

Mitochondrial Genome of *Garcinia mangostana* L. variety Mesta

Ching-Ching Wee^{1,2}, Nor Azlan Nor Muhammad¹, Vijay Kumar Subbiah², Masanori Arita³,
Yasukazu Nakamura³, Hoe-Han Goh^{1*}

¹Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.

²Biotechnology Research Institute, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia.

³National Institute of Genetics, Mishima, 411-8540 Shizuoka, Japan.

***Correspondence:**

E-mail: gohhh@ukm.edu.my

Short Title: Mitogenome of mangosteen variety Mesta

Abstract

Fruits of *Garcinia mangostana* L. (mangosteen) are rich in nutrients with xanthenes found in the pericarp having great pharmaceutical potential. Mangosteen variety Mesta is only found in Malaysia, which tastes sweeter than the common Manggis variety in Southeast Asia. In this study, we report the complete mitogenome of *G. mangostana* L. variety Mesta with a total sequence length of 371,235 bp of which 1.7% could be of plastid origin. The overall GC content of the mitogenome is 43.8%, comprising 29 protein-coding genes, 3 rRNA genes, and 21 tRNA genes. Repeat and tandem repeat sequences accounted for 5.8% and 0.15% of the Mesta mitogenome, respectively. There are 333 predicted RNA-editing sites in Mesta mitogenome. These include the RNA-editing events that generated the start codon of *nad1* gene and the stop codon of *ccmFC* gene. Phylogenomic analysis with a maximum likelihood method showed that the mitogenome of mangosteen variety Mesta was grouped under Malpighiales order. This is the first complete mitogenome from the *Garcinia* genus for future evolutionary studies.

Keywords: *Garcinia mangostana*; mitogenome; SMRT PacBio; structural variants; RNA editing

Introduction

Mitochondria are the main organelle of energy production for cell sustainability. Besides chloroplast genomes (plastomes), mitochondrial genomes (mitogenomes) serve as important genetic information for phylogenetic and evolutionary studies¹. The first land plant mitogenome sequenced was *Marchantia polymorpha*². Mitogenome sizes are more variable, which range from 66 kb³ to 5.5 Mb⁴ compared to plastomes that are more conserved with the length range from 110 kb to 200 kb⁵. Plant mitogenomes are complex due to the rearrangement, duplication, recombination, and horizontal gene transfer between nucleus and organelles (plastids and mitochondria)⁶⁻⁸.

With the advancement of third-generation sequencing technology such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) which can sequence over 100 kb per molecule and up to 1 million bp, long-read sequencing is becoming more commonly applied in mitogenome assembly^{9,10}. The repetitive regions and rearrangement events that are common in mitogenomes hinder assembly with short-read sequencing¹¹.

Mangosteen (*Garcinia mangostana* L.) is well-known as the “Queen of fruits” with sweet and juicy fruit pulp. It is a tropical fruit under the Clusiaceae family¹² that can be found in Southeast Asia countries¹³. In Malaysia, there is a unique variety Mesta, which is characterized by an oblong shape, thicker mesocarp, and relatively fewer and smaller seeds compared to the common Manggis variety. Despite eight published chloroplast genomes from *Garcinia* species, there is no mitogenome from *Garcinia* species reported to date.

In this study, we assembled a complete mitogenome of Mesta using PacBio data¹⁴ and polished using Illumina short reads¹⁵. We also compared its structure and gene contents with five closely related species from the same order of Malpighiales and another two species from the Brassicales order, namely *Arabidopsis thaliana* and *Carica papaya* (reference used during assembly). The study provides a reference for future evolutionary studies of the *Garcinia* genus.

Results

General features of *Garcinia mangostana* var. Mesta mitogenome

De novo assembly using Organelle_PBA generated two mitogenome contigs with the length of 389,277 bp (scf7180000000010) and 20,340 bp (scf7180000000011), respectively. The smaller contig was the subset of the larger (master) contig (Supplementary Figure S1). For the

larger contig that was circular, manual curation was done by removing one of the identical ends (~18 kb) and a total of 63 bases were added based on the detected variants when short reads were aligned to the trimmed mitogenome (Supplementary Figure S2). The final complete mitogenome of *Garcinia mangostana* var. Mesta was 371,235 bp which was slightly larger than *Arabidopsis thaliana* mitogenome (367,808 bp) but smaller than *Carica papaya* mitogenome (476,890 bp) (Table 1). The average Mesta mitogenome coverage was 129 × using PacBio subreads (Supplementary Figure S3). The mitogenome comprising 29 protein-coding genes, 3 rRNA genes (*rrn5*, *rrn18*, and *rrn26*), and 21 tRNA genes (Figure 1 & Table 2). The total length of protein-coding genes was 28,113 bp, which accounted for 7.6% of the mitogenome. There were only five ribosomal proteins (*rpl5*, *rpl10*, *rpl16*, *rps3*, and *rps4*) found in the Mesta mitogenome.

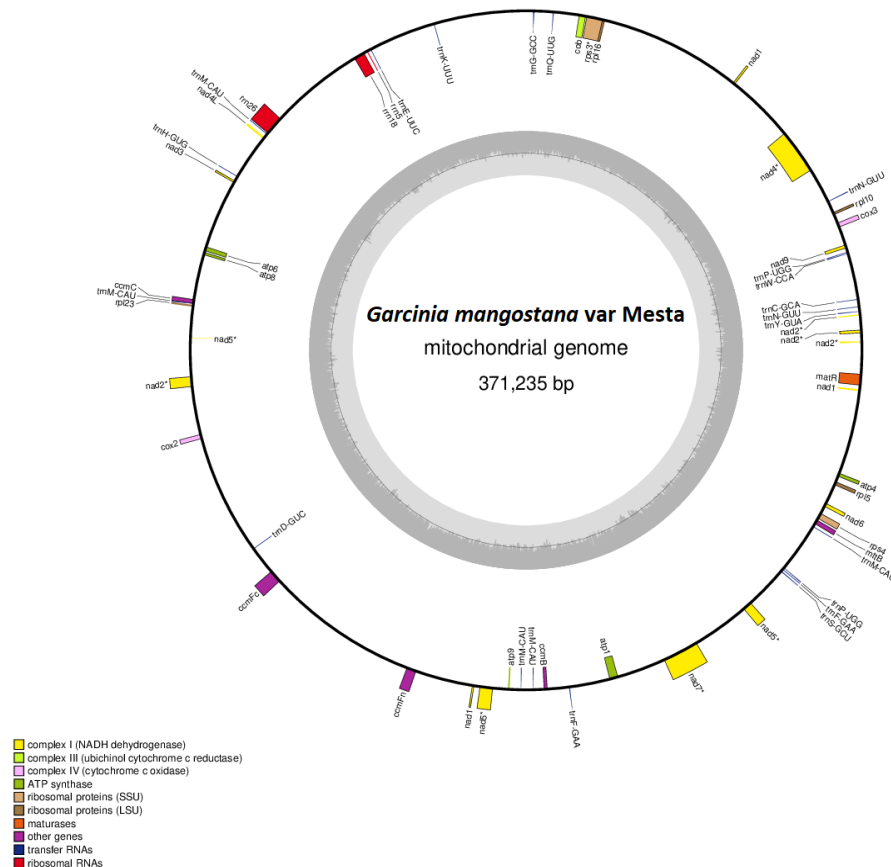


Figure 1. The circular mitochondrial genome of *G. mangostana* variety Mesta. Genes inside the circle are transcribed clockwise while genes outside the circle are transcribed anti-clockwise. The gray bars inside the circle represent the GC content of the sequence. Asterisks (*) indicate genes containing intron(s).

Table 1. Comparison of the gene content in mitogenomes of different species.

Order	Family	Features	Accession number	Size (bp)	GC content (%)	Number of gene	Protein-coding	rRNA	tRNA
Brassicales	Brassicaceae	<i>A. thaliana</i>	BK010421	367,808	44.8	58	33	3	22
	Caricaceae	<i>C. papaya</i>	NC_012116.1	476,890	45.1	61	39	3	19
Malpighiales	Clusiaceae	<i>G. mangostana</i>	OM759996	371,235	43.8	55	29	3	21
	Euphorbiaceae	<i>R. communis</i>	NC_015141.1	502,773	45.0	60	37	3	20
	Salicaceae	<i>P. alba</i>	NC_041085.1	838,420	44.8	58	33	3	22
		<i>P. davidiana</i>	NC_035157.1	779,361	44.8	58	33	3	22
		<i>P. tremula</i>	NC_028096.1	783,442	44.7	58	33	3	22
Passifloraceae	<i>P. edulis</i>	NC_050950.1	680,480	44.7	73	41	3	29	

Table 2. List of genes in *Garcinia mangostana* var. Mesta mitogenome.

Group	Gene
Complex I	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
Complex III	<i>cob</i>
Complex IV	<i>cox1, cox2, cox3</i>
Complex V	<i>atp1, atp4, atp6, atp8, atp9</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFc, ccmFn</i>
Ribosome large subunit	<i>rpl5, rpl10, rpl16</i>
Ribosome small subunit	<i>rps3, rps4,</i>
Intron maturase	<i>matR</i>
SecY-independent transporter	<i>mttB</i>
rRNA genes	<i>rrn5, rrnL, rrnS</i>
tRNA genes	<i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA (X2), trnG-GCC, trnH-GUG, trnK-UUU, trnM-CAU (X5), trnN-GUU (X2), trnP-UGG (X2), trnQ-UUG, trnS- GCU, trnW-CCA, trnY-GUA</i>

Comparison of Mitogenome Gene content of different species

Comparison of the mitogenomes gene content of different species (Supplementary Table S1) showed that all the mitogenomes encoded the basic core set of 24 protein-coding genes. However, the mitogenome of *Passiflora edulis* encoded two copies of the genes *ccmB*, *nad4L*, *nad6*, and *nad7* and four copies of the gene *cox2*. Mitogenome of *Arabidopsis thaliana* also encoded two copies of *atp6* gene with the length of 1158 bp and 1050 bp, respectively. Two copies of *ccmFN* gene were also found in the mitogenome of *Carica papaya*. The *nad1* gene in Mesta was found only consisted of 3 exons instead of 5 as observed in other species despite the total length of its CDS sequence being almost like the other species. Most of the ribosomal proteins were found missing in both Salicaceae and Clusiaceae families from the order Malpighiales. Noticeably, gene *rps12* was missing from Mesta.

Distribution of tRNAs

The 21 tRNAs identified in Mesta mitogenome only code for 14 amino acids (Ser, Phe, Asn, Met, Pro, Gly, Lys, Gln, Tyr, His, Trp, Asp, Glu, Cys). Two out of 21 tRNAs had a chloroplast-origin while the rest were mitochondrial-origin. However, the tRNA genes code for the other six amino acids (Leu, Ile, Thr, Ala, Val, Arg) were not detected. Among the 21 tRNAs, one of them was predicted to have one intron (Supplementary Table S2).

Plastome-derived sequences

There were five plastome-derived sequences with an identity of more than 80% and a sequence length of at least 100 bp (Table 3) found in the mitogenome. This accounted for a total length of 6,214bp which was 1.7% of the mitogenome. The plastome genes contained were *rpl2* (partial), *rpl23*, *trnI-CAU*, *ndhA* (partial), *ndhH*, *rps15* (partial), *atpE* (partial), *atpB*, *rps3* (partial), and *trnD-GUC*. Three of the fragments were found at the plastome large single-copy (LSC) region and one from the single-copy (SSC) region and inverted repeat (IR) regions, respectively.

Table 3. Plastid gene insertions in the mitochondrial genome of Mesta.

Plastid gene	Length (bp)	Position	Identity (%)	Location
<i>rpl2</i> [^] , <i>rpl23</i> , <i>trnI-CAU</i>	1,418	178,888-177,471	97.77	IR
<i>ndhA</i> [^] , <i>ndhH</i> , <i>rps15</i> [^]	2,476	21,314-18,839	97.68	SSC
<i>atpE</i> [^] , <i>atpB</i>	1,882	72,995-71,114	97.54	LSC
<i>rps3</i> [^]	289	310,031-310,319	87.06	LSC
<i>trnD-GUC</i>	149	222,409-222,557	85.81	LSC

[^] partial

Introns and RNA Editing

There were 20 introns distributed among 8 protein-coding genes (*ccmFC*, *cox1*, *nad1*, *nad2*, *nad4*, *nad5*, *nad7*, and *rps3*) in Mesta mitogenome. Among them, *nad1*, *nad2*, and *nad5* genes were trans-spliced. The maximum number of introns found in a gene were four which can be found at certain *nad* gene families such as *nad2*, *nad5*, and *nad7*.

A total of 333 putative RNA-editing sites had been predicted using PREP-MT (Supplementary Table S4). Among the CDS, gene *nad4* contained the highest number of predicted editing sites (35) while there was no editing site found at the *atp9* gene. A total of three annotated genes did not start with the start codon ATG (*atp6*, *nad1*, and *rpl16*) (Supplementary Table S3). Among them, the ACG site at the beginning sequence of *nad1* was putatively predicted to be one of the editing sites that converted it into the ATG start codon. Similarly, early termination was predicted at the gene *ccmFC* sequence which converted the CGA into the stop codon TGA.

Repeat sequences of Mesta mitochondrial DNA

A total of 64 repeat sequences comprising the forward and palindromic repeats were detected. The repeat sequences ranged from 51 bp to 4,172 bp with a total size of 21,356 bp which accounted for 5.8% of the total Mesta mitogenome (Supplementary Table S5). There were

three forward repeat sequences with lengths of 4,172 bp, 2,031 bp, and 1,641 bp. As for the tandem repeats, it constituted 0.15% of the mitogenome (Supplementary Table S6).

Phylogenomic Analysis

A total of 22 (excluding *ccmFn* and *mttB*) out of 24 basic core sets of protein-coding genes were used for phylogenomic analysis, which separated the 15 species into different groups based on the order. Mesta was grouped under Malpighiales together with *P. edulis*, *S. suchowensis*, *S. purpurea*, *P. alba*, *P. tremula*, and *P. davidiana*.

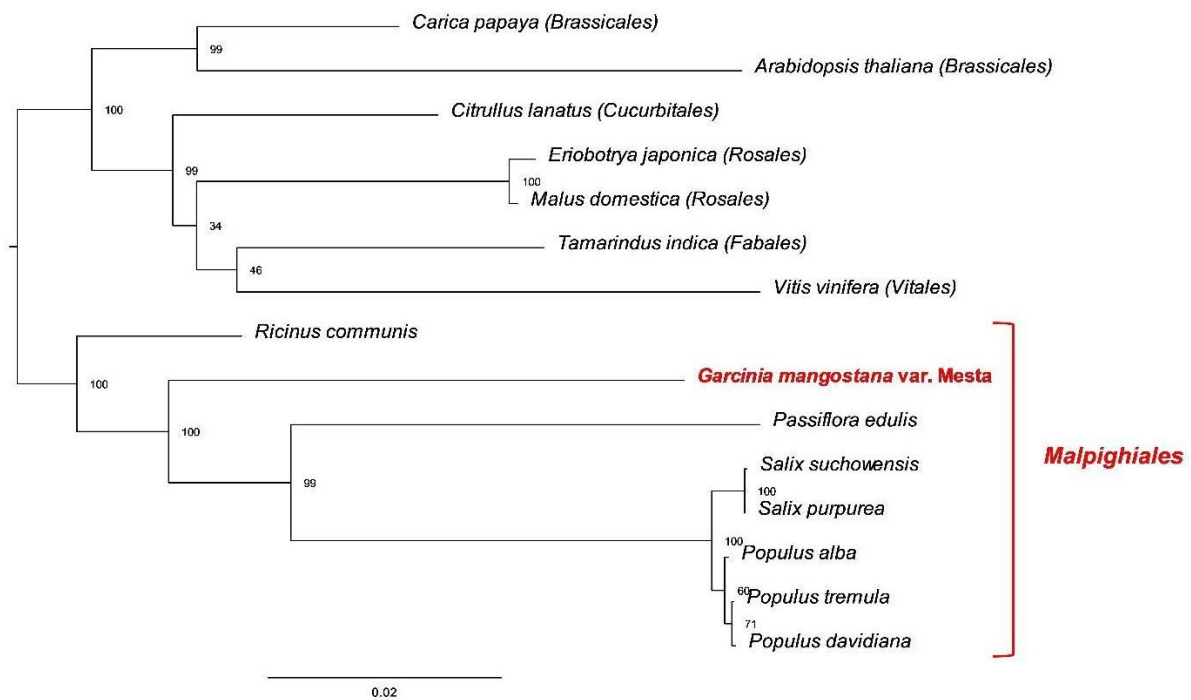


Figure 2. Phylogenomic tree (maximum likelihood) of 15 species based on 22 protein-coding genes. The parentheses and red line indicate the order name.

Discussion

Due to the high complexity of plant mitogenome with large repetitive regions, long-read sequencing is superior in mitogenome assembly^{4,16}. In this study, a total of two Mesta mitogenome contigs were obtained using PacBio data. The shorter contig was a subset of the longer one with the size of 371,235 bp (after manual curation) and was considered as the complete Mesta mitogenome. It is not uncommon to have multiple mitogenome contigs in plants to exist in both circular and linear structures due to intramolecular recombination events^{11,17,18}. For instance, there were 10 contigs in *Fagopyrum esculentum*¹⁹ and 13 contigs in *Picea sitchensis*⁴.

Mesta mitogenome encoded the basic core set of 24 protein-coding genes²⁰ that include genes encoding proteins of the electron transport chain (complex I, III, IV, V, and cytoplasmic membrane proteins) commonly found in plant mitogenome²¹. However, Mesta mitogenome size was relatively small compared to other plant mitogenomes²² from the same order, Malpighiales (Table 1), due to the reduced number of ribosomal proteins and missing genes encoding respiratory chain complex II, *sdh3* and *sdh4* (Supplementary Table S3). These protein-coding genes could be lost during evolution and might be transferred to the nuclear genome as observed in other angiosperm mitogenomes²³⁻²⁵ such as *S. latifolia*²⁶, *S. noctiflora*²⁷, *P. dactylifera*, and *A. indica*^{21,25}. For instance, *rps12* was not found in Mesta mitogenome as well as *Oenothera* and *Zostera marina*^{28,29}.

A complete set of tRNAs coding for 20 amino acids is required for protein translation in plant mitogenomes. However, currently, there was no complete set of tRNA genes found in the mitogenome of angiosperm³⁰. For the Mesta mitogenome, a total of six amino acids encoded by tRNAs (Leu, Ile, Thr, Ala, Val, Arg) were not detected and these tRNAs types were generally reported missing in angiosperm mitogenomes³⁰. In the mitogenome of *S. latifolia*, the majority of the tRNAs were reported lost with only nine types of amino acids²⁶. The loss of tRNAs can be either replaced by tRNAs from the chloroplast or nuclear genome^{26,31}.

There are several factors attributed to the differences of plant mitogenome lengths, including the integration of nuclear and plastid genomes as well as the number and length of non-coding regions²⁰. Chloroplast sequences can be found in plant mitogenomes as there were integration events during evolution. The integration of plastome sequences into mitogenomes can range from 1 to 12%³². For the Mesta mitogenome, the integration rate was 1.7% which was smaller compared to other plants such as watermelon (7.6%)³³, *P. dactylifera* (10.3%)²¹, and *C. pepo* (11.6%)³⁴.

The repetitive regions found in the intergenic regions of mitogenome were in variable types such as short repeats, tandem repeats, and long complex repeats^{21,22,35,36}. The large repeats (>1 kb) might cause homologous recombination and eventually lead to the different configuration of the mitogenomes³⁷. Apart from large repeats, both direct and inverse repeats also contribute to the subgenomic molecules⁷. Repeats detected in Mesta mitogenome using web-based REPuter was low (5.8%) compared to *C. melo* (42.7%)³⁸, *V. vinifera* (6.8%)³⁹, and *N. colorata* (48.89%)⁴⁰ but higher than *P. dactylifera* (2.3%)²¹. Similarly, the tandem repeats detected in Mesta mitogenome were also low which was 0.15% compared to 0.33% in *P. dactylifera*²¹.

RNA-editing events are essential in plant development and stress response⁴¹. The most common RNA editing events in plant organelles (mitochondria and plastids) were the conversion of C-to-U⁴². RNA editing can lead to the start codon/stop codon generation, eliminate premature stop codon, change the splicing site, affect the RNA structure, and cause instability of RNAs⁴¹. It is predicted that RNA-editing events generated the start codon in *nad1* and the stop codon in *ccmFC* genes of Mesta mitogenome. The start codon of *nad1* gene in several species such as *A. alpina*⁴³, *B. stricta*⁴⁴, and *C. rubella*⁴⁵ was also formed by RNA editing. On the other hand, stop codon prediction in *ccmFC* gene sequence had also been reported in *A. thaliana* and *C. bursa-pastoris*¹. The GTG in *rpl16* might be a translation start codon as similar observations were found in maize, *Marchantia*, and *Petunia* mitogenomes⁴⁶.

Conclusion

The complete mitogenome of Mesta was successfully assembled. The Mesta mitogenome length was relatively smaller than other species in the same order due to the loss of the majority of ribosomal proteins and both *sdh* genes. Phylogenetic analysis based on the 22 protein-coding genes among the 15 selected species showed that Mesta was clustered within the Malpighiales order. The mitogenome can serve as a good reference to study the regulation of the mitogenome genes.

Materials and Methods

Mitochondrial Genome Assembly

Genome sequences of Mesta variety were obtained from the NCBI SRA database with the accession numbers SRX2718652 to SRX2718659 for PacBio long-read data (9.5 Gb)¹⁴ and SRX270978 for Illumina short reads (50.2 Gb)¹⁵. CANU v2.0⁴⁷ was used to perform PacBio raw data correction and trimming using default parameters. Next, non-mitogenome reads were removed by sequence alignment of each read against a *Carica papaya* mitogenome (accession no. NC_012116.1) which was used as the reference genome. Then, *de novo* assembly was performed using Organelle_PBA software⁴⁸. Manual curation was performed to obtain the complete Mesta mitogenome.

Genome annotation

Gene annotation was conducted using both GeSeq⁴⁹ and MITOFY web server³⁴. Web-based tRNA-scan v2.0 server (<http://lowelab.ucsc.edu/tRNAscan-SE/index.html>)⁵⁰ was used to

annotate tRNA genes. The physical mitogenome map was generated using Organellar Genome DRAW (OGDRAW v1.3.1) program with default parameters⁵¹.

Identification of plastome derived sequences

Plastome-derived sequences were identified by aligning plastome (accession number: MZ823408) and mitogenome of *G. mangostana* var. Mesta (accession number: OM759996) using NCBI-Nucleotide BLAST (BLASTN) webserver (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) with at least 80% sequence identity and alignment length greater than 100 bp.

Analysis of RNA-Editing and Substitution Rate

Putative RNA-editing sites in protein-coding genes of Mesta mitogenome were predicted using PREP-mt web-based program (<http://prep.unl.edu/>)⁵². The cut-off value was set at 0.6 to obtain an accurate prediction.

Analysis of Repetitive Sequences

The repeat sequences were identified using web-based REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>)⁵³ with a minimum length of repeat size of 50 bp. Web-based tool Tandem Repeats Finder version 4.09⁵⁴ (<https://tandem.bu.edu/trf/trf.basic.submit.html>) was used to identify tandem repeats in Mesta mt using default parameter.

Phylogenomic Analysis

To examine the evolutionary relationship of *G. mangostana* var. Mesta, a maximum likelihood (ML) phylogenomic tree was inferred based on 22 protein-coding genes (Figure 1). The amino acids sequences of these 22 protein-coding genes were concatenated and aligned using MAFFT version 7 online tool (<https://mafft.cbrc.jp/alignment/server/>)⁵⁵. The maximum likelihood (ML) analysis was performed using RAxML-NG v1.0.2 tool⁵⁶ based on the selected model STMTREV+I+G4+F in the ModelTest-NG v0.1.6⁵⁷.

Data Availability Statement

The complete mitogenome sequence of *Garcinia mangostana* var. Mesta has been submitted to GenBank with the accession number OM759996.

References

- 1 O’Conner, S. & Li, L. Mitochondrial Fostering: The Mitochondrial Genome May Play a Role in Plant Orphan Gene Evolution. *Frontiers in Plant Science* **11**, doi:10.3389/fpls.2020.600117 (2020).
- 2 Oda, K. *et al.* Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *Journal of molecular biology* **223**, 1-7 (1992).
- 3 Skippington, E., Barkman, T. J., Rice, D. W. & Palmer, J. D. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences* **112**, E3515-E3524 (2015).
- 4 Jackman, S. D. *et al.* Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. *Genome biology and evolution* **12**, 1174-1179 (2020).
- 5 Sugiura, M. The chloroplast genome. *Essays in biochemistry* **30**, 49-57 (1995).
- 6 Bergthorsson, U., Richardson, A. O., Young, G. J., Goertzen, L. R. & Palmer, J. D. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences* **101**, 17747-17752 (2004).
- 7 Møller, I. M., Rasmusson, A. G. & Van Aken, O. Plant mitochondria—past, present and future. *The Plant Journal* (2021).
- 8 Sanchez-Puerta, M. V. *et al.* Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). *Molecular phylogenetics and evolution* **132**, 243-250 (2019).
- 9 Midha, M. K., Wu, M. & Chiu, K.-P. Long-read sequencing in deciphering human genetics to a greater depth. *Human genetics* **138**, 1201-1215 (2019).
- 10 Xiao, T. & Zhou, W. The third generation sequencing: the advanced approach to genetic diseases. *Transl Pediatr* **9**, 163-173, doi:10.21037/tp.2020.03.06 (2020).
- 11 Sloan, D. B. One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure. *New Phytologist* **200**, 978-985 (2013).
- 12 Nazre, M., Newman, M., Pennington, R. & Middleton, D. Taxonomic revision of *Garcinia* section *Garcinia* (Clusiaceae). *Phytotaxa* **373**, 1-52 (2018).
- 13 Aizat, W. M., Jamil, I. N., Ahmad-Hashim, F. H. & Noor, N. M. Recent updates on metabolite composition and medicinal benefits of mangosteen plant. *PeerJ* **7**, e6324, doi:10.7717/peerj.6324 (2019).
- 14 Midin, M. R. *et al.* SMRT sequencing data for *Garcinia mangostana* L. variety Mesta. *Genomics Data* **12**, 134-135, doi:10.1016/j.gdata.2017.04.003 (2017).
- 15 Abu Bakar, S., Kumar, S., Loke, K. K., Goh, H. H. & Mohd Noor, N. DNA shotgun sequencing analysis of *Garcinia mangostana* L. variety Mesta. *Genomics Data* **12**, 118-119, doi:10.1016/j.gdata.2017.05.001 (2017).
- 16 Omelchenko, D. O. *et al.* Assembly and Analysis of the Complete Mitochondrial Genome of *Capsella bursa-pastoris*. *Plants* **9**, 469 (2020).
- 17 Kozik, A. *et al.* The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS genetics* **15**, e1008373 (2019).
- 18 Mower, J. P., Case, A. L., Floro, E. R. & Willis, J. H. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the

- mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biol Evol* **4**, 670-686, doi:10.1093/gbe/evs042 (2012).
- 19 Logacheva, M. D., Schelkunov, M. I., Fesenko, A. N., Kasianov, A. S. & Penin, A. A. Mitochondrial genome of *Fagopyrum esculentum* and the genetic diversity of extranuclear genomes in buckwheat. *Plants* **9**, 618 (2020).
- 20 Zardoya, R. Recent advances in understanding mitochondrial genome diversity. *F1000Research* **9** (2020).
- 21 Fang, Y. *et al.* A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PloS one* **7**, e37164 (2012).
- 22 Wynn, E. L. & Christensen, A. C. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3: Genes, Genomes, Genetics* **9**, 549-559 (2019).
- 23 Adams, K. L., Qiu, Y.-L., Stoutemyer, M. & Palmer, J. D. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences* **99**, 9905-9912 (2002).
- 24 Asaf, S. *et al.* Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. *PloS one* **11**, e0152937 (2016).
- 25 Choi, K.-S. & Park, S. Complete Plastid and Mitochondrial Genomes of *Aeginetia indica* Reveal Intracellular Gene Transfer (IGT), Horizontal Gene Transfer (HGT), and Cytoplasmic Male Sterility (CMS). *International Journal of Molecular Sciences* **22**, 6143 (2021).
- 26 Sloan, D. B., Alverson, A. J., Štorchová, H., Palmer, J. D. & Taylor, D. R. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evolutionary Biology* **10**, 1-15 (2010).
- 27 Sloan, D. B., MacQueen, A. H., Alverson, A. J., Palmer, J. D. & Taylor, D. R. Extensive Loss of RNA Editing Sites in Rapidly Evolving *Silene* Mitochondrial Genomes: Selection vs. Retroprocessing as the Driving Force. *Genetics* **185**, 1369-1380, doi:10.1534/genetics.110.118000 (2010).
- 28 Grohmann, L., Brennicke, A. & Schuster, W. The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. *Nucleic acids research* **20**, 5641-5646 (1992).
- 29 Petersen, G. *et al.* Mitochondrial genome evolution in Alismatales: Size reduction and extensive loss of ribosomal protein genes. *PLoS One* **12**, e0177606 (2017).
- 30 Michaud, M., Cognat, V., Duchêne, A. M. & Maréchal-Drouard, L. A global picture of tRNA genes in plant genomes. *The Plant Journal* **66**, 80-93 (2011).
- 31 Dietrich, A., Small, I., Cosset, A., Weil, J. & Maréchal-Drouard, L. Editing and import: strategies for providing plant mitochondria with a complete set of functional transfer RNAs. *Biochimie* **78**, 518-529 (1996).
- 32 Mower, J. P., Sloan, D. B. & Alverson, A. J. Plant mitochondrial genome diversity: the genomics revolution. *Plant Genome Diversity Volume 1*, 123-144 (2012).
- 33 Cui, H. *et al.* Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Scientific reports* **11**, 1-9 (2021).
- 34 Alverson, A. J. *et al.* Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* **27**, 1436-1448, doi:10.1093/molbev/msq029 (2010).

- 35 Alverson, A. J., Zhuo, S., Rice, D. W., Sloan, D. B. & Palmer, J. D. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS one* **6**, e16404 (2011).
- 36 Figueroa-Martinez, F., Nedelcu, A. M., Smith, D. R. & Reyes-Prieto, A. The plastid genome of *Polytoma uvella* is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to nonphotosynthetic lifestyles. *Plant physiology* **173**, 932-943 (2017).
- 37 Gualberto, J. M. *et al.* The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**, 107-120 (2014).
- 38 Rodríguez-Moreno, L. *et al.* Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC genomics* **12**, 1-14 (2011).
- 39 Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular biology and evolution* **26**, 99-110 (2009).
- 40 Dong, S. *et al.* The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics* **19**, 614, doi:10.1186/s12864-018-4991-4 (2018).
- 41 Tang, W. & Luo, C. Molecular and functional diversity of RNA editing in plant mitochondria. *Molecular biotechnology* **60**, 935-945 (2018).
- 42 Hao, W. *et al.* RNA Editing and Its Roles in Plant Organelles. *Frontiers in Genetics*, 1747 (2021).
- 43 Xu, Y. & Bi, C. The complete mitochondrial genome sequence of an alpine plant *Arabis alpina*. *Mitochondrial DNA Part B* **3**, 725-727 (2018).
- 44 Li, J., Bi, C., Tu, J. & Lu, Z. The complete mitochondrial genome sequence of *Boechea stricta*. *Mitochondrial DNA Part B* **3**, 896-897 (2018).
- 45 Lin, H. & Bai, D. The complete mitochondrial genome of a highly selfing species *Capsella rubella*. *Mitochondrial DNA Part B* **4**, 1907-1908 (2019).
- 46 Bock, H., Brennicke, A. & Schuster, W. *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria? *Plant Mol Biol* **24**, 811-818, doi:10.1007/bf00029863 (1994).
- 47 Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).
- 48 Soorni, A., Haak, D., Zaitlin, D. & Bombarely, A. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC genomics* **18**, 49-49, doi:10.1186/s12864-016-3412-9 (2017).
- 49 Tillich, M. *et al.* GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* **45**, W6-w11, doi:10.1093/nar/gkx391 (2017).
- 50 Chan, P. P. & Lowe, T. M. in *Gene prediction* 1-14 (Springer, 2019).
- 51 Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**, W59-W64, doi:10.1093/nar/gkz238 (2019).
- 52 Mower, J. P. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Research* **37**, W253-W259, doi:10.1093/nar/gkp337 (2009).
- 53 Kurtz, S. *et al.* REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research* **29**, 4633-4642, doi:10.1093/nar/29.22.4633 (2001).

- 54 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-580, doi:10.1093/nar/27.2.573 (1999).
- 55 Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics* **20**, 1160-1166 (2019).
- 56 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455, doi:10.1093/bioinformatics/btz305 (2019).
- 57 Darriba, D. *et al.* ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution* **37**, 291-294, doi:10.1093/molbev/msz189 (2020).

Acknowledgments

We would like to acknowledge the support of this research by Universiti Kebangsaan Malaysia (UKM) Research University grant AP-2012-018. The group is currently supported by UKM Research University grant DIP-2020-005 (H-HG) and NIG-JOINT grant 2021 (2A2021) (YN and H-HG), Japan.

Author Contribution Statement

C.C.W. and H.H.G. conceived and planned the experiment. C.C.W. wrote the paper and analyzed the data. Y.N. contributed to the data analysis. H.H.G., N.A.N.M, V.K.S., M.A., and Y.N. reviewed and edited the manuscript. All authors read and approved the manuscript.

Additional information

Supplementary Tables: Figure S1-S3

Supplementary Figures: Table S1-S6

Conflict of Interest

The authors declare no competing interests.

Legends

Figures

Figure 1. The circular mitochondrial genome of *G. mangostana* variety Mesta. Genes inside the circle are transcribed clockwise while genes outside the circle are transcribed anti-clockwise as indicated by the gray arrows. The gray bars inside the circle represent the GC content of the sequence. Asterisks (*) indicate genes containing intron(s).

Figure 2. Phylogenomic tree (maximum likelihood) of 15 species based on 22 protein-coding genes. The parentheses and red line indicate the order name.

Tables

Table 1. Comparison of the gene content in mitogenomes of different species.

Table 2. List of genes in *Garcinia mangostana* var. Mesta mitogenome.

Table 3. Plastid gene insertions in the mitochondrial genome of Mesta.