



## Data Article

# First high-quality genome assembly data of sago palm (*Metroxylon sagu* Rottboll)

Leonard Whye Kit Lim<sup>a,\*</sup>, Melinda Mei Lin Lau<sup>a</sup>,  
Hung Hui Chung<sup>a,\*</sup>, Hasnain Hussain<sup>b</sup>, Han Ming Gan<sup>c,d</sup>

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

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

<sup>c</sup> GeneSEQ Sdn Bhd, Bukit Beruntung, 48300 Rawang, Selangor, Malaysia

<sup>d</sup> Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia

## ARTICLE INFO

## Article history:

Received 18 November 2021

Accepted 4 January 2022

Available online 6 January 2022

## Keywords:

*Metroxylon sagu*

Sago palm

Genome annotation

Starch synthesis

## ABSTRACT

The sago palm (*Metroxylon sagu* Rottboll) is a tropical halophytic starch-producing, economically important crop palm mainly located in Southeast Asian countries. Recently, a genome survey was conducted on this palm using the Illumina sequencing platform, with a very low (21.5%) BUSCO genome completeness score, and most of them (~78%) are either fragmented or missing. Thus, in this study, the sago palm genome completeness was further improved with the utilization of the Nanopore sequencing platform that produced longer reads. A hybrid genome assembly was conducted, and the outcome was a much complete sago palm genome with BUSCO completeness achieved at as high as 97.9%, with only ~2% of them either fragmented or missing. The estimated genome size of the sago palm is 509,812,790 bp in this study. A sum of 33,242 protein-coding genes was revealed from the sago palm genome and around 96.39% of them had been functionally annotated. An investigation on the carbohydrate metabolism KEGG pathways

**Abbreviations:** aa, amino acid; bp, base pair; BUSCO, Benchmarking Universal Single Copy Orthologs; KEGG, Kyoto Encyclopedia of Genes and Genomes; SDS, Sodium dodecyl sulfate.

\* Corresponding authors.

E-mail addresses: [19010029@siswa.unimas.my](mailto:19010029@siswa.unimas.my) (L.W.K. Lim), [hhchung@unimas.my](mailto:hhchung@unimas.my) (H.H. Chung).

Social media:  (L.W.K. Lim)

<https://doi.org/10.1016/j.dib.2022.107800>

2352-3409/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)