



Genome survey of sago palm (*Metroxylon sagu* Rottboll)

Leonard Whye Kit Lim^{a,*}, Hung Hui Chung^{a,*}, Hasnain Hussain^b, Han Ming Gan^{c,d}

^a Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

^b Centre for Sago Research (CoSAR), Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

^c GeneSEQ Sdn Bhd, Bukit Beruntung, 48300 Rawang, Selangor, Malaysia

^d Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia

ARTICLE INFO

Keywords:

Metroxylon sagu Rottboll
Genome survey
Microsatellite
K-mer
Heterozygosity

ABSTRACT

The sago palm (*Metroxylon sagu* Rottboll) is a halophytic, underrated starch-producing palm that is both disease resistant and capable of producing a high yield of storage starch. It is deemed the future tree of life as it has a high potential to be cultivated for its starch to solve global food security issues in the future. Due to low scientific interest of this palm, scarcity of its genomic information has halted our comprehension on the mechanism behind its starch biosynthesis as well as salt and disease resistance pathways. In this study, we report on the genome survey and microsatellite analysis of sago palm. Based on the k-mer statistical approach, the sago palm has an estimated genome size of 464 Mbp to 616 Mbp, consisting of 31.9 to 35.7% repetitive sequence. The heterozygosity of the individual sequenced, a true sago palm clonal breed, was 0.63 to 0.66%. A sum of 123,430,321,878 clean bases were yielded from the 124,258,713,300 raw bases. BUSCO analysis revealed that only around 22% of the 1375 single copy plant genes are complete in both sago palm genome scaffolds and contigs whereas the remaining are missing (~45%) and fragmented (~32%). The fragmented genome assembly generated from short-read Illumina data is likely due to the high repetitiveness of the genome. It is recommended that a long-read sequencing to be carried out to improve the genome information of sago palm. The Illumina data generated in this study will be instrumental for polishing the genome assembly generated from a long-read assembly that is known to be error-prone.

1. Introduction

The sago palm (*M. sagu* Rottboll) is one unique starch-producing halophytic palm found mainly in Sarawak state of Malaysia, Indonesia, Papua New Guinea and the Philippines. It is known as the future tree of life for its diverse functions yieldable from all organs above its roots, namely frond, leaf petiole, trunk, sap, trunk starch, pith and fibrous residue. This palm is a pivotal eco-economic plant equipped with relatively high salt tolerance capability (Lim and Chung, 2020) coupled with its phenomenally higher starch yield in contrast to a myriad of commercial food crops established to date (Karim et al., 2008). It can survive well in tropical rainforests as well as lowland freshwater swamp (Johnson, 1977). The sago starch harvested from its trunk is valuable raw material input for the food, polymer, textile, pharmaceutical industries (Radley, 1976; Nuttanan et al., 1995; Ishiaku et al., 2002; Purwani et al., 2006). The sago palm fronds are durable for thatching and papermaking (Jamaludin et al., 1995). The sago fibrous residue is nothing close to what its name suggests because of its vast applications

across numerous manufacturing industries involving animal feed, bio-ethanol, fermentable sugar, biosorbent as well as microbial conversion agent, to name a few (Vickineswary et al., 1994; Bujang et al., 1996; Kumaran et al., 1997; Awg-Adeni et al., 2010). Above all, what makes this palm superior to its other food crop counterparts is its ability to withstand salinity thrice as much as the other starch producers (Yoneta et al., 2006). The sago palm has been the centre of research recently (reviewed by Lim et al., 2019), picking up its pace to emerge as one of the most versatile food crops to solve food security and scarcity issues in the future (Ehara, 2009).

Although this palm is one of the main starch producing crops in the region of Borneo, Indonesia, Papua New Guinea and the Philippines, very little of molecular study has been done for this plant (Lim et al., 2019; Nisar and Hussain, 2020). The central dogma of molecular biology is known as the source of life for all organisms and genes play a major part in it (Jee et al., 2017; Lim et al., 2018; Aminan et al., 2020; Lai et al., 2021; Md Yusni et al., 2020; Yeaw et al., 2020; Lim et al., 2021c; Lau et al., 2021b). With the booming of omics technology at an

* Corresponding authors.

E-mail addresses: 19010029@siwa.unimas.my (L.W.K. Lim), hychung@unimas.my (H.H. Chung).

<https://doi.org/10.1016/j.plgene.2021.100341>

Received 25 March 2021; Received in revised form 8 October 2021; Accepted 20 October 2021

Available online 24 October 2021

2352-4073/© 2021 Elsevier B.V. All rights reserved.