



Faculty of Resource Science and Technology

Genomic Landscape of Sago Palm (*Metroxylon sagu* Rottb.)

Lim Whye Kit Leonard

**Doctor of Philosophy
2022**

Genomic Landscape of Sago Palm (*Metroxylon sagu* Rottb.)

Lim Whye Kit Leonard

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

(Genetics)

Faculty of Resource Science and Technology

UNIVERSITI MALAYSIA SARAWAK

2022

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Lim Whye Kit Leonard

Matric No.: 19010029

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date: 7 February 2022

ACKNOWLEDGEMENT

This thesis could not have been come to fruition without the robust assistance and encouragement from many individuals. I would like to take this opportunity to express my heartfelt appreciation to each of them.

First and foremost, I would like to thank God for all the spiritual strength, wisdom and good health granted to me throughout this research.

My deepest gratitude goes to my supervisor, Dr. Chung Hung Hui who have persistently and selflessly mentored me with his overwhelming support and guidance. His never-ending impartment of expertise and knowledge are the keys towards the success of this project. I also would like to thank my co-supervisor, Assoc. Prof. Dr. Hasnain Hussain for his priceless insights and endless encouragement.

Next, I would like to express my exceptional appreciation to my lab mates: Melinda, Mindy, Tiing Tiing, Nabeel, Danial, Joslyn, Pei Ni, Zi Xuan and Nabila, who have voluntarily offered numerous types of aid and inspiration during my hard times. My sincere thank you goes to my friends: John Wong, Delvino, Anarrin, John Chua, Yew, Carlson, Mugu, Simon, Soon Sen and Chia Hong for their motivations.

Last but not least, I am greatly beholden to my dear family members: my wife Teh Zy Ying, my dad Lim Loong Kien, my mum Liew Hong Yen and my sister, Lim Zhi Vern Joanne for their unceasing love. Without them, my journey as a Doctor of Philosophy student would be very much tougher and challenging. Thank you once again to everyone who have helped in making this thesis a reality.

ABSTRACT

The sago palm (*Metroxylon sagu* Rottboll) is a halophytic tree and the ultimate solution to two major global issues: food security and the rapid loss of saline free agricultural lands. This palm can produce high amount of starch, besides being resistant to saline and almost disease free. Unfortunately, this valuable palm is currently underappreciated by the scientific community and the genomic landscape of this palm remained scarce to date. The aim of this study is to sequence and characterise the sago palm chloroplast genome, document the organellar genome copy number changes across various developmental stages, organs, localities and phenotypes as well as sequence the entire nuclear genome of sago palm. The sago palm chloroplast and nuclear genome sizes determined in this study are 157,300 bp and 616,599,600 bp respectively, yielded via next generation sequencing technologies. The organellar genome copy number investigation employing the combinatorial approach of both qPCR and long-PCR assays had unearthed that the mature leaf is generally most enriched with unimpeded plastome and mitogenome among other organs. The *Plawei Manit* growth stage was discovered to have the highest unimpeded organellar genomes copy numbers (over 180,000) among the five major growth stages. The trunking phenotype was proven to be superior to that of the spiny and non-trunking phenotypes in terms of organellar genomes copy numbers (over 188,000). Interestingly, the organellar genome copy numbers across four main Sarawak localities in Malaysia are indifferent statistically. This study serves as a groundwork for preliminary sago palm genomic landscape to support future haplotype and genome wide association studies.

Keywords: *Metroxylon sagu*, chloroplast genome, nuclear genome, organellar genome, genomic landscape

Lanskap Genomik Palma Sagu (Metroxylon sagu Rottb.)

ABSTRAK

Palma sagu (Metroxylon sagu Rottboll) ialah pohon halofit dan penyelesaian dua masalah global utama: keselamatan makanan dan kehilangan tanah pertanian bebas garam yang pantas. Pokok palma ini dapat menghasilkan banyak kanji, selain lasak terhadap garam dan hampir bebas penyakit. Malangnya, palma ini kini kurang dihargai oleh komuniti saintifik dan landskap genomik palma ini masih belum ada. Tujuan kajian ini adalah untuk menjujuk dan mencirikan genom kloroplas sagu, mendokumentasi perubahan bilangan salinan genom organel di pelbagai peringkat perkembangan, organ, kawasan dan fenotip serta menjujuk keseluruhan genom nuklear palma sagu. Saiz genom kloroplas dan nuklear palma sagu yang ditentukan dalam kajian ini ialah 157,300 bp dan 616,599,600 bp, dihasilkan melalui teknologi penjujukan generasi depan. Penyelidikan salinan genom organel menggunakan cara kombinatori ujian qPCR dan PCR panjang telah menemukan bahawa daun matang pada umumnya paling diperkaya dengan plastom dan mitogenom antara organ-organ lain. Tahap Plawei Manit didapati mempunyai salinan genom organel tertinggi antara lima peringkat pertumbuhan utama (melebihi 180,000). Fenotip berbatang terbukti lebih unggul daripada fenotip berduri dan bukan berbatang dari segi salinan genom organel (melebihi 188,000). Menariknya, nombor salinan genom organel di empat kawasan utama Sarawak di Malaysia tidak berbeza secara statistik. Secara keseluruhan, kajian ini berfungsi sebagai landasan dalam membentuk lanskap genomik awal palma sagu untuk menyokong kajian haplotaip dan seluruh genom pada masa depan.

Kata kunci: *Metroxylon sagu, genom kloroplas, genom nuklear, genom organel, lanskap genomik*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	iv
TABLE OF CONTENTS	v
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xv
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Objectives	3
1.4 Sago Palm (<i>Metroxylon sagu</i> Rottb.)	4
1.5 Organellar Genomes	5
1.5.1 Chloroplast Genome	7
1.5.2 Mitogenome	8
1.5.3 The Significance of Organellar Genome Copy Numbers on Sago Palm	10
1.6 Research Landscape of Sago Palm	12

1.6.1	Sago Food Manufacturing and Polymer Production	13
1.6.2	Microbiology of Sago Palm	14
1.6.3	Sago Bioprocess Operation	16
1.6.4	Sago Waste Superintendence	18
1.6.5	Sago Palm Pest Control	19
1.6.6	Sago Palm Computational Biology	20
1.6.7	Phenotypic Variation Studies of Sago Palm	21
1.6.8	Sago Palm Population Studies	22
1.6.9	Sago Palm Genetics and Omics	24
1.6.10	The Importance of Sago Palm Genetic Research	26
CHAPTER 2: CHLOROPLAST GENOME SEQUENCING AND CHARACTERIZATION		28
2.1	Introduction	28
2.2	Methodology	30
2.2.1	Genome Sequencing, Assembly and Validation	30
2.2.2	Genome Annotation	31
2.2.3	Microsatellites and Long Repeats Analysis	31
2.2.4	Internal Repeat Expansion and Contraction Analysis	32
2.2.5	Comparative and Divergence Genomic Analysis	32
2.2.6	RNA Editing Analysis	33

2.2.7	Phylogenetic Analysis	33
2.3	Results	33
2.3.1	Genome Assembly and Validation	33
2.3.2	Chloroplast Genome Organisation	35
2.3.3	Analysis of Long and Short Repeats	41
2.3.4	Inverted Repeat Contraction and Expansion	43
2.3.5	Comparative Chloroplast Genome Analysis	45
2.3.6	RNA Editing	54
2.3.7	Phylogenetic Analysis	58
2.4	Discussion	61
2.4.1	Chloroplast Genome Organisation	61
2.4.2	Nucleotide Variation Analysis	61
2.4.3	Phylogenetic Analysis	64
CHAPTER 3: ORGANELLAR GENOME PROFILES ACROSS ORGANS, GROWTH STAGES, PHENOTYPES AND LOCALITIES		65
3.1	Introduction	65
3.2	Methodology	67
3.2.1	Sampling and DNA Extraction	67
3.2.2	Long-PCR	68
3.2.3	Real-time PCR	69

3.2.4	Statistical Analysis	71
3.3	Results	71
3.3.1	Visualization of Long-PCR Bands	71
3.3.2	Organellar Genome Analysis Across Different Organs	72
3.3.3	Organellar Genome Analysis Across Different Growth Stages	74
3.3.4	Organellar Genome Analysis Across Different Phenotypes	75
3.3.5	Organellar Genome Analysis Across Different Localities in Sarawak	77
3.4	Discussion	80
3.4.1	Organellar Genome Analysis Across Different Organs	80
3.4.2	Organellar Genome Analysis Across Different Growth Stages	82
3.4.3	Organellar Genome Analysis Across Different Phenotypes	83
3.4.4	Organellar Genome Analysis Across Different Localities in Sarawak	84
3.4.5	Interspecies Comparison of Organellar Genome Counts	85
	CHAPTER 4: GENOME SURVEY ANALYSIS	86
4.1	Introduction	86
4.2	Methodology	87
4.2.1	Sampling and DNA extraction	87
4.2.2	Sequencing and Quality Control	87
4.2.3	K-mer Analysis	88
4.2.4	<i>De novo</i> Genome Assembly and Genome Completeness Assessment	88

4.2.5	Microsatellite Analysis	89
4.2.6	Starch-related Homology Search	89
4.2.7	Identification of rRNA Cistron Components and Phylogenetic Analysis	90
4.3	Results	90
4.3.1	Sequencing Data Statistics	90
4.3.2	K-mer Analysis	91
4.3.3	GC Content and Distribution Pattern	93
4.3.4	<i>De novo</i> Genome Assembly and Genome Comprehensiveness Assessment	94
4.3.5	Small Sequence Repeat (SSR) Analysis	94
4.3.6	Starch-related Homology Analysis	95
4.3.7	Characterisation of rRNA Cistrons and Phylogenetic Analysis	98
4.4	Discussion	100
4.4.1	Genome Length Approximation	100
4.4.2	Genome Heterozygosity, Repeated Sequences Percentage and GC Content	101
4.4.3	Genome Comprehensiveness Assessment	103
4.4.4	Genome-wide Microsatellite Analysis	103
4.4.5	Starch-related Homology Analysis	104
4.4.6	Identification of rRNA Cistron Components and Phylogenetic Analysis	105
	CHAPTER 5: CONCLUSION AND RECOMMENDATIONS	108
5.1	Conclusion	108

5.2	Recommendations	109
	REFERENCES	110
	APPENDICES	157

LIST OF TABLES

	Page
Table 2.1	Genes encoded by the sago palm chloroplast genome. 36
Table 2.2	The sago palm chloroplast genome features. 37
Table 2.3	The list of genes along with their respective introns in the sago palm chloroplast genome, depicting exon and intron lengths. 38
Table 2.4	Codon usage of both <i>Metroxylon</i> species chloroplast genomes. Ms: <i>M. sagu</i> ; Mw: <i>M. warburgii</i> . 38
Table 2.5	The frequency of discovered microsatellite motifs across the <i>M. sagu</i> and <i>M. warburgii</i> chloroplast genomes. 43
Table 2.6	SNP identified from the comparison between <i>M. warburgii</i> and <i>M. sagu</i> chloroplast genomes. 49
Table 2.7	Indel comparison across both <i>Metroxylon</i> species chloroplast genomes. Ms: <i>M. sagu</i> ; Mw: <i>M. warburgii</i> . 53
Table 2.8	RNA editing sites in sago palm chloroplast transcripts. 55
Table 4.1	List of protein sequences utilised for DIAMOND protein BLASTx with their respective GenBank accession numbers. 89
Table 4.2	The summary of general genomic information of sago palm nuclear genome. 91
Table 4.3	Leading similarity scoring scaffolds (cut-off value 70%) with their aligned length greater than 100 aa, along with their respective GenBank accession number, E-value, identity percentage and amino acid length. 97
Table 4.4	Leading nucleotide hits of sago palm rRNA cistrons, alongside their similarity percentage, E-value and amino acid length. 98

LIST OF FIGURES

	Page
<p>Figure 2.1 The circular gene map of the sago palm chloroplast genome. The genes found on the internal of the bigger circle are transcribed clockwise whereas the genes located on the external of the bigger circle are transcribed anticlockwise. The GC content was illustrated by the darker grey region whereas the AT/U content was depicted by the lighter grey region. LSC, large single copy region; IR, inverted repeat region; SSC, small single copy region.</p>	34
<p>Figure 2.2 The amino acid frequencies in protein-coding genes of both <i>M. sagu</i> and <i>M. warburgii</i> chloroplast genomes.</p>	40
<p>Figure 2.3 The distribution of SSRs in the chloroplast genomes of <i>Metroxylon sagu</i> and <i>Metroxylon warburgii</i>. (A) Number of varying SSR types in both <i>Metroxylon</i> species chloroplast genomes; (B) The distribution pattern of SSRs across different chloroplast genomic regions in both <i>Metroxylon</i> species genomes; (C) The distribution pattern of SSRs in coding and non-coding regions of the two <i>Metroxylon</i> species chloroplast genomes.</p>	42
<p>Figure 2.4 Long repeat sequences analysis on <i>M. sagu</i> and <i>M. warburgii</i> chloroplast genomes. (A) Long repeats types and frequency; (B) Long repeats frequency by length.</p>	43
<p>Figure 2.5 Comparison of IR/LSC and IR/SSC border positions across five Arecaceae family members chloroplast genomes. JLB: LSC/IRb junction; JSB: IRb/SSC junction; JSA: SSC/IRa junction; JLA: IRa/LSC junction.</p>	44
<p>Figure 2.6 Comparison of chloroplast genomes across the five Arecaceae family members, with <i>M. sagu</i> as reference, employing mVISTA alignment program. Dark grey arrows represent gene orientations, dark grey bars indicate exons, light grey bars depict non-coding sequences and white spaces represent differences across the chloroplast genomes.</p>	46
<p>Figure 2.7 Sliding window analysis of the complete chloroplast genomes. Window length: 600 bp; Step size: 200 bp. X-axis: midpoint positions of a window; Y-axis: nucleotide diversity of each window. (A) Nucleotide diversity across <i>M. warburgii</i> and <i>M. sagu</i>. (B) Nucleotide diversity across <i>E. tristis</i>, both <i>Metroxylon</i> species, <i>P. elata</i> and <i>M. flexuosa</i>.</p>	48
<p>Figure 2.8 The distribution of RNA editing sites across sago palm chloroplast genes.</p>	58

- Figure 2.9 The maximum likelihood phylogenetic tree of Arecaceae family members inferred from chloroplast protein-coding genes, with two outgroup gymnosperms (*Typha latifolia* and *Hanguana malayana*) included into analysis. The labels on the right represent the subfamilies of the Arecaceae family whereas the labels on the left represent Calamoid and Coryphoid tribes, with their respective genome sizes respectively. 60
- Figure 3.1 Long-PCR amplicons of mitogenome and plastome DNA across a variety of organs, growth stages and phenotypes of sago palm. Pt: plastome DNA; Mt: mitogenome DNA; ML: mature leaf; NT: non-trunking; S: spiny; YS: young shoot; P: pneumatophore; AL: aged leaf; YR: young root; G1: *Plawei* stage; G2: *Plawei Mani* stage; G3: *Bubul* stage; G4: *Angau Muda* stage; G5: *Angau Tua* stage. 72
- Figure 3.2 (A) The organellar genome copy number across different organs of sago palm established via long-PCR and qPCR. Statistical analysis was implemented using one-way ANOVA and post-hoc Dunn's Test with $p < 0.05$. Significantly distinct data are denoted with different alphabets and the fold differences to that of the nuclear genome copy number are depicted beside the alphabets. (B) The long-PCR/qPCR ratio graph displaying the plastome and mitogenome long-PCR/qPCR ratio across different organs of sago palm. mt: mitogenome; pt: plastome. 73
- Figure 3.3 (A) The organellar genome copy number across different growth stages of sago palm established via long-PCR and qPCR. Statistical analysis was implemented using one-way ANOVA and post-hoc Dunn's Test with $p < 0.05$. Significantly distinct data are denoted with different alphabets and the fold differences to that of the nuclear genome copy number are depicted beside the alphabets. (B) The long-PCR/qPCR ratio graph displaying the plastome and mitogenome long-PCR/qPCR ratio across different growth stages of sago palm. mt: mitogenome; pt: plastome. 75
- Figure 3.4 (A) The organellar genome copy number across different phenotypes of sago palm established via long-PCR and qPCR. Statistical analysis was implemented using one-way ANOVA and post-hoc Dunn's Test with $p < 0.05$. Significantly distinct data are denoted with different alphabets and the fold differences to that of the nuclear genome copy number are depicted beside the alphabets. (B) The long-PCR/qPCR ratio graph displaying the plastome and mitogenome long-PCR/qPCR ratio across different phenotypes of sago palm. mt: mitogenome; pt: plastome. 77
- Figure 3.5 (A) The organellar genome copy number across different main localities of sago palm in Sarawak, Malaysia established via long-PCR and qPCR. Statistical analysis was implemented using one-way ANOVA and post-hoc Dunn's Test with $p < 0.05$. Significantly

different data are conveyed with different alphabets and the fold differences to that of the nuclear genome copy number are indicated beside the alphabets. Non-trunking phenotype samples were not harvested from Pusa and Samarahan due to unobtainability. (B) The long-PCR/qPCR ratio graph illustrating the mitogenome and plastome long-PCR/qPCR ratio across different main localities of sago palm in Sarawak, Malaysia. Non-trunking phenotype samples were not harvested from Pusa and Samarahan due to unobtainability. mt: mitogenome; pt: plastome. Bars denote Sarawak localities. 79

- Figure 4.1 The distribution curves of K-mer. (A) K21: K-mer value 21; (B) K25: K-mer value 25; K31: (C) K-mer value 31. GenomeScope parameters: read length = 150 bp; max K-mer coverage = 1000. 92
- Figure 4.2 The GC depth graph of sago palm. X-axis: GC percentage; Y-axis: sequencing depth. 93
- Figure 4.3 The constitution of dinucleotide, trinucleotide, tetranucleotide and pentanucleotide SSRs identified from the sago palm genome paired-end sequencing reads. 95
- Figure 4.4 The SSR percentage bar graphs with the chosen top six frequency SSRs from each group: (A) dinucleotide repeats, (B) trinucleotide repeats, (C) tetranucleotide repeats and (D) pentanucleotide repeats. 95
- Figure 4.5 The genome-wide starch-related protein hits predicted utilising the sago palm genome scaffolds. 97
- Figure 4.6 The ITS2 maximum likelihood phylogenetic tree of 120 palms (containing sago palm) with Kimura 2 + G model and 1000 bootstrap repetitions. 99

LIST OF ABBREVIATIONS

CTAB	Cetyl Trimethylammonium Bromide
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
IR	Inverted Repeat
ITS	Internal Transcribed Spacer
LSC	Large Single-Copy Region
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SSC	Small Single-Copy Region
SSR	Short Sequence Repeat
TAE	Tris Acetate EDTA

CHAPTER 1

INTRODUCTION

1.1 Background

The sago palm (*Metroxylon sagu* Rottboll) is one of the most underutilised and underrated eco-economic food crop of the future. It has several superior characteristics that are unique and advantageous as compared to other commercial and conventional food crops. To name a few, its high starch yield as well as its hardy characteristics like salt resistance and disease tolerance (Karim et al., 2008; Ehara, 2009). Its starch is completely protected within the woody trunk of the tree and it is less vulnerable to pest manifestation. Previously, many researchers have been focusing on studying its starch quality, salt resistance, pest resistance as well as disease resistance physiologically (Schuiling & Flach, 1985; Haryanto & Pangloli, 1992; Notohadiprawiro & Louhenapessy, 1992; Djoefric, 1999; Purwani et al., 2006; Agus & Subiksa, 2008; Bintoro et al., 2010; Okazaki & Sasaki, 2018; Purwoko et al., 2019; Roslan et al., 2020) but the focus on the genomic aspect is still lacking to date.

The genetic and genomic aspects of plants have been proven powerful in elucidating various underlying functional pathways and mechanisms (Li et al., 2006; Kumar et al., 2014; Ma & Li, 2015). The success of employing both organellar and nuclear genome data in enhancing our comprehension and elucidation of the essential crop metabolic pathways was observed in various key crops like rice, corn, cassava, wheat and sorghum. Distinctive starch bioproduction pathways of cassava were unearthed across varying root developmental stages by Saithong et al. (2013) with the availability of the

entire genomic sequences. Campbell et al. (2016) had successfully identified the effects of domestication on the starch biosynthesis pathway based on whole sorghum genome data. The identification of the contributors towards the starch content regulation pathway in corn kernels was also succeeded through the genome-wide association study (Liu et al., 2016). The detection of the differential starch accumulation rate in endosperm of different rice cultivars, *japonica* and *indica*, was also made possible with the presence of whole genome data (Inukai, 2017). The genome-wide approach had also magnificently assisted in the discovery of key transcription factors orchestrating the reserve starch biosynthesis in wheat (Gu et al., 2021).

1.2 Problem Statement

The sago palm is one imperative player in various fields of industry, contributing significantly towards food security and economy in the past, present and future. However, there is several gaps in research on this cash crop that remained earthed to date and this had impeded the progress to establish this food crop as one of the most promising commodities to human race. Two significant aspects that requires dire attention are none other than the genetic and genomic aspects of the sago palm whereby there is only a handful of research tapping on only the tip of the iceberg.

Interestingly, the sago palm starch yield was previously found to be three to four times more than that of corn, wheat and rice as well as 17 times more than that of the cassava (Karim et al., 2008). Adding to its advantage is that the sago palm is highly tolerable to brackish environment unlike all the aforementioned conventional food crops. The research question is whether the genomic landscape of sago palm could provide us

with more information on the mystery behind its high starch yield coupled with high salt and disease resistance. It is hypothesised that understanding the basic food-producing and defence capability of the sago palm would lead to the comprehension of the contributors orchestrating the metabolic pathways that caused the large discrepancy in starch yield, salt tolerance and disease resistance. This is where the genetic and genomic aspects of the sago palm come into picture as indispensable puzzle pieces in deciphering its unknown starch biosynthesis and salt resistance pathways that make it superior of other food crops.

1.3 Objectives

The genetic and genomic aspects of the sago palm are huge knowledge gaps waiting to be filled urgently. To answer the research question above, the hypothesis of this research was centered around deciphering the genomic landscape of the sago palm via sequencing and characterization approaches. The scope of this study covers various genomic aspects of the sago palm. Therefore, the objectives of this present study are:

- i. To sequence and characterise the entire chloroplast genome of the sago palm;
- ii. To elucidate the organellar genomes copy number at different organs, growth stages, localities and phenotypes of the sago palm; and
- iii. To conduct genome survey on the entire nuclear genome of the sago palm.

1.4 Sago Palm (*Metroxylon sagu* Rottb.)

The sago palm (*Metroxylon sagu* Rottb.) is a true palm grouped in the order Arecales, family Palmae and subfamily Calamidae (Johnson, 1977). Native to Southeast Asia including Indonesia, Papua New Guinea and Malaysia, this palm can thrive well in clumps and in pure stands, in low-land freshwater swampy areas and tropical rain forests (Johnson, 1977; Husaini et al., 2016). It is deemed as one of the potential underutilised food palms that stores starch right within the stem of palm, unlike other crops that normally store starch in the form of cereals, legumes or tubers (Husaini et al., 2016). Its hardy characteristics are probably one of the reasons that explains for its high starch yield per unit: three to four-fold higher than that of corn, rice and wheat and astonishingly 17 times greater than that of cassava (Karim et al., 2008; Ehara, 2009).

The sago palm can grow up to the height of 15 metres and 75 centimetres thickness topped with a massive crown, trunk formation is only initiated at the third to fourth year of palm growth (Kiew, 1977; Kueh, 1977). Within their vegetative phase of seven to fifteen years, photosynthetically produced nutrients from the leaves are channelled to the trunk for storage in the form of starch which upon palm maturity, saturates the trunk up till the crown (Lim, 1991). Flowering occurs at maturity (15 years) and is followed by fruit formation, the falling of mature fruits off the palm is an early indication of palm death (Kueh et al., 1987). In Sarawak, five growth stages have been characterised on mature sago palm, namely “Plawei” (palm at maximum vegetative growth), “Plawei Manit” (emergence of inflorescence), “Bubul” (development of inflorescence), “Angau Muda” (flowering) and “Angau Tua” (fruiting) (Lim, 1991). The “Angau Muda” stage scores the greatest starch content per trunk (39% to 41% on dry weight basis) (Lim, 1991; Pei-Lang et al., 2006) among all other stages, yielding 216 to 219 kg of starch per dry weight (Hamanishi et al.,

1999) whereas the other three stages before the “Angau Tua” stage had showed no significant differences in terms of starch yield (Lim, 1991).

The sago palm is a palm with high economic value where different parts such as the pith, sap, leaf petiole, frond and even the “hampas” (fibrous residue) can be harvested for valuable uses (Singhal et al., 2008). The living tree itself is considered as a countermeasure to the greenhouse effect (Quat Ng, 2007). The sago starch had been placed into the limelight of the food, textile, polymer and pharmaceutical industries with promising potentials (Radley, 1976; Nuttanan et al., 1995; Ishiaku et al., 2002; Purwani et al., 2006). Besides, the sago fronds are excellent materials for papermaking and thatching with amazing durability (Jamaludin et al., 1995). The sago “hampas” is nothing close to what its name suggests because of its wide potential ranging from substrate for microbial conversion via solid state fermentation (Kumaran et al., 1997) to its role as biosorbent (Vickineswary et al., 1994) and activated carbon for industrial wastewater treatment (Kardirvelu et al., 2004; Vennilamani et al., 2005). With more functions of this palm being added to the list at an unprecedented pace, the genetic makeup of this palm might be the key towards better understanding of the driver behind these superb phenotypes of the sago palm.

1.5 Organellar Genomes

The organellar genomes, namely the mitogenome and plastome, are indispensable in various biological and developmental processes. These genomes are exposed to damage caused by various reactive by-products such as reactive oxygen species (ROS) and UV light, because of the metabolism pathways they are involved in (Boecs et al., 2010;

Alexeyev et al., 2013; Kumar et al., 2014). Unlike their nuclear counterpart, their highly polyploid nature favour for the degradation of damaged molecules instead of the DNA repair pathway as seen in the nuclear genome (Liu & Demple, 2010). This accounts for their varied copy numbers in terms of mitochondrial DNA and plastid DNA across various developmental stages of plant.

The high copy number of organellar DNA is a strong indication of great demand for organellar ribosomes and the possible way to fulfil this demand is no other than the elevation of rRNA level from the event of genome multiplication (Bendich, 1987). The copy number variations in plants are one of the major drivers in phenotypic variations, orchestrating heritable dissimilar traits between individuals (Zhang et al., 2009; Iskow et al., 2012). These traits enable plants to possess efficient maintenance pathways essential for the battle against oxidative pressure, environmental damage and replication defaults (Gualberto et al., 2014). Along the process, when not lethal to host plant, mitogenome rearrangements takes place, which can lead to severe phenotypes or induce cytoplasmic male sterility (CMS) (Gualberto et al., 2014).

Energy metabolism impacts the mitochondrial and chloroplast genome greater than the nuclear genome. Both fully functional organellar DNA and nuclear DNA are required for photosynthesis and respiration processes at early stages of organellar development (Kumar et al., 2014; Lim et al., 2020). The high copy of organellar DNA deviates from that of the stable nuclear copy, depicting a decrement trend as leaves start to develop and physiological roles of organellar DNA changes (Kumar et al., 2014). Fully functional plastome is preserved to contribute to chloroplast development until the greening of plant where its degradation is facilitated by ROS. Similarly, the fully functional mitogenome is maintained up till the photosynthesis stage before it is not required anymore in mature

green cells (Kumar et al., 2014). In general, the relinquishment of organellar DNA from the plastome is drastic in maize and gradual in other dicot plants such as tobacco, pea and barrel medick (Oldenburg et al., 2006; Shaver et al., 2006; Rowan & Bendich, 2009).

1.5.1 Chloroplast Genome

The chloroplast is one important metabolic powerhouse that generates carbohydrates from solar energy via the process known as photosynthesis. This organelle also plays indispensable roles in essential plant development and physiology processes, namely in the synthesis of phytohormones, fatty acids, amino acids, nucleotides, vitamins and a myriad of other metabolites as well as in the assimilation of nitrogen and sulphur (Daniell et al., 2016; Lim et al., 2020). These metabolites are utilised greatly in aiding the host plant to interact with its environment, for instances, in response to salt, heat, drought, light and viral invasions (Daniell et al., 2016; Lim et al., 2020).

The first chloroplast genome that was sequenced was that of the tobacco (*Nicotiana tabacum*), which was completed in the year 1986 (Shinozaki et al., 1986). Since then, the number of chloroplast genomes available in the public gene database had elevated exponentially with the aid of the advancement of high-throughput sequencing technologies. The chloroplast genome is one massive reservoir with the complete collection of essential genes involved in plant biology and diversity. Furthermore, it provides the resolution in plant species identification and evolutionary relationship. It is also one important puzzle piece to the comprehension of adaptation mechanism of the host plant towards threats like climate and salinity stress, especially in crops. The chloroplast genome is also made interesting with the study of gene transfer events across the other two genomes, mitogenome and nuclear genome.