



EVOLUTIONARY GENETICS IN *Salacca* SPECIES BASED ON CHLOROPLAST GENOME, GENES AND INTERGENIC SEQUENCE DIVERSITIES

ARSLAN ARSHAD



PLANT BREEDING AND BIOTECHNOLOGY STUDY PROGRAM
FACULTY OF AGRICULTURE
INSTITUT PERTANIAN BOGOR
BOGOR
2024

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

STATEMENT REGARDING THE THESIS AND SOURCES OF INFORMATION AND ASSIGNMENT OF COPYRIGHT

I hereby declare that the research proposal entitled "Evolutionary Genetics in *Salacca* Species Based on Chloroplast Genome, Genes and Intergenic Sequence Diversities" is my work with the direction of the supervisory commission and has not been submitted in any form to any university. Sources of information originating or quoted from published or unpublished works of other authors have been mentioned in the text and included in the Bibliography at the end of this thesis.

I hereby assign the copyright of my written work to the Bogor Agricultural Institute.

Bogor, July 2024

Arslan Arshad
A2503222801

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

Arslan Arshad. Genetika Evolusioner pada Spesies *Salacca* Berdasarkan Genom Kloroplas, Gen, dan Keanekaragaman Urutan Intergenik Supervised by Prof. Dr. Ir. Sudarsono, M.Sc., Dr. Megayani Sri Rahayu, M.S. and Dr. Azis Natawijaya, SP., M.Sc.

Tesis ini mengeksplorasi genom kloroplas (Cp) dan keragaman urutan di wilayah genomik utama pada tujuh spesies *Salacca*, yang dikenal sebagai buah salak. Analisis awal melibatkan konstruksi dan pemeriksaan genom Cp menggunakan data dari platform BGISEq-500 dan assembler GetOrganelle, yang mengungkapkan struktur yang konsisten di seluruh genom dengan panjang bervariasi antara 157.047 dan 158.182 pasang kilobase. Analisis genomik komparatif yang mendetail mengidentifikasi gen *ycf1* sebagai gen dengan jumlah polimorfisme nukleotida tunggal (SNP) tertinggi, dengan temuan signifikan seperti asam amino yang hilang pada *Salacca affinis* dan prevalensi motif SSR mononukleotida. Selain itu, terdapat variasi pada pertemuan wilayah salinan tunggal besar (LSC), pengulangan terbalik (IR), dan salinan tunggal kecil (SSC), terutama pada *S. wallichiana*. Studi ini juga menyelidiki keragaman urutan pada wilayah *rbcL*, *matK*, *ITS*, *trnL-F*, dan *atpH-F*, menggunakan sekuensing Sanger dan primer khusus gen untuk mengatasi tantangan identifikasi spesies akibat kesamaan morfologis dan taksonomi yang kompleks. Pengembangan 26 primer Single Nucleotide Amplified Polymorphism (SNAP) menunjukkan spesifisitas dan efisiensi yang tinggi, memfasilitasi identifikasi spesies secara tepat. Analisis filogenetik mengungkapkan bahwa spesies *Salacca* membentuk klad yang berbeda dengan nenek moyang bersama yang baru-baru ini, menekankan hubungan evolusioner yang dekat. Penelitian ini menyoroti kegunaan informasi genom kloroplas dan wilayah genomik spesifik dalam meningkatkan pemahaman kita tentang keragaman genetik, identifikasi spesies, biologi evolusioner, dan upaya konservasi dalam genus *Salacca*. Temuan ini menegaskan adanya hubungan genetik antara *Salacca* dan *Cocos nucifera* serta menyoroti perbedaan dengan *Phoenix dactylifera*, memberikan wawasan penting untuk studi dan program pemuliaan di masa depan.

Kata Kunci: Analisis filogenetik, anotasi genom, DNA Barcoding, GetOrganelle, identifikasi spesies, SNPs



SUMMARY

Arslan Arshad. "Evolutionary Genetics in *Salacca* Species Based on Chloroplast Genome, Genes and Intergenic Sequence Diversities" Supervised by Prof. Dr. Ir. Sudarsono, M.Sc., Dr. Megayani Sri Rahayu, M.S. and Dr. Azis Natawijaya, SP., M.Sc.

This thesis explores the chloroplast genomes (Cp) and sequence diversity of key genomic regions in seven *Salacca* species, commonly known as snake fruit. The initial analysis involves constructing and examining Cp genomes using data from the BGISEq-500 platform and GetOrganelle assemblers, revealing a consistent structure across genomes with lengths varying between 157,047 and 158,182 kilobase pairs. A detailed comparative genomics analysis identifies the *ccf1* gene as having the highest number of single nucleotide polymorphisms (SNPs), with significant findings such as missing amino acids in *Salacca affinis* and a prevalence of mononucleotide SSR motifs. Additionally, variability in the junctions of large single copy (LSC), inverted repeat (IR), and small single copy (SSC) regions is noted, particularly in *S. wallichiana*. The study further investigates the sequence diversity of the *rbcL*, *matK*, *ITS*, *trnL-F*, and *atpH-F* regions, employing Sanger sequencing and gene-specific primers to address species identification challenges due to morphological similarities and complex taxonomy. The development of 26 Single Nucleotide Amplified Polymorphism (SNAP) primers shows high specificity and efficiency, facilitating precise species identification. Phylogenetic analysis reveals that the *Salacca* species form distinct clades with a recent common ancestor, emphasizing their close evolutionary relationships. This research highlights the utility of Cp genome information and specific genomic regions in advancing our understanding of genetic diversity, species identification, evolutionary biology, and conservation efforts within the *Salacca* genus. The findings establish a genetic connection between *Salacca* and *Cocos nucifera* and highlight the contrast with *Phoenix dactylifera*, providing crucial insights for future studies and breeding programs.

Keywords: Analisis Filogenetik, Anotasi Genom, DNA Barcoding, GetOrganelle, Identifikasi Spesies, SNPs

Hak cipta milik IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Copyright owned by IPB, 2024
Copyright protected by law

It is prohibited to quote part or all of this written work without citing or citing the source. Quotations are only for educational purposes, research, writing scientific papers, preparing reports, writing criticism, or reviewing a problem and these quotations do not harm the interests of IPB.

It is prohibited to publish or reproduce part or all of this written work in any form without IPB permission.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

EVOLUTIONARY GENETICS IN *Salacca* SPECIES BASED ON CHLOROPLAST GENOME, GENES AND INTERGENIC SEQUENCE DIVERSITIES

ARSLAN ARSHAD

Thesis
As part of the requirements for obtaining a Master's degree
in
Plant Breeding and Biotechnology Study Program

**PLANT BREEDING AND BIOTECHNOLOGY STUDY PROGRAM
FACULTY OF AGRICULTURE
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

@Hak cipta milik IPB University

IPB University



IPB University
Bogor Indonesia

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



@Hak cipta milik IPB University

Examiner's Team for Thesis Examination:

1. Dr. Ir. Diny Dinarti M.Si

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Thesis Title : Evolutionary Genetics in *Salacca* Species Based on Chloroplast Genome, Genes and Intergenic Sequence Diversities
Name : Arslan Arshad
NIM : A2503222801

Approved by

Supervisor 1:
Prof. Dr. Ir. Sudarsono, M.Sc.



Supervisor 2:
Dr. Ir. Megayani Sri Rahayu, M.S.

Supervisor 3:
Dr. Azis Natawijaya, SP., M.Sc.

Acknowledge by:

Head of the Study Program:
Prof. Dewi Sukma, S.P., M.Si.
NIP 19700404199702001

Dean of Faculty of Agriculture:
Prof. Dr. Ir. Suryo Wiyono, M.Sc.Agr.
NIP 196902121992031003



Exam Date: 4th July, 2024.

Graduation Date: 11 JUL 2024

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



FOREWORD

Praise and gratitude to the Almighty God who always gives me grace, strength, and guidance on my Evolutionary Genetics in *Salacca* Species Based on Chloroplast Genome, Genes and Intergenic Sequence Diversities thesis. This thesis is one of the requirements for obtaining a master's degree, and it was conducted from November 2022 to May 2024. The author would like to thank his supervisors, Prof. Dr. Ir. Sudarsono, M.Sc., Dr. Megayani Sri Rahayu, M.S., and Dr. Azis Natawijaya, SP., M.Sc. for their direction and unwavering support throughout this research. The author would also like to express his sincere gratitude to the teaching staff at the Plant Breeding and Biotechnology Study Program, Faculty of Agriculture, Bogor Agricultural Institute, who have provided much direction and guidance in this thesis.

Thank you to my parents, Mr. and Mrs. Muhammad Arshad, my brother Usman Arshad and Senior Redi Aditama, and friends who have supported me throughout my studies.

Bogor, July 2024

Arslan Arshad
A2503222801



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

TABLE OF CONTENTS

TABLE OF CONTENTS	xiii
LIST OF TABELS	xiv
LIST OF FIGURES	xiv
LIST OF ATTACHMENTS	xv
I INTRODUCTION	1
1.1 Background	1
1.2 Statement of Problem	2
1.3 Research Objectives	3
1.4 The Scope of Research	3
1.5 Benefits of the Study	3
1.6 Hypothesis	3
1.7 Outline of research and experimental flow	4
II LITERATURE REVIEW	5
2.1 <i>Salacca zalacca</i>	5
2.2 Chloroplast Genome Structure	6
2.3 Evolutionary studies based on Chloroplast Genome	6
2.4 DNA Barcoding	6
2.5 Single Nucleotide Polymorphism (SNP)	7
2.6 SNP-Based Molecular Marker	7
III UNVEILING THE DIVERSITY: COMPARATIVE STUDY OF CHLOROPLAST GENOMES ACROSS SEVEN <i>Salacca</i> SPECIES	8
3.1 Abstract	8
3.2 Introduction	8
3.3 Materials and Methods	10
3.4 Results and Discussion	12
3.5 Conclusion	29
IV SEQUENCE DIVERSITY OF <i>rbcL</i>, <i>matK</i>, <i>ITS</i>, <i>trnL-F</i> and <i>atpH-F</i> ACROSS <i>Salacca</i> SPECIES REVEAL THEIR USE IN DIVERSITY ANALYSIS	30
4.1 Abstract	30
4.2 Introduction	30
4.3 Material and Methods	31
4.4 Result and Discussion	34
4.5 Conclusion	45
V CONCLUSION AND SUGGESTION	46
5.1 Conclusion	46
5.2 Suggestions	46
BIBLIOGRAPHY	47
BIOGRAPHY	65



LIST OF TABLES

1. <i>Salacca</i> species and their corresponding accession numbers, SRA data size, and remarks for the next generation sequences. The listed <i>Salacca</i> species represent the six largest sizes of <i>Salacca</i> genome SRA downloaded from the NCBI	13
2. Features of the chloroplast genome of seven <i>Salacca</i> species assembled from the downloaded whole-genome short-read archived (SRA) from the National Center for Biotechnology Information (NCBI) SRA Database	15
3. Gene counts in the Cp genomes of various <i>Salacca</i> species	15
4. Gene duplication in subunit gene IDs across various <i>Salacca</i> species functional genes with two or more physical copies in the cp genome of seven <i>Salacca</i> species	16
5. Category and group of genes with gene products and copy numbers across the <i>Salacca</i> species	18
6. Relative synonymous codon usage (RSCU) recognition pattern of all <i>Salacca</i> Cp genome	21
7. Number of codon usage recognition patterns in <i>Salacca</i> chloroplast genome	22
8. Quantitative examination of the proportion of nucleotides in various structural parts of the Cp genomes of seven <i>Salacca</i> species	23
9. Quantitative examination of the frequency of nucleotides in various structural parts of the Cp genomes of seven <i>Salacca</i> species	23
10. Non-synonymous and synonymous SNPs in each Gene of <i>Salacca</i> Cp genome	24
11. The frequency of chloroplast SSR in the coding (Genic CpSSR), and non-coding regions (Intergenic Cp SSR) Of Seven <i>Salacca</i> Species	27
12. <i>Salacca</i> Species used for Sanger Sequencing	31
13. Gene-specific primer based on assembled Cp <i>Salacca</i> genomes	34
14. Identified single nucleotide polymorphism (SNP) sites in <i>Salacca</i> nucleotide sequences based on sanger sequencing results	38
15. List of SNAP primers based on the Sanger Sequencing analysis results from WASP	40

LIST OF FIGURES

1. Outline of research and experimental flowchart	4
2. The circular diagram depicts <i>S. sumatrana's</i> Cp genome, illustrating gene transcription in clockwise and counterclockwise directions and encompassing functional groups like photosystems, cytochrome b/f complex, ATP synthase, and RNA polymerase	14
3. Percentage of SNP and InDel in Cp genomes of <i>Salacca</i>	25
4. Seven-chloroplast genomes IR/LSC and IR/SSC border locations are compared, including <i>S. zalacca</i> , <i>S. sumatrana</i> , <i>S. ramosiana</i> , <i>S. affinis</i> , <i>S. glabrescens</i> , <i>S. secunda</i> and <i>S. wallichiana</i>	26
5. Phylogenetic tree of Cp genome	28

6. Gradient PCR optimization results show optimal annealing temperatures for amplifying Cp genome regions in <i>Salacca</i> species.	35
7. Un-purified PCR product before sending to GSI (Genetika Service Inc.)	35
8. Agarose quantification report	36
9. Primer validation using Geneious Prime	43
10. Phylogenetic relationships of species within the <i>Arecaceae</i> family inferred from five gene sequences of 25 palm accessions and one outgroup (<i>Arabidopsis thaliana</i>) using the Neighbor-Joining method. The bootstrap support values were obtained using 1000 iterations	44

LIST OF ATTACHMENTS

1. Gene-wise Phylogenetic Trees <i>atpH-F</i> (red color shows Sanger results , blue: assembled Cp genome, green: NCBI <i>Salacca</i> accession, brown: outgroup)	59
2. Gene-wise Phylogenetic Trees <i>ITS</i> (red color shows Sanger results , blue: assembled Cp genome, green: NCBI <i>Salacca</i> accession, brown: outgroup)	60
3. Gene-wise phylogenetic trees <i>matK</i> (red color shows Sanger results , blue: assembled Cp genome, green: NCBI <i>Salacca</i> accession, brown: outgroup)	61
4. Gene-wise phylogenetic trees <i>rbcL</i> (red color shows Sanger results , blue: assembled Cp genome, green: NCBI <i>Salacca</i> accession, brown: outgroup).	62
5. Gene-wise phylogenetic trees <i>trnL-F</i> (red color shows Sanger results , blue: assembled Cp genome, green: NCBI <i>Salacca</i> accession, brown: outgroup)	63
6. Genes encoded by the <i>Salacca sumatrana</i> Cp genome	64