres in diameter (Kiew,