Floral Diversity and Genome Size Estimation of Flowering Plants from IIT Guwahati – Plastome Mining of Selected Small Genome Sized Plants

A thesis submitted by

Alok Senapati

For the award of the degree

of

Doctor of Philosophy

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Assam, India- 781039

Dec 2023

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biosciences and Bioengineering
Guwahati – 781039

STATEMENT

I do hereby declare that the matter embodied in this thesis entitled **"Floral Diversity and Genome Size Estimation of Flowering Plants from IIT Guwahati – Plastome Mining of Selected Small Genome Sized Plants"** is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India, under the guidance of Prof. Latha Rangan. In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators.

Alok Senapati

June, 2024

Alok Senapati
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Assam India- 781039

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biosciences and Bioengineering
Guwahati – 781039

CERTIFICATE

It is certified that the work described in this thesis, entitled **"Floral Diversity and Genome Size Estimation of Flowering Plants from IIT Guwahati – Plastome Mining of Selected Small Genome Sized Plants"**, done by Mr. Alok Senapati for the award of the degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India, and this work has not been submitted elsewhere for a degree.

Latha

June, 2024

Prof. Latha Rangan
(Thesis Supervisor)
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Assam India- 781039

ACKNOWLEDGEMENT

I extend my profound gratitude to the individuals who played pivotal roles in guiding and supporting me throughout my doctoral journey. Foremost, I express my sincere thanks to Prof. Latha Rangan for her unwavering scientific and moral guidance as my mentor. My deepest appreciation goes to the esteemed members of the doctoral committee – Prof. Shankar Prasad Kanaujia, Prof. Manish Kumar, and Prof. Chandan Das – for their invaluable suggestions, practical advice, and timely guidance that significantly contributed to the completion of my thesis.

I am also indebted to Prof. Sudip Mitra for his constructive insights and assistance. I extend my heartfelt gratitude and utmost respect to the Head of the Department of Biological Sciences and Bioengineering (BSBE) at IIT Guwahati. My sincere thanks to all the dedicated staff members of the Department of BSBE who supported me in every possible way. The timely assistance and cooperation from the staff at the Central Instruments Facility (CIF) and IITG are duly acknowledged.

I am immensely grateful to the Ministry of Education, Government of India, for their crucial financial assistance during my Ph.D. term, without which this journey would not have been possible.

Special appreciation is reserved for my friends and seniors – Dr. Rahul G. Shelke, Dr. Anuma Singh, Dr. Reshmi Das, Dr. Ishani Chakraborty, Dr. Nihal Gujre, Dr. S. Sanjana, Dr. Manish Kumar Gupta, Dr. Gaurav Bhatt, and Ms. Hiramoni Boro – for their invaluable cooperation and unwavering support. I extend my heartfelt thanks to Mr. Bimal K. Chhetri for his invaluable cooperation throughout the entire process.

The members of the ABL (Applied Biodiversity Lab) – Ms. Nuzelu, Ms. Anamika, Ms. Rashmi Singh, Ms. Rubeka, and Mr. Sonu – have significantly contributed to both my personal and professional development during my tenure at IITG. The group has been not only a stress reliever but also a source of valuable information and collaboration.

My deepest appreciation is reserved for my parents, who granted me the opportunity to explore this fascinating world and validate my scientific endeavors. The prayers and blessings of my parents have been instrumental in helping me overcome life's challenges. I also express my heartfelt gratitude to my sister, brother, and friends for their unwavering moral support, constant encouragement, and unwavering faith in me.

A personal note of gratitude, love, and respect is extended to everyone who has supported me in my Ph.D. thesis work, whether directly or indirectly. Finally, I acknowledge the unwavering support of God Almighty throughout my time at IIT Guwahati.

June, 2024 **Alok Senapati**

CONTENTS

Abstract i

Abbreviations ii

Units iv

List of figures v

List of tables x

CHAPTER 1: INTRODUCTION 1-4

CHAPTER 2: REVIEW OF LITERATURE 5-34

2.1. Floral diversity of IIT Guwahati campus 5

- 2.1.1. Introduction 5
- 2.1.2. Floral diversity of the world 5
- 2.1.3. Floristic diversity of India and NE India 6
- 2.1.4. Benefits of floral diversity 8
- 2.1.5. Floristic studies and inventories 9
- 2.1.6. Endemic, rare and endangered species of NE India 10
- 2.1.7. Medicinal and ethnobotanical importance 10
- 2.1.8. Threats to flora diversity and conservation efforts 11
- 2.1.9. Knowledge gaps and conclusion 12

2.2. Genome size estimation 13

- 2.2.1. Introduction 13
- 2.2.2. Genome size and its relevance 13
- 2.2.3. Various approaches for estimating genome size 15
- 2.2.4. Flowcytometry its advantages and limitations 16
- 2.2.5. History of plant nuclei isolation buffers 17
- 2.2.6. Standardization methods and plant reference standards 19
- 2.2.7. Genome size estimation in angiosperms 20
- 2.2.8. Genome size variation and influencing factors 22
- 2.2.9. The impact of genome size on plant traits 23
- 2.2.10. Challenges and future prospects 24

2.3. Plastome mining of selected small genome sized plants 25

- 2.3.1. Introduction 25
- 2.3.2. Plastome mining and its relevance 25
- 2.3.3. Plastome structure and function 26

2.3.4. Small genome sized plants and their advantages 28
2.3.5. Plastome sequencing and annotation techniques 30
2.3.6. Plastome sequencing and phylogenetics 31
2.3.7. Applications of plastome mining 32
2.3.8. Challenges and future prospects 33

CHAPTER 3: FLORAL DIVERSITY OF IIT GUWAHATI CAMPUS 35-58

3.1. Introduction 35
3.1.1. NE India- a paradise of floral diversity 36
3.1.2. Climate and forest types 36
3.1.3. Importance of floral diversity 37
3.1.4. Importance of community composition 37
3.1.5. Importance of species distribution 37
3.1.6. Novelty of this research 38

3.2. Materials and methods 38
3.2.1. Study area 38
3.2.2. Sample collection 39
3.2.2.1. Sample selection 39
3.2.2.2. Sample collection 39
3.2.3. Data collection 39
3.2.3.1. Sampling methodology 39
3.2.3.2. Data recording 40
3.2.3.3. Plant species identification 40
3.2.3.4. Data variables 40
3.2.4. Data analysis 40
3.2.4.1. Biodiversity indices 40
3.2.4.2. Community indices 41
3.2.5. Ethical considerations 42
3.2.5.1. Permits and permissions 42
3.2.5.2. Ethical guidelines 42
3.2.6. Data quality assurance 42

3.3. Results and discussion 43
3.3.1. IIT Guwahati topography and vegetation 43
3.3.2. Floristic analysis 44
3.3.2.1. Floral diversity 44
3.3.2.2. Identification and documentation 45

3.3.2.3. Exotics and indigenous species 47

3.3.2.4. Common and rare species of the campus 48

3.3.2.5. Species utilitarian value 49

3.3.2.6. Economic plants 51

3.3.3. Familial analysis of the flora 53

3.3.4. Diversity analysis 55

3.4. Conclusion 58

CHAPTER 4: GENOME SIZE ESTIMATION 60-95

4.1. Introduction 60

4.1.1. Background and significance of GS in plants 60

4.1.2. GS diversity in land plants 61

4.1.3. Methodological advances in GS estimation 62

4.1.4. Scope of the study 62

4.2. Materials and methods 63

4.2.1. Selection of plants 63

4.2.2. Sample collection 63

4.2.3. Nuclei isolation buffer 63

4.2.4. Sample preparation 65

4.2.5. Standardization methods 66

4.2.6. Flow cytometric analysis 67

4.2.6.1. Instrument calibration 67

4.2.6.2. Reference standard plants 69

4.2.7. Ethical considerations and quality control 70

4.2.8. Data analysis 71

4.2.9. Correlation among genetic and morphological traits 72

4.2.9.1. Linear correlation test 72

4.2.9.2. Regression analysis 72

4.2.9.3. Artificial Neural Network model 72

4.2.10. Statistical analysis 73

4.3. Results and discussion 73

4.3.1. Sample collection and preparation 73

4.3.2. Buffer selection and composition- modified hypotonic PI buffer 74

4.3.3. Choice of reference standards and standardization methods 76

4.3.4. Flow cytometric analysis 78

4.3.5. Genome size estimation 80

4.3.5.1. GS estimation of herbs 81

4.3.5.2. GS estimation of shrubs 82

4.3.5.3. GS estimation of trees 84

4.3.6. Genome size variation 87

4.3.7. Correlation between GS to various plant traits 89

4.3.7.1. Linear correlation 90

4.3.7.2. Regression analysis 91

4.3.7.3. ANN models 93

4.4. Conclusion 94

CHAPTER 5: PLASTOME MINING OF SELECTED SMALL GENOME SIZED PLANTS 96-156

5.1. Introduction 96

5.1.1. Significance of plastome mining 96

5.1.2. Rationale for studying small genome-sized plants 96

5.1.3. Chloroplast genes and their applications 97

5.1.4. Overview of methodology 97

5.1.5. Anticipated contributions and outcomes 98

5.2. Materials and methods 98

5.2.1. Selection of plants 98

5.2.2. Physical characteristics 98

5.2.3. Sample collection 99

5.2.4. Plastome sequencing and mining of molecular information 99

5.2.4.1. Genome sequencing and library preparation 99

5.2.4.2. Genome assembly and annotation 100

5.2.4.3. Structural comparison 100

5.2.4.4. Tandem and simple sequence repeats 101

5.2.4.5. DNA diversity (Pi) 101

5.2.4.6. Nucleotide substitution rate (Ka/Ks) 102

5.2.4.7. Synonymous codon usage bias 102

5.2.4.8. Phylogenetic analysis 102

5.3. Results and discussion 103

5.3.1. Physical characterization 103

5.3.2. Plastome mining: unveiling molecular insights 104

5.3.2.1. Genome assembly and annotation 105

5.3.2.2. Structural comparison 114

5.3.2.3. Tandem and simple sequence repeats 125
5.3.2.4. DNA diversity analysis (*Pi)* 137
5.3.2.5. Nucleotide substitution rate (Ka/Ks) 141
5.3.2.6. Codon usage bias 147
5.3.2.7. Phylogenetic analysis 154
5.4. Conclusion 158
CHAPTER 6: SUMMARY AND FUTURE PROSPECTS 160-163
6.1. Summary 160
6.2. Future prospects 159
REFERENCES 164-202
LIST OF PUBLICATIONS 203-207
ANNEXURES 208-232

ABSTRACT

Flowering plants exhibit a remarkable diversity with high endemism in the Northeastern (NE) region of India. In this study, we explored the floral diversity and estimated the genome sizes (GS) of selected species found within the campus of the Indian Institute of Technology (IIT) Guwahati. A comprehensive survey of the campus was conducted to identify and collect flowering plant specimens. A total of 434 species were identified from 101 flowering plant families. Our findings revealed a rich floral diversity within the IIT Guwahati campus, which includes 172 herbs, 98 shrubs, 82 trees, 56 aquatic and 26 climbers. Plants were also categorized under medicinal (296 spp.), ornamental (47 spp.), exotic (178 spp.), and 256 native plants among the studied species. The genome size of 110 identified species was estimated using flow cytometry, which included 58 new estimates. Additionally, the genome size estimations shed light on the genetic characteristics and evolutionary trajectories of these plants, potentially contributing to our understanding of their ecological adaptations. Five small genome sized (≤ 2 pg) plants were selected and high-throughput sequencing techniques were employed to obtain plastid genomes (plastomes) of the selected species. The plastome mining approach facilitated the assembly and annotation of complete plastomes for the targeted plant species. These plastomes were used for comparative analysis of tandem repeats, simple sequence repeats (SSRs), nucleotide diversity (*Pi*), nucleotide substitution rate (Ka/Ks ratio), and synonymous codon usage bias among related species. The results of this study serve as a valuable resource for future research on the floral biology, conservation, and evolutionary biology of flowering plants within the region. Furthermore, the plastome sequences generated in this study can be utilized for phylogenetic analyses, comparative genomics, and further investigations into the functional genomics of these small genome sized plants.

Keywords: Floral diversity, Flow cytometry, Genome size, IIT Guwahati, Plastome

ABBREVIATIONS

GS	Genome size
NE	Northeast
MSL	Mean sea level
FCM	Flow cytometry
Sp.	Species (singular)
Spp.	Species (plural)
PVP	Polyvinylpyrrolidone
BME	Betamercaptoethanol
PI	Propidium iodide
LMA	Leaf mass per unit area
LSC	Large single copy
SSC	Small single copy
IR	Inverted repeat
SSR	Simple sequence repeat
NGS	Next-generation sequencing
IMD	Indian meteorological department
BSI	Botanical survey of India
var.	Variety
D	Simpson's Diversity Index
H′	Shannon-Weiner Index
J	Jaccard’s similarity Index
E	Evenness
d	Species richness
qPCR	Quantitative real-time PCR
WGS	Whole genome sequencing
MHPI	Modified hypotonic propidium iodide
n	Haploid chromosome number
ANN	Artificial neural network
hn	Hidden neurons

CV	coefficient of variation
SD	standard deviation
FSC/SSC	Forward/Side scatter
cp	Chloroplast
RSCU	Relative Synonymous Codon Usage
PCG	Protein coding genes
LCB	Local collinear blocks
cpDNA	Chloroplast DNA
GC1	GC at first position of codon
GC2	GC at second position of codon
GC3	GC at third position of codon
MAS	Marker assisted selection

UNITS

°C	Degree Celsius
pg	Picogram
kb	Kilobasepairs
Mb	Megabases
mM	Millimolar
w/v	Weight/Volume
v/v	Volume/Volume
Km2	Square kilometer
mg	Milligram
µg	Microgram
nm	Nanometer

LIST OF FIGURES

Fig. No	Figure description	Page
2.1	State wise area of Brahmaputra River basin (Source: Ministry of Water Resource, 2014)	7
2.2	Genome size (1C-value) variability by lineage in the plant DNA C-values database (Release 7.1, April 2019): Range and mean (•) with species counts (Source: Pellicer and Leitch, 2019)	21
2.3	Schematic representation illustrating the relationship between GS change and trait evolution (Source: Bhadra et al., 2023)	23
2.4	Structure of the *Arabidopsis thaliana* chloroplast genome. (SSCR small single copy region, LSCR large single copy region, and IR-A and IR-B inverted repeats. Length of plastome and its parts in kbp) (Source: Olejniczak et al., 2016)	27
2.5	The graph illustrates how plastid reads are distributed within genomic DNA sequence libraries across a set of 27 diverse green plant species with varying genome sizes. It displays the percentage of sequence reads that align with a reference database containing 100 plant plastomes. The shaded area in grey represents the 95% confidence interval of the regression line. (Source: Twyford and Ness, 2017)	29
3.1	Study area and its division into four of zones (Z-1, Z-2, Z-3, and Z-4)	39
3.2	Representative species of (a) aquatics, (b) climbers, (c) herbs, (d) shrubs and (e) trees	45
3.3	Diversity indices for four zones (a) Shimpon’s index, (b) Shannon-Weiner index, (c) Evenness and (d) Margalef index	56
3.4	Jaccard’s similarity index for six pair of zones	57
3.5	(a) Density map of plants and (b) Distribution percentage (Sea green: aquatics, Green: climbers, Deep green: herbs, Yellow: Shrubs and Red: Trees)	58
4.1	Inhibitors and secondary metabolites interfering in plant flow cytometry	64
4.2	Standardization methods for plant DNA analysis (a) external standardization; (b) pseudo-internal standardization, and (c) internal standardization	66
4.3	Cell cycle analysis through flow cytometry	67
4.4	Flow cytometric calibration for plant DNA analysis (a) voltage settings; (b) data acquisition (c) flow rate; (d) gating; (e) peak and (f) statistics	68

4.5 Standard plants used for GS estimation (a) *R. communis*, (b) *V. radiata*, (c) *O. sativa*, (d) *P. pinnata*, (e) *S. lycopersicum*, and (f) *Z. mays* 70

4.6 Genome size estimation of plant samples through flow cytometric analysis 71

4.7 Microscopic observations of comparative free nuclei isolation among different buffers observed at 20X in *A. oleracia* (a) LB01 buffer, (b) Galbraith's buffer, (c) Ploidy reagent (Sysmex), and (d) Hypotonic PI buffer (upper: bright field image, lower: blue filter) 75

4.8 Internal standardization G0/G1 peaks of standard and sample (a) Overlapping peaks (b) Clear peaks 77

4.9 Various scatter plots analyzed through flow cytometric analysis (a) FSC vs SSC, (b) Fluorescence intensity vs Width, (c) Fluorescence intensity vs FSC, (d) Histogram 79

4.10 Internal standardization histograms of GS estimation from 47 herbaceous plants 82

4.11 Histograms of internal standardization for GS estimation of 12 shrub species obtained through flow cytometric analysis 83

4.12 Flow cytometric histograms of internal standardization for GS estimation of 51 tree species. 85

4.13 Variation in genome size observed in herbs, shrub, and trees 88

4.14 Regression analysis results for herbs, shrubs and trees 93

4.15 ANN models for (a) herbs, (b) shrubs, (c) trees and (d) all plants 94

5.1 Circular map of *C. quadrangularis* whole cp genome along with gene annotations. Clockwise-transcribed genes are displayed inside the circle, and genes positioned outside are transcribed anti-clockwise. Different colour codes represent genes of different functions 107

5.2 A circular map depicting the entire chloroplast genome of *F. hispida* is presented, featuring gene annotations. Genes transcribed in a clockwise direction are illustrated within the circle, while those transcribed in anti-clockwise direction are positioned outside. Various color codes are employed to distinguish genes based on their functions. 109

5.3 A circular map depicting the entire cp genome of *S. rutidosperma* featuring gene annotations. Genes transcribed in a clockwise direction are illustrated within the circle, while those transcribed in an anti-clockwise direction are positioned outside. Color codes are employed to distinguish genes based on their functions. 111

5.4 Circular map of *S. asper* whole cp genome along with gene annotations. Clockwise-transcribed genes are displayed inside the circle, and genes positioned outside are transcribed anti-clockwise. Different colour codes represent genes of different functions. 112

5.5 Circular map of *T. peruviana* whole cp genome along with gene annotations. Clockwise- and anti-clockwise transcribed genes are displayed inside and outside the circle respectively. 113

5.6 MAUVE alignment of cp genomes of six *Cissus* spp. Different colored local collinear blocks represent the gene arrangement in chloroplast genomes. 115

5.7 Comparison of IR junctions among six *Cissus* spp. Colour boxes present the genes. The genes on the upper track represent the negative strand, whereas those below the track denote the positive strand. 116

5.8 Progressive MAUVE alignment of cp genomes of four *Ficus* spp. Different colored local collinear blocks represent the gene arrangement in chloroplast genomes. 117

5.9 Comparison of Inverted Repeat (IR) junctions among four *Ficus* spp. Genes are represented by color boxes. The upper track displays genes on the negative strand, while those below the track represent genes on the positive strand. 118

5.10 Plastome comparison among seven Cleomaceae species through progressive MAUVE alignment. 120

5.11 Comparison of IR junctions among seven Cleomaceae spp. Genes are represented by color boxes. The upper track displays genes on the negative strand, while those below the track represent genes on the positive strand. 120

5.12 Plastome comparison through progressive MAUVE alignment of seven *Streblus* spp. 122

5.13 Comparisons of LSC, SSC, and IR region borders among seven cp genomes of *Streblus* spp., colour boxes present the coding genes. 122

5.14 Plastome comparison using progressive MAUVE alignment of five Plumerieae species 124

5.15 Comparisons of LSC, SSC, and IR region borders among five Plumerieae chloroplast genomes of , colour boxes present different coding genes. 124

5.16 Comparison of tandem repeats of six chloroplast genomes of *Cissus* spp. (a) Number of tandem repeats by tandem repeat types, (b) Total number of tandem repeats in each species, and (c) Number of tandem repeat sequences by length. 126

5.17 Comparison of SSR repeats between *C. quadrangularis* and five chloroplast genomes of *Cissus* species. (a) Number of SSR repeats in regular and 127 compound formation, (b) Total number of SSR repeat types, (c) The number of SSR motifs, and (d) SSR density.

5.18 Tandem repeat mining from four *Ficus* spp. (a) Repeat types; (b) Total no of tandem repeats in each species; and (c) Number of repeats based on repeat length 129

5.19 Simple sequence repeats in *Ficus* spp. (a) Types of SSR; (b) No of SSR based on unit size' (c) Number of SSR based on unit type; and (d) SSR density in four species 129

5.20 Tandem repeat mining from seven Cleomaceae plants (a) Repeat types; (b) Total no of tandem repeats in each species; and (c) Number of repeats based on repeat length 131

5.21 Simple sequence repeats in Cleomaceae *plants* (a) Types of SSR; (b) No of SSR based on unit size' (c) Number of SSR based on unit type; and (d) SSR density in four species 132

5.22 Tandem repeats in seven *Streblus* spp. (a) Types of tandem repeats; (b) Total number of tandem repeats; and (c) Number of tandem repeats based on size. 133

5.23 SSRs found in seven *Streblus* spp. (a) Types of SSR; (b) Number of SSRs based on size; (c) Number of SSRs based on repeat types; and (d) SSR density. 134

5.24 Tandem repeats in five Plumerieae species (a) Types of tandem repeats; (b) Total number of tandem repeats; and (c) Number of tandem repeats based on size 136

5.25 SSRs found in five Plumerieae spp. (a) Types of SSR; (b) Number of SSRs based on size; (c) Number of SSRs based on repeat types; and (d) SSR density 136

5.26 Nucleotide diversity (*Pi*) values of protein coding regions in *Cissus* species 137

5.27 Nucleotide diversity (*Pi*) values of *Ficus* spp. among protein-coding regions 138

5.28 Nucleotide diversity (*Pi*) values of seven Cleomaceae spp. in protein-coding regions 139

5.29 Nucleotide diversity (*Pi*) among protein coding regions in seven *Streblus* spp. 140

5.30 Nucleotide diversity (*Pi*) among protein coding regions in five Plumerieae spp. 141

5.31 Comparison of Ka/Ks ratio generated in protein-coding genes aligned from six *Cissus* species cp genomes. 142

5.32 Comparison of Ka/Ks ratio generated in protein-coding genes from cp genome alignment of four *Ficus* spp. 144

5.33 Comparison of Ka/Ks ratio generated for protein-coding genes from seven aligned Cleomaceae spp. cp genome 145

5.34 Nucleotide substitution rate (Ka/Ks value) in coding regions of seven *Streblus* spp. 146
5.35 Nucleotide substitution rate (Ka/Ks value) in coding regions of five Plumerieae spp. 146
5.36 Codon usage bias of six *Cissus* spp. in terms of relative synonymous codon usage (RSCU). 148
5.37 Codon usage bias in terms of relative synonymous codon usage (RSCU) for four species of *Ficus* spp. 149
5.38 Codon usage bias in seven species of *Cleomaceae* in terms of relative synonymous codon usage. 150
5.39 Codon usage bias among seven *Streblus* spp. through RSCU values 151
5.40 Codon usage bias among five Plumerieae species through RSCU values 153
5.41 Phylogenetic tree using whole chloroplast genomes from 24 Vitaceae species constructed through the maximum likelihood method. 155
5.42 Phylogenetic tree showcasing the relationship of *F. hispida* with 20 other species based on the whole chloroplast genomes. 156
5.43 Phylogenetic tree showcasing the taxonomic position of *S. rutidosperma* from 17 complete chloroplast genomes. 157
5.44 Phylogenetic tree for *T. peruviana* based on the cp genomes from 17 Gentianales and one Lamiales constructed through ML method. 158

LIST OF TABLES

Table No	Table description	Page
2.1	Composition of the different buffers used for flow cytometric estimation of nuclear DNA content	18
2.2	Some established DNA reference standards recommended for the estimation of nuclear DNA estimation	20
3.1	Assessing plant biodiversity of IITG: Family, genus and species counts.	45
3.2	Frequently observed exotic plants in IITG	48
3.3	Top ten common and rare plants	49
3.4	Medicinal plants found in IIT Guwahati	51
3.5	Economically important plants	52
3.6	Some common food plants of different butterflies	54
3.7	Diversity analysis of four zones	55
3.8	Jaccard’s Similarity index	57
4.1	Modified PI buffer components and their role	64
4.2	Reference standards used for GS estimation	69
4.3	Linear correlation test between GS and genomic as well as morphological traits through Pearson’s R value analysis (* Statistical significance ±0.11)	91
4.4	Regression analysis of trees, herbs and shrubs	92
5.1	Physical characteristics of selected small genome sized plants compared to the model plant (*A. thaliana*)	104

Chapter 1

Introduction

CHAPTER 1: INTRODUCTION

Flowering plants, also known as angiosperms, are the most diverse group of plants on Earth, with over 400,000 species identified to date (APG IV, 2016). They play a crucial role in the ecology and economy of the planet, providing food, medicine, and ecosystem services to humans and other organisms. However, the diversity of flowering plants is not evenly distributed around the world, with some regions being more species-rich than others. The Northeastern (NE) region of India is one such area, known for its high floral diversity and endemism (Chatterjee et al., 2006). Indian Institute of Technology Guwahati (IITG), located in the NE Indian state of Assam, is home to many flowering plants. Our research endeavors have focused on exploring the floral diversity of the region, engaging in the identification and comprehensive cataloging of plant species. Furthermore, cutting-edge genomic techniques have been employed to estimate the genome size (GS) of these plants and to delve into their plastome, providing invaluable insights into the molecular characteristics of the local flora.

Flora, comprising the plant species found in a particular region or globally, play a vital role in ecological processes, providing habitats, food sources, and resources for countless organisms, including humans (Ghorbanpour et al., 2017). The diversity and history of flora hold significant importance in understanding the natural world and its evolution. The exploration of floral diversity, along with its historical context, offers valuable insights into the evolution, biogeography, and conservation of plant species (Donoghue, 2008). The world flora encompasses the entire range of plant species found across different continents, ecosystems, and climatic zones. With an estimated 400,000 plant species identified so far, the world flora exhibits remarkable diversity, reflecting the adaptation of plants to diverse environmental conditions (Govaerts, 2001; Christenhusz and Byng, 2016). Plant species have evolved over millions of years, undergoing various processes such as speciation, extinction, migration, and colonization, resulting in the current global distribution of plant species (Daru et al., 2018).

Exploring floral diversity helps one to understand the biogeographic patterns and historical factors that have shaped the distribution of plant species. The study of plant fossils, molecular phylogenetics, and ecological studies aids in reconstructing the evolutionary history of plant lineages and their responses to changing environmental conditions (Daru et al., 2018). India, a vast country with diverse landscapes, is known for its rich and diverse flora. The flora of India comprises approximately 46,000 plant species, representing around 7-8% of the world's total plant diversity (Mao, 2019). India's plant diversity can be attributed to its varied climatic conditions, ranging from tropical rainforests to alpine regions, and its diverse topography, including mountains, plains, plateaus, and coastal areas. The flora of India showcases a remarkable assemblage of plant species, including numerous endemic species that are found nowhere else in the world. Indian flora has been influenced by various historical and biogeographic factors, such as geological events, continental drift, tectonic activities, climatic changes, and interactions with neighboring regions (Mani, 1974).

NE India, encompassing the states of Assam, Arunachal Pradesh, Meghalaya, Nagaland, Manipur, Mizoram, Tripura, and Sikkim, is renowned for its unique and diverse flora. The region is characterized by its rich biodiversity, ranging from tropical rainforests to temperate forests, alpine meadows, wetlands, and riverine ecosystems. The flora of NE India is influenced by the region's geographical isolation, high rainfall, hilly terrains, and the convergence of different biogeographic realms. This region is home to over 8,000 species of flowering plants so far (Hazarika, 2021). Assam, the largest NE state, is particularly known for its exceptional floral diversity. It is home to a significant number of plant species, including rare and endangered taxa. Assam's flora is influenced by its unique climatic conditions, including the monsoon-driven rainfall and the presence of the Brahmaputra River. The region's strategic location at the intersection of the Indian subcontinent and Southeast (SE) Asia has resulted into a rich diverse flora with both Indian and SE Asian elements (Chatterjee et al., 2006). Exploring the flora of NE India, and specifically Assam, provides insights into the adaptation of plants to specific ecological niches, the impacts of geographical barriers on species distribution, and the conservation challenges faced by unique and threatened plant species. The flora of the NE region of India is characterized by its high endemism, some of the notable endemic plant families found in the region include Zingiberaceae (ginger family), Euphorbiaceae (spurge family), Poaceae (bamboo family), and Orchidaceae (orchid family) (Basak et al., 2010; Chakravarty et al., 2012). By studying and understanding these diverse flora, we can deepen our knowledge of plant evolution, inform conservation efforts, and promote sustainable management of our natural heritage (Dikshit and Dikshit, 2014).

IITG established in 1994 is home to more than 13,000 people inside a 700 acres of secluded land is situated in the northern sector of Guwahati city within the Kamrup Rural district of Assam, India with limited external contact, creating a heaven for rich floral diversity. This region is nestled alongside the mighty Brahmaputra river, spanning from 26° 12' 0" to 26° 12' 44" North and 91° 43' 23" to 91° 46' 42" East. The climate in this region is mostly tropical, with an average temperature of 24°C. During the monsoon season, heavy rainfall and thunderstorms are common. The soil in this region is primarily brown and reddish sandy loam, rich in both macro and micronutrients. The topography of the area is undulating, with variations in elevation ranging from 49.5 meters to 100 meters above mean sea level (MSL). These geographical features contribute to the unique environment surrounding IIT Guwahati, fostering a diverse and captivating ecosystem.

Molecular level characterization of plants is a fundamental aspect of modern plant biology that has revolutionized our understanding of these organisms. At the heart of this molecular exploration lies GS, a crucial parameter that underpins the intricate genetic machinery of plants (Gnanesh et al., 2023).

GS, often referred to as C-value, denotes the quantity of DNA contained within a plant's nucleus, encapsulating the blueprint for its growth, development, and interaction with the environment (Carta et al., 2020). Molecular-level characterization has opened new frontiers in plant research, enabling us to decipher the genetic makeup of various plant species with unprecedented precision. Understanding the GS of plants is pivotal because it serves as the foundation upon which all genetic information is encoded (Anderson and Purugganan, 2022). This knowledge is invaluable for unraveling the complex biological processes governing plant growth, development, and adaptation to diverse ecological niches. The GS of plants offers crucial insights into their evolutionary history, genetic diversity, and potential for adaptation to changing environments (Leitch and Leitch, 2012). As we delve deeper into molecular plant biology, GS estimation emerges as an essential tool with far-reaching implications for crop improvement, conservation of plant biodiversity, and overall comprehension of plant biology (Varshney and Dubey, 2009).

GS refers to the total amount of DNA within the cells of an organism and is typically measured in base pairs (bp) or picograms (pg). Despite the lack of a clear correlation between GS and organismal complexity, GS variation has been observed across diverse plant taxa. For instance, some plant species have extremely compact genomes, while others possess genomes that are several orders of magnitude larger. The GS of plants varies greatly, ranging from less than 100 Mb (0.1 pg) in some mosses to over 100 Gb (102.25 pg) in some angiosperms (Pellicer et al., 2018). The variation in GS among plants has fascinated scientists for decades. It has been proposed that these variations are influenced by multiple evolutionary factors, including repetitive DNA elements, transposable elements, and polyploidy events. Repetitive DNA sequences, such as tandem repeats and transposable elements, can contribute significantly to the expansion of plant genomes. Polyploidy, which involves the duplication of entire genomes, can also lead to substantial increases in GS (Bennett and Leitch, 2005c).

Estimating the GS of plant species is essential for understanding the genetic architecture and evolutionary implications of GS variation. GS estimation can provide insights into the mechanisms shaping plant genomes and their functional implications. Furthermore, it can shed light on the evolutionary strategies employed by different plant lineages in response to the environmental changes and ecological interactions (Ohri, 1998). One of the most widely used techniques for GS estimation is flow cytometry (FCM), which allows for high-throughput analysis of plant samples. FCM works by measuring the amount of DNA in individual nuclei using a fluorescent dye. The DNA content of the nuclei is then compared to that of a known standard, allowing for the estimation of GS. FCM has several advantages over other methods of GS estimation, including its speed, accuracy, low sample requirement, and low cost (Pellicer et al., 2014). This work explores the methodologies, significance, and applications of GS estimation, shedding light on its pivotal role in advancing our knowledge of plant biology at the molecular level.

Plastomes, the genomes of plastids, are valuable sources of genetic information for evolutionary studies in plants and algae. Plastids, including chloroplasts, are essential organelles involved in photosynthesis, carbon fixation, and other metabolic processes. The compact size and conserved nature of plastomes make them ideal targets for genome sequencing and analysis (Olejniczak et al., 2016). Chloroplasts are specialized organelles that are responsible for photosynthesis, which is a vital process for plants. The typical structure of the plastome is a circular, double-stranded DNA molecule, with a size range of 120 kb to 170 kb in most angiosperms (Shelke et al., 2022). The plastome is highly conserved across different plant species, which makes it an important resource for phylogenetic analysis and marker development.

Plastome mining is a process that involves the complete sequencing and bioinformatic analysis of the chloroplast genome. It contains a set of conserved genes that encode crucial components involved in photosynthesis, including photosystem proteins, electron transport chain components, and ribosomal RNAs. Apart from these, the plastome also encodes various other genes that play a role in chloroplast function and gene expression. The plastome provides a comprehensive view of the chloroplast genome and offers valuable insights into evolutionary relationships, genetic diversity, and functional genomics (Gao et al., 2010). With the continual development of sequencing technologies and bioinformatics, plastome sequencing will continue to contribute significantly to our understanding of plant evolution, ecology, as well as structural and functional genomics.

Current thesis aims to provide an overview of the floral diversity and GS estimation of flowering plants from IIT Guwahati, with a focus on plastome mining of selected small GS plants. Thesis is built on exploring and dissecting rich floral diversity of IIT Guwahati, with particular emphasis on the distinctive features of the region's plant life. Then it paves the way to discuss the techniques used for GS estimation, with a focus on FCM, a high-throughput technique that has revolutionized the field of plant genomics and finally, culminating with the plastome sequencing and bioinformatic analyses for selected small GS plants.

The work embodied in the thesis is based on the following objectives:

1. Comprehensive floral database of IIT Guwahati
2. Genome size estimation of identified plants
3. Plastome mining of selected small genome sized plants

By achieving these objectives, we aim to enhance our understanding of plant diversity, genome architecture, and the evolutionary processes that have shaped flowering plants. This research will contribute to the broader field of plant biology, ecology, and conservation, providing valuable insights into the genetic characteristics and evolutionary patterns of selected small genome-sized plants from IIT Guwahati.

Chapter 2

Review of Literature

CHAPTER 2: REVIEW OF LITERATURE

2.1. Floral diversity

2.1.1. Introduction

The review of literature on the flora diversity of NE India and Assam reveals the profound significance of this research in the context of biodiversity, ecology, and conservation. Nestled within the northeastern part of the Indian subcontinent, this region showcases a distinctive amalgamation of geographical and ecological factors, resulting in the evolution of a diverse array of plant species. These factors have established NE India and Assam as a pivotal focal point for the exploration and documentation of botanical diversity (Hazarika, 2021). Geographically, the region's unique positioning is characterized by the intersection of the Eastern Himalayas, the expansive Brahmaputra River basin, and the pronounced influence of the monsoon climate. This geographical diversity contributes to a wide spectrum of ecosystems, encompassing tropical rainforests, subtropical forests, wetlands, and grasslands. These ecosystems, in conjunction with the rugged topography ranging from lowland plains to towering mountain peaks, foster a remarkable diversity of plant life (Goswami et al., 2012). It is in this dynamic environment that numerous endemic and endangered plant species have found a niche, accentuating the region's biological significance (Chakravarty et al., 2012).

What makes this review particularly compelling is the exploration of the distinctive endemic flora found exclusively within North East India and Assam. These endemic species offer a window into the region's long history of evolutionary isolation, adaptation, and ecological coexistence (De et al., 2014). By scrutinizing the intricate relationships between these plants and their habitats, researchers gain insights into the mechanisms of evolution and the delicate balance of ecological interconnectedness (Donoghue, 2008). Furthermore, a comprehensive understanding of the region's flora is imperative for the formulation of well-informed conservation strategies, given the potential far-reaching consequences of losing unique species in this biodiversity hotspot (Olejniczak et al., 2016). Consequently, the literature on flora diversity in North East India and Assam not only contributes to regional knowledge but also plays a vital role in advancing global biodiversity science and conservation efforts.

2.1.2. Floral diversity of the world

The reviewed literature encompasses various studies related to plant diversity and its implications for biodiversity conservation and urban environments. Each of these studies offers valuable insights and findings pertinent to the broader understanding of plant species and their roles in different contexts.

Mutke and Barthlott (2005) conducted a comprehensive study examining diversity patterns of land plants at continental and global scales. Their research resulted in revised world maps of vascular plant species richness, mosses, and gymnosperms. The study revealed global centres of plant diversity in tropical and subtropical regions, emphasizing the significance of these geo-diverse areas. In 2010, a pioneering initiative commenced to create a consolidated global checklist of land plants, a remarkable undertaking not seen in over a century. This effort aimed to compile an updated and comprehensive inventory of terrestrial plant species, providing a valuable resource for plant taxonomy and classification through the website http://www.theplantlist.org (Convention on Biological Diversity, 2012). The follow up project to the Plant List was undertaken in 2012 in the name of the World Flora Online (WFO), which is merged with the Plant List in 2019 (Borsch et al., 2020). According to the WFO a total of 383,054 species of plants have been identified till today, from which 20,749 species were found in India (WFO, accessed on 2nd Nov 2023). Based on the latest data from APG IV all these species belong to 416 families from 64 orders (APG IV, 2016). Christenhusz and Byng (2016) reported 374,000 accepted species of plants from which 295,383 are flowering plants comprising of about 72% eudicots (210,008 spp.) and rest are monocots (74,273 spp.).

Lastly, Govaerts et al. (2021) introduced the World Checklist of Vascular Plants (WCVP), a valuable resource for plant diversity research, conservation, and management, which compiles 200,700 scientifically described plant species from various sources, ensuring data quality and accessibility. These studies collectively contribute to our understanding of plant diversity, its conservation, and its impact on various ecosystems and urban settings, offering valuable insights for future research and practical applications in biodiversity management.

2.1.3. Floristic diversity of India and NE India

India is celebrated for its exceptional floral diversity, which encompasses a wide range of plant species and ecosystems. This remarkable botanical richness has solidified India's status as one of the world's foremost biodiversity hotspots (Hinz et al., 2020). In NE India, this diversity reaches its pinnacle, with the region serving as a repository of distinctive and indigenous plant life. The varied landscapes, climatic variances, and geographical attributes of this region have fostered an unmatched abundance of plant species, making it a central focus for researchers and conservationists.

India stands as one of the world's megadiversity nations, boasting an extraordinary array of flowering plant species. With approximately 19,530 documented species, it is home to 5,400 endemics. The Hindustan Region in India is recognized as one of the Vavilovian Centers for agrobiodiversity, contributing 167 agricultural species to the world. Additionally, India harbors 320 wild crop relatives and encompasses 16 major forest types. Notably, 47,000 documented species constitute the diverse flora of India, which faces threats from agricultural expansion, urbanization, and development efforts (Raju et al., 2010). India's botanical wealth spans a vast and diverse landscape, encompassing 328 million hectares from tropical regions to high alpine areas. This diverse topography includes rich wetlands, deserts, coastlines, and mountain ranges like the Western and Eastern Ghats and the Himalayas. The country hosts a rich diversity of plant species, many of which are integral to the livelihoods of its large population (Bhatnagar and Kapoor, 2018). India's unique and extensive biodiversity accounts for 11.4% of all species on Earth, with diverse ecological habitats that support a wide range of ecosystems. Among the vast vegetal cover in the country, approximately 50,012 plant species, including 21,558 angiosperms, contribute to its remarkable biodiversity. Significantly, about 28% of these Indian plants are endemic (Dash et al., 2022).

NE India, comprises eight states, each characterized by its distinct cultural, linguistic, ethnic and floral diversity. These eight states are Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura. The Brahmaputra river basin is one of the highly species rich regions in NE India, spanning over an area of 528,083 square kilometers, largely extends beyond India's borders, with only 33.6% of its territory lying within the Indian boundary (Pradhan et al., 2021) (**Fig. 2.1**). This region is distinguished by the convergence of two significant biogeographic hotspots: the Himalayas and Indo-Burma, which, coupled with its fertile soil and favorable climate, make it a repository of remarkable biodiversity (Pradhan et al., 2021).

Bora et al. reported a rich diversity of 724 species of vascular plants from 492 genera and 142 families in the Brahmaputra valley. Their study also emphasized the floral diversity pattern, phenology, and the significance of various medicinal, fodder, economically important, rare, and endemic plants in Assam (Bora et al., 2003). Government data records 3854 species of angiospermic plants, of which 47% are herbaceous, 22% shrubs, 20% trees, 8% climbers, and 3% undershrubs in Assam. Additionally, Assam hosts 35% of Pteridophytes in the Indian flora (BSI, 2019).

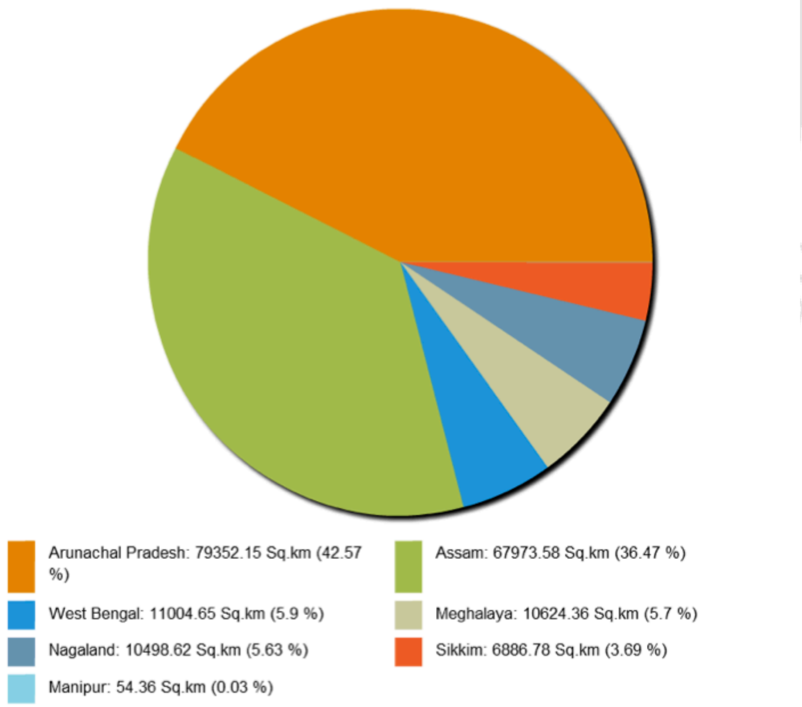


Figure 2.1: State wise area of Brahmaputra River basin (Source: Ministry of Water Resource, 2014)

Bhattacharjya and Sarma (2016) reported 217 weeds from 60 families of angiosperms and pteridophytes in the Nalbari district of Assam through frequent field visits over a period of 3 years. They found Asteraceae as the largest. Saikia et al. (2012) identified *Aquilaria malaccensis* as the dominant tree species in upper Assam, alongside 294 plant species from 217 genera representing 92 families. Since, a large portion of the Indian Brahmaputra valley falls within Assam, the floral diversity is very rich in Assam. Barooah and Ahmed (2014) cataloged 3854 taxa from Brahmaputra valley from 1394 genera, spanning 236 families, with 20% of them being trees. This study revealed a rich floral diversity unique to the Brahmaputra valley. In the valley, there is an extraordinary diversity of flowering plants from various genera, including Legumes, Orchids, and Cucurbits, as well as *Elaeocarpus*, *Dendrobium*, *Garcinia*, Pipers, *Dipterocarpus*, and *Calamus*. In this valley, the cultivation of tea was discovered, leading to the production of some of the world's finest teas, including the indigenous *Camellia sinensis* var. assamica (Barooah et al., 2014).

Gogoi and Nath (2021) identified 462 species of angiosperms from 334 genera under 106 families, out of which 129 species were invasive. Choudhary et al. (2012) presented a checklist of 1,003 taxa from 110 families, including 529 genera and 994 species, along with an abundant diversity of bamboos. In Assam 102 species of endemic plants have been listed by the Botanical Survey of India (Gogoi and Nath, 2021), and it is also considered the 'Creator Center' for *Citrus* and banana due to its rich germplasm. Additionally, aquatic fruits like makhana or gorgon fruit (*Eurale ferox*) are found frequently within the state (Choudhary, 2013).

2.1.4. Benefits of floral diversity

Floral diversity, which encompasses the wide range of plant species and their distinct blossoms, serves as a crucial element in our natural environment, offering numerous advantages that extend well beyond their visual attractiveness. These advantages include supporting pollinators and maintaining ecosystem health, as well as contributing to human well-being through food production and the potential for medicinal uses. Floral diversity enriches our surroundings and is fundamental to biodiversity preservation and human prosperity, both ecologically and economically.

Wilson et al. (1999) conducted a study focusing on the diet of granivorous bird species in European temperate farmland and how agricultural practices affect their food sources. Their findings suggest that intensification of arable land has led to a reduction in the diversity and abundance of food plants, potentially contributing to population declines in these birds. The study emphasizes the significance of certain plant species for these granivorous birds and highlights the adverse effects of herbicide applications and habitat loss. It has been observed in multiple cases that the introduction of exotic species can sometimes destroy the native vegetation and in turn ecosystem imbalance is inevitable. Exotic plants in southern Brazilian landscape was observed to influence native plantations hence decreasing the lichen diversity in the region (Kaffer et al., 2009). The study recommends replacing exotic tree species with native ones to preserve lichen diversity.

Analysis of plant-bee visitation networks in herbaceous semi-natural habitats, the study identified key resource plants for different conservation target groups, highlighting the importance of floral resource complementarity in maintaining diversity (Sutter et al., 2017). The role of ornamental

Geranium spp. in supporting urban pollinators was discussed by comparing these plants in terms of floral phenology, abundance, nectar and pollen production, and insect visitation. The results suggest that these plants could be beneficial for urban beekeeping and urban planting strategies (Masierowska et al., 2018, Oliveira et al., 2020). Another study found that local bee diversity increased with greater floral resource richness, emphasizing the importance of focusing on local-scale diversity in habitat plantings, even in areas dominated by agriculture (Lane et al., 2020).

2.1.5. Floristic studies and inventories

This section encompasses a collection of studies focused on analysing the structure and floristic composition of various ecosystems, providing insights into the biodiversity, conservation, and ecological dynamics of these environments. These studies offer valuable information about plant species, their distribution, and the factors influencing their diversity.

In the Beza Mahafaly Reserve in southwestern Madagascar, 923 plants belonging to 69 species and 43 genera were identified revealing the dominance of native species (Sussman and Rakotozafy, 1994). Strasberg (1996) inventorized the lowland tropical rain forest on La Réunion Island, identifying 43 tree species and mapping their diversity and structure. The study highlights the impact of factors like hurricane disturbance and limited seed dispersal on the forest's composition. The floristic study in Nepal's Terai zone, provided valuable insights into the composition, distribution, uses, and threats to the wetland ecosystems (Siwakoti, 2006). Another study by Flinn et al. (2008) on small wetlands in upland forests in Quebec, Canada revealed their potential to support diverse plant communities and increase species richness. The study highlights the importance of local wetlands in preserving regional biodiversity and categorizes five distinct plant community types.

Xiao-Tao et al. (2011) highlighted the importance of understory vegetation in biodiversity conservation decisions within tropical seasonal rain forests. This study identified 309 species highlighting the significance of the Rubiaceae family in both sapling and herb/seedling layers, emphasizing the need for greater attention to understory vegetation in conservation efforts. The importance of relationship between elevation and herbaceous plant diversity was discussed emphasizing the need for habitat diversity to sustain high species richness along elevation gradients in Veracruz, Mexico (Gómez-Díaz et al., 2017). The study on the composition of tree and shrub species in cemeteries across Muslim, Christian, and Jewish communities in Istanbul, Turkey; highlights the importance of preserving rare and threatened species. It also emphasizes the need for a centralized management system, which hold both natural and cultural significance (Servet and Aktağ, 2019). Killeen et al. (1998) delved into the floristic inventory of a tropical semideciduous forest in Santa Cruz, Bolivia, featuring 501 plant species.

2.1.6. Endemic, rare and endangered species of NE India

A remarkable blend of geographical and environmental elements, including rich soil and a favorable climate, are responsible for the extraordinary botanical diversity and notable prevalence of endemic species, rendering it a notable hub for research and efforts aimed at conserving biodiversity.

The NE India, occupies 7.7% of India's geographical area, housing 50% of its flora, including approximately 8000 species, of which 31.58% (around 2526 species) are endemic. The region is renowned for its rich diversity of orchids, ferns, oaks, bamboos, rhododendrons, and magnolias. Among these, orchids stand out with about 856 species, and 34 of them are identified as threatened in India, many of which are exclusive to this region (De and Medhi, 2014). A survey of North Cachar hills, Assam, NE India revealed 34 species of plants to be threatened. Five species have been included in the Red Data List of the International Union for Conservation of Nature (IUCN). Population of species like *Taxus baccata*, *Renanthera inschootiana*, *Swertia chirata* etc., was found to be declining sharply (Sajem et al., 2008). Sivasagar district, once the capital of the Ahom dynasty, stands out for its natural resources, with six reserve forest areas like Sapekati and Dilli that host a wide range of plant and animal species due to varying climate and soil conditions. This district plays a crucial role in preserving a significant portion of Assam's plant life, supporting local ethnic communities like Tai-Ahom, Tai-Shyam, Mising, and others who rely on these forests for their livelihood (Jiji, 2015).

Rattans, prickly climbing palms comprising over 50% of India's palm species, hold economic importance in handicrafts and furniture due to their strong fibers. However, overexploitation and habitat loss have rapidly depleted rattan resources. In particular, *Calamus nambariensis*, a high-quality cane endemic to Northeast India, faces severe threats to its survival. This study reveals its presence in only two areas in Assam, with a declining population, emphasizing the urgent need for conservation strategies (Deka et al., 2017). This NE region reported to have 127 orchid species, including 6 varieties, endemic to the area. These orchids exhibit a diverse range of growth habits. While many are epiphytic, others are terrestrial or saprophytic. The highest concentration of endemic orchids is found in Arunachal Pradesh and Sikkim (Pal and Singh, 2016). The northeastern region of India harbors a rich diversity of *Citrus* spp., both wild and semi-wild, but many face genetic erosion and extinction due to changing agricultural practices. To safeguard these valuable resources, research efforts are needed for molecular and morphological characterization, evaluation, and screening against various stresses (Hazarika, 2012).

2.1.7. Medicinal and ethnobotanical importance

The Bez community from Barpeta, Assam were employing over 66 plant species for various health-related purposes, including treating dysentery, jaundice, blood pressure, and cough (Das, 2016). Lodh and Swamy (2019) underscored the significance of medicinal plants in NE India for women's health, particularly in addressing gynecological disorders. They discussed various plants and their bioactive compounds, highlighting their role in treating conditions like infertility and ovarian cancer. Tynsong et al. (2020) discussed the Traditional Ecological Knowledge (TEK) held by tribal communities in NE India, highlighting the rich reservoir of knowledge related to species, ecosystems, and sustainable resource management. They emphasize the need to document and preserve this valuable knowledge to prevent its loss. Talukdar et al. (2020) explored the diversity of endophytic fungi in *Houttuynia cordata*, an ethnomedicinal plant in Northeast India. They investigated the antimicrobial metabolites and suggested the potential for developing pharmaceutical agents from these fungi.

Sajem et al. (2008) documented the utilization of various plant species in the daily lives of local communities in NE India. They emphasized on the importance of conserving these species due to their declining populations and provide information on their botanical names, usage, and threat status according to IUCN guidelines. De et al. (2016) highlighted the extraordinary orchid diversity in NE India, emphasizing the region's ecological significance. They noted the various threats, including habitat loss, and the need for conservation measures, including the establishment of protected areas. Deka et al. (2017) discussed the conservation of rare and threatened orchid species, *Vanilla borneensis* and *Calamus nambariensis*, found in Assam. They detailed the distribution, population estimation, and propagation methods to improve their conservation status.

2.1.8. Threats to flora diversity and conservation efforts

The reviewed literature provide insights into various ecological and environmental aspects, encompassing the impacts of human colonization, volcanic history, competition, herbivory, and habitat fragmentation on plant diversity and ecosystems. These studies offer valuable information on the dynamics of plant communities, species richness, and the consequences of environmental changes.

A study on lowland tropical rain forests in La Réunion, highlighted the negative effects of human colonization on forest coverage and the diversity of plant species (Choudhury and Khan, 2010). De and Medhi (2014) emphasized the threats of habitat loss, exotic species introduction, over-exploitation, pollution, global warming, and agricultural commercialization. They discuss India's conservation measures, including regulations and protected areas, to address these challenges. The declining biodiversity in Shivsagar district of Assam, attributed to growing demand, destructive harvesting, over-exploitation, illegal trade, and urban expansion. The authors suggested establishment of medicinal plant farms as a sustainable solution to meet resource demand while preserving biodiversity (Jiji, 2015). Baruah et al. (2020) addressed the conservation of *Mesua assamica*, an endangered plant in Assam, India to prevent its extinction. In other cases volcanic eruptions, grass competition and herbivory, historical and environmental factors, urbanization and habitat fragmentation were reported to be the primary causes of diversity loss (Emerson, 2003; Midoko-Iponga et al., 2005; Kreft and Jetz, 2007; Aronson et al., 2014; Xiao, 2016; Lyseng et al., 2018).

Xiao-Tao et al. (2011) stressed upon the importance of understory vegetation in making biodiversity conservation decisions within tropical seasonal rainforests. They advocated for the importance of Rubiaceae family in both sapling and herb/seedling layers. The NE Indian region is home to 17 crop species representing 47% of the country's crop diversity. However, the region faces challenges due to fragility, marginality, inaccessibility, cultural heterogeneity, and ethnicity. Northeast India is rich in plant diversity, with a high percentage of endemic species and numerous rare and endangered taxa. Conservation efforts are essential to protect this unique biodiversity (Roy et al., 2015). The review by Corlett (2016) emphasized the urgent need for protected areas and the assessment of unevaluated plant species to mitigate threats such as habitat loss and climate change. Rion (2017) investigated vegetation changes in montane fens and wet meadows in the Jura Mountains over several decades. The study highlighted the impact of eutrophication resulting from land-use changes and nitrogen deposition and voiced for conservation efforts. Urban areas, where over 50% of the global population resides, have led to a loss of biodiversity due to land-use changes. Urban green spaces, including urban forests, play a vital role in conserving and managing biodiversity (Dhyani et al., 2020).

2.1.9. Knowledge gaps and conclusion

The state of Assam, India is a treasure trove of floral diversity, encompassing a wide array of plant species, including numerous endemic species. The North Guwahati (IITG) enjoys unique combination of geographical and ecological factors, such as fertile soil and favorable climate, which contributes to its exceptional botanical richness, and high endemism making it a significant location for biodiversity research and conservation. Assam’s plant diversity extends beyond sheer numbers, playing a vital role in the cultural, economic, and ecological fabric of the NE region (Deka et al., 2017; Tynsong et al., 2020).

There are several notable knowledge gaps in the field of biodiversity and its conservation in the NE regions of India. First and foremost, there is a need for comprehensive and up-to-date biodiversity assessments that encompass both flora and fauna (Arora, 2018). Many species in this region remain poorly documented, and the lack of baseline data hinders effective conservation efforts. Additionally, there is limited research on the specific ecological dynamics, interdependencies, and unique adaptations of the region's biodiversity, which is critical for formulating tailored conservation strategies (Kandel et al., 2016). Furthermore, understanding the socio-economic factors and traditional knowledge systems of indigenous communities in the NE region is essential, as their practices can significantly impact local biodiversity (Singh et al., 2010). Finally, the impacts of climate change and land-use changes on the region's biodiversity need to be better understood to develop adaptive conservation measures (Gogoi and Nath, 2021; Habibullah et al., 2022). Addressing these knowledge gaps is crucial for the effective preservation of the rich and unique biodiversity in the NE regions of India.

Biodiversity in North Guwahati, specifically in the vicinity of the Indian Institute of Technology (IIT) Guwahati, has a rich historical backdrop. This region, situated in the NE part of India, is renowned for its exceptional biodiversity and natural beauty. The IIT Guwahati campus itself hosts a diverse range of flora, representing both indigenous and exotic plant species. The presence of this diverse plant life contributes significantly to the ecological balance of the region and provides a natural environment for various organisms, including insects, birds, and mammals. Moreover, the rich flora of IIT Guwahati serves as a valuable resource for scientific research and conservation efforts, as many of these plant species have unique properties and are of immense ecological and economic importance. Understanding the history and composition of this biodiversity in the context of IIT Guwahati is crucial for its effective conservation, utilization and management.

2.2. Genome size estimation

2.2.1. Introduction

The study of GS (C-value), is a critical component of plant genomics and an essential aspect of understanding the genetic diversity and complexity of the plant kingdom (Gregory, 2002). The GS of plants represents the amount of DNA contained within the nucleus of their cells, encompassing all the genetic information necessary for their growth, development, and environmental interactions (Bennett and Leitch, 2005a). GS estimation in flowering plants, a subgroup of the plant kingdom, is a vibrant and evolving field of research that offers valuable insights into the biology, evolution, and diversity of these plants (Bennetzen et al., 2005; Doležel and Greilhuber, 2010).

The quest to estimate and comprehend GS of flowering plants has been greatly facilitated by advancements in molecular biology and sequencing technologies (Álvarez-Borrego et al., 2007; Kelly and Leitch, 2011; Vitales et al., 2019). Accurate measurement of GS allows scientists to not only explore the evolutionary history of plant species but also delve into the ecological and physiological adaptations that have shaped them over time (Albach and Greilhuber, 2004; Canapa et al., 2016). This review of literature on GS in plants, with a particular focus on flowering plants, aims to illuminate the significance of this field of study. It will elucidate the methodologies and techniques employed to estimate GS, emphasizing their role in advancing scientific knowledge, molecular characterization, and the conservation of plant biodiversity.

2.2.2. Genome size and its relevance

GS is a fundamental parameter in molecular plant biology that has garnered significant attention due to its relevance in understanding plant diversity, evolution, and adaptation (Das and Rangan, 2020). GS refers to the amount of DNA contained within a plant's nucleus, encompassing both coding and non-coding regions (Blommaert, 2020). It is generally measured in picograms (pg) or in nucleotide base pairs (bp). One picogram equals to 978 megabases (Dolezel et al., 2003; Bennett and Leitch, 2005a). The estimation of GS has become a crucial tool in plant research, offering insights into various aspects of plant biology (Greilhuber et al., 2005). Estimates of GS are of paramount importance for the construction and screening of genomic and cDNA libraries (Das et al., 2018). In angiosperms 2400-fold variation in GS was reported by Pellicer et al. (2010).

One of the primary areas where GS plays a critical role is in the understanding of plant evolution and diversity. In recent decades, interest in plant GS has been growing exponentially as the biological, evolutionary and ecological significance of this key biodiversity trait is increasingly recognized (Basak et al., 2017; Pellicer et al., 2018). It has been observed that GS varies widely among plant species, even within the same taxonomic group as observed in nineteen turmeric species (Basak et al., 2017). These variations in GS are indicative of the diverse evolutionary histories of different plant lineages (Bennett and Leitch, 2005b). GS estimation can help unravel the relationships between plant species and provide insights into their evolutionary trajectories (Leitch et al., 2009). GS is also closely linked to plant adaptation and environmental responses (Chénais et al., 2012; Das and Rangan, 2020). Plants with larger genomes tend to exhibit slower growth and reproduction rates, which can be advantageous in stable environments (Greilhuber and Leitch, 2012; Suda et al., 2015). Conversely, plants with smaller genomes often display faster growth and reproduction, which can be beneficial in rapidly changing or disturbed habitats (Gregory, 2002). By studying GS, researchers can gain a deeper understanding of how plants adapt to different ecological niches and respond to environmental challenges (Ramesh et al., 2014; Das et al., 2018).

Furthermore, GS estimation is essential in crop improvement and plant breeding programs (Choudhury et al., 2014). Knowledge of the GS of crop species is crucial for selecting appropriate breeding strategies (Edwards and Batley, 2010). It helps in identifying potential barriers to hybridization, understanding the genetic diversity within a crop species, and optimizing breeding techniques to enhance crop productivity and resilience (Miflin, 2000; Bevan et al., 2017). Also, GS has implications for plant conservation (Pellicer et al., 2018). Endangered plant species often have smaller genomes, which can make them vulnerable to genetic bottlenecks and reduced adaptability (Hamabata et al., 2019). Understanding the GS of endangered plants is vital for implementing effective conservation strategies to safeguard their genetic diversity and prevent extinction (Khan et al., 2016; Brandies et al., 2019).

It is now evident that GS can have a substantial impact at various levels, from influencing gene and genome dynamics (Dodsworth et al., 2015) to playing a role in broader ecological contexts, affecting aspects such as plant growth strategies, plant community composition, plant-animal interactions, evolutionary pathways, and ecosystem dynamics (Suda et al., 2015; Simonin and Roddy, 2018; Guignard et al., 2019). In the age of rapidly advancing high-throughput sequencing technologies, C-values also serve as fundamental information essential for estimating the costs of whole-genome sequencing, as the expenses are inherently linked to GS (Li and Harkess, 2018; Pellicer and Leitch, 2019). GS in plants is a powerful tool with multifaceted relevance. It contributes to our understanding of plant diversity, evolution, adaptation, and provides valuable insights for crop improvement and conservation efforts. As molecular techniques continue to advance, GS estimation remains a cornerstone of modern plant biology, offering a molecular-level perspective on the intricate world of plants (Bennett and Leitch, 2005c).

2.2.3. Various approaches for estimating genome size

GS estimation in plants is a crucial aspect of genomics research, as it provides valuable information about the structure and organization of plant genomes. Accurate GS estimation is essential for various applications, including comparative genomics, phylogenetics, and understanding the evolution of plant genomes (Doležel et al., 1998; Ramesh et al., 2014).

Feulgen staining combined with microspectrophotometry involves staining plant tissue sections with a specific dye and then measuring the absorbance of stained nuclei using a microspectrophotometer. This method provides a direct measurement of DNA content in individual nuclei (Baranyi and Greilhuber, 1996; Greilhuber and Obermayer, 1998). The advantages of the method include precise measurement of DNA content, suitablity for a wide range of plant species, and ability to generate spatial information about DNA distribution within the tissue (Prenna et al, 1974; Hardie et al., 2002; Rasch, 2004). But, this method of GS estimation is time-consuming and requires specialized equipment and not suitable for small or precious plant samples (Greilhuber, 1998; Álvarez-Borrego et al., 2007)

Karyotyping involves examining the physical structure of chromosomes through various techniques such as microscopy and fluorescence *in situ* hybridization (FISH). GS can be estimated by counting the number of chromosomes and measuring their lengths (Bennett and Leitch, 2005a; Chen et al., 2013). The advantages of this method are generation of valuable cytogenetic information, ability of integration with other genomic analyses and helpful in taxonomic studies (Greilhuber, 1998; Doležel, et al., 2012). This method is time-consuming and labor-intensive, may not work for species with small chromosomes or similar chromosome sizes, and limited to species with well-characterized karyotypes (Dolezel and Bartos, 2005; Bennett and Leitch, 2005a; Weiss-Schneeweiss and Schneeweiss, 2012).

Another recent and accurate method is whole-genome sequencing (WGS) which can provide an accurate estimation of GS by analyzing the total amount of sequenced DNA. It is particularly useful for non-model species or those with complex genomes (Li and Harkess, 2018; Pucker, 2019). WGS is a versatile and comprehensive method for analyzing the genetic information of various plant species. It offers valuable insights beyond GS estimation (Bolger et al., 2014; Ari and Arikan, 2016; Wickland et al., 2017). Modern sequencing platforms have made it more accessible, but it can be costly and data-intensive. Advanced bioinformatics analysis and substantial computational resources are necessary for meaningful data interpretation. Therefore, it may not be the most cost-effective option for small research projects or for plant species with large, repetitive genomes (Hamilton and Robin, 2012; Ari and Arikan, 2016)

Pulsed-Field Gel Electrophoresis (PFGE) is a technique that separates high-molecular-weight DNA fragments according to their size, achieved by exposing them to an alternating electric field (Sobral et al., 1990; Grothuesi and Tümmler, 1991). It is particularly well-suited for plant species with large genomes and offers insights into genome structure (Davis et al., 2003). However, PFGE necessitates specialized equipment and expertise, making it less suitable for routine GS estimation (Howlett, 1996; Birren and Lai, 2012).

Flow cytometry is one of the most commonly used methods for GS estimation in plants. This technique relies on measuring the fluorescence intensity of nuclei stained with a DNA-specific fluorescent dye as they pass through a laser beam (Darzynkiewicz et al., 2010). The DNA content of the sample is then determined by comparing it to a reference sample with known GS (Doležel and Bartoš., 2005; Pellicer and Leitch, 2014). Flow cytometry has multiple advantages including high throughput, simplicity and cost-effectiveness (Nath et al., 2014; Bourge et al., 2018). The flowcytometry is also suitable for a wide range of plant species (D'hondt et al., 2011). The limitations of FCM include requirement of well-preserved nuclei, making it less suitable for some recalcitrant plant species (Vrána et al., 2014). The accuracy is dependent on sample quality, and presence of contaminants (Doležel et al., 1998; Trávníček et al., 2015). Co-existing endopolyploidy can lead to errors in estimation (Doležel et al., 2007).

2.2.4. Flowcytometry its advantages and limitations

Flow cytometry has emerged as a powerful tool for GS estimation in plants, offering several advantages over traditional methods. This review discusses the principles of flow cytometry, its strengths, and limitations in the context of GS estimation in plant research (Dolezel and Bartos, 2005; Wang et al., 2015).

Flow cytometry measures GS (C-value) by analyzing the fluorescence intensity of individual nuclei labeled with a DNA-specific fluorescent dye as they pass through a laser beam (Macey and Macey, 2007; Manohar et al., 2021). This method allows for rapid, high-throughput analysis of a large number of nuclei, making it particularly attractive for plant genomics studies (Galbraith, 2009). Flow cytometry has gained popularity due to its simplicity and efficiency in estimating nuclear DNA content (Adan et al., 2017). This is a high-throughput method which can analyze thousands of nuclei in a matter of minutes, providing substantial amount of data in a short time (Dolezel and Bartos, 2005). Unlike some traditional methods, flow cytometry is non-destructive, allowing for the preservation of precious plant samples (Heidmann et al., 2016). Flow cytometry is relatively easy to perform, requiring less extensive sample preparation compared to techniques like Feulgen staining or microspectrophotometry (Sliwinska et al., 2022). FCM can detect the presence of subpopulations, which is valuable when studying mixtures of cell types or different cell cycle stages within a sample (Kormelink et al., 2016). FCM can quantify the relative DNA content with high accuracy, with minimal influence from cytosolic compounds, thanks to the appropriate buffer and staining choices (Dolezel et al., 2007).

The choice of buffer, staining period, and dye concentration can impact nuclear DNA content estimates, potentially altering the C-values (Dolezel et al., 2007). FCM is typically performed on young leaves, but seed materials can be used as alternatives. The choice of sample material can affect the results (Sliwinska et al., 2005). Variations in the procedure, even with the same plant, may lead to intraspecies variation, emphasizing the importance of replicating experiments (Suda et al., 2007). FCM can have difficulty interpreting Cot (Conc. X Time) curves due to the presence of repetitive elements in sequences, limiting its use in some genomic analyses (Sarkar et al., 2011).

2.2.5. History of plant nuclei isolation buffers

Accurate DNA extraction and nuclei isolation are fundamental in plant flow cytometry studies for estimating GS. Over time, significant improvements have been made in the nuclear isolation buffer and staining protocols, as well as the choice of buffer solutions. This review summarizes the key advancements and considerations in these areas for reliable GS estimation using flow cytometry.

During early days of GS estimation Thornthwaite et al. (1980) introduced the nuclear isolation medium (NIM), comprising phosphate-buffered isotonic saline with calcium and magnesium. NIM also included NP40 and the DNA fluorochrome DAPI for staining. Rayburn et al. (1989) found that DAPI was more effective than mithramycin (MI) as a fluorochrome stain, which was further supported by Doležel et al. (1992). Loureiro et al. (2006) identified LB01 and Otto's buffer as effective for nuclei isolation, with Otto's buffer being suitable for species with low DNA content. Galbraith's buffer also yielded satisfactory results. However, no single buffer was optimal for all species, as different buffers responded differently to specific challenges.

To enhance the reliability of FCM analysis, modifications to the nuclei isolation buffer have been suggested, such as the addition of PVP and BME, longer incubation times on ice before analysis, or using different floral parts of plants (Loureiro et al., 2021). In a study by Sadhu et al. (2016), a modified buffer (MB01) was compared against eight commonly used nuclei isolation buffers for GS estimation in the Zingiberaceae family. Sharma et al. (2019) also introduced modifications, demonstrating the effectiveness of a custom buffer to counteract the inhibitory effects of phenolic compounds present in tea leaves on nuclei staining with Propidium Iodide (PI). Najafi et al. (2022) utilized a PI kit for nuclear DNA content estimation in *Aegilops* spp. Their study featured an incubation period with a DAPI staining solution, presenting a recent approach in the field. Another hypotonic PI buffer was used to analyse the mammalian cells in FCM by Krishan (1975), which was later modified for plants by Ramesh et al. (2014) and has been used since then for various plant groups with higher success rate than other nuclei isolation buffers (Das et al., 2018)

Despite its effectiveness, the use of a PI buffer presents limitations like toxicity to cells and staining other cellular components, including RNA and proteins, which can complicate data interpretation (Pozarowski et al., 2003). Advancements in DNA extraction, nuclei isolation, and staining protocols have significantly improved the accuracy and reliability of GS estimation using flow cytometry. These developments, along with custom buffer solutions, enable researchers to overcome challenges and limitations associated with various plant species (**Table 2.1**).

Table 2.1. Composition of the different buffers used for flow cytometric estimation of nuclear DNA content

Buffer	**Composition**	**Reference**
Galbraith's buffer	45mM $MgCl_2$; 30mM sodium citrate; 20mM MOPS; 0.1%(w/v) Triton X-100; pH 7.0	Galbraith et al. (1983)
LB01	15 mM TRIS; 2 mM Na_2EDTA; 0.5 mM spermine.4HCl; 80 mM KCl; 20 mM; NaCl; 15 mM β-mercaptoethanol; 0.1 % (v/v) Triton X-100; pH 7.5	Dpoolezel et al. (1989)
Arumuganathan and Earle	9.53 mM $MgSO_4.7H_2O$; 47.67 mM KCl; 4.77 mM HEPES; 6.48 mM DTT; 0.25 % (w/v) Triton X-100; pH 8.0	Arumuganathan and Earle (1991)
Marie's nuclear isolation buffer	50 mM glucose; 15 mM KCl; 15 mM NaCl; 5 mM Na_2EDTA; 50 mM ;sodium citrate; 0.5 % (v/v) Tween 20; 50 mM HEPES; 0.5 % (v/v) β-mercaptoethanol; pH 7.2	Marie and Brown (1993)
Otto buffers	Otto I buffer: 100mM citric acid; 0.5%(v/v) Tween 20 (pH approx. 2.3) Otto II buffer: 400 mM $Na_2HPO_4.12H_2O$ (pH approx. 8.9)	Otto (1990) Dolezel and Gohde (1995)
Tris-$MgCl_2$	200 mM TRIS; 4 mM $MgCl_2.6H2O$; 0.5 % (v/v) Triton X-100; pH 7.5	Pfosser et al. (1995)
Baranyi's	Baranyi solution I: 100 mM citric acid monohydrate 0.5 % Triton -X-100 Baranyi solution II: 400 mM $Na_2PO4.12H2O$; 10mM sodium citrate; 25mM sodium sulfite	Baranyi and Greihuber (1995)
Bergounioux's	Tissue culture salts supplemented with 700 mM sorbitol; 1.0 % (v/v) Triton-X-100; pH 6.6	Bergounioux et al (1986)
Rayburn's	1mM hexylene glycol; 10mM $MgCl_2$; 0.5 % (v/v) Triton- X-100; pH 8.0	Rayburn et al. 1989
Bino's	200 mM Manitol; 10 mM MOPS; 0.05 % (v/v) Triton-X-100; 10mM KCl; 10mM NaCl; 2.5 mM DTT; 10 mM Spermine.4HCl; 2.5mM $Na_2EDTA.2H_2O$; 0.05 % (w/v) sodium azide; pH 5.8	Bino et al. (1993)
De Laat's	15 mM HEPES; 1 mM $EDTANa_2.2H_2O$; 0.2 % (v/v) Triton-X-100; 80mM KCl; 20mM NaCl;15 mM DTT; 0.5 mM Spermine.4HCl; 300 mM sucrose; pH 7.0	de Laat and Blaas (1984)
Modified Hypotonic PI (MHPI) Buffer	PI 25 25 µg/mL; IGEPAL 0.1% (v/v); Trisodium citrate 0.1% (w/v), BME 1% (v/v), PVP-40 1% (W/V); RNase-A 2mg/ml	Ramesh et a. (2014)

2.2.6. Standardization methods and plant reference standards

Flow cytometry, a powerful tool in plant genomics, relies on robust standardization methods in sample preparation for accurate GS estimation (Doležel and Bartoš, 2005). Reference standard plants play a pivotal role in this process. Flow cytometry is a widely used technique for estimating GS in plants (Dolezel et al., 2007; Temsch et al., 2022) . Proper standardization methods are crucial to ensure the accuracy and comparability of GS measurements. Three common standardization approaches are employed in plant flow cytometry: internal standardization, external standardization, and pseudo-internal standardization (Greilhuber et al., 2007).

Internal standardization involves processing the tissues of an internal reference standard of known GS along with the sample and analysis using FCM. This standardization ensures same physical and chemical processing conditions both for sample and reference standard plant tissues minimizing potential errors. It allows for within-study calibration, reducing variability within and among experiments along with cytological stability and uniformity (Choudhury et al., 2014; Temsch et al., 2022). In external standardization, the reference standard and the sample tissues are processed and analyzed separately. It allows for the calibration of flow cytometers and comparability of GS measurements across dissimilar species. Unintentional differences in sample preparation can potentially introducing errors in GS estimation (Sliwinska et al., 2022; Temsch et al., 2022). Pseudo-internal standardization is a hybrid approach that combines aspects of both internal and external standardization. In this method sample and reference standard plants' nuclei are separately extracted and mixed just before analysis. This approach strikes a balance between the advantages of internal and external standardization (Ramesh et al., 2014). Depending on the genetic distance between the sample species and reference standard, some variation in GS estimates may persist (Čertnerová and Galbraith, 2021). The choice between internal, external, or pseudo-internal standardization methods in plant flow cytometry depends on factors such as the availability of suitable standards, the biological similarity between the standards and the samples, and the goals of the study. While internal standardization is ideal and widely accepted, ensuring biological similarity in most of the cases, it may not always be practical (Sliwinska et al., 2022).

Reference standard plants used for flow cytometry should be cytologically stable, uniform, easily available and without intra-specific variation. Ideally, they should have a low level of secondary metabolites and possess an appropriate and well-defined GS obtained by flow cytometry (D'Hondt et al., 2011). The GS of reference standards should not differ significantly from the samples under investigation, preferably no more than four times larger or smaller. The reference standard plants facilitate comparability of results within and among studies, aiding in the assessment of GS variations within plant lineages (Temsch et al., 2022). Some well defines and accepted reference standards reported over time are listed in **Table 2.2.**

Table 2.2: Some established DNA reference standards recommended for the estimation of nuclear DNA estimation

Plant species and cultivars	**2C DNA content (in pg)**	**References**
Oryza sativa	1.01	Price and Johnston (1996)
Raphanus sativus	1.11	Dolezel et al. (1992)
Solanum lycopersicum	1.96	Dolezel et al. (1992)
Glycine max	2.50	Dolezel et al. (1994)
Zea mays	5.43	Lysák and Doležel (1998)
Pisum sativum	9.09	Doležel et al. (1998)
Secale cereal	16.19	Doležel et al. (1998)
Vicia faba	26.90	Doležel et al. (1992)
Allium cepa	34.89	Doležel et al. (1998)

2.2.7. Genome size estimation in angiosperms

Flow cytometry has emerged as a powerful tool for the estimation of GSs in various plant species. A review of the literature reveals several prominent plant GS estimates, shedding light on the diversity and correlations within plant families.

Bennett and Smith (1991) along with Bennett and Leitch (1997) compiled estimates for the nuclear DNA amounts of more than 1000 angiosperm species. Ohri and Kumar (1986) determined the GS of ten different legume plants using FCM and ranged between 1.68 to 4.60 pg. As per the report nuclear DNA content (2C) was estimated to be 3.60 pg in *Pongamia glabra* and 4.0 pg in *Millettia ovalifolia* (Ohri and Kumar, 1986). Choudhury et al. (2014) estimated the GS of *in vitro* regenerated *P. pinnata* with the help of flow cytometric analysis and reported (2.51pg). For *J. curcas* flow cytometry indicates an average 2C value of 0.85 pg (Carvalho et al., 2008). The draft genome sequence and FCM analysis of the oilseed species *R. communis* reported the GS to be 1.04 pg (Houben et al., 2003; Chan et al., 2010). Also, 2C DNA amounts for more than 10,000 angiosperms from different families have been reported in Kew Royal botanical garden's Plant DNA C-value database (https://cvalues.science.kew.org) (Leitch et al., 2019). The range of GS (2C) in plants varies from 0.14 pg in *Genlisea aurea* to 304.46 pg in *Paris japonica* (Pellicer et al., 2010; Fleischmann et al.,2014) .

In the Poaceae family, which includes major cereal crops, GS estimates show significant variation. For instance, *Oryza sativa* (rice) has an estimated 1.0 pg genome, while *Zea mays* (maize) has a much larger genome of 5.4 pg (Bennett, 1976) . These variations within the same family reflect the dynamic nature of GS, which can be attributed to differences in repetitive DNA content and polyploidy events (Sliwinska, 2018). Moving to the Solanaceae family, the model plant *Solanum lycopersicum* (tomato) exhibits a GS of approximately 2.12 pg, whereas its relative Solanum tuberosum (potato) has a larger genome estimated at around 4.2 pg (Bennett and Smith, 1976; Kron and Husabnd, 2012). While these GS are relatively close, they exemplify the intra-family variability and emphasize the need for species-specific GS estimations. The Brassicaceae family, which includes *Arabidopsis thaliana*, showcases a contrast between model plants and their wild relatives. *A. thaliana* has a small genome of about 0.32 pg, while its wild relative *Arabidopsis lyrata* possesses a larger genome estimated at 0.5 pg (Bennett et al., 2003; Lysak et al., 2009). These differences highlight the influence of evolutionary processes on GS and the role of these model plants in understanding genome dynamics.

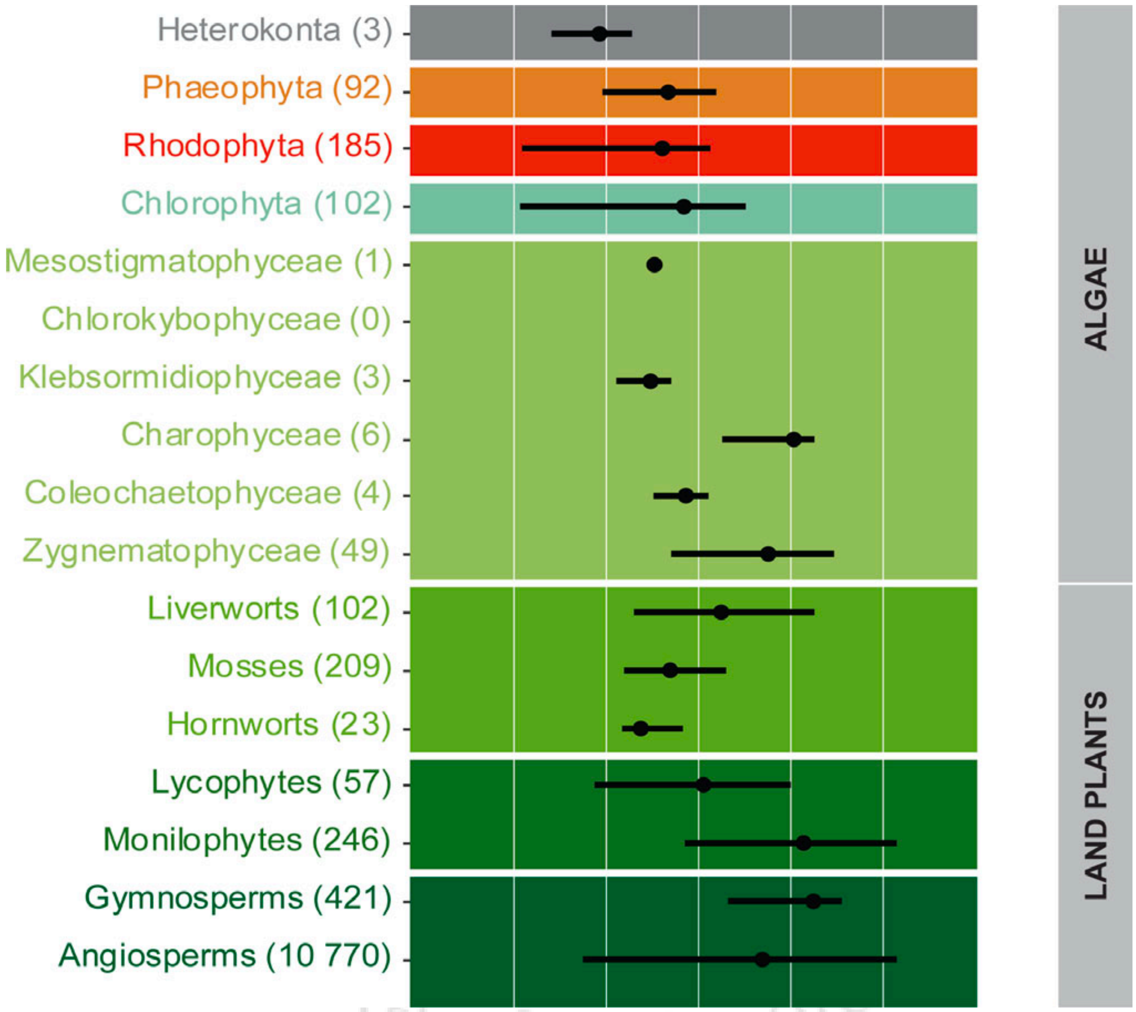


Figure 2.2: Genome size (1C-value) variability by lineage in the plant DNA C-values database (Release 7.1, April 2019): Range and mean (•) with species counts (Source: Pellicer and Leitch, 2019)

The Fabaceae family, known for its economic and ecological importance, also demonstrates notable GS variations. *Glycine max* (soybean) is estimated to have a GS of around 2.26 pg, while *Medicago truncatula*, another model legume, has a smaller genome of approximately 0.94 pg (Arumuganathan and Earle,1991; Greilhuber and Obermayer,1997). GS of more than 1,000 species from another important family Asteraceae has been reported ranging from 0.8 pg in *Leontodon longirostris* to 52.3 pg in *Senecio pedulus* (Marañón and Grubb,1993; Zonneveld et al.,2005). These estimates underscore the impact of GS on plant traits, ecological adaptations, and domestication. Furthermore, the estimation of GS in various plant families contributes to our understanding of genome evolution, polyploidy events, and their correlations with plant diversity. The largest wealth of GS data stored in Plant C-Value database generated by flow cytometry is arranged by lineage of plants is given in **Fig. 2.2** (Pellicer and Leitch, 2019).

2.2.8. Genome size variation and influencing factors

GS in plants, as estimated through various techniques, has revealed fascinating insights into genetic relatedness, population variations, and the presence of non-coding, often considered "junk" or "selfish" DNA. This review explores the factors influencing GS variation and its implications in the plant kingdom.

GS, originally considered bulk or selfish DNA (Ohno, 1972; Cavalier-Smith, 1980), is now recognized to carry phenotypic effects independent of coding functions. GS variation is subject to strong selection pressure concerning phenotypic and phenological factors (Ohri and Kumar, 1986). Studies have shown no direct correlation between GS and organism complexity, challenging earlier perceptions (Doležel and Bartoš, 2005; Sliwinska, 2018). Polyploidy, common in woody taxa, can cause GS variation among species within a genus (Mehra and Bawa, 1969). Ohri et al. reported nuclear-size ratio variation even at the same ploidy levels, especially in tropical hardwoods (Ohri and Kumar, 1986). Variations in nuclear-size ratio are dependent on the amount of repetitive DNA, influenced by micro- and macro-environmental alterations during species evolution (Mohanty and Das, 2006). Diploid cytotypes are linked to warm and humid regions, while polyploids extend into drier and cooler regions (Ohri, 1996).

GS variation is often attributed to processes like polyploidization, whole-genome duplication (WGD), and the presence of transposable elements affecting genome expansion (Neale et al., 2017). The discovery of non-coding RNAs has shed light on mechanisms controlling GS obesity, providing valuable insights into genome dynamics. The pea genome serves as a prime example of intense gene dynamics, potentially associated with GS expansion during the divergence of Fabeae from sister tribes (Kreplak et al., 2019). GS variation in plants is a dynamic and complex process influenced by factors such as polyploidy, transposable elements, whole-genome duplication, and non-coding RNAs. While earlier perceptions suggested a direct link between GS and organism complexity, recent genomic studies have challenged this notion, emphasizing the significance of repetitive elements and WGD in GS variation.

2.2.9. The impact of genome size on plant traits

GS is a fundamental characteristic of organisms, but it is not always directly proportional to their complexity (Markov et al., 2010). This review delves into the intricate relationship between GS and plant phenotypic traits, shedding light on the diverse impact of GS on cellular and ecological aspects.

Contrary to expectations, GS doesn't always correlate with an organism's complexity. Some single-celled organisms have larger GS than certain multicellular organisms. Environmental factors can significantly affect GS, subsequently influencing various phenotypic cell characteristics. Studies have revealed intriguing correlations at the cellular level (**Fig. 2.3**). These include a positive correlation between GS and cell volume and nuclear volume (Knight et al., 2005; Jovtchev et al., 2006). There is also a report of negative correlation between GS and cell cycle duration (Beaulieu et al., 2008; Diallo et al., 2016). While there have been reports of strong positive relationships between GS and cell size in the animal kingdom (Hardie and Hebert, 2003; Organ et al., 2007), the findings in large taxonomical groups have not yielded consistent results. This suggests that the GS-cell size correlation is species-dependent (Greilhuber and Leitch, 2012).

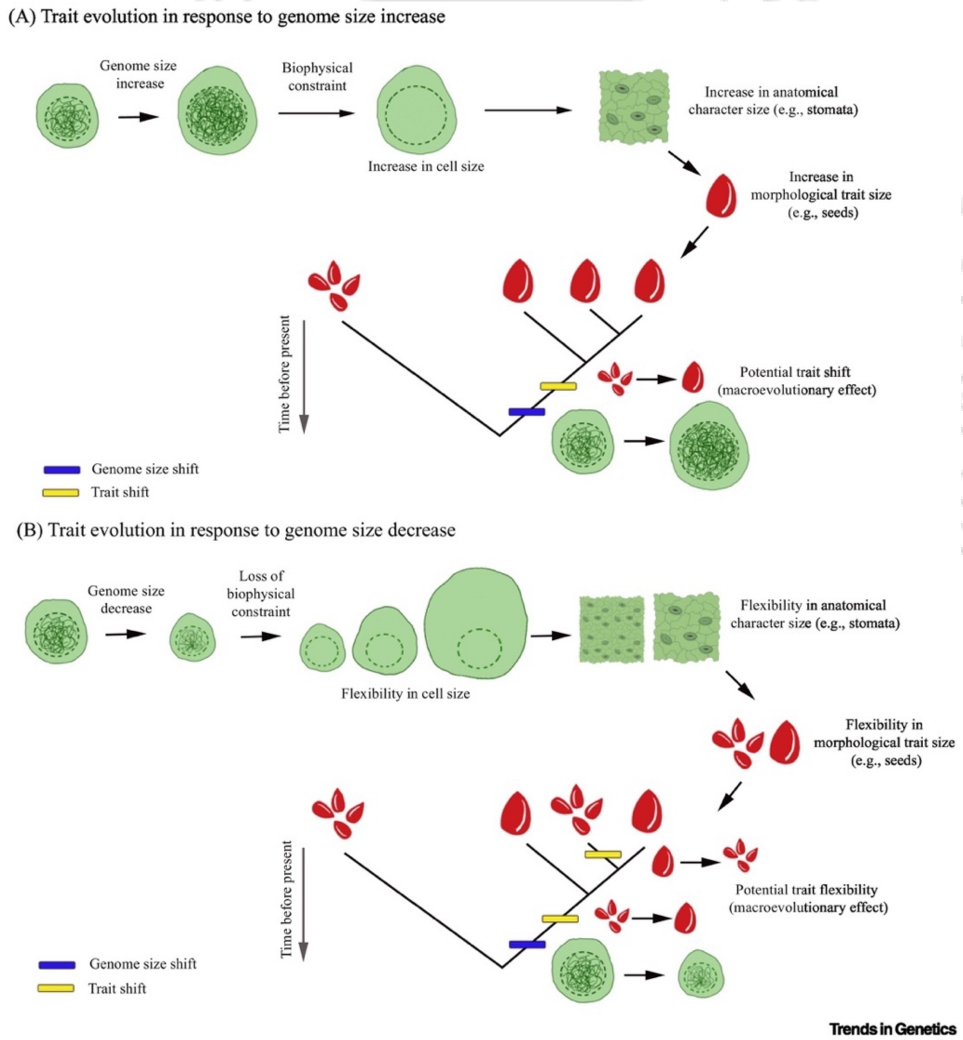


Figure 2.3: Schematic representation illustrating the relationship between GS change and trait evolution (Source: Bhadra et al., 2023)

Stomatal length has been associated with nuclear DNA content. Larger nuclear DNA content is linked to longer guard cells, resulting in increased stomatal length (Khazaei et al., 2010; Jordan et al., 2015; Das and Rangan, 2020). A study by Gruber et al. (2010) explored the relationship between GS and root meristem growth rate, hypothesizing significant consequences for GS expansion in plants. Pollen DNA content has been examined using flow cytometry, aiding in the understanding of pollen grain development and the detection of unreduced gametes (Kron and Husband, 2012). The relationship between GS and leaf mass per unit area (LMA) has been suggested as a genetic driver for this trait. However, the relationship appears to be weak and species-dependent (Beaulieu et al., 2007). A strong positive relationship between GS and seed mass has been proposed. However, further analyses have revealed that this relationship may not be as robust as initially thought. GS has been linked to environmental conditions such as temperature, water availability, latitude, and elevation (Knight and Ackerly, 2002). These environmental predictors may contribute to GS variation as reported in maize (Ohri, 2005; Greilhuber and Leitch, 2012). Significant variation in plant GS within species remains largely unexplained. The exploration of the connection between intraspecies GS variation and plant phenotypic traits is still going on.

2.2.10. Challenges and future prospects

Challenges in the realm of plant GS estimation using flow cytometry are multifold. Firstly, the accuracy of the estimates heavily relies on the availability of appropriate internal standards and reference materials (Dolezel and Bartos, 2005). These standards need to be meticulously characterized and maintained, posing a significant challenge for researchers (Dolezel and Greilhuber., 2010). Additionally, obtaining a representative tissue sample for analysis can be challenging, especially in the case of plants with extensive tissue variations or mixtures of tissues in their organs (Čertnerová, 2022). Maintaining the quality of isolated nuclei or intact cells during the sample preparation process is another hurdle, as any degradation or contamination can lead to inaccurate results. Furthermore, not all plant species are compatible with flow cytometry, and the development of suitable protocols for recalcitrant species remains a constant challenge (Galbraith et al., 2021). The lack of standardization in flow cytometry procedures across laboratories and platforms can also lead to variations in results, emphasizing the need for more robust protocols and inter-laboratory collaborations (Sliwinska et al., 2022).

Looking ahead, the future prospects of plant GS estimation using flow cytometry are promising. The continuous advancement of flow cytometry instrumentation and software tools enables researchers to obtain more precise and reliable measurements (Manohar et al., 2021). The development of reference standards, particularly for non-model plant species, is an active area of research, which will enhance the accuracy and comparability of GS estimates. The integration of flow cytometry with other technologies, such as next-generation sequencing, can provide a more comprehensive understanding of plant genomes (Qiu et al., 2019).

2.3. Plastome mining of selected small genome sized plants

2.3.1. Introduction

Plastome mining, the meticulous sequencing and in-depth analysis of chloroplast genome, stands at the forefront of scientific exploration in the realm of plant biology (Rabah et al., 2017; Qu et al., 2023). This pioneering approach holds immense promise, particularly when applied to small genome-sized plants, as it offers a window into the genetic intricacies that govern these organisms (Sato et al., 1999; Dhingra and Folta, 2005). Chloroplasts, commonly referred to as the cellular powerhouse of plants, are essential for photosynthesis and the synthesis of crucial compounds (Kazakoff et al., 2012). As a result, the chloroplast genome represents a veritable goldmine of genetic data that can unlock a wealth of information concerning the evolutionary history, phylogenetics, and potential applications of these plants in various domains (Daniell et al., 2016; Dobrogojski et al., 2020).

In this comprehensive review of literature, we embark on an in-depth journey to unravel the recent developments and breakthroughs in plastome mining, with a particular focus on selected small genome-sized plants. By delving into the intricacies of this research, we aim to emphasize its paramount importance in advancing our understanding of plant biology, ecology, and biotechnology. Within these pages, we will explore the methodologies employed in chloroplast genome sequencing and the profound implications it carries for the scientific community (Dobrogojski et al., 2020). This analysis not only seeks to shed light on the burgeoning wealth of knowledge amassed in this field but also underscores the multifaceted contributions that plastome mining has made to the broader scientific landscape, showcasing its role in shaping the future of plant biology research (Zhang et al., 2020).

From the elucidation of evolutionary relationships to the development of novel biotechnological applications, plastome mining opens new vistas for scientific inquiry and discovery (Daniell et al., 2016). By unveiling the secrets encoded within the chloroplast genomes of selected small genome-sized plants, we aim to provide an expansive and detailed account of the achievements and ongoing efforts in this dynamic field of study. In doing so, we hope to inspire further exploration, collaboration, and innovation, ultimately enriching our collective understanding of the botanical world and harnessing its potential for the betterment of society (Bi et al., 2018).

2.3.2. Plastome mining and its relevance

Plastome mining, a specialized research approach, has gained prominence in the realm of plant biology, particularly in understanding the role of chloroplast genome in various aspects of plant biology (Daniell et al., 2016). Chloroplast, essential organelle responsible for photosynthesis and the synthesis of crucial compounds in plant cells, possess their own unique genome, making them a focus of significant scientific interest (Boardman et al., 1971; Waters and Langdale., 2009). By meticulously sequencing and analyzing chloroplast genomes, researchers have been able to make substantial contributions to the elucidation of the evolutionary history, taxonomy, and phylogenetic relationships within different plant species (Shelke and Rangan, 2022).

Plastome mining primarily distinguishes itself through its contributions to the discipline of plant taxonomy and phylogenetics as reported in Poaceae, Orchidaceae and Asteraceae (Burke et al., 2016; Zhang et al., 2019; Kim et al., 2020). Chloroplast genomes harbor valuable genetic data that assists in the determination of evolutionary relationships among plant species, thereby enhancing our understanding of plant diversity and origins (Simeone et al., 2016). Furthermore, plastome mining is instrumental in resolving taxonomic ambiguities, facilitating precise classification of plant species, and supporting conservation efforts (Dhingra and Folta, 2005). Beyond its taxonomic implications, plastome mining also offers practical applications within the realms of biotechnology and plant breeding (Rogalski et al., 2015; Řepková, 2010). The genetic information extracted from chloroplast genomes can be harnessed to develop markers for plant identification, genetic mapping, and the genetic improvement of crop species (Bansal and Saha, 2012). This accelerates the breeding of agriculturally significant plants, enhancing their resilience to environmental stressors and improving crop yields (Olejniczak et al., 2016). Additionally, plastome mining has the potential to identify unique plant traits and compounds, which may find applications in fields such as medicine and biofuel production (Shelke et al., 2022).

In recent times, plastome mining has emerged as a crucial tool for the comprehensive exploration of plant biology, offering insights into the genetic nuances encoded within chloroplast genomes (Rabah et al., 2017). Its significance ranges from advancing our comprehension of plant evolutionary history and taxonomy to providing practical solutions for biotechnology and plant breeding (Zhang et al., 2019). This dynamic field is poised to continue unlocking the mysteries of the plant kingdom, with profound implications for scientific research and broader applications in the domain of plant biology (Dobrogojski et al., 2020).

2.3.3. Plastome structure and function

The plastome, also known as the chloroplast genome, plays a vital role in plant biology. It's closely connected to how plant cells work and is like a mini-version that reflects plant evolution, how they adapt, and their ability to perform photosynthesis (Olejniczak et al., 2016). This makes it a subject of a lot of research and investigation in the scientific community. The availability of complete plastome sequences for terrestrial plants has greatly improved due to the advancements in next generation sequencing (NGS) methods and the utilization of sophisticated bioinformatics tools (Vera-Paz et al., 2022).

The structure of the plastome exhibits remarkable conservation across a wide range of plant species. It typically comprises a circular, double-stranded DNA molecule with a size that can vary between 120 to 160 kb among land plants (Li and Zheng, 2018). The genome is generally a quadripartite structure with a small single copy (SSC) region, large single copy (LSC) region and two inverted repeats (IRs) in almost all cases (Olejniczak et al., 2016; Wang and Lanfear, 2019) (**Fig. 2.4**). While the plastome is relatively small in comparison to the nuclear genome, it plays a crucial role in governing the biology of the chloroplast, a vital organelle responsible for photosynthesis. The plastome is compact, containing a set of essential genes, organized into clusters, and surrounded by non-coding regions. These genes encode proteins, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) that are essential for chloroplast function (Haris et al., 1994; Alshegaihi et al., 2023).

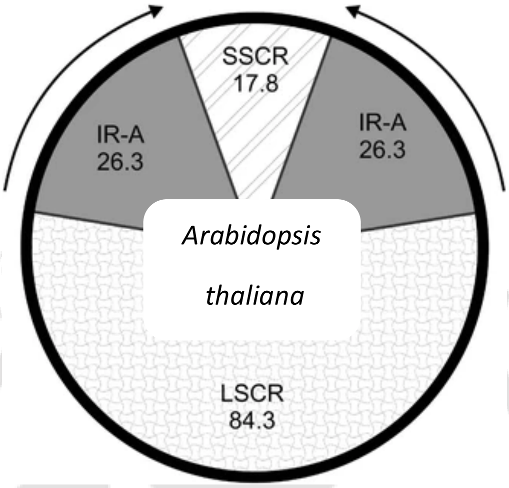


Figure 2.4: Structure of the *Arabidopsis thaliana* chloroplast genome. (SSCR small single copy region, LSCR large single copy region, and IR-A and IR-B inverted repeats. Length of plastome and its parts in kbp) (Source: Olejniczak et al., 2016)

The function of the plastome is primarily cantered around photosynthesis, a process by which plants convert light energy into chemical energy in the form of glucose (Palmer, 1985). This is achieved through the synthesis of chlorophyll, which captures light energy, and a cascade of biochemical reactions in which the plastome-encoded genes play a pivotal role (Howe et al., 2003). Moreover, the plastome participates in the production of various metabolites, such as amino acids, lipids, and secondary compounds, that are integral to plant growth and defence mechanisms (Zhang et al., 2020). Beyond photosynthesis, the plastome is also involved in critical cellular processes such as the regulation of redox balance, gene expression, and responses to environmental stimuli (Llamas et al., 2017; Zhang et al., 2023). Recent studies in *Oryza sativa*, and Amaryllidaceae and many more have revealed the plasticity of plastome structure, enabling plant adaptation to diverse environmental conditions (Gao et al., 2019; Scobeyeva et al., 2021). In multiple cases, allowing for the capture of foreign DNA sequences through horizontal gene transfer from nucleus and vice-versa has also been observed (Bock and Timmis, 2008; Cullis et al., 2009). As research continues to unravel the complexities of the plastome, we gain further insights into the fundamental mechanisms that drive plant growth, development, and adaptation in a changing world.

2.3.4. Small genome sized plants and their advantages

Small genome-sized plants, characterized by their compact and streamlined genetic makeup, have garnered significant attention in the field of plant genomics, especially in the context of chloroplast sequencing (Kelly et al., 2012). As pointed out by Claros et al. assembling the plant genome is quite a challenging task, where small genome sized plants have an advantage (Claros et al., 2012).

While the size of the GS may not always indicate genome complexity, having a smaller GS offers several advantages. These include a shorter life cycle, faster growth rate, genetic stability, and generally more straightforward downstream bioinformatics analysis in the majority of cases (Grotkopp et al., 2004; Lavergne et al., 2010; Michael, 2014). This simplicity in their genomic organization accelerates the process of DNA sequencing, making it more cost-effective and less time-consuming compared to larger genome species (Li and Harkess, 2018). Furthermore, the compact genomes of these plants often lack the intricate repeat sequences that can pose challenges in assembly and annotation (Tyagi et al., 2020). This makes them ideal candidates for chloroplast sequencing, as it allows for the accurate determination of the chloroplast genome structure and gene content (Jian et al., 2018). Small genome-sized plants are particularly advantageous in chloroplast sequencing due to their potential to yield highly accurate and complete chloroplast genome sequences. The smaller genome size results in fewer gaps or ambiguities in the assembled sequences, thereby enhancing the quality of the data obtained. This precision is vital for various applications, including phylogenetic studies, molecular marker development, and evolutionary analyses (Li and Harkess, 2018).

It has also been reported by Twyford and Ness (2017) that the amount of plastid data in a gDNA library varies depending on GS of the species, in sugarcanes it is < 0.5% (Hoang et al., 2015) whereas, it is over 20% in milkweeds (Straub et al., 2012) (**Fig. 2.5**). Moreover, the chloroplast genome of small-sized plants tend to be more conserved, making them highly informative for comparative genomics and phylogenetic investigations as reported in Actinidiaceae and Poaceae plants (Salse and Feuillet, 2007; Yang et al., 2022). Researchers can utilize these sequences to explore the evolutionary relationships between different plant species and infer the processes that have shaped their genomes over time (Daniell et al., 2016). Small genome-sized plants also exhibit greater potential for uncovering unique genetic features, such as structural variations, gene rearrangements, or novel functional elements within the chloroplast genome (Yue et al., 2008; Xu et al., 2023). These advantages contribute to the broader understanding of plant biology and have far-reaching implications in areas such as taxonomy, conservation, and biotechnology (Bansal and Saha, 2012).

Tandem repeats within chloroplast genomes have gained significance due to their association with structural rearrangements and their utility in taxonomic investigations. In a recent study conducted by Chen et al. (2023), conserved sequences derived from chloroplast genomes were identified as potential DNA barcodes for the discrimination of *Clerodendrum* spp. within the Lamiaceae family. Similar investigations have been undertaken in various plant taxa, including *Allium* spp. by Huo et al. (2019), five species within the Solanaceae family by Mehmood et al. (2020), four medicinal plants from

Salvia spp. by Liang et al. (2019), and in seven species of *Cissus* spp. by Senapati et al. (2023), all of which have reported the discovery of valuable tandem repeat sequences.

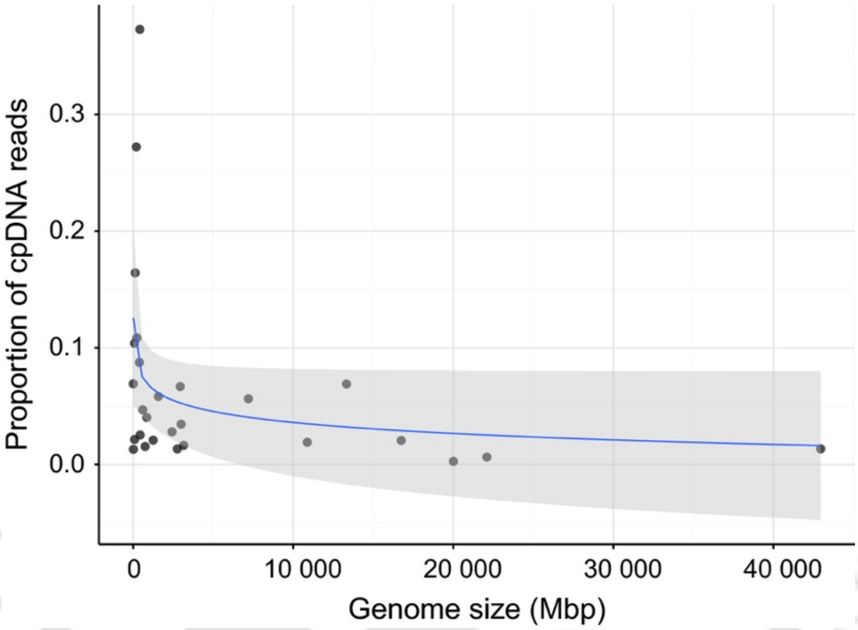


Figure 2.5: The graph illustrates how plastid reads are distributed within genomic DNA sequence libraries across a set of 27 diverse green plant species with varying genome sizes. It displays the percentage of sequence reads that align with a reference database containing 100 plant plastomes. The shaded area in grey represents the 95% confidence interval of the regression line (Source: Twyford and Ness, 2017)

Simple sequence repeats (SSRs), commonly referred to as microsatellites, are tandem repeats characterized by their short repetitive units within the chloroplast genome. These SSRs have demonstrated their utility in genetic diversity assessments and population studies. Notably, Gu et al. (2019) employed SSR analysis in 22 species and Zheng et al. (2020) in 13 species of Lythraceae to elucidate relationships among the members. Furthermore, Zhao et al. (2021) identified DNA markers and resolved the phylogenetic relationships within 14 species of *Zanthoxylum* spp. Additional studies have explored SSRs in *Messua ferrea* (Shelke and Rangan, 2022), *Callitropsis funebris* (Ping et al., 2021), and three *Achyranthes* spp., emphasizing the importance of SSRs in marker development and the assessment of genetic diversity.

The analysis of genetic diversity within chloroplast genomes serves as a valuable avenue for gaining insights into the evolutionary history and adaptive potential of plants. Recent studies frequently employ diversity metrics like nucleotide diversity (*Pi*), as demonstrated by He et al. (2020), being a widely utilized tool in assessing genetic variation. *Pi* analysis has been applied to examine genetic diversity in various plant species, including watermelon (Cui et al., 2020), Japanese apricots (Huang et al., 2022), thirty-six species within the Asteraceae family (Loeuille et al., 2021), and e Rosaceae family (Mu et al., 2021). Additionally, a study focusing on seven Fagaceae plants, conducted by Shelke et al. (2022), revealed five distinct hotspot coding and noncoding regions, indicating the presence of unique genetic diversity. Ka/Ks analysis, which measures the non-synonymous to synonymous substitution rate, has been employed to investigate the selective pressures acting on protein-coding genes within chloroplast genomes throughout species evolution. Wu et al. (2020) utilized Ka/Ks analysis to study six *Chrysosplenium* spp. shedding light on the selective pressures influencing these genes. Furthermore, the investigation of selective pressure on essential genes within *Eruca sativa* (Brassicaceae) (Zhu et al., 2021), *Amomum* spp. (Gong et al., 2022), and *Cissus* spp. (Senapati et al., 2023) has provided valuable insights on the evolutionary history of these plants.

Codon usage analysis assumes significance in comprehending the protein-coding genes within chloroplast genomes, offering insights into evolutionary processes where codon usage bias and preferences play a crucial role. A recent study by Li et al. (2023) delved into the codon usage bias within the chloroplast genomes of nine *Elaeagnus* spp. (Elaeagnaceae) and observed a preference for A/T bases in the third position. This trend mirrors findings in *Panicum* spp. (Li et al., 2021), *Camellia* spp. (Yengkhom et al., 2019), and *Cissus* spp. (Senapati et al., 2023). Wang et al. (2020) reported that natural selection, rather than mutation pressure, underlies codon usage bias in Euphorbiaceae plants. Similar patterns of codon usage bias have been identified in *Porphyra umbilicalis* (Li et al., 2019), Theaceae plants (Wang et al., 2022), *Panicum* spp. (Li et al., 2021), and *Malus* spp. (Li et al., 2023). These studies collectively underscore the value of codon usage analysis in unraveling the intricacies of chloroplast genome evolution. These studies showcase the diversity of analyses conducted on chloroplast genomes of small-sized plants and their applications in understanding plant evolution, taxonomy, and adaptation. Please verify the sources' credibility and check for any updates or related publications in this field, as the research landscape is continuously evolving.

2.3.5. Plastome sequencing and annotation techniques

The sequencing and annotation of plastomes (chloroplast genomes) in plants have witnessed significant advancements in recent years, thanks to a plethora of innovative methods and techniques (Dobrogojski et al., 2020). These approaches have revolutionized our understanding of plant biology, phylogenetics, and the molecular basis of various traits (Kim et al., 2020). Here, we review several methods of plastome sequencing and annotation, shedding light on the state-of-the-art techniques that have become essential in modern plant genomics.

One of the traditional methods of plastome sequencing involves Sanger sequencing (Wysocki et al., 2014). Although it has been largely superseded by NGS technologies, Sanger sequencing remains relevant in cases where ultra-high accuracy is required (Jansen et al., 2005). It is particularly useful for finishing and validating plastome assemblies generated through NGS (Straub et al., 2012). The Illumina platform, a prominent NGS technology, has played a pivotal role in the rapid and cost-effective sequencing of plastomes (Twyford and Ness, 2017). Its high throughput and short-read sequencing capabilities make it well-suited for this purpose. However, plastome assembly from short reads can be challenging due to the presence of repeated sequences in the chloroplast genome (Straub et al., 2012). The GetOrganelle toolkit represents an advanced and freely accessible resource for the precise reconstruction of organelle genomes from whole genome sequencing data. It employs a refined methodology known as "baiting and iterative mapping" to selectively identify organelle-associated reads. Following this, it proceeds with de novo assembly, assembly graph refinement, and the generation of all conceivable circular organelle genome configurations. Notably, the accuracy of these reconstructed organelle genomes is notably superior in comparison to alternative methods (Jin et al., 2020).

Another significant NGS technology, long-read sequencing using platforms like PacBio and Oxford Nanopore, has become indispensable for de novo plastome assembly (Syme et al., 2021; Zhou et al., 2023). The longer reads produced by these platforms aid in resolving complex regions, such as repetitive elements and inverted repeats, which are prevalent in plastomes. This results in more complete and accurate plastome sequences (Scheunert et al., 2020). Moreover, these long-read technologies facilitate the detection of structural variations as reported in *O. sativa*, and *Acacia pycnantha*, which is crucial for studying chloroplast genome evolution and phylogenetics (Takamatsu et al., 2018; Syme et al., 2021; Pucker et al., 2022).

Plastome annotation has also seen substantial innovation. Homology-based annotation, comparing newly sequenced plastomes to well-annotated references, remains a standard approach (Kahraman and Lucas, 2019). However, as the number of available plastome sequences continues to grow, automated annotation tools have become indispensable. Tools like DOGMA (Dual Organellar Genome Annotator) (Wyman et al., 2004), Mitofy (Alverson et al., 2010), CpGAVAS (Liu et al., 2012), Verdant (McKain et al., 2017), and GeSeq (Tillich et al., 2017) offer rapid and reliable annotation, incorporating structural and functional information. In recent times, hybrid annotation methods that combine homology-based and ab-initio predictions have gained prominence. These approaches leverage the strengths of both techniques to enhance annotation accuracy (Liu et al., 2012). Additionally, machine learning-based methods are emerging as powerful tools for gene prediction, particularly in plastomes with non-canonical genetic codes or unique structural features (Mahood et al., 2020). With NGS technologies, long-read sequencing, and innovative annotation tools plastome sequencing and annotation techniques have evolved significantly. These methods have not only facilitated the rapid generation of plastome sequences but have also improved their quality and accuracy.

2.3.6. Plastome sequencing and phylogenetics

Phylogenetic and evolutionary studies of plant species have greatly benefited from the advent of plastome sequencing, providing valuable insights into the evolutionary history, relationships, and adaptation of various plant lineages (Gitzendanner et al., 2018). The significant contributions that plastome sequencing has made to our understanding of plant phylogenetics and evolution, backed by pertinent examples and scientific sources (Gao et al., 2010).

Plastome sequencing has emerged as a potent tool in addressing longstanding inquiries within the realm of plant phylogenetics. In a study conducted by Kim et al. (2019), 28 plastome sequences were harnessed to provide a definitive classification of the *Lilium* genus. Another research endeavour by Wang et al. (2021) elucidated the interrelationships among eight subfamilies within the Malvaceae family, employing 78 protein-coding genes. Furthermore, an extensive plastome phylogenetic analysis encompassing 124 orchid species, not only confirmed the relationships within the Apostasioideae subfamily but also established the phylogenetic associations among 17 distinct tribes (Kim et al., 2020). The reconstruction of ancestral plastome structures has been a notable outcome of plastome sequencing, enabling scientists to gain insights into the evolutionary events that have shaped the diversity of plant lineages. For example, the loss of the IR region in the plastome, known as the IR-lacking clade, has been a subject of interest. The study by Ruhfel et al. (2014) reconstructed the ancestral plastome of land plants and identified multiple independent losses of the IR region during land plant evolution (Ruhfel et al., 2014).

Plastome sequencing has played a pivotal role in the exploration of adaptive evolution among plant species. Convergent evolution, a phenomenon wherein unrelated species develop akin traits due to similar environmental pressures, has been a focal point of investigation. Notably, the examination of plastome data has revealed a higher incidence of convergent evolution in 64 C4 plants compared to C3 plants, as established by Casola and Li (2022). Furthermore, plastome-based analyses have been employed to unravel the evolutionary histories of various plant species. For instance, a study by Shen et al. (2020), which cantered on plastome data from four Asteraceae species, shed light on the placement of Dolomiaea as a clade within the Saussureinae subtribe. Another significant facet of plastome sequencing is the detection of gene loss events during the course of plant evolution. As exemplified by Gao et al. (2010), certain parasitic plant species have undergone specific losses of plastid genes associated with photosynthesis, attributed to the relaxation of functional constraints. With the continued sequencing and analysis of plastomes from various plant species, we can anticipate further significant advancements in our comprehension of the intricate evolutionary narratives and adaptive mechanisms within the plant kingdom.

2.3.7. Applications of plastome mining

Plastome mining, or the sequencing and analysis of chloroplast genomes, has ushered in a new era of research with far-reaching applications in plant biology. This section delves into the diverse applications of plastome mining, supported by accurate examples and corresponding sources, highlighting the transformative impact of this approach on the study of plant genomes.

Plastome mining has proven instrumental in resolving intricate phylogenetic relationships among plant species. A prime example is the study by Niu et al. (2017), where plastome sequences were employed to elucidate the complex evolutionary history of the orchid family (Orchidaceae). Plastome sequencing has facilitated the accurate classification of plant species, particularly in groups with challenging taxonomic boundaries. The accurate classification of Liliaceae, Polygonaceae, and Lauraceae species was achieved using plastome data (Kim et al., 2019; Song et al., 2020; Liu et al., 2021). Zhao et al. (2021) utilized plastome sequences to revise the classification of the Lamiaceae.

Plastome mining has revolutionized the development of molecular markers for plant identification and population genetics. For instance, a study by Park et al. (2019) harnessed plastome sequences to identify markers for the medicinally important genus *Angelica* sp. These markers are crucial for assessing genetic diversity and assisting breeding programs (Park et al., 2019). Plastome mining is indispensable in studies of plant phylogeography and conservation genetics. A study by Dong et al. (2017) was focused on plastome data to investigate the genetic diversity and population structure of the endangered tree species *Juglans hopeiensis*. The insights gained from plastome analysis contributed to conservation efforts for this rare species. Plastome sequences have been harnessed for genetic engineering and biotechnological applications. Plastome mining has catalysed a multitude of applications in plant biology, spanning phylogenetic reconstruction, taxonomy, molecular marker development, conservation genetics, and biotechnology.

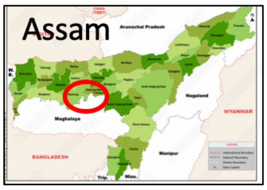
2.3.8. Challenges and future prospects

Plastome sequencing has undoubtedly revolutionized our understanding of plant genomics, yet it is not without its challenges and future prospects. Plastome sequences are typically conserved within a species. However, recent studies, such as that by Cho et al. (2021) have highlighted instances of intraspecific plastome variation in *Malus* spp. This challenges the assumption of plastomes being highly conserved and necessitates a more comprehensive sampling of individuals within a species. Plastome sequences can exhibit structural rearrangements, such as inversions and rearrangements of genes as observed in *Medicago* and its relatives (Wu et al., 2021). Identifying and correctly assembling these variations in plastomes can be challenging. Plastome sequencing is biased towards certain plant groups, leaving gaps in our understanding of plastome diversity across the entire plant kingdom. For example, while there are extensive plastome data available for angiosperms, gymnosperms and non-seed plants are underrepresented.

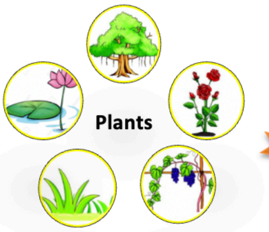
Although chloroplast genome sequencing has advanced quite well in recent times advancements in high-throughput sequencing technologies like Illumina and long-read sequencing platforms like PacBio and Oxford Nanopore, hold great promise for overcoming current challenges. The application of metagenomics and environmental sequencing to plastome research has the potential to uncover plastomes from diverse environments, offering insights into the plastome diversity of non-model species (Olejniczak et al., 2016). Future prospects for plastome research extend to functional genomics and biotechnological applications. The development of transplastomic plants was achieved for the production of high-value crops (Yarra, 2020). plastome sequencing has made significant strides, but challenges persist, particularly related to intraspecific variation, structural complexities, and taxonomic coverage. However, with ongoing developments in sequencing technologies and innovative approaches such as metagenomics and functional genomics, the future of plastome research is promising.

Chapter 3

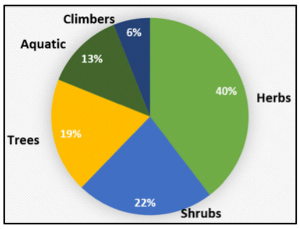
Floral Diversity of IIT Guwahati



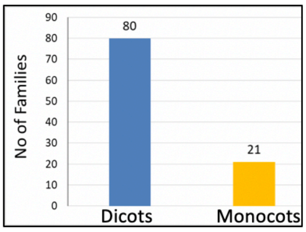
(a) Study area



(b) Collection and identification of plants



(c) Floral composition



(d) Distribution of monocots and dicots

Floral diversity of North Guwahati

Taxa	Dicots	Monocots	Total
Family	91	21	112
Genus	257	67	324
Species	352	103	455

(e) Statistics of flora

(f) Inventorization

3.3. Introduction

India, one of the major biodiverse nations occupies only 2.4% land area of the earth and harbors 7-8% of world flora. While following the path of development, the nation has been sensitive towards the conservation of its rich plant genetic resources. Currently, the country is home to 18,666 species of flowering plants under 2,991 genera and 251 families (Mao, 2019). India has secured the fourth position in Asia and tenth in the world in terms of plant diversity (Sinha et al., 2010). There are 17 megadiverse countries identified in the world and India is one among them harboring about 12% of the world's flora (UN-WCMC). Four biodiversity hotspots out of 34 globally, fall in India or parts of it. The Himalaya hotspot falls in the north and north eastern (NE) parts of the country, the Indo-Burma hotspot is limited to NE states, the Western Ghats- Sri Lanka hotspot is located in the south western (SW) part of the country and the Sundaland falls in the eastern islands of Nicobar. A large number of crops like maize, millets, rice etc. display huge diversity within India and about 23% of them are endemic to this country (Arisdason et al., 2023). The reason behind such a huge diversity of plants is the geographical location and climatic condition of the country (Singh and Chaturvedi, 2017). This country is a combination of all possible terrain types from deserts to snow lands with number of rivers and extensive coastline supporting different types of ecosystems including deserts, forests, wetlands, grasslands and marine (Roy and Purohit, 2018). NE region is one of the richest biodiversity sources found in the country with highest orchid diversity (Tripathi et al., 2016; Ninawe and Swapna, 2017).

The NE region, located between 87°32' E to 97°52' E longitudes and 21°34' N to 29°50' N latitude with tropical to subtropical climate influenced by SW monsoon comprises Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura states. The NE India is considered as paradise of biodiversity in India carrying almost half of the country's floral wealth and falls under two biodiversity hotspot regions i.e. the Himalaya and the Indo-Burma hotspot (Tripathi et al., 2016). The valuable bioresources of this region can be a center of attraction for the economic activities of the country with proper planning and management. In last few decades, the loss of biodiversity and forest degradation has increased remarkably due to the rising demand for natural resources for the huge population in different ways like industrialization, overexploitation, deforestation etc.

The NE floral wealth is rich in medicinal, crop, timber, fruit and flower yielding plants including a number of rare and endangered species (Begum et al., 2010; Dikshit and Dikshit, 2014). Floristic diversity studies are necessary for recognizing evolution, endemism, speciation and more importantly forest management of the region (Gogoi and Nath, 2021). In this study, inventorization of floral diversity of North Guwahati region on the bank of River Brahmaputra along with the species distribution and their uses has been carried out and to the best of our knowledge, this is the first attempt to carry out such a study.

3.3.1. NE India- a paradise of floral diversity

According Ministry of Environment and Forests (MoEF, 2014) the Andaman and Nicobar Islands, Western Ghats, parts of Eastern Ghats and NE India are considered as biodiversity rich regions in India (MoEF, 2014). The Indo-Burma hotspot region falls in NE Indian states, central Nepal, parts of southern China, Myanmar, Thailand and islands of Andaman and Nicobar including parts of Malaysia. The Indo-Burma hotspot is the second largest beside Mediterranean basin and all eight states of NE India come across either Himalaya or Indo-Burma hotspot (Gillette et al., 2022). Even after decades of biodiversity degradation due to natural and anthropogenic activities, the NE India still remains the paradise of biodiversity in India, along with the highest area of forest cover (Tripathi et al., 2016).

Due to its unique geography and highly favorable climatic conditions, NE India constitutes few of the highly diverse and globally recognized biomes. This region is home to almost half of the Indian flora growing in various types of forests like evergreen forests, temperate forests, alpine forests, including grass lands. The variety of forest types favors the growth and development of diverse plant groups. Orchids, bamboos, zingers, canes, and many crop species exhibit extensive diversity in this region. The NE region considered as the origin of many cultivated plants like rice, citrus, bamboos etc. (Barooah and Borthakur, 2003). This region harbors almost 50% flowering plants of the Indian subcontinent (Tripathi et al., 2016)

3.3.2. Climate and forest types

Predominantly tropical climate is observed in the valleys of NE India due to close proximity to Tropics. Three seasons can be perceived in this region, summer, rainy and winter. The summer starts from June till September, and the rainy season coincides with summer with June experiencing highest rainfall mainly due to the southwest monsoon. The climate is different in valleys and mountains with mean January temperature of 16 °C in valley regions. Whereas, the highest of 14 °C with minimum of sub-zero temperature can be observed in mountainous regions of Arunachal Pradesh, Sikkim and Nagaland. The hilly regions have a mean temperature of 20 °C in contrast to plains with mean temperature around 30 °C in summer days. Snowfall is limited to parts of Arunachal Pradesh and Sikkim (Dikshit and Dikshit, 2014).

All the NE states receive more than 1,000 mm of rainfall annually. Mawsynram, of East Khasi Hills district in Meghalaya has replaced Cherapunji to be the place with highest rainfall in the earth with annual average rainfall of 12,717 mm in 2022 (IMD). The Brahmaputra valley receives an average of 2,000 mm of rainfall, Guwahati being in the rain shadow of Meghalaya plateau gets an average 1,717 mm of rainfall annually (Dikshit and Dikshit, 2014). Broadly five types of vegetations viz. 1. Tropical, 2. Subtropical, 3. Temperate, 4. Sub-alpine, and 5. Alpine are observed in NE region. Through remote sensing and investigation of databases Tripathi et al. (2016) identified 19 subtypes of forests in NE India and reported moist deciduous forest as the dominant forest covering 33,900 km^2. Other leading forest types include semi-evergreen forests spreading across 13,942 km^2, Himalayan moist temperate forest over an area of 24,559 km^2, and subtropical broadleaf forests with 26,241 km^2 coverage in NE India.

3.3.3. Importance of floral diversity

The diversity can be summed as compositional diversity, structural diversity and the functional diversity present within a forest. The magnitude of today's biodiversity is a result of migration, speciation, extinction and other human activities over 3.5 billion years. In Latin, flora means "the goddess of plants". According to botanical survey of India (BSI) floral diversity is the presence of various type of plants in a specific region in a particular era. Biodiversity degradation hampers sources of income, food supply chain, medicine industries and energy sources directly or indirectly (Saikia et al., 2006; Tripathi et al., 2016). Biodiversity acts as the cushion for life forms during the adaptation to changes. There are many plants used for treatment of different diseases and ailments by local communities, which is only possible if the biodiversity is rich in the region. NE India Assam in particular is rich in biodiversity and some of the highly diverse crops like citrus (52 var.), maize (15 var.), banana (14 var.), bamboos (60 var.), sugarcane (15 var.) and rice (more than 9000 var.) are found here (Roy et al., 2015).

3.3.4. Importance of community composition

Community composition can be defined as the identification and relative abundances of all taxa in a specific community (Hanson et al., 2013). Species richness, and species evenness are the important components in community composition analysis. Species richness refers to the number of species present in a particular area and evenness refers to the distribution pattern of a species. Species diversity, another important aspect of community composition, depends on species richness and species evenness. Many factors like species interactions, abiotic factors, disturbance levels, and anthropogenic events influence the composition of a community. Some species, such as foundation species, which have a strong role in the community and can occupy any trophic level (Dayton, 1972), and keystone species, which defines the community, are two very important factors in determining the composition of a community.

In case of plants, the community can be divided into trees and timber yielding plants, edible plants, economically important plants, orchids, medicinal plants, parasitic plants, invasive plants etc. (Bora et al., 2003; Parida et al., 2020). Plant community and its heterogeneity influences the survival and functioning of other species. Any change in composition of a community may result changes in the species richness resulting in changes to the native biodiversity as all the organisms are interdependent (Tripathi et al., 2016).

3.3.5. Importance of species distribution

Species distribution is the spatial arrangement of any taxon in a particular area. It can be uniform, random, or clumped. Distribution pattern is dependent on biotic, abiotic factors, and human activities.

The abiotic factors include sunlight, humidity, temperature, soil pH, water availability etc. and the biotic factors include disease, competition, breeding, predation etc. (Leach et al., 2016; Zhang et al., 2021). Species distribution pattern is important in determining future landscape changes and necessary steps required for biodiversity conservation. It also provides information regarding dispersal, survival and growth of a species; in return development of projection models to tackle future climate and biodiversity changes (Srivastava et al., 2019). Mapping plant distribution patterns is considered as an reliable approach for conservation of plant biodiversity and endemism (Spiers et al., 2018).

3.3.6. Novelty of this research

The rich biodiversity of the country imposes a huge task on ecologists and environmentalists for their biodiversity assessment and conservation strategies (Tripathi et al., 2016). NE region with diverse terrain and forest types, occupies only 7% land and harbors about 850 species of the endemic floral wealth of the country (De and Medhi, 2014). So, complete assessment of the biodiversity in this region is difficult as well as demanding. The rich plant diversity in this region has gone through many thick and thin in recent years due to various abiotic, biotic factors and anthropogenic activities. The increase in population, developmental projects, extraction of oil and natural resources, as well as natural calamities are affecting the biodiversity directly or indirectly (Buragohain, 2011). The objectives of this research include highlighting the floral diversity observed in the NE Brahmaputra valley, and studying the relatively less-explored North Guwahati region, focusing on documenting the composition of the local flora, the distribution of plant species, and the traditional uses of these plants for sustainable resource utilization and conservation. Therefore in present study, we aimed to create a comprehensive floristic inventory of IIT Guwahati, with the following specific goals: (i) a thorough examination of the floral diversity in the IIT Guwahati, and (ii) the implementation of spatial analysis using Geographic Information Systems (GIS) to identify the distribution and types of plant species within the study area.

3.4. Materials and methods

3.4.1. Study area

The process of documenting the floral diversity of north Guwahati was limited to flowering plants only. The key locations where the studies were conducted include hilltops (approach towards viewpoint), hill base and surroundings, residential areas, lakes, roads and surroundings. The study area was divided into four zones namely Z1, Z2, Z3 and Z4 **(Fig. 3.1).** The Z1 includes south western part of the IITG, Z2 includes north western areas including waterbodies, Z3 is in north eastern boundary, major portion of which are swamps, and the last zone Z4 includes residential areas and Ghoramara hill in south eastern region adjacent to Brahmaputra.

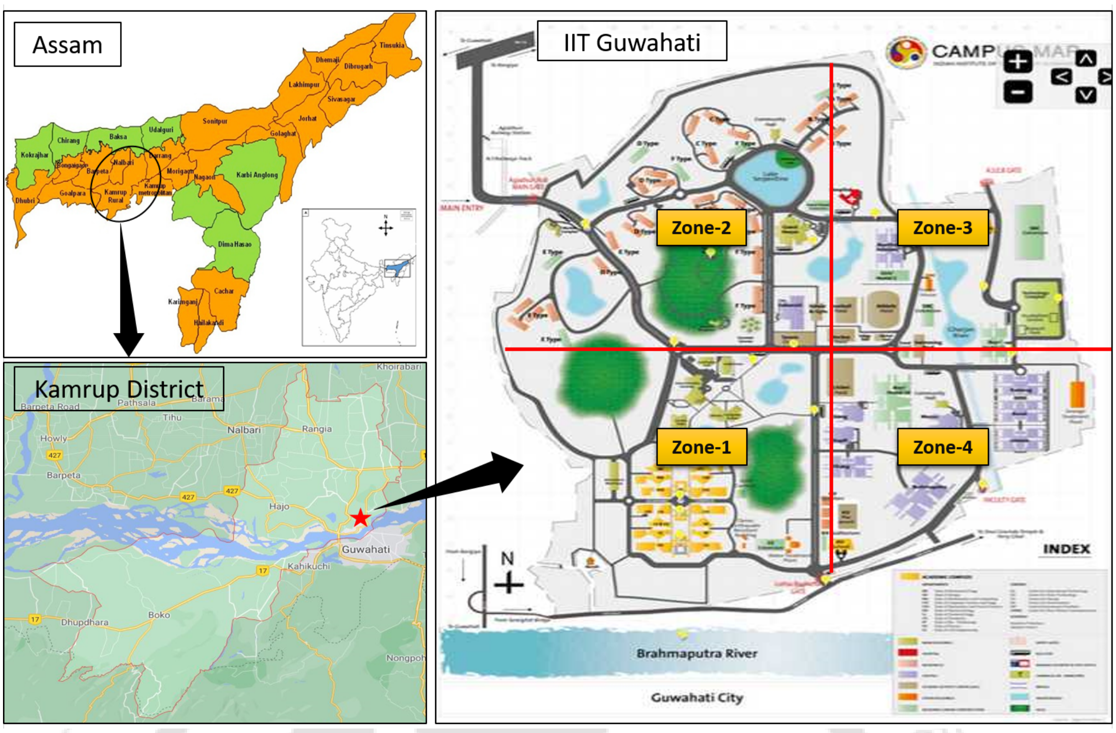


Figure 3.1: Study area and its division into four of zones (Z-1, Z-2, Z-3, and Z-4)

3.4.2. Sample collection

3.4.2.1. *Sample selection*: This study cantered on the analysis of flowering plants at IIT Guwahati, encompassing a wide range of species, including herbs, shrubs, trees, aquatic plants, and climbers. Special attention was given to seasonal plants, particularly herbaceous species found in shallow swampy areas, due to their transient nature. Additionally, we closely observed and captured photographs of plants that maintained their presence throughout the year at regular intervals.

3.4.2.2. *Sample collection:* Numerous plant species, such as trees and large shrubs, exhibit year-round availability, while for perishable and seasonal plants, we employed a strategy of collecting field samples and storing them for subsequent analysis. Our sample site selection was meticulous, focusing on representative locations within the study area, and the plant specimens chosen were in good health. This approach ensured a systematic and impartial method for sample collection, enhancing the reliability of our data for analysis.

3.4.3. Data collection

3.4.3.1. *Sampling methodology:* Since the terrain in IIT Guwahati is highly diverse including hills, lakes, swamps, estate buildings and small forest areas it was quite challenging to follow standard sampling methodologies. So, we have developed a modified method, in which the study area was divided into four zones and then followed the line transact method (Williams et al., 2019) to analyse the flora of this region.

3.4.3.2. *Data recording:* Systematic surveys to record sightings and occurrence of various flora species were conducted around the year. Efforts were made not to miss the flowering stage of various plants, which is the key for identification and differentiation between closely related species, or else highly misleading. During this process, field notebooks were used to record detailed information about each plant species, including the collection date, local and scientific names, family, locality, altitude, habit, habitat, local uses, vegetation, and collector's name. Additionally, special features like the presence of latex, flower color, size, scent, stem characteristics, hairs, plant height, thorns, species associations, and root and rhizome traits were also documented for a comprehensive understanding of each specimen. Clear and high definition photographs of flowers, fruits, and whole plants were also captured for getting maximum information. Thus, the primary data was based on existing information related to the flora observed through direct sightings and secondary sources (Naidu and Kumar, 2016).

3.4.3.3. *Plant species identification*: Taxonomic keys, photographs of identifying features, and literatures were employed for the identification of plants. Identified plants were also cross-checked by taxonomic experts and then inventorized.

3.4.3.4. *Data variables:* The selection of specific variables in biodiversity analysis serves to provide a comprehensive understanding of the ecosystem and its plant diversity. In this study the specific variables like species richness, evenness, abundance, and distribution were investigated. Species richness is a fundamental metric which measures the number of different plant species in an area and reflects the diversity of species (Schuldt et al., 2019). Assessing the evenness of species abundance helps in understanding whether some species dominate while others are rare, providing insights into the distribution of resources and the ecosystem's stability (Walter et al., 2021). Quantifying the abundance of each species allows us to identify dominant and rare species, which can shed light on the ecosystem's structure and the potential impacts of invasive or keystone species (Roswell et al., 2021). Studying the spatial distribution of plant species provides information about the range and habitat preferences of different species, which can be vital for conservation and management (Srivastav et al., 2019).

3.4.4. Data analysis

3.4.4.1. *Biodiversity indices:* An important diversity index is Simpson's Diversity Index (D). The Simpson’s index includes the number of species present and the relative abundance of each species. It is directly proportional to species richness and evenness. The D value falls within 0 to 1, where 1 means infinite diversity and 0 presents no diversity. The D value can be calculated using the formula below (Simpson, 1949).

$$D = 1 - \frac{\sum n(n-1)}{\mathrm{N(n-1)}}$$

Where,

n = Number of individuals in each family

N = Total no of individuals in all families

Shannon-Weiner Index (H′) is a quantitative measure of types of species present in a specific area. It is calculated to find the species diversity using the following formula (Shannon and Weaver, 1963).

$$H' = -\sum \left({}^{ni}/_{n}\right) \ln \left({}^{ni}/_{n}\right)$$

Where,

n_i/n = Importance probability of each species in a population

n_i = Importance value of species

n = Total number of individuals of all species.

Adjacent ecosystems are expected to be more similar in composition than others. In this study the similarity between all the four zones were evaluated by calculating Jaccard's similarity Index (J) for every pair of zones using the given formula.

$$J = |X \cap Y| / |X \cup Y| * 100$$

Where,

J = Jaccard's similarity index

X= Number of species found in both zones

Y= Total number of species present in both zones

3.4.4.2. *Community indices:* The structural diversity was analyzed by comparing the distribution of families. Community indices like species diversity (H), evenness (E), and species richness (d) were calculated. Species diversity is one of the critical components of biodiversity and is defined as the variety of species present in a specific area. Different species in an ecosystem are dependent among themselves. Higher species diversity in an ecosystem maintains its balanced and productivity. There are two essential attributes of species diversity, namely species richness and species evenness. Species richness is directly related to the number of species present whereas, species evenness is linked with the number of individuals present in a species.

Equitability of evenness refers to the degree of relative dominance and mathematically quantifies how much equal a specific taxa is. It was calculated according to Pielou's formula (Pielou, 1966).

$$\text{Evenness (E)} = {}^{H'}/_{\text{Log S}}$$

Where,

H′ = The Shannon index

S = Number of species.

The species richness is the number of various species found in an ecosystem. Species richness is a numerical attribute that is independent of species abundance. The species richness was determined by Margalef index using the formula given below (Margalef, 1968).

$$\text{Species richness (d)} = {}^{(S-1)}/_{\ln N}$$

Where,

S = The number of species

N = The total number of individuals in the sample.

A density map of distribution of plants in different zones was plotted using Arc GIS (V10.3) software. To plot this map the geographical location of individual species in terms of latitude and longitude value were collected. The co-ordinates were interpolated to provide the information on distribution pattern of herbs, shrubs, trees, climbers, and aquatic plants. Colour codes were used to highlight the density of different plant types (Jaccard, 1912)**.**

3.3.5. Ethical considerations

3.3.5.1. *Permits and permissions*: All the permissions for data collection from private and restricted areas were obtained from relevant authorities wherever necessary. For public places and wild areas no permission was required but caution was observed.

3.3.5.2. *Ethical guidelines*: Adherence to ethical guidelines during plant biodiversity analysis was paramount, encompassing responsible plant collection practices, transparent data sharing, and a keen awareness of potential impacts on local ecosystems. Proper permits, respect towards protected species and habitats, and minimizing disturbances during plant collection was taken into consideration. By sharing data openly and responsibly, we promote collaboration and support broader conservation efforts while safeguarding the delicate balance of local ecosystems.

3.3.6. Data quality assurance

To uphold the quality of data rigorous taxonomic verification to confirm plant species identification, strict adherence to standardized data collection protocols, and the implementation of validation procedures to reduce errors were employed. Good quality data ensures validity of the findings and make valuable contributions towards the understanding of plant biodiversity and ecological conservation

3.4. Results and discussion

3.3.1. IIT Guwahati topography and vegetation

North Guwahati is located in the northern part of Guwahati city in the district of Kamrup Rural, Assam, India. This region is located in the bank of the mighty Brahmaputra and lies at 26° 12' 0" to 26° 12' 44" N and 91° 43' 23" to 91° 46' 42" E at an average elevation of 54 m and spread across more than 2,500 acres of land. The climate is tropical with average temperature of 24 °C, IMD recorded 31 °C maximum and 12 °C minimum temperature in this location. There is heavy rainfall along with thunderstorms during the typical monsoon season. Soil type is brown and reddish sandy loam rich in both macro and micronutrients. The topography of the region is undulating with variation in elevation from 49.5 m to 100 m above MSL. The study area include IIT Guwahati spread across more than 700 acres of land which is part of Kamrup Rural district. The IIT campus is a relatively protected area having negligible contact with the outer world, making it a hub of rich floral diversity. There are two hills near the IIT campus; one is Kala Pahar (380 m above MSL), situated on the northwestern side, and Ghoramara Pahar (100 m above MSL) in the southern part of the campus.

The complex environmental features, including climates, geology, soil, forest biota, and phyto-sociological history, govern an area's vegetation. The diverse physiography of the region helps to harbor rich vegetation. The land is scattered with many hills and lakes within the premises. It is an ideal example for the coexistence of various types of mini ecosystems and these include open grasslands, wetlands, forest patches, aquatic ecosystems and patches of scrub jungles. The vegetation of the region can broadly be classified into: 1. **Degraded alluvial grasslands:** this is a characteristic type of vegetation of the wetlands, which remains nearly underwater in most places during the summer season. 2. **Aquatic vegetation**: quite a large area of the region has water bodies, which harbor a good number of aquatic plants. 3. **Highland vegetation:** this vegetation observed on the hillock slopes and is comparatively undisturbed. The hilly areas mostly covered with forests of various formations ranging from moist deciduous forests to bamboo brakes. There is abundant moisture, and the soil type is rich in minerals and nutrients. The leaves dropped from trees provide a steady source of organic material for the soil. 4. **Secondary scrub forests:** low-lying, dense underbrush, an area of little precipitation, plenty of continuous winds, poor drainage and medium to poor soil quality. 5. **Wasteland vegetation**: this type of vegetation is continued in the high wasteland and frequently disturbed by anthropogenic pressure. As a result, the species richness is excessive when compared to other vegetations.

Ashoka, Arjun, Eucalyptus, Areca palm, Foxtail palm, Bakul, Nahor, Krishnasura, Radhasura and several varieties of bamboo amid nature are found in this area. There are several avenues on the IIT Guwahati campus named after trees, for example, Suryamukhi Road (sunflower) and Radhasura Road (pink cassia). Other avenue trees in some of the important lanes of the campus are *Peltophorum pterocarpum* (copper pod tree), *Samanea saman* (rain tree), *Jacaranda mimosifolia*, *Butea monosperma,* and *Delonix regia* (gul mohar). In shallow areas the soil is poor in nutrients, resulting in stunted growth of bamboo and scrubs. Till date, there is no exclusive data on the floral diversity of the IIT Guwahati campus and its adjacent regions, whichever data available are insufficient and incomplete.

3.3.2. Floristic analysis

3.3.2.1. *Floral diversity*

A systematic exploration of the north Guwahati with a rich assemblage of plant species has resulted in the documentation of 434 species of Angiospermic taxa from 101 families **(Table 3.1.)**. As far as floristic analysis in the current study is concerned, 82 species of trees belonging to 30 families, 98 species of shrubs belonging to 32 families and 254 species of herbs including climbers and aquatics belonging to 74 families were documented (**Annexure I**). The documentation revealed 172 herbaceous species (128 dicots and 44 monocots); 98 shrubs (93 dicots and 5 monocots); 82 trees (77 dicots and 5 monocots); 26 climbers (22 dicots and 4 monocots) and 56 species of aquatic plants (21 dicots & 35 monocots). Grasses on the campus account for 30 species. Most of the species were recorded from plains, followed by hillslopes, roadsides, marshy lands, and wastelands. Nearly 206 species have been recorded from plains; 80 from hillslopes; 56 from roadsides; 63 species recorded from the marshy area; 26 species recorded from the wasteland. About 348 species are found to propagate through seeds, 19 by root, 48 by rhizome and/or tuber, along with 16 species having the capability to regenerate either from stem, leaf, stolon or suckers. A number of anthropogenic activities like construction activities, deforestation, and overuse of resources are causing disturbances for vegetation in IITG but animal grazing has not been reported (Kar and Borthakur, 2007). Some of the representing species from different plant groups are given in **Fig. 3.2.**

In line with previous floristic investigations across NE India and Assam, congruent patterns of diversity were observed. For instance, the Dibrugarh district of Assam documented 462 angiosperm species distributed across 106 families (Gogoi and Nath, 2021). Similarly, an inventory of flowering plants in the East Siang district of Arunachal Pradesh revealed 508 taxa encompassing 102 families (Taram et al., 2020). In another study Choudhury and Kalita (2015) recorded 313 flowering plants from 94 familes in Greater Sualkuchi area of Kamrup district, Assam. Furthermore, a comprehensive analysis of floristic diversity and vegetation in the Upper Siang district of NE India recorded an impressive 1003 taxa representing 110 families and 529 genera (Choudhary et al., 2012). In a comparable study conducted near IIT Guwahati, specifically in Gauhati University, characterized by similar climatic conditions, 337 species of flowering plants within 83 families were reported. This study also delineated the composition of herbs, shrubs, trees, and climbers, noting 213 herbaceous species, 57 shrubs, 42 trees, and 25 climbers (Kar and Borthakur, 2007). It's noteworthy that while several studies have explored various regions of Assam, there remains a gap in reports from North Guwahati, emphasizing the need for comprehensive assessments of biodiversity in this particular area.

Table 3.1: Assessing plant biodiversity of IITG: Family, genus and species counts

Taxa	Dicotyledons	Monocotyledons	Total
Family	81	21	102
Genus	263	71	334
Species	338	93	431

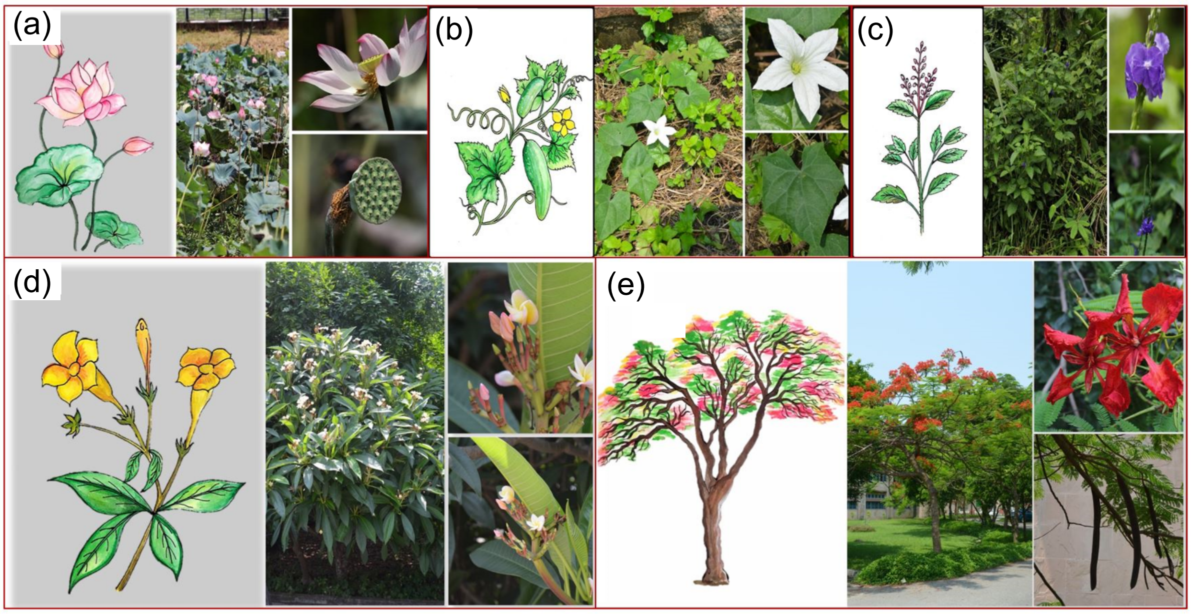


Figure 3.2: Representative species of (a) aquatics, (b) climbers, (c) herbs, (d) shrubs and (e) trees

3.3.2.2. *Identification and documentation*

The modified sampling methodology implemented at IIT Guwahati, considering its diverse terrain, has yielded insightful results in the documentation and identification of plant species within the campus. The division of the study area into four zones, coupled with the line transact method, facilitated a systematic analysis of the region's flora, overcoming the challenges posed by the varied landscape. Similar to studies by Oktaviani et al. (2019), systematic surveys conducted throughout the year ensured a comprehensive record of sightings and occurrences of diverse plant species. A particular emphasis on capturing plants during their flowering stage proved crucial for accurate identification and differentiation among closely related species, as reported before by Mir et al. (2014) in Meghalaya and Gogoi and Nath (2021) in Dibrugarh district of Assam. The utilization of field notebooks allowed for the meticulous recording of detailed information for each plant species, capturing key characteristics such as local and scientific names, family, locality, altitude, habit, habitat, local uses, and vegetation. The inclusion of special features, including the presence of latex, flower color, size, scent, stem characteristics, hairs, plant height, thorns, species associations, and root and rhizome traits, provided a nuanced understanding of each specimen similar to other diversity studies (Dutta and Devi, 2013; Hawthorne and Lawrence, 2013). The incorporation of clear and high-definition photographs further facilitated correct identification, ensuring a robust documentation of the flora.

Comparisons with similar studies by Gurung et al. (2019) in Kanchanjunga, Elfrida et al. (2020) in Barak Valley, Assam underscored the significance of the modified methodology employed at IIT Guwahati. The approach aligns with established practices in biodiversity research, emphasizing the importance of thorough documentation and systematic analysis for accurate plant species identification as reported by Hawthorne and Lawrence (2013). The collaboration with taxonomic experts for cross-verification enhances the reliability of the results, a practice consistent with studies prioritizing the validation of taxonomic assignments (Naidu and Kumar, 2016). The inclusion of special features and comprehensive documentation aligns with the methodologies employed in comparable studies, ensuring a holistic understanding of plant diversity (Panda et al., 2013). The utilization of taxonomic keys, photographs, and literature in the identification process is in line with scientific standards, contributing to the robustness of the findings (Kirchoff et al., 2014). Overall, the results obtained through this method at IIT Guwahati contribute valuable insights into the campus's plant diversity, aligning with and building upon the methodologies and practices of previous studies in the field.

The IIT Guwahati campus encompasses an elevation range of 49 to 100 meters above MSL. Notably, Zones 2 and 3 exhibit a higher average elevation compared to the others. These zones are characterized by the prevalence of wild plants such as *Albizia saman*, *Cascabela thevetia*, *Cassia* spp., *Pongamia pinnata*, *Ricinus communis*, and *Spilanthes* spp. which is in accordance with diversity studies along altitude gradients in Himalayas (Sharma et al., 2019; Sekar et al., 2023), Zone 3 stands out for the presence of aquatic bodies and a rich diversity of aquatic species, including *Ipomoea* spp. and *Nymphoides indica* similar to diversity in western Ghats (Bhat et al., 2020). Conversely, in Zone 4, the landscape is influenced by the presence of numerous buildings, leading to a prevalence of herbaceous plants and climbers like *Clerodendrum thomsoniae*, *Smilax* spp., and various grasses as observed in seven temperate forests (Dar and Sundarapandian, 2016).

The plant diversity across four zones of the IIT Guwahati campus was assessed, revealing distinct patterns in each zone. In Zone 1, herbs dominated with 102 species, followed by shrubs (60 species) and trees (42 species). Zone 2 exhibited a similar trend, with herbs leading at 74 species, followed by trees (45 species) and shrubs, while also noting the presence of five aquatic plants. Zone 3 mirrored these findings, reporting 96 herb species, 24 tree species, and 20 shrub species. In Zone 4, a total of 154 herbaceous species, 33 tree species, and 28 shrub species were identified, along with five aquatic plants. This comprehensive analysis provides valuable insights into the diverse plant life thriving within the IIT Guwahati campus which was also confirmed from GPS data. The GPS data played a pivotal role in spatial mapping, enabling site-specific conservation efforts, and facilitating long-term monitoring as reported by Saadi et al. (2020) on medicinal plants of West Bengal, India. The integration of GPS data in the plant diversity analysis of the IIT Guwahati campus provides a holistic understanding of distribution patterns. This, in turn, proves instrumental in devising effective conservation and management strategies tailored to the diverse plant species across different elevations.

3.3.2.3. *Exotics and indigenous species*

Exotic or introduced plants are found in significant numbers inside the IIT Guwahati campus, 178 species of exotic plants were identified during this study. Several of these species are trees and introduced for their attractive foliage or for their ornamental value. Species such as *Cassia javanica* (pink cassia), *D. regia* (flame of the forest), *Jacaranda mimosifolia* (blue Jacaranda), *Bahunia accuminata* (white orchid tree) and shrubs like *Allamanda cathartica* (golden trumphet), *Calliandra surinamensis* (surinam powder puff) have been planted for their brightly coloured flowers while *Ficus religiosa* and *Callistemon viminalis* (bottle brush) have been planted for their foliage. Along with the exotics, 253 native plants were also recorded like *Cassia fistula* (Indian laburnum), *Magnolia champaca* (champak) and *Saraca asoca* (ashoka) **(Table 3.2).**

In a comprehensive analysis by Saikia et al. (2012), findings parallel to our study were reported in Upper Assam. They noted a higher proportion of native plant species in categories such as fuelwood, timber, fruits, and medicinal plants, contrasting with a greater prevalence of non-native plants in the ornamental, weeds, grasses, and vegetable categories. Interestingly, the order Lamiales emerged as the most extensive, harboring 35 native plant species. In this study, the majority of species, totaling 60.8%, were identified as native, while the remainder were categorized as invasive (Saikia et al., 2012). In a study conducted by Das and Duarah (2013), an investigation into exotic plants along the roadsides of Jorhat, Assam, brought to light a total of 18 alien plant species, originating from 10 diverse families. These included species such as *Alternanthera* sp., *Cassia* sp., and *Acmella* sp., echoing the findings of our current study. Barua et al. (2013) underscored the significance of invasive species, exemplified by *Ageratum* sp., *Chromolaena odorata*, *Lantana camara*, *Parthenium* sp., and others, on the indigenous flora of Assam. This resonates with our own observations regarding the presence of invasive plant species in the vicinity of IIT Guwahati. Furthermore, similar results were reported by Sarma et al. (2019) in their examination of Amchang wildlife sanctuary in Kamrup district of Assam, where 57 invasive species belonging to 22 different families were identified.

Notably, among the native plants identified, several hold ethnomedicinal significance. as Species like *Dillenia indica*, *Aegle mermelos*, *Allium cepa*, *Azadirachta indica*, *Bacopa monnieri*, *Clerodendrum* sp., and *Curcuma* sp. were found to be of particular importance, aligning with their documented ethnomedicinal value (Kalita et al., 2014; Bhattacharyya and Bhattacharyya, 2016; Gogoi et al., 2020). These species, which have been observed in the context of IIT Guwahati, contribute to the broader understanding of the region's biodiversity and its potential applications in ethnomedicine.

Table 3.2: Frequently observed exotic plants in IITG

S No	Exotic plants	Scientific name	Nativity (Origin)
1	Arrowhead vine	*Syngonium podophyllum*	Latin America
2	Balloon plant	*Cardiospermum halicacabum*	South central America
3	Bleeding heart glorybower	*Clerodendrum thomsoniae*	Tropical Africa
4	Blue jacaranda	*Jacaranda mimosifolia*	South America
5	Bottle brush	*Callistemon viminalis*	Australia
6	Creeping spotflower	*Acmella repens*	North America
7	Crown of thorns	*Euohorbia milii*	Madagascar
8	Dwarf Poinciana	*Caesalpinia pulcherrima*	Tropical America
9	False nut sedge	*Cyperus strigosus*	United States
10	Field dodder	*Cuscuta campestris*	North America
11	Flame of the forest	*Delonix regia*	Madagascar
12	Fringed spider flower	*Cleome rutidosperma*	Africa
13	Giant cabuya	*Furcraea foetida*	Caribbean
14	Giant crego aster	*Callistephus chinensis*	China
15	Graveyard plant	*Catharanthus roseus*	Madagascar
16	Indian shot	*Canna indica*	America
17	Litchi	*Litchi chinensis*	Southern China
18	Lobster claw	*Heliconia rostrata*	Peru
19	Miracle leaf	*Bryophyllum pinnatum*	Madagascar
20	Physicnut	*Jatropha curcus*	Central America
21	Pink cassia	*Cassia javanica*	Thailand
22	Saman tree	*Albizia saman*	South America
23	Santa maria	*Parthenium hysterophorus*	Central America
24	Smooth chaff flower	*Alternanthera triandra*	Central south America
25	Water snow flake	*Nymphoides indica*	Puerto rico

3.3.2.4. *Common and rare species of the campus*

Some of the common trees found in IITG were *A. saman* (rain tree), *Cascabella thevetia* (beestill tree), *Cassia fistula* (Indian laburnum), *Terminalia arjuna* (arjun), *P. pinnata* (pongam oil tree) and *Dalbergia sisoo* (Indian rosewood). On the other hand, tree species like *Dillenia indica* (elephant apple), *Furcraea foetida* (giant cabuya), and *Pachypodium lamerei* (Madagascar palm) have only a single specimen on the campus. Other rare species include *Acacia auriculiformis* (black wattle), *Anacardium occidentale* (cashew), *Magnolia champaca* (champak), *Phyllanthus embelica* (amla), *Stericulia villosa* (elephant rope tree) and *Litchi chinensis* (litchi).

In the diverse ecosystems of Assam, both common and rare plant species play pivotal roles in sustaining local biodiversity and ecosystem functions. Common plants, often abundant and easily encountered, underpin ecosystem stability by serving as primary food sources, contributing to soil fertility, and providing resources with traditional, medicinal, and industrial value. One such common species, *Azadirachta indica* (Neem), has garnered substantial scientific attention for its versatile applications and plentily available in IITG. With a well-documented history of traditional medicinal use, Neem has been extensively investigated for its antiviral, antibacterial, antifungal, and anti-inflammatory properties, underscoring its significance in both traditional and modern medicine (Quraishi et al., 2018). Our findings align with earlier studies in Assam, as we also observed the presence of common species such as *Areca catechu* (areca nut), *Mangifera indica* (mango), *Camellia sinensis* (tea), and *Artocarpus heterophyllus* (jackfruit), confirming their prevalence in the region (Saikia et al., 2012).

Table 3.3: Top ten common and rare plants

S No	Common species	S No	Rare species
1	*Acmella repens*	1	*Abroma augustum*
2	*Ageratum conyzoides*	2	*Acacia auriculiformis*
3	*Albizia saman*	3	*Dillenia indica*
4	*Alpinia nigra*	4	*Furcraea foetida*
5	*Cascabella thevetia*	5	*Heliconia rostrate*
6	*Cassia fistula*	6	*Jacaranda mimosifolia*
7	*Cassia javanica*	7	*Macaranga denticulata*
8	*Delonix regia*	8	*Pachypodium lamerei*
9	*Eichhornia crassipes*	9	*Phoenix dactylifera*
10	*Lantana camara*	10	*Stericulia villosa*

In contrast, some rare plant species face unique challenges, especially those of forest origin, like *Aquilaria malaccensis*. This endangered species, dominant in Assam's Nalbari district, significantly contributes to the local economy (Saikia et al., 2012). However, rare species may confront genetic diversity constraints in managed ecosystems due to factors such as limited gene flow, inbreeding, and selective pressures (Chung et al., 2021). Notably, we also documented the presence of a few rare species, including *Acorus calamus*, *Aquilaria malaccensis*, *Costus speciosus*, and *Clerodendrum* sp. These species are distinct in that they are generally not cultivated or introduced into the local ecosystem. As the information in **Table 3.3** illustrates, our study, like previous research in Assam, highlights the coexistence of both common and rare plants, each contributing to the region's unique biodiversity and ecological dynamics (Mazumdar, 2014).

3.3.2.5. *Species utilitarian value*

The IIT Guwahati campus is not only a hub of academic excellence but also a fascinating repository of plant species, each contributing to the ecosystem and serving human needs. These plant species, found in abundance, hold utilitarian value, contributing to local biodiversity, economy, and human well-being.

These encompass fruit-yielding plants, ornamental species, and others with diverse utilities. Some of the species with high utility value (list not exclusive) are fruit-yielding plants: These are of paramount importance, not only for their delicious produce but also for their role in supporting biodiversity. Prominent examples within the campus include *Mangifera indica* (mango), *Artocarpus heterophyllus* (jackfruit), *Psidium guajava* (guava), *Syzygium cumini* (jamun), *Citrus* spp. (orange), and

Litchi chinensis (litchi). As per Saikia et al. (2012) about 18% of total flora in Assam are of high value medicinal plants, which is in accordance with current investigation. These species serve as primary food sources for various wildlife, helping to maintain ecological balance and are commonly available throughout NE India (Hazarika, 2012; Barua et al., 2019; Barua et al., 2021).

Cassia fistula (Indian laburnum), *C. javanica* (pink cassia), *Polyalthia longifolia* (false ashoka), *M. champaca* (champak), *Lagerstroemia indica* (crepe myrtle) and *Plumeria rubra* (temple tree) are some of the ornamental plants observed in IITG and similar results were also observed in Upper Assam except ornamental bananas (Joe and Sabu, 2016). Rani et al. (2012) reported the wide distribution of *Nyctanthes arbor-tristis* (night jasmine) which was also available in studied region. A good number of ornamental herbaceous plants were observed in the campus, which can be categorized as climbers (*C. thomsoniae, Thunbergia grandiflora, Ipomoea* sp., *Clitoria ternatea, Syngonium podophyllum*), hedges (*Callistephus chinensis, Canna indica, Catharanthus roseus, Tridax procumbens, Euphorbia milii*), and lawns (*Tradescantia fluminensis, Spilanthes acmella, Cuphea hyssopifolia*), these plants were well documented and readily available all over the NE region (Jaiswal, 2010; Mao and Roy, 2016). Ornamental plants not only enhance the visual appeal of the campus but also provide essential ecosystem services. Beyond aesthetics, ornamental plants support pollinators and beneficial insects, contributing to ecosystem health.

Saraca asoka and *Neolamarkia cadamba* are sacred and medicinal plants reported to be available all over India and found within the campus of IITG as well (Pandey and Negi, 2016; Fathima et al., 2023). From Kamrup district, Assam 50 medicinal plants from 33 families were reported, similar to the current study, which are being used by local communities long ago (Deka and Kalita, 2013). *Azadirachta indica* is an important medicinal plant with cosmopolitan distribution and is of enormous importance from medicinal point of view and has been well characterized in Assam (Rai and Lalramnghinglova, 2011; Saikia et al., 2012). Medicinal plants are invaluable resources for traditional and modern medicine. They offer remedies for a range of ailments, from common stomachaches to kidney disorders (Bordoloi et al., 2023). The presence of these species within the campus highlights their ecological importance and the need for their conservation (**Table 3.4**).

Other plants from IITG with high utilitarian values include *Bombax ceiba* (silk cotton tree) is an economically important crop for textile industries previously characterized from Barak Valley in Assam (Das and Kalita, 2016). The domestic use of some plants includes the use of grasses such as *Brachiaria mutica*, *Pennisetum pedicellatum* and all lawn grasses as fodder and *Phoenix* sp. and *Chrysopogon* sp. for making brooms by the locals and also used for soil management (Dogan et al., 2008). Some woody shrubs or trees like *Lantana camara, Acacia auriculiformis, Solanum torvum,* and *Albizia lebbeck* are used as fuelwood. Species such as *Eupatorium odoratum* (Mexico), *Parthenium* (South America), *Ageratum conyzoides* (Tropical America), and *Lantana camara* (South America) are termed as weeds due to their uncontrolled proliferation and some of them are common allergic plants to human. Their pollen is supposed to be allergenic, causing bronchial disorders. The botanical diversity within the IIT Guwahati campus showcases the intricate relationships between plant species, the environment, and human communities. Recognizing their utilitarian significance is essential for their conservation, sustainable management, and continued contributions to the local ecosystem and human well-being.

Table 3.4: Medicinal plants found in IIT Guwahati

S No	Common species	Scientific name	Medicinal uses
1	Amla	*Phyllanthus emblica*	Improves immunity, hair care, Blood purifier *etc.*
2	Arjun	*Terminalia arjuna*	Traditionally used for heart disease
3	Balloon vine	*Cardiospermum helicacabum*	Rheumatism, itchy skin, diaphoretic, diuretic *etc.*
4	Bamboo-leaved galangal	*Alpinia nigra*	Aphrodisiac, diuretic analgesic *etc.*
5	Bengal dayflower	*Commelina benghalensis*	Stem and leaves used to stop bleeding
6	Catechu	*Acacia catechu*	Extract used for treating osteoarthritis
7	China rose	*Hibiscus rosa-sinensis*	Checks bleeding, soothes irritated tissues, helpful in painful menstruation
8	China smilax	*Smilax anceps*	Anti-inflammatory and diuretic
9	False daisy	*Eclipta prostrata*	Helpful in range of skin disorder, liver problems and anti-venomous
10	Giant calotrop	*Calotropis gigantea*	Useful in asthma, liver and spleen disease
11	Goat weed	*Ageratum conyzoides*	Used against epilepsy and wounds
12	Graveyard plant	*Catharanthus roseus*	Leaf extract has anti-cancer and anti-diabetic properties
13	Holy basil	*Ocimum sanctum*	All parts of the plant used to treat bronchitis, bronchial asthma, malaria, dysentery *etc.*
14	Insulin plant	*Costus igneus*	Leaves help to produce insulin in body
15	Manimuni	*Centella asiatica*	To cure dysentery
16	Nagakesara	*Mesua ferrea*	Relives joint pain, stomach problems
17	Siris tree	*Albizia lebbeck*	An astringent, bark used to treat inflammation
18	Suho nibon	*Spilanthes acmella*	Leaves used for mouth ulcer
19	Thumbai	*Leucas aspera*	Antifungal, antimicrobial and cytotoxic
20	Water hyacinth	*Eichhornia crassipes*	Rhizome eaten in empty stomach to cure asthma
21	Water snowflake	*Nymphoides indica*	Plant juice used in jaundice

3.3.2.6. *Economic plants*

Economic plants are plant species that provide significant value and benefits to human societies due to their various practical applications. These plants are cultivated or harvested for their direct contributions to the economy, which can include their use as food, medicine, fibers, building materials, fuel, and raw materials for various industrial processes. The value of economic plants is derived from their ability to support livelihoods, enhance food security, contribute to healthcare, and stimulate economic development through trade and industry. Their importance extends to both local and global scales, making them integral to agricultural, pharmaceutical, and industrial sectors. Plants like *A. capilaris, B. vulgaris*, *C. olitorius*, and *C. sativa*, valued for their strong fibers used in textiles and ropes. Significant beverage plants such as *C. sinensis*, which produces Assam tea, and *C. arabica*, known for its coffee beans were also observed. Other beverage plants included *C. nucifera* and *P. dactylifera*. A total of 17 species of herbs and spices were identified for their culinary heritage including herbs and spices like *C. tamala* (Indian bay leaf) and *O. sanctum* (holy basil), which are widely used in Assamese cuisine for their distinct flavors and medicinal properties. Additionally, culinary spices such as *C. chinense* (bhut jolokia), one of the hottest chilies, and *C. longa* (turmeric), known for its culinary versatility and health benefits (**Table 3. 5**).

Table 3.5: Economically important plants

S No	**Fibre plants**		
1	*Cannabis sativa*	13	*Zingiber zerumbet*
2	*Corchorus olitorius*	14	*Elettaria cardamomum*
3	*Agrostis capillaris*	15	*Piper nigrum*
4	*Bambusa vulgaris*	16	*Syzygium aromaticum*
5	*Coix lachryma-jobi*	17	*Myristica fragrans*
6	*Phragmites karka*		
			Toxic plants
	Beverage plants	1	*Ricinus communis*
1	*Camellia sinensis*	2	*Nerium oleander*
2	*Coffea arabica*	3	*Calotropis gigantea*
3	*Coffea benghalensis*	4	*Datura metel*
4	*Mangifera indica*	5	*Datura stramonium*
5	*Cocos nucifera*	6	*Atropa belladonna*
6	*Phoenix dactylifera*	7	*Solanum nigrum*
		8	*Solanum khasianum*
	Spices	9	*Solanum rostratum*
1	*Capsicum annuum*	10	*Solanum torvum*
2	*Capsicum chinense*	11	*Solanum xanthocarpum*
3	*Capsicum frutescens*	12	*Cascabela thevetia*
4	*Curcuma longa*	13	*Holarrhena antidysenterica*
5	*Curcuma amada*	14	*Catharanthus roseus*
6	*Curcuma aromatica*	15	*Plumeria rubra*
7	*Curcuma caesia*	16	*Rauvolfia canescens*
8	*Curcuma zedoaria*	17	*Tabernaemontana divaricata*
9	*Ocimum basilicum*	18	*Wrightia tomentosa*
10	*Ocimum sanctum*	19	*Heliotropium indicum*
11	*Cinnamomum tamala*	20	*Jatropha curcas*
12	*Zingiber officinale*	21	*Jatropha gossypifolia*

Among the toxic plants, *R. communis* produces ricin, a highly toxic protein that can be fatal in small doses but has potential applications in cancer treatment when used to target and kill malignant cells. *N. oleander* another toxic plants contains cardiac glycosides, which can be lethal if ingested but are also studied for their potential in heart disease treatment. *Calotropis* spp. produces calotropin, which can cause severe gastrointestinal distress but has been researched for its potential in pharmaceutical formulations. *Datura* spp. contain tropane alkaloids such as scopolamine and atropine, which can cause delirium and hallucinations; however, atropine is used medicinally to treat certain types of poisoning. *Solanum* spp. produce solanine, which can cause gastrointestinal and neurological disorders but has pesticidal properties. While these plants' toxins pose significant risks to human and animal health, they can also play a role in controlling pest populations and contributing to medical research. In the local ecosystem, these toxins can deter herbivores and invasive species, helping to maintain ecological balance. However, they can also lead to accidental poisoning of non-target species, including livestock and humans, necessitating careful management and education on their potential dangers and benefits.

3.4.3. Familial analysis of the flora

Asteraceae emerged as the most speciose family with 38 species. Other speciose families include Fabaceae (33 species), Poaceae (30 species), Solanaceae (22 species), Euphorbiaceae (22 species), Malvaceae (17 species) Cyperaceae (16 species), Apocynacea and Zingiberaceae (13 species each), Lamiaceae (10 species), Araceae (7 species), Rubiaceae, Moraceae and Verbenaceae (each represented by 8 species). A similar report on plant diversity was also reported in Gauhati University where the Asteraceae was the dominant family with 17 species followed by Fabaceae (14 species) (Kar and Borthakur, 2007). Families like Fabaceae, Bignoniaceae and Moraceae are represented by both introduced as well as indigenous species. On the other hand, families like Proteaceae, Araliaceae and Rhamnaceae are represented only by exotic species and Sapindaceae, Lythraceae, Rutaceae, Magnoliaceae are families represented only by indigenous species, a similar report supports our findings in this aspect (Choudhary, 2013).

In another report floristic composition of upper Assam revealed Euphorbiaceae as the most dominant family with 14 species, followed by Moraceae and Poaceae (12 species each). While Moraceae (12 species) was the most dominant tree family, Cucurbitaceae (9 species) was the most dominant family of herbs, and Malvaceae and Solanaceae (7 species) the most dominant families of shrubs which was different from the urban campus of IITG (Saikia et al., 2012). Gogoi and Nath (2021) documented the floristic diversity of Dibrugarh district of Assam and inventorized 462 species distributed under 334 genera, representing 106 families. Similar to current study the most specious family observed was Asteraceae with 32 species, followed by Fabaceae (23 species), and Poaceae (19 species). Habit analysis showed that herbs were represented by the highest proportion with 211 species

(46%) whereas in IITG it was 254 species (59%) suggesting more diversity among herbs (Gogoi and Nath, 2021).

Interestingly, Cannaceae, a close relative to Zingiberaceae, and Musaceae having a single genus *Canna* was observed in current study. This genus is native to tropical regions and southern United States and found to be present on IITG campus in several places as *Canna indica,* which is considered an invasive species. Another exciting family Euphorbiaceae, the fifth most prominent family of flowering plants, consists of most of the shrubs represented by 22 species in IITG and was found to be present in different parts of NE India (Roy and Saran, 2004). Orchids are believed to be evolved from Assam, Out of 1332 orchids NE India sustains about 856 species (De and Medhi, 2014), from which two species *Rhynchostylis retusa* and *Vanda roxburghii* were observed within IITG. In the same way, family Arecaceae or the palm family are commonly perennial trees covered by fronds at the top and can grow in diverse habitats from deserts to rainforests (Salehi et al., 2020). Trees belonging to this family are represented in the region by few exogenous species like *Cocous nucifera, Phoenix dactylifera* etc.

Non-woody flora (including shrubs, herbs, and grasses) are abundantly present in this area and they have representation in all habitat types. Some of the larger families of herbaceous species are Asteraceae (38 species), Poaceae (30 species), Euphorbiaceae (22 species) and Cyperaceae (16 species) with ample representation which is in accordance with previous studies (Singh, 2020). Heliconiaceae (1 sp.), Aparagaceae (1 sp.) and Tropaeolaceae (1 sp.) are families with less representation. Most species on the region are insect-pollinated. Some like *B. monosperma* (flame of the forest), *Erythrina stricta* (coral tree) and *Bombax ceiba* have bright and showy flowers, and are adapted to bird pollination (Krauss et al., 2017). Figs (*Ficus* spp.) have specific pollinators (wasps); their biology 'fig-wasp association' is quite interesting and most of the palm species are wind-pollinated (Jander and Herre, 2010). Unisexual species like *Brousonetia papyrifera* release pollen dust during flowering season, which is wind-pollinated. Some hemi-parasitic families like Loranthaceae plants have species like *Dendrophthoe falcata* and depend on birds (sunbird) for their pollination and dispersal (Guerra et al., 2014). Being food plants of many butterflies is another important ecological role of these herbaceous plants (**Table 3.6**).

Table 3.6: Some common food plants of different butterflies

S No	Plants	Butterflies
1	*Acmella* spp.	Dogface
2	*Ageratum conyzoides*	Orange sulphur
3	*Asclepias* spp.	Monarch butterflies, Spring azure, American copper, Eastern black swallowtail
4	*Ficus* spp.	Green comma
5	*Helianthus annuus*	Eastern tiger swallowtail, Vaiegated fritlaria
6	*Hibiscus* spp.	Variegated fritillary, Cloudless sulphur, Zebra longwing, Gulf fritillary
7	*Lantana camara*	Giant swallowtail, Great spangled fritillary and Zebra swallowtail

3.4.4. Diversity analysis

To understand the floristic diversity of the region, various diversity indices were calculated and analyzed **(Table 3.7)**. Simpson's diversity index (D) is one of the indices which is directly proportional to species diversity and was calculated to measure and compare the species diversity between the zones. It depends on the number of species present and the relative abundance of each species. Taking the D value into consideration, maximum diversity was found to be in Z4 with D-value 0.9774. Z2 came second in D-value (0.9769), indicating 2nd most diverse zone and the Z1 and Z3 came 3rd and 4th with D-values of 0.9767 and 0.9759, respectively. So, the least varied region in the campus was found to be Z3 according to Simpson's index and highest diversity observed in Z4 (Simpson, 1949). This difference could be due to presence of inaccessible swamps and abundant number of buildings in Z3, whereas Z4 doesn't have any inaccessible ecosystem making documentation easier **(Fig. 3.3a).**

The D-values for the four zones demonstrate a high level of plant diversity across all areas, with Z4 showing the highest diversity and Z3 the least. These indices, which are all near 1, reflect a balanced distribution of species, indicating that no single species predominates in any zone. High plant diversity contributes to ecosystem stability, resilience, and productivity, as it enables the ecosystem to better endure environmental stresses and support a broader range of ecosystem functions. These diversity measurements are crucial for conservation, helping to identify biodiversity hotspots for preservation and guiding restoration projects to maintain species richness and evenness, which are essential for ecological health and sustainability.

Table 3.7: Diversity analysis of four zones

Diversity indices	**Zone-1**	**Zone-2**	**Zone-3**	**Zone-4**
Simpson index (D)	0.9767	0.9769	0.9759	0.9774
Shannon-Weiner index (H′)	4.084	4.076	4.047	4.051
Evenness (E)	0.8988	0.8993	0.8972	0.9047
Margalef index (d)	16.58	16.31	16.05	15.67

The Shannon-Weiner index is also used to measure the species diversity, but excluding the species abundance (Shannon and Weiner, 1963). In this study, the highest value of Shannon diversity index (H′) was found to be 4.084 in Z1, representing the highest species diversity and the 2nd highest value of H′ is recorded to be 4.076 in Z2. Z4 acquired 3rd position in species diversity by scoring 4.051, whereas Z3 recorded the least species diversity with a score of 4.047. These values suggest a rich and complex plant community structure, where species are fairly evenly distributed, minimizing the dominance of any single species. High Shannon-Weiner index values are indicative of healthy ecosystems, capable of providing various ecosystem services and maintaining ecological balance. These diversity metrics are essential for ecological research and conservation strategies, as they help identify key areas for biodiversity preservation and inform management practices aimed at sustaining ecosystem function and resilience. A similar report by Rahman et al. (2011) reported that tree dominated ecosystems produce higher diversity and greater H′ value which is found to be relevant for Z1 of IITG **(Fig. 3.3b).**

Zone 4 ranked 1st with 0.9047 score in evenness (E), indicating the number of individuals in each species in this region are comparatively more equally distributed than other zones. Other three zones scored almost identical score for evenness with Z3 scoring the lowest (0.8972) **(Fig. 3.3c).** Evenness directly affects ecosystem functions such as process rates via species identity and evenness, and indirectly alters the relationship between process rates and species richness (Hillebrand et al., 2008). The species richness referring to the number of individuals in each species and independent of evenness, is calculated using Margalef index (d). The highest d-value (16.58) was observed in Z1, indicating the highest species richness in this zone. The 2nd highest d-value (16.31) observed in Z2, the Z3 got 3rd position with d-value of 16.05 and the lowest species richness was observed in Z4 with d-value 15.67. This results identify Z1 as more species rich than other zones **(Fig. 3.3d).**

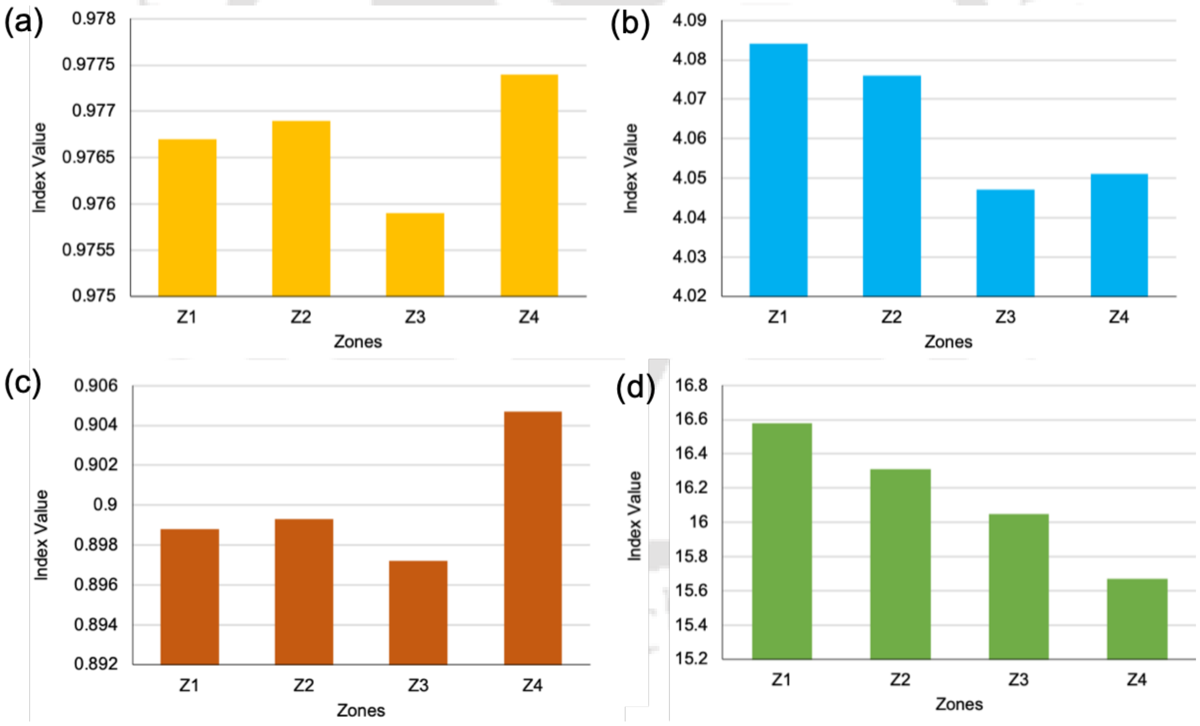


Figure 3.3: Diversity indices for four zones (a) Shimpon’s index, (b) Shannon-Weiner index, (c) Evenness and (d) Margalef index

The E and d values for the four zones provide insights into the distribution and richness of plant species in these areas. The evenness values indicate a relatively uniform distribution of species within each zone, with Z4 having the highest evenness, suggesting the most balanced species distribution, and

Z3 the lowest. This high level of evenness across all zones suggests minimal dominance by any single species, contributing to ecological stability and resilience. The Margalef index values reflect the species richness within each zone, with Z1 showing the highest richness and Z4 the lowest. These values, although showing some variation, still indicate a rich species composition in all zones, which is beneficial for ecosystem functionality and adaptability. These indices are valuable tools in ecological research and conservation planning, helping to identify areas with high biodiversity that may require protection and guiding efforts to enhance species diversity and ecosystem health.

Jaccard's similarity index (J) measures similarity in terms of diversity between two ecosystems or zones. J-value is crucial in identifying similar regions among more than one. The Jaccard's index value is dependent upon the number of common species present between the two areas. The J-value is generally measured in percentage (%) and falls within 0 to 1. The value 1 represents infinite similarity, and 0 means no similarity (Jaccard, 1912). From this value it is found that Z1 and Z4 are most similar among all others with 80% similarity value. The Z2 and Z4 are found to have second highest similarity among them with 75% score. The 3rd highest similarity was found in between Z1 and Z2 (74%). The 4th pair was Z3 and Z4 with 73.65% score. The pair Z2 and Z3 were found to be less similar with 70.7% score and the least similar pair was Z1 and Z3 with only 63.5% similarity **(Table 3.8; Fig. 3.4).**

Table 3.8: Jaccard's Similarity index

S. No	Representation	Z1-Z2	Z1-Z3	Z1-Z4	Z2-Z3	Z2-Z4	Z3-Z4
1	No. of common sps.	197	176	173	198	167	189
2	Total no. of sps.	266	249	235	248	263	252
3	J- Value (%)	74	63.5	80	70.7	75	73.65

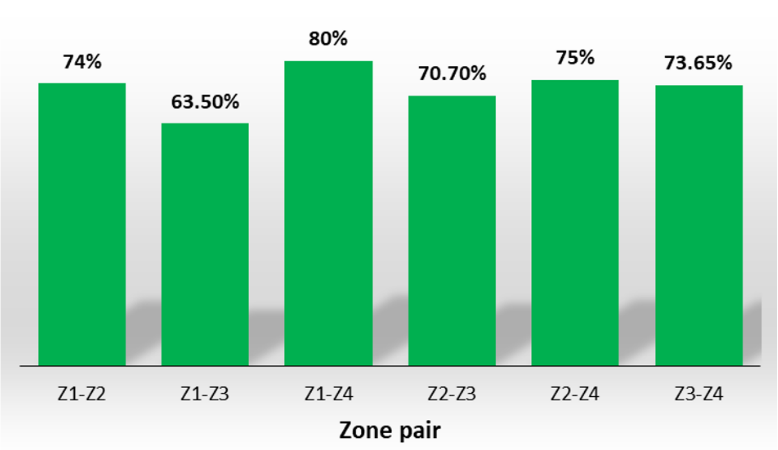


Figure 3.4: Jaccard's similarity index for six pair of zones

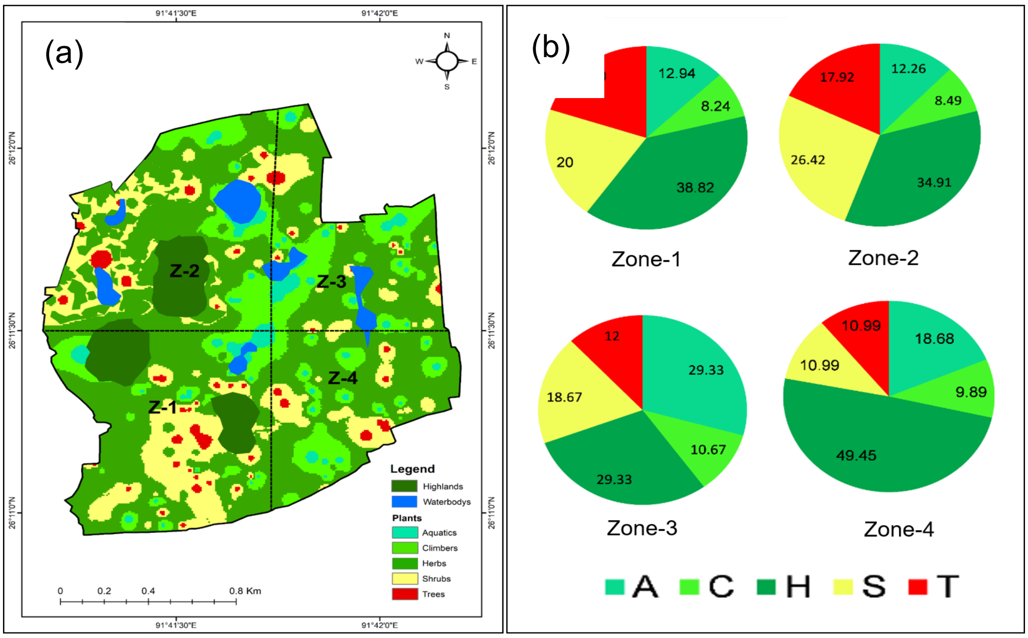


Figure 3.5: (a) Density map of plants and (b) Distribution percentage (Sea green: aquatics, Green: climbers, Deep green: herbs, Yellow: Shrubs and Red: Trees)

The colour-coded density map for aquatics (blue), climbers (green), herbs (dark green), shrubs (yellow) and trees (red) was prepared to highlight the dominant group of plants in specific location of the region (**Fig. 3.5**). The highlands with mountains (darker green) and water bodies (dark blue) with lakes and ponds were highlighted. Most of the areas in the region is covered by herbaceous plants represented by the abundant green colour on the map. In Z1 the shrubs are also found frequently. The climbers and trees in Z1 are scattered and aquatic plants are only located in the water bodies. It is also recorded that the flora in Z1 consists of 12.94% aquatic plants, 8.24% climbers, 38.82% herbs and 20% shrubs and trees each. In Z2, shrubs (26.42%) are also more abundant than herbs (34.91%) and indicated by yellow patches in the map. Trees come in 3rd place in Z2 with 17.92% representation; aquatic plants (12.26%) are also more abundant than climbers (8.49%) in this zone.

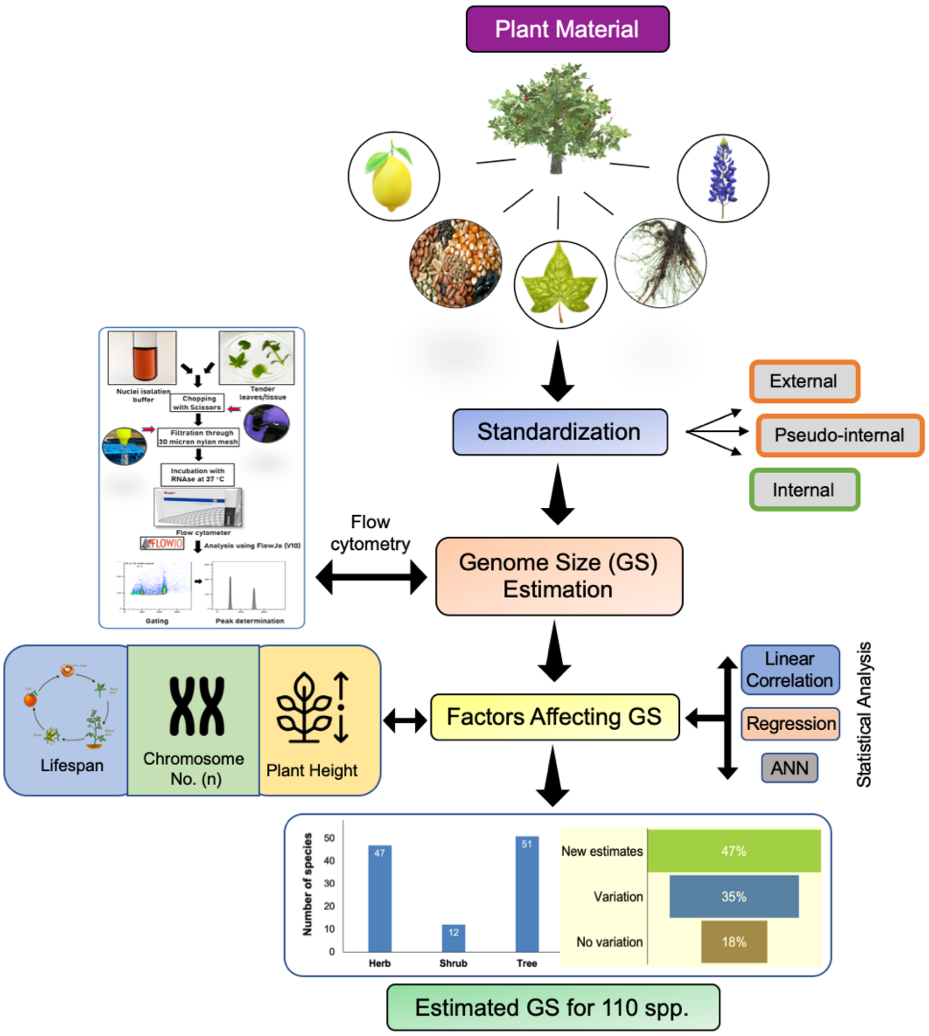
3.5. Conclusion

The IIT Guwahati is undergoing rapid development, marked by extensive construction, deforestation, and over-exploitation, posing a significant threat to ecosystem diversity. This has led to the loss of density in plant species like *F. foetida*, *D. indica*, *A. augustum* etc., with the introduction of exotic plants including *Cardiospermum* spp., *A. repens*, *Parthenium* spp. etc., further altering the region's flora. Recognizing the integral role of plants for various organisms, including animals, birds, and insects, the involvement of policymakers, authorities, communities, and experts is crucial for effective conservation planning (Singh, 2017).

This study emphasizes the urgency of preserving plant diversity in the context of the highly diverse and beautiful IIT Guwahati. Documenting the floristic diversity not only serves as a record but also highlights the importance of different plant species in fostering a healthy ecosystem (Semeraro et al., 2021). The findings of this study contribute valuable insights that can guide conservation efforts and aid in the sustainable management of this ecologically rich area. To ensure the well-being of both natural landscapes and urban areas, policymakers and communities need to prioritize the conservation of forests and other ecosystems in the region (Tian et al., 2020). The study underscores the interconnectedness of plant diversity and ecosystem services, emphasizing that a higher diversity of plants enhances productivity and fauna diversity, contributing to a more vibrant and sustainable environment as supported by Seddon et al. (2021). This scientific approach calls for a holistic perspective in future planning, recognizing the intricate relationships within ecosystems and the need for balanced development to maintain biodiversity and ecological stability in the IIT Guwahati.

Chapter 4

Genome Size Estimation



4.1. Introduction

Within the intricate landscape of plant biology, delving into the GS of vascular plants becomes a pivotal pursuit, offering a profound understanding of the genetic intricacies that govern their evolution, adaptation, and ecological roles. GS, encapsulating the entirety of DNA within a cell, emerges as a fundamental parameter shaping the genomic foundation of vascular plants (Bennett and Leitch, 2005a). The significance of unraveling the GS of vascular plants lies in its direct impact on phenotypic traits, ecological adaptations, and the broader evolutionary trajectories of these organisms (Greilhuber and Leitch, 2012). As custodians of genetic information, vascular plants have developed complex genomic architectures that orchestrate responses to environmental stimuli and interactions within diverse ecosystems. Through the lens of GS estimation, scientists embark on a journey to decode the genetic blueprints dictating the intricacies of vascular plant biology.

This scientific endeavor employs advanced methodologies, with a particular focus on techniques such as flow cytometry, ensuring the relative quantification of DNA content (Doležel and Bartoš, 2005). The choice of methodologies is critical, guaranteeing data reliability and opening avenues for comprehensive analyses. Beyond basic research, the exploration of vascular plant GS carries practical implications for biodiversity conservation and agricultural advancements. This study establishes the foundation for a comprehensive exploration into the genetic underpinnings of 110 vascular plants identified from IIT Guwahati campus, promising insights that transcend the boundaries of traditional botanical inquiry.

4.1.1. Background and significance of GS in plants

Genome size in plants refers to the total amount of DNA contained within one copy of a single complete set of chromosomes, is a fundamental and dynamic feature in the realm of plant biology. The diversity observed in GS across different plant species has intrigued scientists for decades, as it reflects a complex interplay of evolutionary forces and ecological pressures (Bennetzen et al., 2005). Understanding the background of GS in plants involves unraveling the mechanisms underlying genome expansion and contraction, exploring the implications for plant physiology, and discerning the functional significance encoded within the vast stretches of DNA. The intricacies of GS are underscored by the fact that plants exhibit an extraordinary range of GS, from the compact genomes of carnivorous *Genlisea tuberosa* (2C=0.12 pg) to the massive and repetitive genomes of certain plants like *Paris japonica* (2C= 304.7 pg) (Pellicer et al., 2018).

The significance of GS in plants extends beyond mere quantitative measures, as it profoundly influences various aspects of plant biology. One key aspect is the impact on plant development and morphology, where it correlates with cell size, cell division rates, and overall growth patterns

(Geilhuber and Leitch, 2012). Moreover, GS is intimately linked to the adaptability of plants to diverse environmental conditions (Faizullah et al., 2021). Plants with larger genomes may possess a reservoir of genetic material that enables them to respond more flexibly to environmental changes, potentially influencing their ability to thrive in varied ecological niches (Mei et al., 2017). Additionally, the study of GS in plants provides valuable insights into the mechanisms of genome evolution, including the role of repetitive elements, polyploidy events, and other genomic rearrangements that contribute to the vast genomic diversity observed across the plant kingdom (Pellicer et al., 2018).

From an applied perspective, the significance of GS in plants has implications for agriculture, as it can influence breeding strategies, crop yields, and the development of stress-resistant varieties (Kim et al., 2020). The GS is gaining importance in deciphering the genetic basis of desirable traits, guiding crop improvement efforts, and ensuring food security in the face of a changing climate (Bevan et al., 2017). The background and significance of GS in plants encapsulate a multifaceted scientific endeavor that spans from understanding the molecular intricacies of DNA composition to unraveling the ecological and agricultural implications that underscore the importance of it in shaping the plant kingdom.

4.1.2. GS diversity in land plants

The narrative of GS diversity in land plants is a captivating saga that unfolds across the vast tapestry of the plant kingdom. From the microscopic mosses to the towering trees, land plants exhibit remarkable variation of GS, a testament to the dynamic forces shaping their genetic landscapes (Leitch and Leitch, 2012). The story begins with the realization that GS is not a static attribute but a variable trait subject to evolutionary forces. This diversity is particularly pronounced among land plants, with genome sizes varying over 2,400 fold (Pellicer et al., 2018). The available data shows that large eukaryotic genomes tend to converge around 150,000 Mb, implying the influence of evolutionary forces limiting GS expansion. This observation suggests a potential upper limit to GS, indicating intricate mechanisms or constraints at play (Hidalgo et al., 2017). This diversity delicately woven into the fabric of plant evolution, underscores the adaptability and resilience of land plants in the face of ever-changing ecological landscapes.

As we delve deeper into the scientific narrative, the variation in GS diversity serves as a molecular signature, reflecting the evolutionary history and adaptive strategies of different plant lineages. This extends beyond the confines of the laboratory, reaching into the heart of ecosystems where plant species with diverse GS navigate the challenges of their environments (Michael, 2014). The reason behind this diversity lie in the complex interplay of factors such as repetitive DNA elements, polyploidy events, and genome expansions or contractions (Pellicer et al., 2018). GS diversity in land plants emerges as a source of inspiration for applied research and conservation efforts. Understanding the genetic architectures that underpin this diversity opens avenues for harnessing the potential of plant species with unique GS in agriculture, medicine, and ecological restoration. The story continues to unfold as scientists unravel the mysteries embedded in the DNA sequences of land plants, a narrative that intertwines the threads of evolutionary history, ecological adaptation, and the potential for sustainable coexistence between plants and their environments.

4.1.3. Methodological advances in GS estimation

GS estimation is the process of determining the total amount of DNA contained within a single complete set of an organism's chromosomes. This measurement is typically expressed in units such as picograms (pg) or megabase pairs (Mb) and provides insights into the genetic complexity of the organism. GS estimation in plants has evolved over time, with the development of various methods aimed at quantifying the amount of DNA. Historically, early attempts involved labor-intensive procedures such as Feulgen staining and microdensitometry (Vilhar et al., 2001; Greilhuber et al., 2010). While these methods laid the groundwork for GS research, they were time-consuming and lacked the precision needed for detailed analyses. Another notable method in GS estimation is the use of fluorescence microscopy, which involves visualizing stained nuclei under a microscope providing valuable insights into nuclear organization, and is often limited by its labor-intensive nature and subjectivity in interpreting results (Cheng et al., 2019). Quantitative real-time PCR (qPCR) has also emerged as a method for relative GS estimation by comparing the amplification of a target DNA sequence to a reference gene. This approach is advantageous in its sensitivity but may be influenced by factors such as primer specificity and amplification efficiency (Gregory et al., 2013).

Advancements in technology have led to the development of high-throughput sequencing-based methods, including whole-genome sequencing (WGS) and k-mer analysis. These methods provide comprehensive insights into GS and structure but come with the drawback of high cost and computational demands (Hesse, 2023). GS quantification witnessed a transformative shift with the advent of flow cytometry in the 1970s (Wakamiya et al., 1973), which significantly increased the efficiency of GS estimation, enabling researchers to study large plant populations with greater accuracy. In recent years, the field of plant flow cytometry has experienced notable advancements, including the utilization of high-throughput flow cytometry, imaging cytometry, and single-cell genomics, signifying a revolutionary phase in the discipline (Vrána et al., 2021; Zhang et al., 2021).

4.1.4. Scope of the study

The scope of the study involves a comprehensive exploration into the GS of 110 vascular plants using the advanced technique of flow cytometry. This research aims to contribute valuable insights into the genetic diversity and dynamics present within the selected vascular plant species. By employing flow cytometry, the study seeks to provide accurate and efficient measurements of DNA content, allowing for a nuanced understanding of the GS across this diverse plant cohort. The outcomes of this research endeavor hold significance not only in advancing the field of plant genomics but also in providing essential information for ecological and evolutionary studies.

Furthermore, the study's scope extends to the potential implications of GS variation in the adaptive strategies and ecological roles of vascular plants. The data generated from this research may shed light on correlations between GS and specific internal factors, contributing to a broader understanding of how these plants respond to their surroundings. The knowledge gleaned from this study could further inform conservation efforts, as well as agricultural and biotechnological applications by identifying key genomic attributes associated with specific plant species. In essence, the scope of this study transcends mere quantification, aiming to unravel the intricate relationships between GS and the ecological and evolutionary contexts of vascular plants through the lens of flow cytometry.

4.2. Materials and methods

4.2.1. Selection of plants

Following the meticulous identification of plant species from IITG campus by collaborating with taxonomists and botanists, 110 vascular plants with unknown GS were selected. The selection process aimed to encompass a diverse array of ecological niches, families, and morphological characteristics. Priority was given to including plants from various habitats to capture the broad spectrum of genomic diversity within the campus flora. Special attention was paid to inclusion of both common and rare species without genomic information, contributing to GS estimation of the local plants.

4.2.2. Sample collection

In the pursuit of estimating genome size for subsequent genomic studies, a systematic approach was employed for plant sample collection. Representative specimens, reflecting the genetic diversity within the species, were carefully chosen from diverse populations within the campus. Priority was given to young, healthy leaves for their abundance in nuclei, ensuring a reliable source of genomic material. Stringent measures including sterile tools, hand gloves, single use sample containers and sterile cleaning materials were implemented to prevent contamination and maintain genetic integrity of the plant sample throughout the collection process.

4.2.3. Nuclei isolation buffer

Nuclei isolation buffer is instrumental in extracting nuclei from plant tissue and preserving their integrity during the subsequent staining process. Typically composed of salts, buffering agents, and detergents, the isolation buffer facilitates the breakdown of cell walls, aiding in the release of nuclei into suspension. Its significance extends to stabilizing the nuclei, preventing clumping or degradation during staining, thereby ensuring accurate measurements of DNA content through flow cytometry. Achieving optimal release of nuclei demands careful fine-tuning of the isolation buffer composition, a challenge that becomes pronounced when dealing with samples containing staining inhibitors such as phenols, alkaloids, tannins, and other secondary metabolites in the cytosol (Price et al., 2000).

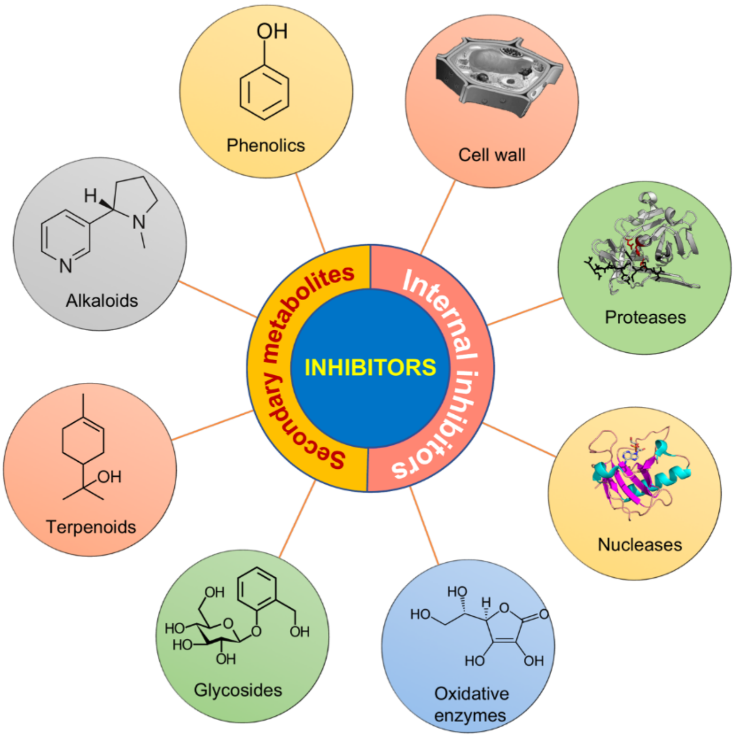


Figure 4.1: Inhibitors and secondary metabolites interfering in plant flow cytometry

Table 4.1: Modified hypotonic PI buffer components and their role

Components	Concentration	Activity
Propidium iodide	25 µg/ml	DNA intercalating dye (Fluorochrome)
RNase-A	0.5 mg/ml	Catalyzes RNA degradation
IGEPAL	0.3% (V/V)	Non-ionic, non-denaturing detergent breaks cell membrane and dissolves lipid bi-layer
Citric acid	0.1% (W/V)	Buffering nuclei integrity
Sodium citrate	0.1% (W/V)	Chelating and buffering
PVP-40	1% (W/V)	Membrane disruption and removal of phenolic compounds
2-ME	1% (V/V)	Reducing agent

In addressing these challenges, intrinsic inhibitors present in the sample can be mitigated through the use of cell wall degrading enzymes (pectinases), protease inhibitors (EDTA), nucleases inhibitors (RNase and DNase), oxidative enzyme inhibitors, and chelating agents (**Fig. 4.1).** Employment of non-ionic detergents, chemical precipitation, antioxidants, chelating agents, and organic solvents as inhibitors to minimize the impact of secondary metabolites was done (Ramesh et al., 2014). Despite extensive studies, a universal method to counteract interfering substances is yet to be established, reflecting the complex nature of staining inhibitors and their mechanisms of action (Greilhuber et al., 2007). The integration of antioxidants like polyvinylpyrrolidone (PVP), β-mercaptoethanol (BME), or dithiothreitol (DTT) into the buffer has been identified to enhance the isolation of intact nuclei and improve flow cytometry results.

Previous observations highlight the unique responses of different tested nuclei isolation buffer to specific challenges, emphasizing the absence of a one-size-fits-all solution. In this study, the adoption of the modified HPI buffer containing PVP for plant nuclei extraction has been found to be instrumental (**Table 4.1**). PVP, a water-soluble polymer, serves to eliminate polyphenols and contaminants from plant tissues. The MHPI buffer comprises 25 µg/mL of PI in a 0.1 M hypotonic trisodium citrate dehydrate solution with 0.3 µl/mL of the non-ionic detergent IGEPAL CA-630. Further supplementation includes 1% PVP-40 and 2 mg/ml of RNase-A in the hypotonic solution. The buffer is finalized by the addition of 25 µg/ml of PI and is stored at 4°C. To prevent self-oxidation, 1% BME is added just before use, and due to PI's light sensitivity, the buffer is stored in darkness. Additionally, the pH of the HPI buffer was maintained at 7 to meet physiological conditions.

4.2.4. Sample preparation

Fresh leaf samples of fully developed leaf sheaths from each plant species were transported immediately to the laboratory in coolers containing ice packs to preserve cellular integrity. In the laboratory, leaf tissues (1 cm^2) were finely chopped in a Petri dish with 1 ml of modified hypotonic PI (MHPI) nuclei isolation buffer using scissors. Nuclei were then isolated using a modified nuclei isolation buffer containing a DNA-specific fluorochrome, propidium iodide (PI), and treated with ribonuclease-A and detergents to remove RNA and break cell walls. The isolated and intact nuclei were obtained using 30µm nylon mesh after removing unwanted debris and tissues. The nuclei were incubated for 10-30 minutes before analyzing in flow cytometry. The MHPI was a nuclei isolation buffer developed from hypotonic citrate buffer by Krishan (1975).

4.2.5. Standardization methods

A pivotal facet of flow cytometry lies in the methodology of standardization, a process integral to ensuring the consistency and reliability of flow cytometry data. This standardization can be achieved through the adoption of external, internal, or pseudo-internal standards, each contributing distinct advantages to the analytical process. In the external standard approach, both the sample plant and the reference standard undergo discrete chopping procedures and subsequent individual analyses. Conversely, the internal standard method involves the simultaneous chopping of both sample and reference plants within the same Petri plate, followed by joint analysis. Lastly, the pseudo-internal standard technique encompasses the separate extraction of nuclei from both the sample and reference standard plants, with subsequent mixing just before analysis (**Fig. 4.2**).

The internal standardization method emerges as particularly reliable, as it involves identical conditions for sample preparation in both sample and standard plants, ensuring uniformity and accuracy in flow cytometry analyses (Ramesh et al., 2014). Consequently, this study opted for the internal standardization approach for all analyses, thereby providing a robust and dependable estimation of GS in vascular plants. This methodological choice aligns with the pursuit of meticulous and standardized procedures, enhancing the accuracy and interpretability of the flow cytometry data obtained in this investigation.

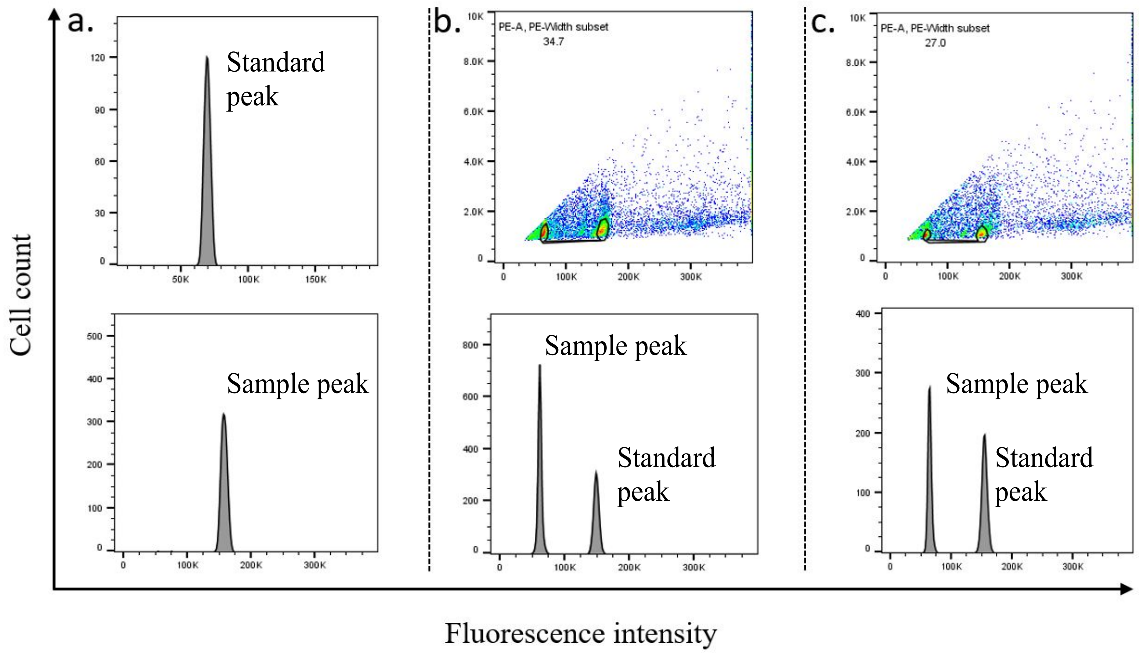


Figure 4.2: Standardization methods for plant DNA analysis (a) external standardization; (b) pseudo-internal standardization, and (c) internal standardization

4.2.6. Flow cytometric analysis

GS estimation was carried out employing a high-performance flow cytometer (CytoFlex, Beckman Coulter) utilizing cell cycle analysis (Das et al., 2018) (**Fig. 4.3**). The fluorescence intensity of stained nuclei was quantified as they traversed a laser beam during analysis. To ensure precision and dependability of measurements, a reference standard with a known genome size was incorporated for calibration in each analytical run. Data collection for each plant species was executed in triplicate to guarantee the reproducibility of results, adhering to rigorous scientific standards in the flow cytometric assessment of GS.

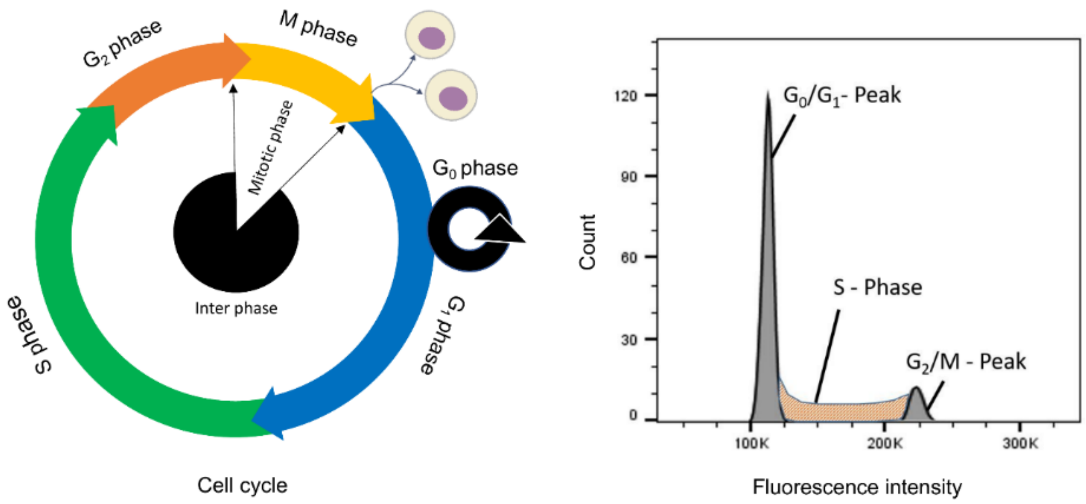


Figure 4.3: Cell cycle analysis through flow cytometry

4.2.6.1. *Instrument calibration*

Instrument parameters in the flow cytometer are crucial for obtaining accurate and reliable data **(Fig. 4.4**). Here are some of the critical instrument parameters:

(i) Dot plots display the light scattered by the cells or fluorescence signals. The choice of parameters for a dot plot's X and Y axis can affect the data's visualization and analysis. For example, a dot plot of forward scatter (FSC) versus side scatter (SSC) can identify cell populations based on their size and granularity.

(ii) Flow rate: The flow rate of the sample through the flow cell affects the rate at which the cells pass through the lasers and detectors. A slower flow rate can increase the sensitivity and resolution of the data but may also decrease the number of cells that can be analyzed per unit of time.

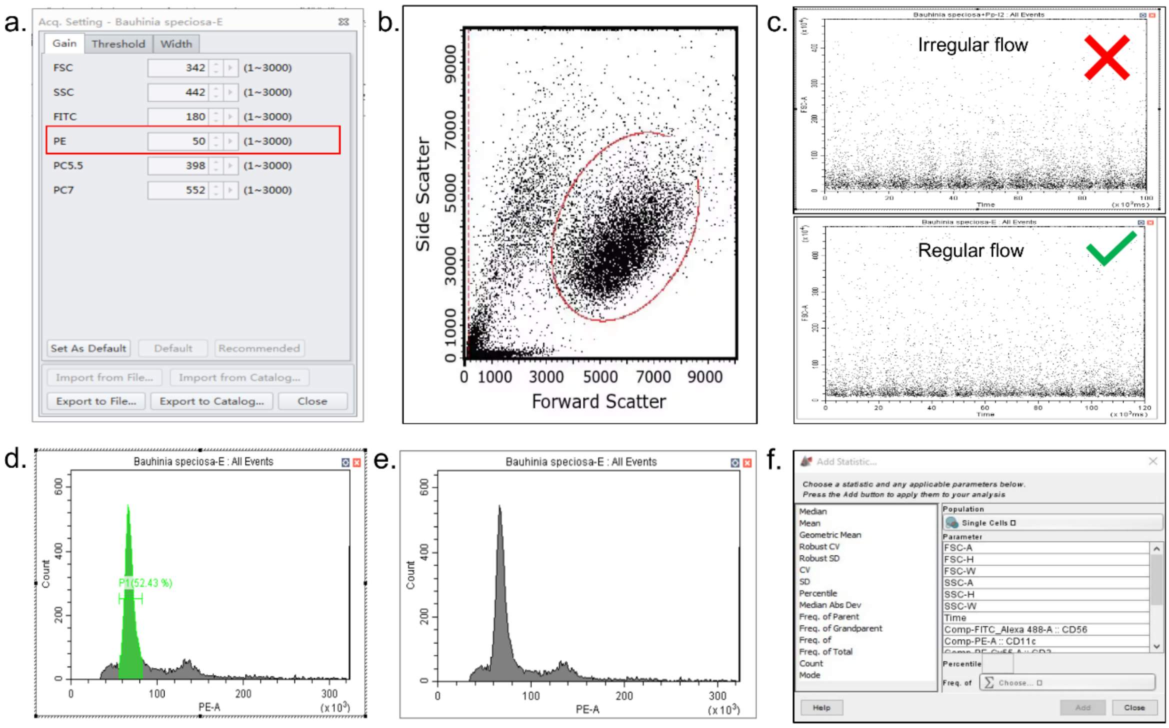
(iii) Voltage settings: Voltage settings determine the sensitivity and range of the fluorescence detectors. Setting the voltage high enough to detect signals above the background noise is essential but not so high as to cause detector saturation. Each parameter's voltage settings should be optimized for accurate and reliable data.

(iv) Gating refers to selecting a subset of cells based on their position within a dot plot or histogram. Gates can be used to remove debris, dead cells, or doublets from the analysis or to identify specific cell populations based on their fluorescence or scatter properties.

(v) Peaks: Peaks in histograms represent the frequency of cells with a particular fluorescence or scatter signal level. The position and shape of peaks can provide information about the distribution and heterogeneity of the cell population.

(vi) Gating refers to selecting a subset of cells based on their position within a dot plot or histogram. Gates can be used to remove debris, dead cells, or doublets from the analysis or to identify specific cell populations based on their fluorescence or scatter properties.

However, general instrument parameters and their optimal settings may vary depending on the specific experiment and sample type. Optimization experiments and following the manufacturer's instrument parameters guidelines can help obtain high-quality and reliable data.



Acq. Setting - Bauhinia speciosa-E

FSC 342 (1~3000)
SSC 442 (1~3000)
FITC 180 (1~3000)
PE 50 (1~3000)
PC5.5 398 (1~3000)
PC7 552 (1~3000)

Add Statistic...

Choose a statistic and any applicable parameters below.
Press the Add button to apply them to your analysis

Population: Single Cells

Figure 4.4. Flow cytometric calibration for plant DNA analysis (a) voltage settings; (b) data acquisition (c) flow rate; (d) gating; (e) peak and (f) statistics

4.2.6.2. *Reference standard plants*

In the context of flow cytometry for plant cell DNA content analysis, the utilization of plant DNA standards is imperative. These standards, featuring a compendium of plants with well-established DNA content, serves as a cornerstone for ensuring the accuracy and reliability of results. Beyond GS quantification, these reference standards find application in ploidy determination and cell cycle analysis, playing a pivotal role in fortifying the precision and reproducibility of flow cytometry measurements within the realm of plant research.

The selection of an appropriate reference standard entails fulfilling specific prerequisites. It should closely approximate the GS of the target species to mitigate the risk of nonlinearity in measurements. In particular the GS of standard plant should fall within four-fold of the sample in both the directions (Temsch et al., 2022). Additionally, the chosen standard should exhibit genetic stability, maintaining a consistent GS, while also being readily available in ample quantities (Dolezel and Greilhuber, 2010). Meeting these criteria poses a challenge, leading to instances where researchers resort to standards other than those reported for the respective plant species. Given the extensive range of GS in plants, spanning over a thousand-fold, a curated set of reference standards becomes imperative for robust GS estimation (Dolezel and Bartos, 2005).

Table 4.2: Reference standards used for GS estimation

S. No.	Species	2C DNA content (in pg)
1	*Vigna radiata*	1.0
2	*Oryza sativa*	1.01
3	*Ricinus communis*	1.3
4	*Solanum lycopersicum*	1.96
5	*Pongamia pinnata*	2.66
6	*Zea mays*	5.43

In the present study, six reference standards for GS estimation were used namely, *P. pinnata* (NGPP46), *R. communis*, *O. sativa*, *S. lycopersicum*, *V. radiata*, and *Z. mays* (**Table 4.2).** The *P. pinnata* and *R. communis* were characterized and tested extensively by our laboratory before being accepted as reference standards. The other four standard plants were grown in green house conditions, thanks to Prof. Dolezel, who kindly provided the seeds. The young and tender leaves (1 cm^2) of the desirable standard plants were used along with the sample plant tissues for internal standardization and consequently GS estimation of sample plants. The rationale behind choosing these plants as reference standards was their universal availability and ease of maintenance under greenhouse conditions, underscoring their suitability for ensuring the fidelity of flow cytometry analyses (**Fig. 4.5**).

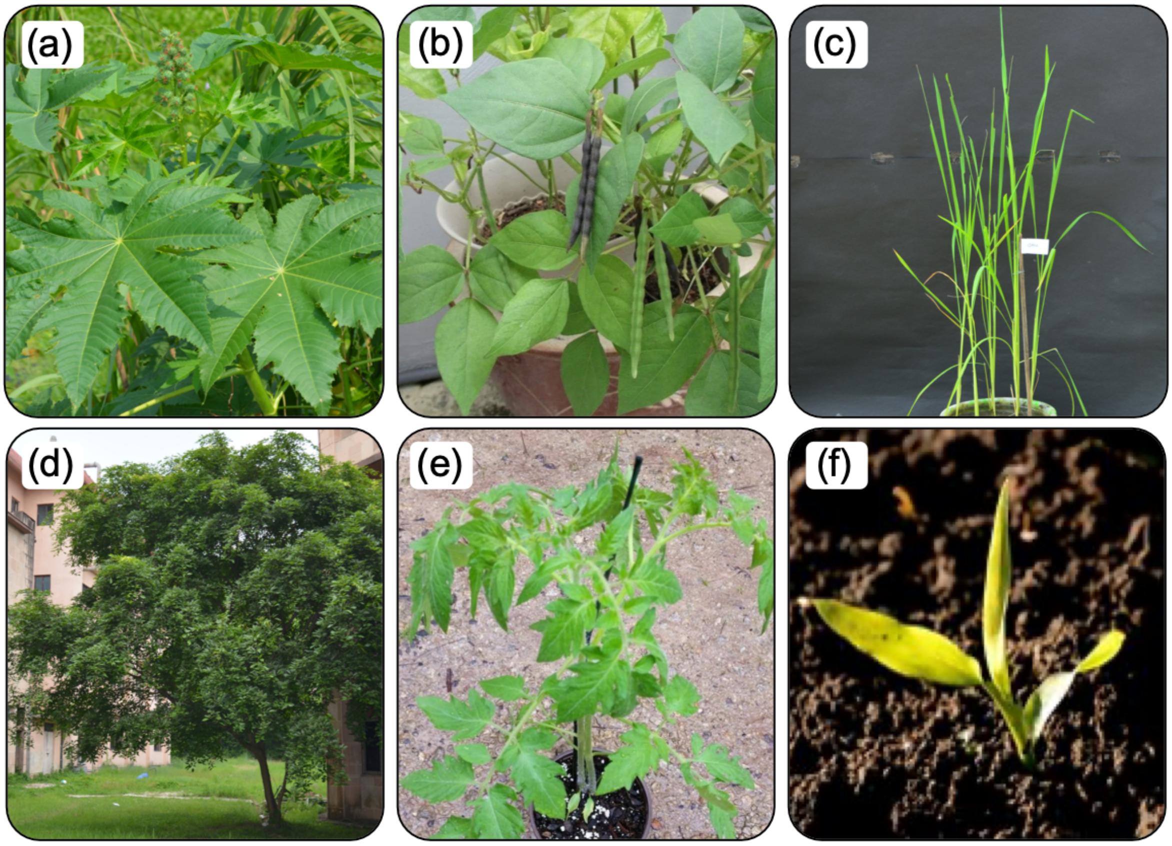


Figure 4.5: Standard plants used for GS estimation (a) *R. communis*, (b) *V. radiata*, (c) *O. sativa*, (d) *P. pinnata*, (e) *S. lycopersicum*, and (f) *Z. mays*

4.2.7. Ethical considerations and quality control

As per Wildlife Protection Act, 1972 (India); section 2(16), section 9 and 11, harvesting and collection of specific plants is regulated from protected areas. According to Biodiversity Act, 2002 adherence to ethical guidelines and consulting authorities is suggested. Since, IIT Guwahati doesn't fall under protected areas all plant collections were performed complying with ethical guidelines and necessary permissions were diligently secured from institute authorities wherever necessary along with continuous efforts to mitigate potential ecological impacts. This ethical and conscientious efforts underscored the commitment to responsible scientific conduct throughout the plant collection process.

Ensuring the precision of flow cytometry analyses mandated rigorous adherence to quality control protocols. These encompassed routine instrument calibration, incorporation of internal controls, and systematic evaluation of the coefficient of variation both within and between analytical runs.

Rigorous scrutiny of data points was undertaken, and any identified outliers were subject to careful examination, prompting re-analysis when necessary. The meticulous execution of these quality control measures underscored the commitment to maintaining accuracy and reliability throughout the flow cytometry procedures.

4.2.8. Data analysis

The raw flow cytometry data, including DNA content histograms and fluorescence peaks, were processed using specialized software (FlowJo V10). The mean and coefficient of variation for each sample were calculated. GS was estimated based on the fluorescence intensity relative to the reference standard plant. Statistical analyses, including correlation studies and comparisons between vascular plant families, were performed to elucidate patterns in genome size variation. The nuclear DNA content is estimated by comparing the mean fluorescence intensity of nuclei of the sample material with that of the reference standard and obtained by multiplying the nuclear DNA content of standard species by the ratio of their fluorescence intensities (**Fig. 4.6**).

$$\text{2C Sample} = \text{2C standard} \times \frac{\text{G1 sample}}{\text{G1 standard}}$$

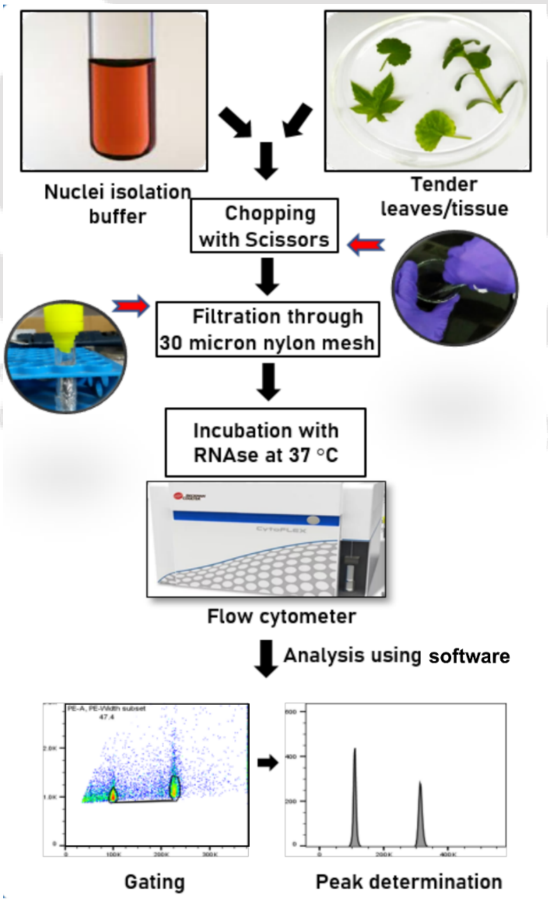


Figure 4.6: Genome size estimation of plant samples through flow cytometric analysis

In this context, the 2C sample denotes the 2C DNA amount of the test sample (in pg), the G1 sample represents the mean fluorescence intensity of the G1 peak (2C) of the sample, the 2C standard signifies the 2C DNA amount of the reference standard, and G1 standard indicates the mean fluorescence intensity of the standard G1 peak (2C). The GS is convertible between picograms and base pairs through the utilization of the formula outlined by Dolezel et al. (2003). This formula given below, facilitates the accurate inter-conversion of GS measurements, allowing for seamless comparisons and analyses in the context of plant flow cytometry studies.

$$\text{DNA content (pg)} = \frac{(\text{Genome size(bp)}}{(0.978 \text{ x } 10^9\ )}$$

Where, 1 pg DNA=978 Mbp

4.2.9. Correlation among genetic and morphological traits

The correlation between genetic trait like haploid chromosome number (n) and morphological traits like life span and plant height was compared to GS of the plants through statistical methods like linear correlation test and regression analysis along with machine learning (artificial neural network) on 95 herbs, 33 shrubs and 71 trees. The information on haploid chromosome number of all the species was obtained from Chromosome Counts Database (CCDB) (accessible through http://ccdb.tau.ac.il/) or from scientific literatures. The morphological information was obtained from various databases including Flora of India (https://bsi.gov.in/page/en/flora), the WFO Plants List (https://wfoplantlist.org/plant-list/) , Plant GDB (https://www.plantgdb.org) etc.

4.2.9.1. *Linear correlation test:* The examination of potential correlations among GS, haploid chromosome number (n), life span, and plant height was conducted through the implementation of a linear correlation test through Pearson's R test.

4.2.9.2. *Regression analysis:* The regression analysis was also performed through statistical package in MS-Excel by multiple R analysis. In this analysis Multiple R value, R square value, and adjusted R square values were calculated along with standard error for herbs, shrubs and trees separately. Also, coefficients of variables were calculated for chromosome number, plant height and life span.

4.2.9.3. *Artificial Neural Network (ANN) modeling:* An ANN model was employed to investigate and model the intricate relationships among GS, haploid chromosome number, life span, and plant height through the software Python language (v3.0.0) available through www.python.org. In this model number of hidden neurons (hn) were shuffled between 1 to 100 to obtain the best fitting model, where hn refers to the size of the matrix inter-linking variables to that of response.

4.2.10. Statistical analysis

In the statistical analysis of GS estimation, a variety of metrics including mean, median, and mode of G0/G1 histogram peaks were determined using FlowJo (v10) software from internal standardization histograms for each species. The mean, represented the average fluorescence intensity, while the median served as a measure of central tendency, and mode provided information about the maximum frequency of the fluorescence. For each species flow cytometric analysis in triplicates was performed to obtain consistent results which was confirmed by coefficient of variation (CV) analysis. The CV was used to determine the relative variability among estimated GS, and standard deviation (SD) were also taken into consideration for finding out dispersion of data points. These statistical parameters were incorporated to enhances the robustness and interpretability of the methodology.

4.3. Results and discussion

4.3.1. Sample collection and preparation

Plant specimens were collected from diverse ecological niches, representing a broad range of plant species from IITG campus including wild and forest regions to organized gardens and backyards. Successful flow cytometric analysis of a subset of 110 plants was done based on types of plants, resource constraints and availability of samples. Although, GS estimation has been reported from various tissues of plants including roots, shoot tips, cotyledons, embryonic tissues, pollen grains and seeds in some plants (Pellicer and Leitch, 2014; Krahulcova et al., 2017; Ejaz et al., 2021). But, it is difficult and tricky to obtain suitable sample in all these cases. In other hand, leaf tissues were predominantly collected for genome size estimation due to their reliability and uniform representation of nuclear DNA content (Pellicer et al., 2021). In particular young leaf tissues have several advantages including easy availability throughout the growing season, actively dividing cells in proliferative phase of cell cycle, consistency in nuclei content and reduced cytoplasmic content (Temsch et al., 2022; Loureiro et al., 2023). Other advantages of young leaves include low levels of secondary metabolites, consistency in tissues, and ease of collection and handling making then the best choice for flow cytometric analysis as reported by Sliwinska et al. (2022) in best practices in plant flow cytometry, hence in all our analysis young tender leaves has been used for GS estimation.

First and foremost the collected plant samples were stored in sterile polythene bags, sandwiched in-between wet tissue papers to preserve cell structure and keep cells alive (Qiu et al., 2020). Subsequent freezing of the plant samples ensured minimal cellular metabolism, prevention of DNA degradation, retention of nuclear integrity, enhanced sample stability, extended shelf life and enhanced reproducibility (Sliwinska et al., 2022; Fomicheva and Domblides, 2023). Rapid freezing ensured the retention of sample consistency, reduction of microbial growth and minimized potential changes in DNA content (Loureiro et al., 2023).

Prior to nuclei extraction, fresh/frozen leaf tissues of 1 cm^2 of standards and test plants were finely chopped using a pair of sterile curved scissors within the nuclei isolation buffer for efficient disruption of cell walls and release of nuclei (Zhou et al., 2023). It was ensured that young tender leaves were used and where not available rough/waxy and leathery leaves having midrib and strong veins were removed before being processed. For efficient staining of the nuclei the samples were incubated for 30 mins in dark conditions to prevent degradation of the photosensitive dye (PI) (Dolezel et al., 2007). Unanalyzed specimens ware preserved in low temperatures for potential future investigations, allowing for expanded analyses or validation of findings. This scientific report outlines the comprehensive process of plant specimen collection, and sample preparation, emphasizing the significance of careful handling and methodological choices in ensuring accurate and meaningful results.

For some high secondary metabolite containing and mucilaginous plants, FCM analysis was unsuccessful due to difficulties in isolation of free nuclei (Loureiro et al., 2021). The modified methodology helped in successful GS estimation of some high secondary metabolite containing plants like *C. tamala*, *B. monnieri*, and *C. sinensis* and it was majorly unsuccessful in *Rhododendron* spp. and in Zingiberaceae plants (Sliwinska, 2018). Additionally, we were able to solve the problem of mucilaginous plants in few cases including *Abelmoschus* sp., but it was highly challenging to obtain free nuclei and blocked sample flow in instrument for species like *H. rosa-sinensis* and cactus plants (Raghu et al., 2019). Some other specimens were excluded due to insufficient nuclei yield, limited time and resource, small and fragile nuclei as found in multiple Orchids, or sample degradation during collection and processing.

4.3.2. Buffer selection and composition- modified hypotonic PI buffer

The plant nuclei isolation buffer in this study was developed from the hypotonic citrate buffer developed by Krishan (1975) for mammalian cell cycle analysis. Addition of PVP-40, 2-ME and inclusion of RNase-A directly to the buffer was done for specific requirements of plants. Concentration of different components of the modified buffer was carefully optimized for better extraction of intact nuclei from wide range of plant tissues for subsequent flow cytometric analysis. Each component of the buffer plays a crucial role in facilitating the isolation process and preserving the quality of the nuclei. Compared to previously used buffers (LB01 and Galbraith's) and commercially available buffer (Ploidy reagent, Sysmex), both the nuclei isolation quality and quantity was found to be better in modified HPI buffer **(Fig. 4.7)**.

The specific activities of each component in the HPI buffer was as follows: The PI (25 µg/ml), was a vital component, serving as a DNA intercalating dye, which is excited by light in the blue range (around 488 nm), and upon binding to DNA, it emits fluorescence in the red-orange range (around 617 nm) (Choudhury et al., 2014). The emitted light is then captured by detectors in the red channel of the flow cytometer. The binding of PI to DNA is stoichiometric, meaning the fluorescence intensity is proportional to the amount of DNA present in the cell, allowing accurate determination of DNA content in the isolated nuclei (Frydrych, 2021). Also, PI is a well-known carcinogen and light sensitive chemical, utmost care was taken while handling the chemical and dark conditions was established to prevent degradation (Dey, 2021). The RNase-A (0.5 mg/ml), a RNA degradation catalyst, played the critical role of degrading RNA contaminants present in the nuclei isolation. This enzymatic activity ensured that the subsequent flow cytometric analysis focuses specifically on DNA content without interference from residual RNA (Yang et al., 2019). Another important component for membrane disruption was a non-ionic detergent IGEPAL (0.3% v/v). This detergent efficiently disrupted the cell membrane and dissolves the lipid bi-layer. This step was essential for releasing intact nuclei from the cells, allowing for their subsequent isolation (Choudhury et al., 2014).

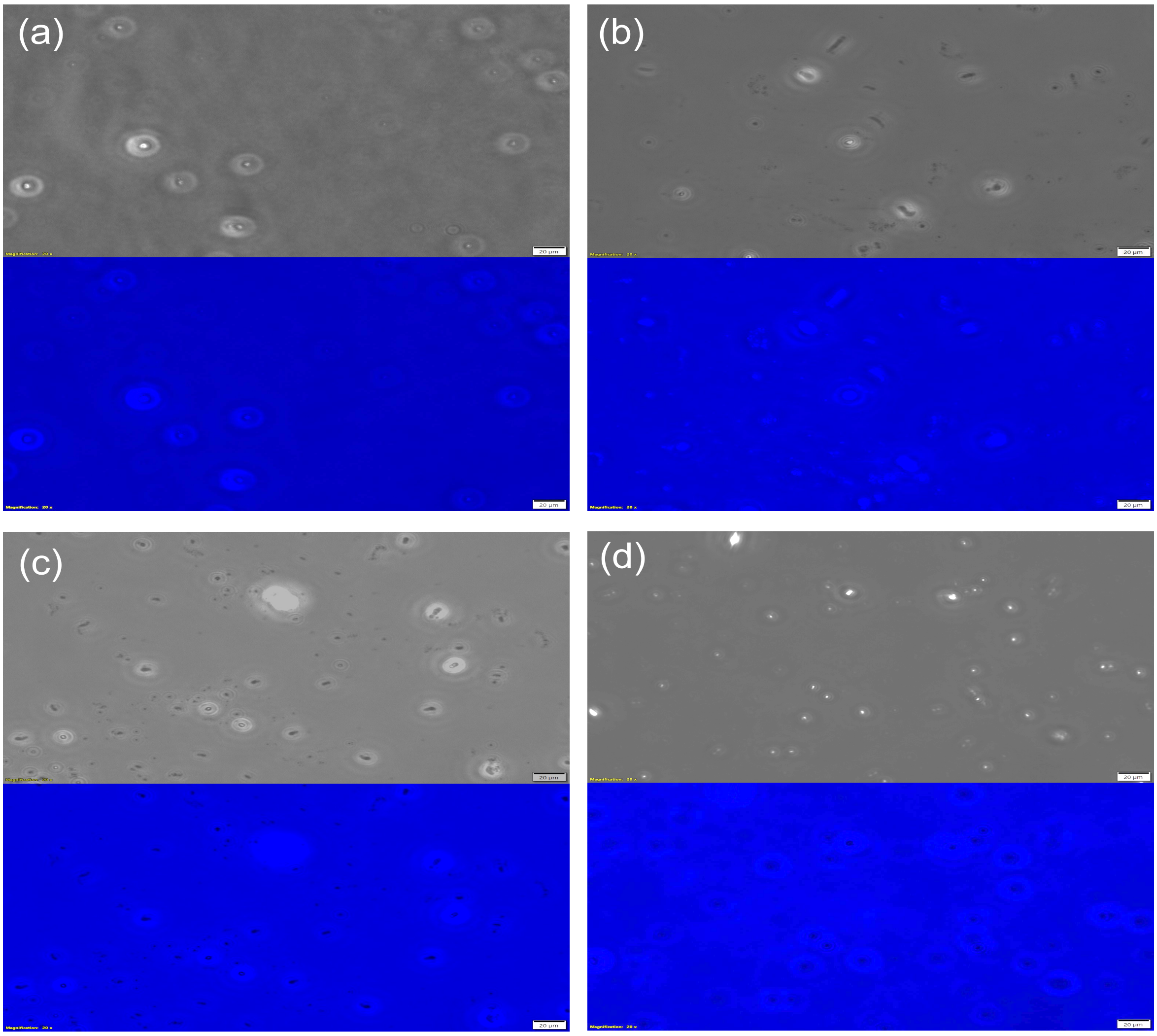


Figure 4.7: Microscopic observations of comparative free nuclei isolation among different buffers observed at 20X in *A. oleracia* (a) LB01 buffer, (b) Galbraith's buffer, (c) Ploidy reagent (Sysmex), and (d) Modified hypotonic PI buffer (upper: bright field image, lower: blue filter)

Citric acid (0.1% w/v) served as a buffering agent to maintain the integrity of the isolated nuclei. It helped in stabilizing the pH of the solution, creating an environment conducive to preserving the structural integrity of the nuclei and has been used previously in Otto's buffer and Baranyi's nuclei isolation buffer (Otto, 1990; Baranyi and Greihuber, 1995). Sodium citrate (0.1% w/v) acted as both a chelating agent, preventing the interference of metal ions, and a buffering agent, contributing to the overall stability of the buffer, which was also used in well known Galbraith's buffer (Galbraith et al., 1983). This dual function was crucial for maintaining the quality of the isolated nuclei. Another important component of HPI buffer was PVP-40, which is a polymer with an average molecular weight of 40 kDa, from polyvinylpyrrolidones group. This compound replaced Tween-20 from previous buffers and played a dual role at 1% w/v in disrupting cellular membranes and removing potentially interfering phenolic compounds (Gautam, 2022). This contributed to the purity and improved quality of the isolated nuclei for subsequent flow cytometric analysis. 2-Mercaptoethanol (2-ME) (1% v/v) is a commonly known reducing agent, prevented the oxidation of cellular components during the isolation process. 2-ME is also light sensitive, hence used in dark experimental conditions. This is crucial for maintaining the structural and functional integrity of the isolated nuclei (Loureiro et al., 2021).

In conclusion, the carefully formulated plant nuclei isolation buffer, incorporating these components, ensured the successful extraction of intact nuclei for flow cytometric analysis. Each component contributed to the overall efficacy of the buffer, resulting in high-quality data with minimal interference from contaminants. The optimized composition of the buffer enhanced the accuracy and reliability of GS estimation through flow cytometry in different types of plants.

4.3.3. Choice of reference standards and standardization methods

The choice of reference standards for flow cytometric analysis involved a meticulous selection process, considering several key factors to ensure accuracy and reliability. Those factors included genome stability, availability of the plant, GS range, ease of propagation and maintenance, and consistency across laboratories (Temsch et al., 2022). Based on GS plants have already been classified into different categories: very small (<1.33 pg), small (>1.33 pg and <3.48 pg), intermediate (>3.48 pg and <14 pg), and large (>14 pg and above) (Michael, 2014). In this study, the reference standards *V. radiata* (2C = 1pg), *O. sative* (2C = 1.01 pg), and *R. communis* (2C = 1.3 pg), were well characterized and accepted reference standards for GS estimation of small GS plants. For estimation of medium sized genomes *S. lycopersicum* (2C = 1.96 pg), and *P. pinnata* (2C = 2.66 pg) were used. The *P. pinnata* is a well characterized reference standard with consistent GS reported from our laboratory (Ramesh et al., 2014). The *Z. mays* (2C = 5.43 pg) was used to provide a reference standard for species with larger GS and greater complexity (Temsch et al., 2022).

To ensure accurate measurements in flow cytometry, it is crucial to achieve a linear response from the instrument and the modern instruments generally offer satisfactory linearity (Temsch et al., 2022). Suda and Leitch (2010) recommend positioning the DNA peak of the target species between the G1/G0 and G2 peaks of the standard, with no more than a four-fold difference in GS to address both minimum and maximum difference between the peaks. A threefold difference was suggested by Sliwinska et al. (2021) emphasizing the importance of minimizing discrepancies, which is a significant restriction in commonly used linear scale, as the use of logarithmic scale is discouraged for GS measurements (Temsch et al., 2021). Overlapping histogram peaks can lead to inaccuracies in determining the DNA content of the sample, affecting the precision of GS estimates. This can result in misinterpretations of ploidy levels and hinder the identification of subtle variations in DNA content. The proximity of peaks may make it challenging to distinguish between different cell populations or phases of the cell cycle, limiting the resolution of the analysis. Additionally, close peaks can increase the likelihood of measurement errors, impacting the reliability and reproducibility of flow cytometry data **(Fig. 4.8)**.

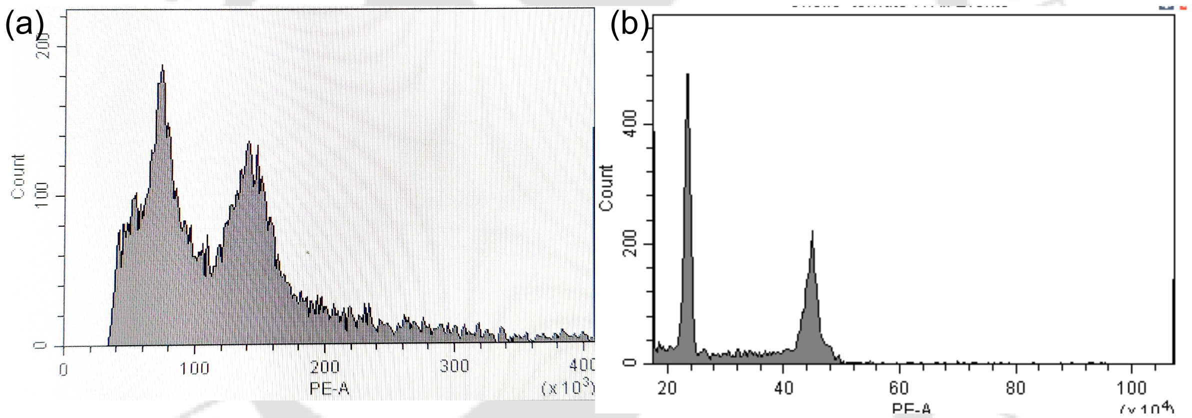


Figure 4.8: Internal standardization G0/G1 peaks of standard and sample (a) Overlapping peaks (b) Clear peaks

The rationale behind the selection of these standard plants lied in the need for a diverse set of references that cover a wide range of GS commonly encountered in plant biology. The small GS standards *V. radiata*, *O. sativa*, and *R. communis* offered stability and precision for smaller genomes. Whereas, *S. lycopersicum* and *P. pinnata*, with their intermediate GS, provided a bridge between smaller and larger genomes, aiding in accurate calibration for a majority of plant species. *Zea mays*, with its larger GS, extended the calibration range to include species with more complex genomes (Temsch et al., 2022). The combination of these plants as reference standards facilitates inter-laboratory comparability and enhances the robustness of flow cytometric analyses across different plant species (Zonneveld, 2021).

In this study three distinct standardization methods were considered: external standardization, pseudo-internal standardization, and internal standardization (Dolezel et al., 2007; Temsch et al., 2022). External standardization involved isolating, staining, and analyzing nuclei separately, leading to potential variations in nuclear fluorescence intensity due to intrinsic process and instrument drifts (Skaptsov et al., 2017). Pseudo-internal standardization prepared nuclei separately for the unknown sample and reference standard, merging them later. While suitable for certain situations, this method was deemed imprecise due to potential influences on sample and standard nuclei in different ways (Temsch et al., 2022). On other hand, the internal standardization, where nuclei of the unknown sample and reference standard were isolated, stained, and measured simultaneously, emerged as the most reliable method (Ramesh et al., 2014). This approach ensured identical treatment conditions for both types of nuclei throughout the entire sample preparation and measurement process. Internal standardization was identified as the only acceptable method for accurate genome size measurement, hence used for all the FCM analysis in this study (Temsch et al., 2022).

The comparative analysis of standardization methods underscored the superiority of internal standardization in flow cytometric GS measurement. Unlike external and pseudo-internal standardization, internal standardization minimizes variations caused by process-related and instrument-related variations (Dolezel et al., 2007). This standardization is essential for accurate measurements of GS, ploidy detection, and DNA base composition analysis, providing a robust foundation for reliable flow cytometric data (Vrana et al., 2014). The variation caused by external and pseudo-internal standardization methods, has already been proved in multiple cases including *P. sativum* and *S. cereal* making them unsuitable for GS estimation (Temsch et al., 2021). Whereas, the ability of internal standardization to mitigate variations introduced by sample and instruments made it the preferred choice for reliable data. In conclusion, the careful selection of reference standards and the adoption of a fluorescence intensity-based standardization method collectively contribute to the reliability and applicability of flow cytometric analysis in plant biology research.

4.3.4. Flow cytometric analysis

The flow cytometric analysis demands meticulous calibration of instrument settings to ensure optimal performance and data accuracy. This involved adjusting the photomultiplier tube voltages, detector gains, and instrument sensitivity settings. Calibration was performed using calibration beads with known fluorescence intensities to standardize the instrument response. To maintain data quality and reproducibility, stringent quality control measures were implemented throughout the experimental process. Regular checks and calibrations were performed using reference standards to monitor instrument stability and performance.

Various types of scatter plots and histograms were generated during the flow cytometric analysis of cell cycle. These included (i) FSC vs SSC: to determine and segregate nuclei based on their size and granularity, (ii) Fluorescence Area vs Width: to visualize doublet and unrelated nuclei, (iii) Fluorescence vs FCS: was particularly helpful in visualizing and removing debris and autofluorescence and (iv) Fluorescence vs Count histograms: to obtain exact peak positions and determination of fluorescence intensity values following proper gating (Choudhury et al., 2014) **(Fig. 4.9).** The histogram plot of DNA content was instrumental in determining the G1/G0 phase of the cell cycle. Peaks corresponding to different cell cycle phases were discerned, and the G1/G0 peak was accurately identified for subsequent analysis of DNA content and cell cycle progression, which is a standard process in plant flow cytometry (Koutecky et al., 2023). Gating was employed to selectively analyze specific populations within the sample. Gate settings were established based on scatter plots to exclude debris and doublets, ensuring that the analysis focused on single, intact cells. Gating parameters were consistently applied across all samples to maintain uniformity (Wang et al., 2015). Autofluorescence, a potential source of background noise, was eliminated or reduced through careful sample preparation, spectral compensation and instrument calibration. Control samples lacking fluorochrome labeling were used to establish baseline autofluorescence levels, and this information was subtracted from experimental samples to enhance signal specificity (Dolezel et al., 2007).

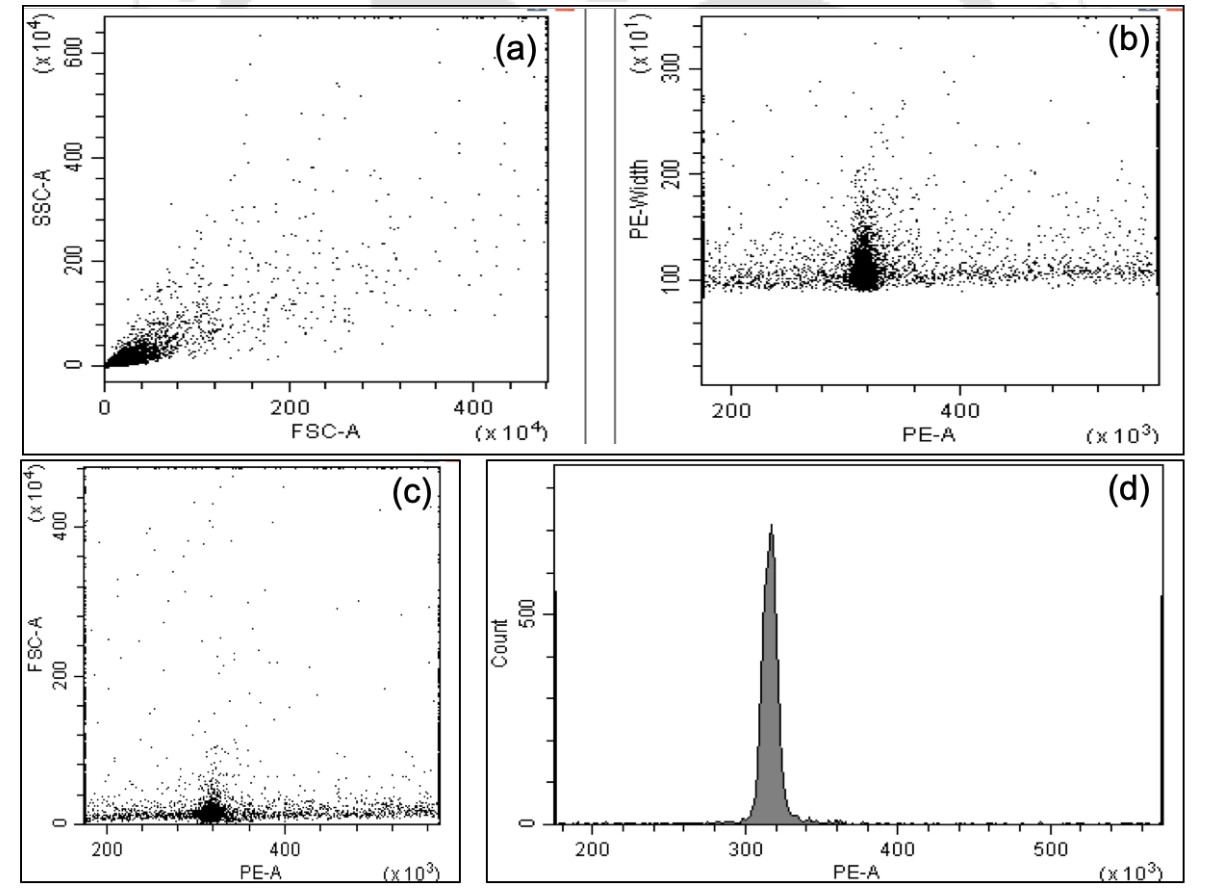


Figure 4.9: Various scatter plots analyzed through flow cytometric analysis (a) FSC vs SSC, (b) Fluorescence intensity vs Width, (c) Fluorescence intensity vs FSC, and (d) Histogram

Although, mean fluorescence intensity has been used in numerous cases for GS estimation in plants, we have used the median fluorescence intensity as a robust measure of central tendency for the fluorescence distribution within the analyzed population, which was reported to be more stable (Giorgi et al., 2013; Sadhu et al., 2016) . It provided a representative value that was less influenced by outliers, offering a reliable indicator of the overall fluorescence intensity. Other statistical parameters, including CV and SD were calculated to characterize the variability of the data. CV reflected the dispersion of data points, and results with CV less than 5% were accepted for further analysis (Temsch et al., 2021).

The comprehensive calibration of instrument settings, rigorous quality control measures, and strategic data analysis, including gating and elimination of autofluorescence, collectively ensured the reliability and accuracy of the flow cytometric analysis. The use of histogram plots facilitated the identification of cell cycle phases, with a focus on the G1/G0 peak for detailed DNA content analysis. Gating strategies were pivotal in excluding unwanted events, contributing to the specificity of the analysis (Sliwinska et al., 2022). The effective elimination of autofluorescence reduced background noise, enhancing the signal-to-noise ratio and improving the accuracy of fluorescence measurements. The utilization of statistical parameters, including median values and CV, provided a robust quantitative assessment of the fluorescence distribution (Koutecky et al., 2023). The employed flow cytometric analysis methodology, encompassing careful calibration, quality control, and strategic data analysis, ensured the generation of reliable and biologically meaningful results. The incorporation of statistical parameters further enhanced the interpretation of fluorescence data, contributing to the overall robustness of the experimental approach (Temsch et al., 2022).

4.3.5. Genome size estimation

The GS (2C DNA) estimation of 110 vascular plants belonging to 53 families from the IIT Guwahati campus, employing flow cytometry, revealed a fascinating spectrum of genomic diversity within the plant species. GS of highest number of species was estimated from Fabaceae (18), followed by Amaranthaceae, Asteraceae, and Moraceae each with seven species. The analyses produced a comprehensive dataset detailing the DNA content of each plant, allowing for a meticulous examination of GS across various families and ecological niches. The results unveiled notable variations in GS, with some species exhibiting larger genomes associated with potential adaptive advantages, while others showcased more compact genomes (Bock et al., 2014). The range of GS varied from 0.43 pg in *Calotropis procera* (Ascelpediaeae) to 42.5 pg in *Sagittaria sagittifolia* (Alismatacreae), which underscored the dynamic nature of plant genomics, highlighting the intricate interplay of factors influencing the evolution and adaptation of these vascular plants within the campus ecosystem. The observed variations in GS provide a captivating platform for scientific inquiry. The correlation analyses among GS, haploid chromosome number, life span, and plant height shed light on the complex relationships inherent in plant biology.

4.3.5.1. *GS estimation of herbs*

The exploration of GS across 47 herbaceous plant species representing 28 families was performed through flow cytometry (**Fig. 4.10**), a striking spectrum of variation unfolded ranging from 0.45 pg in *Typha angustifolia* (Typhaceae) to 42.5 pg in *S. sagittifolia* (Alismatacreae) with average GS of 5.856 pg. Rest of the herbaceous species were excluded on the basis of various factors including sample availability, unavailability of suitable reference standards, already available genomic information etc. Notably, 29 species underwent GS estimation for the first time, unraveling new insights into the genomic dimensions of these plant species. Among these, six species exhibited notable variation in GS, while 12 species displayed a remarkable stability in their genomic content (**Annexure-IIa).** For GS estimation of herbaceous plants *P. pinnata* and *R communis* were used as reference standard for 13 species each, *Z. mays* for 14 species, *S. lycopersicum* for 6 species and Rice was used for only for *S. acuta.*

GS estimation was accomplished for a singular representative in 22 families. Name to few among these were *Centella asiatica* (2C = 1.3 pg) from Apiaceae, *Cannabis indica* (2C = 2.05 pg) from Cannabaceae, *Cleome rutidosperma* (2C = 1 pg) from Cleomaceae, and *Sida acuta* (2C = 1.23 pg) from Malvaceae. Furthermore, GS estimation extended to two species each in Crassulaceae and Poaceae, three species in Araceae, and four species in Solanaceae. Amaranthaceae and Asteraceae underwent comprehensive GS quantification in herbaceous plants, encompassing seven species each. Intriguingly, variations in GS were observed in species such as *Curcuma zedoaria* (1.82 pg), deviating from its previously reported size of 3.26 pg (Islam, et al., 2007). Similar variations were noted in *Pennisetum purpureum* (Poaceae), *Ageratum conyzoides* (Asteraceae), and *Amaranthus spinosus* (Amaranthaceae), adding a layer of complexity to the genomic dynamics of these herbaceous plants (Bennett et al., 1997; Animasaun et al., 2019; Waquar et al., 2023). Based on the GS, highest number of species exhibited GS in medium range (18 spp.), followed by small GS plants (14 spp.), and eleven herbs exhibited very small genomes, whereas the least number of species (04) exhibited large genome size. In this analysis CV was found to between 0.02% in *A. viridis* to 4.8% in *A. philoxeroides*, which lies within the accepted range below 5% in all cases.

This comprehensive exploration of GS across herbaceous plants revealed a remarkable span of variation, illustrating the genomic diversity inherent in these plant populations. Among the analyzed herbaceous plants GS of six species exhibited substantial variation to reported GS, while 12 species demonstrated stable genomic content, which is a common phenomenon observed on land plants (Bennetzen et al., 2005; Lee and Kim, 2014). The GS variability within these herbaceous plants emphasizes the intricate nature of genomic dynamics, wherein certain species exhibit notable plasticity to counter environmental stress, while others maintained remarkable stability (Scholes et al., 2015). The systematic GS estimation also encompassed representatives from 22 herbaceous plant families, providing valuable insights into the genomic landscape of various taxa.

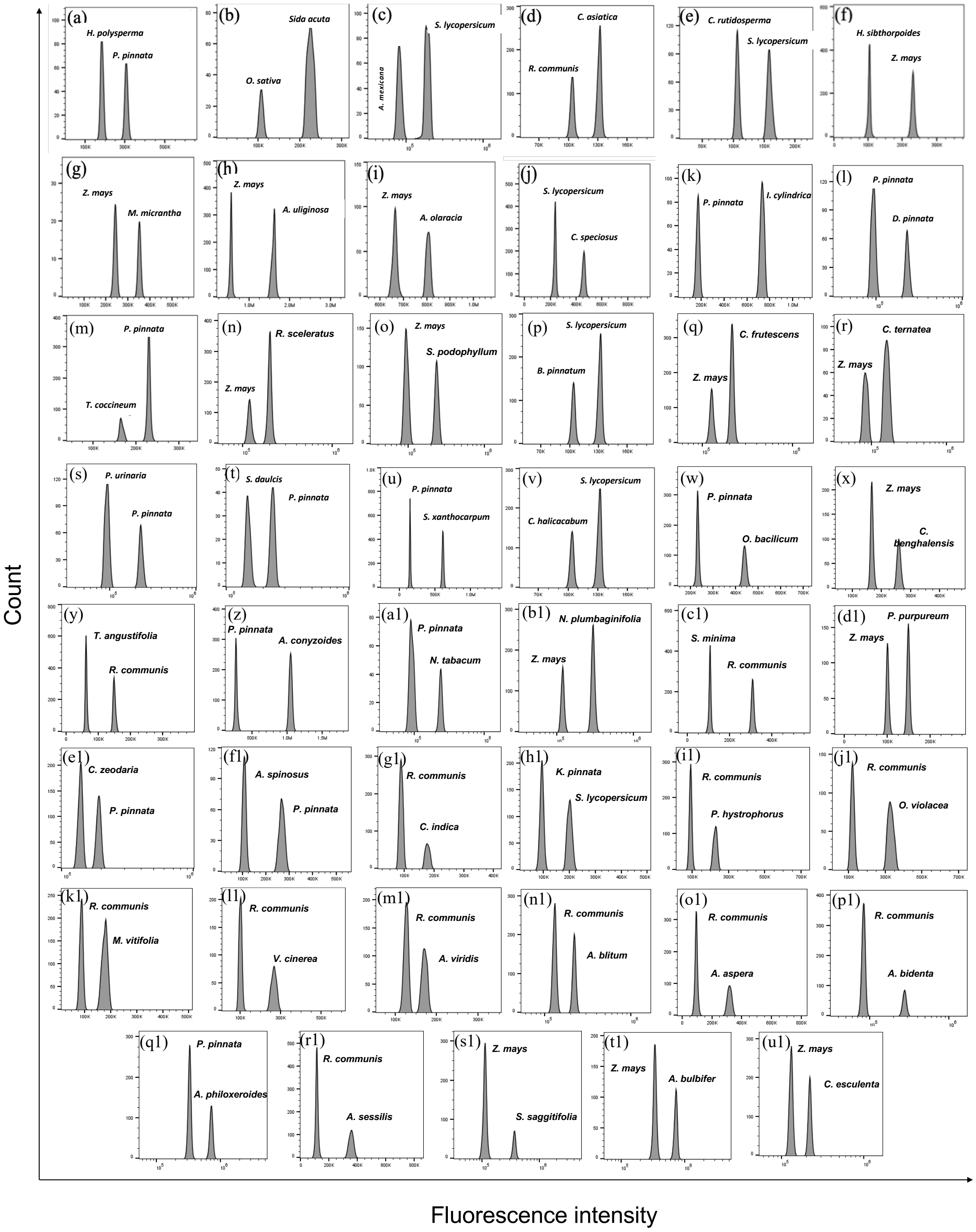


Figure 4.10: Internal standardization histograms of GS estimation from 47 herbaceous plants

4.3.5.2. *GS estimation of shrubs*

The GS (2C) of 12 shrub species spanning 10 diverse families was quantified through histogram analysis in flow cytometry. This provided crucial genetic information, unveiling a spectrum of variation in GS of shrubs (**Fig. 4.11**). The exclusion of the remaining species from GS estimation was based on prioritization aligned with research objectives, resource constraints, and the need to ensure representative sampling while considering logistical and practical considerations. A set of five reference standards were used based on suitability of analysis, *P. pinnata* was used for maximum five species, followed by *S. lycopersicum* for three species, *V. radiata* for two species (*C. procera* and *T. divaricate*) and both Rice and Maize for single species each. In this analysis *C. procera* (Ascelpediaceae) exhibited the smallest GS at 0.43 pg, succeeded by *Cascabela thevetia* from Apocynaceae with 0.63 pg. Conversely, *Codiaenum* sp. (Euphorbiaceae) showcased the largest GS at 15.41 pg, followed by *Yucca gigantea* (12 pg) and *Rotheca serreta* (11.42 pg) from Aparagaceae and Lamiaceae, respectively (**Annexure-IIb).** Among the analyzed species, four shrubs each displayed GS in the medium and small range respectively, and three shrubs exhibited very small genomes, while the single species *Codiaenum* sp. demonstrated large genome sizes. Throughout this analysis, the CV ranged from 2% in *Y. gigantea* to 4.8% in *U. lobata*, all falling within the accepted range below 5% in all cases.

Notably, *Nerium oleander* (Apocynaceae) displayed a remarkably small GS of 0.76 pg, while all other species exhibited sizes surpassing 1 pg. The average GS among shrubs stood at 5.64 pg, akin to herbs, with five species exceeding this average. Eight species underwent GS estimation for the first time, revealing genetic information in *Bougainvillea glabra* (8.2 pg), *Gloriosa superba* (8.96 pg), and *Y. gigantea* (12 pg). *Clerodendrum glandulosum*, in contrast, exhibited a larger GS (1.68 pg) than previously reported (1 pg). Each family was represented by a single species, except Apocynaceae, which included three species in this study.

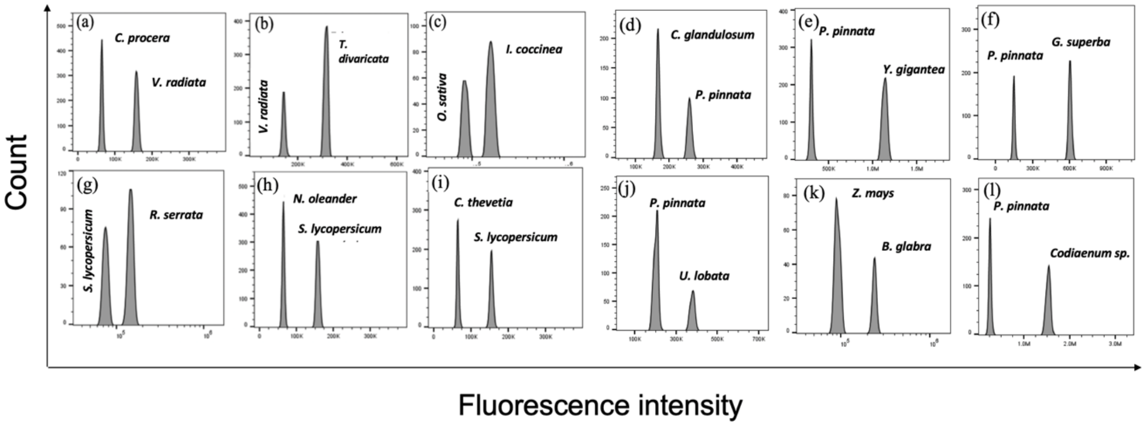


Figure 4.11: Histograms of internal standardization for GS estimation of 12 shrub species obtained through flow cytometric analysis

The GS assessment conducted on 12 shrub species provided valuable insights into the genomic dynamics within shrubs. The observed spectrum of variation, ranging from 0.43 pg in *C. procera* to a substantial 15.41 pg in *Codiaenum* sp., underscores the inherent diversity in GS across shrub species. This variation is exemplified by the smaller GS observed in *C. thevetia* (Apocynaceae) and the larger sizes in *Y. gigantea* (Aparagaceae) and *R, serreta* (Lamiaceae). The substantial variation in range of GS, as evidenced by these examples, highlights the complex genetic architecture within shrub species, potentially influenced by ecological, evolutionary, and adaptive factors as reported in Malvaceae and Aizoaceae plants (Lattier et al., 2019; Powell et al., 2020). Notably, *N. oleander* (Apocynaceae) stands out with a remarkably small GS of 0.76 pg, with potential advantages in adaptation, efficiency and genome stability (Suda et al., 2015).

The majority of the assessed species exhibited GS exceeding 1 pg, emphasizing the prevalence of larger genomes in shrubs as reported in Cupressaceae and other angiosperm families (Beaulieu et al., 2010; Farhat et al., 2019). The average GS of 5.64 pg among shrubs, akin to that of herbs, signifies a general genomic consistency within these plants. GS for eight species was estimated for the first time, expands our genetic understanding of these species. Additionally, the observed deviation in GS for *C. glandulosum*, exceeding the previously reported value, points to the dynamic nature of genomic information in shrubs (Albach and Greilhuber, 2004). The consistency in standard deviation across cases, with *U. lobate* recording the highest at 4.6%, suggests overall reliability in the GS estimation methodology. Overall, these findings contribute to our understanding of the genomic intricacies in shrub species, paving the way for further investigations into the underlying mechanisms governing GS variation in this diverse botanical group.

4.3.5.3. *GS estimation of trees*

Utilizing flow cytometric histogram analysis, the GS of 51 tree species spanning 22 families was quantified (**Fig. 4.12**). Certain tree species were excluded from GS estimation due to logistical constraints, prioritization based on research objectives, already available genomic information, and unavailability of suitable standards. The representation across families varied, with fourteen families featuring a solitary species, five families comprising two species each, Myrtaceae contributing three, Moraceae with seven, and an extensive seventeen species from Fabaceae. Estimated GS ranged from 0.67 pg in *Aegle marmelos* (Rutaceae) to 7.4 pg in *Terminalia chebula* (Combretaceae), with an average of 1.97 pg among trees, a notable reduction compared to herbs and shrubs. Seventeen species exceeded the average GS, while the majority fell below it (**Annexure-IIc**). In the studied species, the maximum of 23 species exhibited GS in small range (<3.48 pg), followed by 22 species of very small GS plants (<1.33 pg), and only seven species of medium GS plants (<14 pg) were observed, while no large genome sized (>14 pg) trees were reported (Michael., 2014). Notably, the CV spanned from 0.01% in *F. religiosa* to 4.8% in *E. stricta* throughout the analysis, consistently falling within the acceptable range of below 5% for all instances.

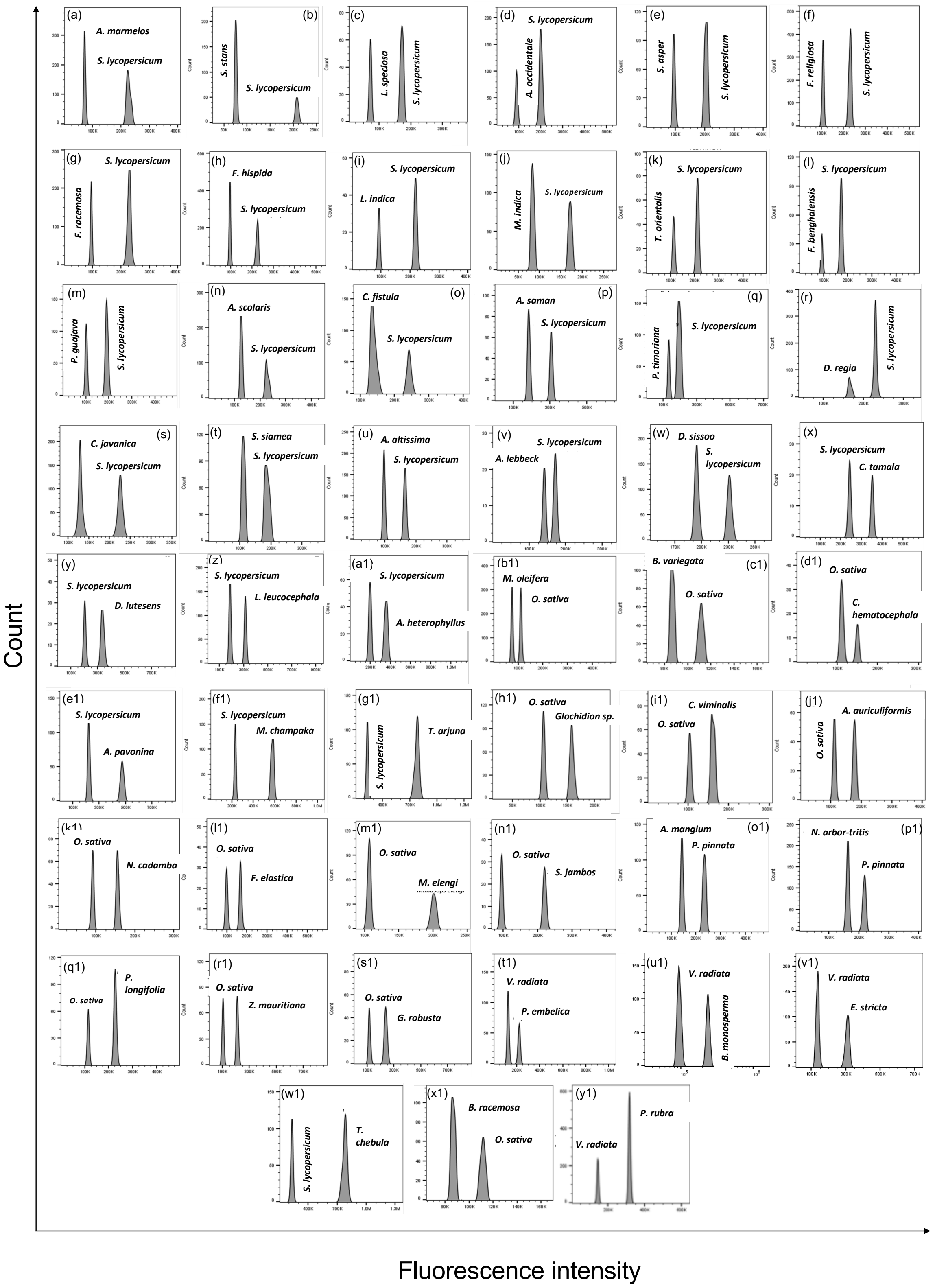


Figure 4.12: Flow cytometric histograms of internal standardization for GS estimation of 51 tree species

Notably, 12 species exhibited GS values below 1 pg, including four from Moraceae and two from Fabaceae. Twenty-one species underwent GS estimation for the first time, from rest of the plants slight deviations observed in *T. arjuna* (7.1 pg) and *Erythrina stricta* (2.2 pg) from reported values (Ohri, 1996; Ohri et al., 2004). Major deviations were identified in 28 species, with the most significant in *Adenanthera pavonina*, presenting an estimated GS (4.58 pg) nearly four times higher than the reported GS (1.36 pg) (Ohri, 2004). Analysis of 17 Fabaceae plants revealed significant GS variation, ranging from 0.77 pg in *Bauhinia variegata* to 4.58 pg in *A. pavonina*. Average GS in Moraceae plants was comparatively smaller than other studied families, ranging from 0.87 pg in *Ficus racemosa* to 3.79 pg in *Artocarpus heterophyllus*. Overall, tree GS was observed to be smaller than that of herbs and shrubs.

The GS estimation across families revealed a varied distribution, with some families represented by single species, while others contribute multiple representatives. Notably, the Fabaceae family was represented by maximum of 17 species, showcasing substantial representation in the campus. The reduced average GS among the trees may reflect the diverse evolutionary strategies employed by trees, with potential implications for their adaptation, growth, and ecological success as reported by Chen et al (2014). The lower average GS in some families, such as Moraceae (ranging from 0.87 pg to 3.79 pg), and the extensive representation of species in Fabaceae, may indicate unique genomic characteristics within these lineages. The observation that 12 species exhibit GS values below 1 pg, including representatives from Moraceae and Fabaceae, highlights the existence of compact genomes in certain tree species (Pellicer et al., 2018). These compact genomes may confer advantages such as energy efficiency and genomic stability (Suda et al., 2015).

The deviations in GS values for some species underscored the importance of validating and updating genomic information. Major discrepancies, as seen in *A. pavonina*, where the estimated GS is nearly four times higher than reported values, emphasize the need for rigorous methodology and continuous refinement of genomic databases (Leitch et al., 2019). Comparative analysis across families reveals intriguing patterns, such as the smaller average GS in Moraceae compared to other families. This discrepancy prompts further investigation into the ecological and evolutionary implications of GS variation within and between families (Valles et al., 2013). In conclusion, the comprehensive analysis of GS in a diverse array of tree species sheds light on the intricate genomic landscape of the plant kingdom. These findings contribute to our understanding of genomic diversity, adaptation, and evolution in trees, with potential applications in areas such as conservation, and sustainable resource management (Suda et al., 2015).

4.3.6. Genome size variation

GS variation in plants exhibits a complex spatial and familial dynamic and frequently observed in angiosperms. Across geographical locations, plants often demonstrate considerable variation in GS, a phenomenon attributed to environmental pressures, adaptive evolution, and localized genetic diversity (Lee and Kim, 2014). This variation is not only evident among different species but also within families, highlighting the intricate interplay of genetic factors. Within a given family, variation in GS was observed in current study, suggesting the role of evolutionary forces, ecological niches, and specific genetic adaptations in shaping the diversification of GS among closely related species (Michael, 2014). This intrafamilial diversity underscores the need for comprehensive studies to unravel the underlying mechanisms governing genome size variations, integrating insights from both ecological and genetic perspectives (Greilhuber et al., 2005). Understanding the spatial and familial nuances of GS variation is crucial for elucidating the adaptive strategies and evolutionary processes that shape plant genomes across diverse landscapes.

In this study of GS estimation across 52 plant species, encompassing 18 herbs, 4 shrubs, and 30 trees, notable findings emerged through flow cytometry analysis. Among the studied species, 16 exhibited consistent GS with no observed variation, comprising 3 shrubs, a single tree, and 12 herbaceous plants. This uniformity in GS across certain species suggests a high degree of genomic stability within these particular plant groups as supported by Schubert and Vu (2016). In contrast, the remaining 36 plants displayed varying degrees of GS variation compared to reported values (**Fig. 4.13**). The extent of variation spanned from 0.01 pg in *T. angustifolia* to 6.3 pg in *C. javanica*. Among herbs, the maximum GS variation was noted in *A. conyzoides* (5.5 pg), highlighting the genomic plasticity within this herbaceous group, similar result was also reported in willow herbs by Abbasi-Karin et al. (2022). In contrast *T. angustifolia* exhibited the minimum variation (0.01 pg) suggesting a more conserved GS in particular species.

Among trees the highest variation was observed in *C. javanica* (6.3 pg), and least in *E. stricta* (0.08 pg) suggesting the ability to adopt environmental adversities, as reported in Pines and Fagaceae plants (Grotkopp et al., 2004; Chen et al., 2014). The average GS variation among the studied species was computed at 1.06 pg, providing a quantitative measure of the degree of genomic evolution or flexibility within plants, as observed in Maize (Diez et al., 2013). These findings underscore the dynamic nature of GS in plants, emphasizing the need for precise and comprehensive re-evaluations to capture the inherent variability within diverse botanical species (Moeglein et al., 2020), which in turn can have significant implications in various fields, including plant biology, ecology, and genomics. Understanding the factors contributing to GS variation can enhance our comprehension of plant evolution, adaptation, and ecological interactions.

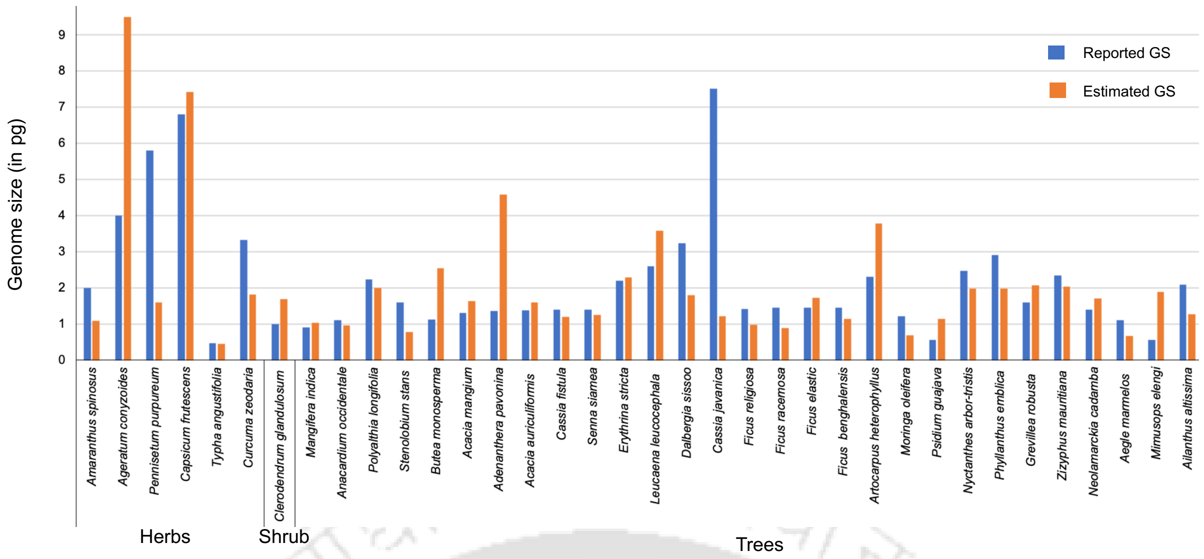


Figure 4.13: Variation in genome size observed in herbs, shrub, and trees

The comprehensive GS estimation from 110 plants has provided valuable insights into the extent of variation within families, illuminating the diversity of genomic content across different plant types. The study revealed that, in the case of 35 families with a single representative species, no GS variation was observed, emphasizing the genetic homogeneity within these families. However, within families featuring multiple species, substantial GS variation was evident, highlighting the complex genomic dynamics inherent in closely related plant species.

In this study, Anacardiaceae family, representing two species (*A. occidentale* and *M. indica*), exemplified GS variation with GS of 0.95 pg and 1.03 pg, respectively. This intraspecific variation underscores the unique genomic characteristics even within a single botanical family, emphasizing the need for a nuanced understanding of genomic diversity as suggested by Zerey-Belaskri et al. (2018). Similarly, the Amaranthaceae family, represented by seven herbaceous plants, exhibited a range of GS values. Notable differences were observed in GS of different species like *A. spinosus* (1.08 pg), *A. viridis* (1.41 pg), and *A. philoxeroides* (5.76 pg), illustrating significant genomic diversity even among closely related herbaceous species. This variation may have implications for the adaptive strategies of these plants in different ecological niches (Vit et al., 2016). The Apocynaceae family, accommodating both shrubs and trees, demonstrated a wide range of GS values. For instance, *C. thevetia* displayed a GS of 0.63 pg, while *Alstonia scholaris* exhibited a GS of 1.19 pg. These variations within the same family suggest potential adaptations to diverse environmental conditions and ecological roles (Santos et al., 2020).

In the Fabaceae family, which featured 17 tree species and a single herb, a large range of GS was observed, from 4.58 pg in *A. pavonina* to 11.08 pg in *C. ternatea*. This substantial intraspecific variation highlights the genomic plasticity within the Fabaceae family, possibly reflecting different evolutionary strategies among its members (Macas et al., 2015). Similarly, the Moraceae family, with seven tree species, exhibited GS variations ranging from 0.875 pg in *Ficus racemosa* to 3.788 pg in *A. heterophyllus*. This variation may be indicative of diverse ecological adaptations within the family, contributing to the overall success and persistence of Moraceae species in different environments (Souza et al., 2020). A comparable pattern was observed in the Asteraceae family, featuring seven herbaceous plants with GS varying from 2.53 pg (*P. hysterophorus*) to 15.72 pg in *A. ulignosa*. The notable GS differences within this family highlight the considerable genomic diversity that may play a role in the ecological niche differentiation and adaptive strategies of these plants (Valles et al., 2013).

The observed GS variation across a diverse array of plant families underscores the nuanced and intricate genomic dynamics inherent in the plant kingdom. The absence of GS variation within families featuring a single representative species accentuates the genetic homogeneity characterizing such lineages. However, contrasting scenarios emerge within families housing multiple species, revealing substantial GS variations. The intraspecific GS differences gave an idea about the unique genomic characteristics even among closely related species noted in families like Anacardiaceae, Amaranthaceae, Apocynaceae, Fabaceae, Moraceae, and Asteraceae. This diversity hints at potential adaptive strategies shaped by environmental conditions and ecological roles. The wide variation in GS within families such as Fabaceae suggests genomic plasticity, reflecting diverse evolutionary strategies among its members. Similarly, GS variations in the Moraceae and Asteraceae families indicate ecological adaptations contributing to the overall success and persistence of species in different environments. Overall, these findings illuminate the rich genomic tapestry of plants, suggesting that GS variation plays a pivotal role in shaping their ecological niches, adaptive strategies, and evolutionary trajectories.

4.3.7. Correlation between GS to various plant traits

The correlation between GS and various plant traits, including haploid chromosome number (n), plant life span (t), and plant height (h), provides valuable insights into the intricate relationships governing plant biology (Knight and Beaulieu, 2008; Lipnerova et al., 2013). Through statistical analysis, correlations have been studied offering a glimpse into the underlying mechanisms shaping the diversity of plant genomes. In some cases, a negative correlation might suggest that smaller GS are associated with longer life spans or increased plant height, while a positive correlation could imply the opposite. These correlations open avenues for further exploration, inviting researchers to delve into the molecular intricacies that govern GS variations and their implications on plant traits (Greilhuber and Leitch, 2012).

4.3.7.1. *Linear correlation*

The Pearson's correlation coefficients calculated for GS (2C), haploid chromosome number (n), life span, and plant height in herbs, shrubs, and trees reveal interesting insights into the relationships among these traits. For herbs, there is a very weak negative correlation between GS and haploid chromosome number (r = -0.0767), indicating a slight tendency for smaller genomes in species with a higher chromosome count (**Table 4.3**) (Sedgwick, 2012). Which was found to be valid in case of *Cyperus iria* (1.6 pg), *C. rotundus* (1 pg), and *C. rostrata* (0.74 pg) having 56, 54, and 38 haploid chromosomes each. No significant correlations were observed between GS and life span or height in herbs. This lack of correlation suggests that, in herbaceous plants, GS may not be a decisive factor in determining life span or height.

In shrubs, a weak positive correlation similar to herbs was found between GS and haploid chromosome number (r = 0.1784), suggesting a tendency for larger genomes in species with more chromosomes. In *Polygala arvensis* (0.86 pg), *C. glandulosum* (1.68 pg) and *D. indica* (1.02 pg) with haploid chromosome numbers 24, 26 qnd 27 respectively this correlation was found to be significant. Interestingly, there is a weak negative correlation observed between GS and life span (r = -0.0745) as well as GS and plant height (r = -0.1082), both of which was observed in *G. superba* (8.96 pg), *L. camara* (5.5 pg), and *E. pulcherrima* (2.6 pg) to be decreasing with increase in GS. These correlations imply that, in shrubs, larger genomes may be associated with shorter life spans and taller heights.

For trees, a moderate positive correlation between GS and haploid chromosome number (r = 0.1576) was observed, indicating a trend for larger genomes in species with more chromosomes. Which is found to be satisfactory in *L. leucocephala* (3.58 pg), *A. heterophyllus* (3.78 pg), and *Z. jujuba* (2.03 pg) with 52, 28 and 24 haploid chromosomes respectively. Additionally, contrast to shrubs a moderate positive correlation was observed between GS and life span (r = 0.1662), as found in *M. indica* (9.2 pg) living more than 100 years, *C. nucifera* (5.6 pg) with a life span of about 60 years and *M. champaka* (5.4 pg) with average life span of 80 years, suggested trees with larger genomes may have longer life span. A strong positive correlation between GS and plant height (r = 0.3473) was established, implying taller trees tend to have larger genomes. Which is found to be accurate in case of *M. indica* (9.2 pg), *T. chebul* (7.4 pg),*T. arjuna* (7.1 pg), *C. nucifera* (5.6 pg) and *M. champaka* (5.4 pg) each having a height of more than 80 feet.

Overall, these correlations suggest that the relationships among GS, haploid chromosome number (n), life span, and plant height vary across herbaceous, shrub, and tree life forms. While herbaceous plants show minimal associations between these traits, shrubs and trees exhibit more noticeable correlations, indicating potential evolutionary and ecological influences on these botanical characteristics within different plant forms.

Table 4.3: Linear correlation test between GS and genomic as well as morphological traits through Pearson's R value analysis (* Statistical significance ±0.11)

Pearson's R value		**Genome size (2C)**	**Haploid chr. no (n)**	**Life span (t)**	**Height (h)**
Herbs	Genome size (2C)	1			
	Haploid chromosome no (n)	-0.076707757	1		
	Life span	-0.012060561	0.00967099	1	
	Height	-0.082522279	-0.049533408	0.143644688	1
Shrubs	Genome size (2C)	1			
	Haploid chromosome no (n)	0.178358837	1		
	Life span	-0.07447762	0.10616896	1	
	Height	-0.108175596	0.16209491	0.441964728	1
Trees	Genome size (2C)	1			
	Haploid chromosome no (n)	0.157610743	1		
	Life span	0.166244584	0.029629756	1	
	Height	0.347253229	-0.013424173	0.389771708	1

4.3.7.2. *Regression analysis*

The multiple regression analysis results for herbs, shrubs, and trees provide insights into the relationships between multiple variables (Beaulieu et al., 2007). In herbs, the multiple R value is 0.116, indicating a weak positive correlation between the predictors (n, t, and h) and the response variable (GS). However, the low R Square value (0.013) suggests that only a small proportion of the variability in the response variable is explained by the predictors (**Table 4.4**). The negative adjusted R square (-0.019) indicates that the model may not be a good fit, and the standard error is relatively high at 5.326 (**Fig. 4.14**). For shrubs, the multiple R value is 0.229, indicating a moderate positive correlation between the predictors and the response variable. The R square value (0.053) suggests that approximately 5.3% of the variability in the response variable is explained by the predictors. The negative adjusted R square (-0.045) indicates potential overfitting, and the standard error is 3.037.

In trees, the multiple R value is 0.384, indicating a relatively strong positive correlation between the predictors and the response variable (Basak et al., 2019). The higher R square value (0.148) suggests that approximately 14.8% of the variability in the response variable is explained by the predictors. The positive adjusted R square (0.110) indicates a better fit compared to herbs and shrubs, and the standard error is the lowest at 1.623. Further, the coefficient analysis for herbs, shrubs and trees gave coefficients for each predictors to correlate with the response variable, which was used to obtain a general equation in each group (**Eq. 1-3).**

The differing regression results across herbs, shrubs, and trees highlight the varying degrees of predictability in these plant forms. The weaker models for herbs and shrubs suggest that the predictors may have limited explanatory power in these groups, while the stronger model for trees implies a more robust relationship between the predictors and the response variable. These findings emphasize the importance of considering plant life forms separately in regression analyses to capture the nuanced relationships within each category.

Table 4.4: Regression analysis of trees, herbs and shrubs

Regression Statistics			
	Herbs	Shrubs	Trees
Multiple R	0.116	0.229	0.384
R Square	0.013	0.053	0.148
Adjusted R Square	-0.019	-0.045	0.110
Standard Error	5.326	3.037	1.623
Observations	95	33	71
		Coefficients	
Intercept	5.771	2.002	0.472
Haploid Chromosome no (n)	-0.039	0.112	0.036
Life span (t)	0.002	-0.008	0.001
Height (h)	-0.095	-0.037	0.017

The equations obtained from regression analysis relating predictors and response variable are as follows

For herbs:

$Y= -0.039n + 0.002t - 0.095h + 5.7$ ……………………..……………………………….(Eq.1)

For shrubs:

$Y= 0.112n - 0.008t - 0.037h + 2.002$ ……………………..……………………………….(Eq.2)

For trees:

$Y= 0.036n + 0.001t + 0.017h + 0.472$ …………………………………………………..….(Eq.3)

Where:

Y= Genome size

n= Haploid chromosome number

t= Lifespan of the plant

h= Plant height

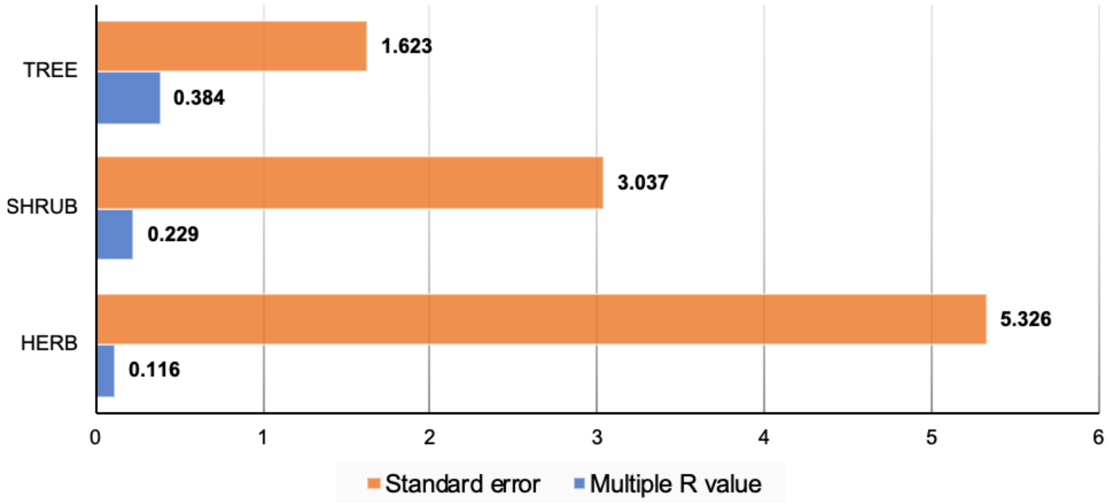


Figure 4.14: Regression analysis results for herbs, shrubs and trees

4.3.7.3. *Artificial Neural Network (ANN) model*

In the artificial neural network (ANN) modeling assessing the relationships between GS and various botanical traits—haploid chromosome number (n), life span (t), and plant height (h)—distinct outcomes were observed for herbs, shrubs, and trees (Patil et al., 2023). For the herbs, the neural network yielded an R value of 0.52944, with the training set achieving 0.69844 and the testing set at 0.4295, showcasing a moderate predictive performance. In the case of shrubs, the best R value was obtained with 84 hidden neurons, where the training set reached 0.829, the testing set 0.7357, and an overall R of 0.724, indicating a strong predictive capacity for this plant group. For trees, the optimal R value occurred at 29 hidden neurons, with the training set reaching 0.6938, testing set at 0.4556, and overall R at 0.553, suggesting a reasonable predictive accuracy. Notably, the overall dataset, encompassing all plant forms, displayed a training R of 0.343, testing R of 0.405, and an overall R of 0.297 at 43 hidden neurons, indicating a modest predictive capability (**Fig. 4.15**).

These results suggest that the neural network models exhibit varying degrees of predictive performance across different plant forms. While the models for shrubs demonstrate strong predictive accuracy, the herbs and trees models show moderate predictive capabilities. The overall dataset presents a more modest performance, indicative of the complex and multifaceted nature of the relationships between GS and botanical traits across diverse plant forms. Further refinement and optimization of the neural network models may be warranted to enhance predictive accuracy and gain deeper insights into the intricate relationships within these plant datasets.

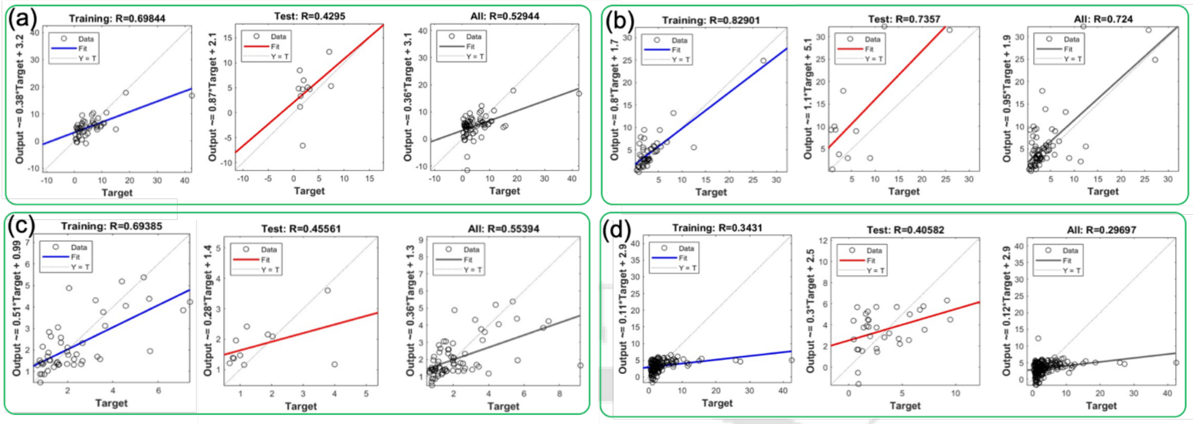


Figure 4.15: ANN models for (a) herbs, (b) shrubs, (c) trees and (d) all plants

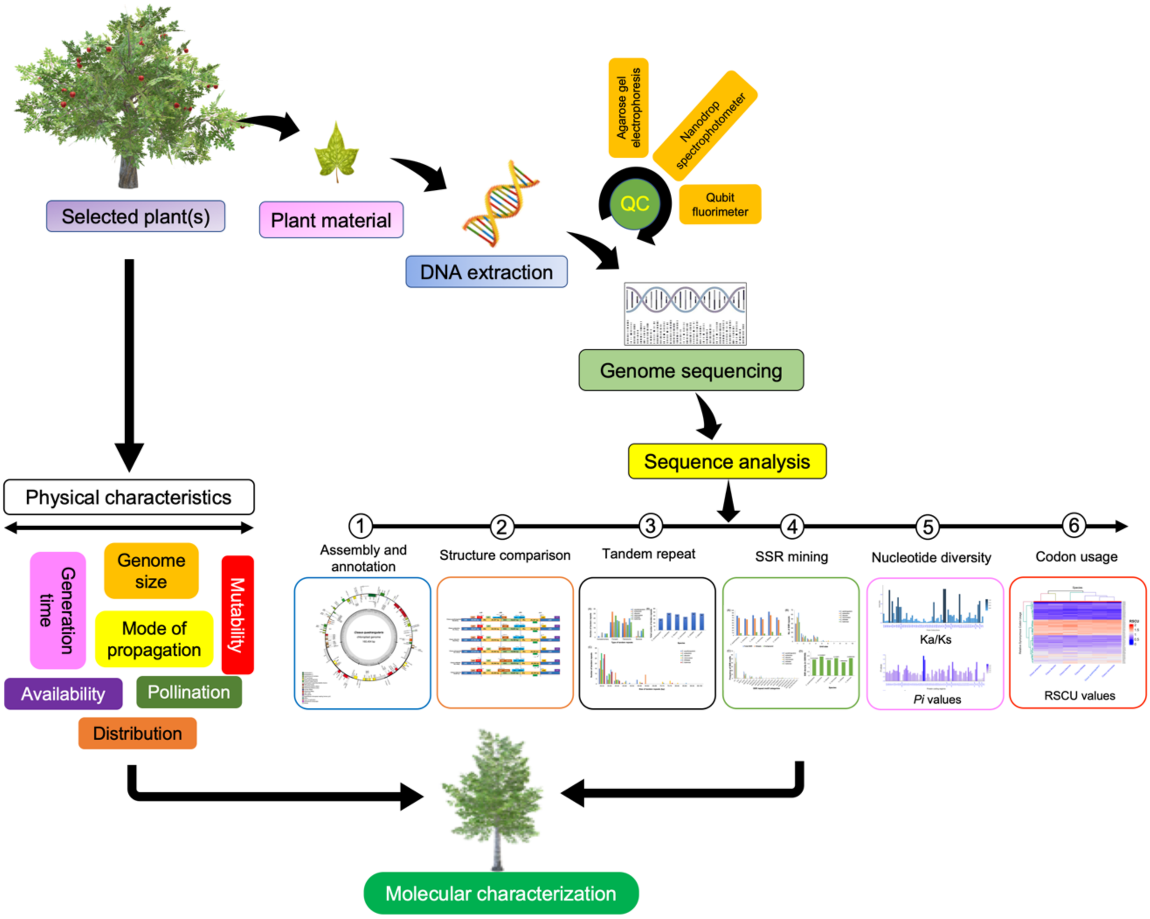
4.3.8. Conclusion

In conclusion, the GS estimation of 110 plants spanning herbs, shrubs, and trees has unveiled a rich tapestry of genomic diversity across various plant families. The meticulous flow cytometric analyses have provided crucial data, highlighting not only the range of GS within individual species but also shedding light on variations from reported values along with methodological developments for better analysis. The extensive exploration of GS within families has been particularly illuminating, with certain families displaying remarkable consistency, while others showcase significant intrafamilial diversity. This intricate mosaic of GS variation reflects the complex interplay of evolutionary forces and environmental adaptations that shape plant genomes. The statistical models provided valuable insights on the degree and direction of relationships between GS and other morpho-genomic traits. Particularly, Linear correlation test was helpful in understanding bilateral relationships among GS and other variables, whereas Regression analysis uncovered the multilateral relationships. The ANN was helpful in capturing subtle dependencies that conventionally overlooked. The statistical models contributed a layer of depth to understand the multifaceted connections between genomic and morphological traits.

Looking forward, the future of GS research in plants hold promising avenues for further exploration. Continued advancements in high-throughput sequencing technologies and refined cytometric methodologies offer the potential for more comprehensive and detailed analyses. Harnessing these tools can enable researchers to delve deeper into the molecular mechanisms governing GS dynamics, unraveling the factors contributing to variability within families and across diverse plant taxa. Moreover, the correlation analyses between GS and key plant traits, such as haploid chromosome number, life span, and plant height, open doors to a more holistic understanding of plant biology. Ultimately, these endeavors contribute to a broader comprehension of plant genomic diversity, with implications for both basic scientific knowledge and applied research in agriculture, horticulture, agro-forestry, and conservation.

Chapter 5

Plastome Mining of Selected Small Genome Sized Plants



CHAPTER 5: PLASTOME MINING

5.1. Introduction

The study of plastomes, or chloroplast genomes, has become increasingly pivotal in plant genomics, providing valuable insights into evolutionary processes, genetic diversity, and adaptive mechanisms (Olejniczak et al., 2016). The chloroplast, a semi-autonomous organelle in plant cells, plays a crucial role in photosynthesis and other essential metabolic processes (Kirchhoff, 2019). As advancements in genomics continue to unfold, we have undertaken a comprehensive investigation into the plastomes of five small genome-sized plants namely *Cissus quadrangularis* (1.41 pg)*, Ficus hispida* (0.9 pg)*, Sierrula rutidosperma* (1.0 pg)*, Streblus asper* (0.97 pg)*, and Thevetia peruviana* (0.63 pg). This multifaceted study involves chloroplast genome assembly, annotation, structural comparison, and other important features (Dobrogojski et al., 2020). The aim is to unravel the intricacies of these plant plastomes, shedding light on their unique genomic landscapes and evolutionary dynamics.

5.1.1. Significance of plastome mining

Understanding the plastome is pivotal not only for unraveling the molecular basis of plant biology but also for applications in agriculture, ecology, and conservation (Kirchhoff, 2019). The chloroplast genome, being maternally inherited in most angiosperms, serves as a powerful tool for phylogenetic studies, population genetics, and forensic analysis (Yao et al., 2015). Moreover, the conserved nature of certain chloroplast regions, combined with the highly variable nature of others, makes plastome sequences invaluable for resolving evolutionary relationships among plants (Zhai et al., 2019). Plastome mining has emerged as a specialized approach to extracting meaningful information from these genomes. This involves a comprehensive analysis of structural features, repetitive elements, nucleotide substitution rates, and codon usage biases (Song et al., 2022). By delving into these aspects, researchers can gain insights into the adaptive strategies of plants, identify potential markers for species discrimination, and discern the forces driving chloroplast genome evolution (Gao et al., 2019; Mo et al., 2020).

5.1.2. Rationale for studying small genome-sized plants

The choice of small genome-sized plants for this plastome mining study is not arbitrary; rather, it is grounded in the scientific rationale of maximizing insights while minimizing genomic complexity. Small genome-sized plants often exhibit unique genomic features, adaptive traits, and specialized ecological niches (Pyšek et al., 2023). By focusing on these species, we aim to unravel specific adaptations encoded in their plastomes, contributing to our broader understanding of plant evolution and ecology.

5.1.3. Chloroplast genes and their applications

Chloroplast genes are integral to the process of photosynthesis, playing a pivotal role in capturing and converting light energy into chemical energy (Jensen and Leister, 2014). Beyond their fundamental biological importance, chloroplast genes have practical applications, particularly in the realm of genetic engineering and crop improvement (Gao et al., 2019). Scientists have been exploring the manipulation of chloroplast genes to enhance photosynthetic efficiency, improve stress tolerance, and increase crop yields. Such advancements hold promise for addressing global food security challenges by creating more resilient and productive plant varieties (Bansal and Saha, 2012).

Furthermore, chloroplast genes are gaining attention in the field of phylogenetics and molecular taxonomy. These genes offer a unique source of genetic information that can be utilized for species identification and evolutionary studies (Zhang et al., 2016). The use of chloroplast DNA barcoding has emerged as a valuable tool for accurate and rapid species identification, aiding in biodiversity conservation efforts (Shneyer and Rodionov, 2019; Senapati et al., 2022). This application has far-reaching implications, enabling researchers to monitor and protect endangered plant species, study ecological relationships, and assess the impact of environmental changes on plant populations (Gao et al., 2019; Shang et al., 2022). As our understanding of chloroplast genes deepens, their potential applications continue to expand, contributing to advancements in both basic plant biology and applied agricultural and environmental sciences.

5.1.4. Overview of methodology

The methodology employed in this study encompasses a series of intricate steps designed to comprehensively analyze the plastomes of the selected plant species. Chloroplast genome assembly serves as the initial step, involving the reconstruction of complete plastome sequences from high-throughput sequencing data. Subsequent annotation allows for the identification of key genes, non-coding regions, and structural features essential for understanding plastome functionality (Shelke et al., 2022).

Structural comparison of cp genome of five selected plants with their close relatives aims to elucidate variations and conserved regions across the plastomes (Park and Oh, 2020). Tandem and SSR repeat mining provide insights into the repetitive elements shaping plastome architecture, while the identification of divergent hotspot regions (*Pi*) offers a glimpse into regions undergoing accelerated evolution (Xiao-Ming et al., 2017). Nucleotide substitution rate analysis, quantified by the Ka/Ks ratio, delves into the selective pressures acting on protein-coding genes within the plastomes (Mo et al., 2020). Lastly, synonymous codon usage bias analysis provides a nuanced understanding of the evolutionary forces influencing codon usage patterns in these plants (Geng et al., 2022).

5.1.5. Anticipated contributions and outcomes

The anticipated contributions of this study are manifold. Firstly, a detailed exploration of the plastomes of five small genome-sized plants will enhance our understanding of the genomic variations and adaptive strategies inherent in these species. The identification of repetitive elements, divergent hotspot regions, and codon usage biases will provide valuable markers for future research on plant evolution, population genetics, and species discrimination.

Furthermore, the comparison of plastomes across these five plant species will shed light on the shared ancestry and evolutionary divergence, contributing to the broader field of plant phylogenomics (Song et al., 2022). The nucleotide substitution rate analysis will uncover genes experiencing positive selection, providing insights into the adaptive evolution of chloroplast-encoded proteins. This study promised to advance our understanding of plant genomics, with implications for diverse fields ranging from evolutionary biology to conservation and beyond. By delving into the intricacies of chloroplast genomes in small genome-sized plants, we aim to unravel the unique stories encoded in these vital organelles, contributing to the ever-expanding tapestry of plant biology.

5.2. Materials and methods

5.2.1. Selection of plants

In this study, five small genome-sized (GS ≤ 1 pg) plants were meticulously chosen from the rich botanical diversity of the IIT Guwahati campus. The selected species included *C. quadrangularis* (0.56 pg), *F. hispida* (0.9 pg), *S. rutidosperma* (1.0 pg), *S. asper* (0.97 pg), and *T. peruviana* (0.63 pg). To ensure taxonomic accuracy, the identification of these plants was carried out employing well-established taxonomic keys and referring to the authoritative "Flora of Assam" by Kanjilal (1934). The sampling process involved collecting specimens directly from their natural habitats on the IIT Guwahati campus, capturing the inherent genetic diversity within these species. To complement the natural habitat sampling, the plants were also cultivated in the controlled environment of the greenhouse at the Department of Biosciences and Bioengineering, IIT Guwahati, Assam, India. This dual-source approach ensured all renewable and accessible source of plant material for ongoing research.

5.2.2. Physical characteristics

Before initiating chloroplast genome sequencing, a meticulous physical characterization of the selected plant species was conducted to provide a comprehensive context for subsequent genomic analyses. Generation time, a critical parameter influencing experimental timelines, was determined through careful observation of life cycle stages and reproductive patterns. Genome size, a pivotal factor in sequencing strategy selection, was assessed using flow cytometry to ensure the compatibility of sequencing platforms with the target genomes. The mode of propagation, whether sexual or asexual, was identified to gauge the potential impact of recombination events on genomic diversity. The availability of plant material was evaluated, considering factors such as geographical distribution and accessibility, to ensure a representative sampling. Additionally, the type of pollination whether self-pollinating or cross-pollinating was determined, as this information aids in understanding genetic diversity patterns (Feuillet et al., 2011). These physical characteristics collectively formed the foundation for a strategic and informed approach to subsequent chloroplast genome sequencing, ensuring that the experimental design aligns with the biological nuances of each plant species under investigation.

5.2.3. Sample collection

The collection of plant samples for chloroplast genome sequencing at IIT Guwahati adhered to stringent protocols, ensuring the acquisition of high-quality material for subsequent genomic analyses. Strategic sampling of plant specimens representing the selected species was conducted within their natural habitats on the campus. Thorough attention was given to selecting healthy, disease-free individuals to capture the dynamic genetic landscape inherent in each species. Detailed ecological information, was meticulously recorded for each specimen, facilitating future ecological and population genetic studies. To mitigate contamination risks, sterile tools were utilized during collection, and samples were promptly placed in individual sterile bags, preserving their integrity. The preference for young and healthy leaves ensured the retrieval of high-quality DNA. Special consideration was given to obtaining sufficient quantity of DNA, crucial for the success of downstream sequencing processes (Godden et al., 2012).

Post-collection, leaves exhibiting optimal health and devoid of any signs of ailments or physical impairment were chosen through meticulous procedures. Subsequently, the samples were swiftly transported to the laboratory, ensuring controlled low temperatures during transit. Upon arrival, the samples underwent immediate flash freezing through immersion in liquid nitrogen, preserving them at an ultra-low temperature of -80°C until DNA extraction. The meticulous methodology employed in sample collection at IIT Guwahati laid a robust groundwork for a comprehensive chloroplast genome sequencing project, ensuring the generation of reliable and scientifically sound genomic data.

5.2.4. Plastome sequencing and mining of molecular information

5.2.4.1. *Genome sequencing and library preparation*

The genomic DNA was used for sequencing of chloroplast genome of selected plants. The quality and quantity of the extracted DNA was determined using Nanodrop spectrophotometer (ND-2000C, ThermoFisher) in Applied Biodiversity Laboratory, BSBE, IITG. The subsequent step encompassed the preparation of genomic libraries, a process facilitated by the employment of the TruSeq sample preparation kit. The resultant libraries were subjected to sequencing procedures on an Illumina NextSeq500 platform for four species excluding *C. quadrangularis*, employing paired-end 2X150 bp PE chemistry with 400–450 bp insert size, which was performed by Eurofins Genomics Pvt. Ltd. (India). Whereas, the *C. quadrangularis* reads were generated on Novogene's Illumina HiSeq X Ten platform as paired-end 2X150 bp sequences with 400–450 bp insert size (Gichuki et al., 2019).

5.2.4.2. *Genome assembly and annotation*

For four of the selected plants raw illumine reads were obtained after sequencing except for *C. quadrangularis*, where we retrieved raw reads (SRR8573652) from the Sequence Read Archive (SRA) database submitted by Gichuki et al. (2019). All the raw Illumina reads were quality inspected using the FastQC (0.11.8) tool before being *de novo* assembled using the NOVOplasty (v4.3.1) assembler with a *rbc*L sequence as a seed (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Using the NCBI BlastN tool, the resulting chloroplast genome assembly sequence was checked against the non-redundant (nr) database. GeSeq tool was employed for initial annotation, followed by a manual curation step in adherence to established protocols (Tillich et al. 2017). The annotation process encompassed the identification of various genetic elements, including protein-coding genes, tRNA and rRNA genes introns and intergenic regions. The circular genome maps were generated utilizing the OGDRAWv1.2 tool (Greiner et al., 2019).

5.2.4.3. *Structural comparison*

In this investigation, we conducted a detailed analysis of the structural characteristics inherent in the chloroplast (cp) genomes of several plant species, comparing them with their closely related counterparts for a comprehensive understanding. For *C. quadrangularis*, six cp genomes of *Cissus* spp. were scrutinized for structural comparisons, including *Cissus antarctica* (NC_061724.1), *Cissus discolor* (NC_061723.1), *Cissus microcarpa* (NC_061722.1), *Cissus trifoliata* (NC_061720.1), *Cissus tuberosa* (NC_061719.1), and *C. quadrangularis* (OP414588). Similarly, within the Cleomaceae family, seven species were comparatively analyzed, encompassing *Thulinella chrysantha* (NC_053524), *Cleomella lutea* (NC_049613), *Coalisina paradoxa* (NC_066812), *Cleome pallida* (NC_054213), *Gynandropsis gynandra* (MW123058), *Tarenaya hassleriana* (NC_034364), and *S. rutidosperma* (OQ220335), shedding light on the structural intricacies of the latter. Additionally, a comparative analysis involving four distinct species from the Moraceae family, namely *Ficus hispida* (OQ220336), *Ficus auriculata* (NC_053837.1), *Ficus racemosa* (KT368151.1), and *Ficus squamosa* (OK077764.1), was performed to elucidate the structural nuances of the *F. hispida* cp genome.

Within the Plumerieae tribe, a comparative plastome analysis was conducted on five species, namely *T. peruviana* (OQ376289), *Cerbera manghas* (MT527963.1), *Plumeria krugii* (MW392093.1), *Plumeria obtusa* (NC_069180.1), and *Plumeria rubra* (MN812495.1). A similar approach was undertaken for *S. asper* within the Moraceae family, involving seven species, including *S. asper* (OQ281590), *Streblus banksii* (MW238798), *Streblus glaber* (MW238799), *Streblus heterophyllus* (MW238800), *Streblus indicus* (MN065161), *Streblus pendulinus* (MW238801), and *Streblus smithii* (MW238802). To facilitate synteny studies of cp genomes, the MAUVE tool (https://darlinglab.org/mauve/mauve.html) was employed. Furthermore, we examined the boundaries between the small single copy (SSC), large single copy (LSC), and the inverted repeat regions (IRa and IRb) among related species. This boundary analysis was facilitated using the IRscope tool (V0.1.R) (Amiryousefi et al., 2018), providing valuable insights into the evolutionary dynamics of genes located in the IR boundary regions.

5.2.4.4. *Tandem and simple sequence repeat (SSR)*

In our investigation of five small genome sized plants, we utilized the REPuter tool (Kurtz et al., 2001) to explore the presence of tandem repeats within their chloroplast (cp) genomes. The identification of forward, reverse, palindromic, and complementary repeats was conducted with specific parameters, including a minimum length of 30 base pairs and a hamming distance of three. The REPuter tool (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) facilitated a comprehensive analysis of the repetitive elements present in the cp genomes of the selected species.

To discern simple sequence repeats (SSRs) within the cp genomes of the five chosen species and their close relatives, the MISA software (v1.0.6) was employed. This analysis aimed to identify mono-, di-, tri-, and other nucleotide repeats, employing specific parameters: 10 repeats for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for each of the following nucleotide categories: pentanucleotides, hexanucleotides, septanucleotides, octanucleotides, nonanucleotides, and decanucleotides. The MISA software, (https://pgrc.ipk-gatersleben.de/misa/misa.html) facilitated a meticulous exploration of the SSR landscape in the chloroplast genomes under investigation. This dual approach provided a comprehensive understanding of both tandem and simple sequence repeats, contributing valuable insights into the repetitive elements present in the chloroplast genomes of the studied plant species.

5.2.4.5. *Genome diversity (Pi)*

In our pursuit of identifying divergent hotspot regions within chloroplast (cp) genomes, a comparative analysis was conducted by juxtaposing genome annotations of the selected five plants with those of closely related species. The annotation of each selected plant served as a reference for this comparative study. To quantify the level of nucleotide diversity within both genic and intergenic regions of the chloroplast genomes, the DnaSP tool (v5.10) (http://www.ub.edu/dnasp/) was employed. This analysis was executed separately for each of the five selected plants, calculating *Pi* values as a measure of nucleotide diversity. The *Pi* values were determined for each individual cp genome, shedding light on the genetic variation present in both coding and non-coding regions. This approach allowed us to discern divergent hotspot regions, providing valuable insights into the evolutionary dynamics and genetic variation within the chloroplast genomes of the studied small genome-sized plants.

5.2.4.6. *Nucleotide substitution rate (Ka/Ks)*

In this study, the evaluation of selection pressure on genes within the chloroplast genomes of five small genome-sized plants was conducted through the nucleotide substitution rate analysis, specifically employing the Ka/Ks ratio. The Ka/Ks ratio serves as a crucial metric, assessing the relative rates of non-synonymous mutations (Ka) to synonymous mutations (Ks) in protein-coding genes. This ratio is indicative of the selective forces acting on specific genes, where a Ka/Ks ratio greater than 1 suggests positive selection, equal to 1 indicates neutral evolution, and less than 1 implies purifying or negative selection. The Ka/Ks ratios were meticulously calculated using the KaKs_calculator 3.0, developed by Zhang (2022). This computational tool facilitated a comprehensive examination of the evolutionary dynamics and selective pressures acting on the chloroplast protein-coding genes of the selected small genome-sized plants, providing valuable insights into the adaptive strategies and functional evolution of these essential genomic elements.

5.2.4.7. *Synonymous codon usage bias*

To investigate synonymous codon usage bias in this study, an analysis was conducted on common protein-coding genes extracted from the chloroplast genomes of the selected five small genome-sized plants. The assessment involved the determination of the average guanine-cytosine (GC) content and codon usage bias for all protein-coding genes. This analysis was carried out through comprehensive examination of nucleotide composition and Relative Synonymous Codon Usage (RSCU). The MEGA X program (https://www.megasoftware.net/) was employed for these analyses (Stecher et al., 2020). The RSCU values provide insights into the preferential usage of synonymous codons, shedding light on the potential evolutionary forces influencing codon usage patterns within the chloroplast genomes of the selected plant species. This analytical approach contributes valuable information to our understanding of the molecular evolution and adaptive strategies encoded in the protein-coding regions of these small genome-sized plants.

5.2.4.8. Phylogenetic analysis

The compared sequences were first aligned using MAUVE to find the locally collinear blocks (LCBs) in HomBlocks pipeline (Bi et al. 2018). Then the model of substitution was figured out with the assistance of the MEGA-X tool (version 10.2.4), and it was done so in accordance with what the Akaike information criterion (AIC) indicated. Finally, the phylogenetic tree was built using the maximum-likelihood method (ML) in MEGA X programme with 500 bootstraps (https://www.megasoftware.net/). The resulting phylogenetic tree was then depicted visually via the iTOL server (https://itol.embl.de/).

5.3. Results and discussion

5.3.1. Physical characterization

The comparative analysis of key physical characteristics among five small GS plant species provided insights into their diverse biological profiles. These characteristics include family association, habit, generation time, pollination methods, availability, distribution, propagation mechanisms, mutability, and conservation status. The species belong to distinct families: Brassicaceae, Vitaceae, Moraceae, Cleomaceae, and Apocynaceae, reflecting their evolutionary and genetic disparities. Growth forms vary from herbs (*C. quadrangularis*, *T. peruviana*) to trees (*F. hispida*, *S. asper*) and shrubs (*S. rutidosperma*), indicating different ecological roles and adaptations. Generation times differ, with *A. thaliana* having a short life cycle of 42 days and *S. rutidosperma* a longer cycle of 120 days. Pollination mechanisms range from self/cross-pollination in several species to wasp-mediated in *F. hispida* and insect-mediated in *T. peruviana*. GS variations range from 0.32 pg in *A. thaliana* to 1 pg in *S. rutidosperma*, indicating significant genetic material differences. Availability and distribution emphasize ecological adaptability, while propagation methods, mutability, and conservation status highlight unique biological attributes (**Table 5.1**).

These differences in characteristics reflect evolutionary adaptations and ecological strategies, showcasing the diverse niches these species inhabit. Variations in family affiliations highlight evolutionary divergence, while differences in habit, generation time, and pollination methods underscore adaptive flexibility and reproductive strategies. Genome size variations are influenced by ecological niches and evolutionary history, with smaller GS potentially conferring energy efficiency. Availability, distribution, propagation methods, mutability, and conservation status contribute to the ecological status and resilience of each species.

Table 5.1: Physical characteristics of selected small genome sized plants compared to the model plant (*A. thaliana*)

S. No	Characteristics	*A. thaliana*	*C. quadrangularis*	*F. hispida*	*S. asper*	*S. rutidosperma*	*T. peruviana*
1	**Family**	Brassicaceae	Vitaceae	Moraceae	Moraceae	Cleomaceae	Apocynaceae
2	**Habit**	Herb	Herb	Tree	Tree	Herb	Shrub
3	**Generation time**	42 days	1-2 years	5-7 years	5-10 years	120 days	5-7 years
4	**Pollination**	Self/Cross	Self/Cross	Cross (Wasp)	Cross (Wind)	Cross (Bees)	Cross/Self (Insect)
5	**GS (2C in pg)**	0.32	0.56	0.9	0.97	1	0.63
6	**Availability**	Available	Ornamental/Available	Wild/Available	Native/Available	Wild/Available	Ornamental/Available
7	**Distribution**	Cosmopolitan	Cosmopolitan	Tropical regions	Cosmopolitan	Cosmopolitan	Cosmopolitan
8	**Propagation**	Seed	Seed/Stem	Seed	Seed	Seed	Seed
9	**Mutability**	High	High	NA	NA	High	Average
10	**Conservation status**	NE	NE	NE	NE	NE	NE

5.3.2. Plastome mining: unveiling molecular insights

Plastome mining has emerged as a powerful tool in unraveling the genetic intricacies of plant species. The chloroplast, a vital organelle involved in photosynthesis, houses a relatively conserved genome, making it an ideal target for molecular characterization (Dobrogojski et al., 2020). The process involves determining the complete nucleotide sequence of the chloroplast genome, enabling researchers to explore its structural organization, gene content, and evolutionary patterns (Nock et al., 2011). It has multifaceted application in various fields such as phylogenetics, population genetics, DNA barcoding and molecular ecology. By comparing plastomes across different plant species, reconstruction of evolutionary relationships can be done, providing valuable insights into the taxonomy and evolutionary history of plants (Dobrogojski et al., 2020). Additionally, plastome data contributes to the development of DNA barcodes, facilitating accurate and efficient species identification, a crucial aspect in biodiversity studies and conservation efforts. Due to conserved nature of cp genomes, the design of specific/universal primers across diverse taxa is possible, which streamlines the sequencing process and facilitates comparative analyses (Li et al., 2015).

Chloroplasts are maternally inherited in plants, providing a unique opportunity for studying maternal lineages and tracing the maternal history of plant populations (Granick, 1961). Plastomes also exhibit structural variations, rearrangements, or insertions, providing uniqueness to specific plants (Olejniczak et al., 2016). The molecular characterization of plants through plastome mining aids in the accurate identification and classification of plant species, crucial for biodiversity studies and conservation planning (Song et al., 2020). It is also, provides insights into plant adaptation and diversification as well as development of molecular markers, enhancing our ability to trace and monitor plant populations in the face of environmental changes and human activities. Its applications extend across diverse fields, offering insights into evolution, taxonomy, and conservation (Zhao et al., 2018; Dobrogojski et al., 2020).

This study unraveled the intricate structural variations and functional nuances embedded within the chloroplast genomes of studied species. This comprehensive analysis spans multiple facets, including genome assembly, annotation, structural comparison, repeat analysis, and investigation of codon usage bias. By employing state-of-the-art methodologies and bioinformatics tools, this study aims to contribute valuable insights into the evolutionary dynamics and functional adaptations encoded within the chloroplast genomes of small genome sized plants, fostering a deeper understanding of their genetic diversity and ecological significance.

5.3.2.1. *Genome assembly and annotation*

Chloroplast genome assembly and annotation represent crucial facets of molecular biology research, offering insights into the genetic architecture of chloroplasts, the cellular organelles responsible for photosynthesis in plant cells (Wanichthanarak et al., 2023). The chloroplast genome, typically a circular DNA molecule, encodes essential genes involved in photosynthetic processes, carbon fixation, and other vital metabolic pathways (Dobrogojski et al., 2020). Assembling and annotating the chloroplast genome involves the intricate task of piecing together its genetic information and identifying specific genes and functional elements. Advances in high-throughput sequencing technologies have significantly accelerated chloroplast genome research, enabling scientists to unravel the complexities of these organelles with unprecedented precision (Tillich et al., 2017). The assembly and annotation of chloroplast genomes not only contribute to our understanding of plant biology but also essential for downstream molecular analysis (Kahraman and Lucas, 2019). In the current research, *de novo* assembly of raw genomic data (Illumina reads) and annotation of genes were accomplished for five plants with small genome sizes, as discussed below.

C. quadrangularis

C. quadrangularis, a perennial plant from the grape family (Vitaceae), is indigenous to the Indian subcontinent, Southeast Asia, and parts of Africa. Recognized by common names such as "Veldt grape" or "Devil's backbone," this succulent plant features quadrangular stems and has a history of traditional use in various medicinal practices majorly as anti-inflammatory and analgesic. The complete cp genome of *C. quadrangularis* was assembled using the NOVOplasty assembler and leveraging a dataset comprising 18,978,233 raw paired-end reads, resulting in a circular genome spanning 160,404 bp (**Fig. 5.1**). The assembled chloroplast genome sequence was deposited in GenBank under the accession number OP414588. The genomic organization adheres to the typical quadripartite structure observed in angiosperm chloroplast genomes, featuring two Inverted Repeat (IR) regions (26,397 bp) separated by the Large Single Copy (LSC) region (88,988 bp) and the Small Single Copy (SSC) region (18,621 bp). The overall GC content in the entire chloroplast genome was calculated at 37.2%, with variations observed in the IR (42.8%), LSC (35.07%), and SSC (31.3%) regions.

Annotation of the chloroplast genome revealed a total of 133 genes, comprising 87 protein-coding genes, 38 tRNA genes, and 8 rRNA genes (**Annexure-IIIa**). Notably, duplications were detected in four rRNA genes, seven tRNA genes, and nine protein-coding genes. Within the IR regions, 18 genes were found, including seven protein-coding genes, four rRNA genes, and seven tRNA genes. Furthermore, seven protein-coding genes (*ndh*B, *rpl*2, *rpl*23, *rps*7, *rps*12, *ycf*2, and *ycf*15), four rRNA genes (rRNA 4.5, rRNA 5, rRNA 16, and rRNA 23), and seven tRNA genes were identified as duplicated in the IRs. Additionally, two genes (*clp*P and *ycf*3) exhibited double introns, while 20 genes featured a single intron.

The circular nature of the *C. quadrangularis* chloroplast genome, exhibiting the typical quadripartite structure consistent with angiosperm chloroplast genomes, aligns with expectations (Shelke and Rangan, 2022). The genome size of ~160 kb falls within the predicted range for angiosperm chloroplast genomes (107–218 kb) (Daniell et al., 2016). Comparative analyses revealed the presence of four additional protein-coding genes (*atp*H, *pet*B, *psb*L, and a duplicated copy of the *ycf*1 gene) that had been overlooked in previous annotations of *Cissus* cp genomes. Intriguingly, the study identified two copies of the *ycf*15 gene in *Cissus* spp. genomes, a rarity in angiosperms, yet observed in members of the Vitaceae family and other plant genera (Jin et al., 2020; Shelke and Rangan, 2022; Wang et al., 2021). Notably, the comprehensive assessment confirmed the absence of any protein-coding gene losses in the *C. quadrangularis* cp genome. The presence of introns, impacting gene expression, was noticed. Additionally, the observation of two copies of the *ycf*15 gene in *Cissus* spp. genomes, a gene typically absent in many angiosperms, offers a novel perspective on the genomic peculiarities within this genus and also found to be present in full copies in the plastome of *Nicotiana*, *Epifagus*, *Cuscuta* etc. (Schmitz-Linneweber et al., 2001).

Figure 5.1: Circular map of *C. quadrangularis* whole cp genome along with gene annotations. Clockwise-transcribed genes are displayed inside the circle, and genes positioned outside are transcribed anti-clockwise. Different colour codes represent genes of different functions

F. hispida

Using the NOVOplasty assembler, we successfully assembled the complete chloroplast genome of *F. hispida*. The total length of the *F. hispida* cp genome was determined to be 160,162 bp (**Fig. 5.2**).

Comparative analysis using BLASTN revealed substantial sequence similarity, ranging from 87% to 99%, when compared to other *Ficus* chloroplast genomes available in the GenBank database. To enhance accessibility and serve as a reference, the entire sequence of the *F. hispida* chloroplast genome has been deposited in the GenBank database under the accession number OQ220336. The structural organization was characterized by two IR regions spanning 25,837 bp each, with the LSC region measuring 88,452 bp and the SSC region measuring 20,036 bp, positioned between the IRs. The overall GC content of the *F. hispida* chloroplast genome was determined to be 35.94%. Notably, this percentage exhibited variation across distinct regions, with the IR region displaying a higher content of 42.65%, while the LSC and SSC regions showed percentages of 33.59% and 28.99%, respectively. This comprehensive analysis offers valuable insights into the genomic structure and characteristics of the *F. hispida* chloroplast genome, contributing to our understanding of the Moraceae family's broader genomic context.

A comprehensive examination of the gene content in the chloroplast genome of *F. hispida* revealed a total of 132 genes. Among these, 87 genes were protein encoding, while the remaining 37 and 8 genes corresponded to tRNA and rRNA respectively (**Annexure-IIIb)**. Notably, the genome exhibited duplications in four rRNA genes, seven tRNA genes, and eight protein-coding genes. A symmetrical distribution of 19 genes was observed in each of the IR regions. These duplicated genes encompassed eight protein-coding genes, four rRNA genes, and seven tRNA genes. Similar to *C. quadrangularis* presence of double introns in two genes (*clp*P and *ycf*3), and single introns in twenty-one genes observed. An interesting observation was the trans-splicing of the *rps*19 gene, present in both the LSC and IR regions. Additionally, the *mat*K gene was identified within the largest intron of the *trn*K-UUU gene.

The successful assembly of the complete cp genome of *F. hispida* provides valuable insights into the genomic characteristics of this species within the Moraceae family. The determined total length of 160,162 bp aligns with typical chloroplast genome sizes observed in angiosperms (Daniell et al., 2016). The employed comparative analysis, specifically using BlastN, revealed a substantial sequence similarity compared to other *Ficus* chloroplast genomes and this level of similarity underscores the conservation of chloroplast genomes within the *Ficus* genus. The structural organization of the *F. hispida* chloroplast genome conforms to the common quadripartite configuration observed in land plants (Wang et al., 2019). The variability in GC content in different regions of the cp genome is a characteristic feature having implications for the stability and functionality of these regions (Guo et al., 2021). The observed duplications in rRNA, tRNA, and protein-coding genes indicate a dynamic evolutionary process, which might involve gene rearrangements or genomic rearrangements over time. Similar findings have been reported in the cp genomes *F. racemosa* and *F. auriculata* where duplications in rRNA, tRNA, and protein-coding genes are recurrent themes in *Ficus* cp genomes, suggesting shared evolutionary patterns within the genus.

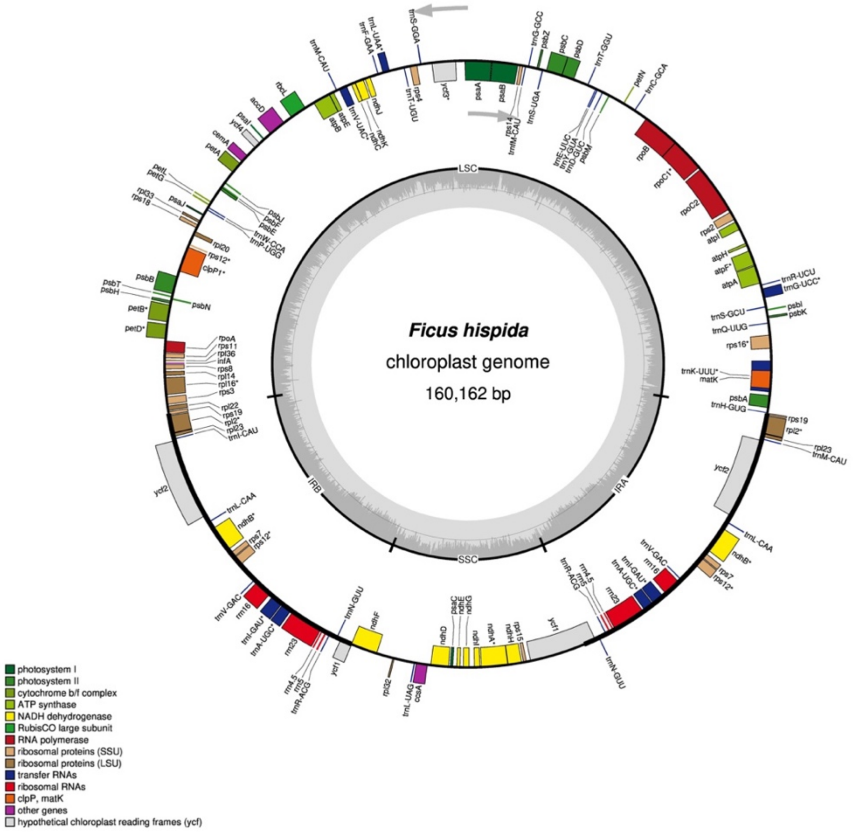


Figure 5.2: A circular map depicting the entire chloroplast genome of *F. hispida* is presented, featuring gene annotations. Genes transcribed in a clockwise direction are illustrated within the circle, while those transcribed in anti-clockwise direction are positioned outside. Various color codes are employed to distinguish genes based on their functions

Furthermore, the presence of introns in certain genes, including double introns in *clp*P and *ycf*3, is consistent with intron patterns observed in other plant chloroplast genomes. The trans-splicing of the *rps*19 gene and the localization of the *mat*K gene within the largest intron of *trn*K-UUU add to the genomic intricacies of *F. hispida*, showcasing features with functional significance (Ruhlman and Jansen, 2014). The assembly and annotation of the *F. hispida* chloroplast genome provide a foundation for understanding the genetic makeup within the Moraceae family. The observed similarities with other *Ficus* spp. and related genera support the notion of conserved genomic features, while the identified variations contribute to our understanding of species-specific adaptations.

S. rutidosperma

S. rutidosperma, commonly known as the fringed spider flower or purple cleome and belonging to the Cleomaceae family, is indigenous to tropical Africa. This invasive herbaceous plant thrives in lowland wet tropical regions of Asia and Australia. Employing the NOVOplasty assembler, a total of 68,555,455 raw paired-end reads were utilized to assemble the complete chloroplast genome of *S. rutidosperma*, yielding a circular genome of 157,073 bp (**Fig. 5.3**). The newly assembled cp genome demonstrated a high degree of similarity (93.39%–98.7%) to other Cleomaceae cp genomes. The entire cp genome sequence has been deposited in GenBank under the accession number OQ220335. The quadripartite organization of the chloroplast genome consists of IR regions (26,083 bp) separated by the LSC region (86,423 bp) and the SSC region (18,484 bp). The overall GC content of the chloroplast genome is 36.03%, with higher percentages observed in the IR region (42.43%), LSC region (33.68%), and SSC region (28.95%).

Similar to the previously assembled *F. hispida*, the chloroplast genome of *S. rutidosperma* was annotated to contain 132 genes, comprising 87 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (**Annexure-IIIc**). Duplications were identified in four rRNA, five tRNA, and eight protein-coding genes. Within each IR region, a total of 18 genes were observed, including seven protein-coding genes, four rRNA genes, and seven tRNA genes. Consistent with previous observations, both *clp*P and *ycf*3 contained double introns, while eighteen genes featured a single intron. The trans-spliced *rps*12 gene was found in both the LSC and IRa regions, resembling the configuration in *F. hispida*.

This study provides valuable insights into the structural arrangement of chloroplast genomes and the extent of gene conservation across seven species of the Cleomaceae family. The identified gene arrangement patterns hold potential implications for understanding plant evolution, phylogenetics, and the development of molecular markers for plant identification and classification. Future research endeavors should delve into exploring the functional significance of these gene arrangements and their potential applications in the realms of plant breeding and biotechnology. Notably, the chloroplast genome of *S. rutidosperma* adheres to the quadripartite and circular pattern characteristic of angiosperm chloroplast genomes, with a length of 157,073 bp falling within the expected range for angiosperms (107–218 kb) (Shelke and Rangan, 2022). Among Cleomaceae genomes, *C. paradoxa* boasts the largest genome size at 159,393 bp, while *C. lutea* has the smallest at 154,124 bp.

In our analysis of the chloroplast genome of *S. rutidosperma*, we observed a shared gene content with other angiosperms and Cleomaceae species, encompassing 37 tRNA genes and 8 rRNA genes (Daniell et al., 2016; Shelke and Rangan, 2022). Notably, the *psb*G gene was absent in six species, including *S. rutidosperma*, but present only in *C. lutea*. Additionally, the role of introns in gene expression was evident, with typical angiosperms containing approximately 23 introns in tRNA and protein-coding genes (Wang et al., 2022). In the *S. rutidosperma* genome, we identified 20 genes with introns. However, we noted the rarity of hypothetical protein-coding genes like *ycf*15 and *ycf*68 in Cleomaceae species, being entirely absent in *S. rutidosperma*.

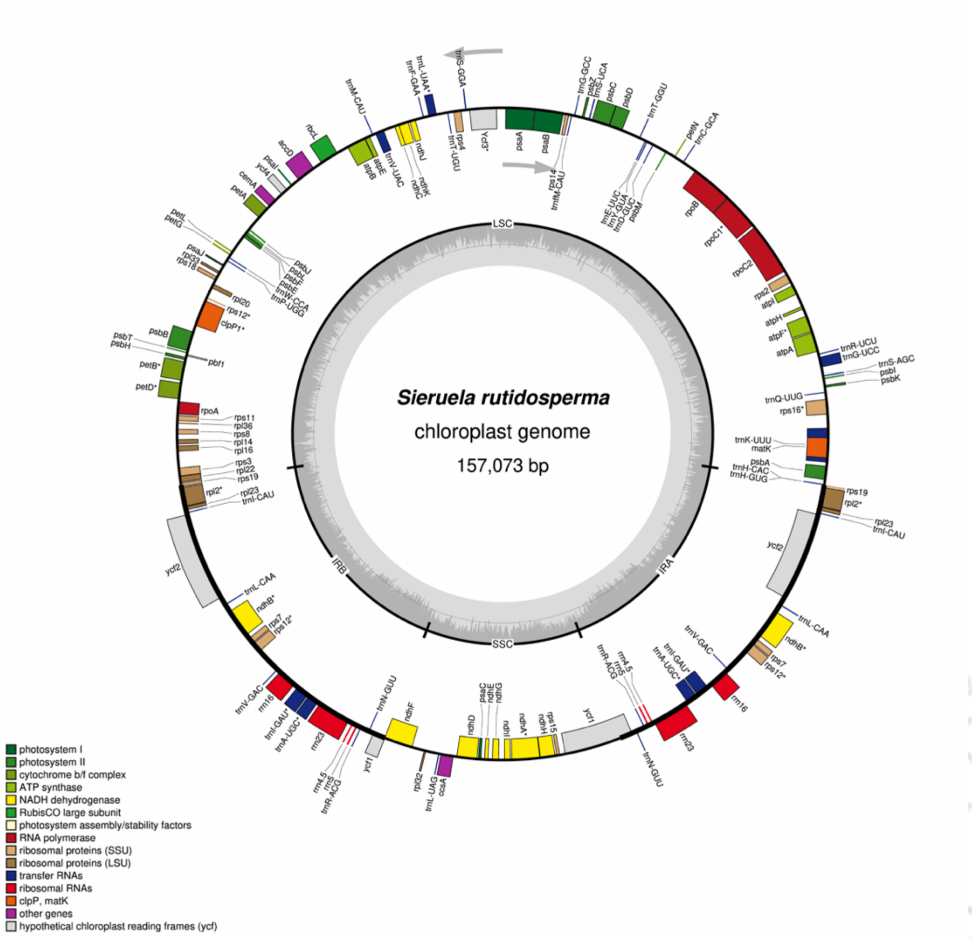


Figure 5.3: A circular map depicting the entire cp genome of *S. rutidosperma* featuring gene annotations. Genes transcribed in a clockwise direction are illustrated within the circle, while those transcribed in an anti-clockwise direction are positioned outside. Color codes are employed to distinguish genes based on their functions

S. asper

Embarking on a journey into the cp genome of *S. asper*, we harnessed the power of NOVOplasty to assemble its complete genetic blueprint, revealing a circular marvel spanning 161,011 base pairs (**Fig. 5.4**). This genome displayed a quadripartite structure, showcasing a LSC region (90,004 bp), a SSC region (19,551 bp), and two IRs measuring 25,728 bp each. The annotation unveiled a rich ensemble of 129 genes within the chloroplast genome of *S. asper*. Among this genomic orchestra, 85 genes were protein coding, while 36 transfer RNA genes, and 8 ribosomal RNA genes were observed (**Annexure-IIId**). Within the IR regions, 18 genes were duplicated, encompassing seven protein coding genes, seven tRNA genes, and four rRNA genes. We uncovered 18 genes housing the introns within the genome while, three coding genes – *ycf*3, *clp*P, and *rps*12 – captivated us with their double-intron.

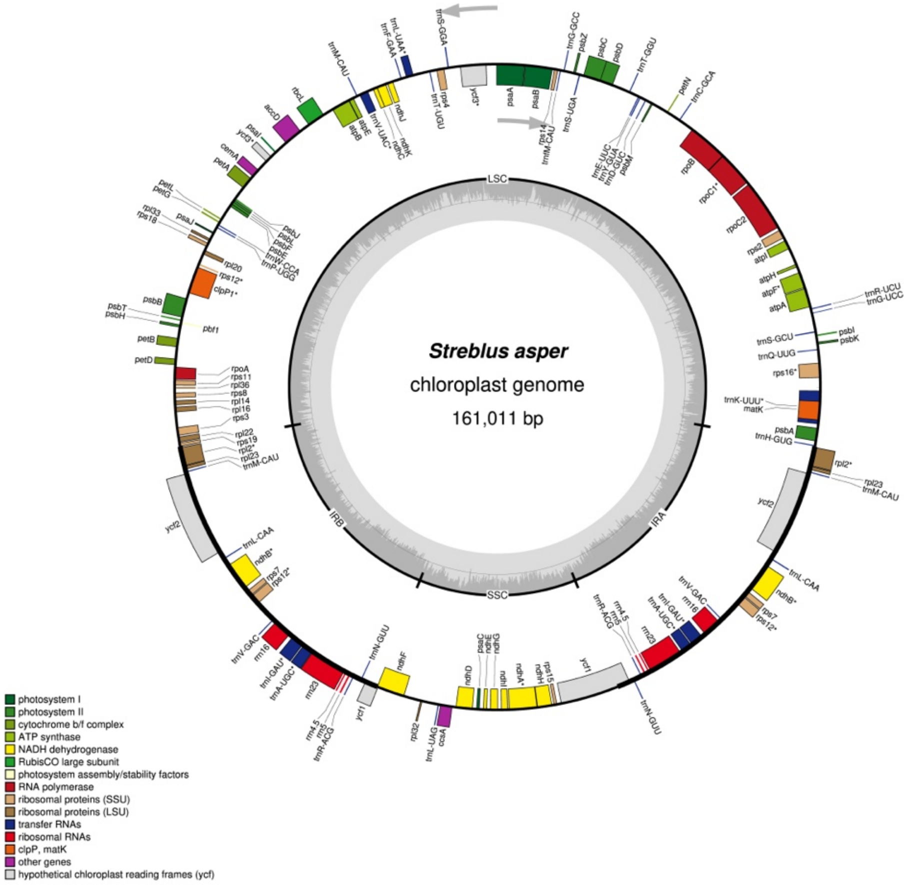


Figure 5.4: Circular map of *S. asper* whole cp genome along with gene annotations. Clockwise-transcribed genes are displayed inside the circle, and genes positioned outside are transcribed anti-clockwise. Different colour codes represent genes of different functions

The assembly and annotation of the complete cp genome of *S. asper* represent a significant contribution to our understanding of the cp genome structure in this species. The determined length of 161,011 bp aligns with the typical size range observed in angiosperms, and the quadripartite organization, consisting of a LSC region, a SSC region, and two IRs, conforms to the conserved structural pattern found in most plant cp genomes (Daniell et al., 2016, Senapati et al., 2023). The presence of 18 duplicated genes within the inverted repeat regions indicates the typical arrangement seen in angiosperm cp genomes (Dobrogojski et al., 2020). This redundancy is known to play a crucial role in maintaining genome stability and ensuring proper functioning of the photosynthetic apparatus (Jin et al., 2020). Comparative analysis with other related species reveals both conservation and divergence in cp genome structure. The size and organization of the chloroplast genome in *S. asper* share similarities with other members of the Moraceae family including *Broussonetia* spp., *Ficus* spp., and *Artocarpus heterophyllus* (Liu et al., 2018; Yang et al., 2022; Zhang et al., 2022). However, variations in gene content, intron presence, and the number of duplicated genes highlight the species-specific characteristics within this taxonomic group. Future research could explore the functional implications of these genomic differences and their potential roles in adaptation and evolution within the Moraceae family. The identified genes, their duplication patterns, and the presence of introns provide valuable information for understanding the genetic basis of traits in *S. asper* and contribute to the broader field of plant genomics as reported in *Ficus* spp. (Liu et al., 2019; Zhang et al., 2022). Further investigations into the functional significance of these genomic features and their correlation with ecological and physiological traits will enhance our understanding of the adaptive strategies within this plant lineage.

T. peruviana

The complete plastid genome of *T. peruviana* was assembled by NOVOplasty, determining its length to be 155,148 bp (**Fig. 5.5**). Comparative sequence analysis with plastid genomes of the Plumerieae tribe indicated a sequence similarity ranging from 89.32% to 99.69%. The plastomes of *T. peruviana* adhered to the canonical circular quadripartite structure, encompassing a LSC region of 84,491 bp, a SSC region of 18,375 bp, and two IRs of 26,141 bp. In total, 133 genes were annotated from cp genome of *T. peruviana*, comprising 87 protein-coding genes (PCGs), 38 transfer RNA genes, and 8 ribosomal RNA genes (**Annexure-IIIe**).

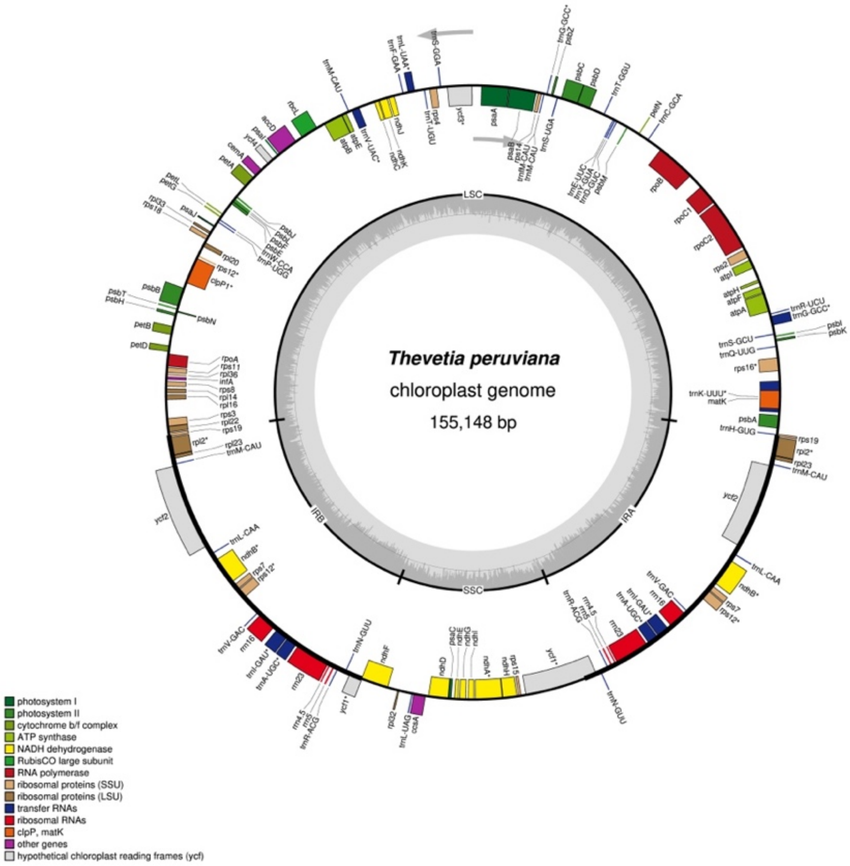


Figure 5.5: Circular map of *T. peruviana* whole cp genome along with gene annotations. Clockwise- and anti-clockwise transcribed genes are displayed inside and outside the circle respectively

Within the IR regions, 17 genes were duplicated, encompassing six PCGs, seven tRNA genes, and four rRNA genes. Furthermore, 18 genes contained introns, with three PCGs (*ycf*3, *clp*P, and *rps*12) harboring two introns each. The plastid genome of *T. peruviana* conforms to the characteristic quadripartite arrangement typical of land plant plastid genomes, as documented in recent studies (Xu et al. 2022; Senapati et al. 2023). Its cp genome size, totaling 155,148 bp, falls within the established range observed in Apocynaceae family, ranging from 150 kb to 176 kb (Liao et al. 2020; Tiernan et al. 2023; Wang et al. 2023). Noteworthy is the high conservation among plastomes within the Apocynaceae family, characterized by consistent genome structures, gene orders, and GC contents, as reported in several recent studies (Xiang et al. 2022; Kim et al. 2023; Tiernan et al. 2023; Wang et al. 2023)

5.3.2.2. *Structural comparison*

Comparing the cp genome structures is necessary to unraveling the intricate details of structural variations among closely related plant species (Wang et a., 2015; Li et al., 2023). The chloroplast genome, a circular DNA molecule, serves as a reservoir of essential genetic information crucial for photosynthesis and various metabolic pathways. By systematically examining the chloroplast genomes of related plant species, researchers can discern variations in terms of gene order, gene content, and the presence of introns. These structural comparisons provide valuable insights into the evolutionary dynamics and divergence within plant lineages (Tang et al., 2004). Additionally, they aid in understanding adaptive processes, as variations in chloroplast genome structure can be indicative of environmental adaptations and evolutionary pressures (Xiong et al., 2021). This comparative analysis not only contributes to our comprehension of plant evolution but also holds significance for applications in taxonomy, phylogenetics, and the broader field of plant biology (Shelke and Rangan, 2022). The structural comparison of cp genomes is being reported for five selected species and their relatives as given.

C. quadrangularis

To elucidate structural variations among various *Cissus* spp., this study engages in a comparative examination of six cp genomes within the genus *Cissus*, namely *C. quadrangularis* (OP414588), *C. antarctica* (NC_061724.1), *C. discolor* (NC_061723.1), *C. microcarpa* (NC_061722.1), *C. trifoliata* (NC_061720.1), and *C. tuberosa* (NC_061719.1). Employing MAUVE alignment, the investigation reveals an absence of significant inversions, except for a distinct 250 bp inversion detected in the LSC region of *C. trifoliata* and *C. tuberosa* (**Fig. 5.6**). Consistently across the six cp genomes, the junction between the IR and SSC regions features two genes, *ycf*1 and *ndh*F. Notably, a distinctive pattern emerges regarding the contraction and expansion of the *ycf*1 gene at the LSC/IRa boundary, demonstrating variability in the number of base pairs (bp) in the SSC and IRa regions, predominantly favoring the SSC region within *Cissus* spp. Specifically, in *Cissus* spp., the *ycf*1 gene spans positions 4447 to 4568 bp in the SSC region and 965 to 1082 bp in the IRa region. The positioning of the *ndh*F gene exhibits variation across species, with *ndh*F located at the IRb and SSC junction in *C. quadrangularis*, *C. tuberosa*, *C. trifoliata*, and *C. microcarpa*, extending 2–38 bp into the IRb region, primarily situated in the SSC region. In contrast, *C. discolor* and *C. antarctica* position *ndh*F entirely within the SSC region.

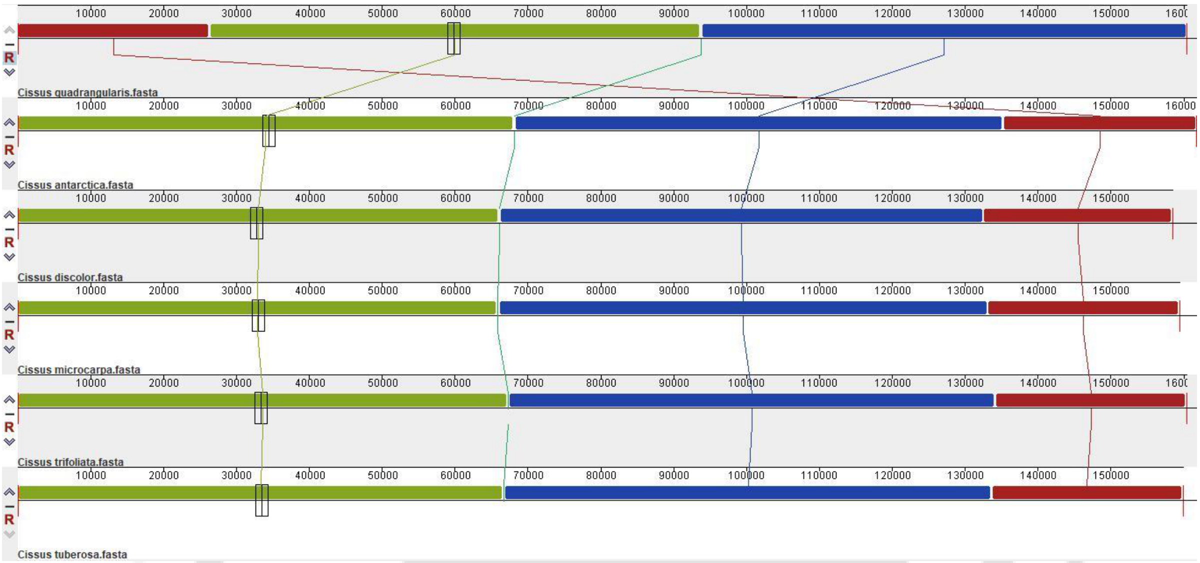


Figure 5.6: MAUVE alignment of cp genomes of six *Cissus* spp. Different colored local collinear blocks represent the gene arrangement in chloroplast genomes

The *rps*19 gene, situated at the LSC/IRb junction, consistently extends 23–45 bp into the IRb region in all species except *C. trifoliata*, where it resides 17 bp inside the LSC region from the border. The complete *rpl*2 gene is consistently found in both the IRa and IRb regions. The *trn*H gene is consistently localized entirely in the LSC region just outside the IRa/LSC border across all investigated species. These findings underscore the shared conserved nature of cp genomes among the six *Cissus* spp., with only minor variations observed at the junctions (**Fig. 5.7**).

Numerous studies have extensively characterized cp genome inversions, highlighting their pivotal role in genome rearrangement (Kim et al., 2005; Jiang et al., 2022; Jin et al., 2020). In our investigation, an inversion approximately 250 bp in length was discerned in both *C. trifoliata* and *C. tuberosa*. A parallel, albeit smaller, inversion in an intergenic region had been previously documented in *Mesua ferrea* (Shelke and Rangan, 2022). In contrast, larger inversions, exceeding 50 kb, have been reported in certain angiosperm species (Jin et al., 2020; Tian et al., 2021; Wu et al., 2022). Notably, the inversion observed here appears conserved, occupying the same genomic position between the *pet*A and *psb*J genes in both *C. trifoliata* and *C. tuberosa* genomes. Such conserved inversions play a crucial role in establishing interspecific differentiation and are likely linked to the formation of plant groupings (Kim et al., 2005).

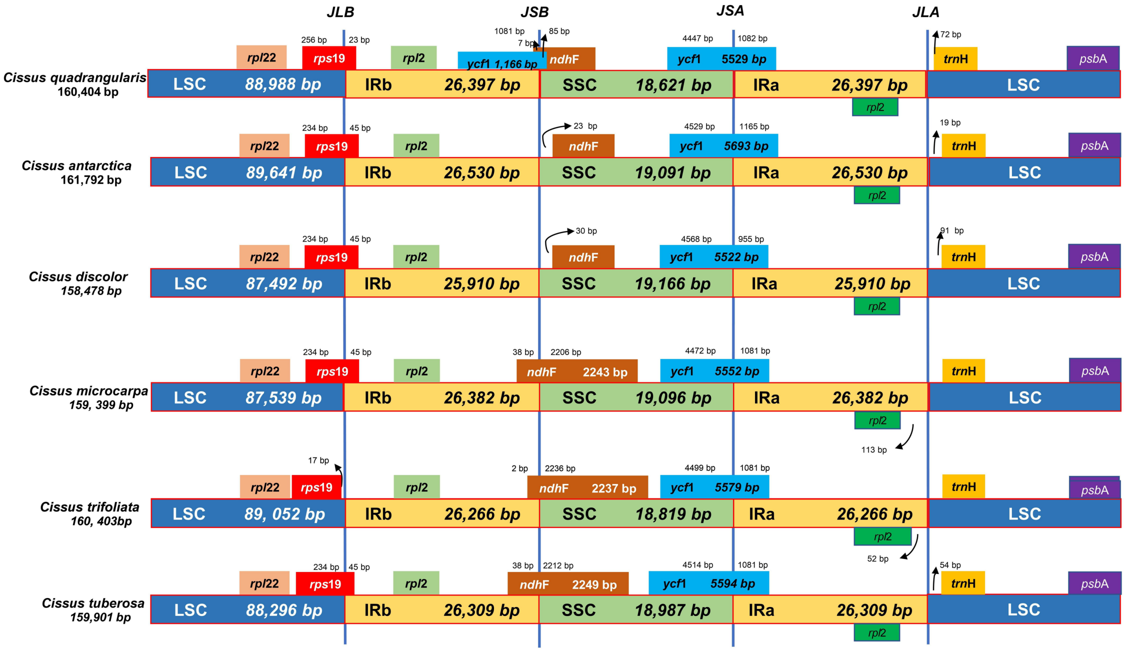


Figure 5.7: Comparison of IR junctions among six *Cissus* spp. Colour boxes present the genes. The genes on the upper track represent the negative strand, whereas those below the track denote the positive strand

The cp genomes of *Cissus* spp. demonstrated identical gene orders, and the sequence identities across compared cp genomes were notably similar among species, indicating substantial conservation (Jin et al., 2020). The expansion and contraction of IR junctions, a common phenomenon in higher plants, contribute to structural changes (Jin et al., 2020). In the comparison of six *Cissus* spp., no evidence of gene loss was observed, except for *ycf*1 gene, which was situated on the IRb/LSC border of *C. quadrangularis*. The high homogeneity observed between LSC/IR and SSC/IR borders in *Cissus* and angiosperm chloroplast genomes, with most of these boundaries located in *rps*19 or *ycf*1, further underscores the conservation within these genomes (Alzahrani et al., 2020; Downie and Jansen, 2015; Shelke and Rangan, 2022).

In *C. quadrangularis*, both *ndh*F and *ycf*1 crossed the IRb/SSC junction, with an overlap of 92 bp in their coding regions. This overlapping sequence was consistent with observations in other plant genera, such as *Andrographis*, *Barleria*, and *Justicia* (Alzahrani et al., 2020). Similar overlapping sequences were also noted for *atp*E_*atp*B, *ndh*K_*ndh*C, and *psb*D_*psb*C genes in *Cissus* and other angiosperm chloroplast genomes (Rossini et al., 2021). The *ycf*1 gene, positioned at the IRa/SSC intersection, penetrated the IRa region with varying lengths (965 to 1082 bp) across compared genomes, a phenomenon common in angiosperms (Alzahrani et al., 2020; Li et al., 2022; Xu et al., 2022). Notably, there were no discernible differences in IR length among *Cissus* genomes. The presence of the *ycf*1 gene was consistent with most Vitaceae species, exhibiting variations in length from 3617 to 5811 bp. Despite its prevalence in many angiosperms, the *ycf*1 gene has not been reported in certain members of the Poaceae family (Goremykin et al., 2005; Guisinger et al., 2011).

F. hispida

A comparative investigation was carried out on the chloroplast genomes of four Moraceae plants, namely *F. hispida* (OQ220336), *F. auriculata* (NC_053837.1), *F. racemosa* (KT368151.1), and *F. squamosa* (OK077764.1). The study is strategically designed to unveil the intricate structural variations and functional nuances present within the chloroplast genomes of these *Ficus* spp. Utilizing progressive MAUVE alignment, a significant inversion event spanning approximately 48 kb was identified in the SSC region of *F. squamosa*. Concurrently, an interesting local colinear block (LCB) containing the *ndh*B gene was observed within the IR region in *F. hispida* and *F. auriculata*. This LCB exhibited an inverted orientation and reduced size in *F. racemosa*, while it was entirely absent in *F. squamosa*. Another distinct LCB, situated at LSC/IR junction in *F. auriculata*, was found to be inversed in *F. squamosa* but absent in the other two species (**Fig. 5.8**).

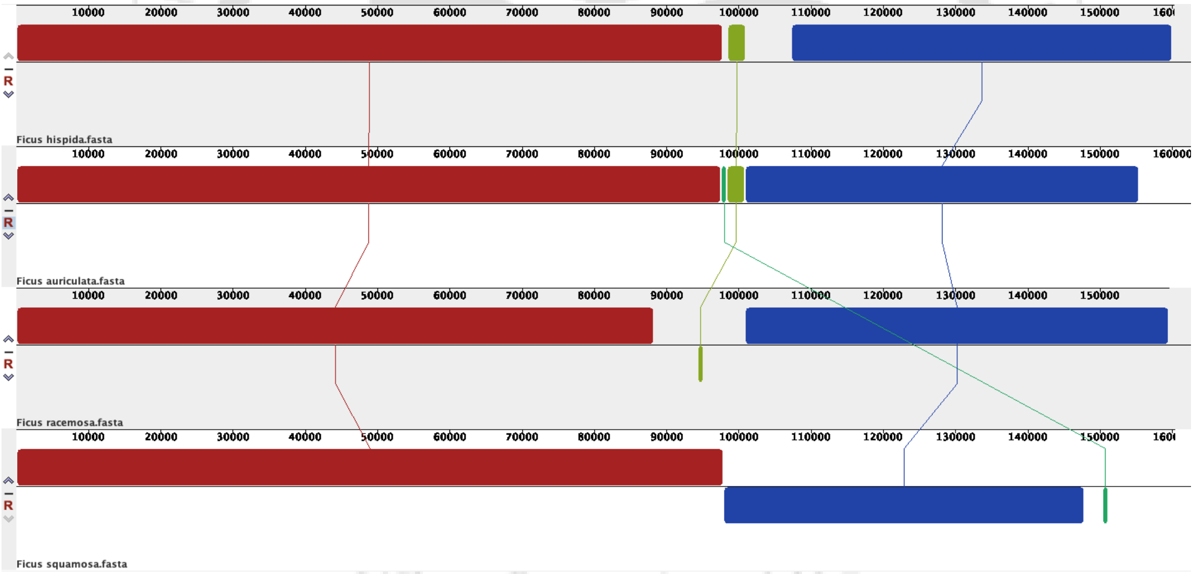


Figure 5.8: Progressive MAUVE alignment of cp genomes of four *Ficus* spp. Different colored local collinear blocks represent the gene arrangement in chloroplast genomes

The boundary between the IR and SSC regions featured two genes, namely *ycf*1 and *ndh*F. In *F. squamosa*, an inversion was evident for both these genes. Specifically, the *ycf*1 gene displayed variations in size at the SSC/IRa boundary, with differing base pair counts in the SSC and IRa regions. Notably, the larger segment of the *ycf*1 gene, sized 4696 bp and 4728 bp, was situated within the SSC region in *F. racemosa* and *F. hispida*, respectively. In contrast, the smaller segment spanned 1014 bp and 1079 bp within the IRa region in *F. squamosa* and *F. racemosa*, respectively. The predominant portion of the *ndh*F gene (2244-2255 bp) out of its total length (2267 bp) resided within the SSC region in all four species.

Consistently, the *rpl*2 gene was located within IR regions near the IR/LSC boundary. The *rps*19 gene was positioned at the boundary, with a segment inside the LSC and another within IRb, except for *F. racemosa*, where it was entirely within the LSC. The entirety of the *rpl*22 and *psb*A genes was confined to the LSC region. The *trn*H gene was primarily situated in the LSC region, positioned just outside the IRa/LSC boundary in all species except *F. racemosa* (**Fig. 5.9**). These findings collectively suggest the conservation of chloroplast genomes within the examined Moraceae species, with subtle deviations in gene arrangements at specific junctions.

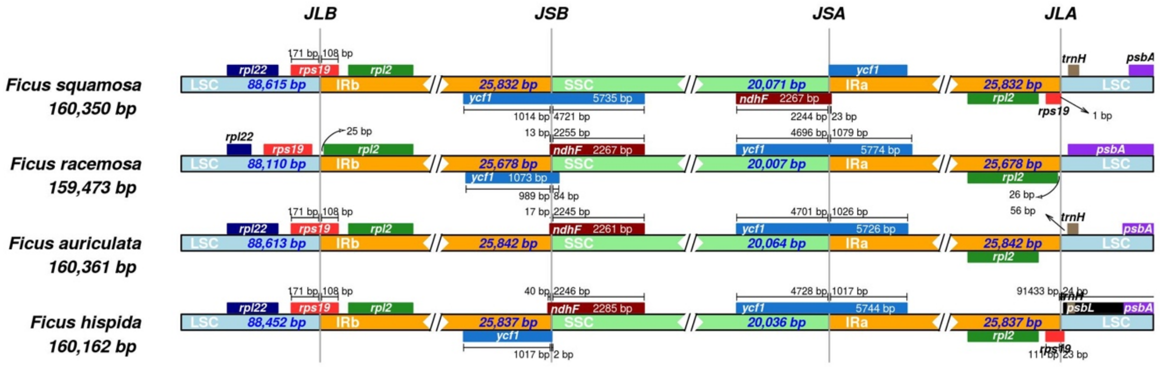


Figure 5.9: Comparison of Inverted Repeat (IR) junctions among four *Ficus* spp. Genes are represented by color boxes. The upper track displays genes on the negative strand, while those below the track represent genes on the positive strand

The comparative analysis of cp genomes of four *Ficus* spp. revealed intriguing variations and conservation patterns. The identification of a substantial inversion of 48 kb in SSC region of *F. squamosa* and the presence of a LCB containing the *ndh*B gene within the IR region in *F. hispida* and *F. auriculata*, with an inverted orientation and reduced size in *F. racemosa*, suggests dynamic evolutionary processes shaping chloroplast genome architecture. The absence of this LCB in *F. squamosa* further emphasizes the genomic divergence within the *Ficus* genus. Another distinct LCB at LSC/IR junction in *F. auriculata*, found to be inversed in *F. squamosa* but absent in the other two species, underscores the species-specific nature of these structural rearrangements. Similar phenomena of inversion events and LCB variations have been observed in the chloroplast genomes of *Cypripedium formosanum* and Orchids (*Paphiopedilum* spp.) other plant species, indicating that such structural changes are not unique to *Ficus* but rather represent recurring themes in plant evolution (Lin et al., 2015; Guo et al., 2021).

The observed inversion in both *ycf*1 and *ndh*F genes in *F. squamosa* suggests a dynamic rearrangement at the SSC/IRa boundary. This phenomenon of gene rearrangement and inversion at IR boundaries is not uncommon in chloroplast genomes and has been reported in various plant taxa, contributing to the understanding of chloroplast genome evolution. The consistent location of the *rpl*2 gene within IR regions is in line with the typical arrangement observed in most angiosperms. The positioning of the *rps*19 gene at the boundary with a segment inside the LSC and another within IRb, indicates variability in the boundary regions of these genomes. Similar observations have been reported in other monocots and *Solanum* spp., suggesting that the location of specific genes at IR/LSC boundaries is subject to evolutionary dynamics (Chung et al., 2006; Wang et al., 2008). The confinement of the *rpl*22 and *psb*A genes to the LSC region is consistent with chloroplast genome organization in as found in *Crocus* spp. (Cay et al., 2022). The primary location of the *trn*H gene highlights subtle deviations in gene arrangements at specific junctions, contributing to the overall diversity in chloroplast genome structures. These findings align with known patterns of chloroplast genome evolution in plants, where inversions, local colinear block variations, and gene rearrangements contribute to the diversity and adaptability of chloroplast genomes.

S. rutidosperma

The progressive MAUVE alignment of cp genomes within the Cleomaceae family identified two major inversions exclusively in *G. gynandra* and *C. paradoxa* within the IRb region, while the remaining compared genomes exhibited no inversions (**Fig. 5.10**). Within the investigated genomes, the junction between the IR and SSC regions featured two genes, *ycf*1 and *ndh*F. The *ycf*1 gene displayed expansion and contraction at the SSC/IRa boundary among Cleomaceae spp., with variations in base pairs between the SSC and IRa regions. The larger portion of the *ycf*1 gene, ranging from 4376 bp to 4412 bp, was situated in the SSC region, while the smaller part varied between 982 bp to 1029 bp in the IRa region. An anomalous arrangement was notably observed at the IRb/SSC boundary in *G. gynandra*, where the *ndh*F gene was located at the IRb/SSC junction in five species and at the SSC/IRa boundary in *C. paradoxa* and *G. gynandra*. It extended 34–37 bp in the IRb area and predominantly remained in the SSC region, whereas it extended 62 bp and 37 bp in IRa in *C. paradoxa* and *G. gynandra* respectively.

The *rps*19 gene spanned 55–112 bp in the IRb region and 167–224 bp in the LSC region across the LSC/IRb junction in all species. Additionally, *rps*19 was situated at the IRa/LSC border in *S. rutidosperma*, *T. chrysantha*, *C. pallida*, and *T. hassleriana*, with the larger portion inside IRa in all four cases. The complete *rpl*2 gene was present in both IRa and IRb regions, while *rpl*22 and *psb*A genes were entirely within the LSC region. The *trn*H gene was primarily located in the LSC region, just outside the IRa/LSC boundary in all species (**Fig. 5.11**). The observed patterns suggest that the cp genomes of Cleomaceae species under investigation were conserved, with only slight variations in gene arrangements at the junctions. Structural alterations in cp genomes are typically observed at the IR junctions, undergoing expansion and contraction in higher plants (Guo et al., 2021). Jin et al. (2020) demonstrated that the length of IR regions influences genome size and structure diversity. In Cleomaceae, no evidence of gene loss at the IR junctions was found, but the presence or absence of *rps*19 and *ycf*1 varied among the studied species. The borders of LSC to IRs and SSC to IRs in angiosperm cp genomes were highly conserved and typically located inside *rps*19 or *ycf*1, as reported in Leguminosae (Bai et al., 2021).

Figure 5.10: Plastome comparison among seven Cleomaceae species through progressive MAUVE alignment

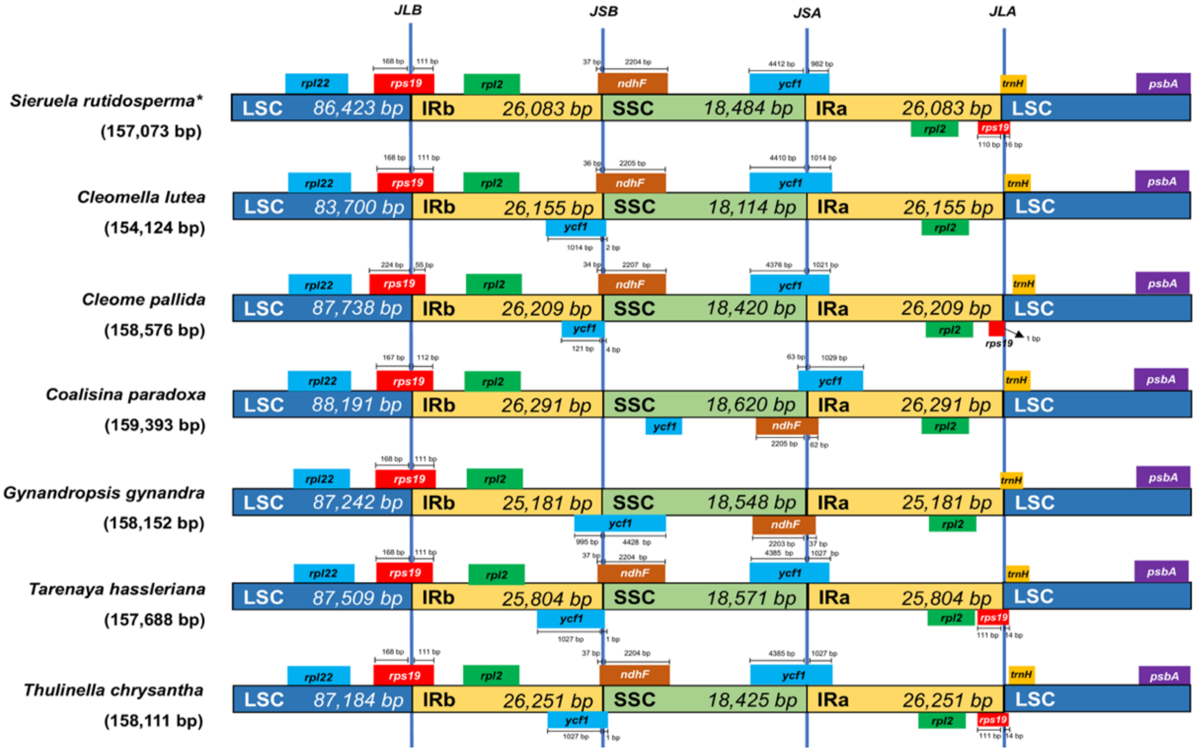


Figure 5.11: Comparison of IR junctions among seven Cleomaceae spp. Genes are represented by color boxes. The upper track displays genes on the negative strand, while those below the track represent genes on the positive strand

In five out of the seven species, *ndh*F and *ycf*1 genes exhibited an overlap at the IRb/SSC junction, a phenomenon observed in various angiosperm species. However, *rpl*2 and *rpl*22 were found to be located entirely within their respective regions without overlap. The *ycf*1 gene was observed to cross the IRa/SSC junction with varying lengths ranging from 982 to 1029 bp, a common feature in angiosperms (Yang et al., 2016). The length of the *ycf*1 gene ranged from 3639 bp to 5424 bp across the studied Cleomaceae genomes, and although *ycf*1 is widely present among angiosperms, it has been confirmed to be absent in some members of the Poaceae family (Huang et al., 2017). Overall, there was no noticeable difference in IR length between the compared Cleomaceae genomes.

S. asper

The comparative sequence analysis of plastid genomes across seven *Streblus* spp. namely *S. asper* (OQ281590), *S. banksii* (MW238798), *S. glaber* (MW238799), *S. heterophyllus* (MW238800), *S. indicus* (MN065161), *S. pendulinus* (MW238801), and *S. smithii* (MW238802) revealed a spectrum of sequence similarities ranging from 88.63% to 99.74%. Notably, the junctions within the IR regions served as precise markers for delineating the boundaries between the IR and SSC regions (Wang et al., 2022). Specifically, the *ycf*1 gene was identified as a consistent marker for demarcating the boundary between IRa and SSC regions across all seven *Streblus* spp. However, a unique observation was made in *S. pendulinus*, where *ycf*1 extended along the IRb/SSC border. Additionally, in line with findings in other species under investigation, the *ndh*F gene was also observed at this border. Further exploration of the IR/SSC border unveiled variations, with the presence of *ccs*A in *S. heterophyllus* and *rps*19 in *S. asper*. Intriguingly, *S. heterophyllus* exhibited absence of the *ycf*1 gene at the IR/SSC border, instead, it was situated at the IR/LSC border, extending 2118 bp into the IR region and 3513 bp into the LSC region. The positioning of *rpl*2 and *ycf*2 along the IR/LSC border, along with the consistent observation of *psb*A and *rps*19 at this border (except in *S. smithii*), highlighted a shared genomic arrangement pattern among the *Streblus* spp. **(Fig. 5.12)**

The distribution of specific genes within the genome also revealed interesting nuances. For instance, *rpl*2 was present in both IRa and IRb regions, while *rpl*22 and *trn*H frequently observed in the LSC region as observed in other Moraceae plants (Huang et al., 2022). However, a notable exception was observed in *S. asper*, where *rpl*22 was positioned within the SSC region. Additionally, the unique presence of *rps*7 at the IR/LSC border in *S. banksii* added to the diversity of cp genome arrangements. In terms of structural variations, a noteworthy discovery was made in *S. indicus* and *S. asper*, where a small LCB of approximately 1.1 kb was identified, which exhibited translocation without inversion. Conversely, in the remaining cases, the overall genome arrangement demonstrated a remarkable similarity. These findings contribute valuable insights into the cp genome organization within the *Streblus* spp., shedding light on both conserved patterns and species-specific variations (**Fig. 5.13**).

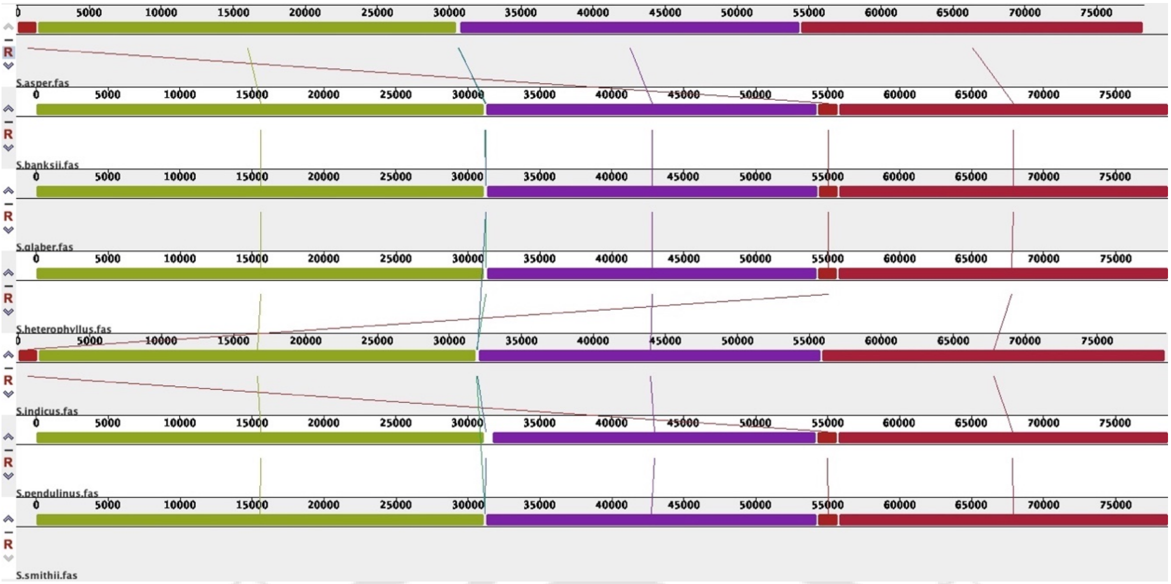


Figure 5.12: Plastome comparison through progressive MAUVE alignment of seven *Streblus* spp.

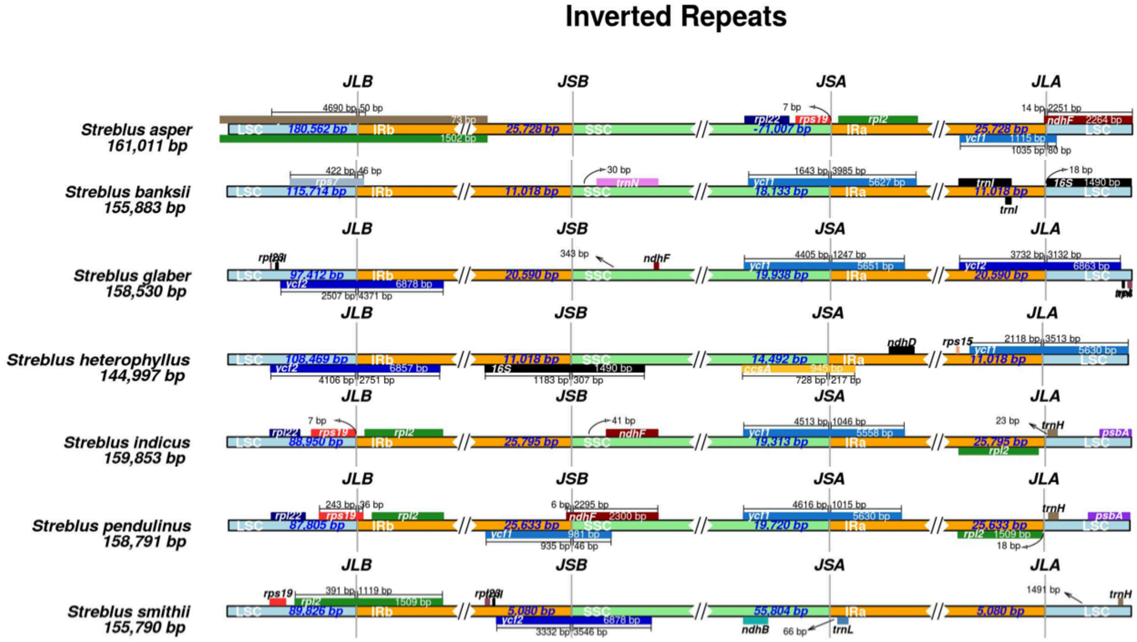


Figure 5.13: Comparisons of LSC, SSC, and IR region borders among seven cp genomes of *Streblus* spp., colour boxes present the coding genes

The comparative analysis of cp genomes across seven *Streblus* spp. provides a comprehensive understanding of the structural nuances and evolutionary dynamics within this genus. The observed spectrum of sequence similarities signifies both conserved elements and species-specific variations in their cp genomes. The consistent identification of the *ycf*1 gene as a marker for the IRa/SSC border across all *Streblus* spp. underscores its evolutionary significance in maintaining genome integrity. The intriguing extension of *ycf*1 along the IRb/SSC border in *S. pendulinus* adds a unique layer to the genomic landscape, suggesting potential adaptive strategies in this particular species. The presence of *ndh*F at this border aligns with observations in related species.

The absence of *ycf*1 at the IR/SSC border in *S. heterophyllus*, coupled with its relocation to the IR/LSC border, unveils a noteworthy genomic rearrangement, highlighting the plasticity of chloroplast genomes within the *Streblus* genus. The shared arrangement pattern of *rpl*2, *ycf*2, *psb*A, and *rps*19 along the IR/LSC border among compared species emphasizes a conserved structural motif (Guo et al., 2021). The exceptional positioning of *rpl*22 within the SSC region in *S. asper* suggests species-specific genomic adaptations (Cay et al., 2022). The unique presence of *rps*7 at the IR/LSC border in *S. banksii* contributes to the intriguing diversity of cp genome arrangements within the *Streblus* spp. Structural variations, exemplified by a translocated LCB in *S. indicus* and *S. asper*, provide insights into evolutionary events shaping the plastid genomes of these species (Wang et al., 2022). Conversely, the remarkable similarity in overall genome arrangement among the remaining species underscores a shared genomic heritage within the *Streblus* spp. Collectively, these findings enhance our understanding of cp genome organization in *Streblus* spp., offering valuable insights into both conserved patterns and species-specific adaptations.

T. peruviana

Examining the junctions in the plastome's IR regions allows us to precisely identify the boundaries between the IR and SSC regions. Across five Plumerieae spp., it was consistently found that two genes, *ycf*1 and *ndh*F, marked the boundary between the IR and SSC regions (**Fig. 5.14**). The *ycf*1 gene showed a unique pattern, with a larger part (4920 bp to 4416 bp) in the SSC region and a smaller part (1020 bp to 1104 bp) in the IRa region. Notably, *P. rubra* was an exception, lacking the *ycf*1 gene. The *psb*A genes were entirely within the LSC region, while *ndh*F and *rpl*22 genes were consistently found in the SSC and LSC regions, respectively, across all species. The *rpl*2 gene was present in both IRa and IRb regions, and the *rpl*22 and *trn*H genes mainly resided in the LSC region, just outside the IRa/LSC boundary in all species. Interestingly, the position of *rps*19 varied, being at the IRa/LSC boundary in *T. peruviana* and *P. obtusa* but absent in *C. manghas*, *P. krugii*, and *P. rubra* **(Fig 5.15**). *T. peruviana* stood out with the longest non-coding sequence (79,197 bp), while *C. manghas* had the shortest (74,157 bp).

The differences in plastome length within the Plumerieae tribe can be attributed to various factors, including changes in the size of IR regions, variations in intergenic regions, and differences in gene numbers. *T. peruviana*, for instance, distinguished itself by having the *rps*19 gene in the IRb region, a feature not seen in other Plumerieae spp. Similar observations have been made in other plant species like *Panax ginseng* and *Cucumis sativus* (Zhao et al., 2015; Liu et al., 2016). Despite variations in the presence and length of certain genes, such as *ycf*1 and *ndh*F, at specific junctions, these differences did not significantly impact the overall size differences of IR regions (Amiryousefi et al., 2018; Xu et al., 2022). Examining the overall plastome size highlighted that *C. manghas* and *T. peruviana* had larger plastomes compared to the other three species, primarily due to the extended length of their non-coding sequences. This consistent trend underscores the significant role of non-coding regions, comprising intergenic spacers, introns, and repetitive sequences, in determining plastome dimensions (Alzahrani et al. 2020; Guo et al. 2021).

Figure 5.14: Plastome comparison using progressive MAUVE alignment of five Plumerieae spp.

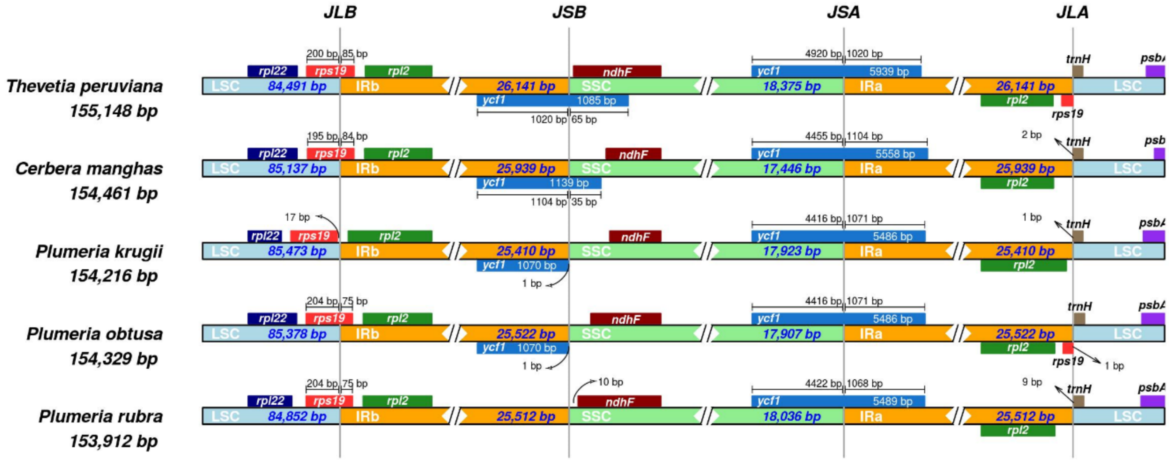


Figure 5.15: Comparisons of LSC, SSC, and IR region borders among five Plumerieae chloroplast genomes of , colour boxes present different coding genes

5.3.2.3. *Tandem and simple sequence repeat*

Repeat elements within cp genomes constitute a captivating area of study with profound implications for understanding plant evolution and species identification. These repetitive sequences, including tandem repeats and SSRs, play a pivotal role in shaping the structural and functional dynamics of cp genomes (Shelke et al., 2022). Tandem repeats, where identical or nearly identical sequences are arranged in close proximity, contribute to genetic variability and can be employed as informative tools for assessing genetic relationships among plant species (Wu et al., 2018). The variability introduced by repeat elements offers valuable insights into molecular markers, enabling the analysis of population genetic structure and evolutionary relationships across distinct species (Shelke and Das 2015; Shelke and Rangan, 2020; Wu et al. 2021). Similarly, SSRs, characterized by short, repeated motifs, offer a wealth of information for species identification due to their variability in length and sequence (Gandhi et al., 2010). Exploring these repeat elements in chloroplast genomes not only deepens our understanding of plant molecular biology but also provides practical applications in delineating species boundaries and unravelling the evolutionary relationships among diverse plant taxa (Liu et al. 2021). The repeat elements within cp genomes of five selected species and their close relatives are discussed as below.

C. quadrangularis

The cp genome of *C. quadrangularis* revealed a total of 32 repeat elements, including 16 forward repeats and 15 palindromic repeats, with no reverse repeats identified (**Fig. 5.16a**). Across the six genomes analysed, the count of tandem repeats ranged from 32 in *C. quadrangularis* to 49 in *C. trifoliata* (**Fig. 5.16b**). Primarily, forward repeats were prevalent in all six genomes, with *C. antarctica* displaying the maximum number (27) and *C. discolor* and *C. quadrangularis* jointly exhibiting the minimum (16). Palindromic repeats were most abundant in *C. antarctica* and *C. trifoliata* (20), but least in *C. microcarpa* (13). Notably, *C. tuberosa* recorded the highest number of reverse repeats (9), while *C. quadrangularis* did not display any reverse repeats. Complementary repeats were absent in both *C. antarctica* and *C. microcarpa*, while *C. discolor* recorded a maximum of five repeats. Tandem repeats in the length range of 30–34 bp were the most frequent, followed by fragments in the 35–39 bp and 40–44 bp ranges (**Fig. 5.16c**). Longer tandem repeats, specifically within the 45–49 bp and 55–59 bp ranges, were uncommon across all species except for *C. quadrangularis*. Additionally, *C. discolor* exhibited nine repetitions within the 60–64 bp range, and repetitions were sporadically observed in the 70–74 bp, 75–79 bp, 80–84 bp, and 85–89 bp ranges.

In the cp genome of *C. quadrangularis*, a total of 70 simple and 9 compound SSRs were identified (**Fig. 5.17a**). Among the simple SSRs, mononucleotide repeats accounted for 50.6%, dinucleotides for 22.8%, trinucleotides for 8.9%, and tetranucleotides for 14.1% (**Fig. 5.17b**). Comparative analysis across species revealed varying SSR counts, with *C. antarctica*, *C. discolor*, *C. microcarpa*, *C. trifoliata*, and *C. tuberosa* containing 67, 81, 72, 95, and 72 SSRs, respectively. *C.*

microcarpa exhibited the highest percentage of mononucleotide SSRs (73.6%), followed by *C. discolor* (67.9%) and *C. trifoliata* (67.37%). In contrast, *C. quadrangularis* had the lowest mononucleotide percentage (50.6%), followed by *C. tuberosa* (59.72%). A/T motifs were abundant in all cases, while C/G repeat motifs were rare. *C. trifoliata* displayed the most A/T-rich repeats (64), followed by *C. discolor* (54), with *C. antarctica* and *C. quadrangularis* having the fewest (38). Notably, AT/AT repeats were more prevalent in *C. quadrangularis* (18) than *C. discolor* (15), while *C. microcarpa* recorded the least (7). AAT/ATT repeats emerged as the most common trinucleotide repeats, while others were infrequent (**Fig. 5.17c**). *C. trifoliata* exhibited a lowest SSR density (0.6 kb/SSR), whereas *C. antarctica* had the highest SSR density (2.42 SSR/kb) (**Fig. 5.17d**).

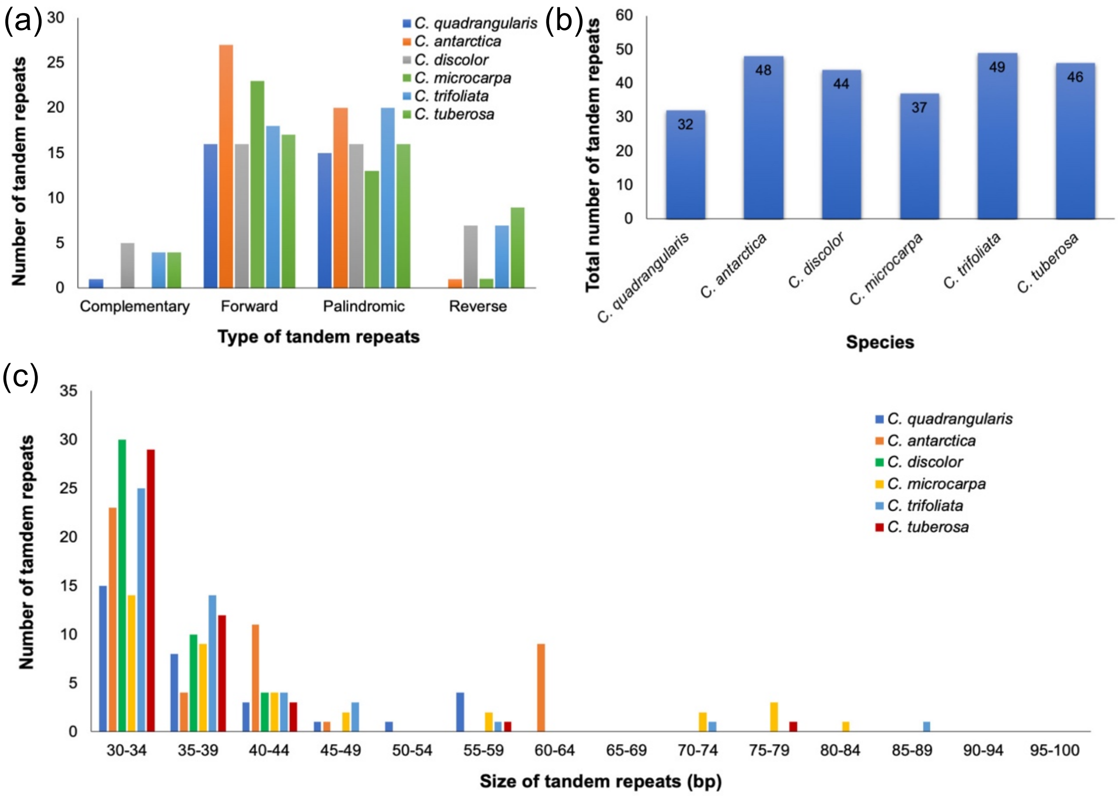


Figure 5.16: Comparison of tandem repeats of six chloroplast genomes of *Cissus* spp. (a) Number of tandem repeats by tandem repeat types, (b) Total number of tandem repeats in each species, and (c) Number of tandem repeat sequences by length

Long tandem repeats, a phenomenon commonly observed in chloroplast genomes, play a role in intermolecular recombination, contributing to sequence diversity (Guo et al., 2021). Within the *Cissus* genomes, the count of tandem repeats varied from 32 in *C. quadrangularis* to 49 in *C. trifoliata*. Notably, despite having the largest chloroplast genome, *C. antarctica* exhibited the fewest number of SSRs among the examined genomes, consistent with previous findings in *M. ferrea* (Shelke and

Rangan, 2022). This implies that the sequence variability introduced by tandem repeats holds potential for designing species-specific markers (Cregan et al., 2020).

SSRs, as a subtype of tandem repeats, are abundantly present in genomes and offer valuable genetic markers (Khan et al., 2019; Shelke and Rangan, 2022; Wang et al., 2021). The count of SSRs ranged from 67 (*C. antarctica*) to 95 (*C. trifoliata*) in the cp genomes of *Cissus* spp. Remarkably, *C. trifoliata* exhibited the highest number of both tandem and SSR repeats among the six *Cissus* spp., while *C. antarctica* displayed the fewest SSRs despite having the largest genome size, consistent with previous findings in *M. ferrea* (Shelke and Rangan, 2022). Notably, despite similar genome sizes in *C. trifoliata* and *C. quadrangularis*, there was a significant difference in the repeat count. This observation is further supported by the significantly lower SSR density in *C. trifoliata* (1.69 SSR/kb). The emergence of numerous SSRs in the genome likely reflects evolutionary adaptations to specific environmental factors, leading to species-specific variations that can be harnessed for barcode marker design (Senapati et al., 2022).

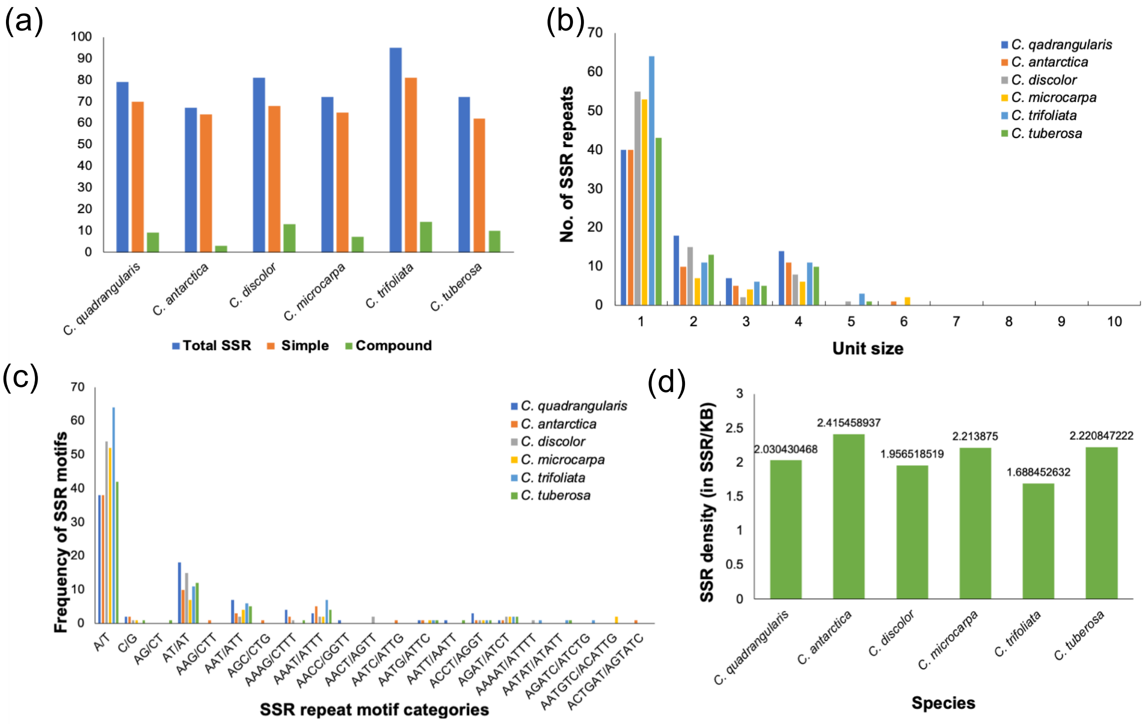


Figure 5.17: Comparison of SSR repeats between *C. quadrangularis* and five chloroplast genomes of *Cissus* species. (a) Number of SSR repeats in regular and compound formation, (b) Total number of SSR repeat types, (c) The number of SSR motifs, and (d) SSR density

F. hispida

Within this research endeavor, an extensive examination of the chloroplast genome of *F. hispida* was undertaken, resulting in the identification of a total of 35 tandem repeats. Among these, 11 were forward repeats, while 21 manifested as palindromic. Additionally, a single complementary and two reverse repeats were also observed. In the comparative context of other Moraceae species, *F. auriculata* displayed the lowest count of tandem repeats, amounting to 18 in total. Palindromic repeats emerged as the most prevalent repeat type across all four species, with *F. racemosa* showcasing the highest abundance of 25, while *F. auriculata* exhibited the lowest with 14. Following in prominence forward repeats were identified as the second most abundant category. Among the species mentioned earlier, *F. racemosa* held the record for the highest occurrence of forward repeats (18), while the lowest (04) was observed in *F. auriculata*.

Notably, the tandem repeat landscape also included reverse repeats, where the highest number (03) was found in *F. racemosa* and *F. squamosa* collectively. *F. hispida* contained two reverse repeats, whereas *F. auriculata* was devoid of such repeats. Complementary repeats were absent in both *F. auriculata* and *F. racemosa*, while *F. hispida* had a solitary instance, and *F. squamosa* exhibited two repeats. The distribution of tandem repeat lengths unveiled distinct patterns. Repeats spanning 30–34 bp emerged as the most frequent, with *F. racemosa* displaying the maximum count of 27. Within the 35–39 bp intervals, *F. racemosa* again took the lead with ten (10) repeats, while *F. auriculata* presented the lowest count of six (06). The subsequent interval of 50–54 bp witnessed a substantial number of repeats, with *F. auriculata* registering the highest tally of six (06). Notably, the 40–44 bp segment also hosted a noteworthy concentration of tandem repeats. Within the intervals of 45–49 bp and 60–64 bp, two repeats were observed each, whereas all other intervals either held a single repeat or none (**Fig. 5.18**). These observations collectively offer insightful revelations concerning the intricate structural variations exhibited by tandem repeats within the chloroplast genome of *Ficus* spp.

Within this investigation, a thorough analysis of the chloroplast genome of *F. hispida* was conducted, leading to the identification of 83 microsatellites, commonly referred as SSRs. Among these, 73 were classified as simple, while the remaining 10 were compound SSRs. Notably, the majority of these SSRs were characterized by mononucleotide repeats, accounting for 61.44% of the total. Dinucleotide repeats followed, constituting 20.48%, trailed by tetranucleotide (10.84%) and trinucleotide (3.6%) repeats. Through a comparative lens encompassing other plant species, distinct trends emerged in the distribution of SSRs. Among the *Ficus* spp. under scrutiny, *F. auriculata* exhibited the highest proportion of mononucleotide SSRs (64.5%), trailed by *F. racemosa* (59.77%) and *F. squamosa* (58.75%). Across all cases, motifs rich in adenine (A) and thymine (T) bases dominated the repeat landscape, whereas motifs featuring cytosine (C) and guanine (G) were notably rare. The A/T-rich repeats were most pronounced in *F. auriculata* (59), followed by *F. racemosa* (51), while *F. squamosa* displayed the fewest (46). Subsequently, AT/AT repeats were notably abundant, with the highest count (18) observed in *F. racemosa* and *F. squamosa* in both, followed by *F. auriculata* (17), with *F. hispida* registering the least at 16.

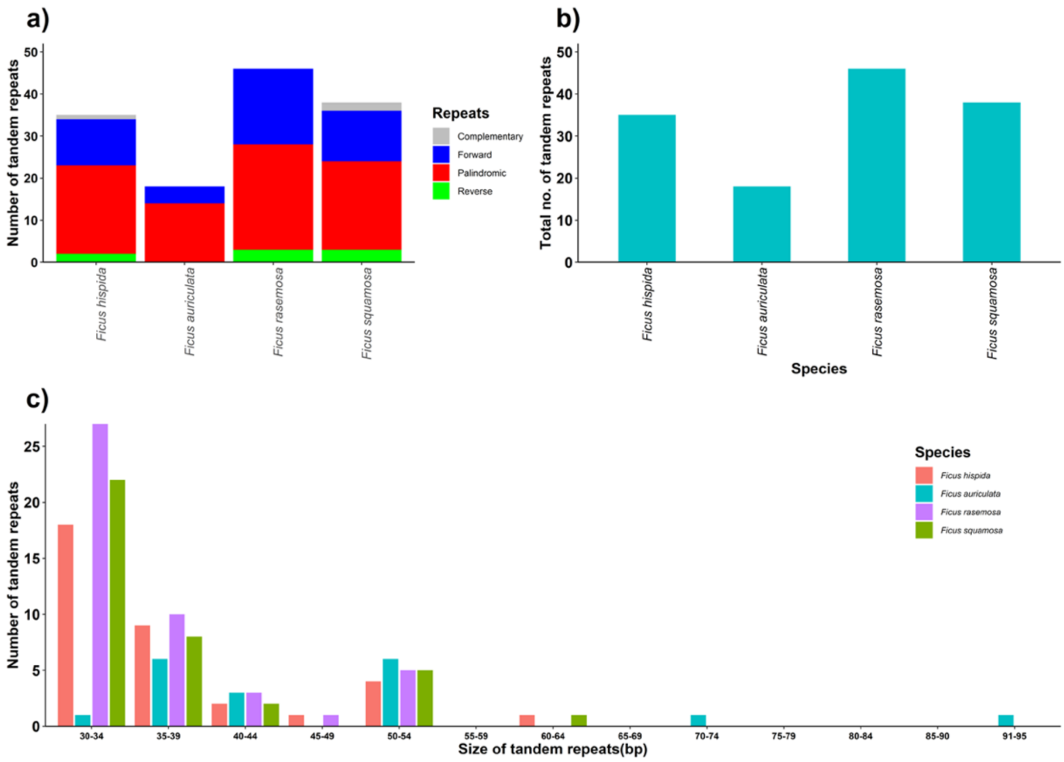


Figure 5.18: Tandem repeat mining from four *Ficus* spp. (a) Repeat types; (b) Total no of tandem repeats in each species, and (c) Number of repeats based on repeat length

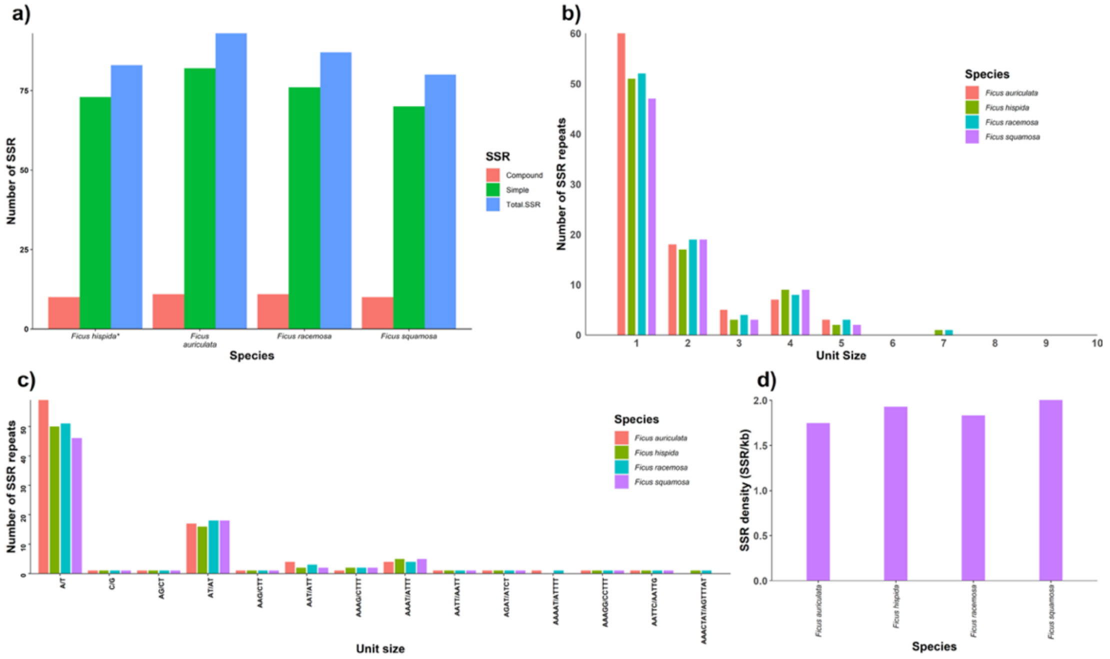


Figure 5.19: Simple sequence repeats in *Ficus* spp. (a) Types of SSR; (b) No of SSR based on unit size' (c) Number of SSR based on unit type, and (d) SSR density in four species

The recurring trinucleotide repeat AAT/ATT emerged as the most prevalent within this category, with *F. auriculata* showcasing the highest frequency of occurrence (4). Furthermore, the density of SSRs, measured in terms of SSRs per kb, unveiled insightful patterns. *F. auriculata* demonstrated the highest SSR density at 0.57 SSR/kb, followed by *F. racemosa* (0.55 SSR/kb), *F. hispida* (0.52 SSR/kb), and lastly, *F. squamosa* with the lowest SSR density of 0.5 SSR/kb (**Fig. 5.19**). Taken together, these findings underscore the abundant and diverse nature of SSRs within the chloroplast genomes of *Ficus* spp. Furthermore, the comprehensive understanding of these microsatellites serves to illuminate the plant species' evolutionary history, thereby enhancing our insights into its genetic characteristics and patterns of variation.

S. rutidosperma

In this study, we examined the cp genome of *S. rutidosperma* and identified 48 tandem repeats, of which 18 were forward repeats, 20 were palindromic, single complement, and three were reverse repeats. In comparison to six other Cleomaceae species, *S. rutidosperma* had the lowest number of tandem repeats (48). Palindromic repeats were the most abundant across all seven species, with the maximum in *C. paradoxa* and the minimum in *T. hassleriana*. Forward repeats were the second most abundant, with the highest numbers observed in *C. paradoxa* and the lowest in *G. gynandra*. The highest number of reverse repeats were observed in *T. hassleriana*, while *C. paradoxa* had the least. Complementary repeats were absent in *C. paradoxa*, while *T. chrysantha* and *G. gynandra* had the highest number of two complementary repeats each. Tandem repeats of length 30–34 bp were the most frequent, followed by 35–39 bp and 40–44 bp fragments. Tandem repeats within the intervals 55–59 bp and 70–74 bp ranges were uncommon in all species, with a maximum of a single repeat if any. Only *G. gynandra* had five repetitions within the 65–69 bp range and *C. paradoxa* had four repeats within the 60-64 range. Additionally, a few repetitions could be observed in the range of 55–59 bp, 60–64 bp, 70–74 bp, and 75–79 bp, respectively (**Fig. 5.20**). These findings provide valuable insights into the structural variation of tandem repeats in the cpt genome of *S. rutidosperma* and other plant species.

In this study, we analyzed the chloroplast genome of *S. rutidosperma* and identified 145 SSRs, including 95 simple and 50 compound SSRs. The majority of the SSRs were mononucleotide (77.2%), followed by tetra-nucleotide (11%), dinucleotide (9.7%), and tri-nucleotide (2%). Comparative analysis of SSRs in other plant species showed that *S. rutidosperma* had the highest percentage of mononucleotide SSRs, followed by *C. paradoxa* and *G. gynandra* (74.8%). A/T motifs were found to be abundant in all cases, while C/G repeat motifs were rare. The most A/T rich repeats were observed in *C. paradoxa* (115), followed by *S. rutidosperma* (112), while *C. lutea* had the fewest (88). AT/AT repeats were highest in *C. lutea* (17) and the second-highest was observed in *C. pallida* (16), whereas the least was found in *C. paradoxa* and *T. hassleriana* (11) jointly. The most common trinucleotide repeat was AAT/ATT, with the highest frequency being present in *G. gynandra* (8). Also, SSR density was the highest in *G. gynandra* (0.98 SSR/kb), followed by *C. paradoxa* (0.97 SSR/kb), while *T.*

chrysantha had the lowest SSR density (0.82 SSR/kb) (**Fig. 5.21**). Our findings suggest that SSRs are abundant and diverse in the chloroplast genome of *S. rutidosperma* and provide valuable insights into the evolutionary history of this plant species.

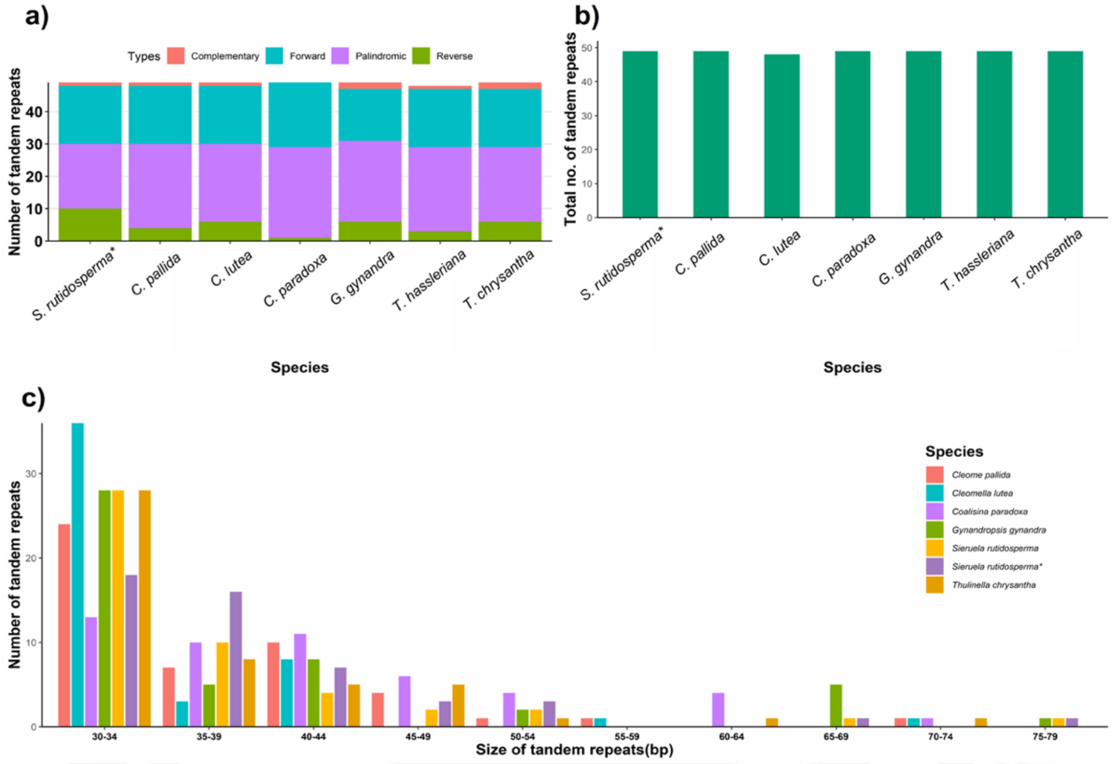


Figure 5.20: Tandem repeat mining from seven Cleomaceae plants (a) Repeat types; (b) Total no of tandem repeats in each species, and (c) Number of repeats based on repeat length

The cp genome is composed of diverse repeated sequences, which include homo-polymeric repeats, SSRs, and long repeats. These sequences play a crucial role in genome evolution and reorganization, as they serve as a source of genetic variation (Das et al., 2018; Shelke et al., 2022). Thus, repeat elements can be utilized as molecular markers for examining population genetic structure and evolutionary relationships among different species (Wu et al., 2021). Intermolecular recombination is known to generate sequence diversity in the chloroplast genome, and long tandem repeats are often associated with this process (Guo et al., 2021). The number of tandem repetitions in all of the Cleomaceae genomes was found to be more or less the same, varying between 48 to 49 (Khan et al., 2019; Shelke and Rangan, 2022). Interestingly, despite being the smallest in size, the genome of *C. lutea* had the same number of tandem repeats as the other genomes.

The number of SSRs in cp genome usually higher than long tandem repeats because of their small sizes (Wang et al. 2021; Shelke and Rangan, 2022). The total number of SSRs in Cleomaceae species ranged from 128 in *C. lutea* to 155 in both *C. paradoxa* and *G. gynandra* resembling their cp genome sizes. Even though the genomes of *C. pallida* and *G. gynandra* are roughly the same size, there was a substantial difference in the number of SSRs between them. It is difficult to compare repeat population results from different studies because of the variable mining parameters employed by different authors. Irrespective of the comparison emergence of a large number of SSRs in the genome is likely due to adaptations that have developed over the course of evolution in response to specific environmental factors (Shelke et al. 2020). These new SSR mutations can lead to species-specific variations and can aid in the development of barcodes as a marker for species delimitation

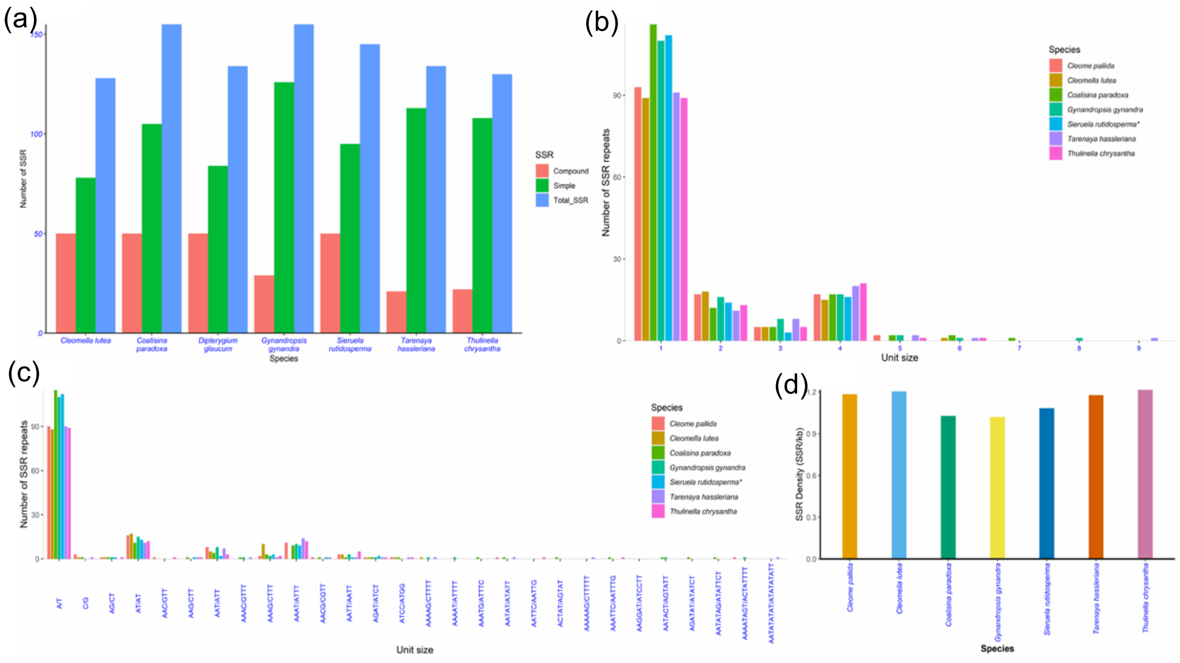


Figure 5.21: Simple sequence repeats in Cleomaceae *plants* (a) Types of SSR; (b) No of SSR based on unit size' (c) Number of SSR based on unit type, and (d) SSR density in four species

S. asper

Through a meticulous examination of the cp genome of *S. asper*, our scrutiny revealed the presence of 47 tandem repeats, comprising 18 forward repeats, 26 palindromic repeats, and three reverse repeats. An intriguing observation emerged in the context of *S. glaber*, where only two complementary repeats were discerned and absent in others, distinguishing it from the other species in the study. Noteworthy variations in the total number of tandem repeats were evident, ranging from 27 in *S. heterophyllus* to 49 in *S. indicus*. Of particular interest was the identification of eight reverse repeats in *S. indicus*, a notably higher count compared to the other species. In contrast, all other species exhibited a significantly lower prevalence of such reverse repeats. The distribution of tandem repeats within specific size intervals revealed a consistent pattern across the *Streblus* spp. The most frequently occurring tandem repeats fell within the 30–34 bp interval, showcasing a shared trend across all seven species as observed in other Moraceae plants (Yang et al., 2022). Following closely, the 35–39 bp interval housed the next most prevalent set of tandem repeats. Longer repeats, a rarity in occurrence, were generally limited to single or double instances. Intriguingly, *S. indicus* stood out with the exclusive presence of three repeats in the length range of 55–59 bp. Furthermore, a solitary repeat within this size range was identified in *S. pendulinus*, while such repeats were conspicuously absent in the remaining species (**Fig. 5.22**).

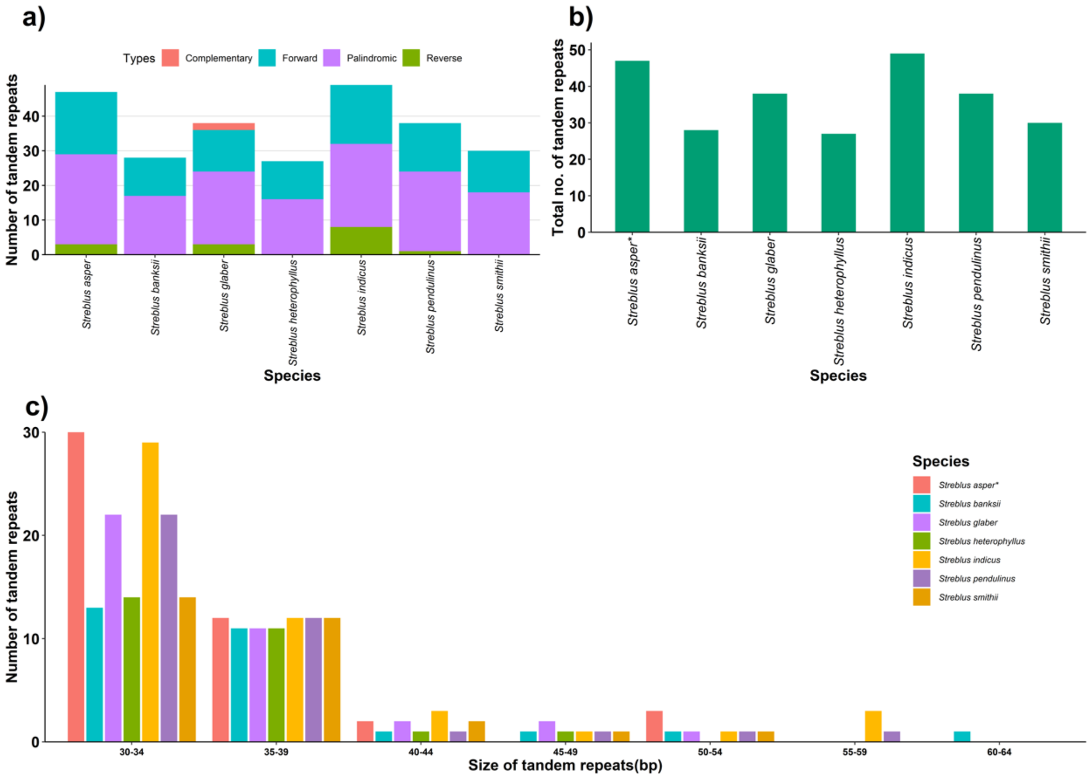


Figure 5.22: Tandem repeats in seven *Streblus* spp. (a) Types of tandem repeats; (b) Total number of tandem repeats, and (c) Number of tandem repeats based on size

In the exploration of the *S. asper* cp genome, a discernment of 82 SSRs was achieved; comprising 75 simple and 7 compound repeats. Notably, *S. pendulinus* exhibited the highest SSR count (104). The prevalence of mononucleotide SSRs dominated, ranging from 48.28% in *S. heterophyllus* to 73.91% in *S. banksii.* Dinucleotide SSRs were also substantial, constituting 8.65% to 17.07%, while other repeat types contributed below 10%, with the exception of tetranucleotides in *S. heterophyllus* (17.24%), *S. pendulinus* (11.53%), and *S. smithii* (10.93%). The SSR density, reflecting the abundance of repeats per kilobase, revealed *S. pendulinus* with the highest density at 0.65 SSR/kb, followed by *S. glaber* (0.57 SSR/kb), *S. indicus* (0.54 SSR/kb), and *S. asper* (0.51 SSR/kb). In contrast, the lowest SSR densities were observed in *S. heterophyllus* (0.18 SSR/kb) and *S. smithii* (0.40 SSR/kb) (**Fig. 5.23**). These findings provide valuable insights into the diversity of SSRs across *Streblus* spp., with potential implications for genomic variation and evolutionary dynamics within the genus.

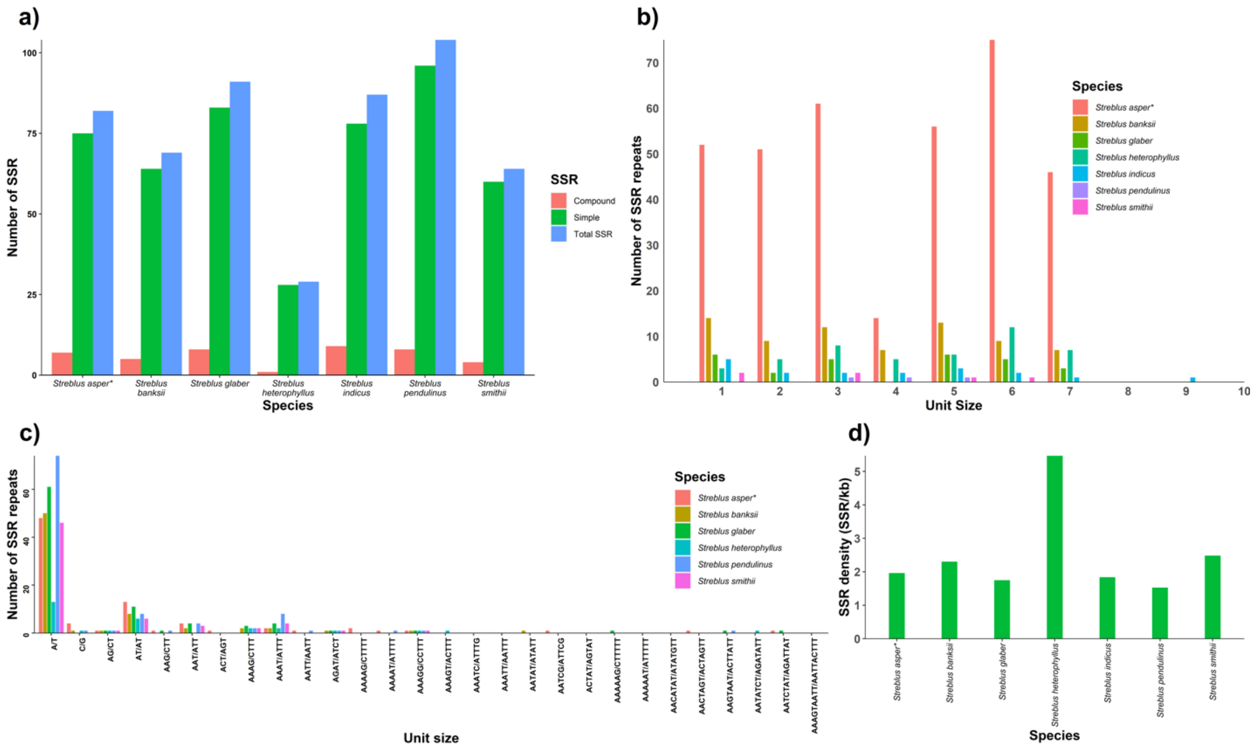


Figure 5.23: SSRs found in seven *Streblus* spp. (a) Types of SSR; (b) Number of SSRs based on size; (c) Number of SSRs based on repeat types, and (d) SSR density

The analysis of tandem repeats within the cp genomes of seven *Streblus* spp. has unveiled intriguing patterns, contributing valuable insights into the structural diversity of these genomes. The presence of 47 tandem repeats in *S. asper*, including forward, palindromic, and reverse repeats, underscores the dynamic nature of repetitive elements in cp DNA. Particularly noteworthy is the species-specific presence of complementary repeats in *S. glaber*, distinguishing it from the others, similar to observations in *Morus* spp. (Kong et al., 2017). Varied total tandem repeat counts, ranging from 27 to 49, suggest unique genomic signatures across the *Streblus* spp. The heightened prevalence of reverse repeats in *S. indicus* raises questions about the functional significance and potential adaptive roles of these elements in this species as found in *Artocarpus* sp. (Souza et al., 2020). The consistent distribution of tandem repeats within specific size intervals, with a preference for 30–34 bp, points towards shared genomic characteristics in the *Streblus* spp., while the rare occurrence of longer repeats introduces a layer of genomic intricacy, notably observed in *S. indicus* and other Moraceae plants including *Ficus* spp. (Xia et al., 2022)

In parallel, the examination of SSRs in *S. asper*'s cp genome provides a comprehensive view of microsatellite abundance and diversity. *S. pendulinus* stands out with the highest SSR count, emphasizing the species-specific nature of these repeats as reported in *Cactus* and *Salvia* spp. (Ning et al., 2020; Oulo et al., 2020). The dominance of mononucleotide SSRs, especially in *S. banksii*, suggests potential evolutionary pressures shaping these motifs. Substantial contributions of dinucleotide SSRs and notable occurrences of tetranucleotide repeats in specific species further enrich our understanding of cp genomic architecture and uniqueness (Kong et al., 2017). The SSR density analysis reveals variations across species, with *S. pendulinus* exhibiting the highest density, potentially indicating a higher degree of genomic instability or adaptive potential. Conversely, the lower SSR densities in *S. heterophyllus* and *S. smithii* suggest genomic stabilities or distinct evolutionary trajectories (Sun et al., 2018). Together, these findings underscore the importance of repetitive elements in shaping cp genome architecture within the *Streblus* spp.

T. peruviana

Upon analyzing the cp genome of *T. peruviana*, we identified a total of 65 tandem repeats, with 38 being forward repeats and 27 palindromic repeats. In comparison, only two complementary repeats were observed in *P. obtusa*. The tandem repeats varied in length, with 30–34 bp being the most common, followed by 35–39 bp and 40–44 bp. Longer repeats, especially in the 40–44 bp and 80–84 bp ranges, were more prevalent in *T. peruviana*, *C. manghas*, and *P. krugii*, while they were less common in *P. obtusa* and *P. rubra*. A comparative analysis across species revealed that *C. manghas* had the highest overall count of tandem repeats (99), followed by *T. peruviana* (65), *P. rubra* (27), *P. krugii* (28), and *P. obtusa* (33) (**Fig. 5.24**). In *T. peruviana* a total of 69 SSRs were detected, with 61 being simple and 8 compound. *T. peruviana* had the lowest proportion of mononucleotide SSRs (52.17%) among the studied Plumerieae spp. The highest SSR density was found in *T. peruviana* (0.44 SSR/kb), followed by *C. manghas* (0.40 SSR/kb) and *P. rubra* (0.30 SSR/kb) (**Fig. 5.25**).

It is a common observation that non-coding regions accumulate more repetitive elements than coding regions, potentially serving as a mechanism to prevent harmful mutations. The types of repeats also differ among species, with *C. manghas* having abundant "Forward" repeats, while *T. peruviana* and *P. krugii* show comparable counts for "Forward" and "Palindromic" repeats. These variations in repeat types can have significance in structural variations and can serve as practical markers for species identification. A notable correlation was found between the number of SSRs and the size of the cp genome. *T. peruviana* and *C. manghas*, with larger genomes, displayed a higher count of SSRs, while *P. rubra*, with fewer SSRs, had a smaller genome. This correlation aligns with similar observations in other plant species. The similarities in SSR motif counts between *T. peruviana* and *C. manghas* suggest potential conservation of genomic features or evolutionary connections between these species, making SSRs useful markers for future evolutionary studies of these plants.

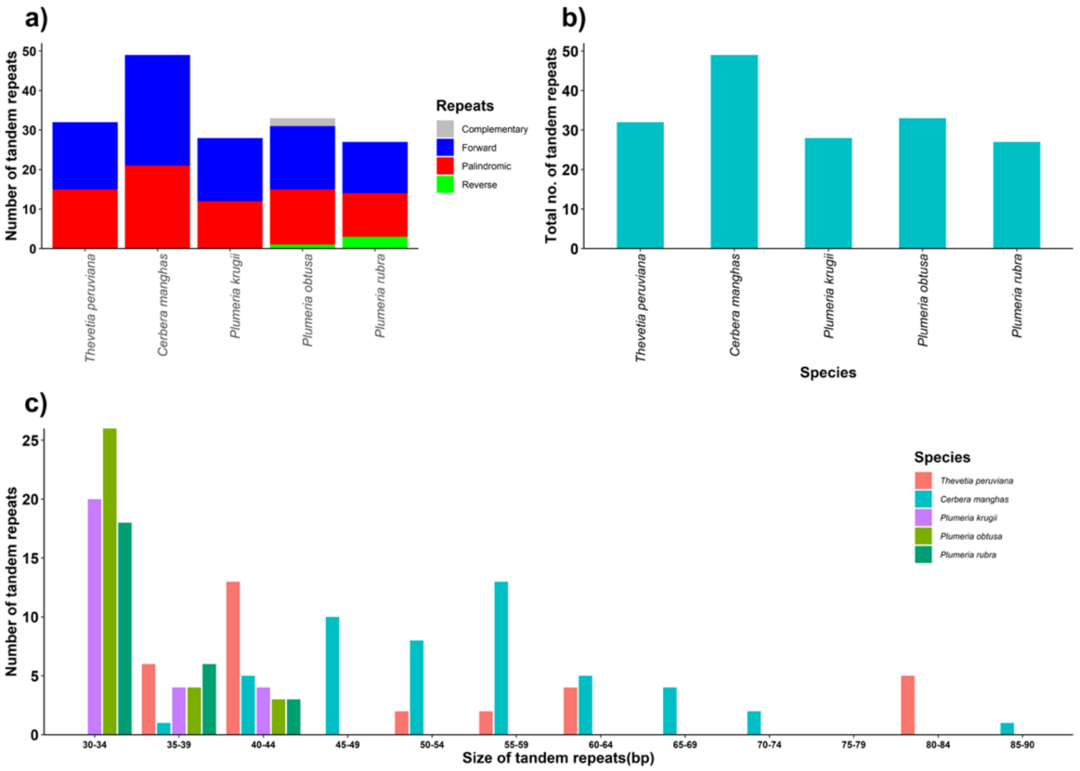


Figure 5.24: Tandem repeats in five Plumerieae species (a) Types of tandem repeats; (b) Total number of tandem repeats, and (c) Number of tandem repeats based on size

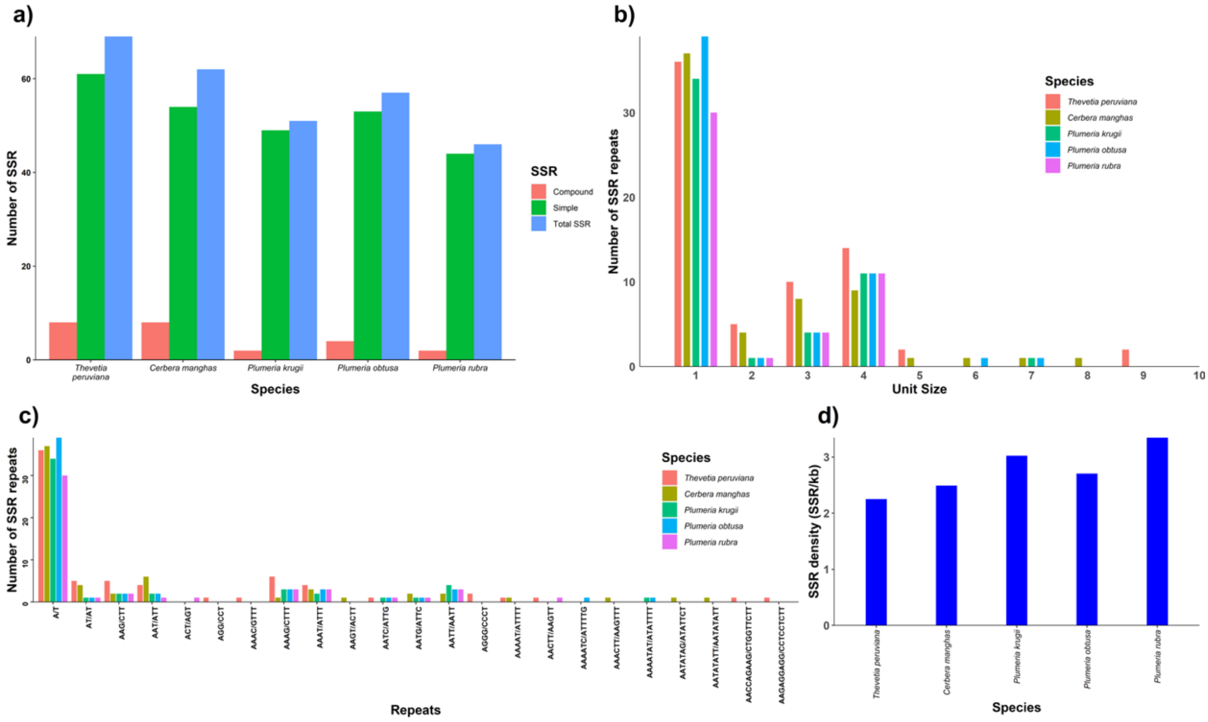


Figure 5.25: SSRs found in five Plumerieae spp. (a) Types of SSR; (b) Number of SSRs based on size; (c) Number of SSRs based on repeat types, and (d) SSR density

5.3.2.4. *DNA diversity analysis (Pi)*

DNA diversity analysis, often measured by the *Pi* value, is a crucial aspect of investigating chloroplast genomes, offering profound insights into the genetic variation within plant populations. The *Pi* value quantifies the average nucleotide diversity among individuals in a given DNA sequence, providing a quantitative measure of the extent of polymorphism (Xiong et al., 2020). In the context of chloroplast genomes, DNA diversity analysis facilitates the understanding of population dynamics, genetic differentiation, and evolutionary processes. The significance of *Pi* values in chloroplast genomes lies in their ability to reveal the historical patterns of speciation, migration, and adaptation within plant species (Li et al., 2018). By unraveling the intricate tapestry of genetic diversity, researchers can gain valuable information about the adaptive potential and resilience of plant populations, aiding in conservation efforts and advancing our comprehension of the dynamic interplay between environmental factors and genomic variation in the plant kingdom. Below, the revelation of DNA diversity in five plants characterized by small genome sizes has been undertaken.

C. quadrangularis

We identified regions of high nucleotide diversity, termed diversity hotspots, in the cp genomes of *Cissus* spp. by assessing nucleotide diversity in coding regions. The *Pi* values for nucleotide diversity in protein-coding genes ranged from 0.0013 (*ndh*B) to 0.0579 (*psa*I), with an average diversity of 0.0191. Noteworthy variations were observed, with *psa*I (0.0579), *psa*J (0.0486), *rpl*32 (0.0439), *ycf*1 (0.0437), and *mat*K (0.03721) representing the most divergent regions, while *ndh*B (0.0013), *rps*7 (0.0027), *pet*N (0.0040), *ycf*15 (0.0048), and *psb*N (0.0050) exhibited the least diversity. This analysis underscored both the highly diverse and conserved nature of coding regions within the cp genome of *Cissus* spp. (**Fig. 5.26**).

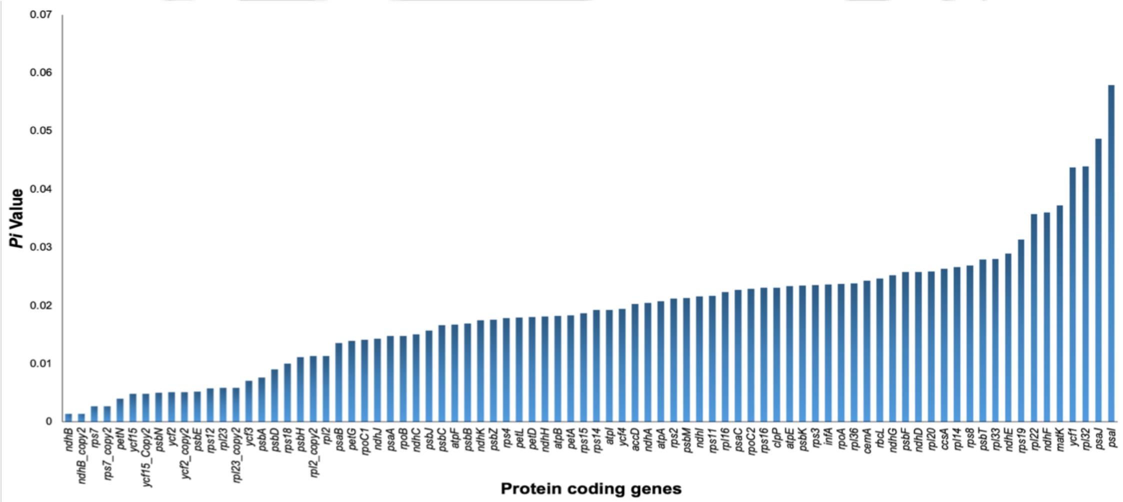


Figure 5.26: Nucleotide diversity (*Pi*) values of protein coding regions in *Cissus* species

Examining the sequence similarity among *Cissus* cp genomes revealed variations in coding region divergence, with the IR regions displaying more conservation compared to the LSC and SSC regions. This observation aligns with patterns seen in other angiosperms (Guan et al., 2022; Shelke and Rangan, 2022). Consistently, the analysis of DNA diversity indicated that coding regions were generally more conserved than non-coding regions (Rossini et al., 2021). Recognizing the challenges associated with universal DNA barcodes in resolving species-level distinctions in many angiosperms, the highly diverse regions identified here hold promise as potential genus-specific DNA barcodes for distinguishing *Cissus* spp. in future studies.

F. hispida

Within the scope of this investigation focused on *Ficus* cp genomes, the delineation of hotspot regions was realized through meticulous analysis of nucleotide diversity encompassing both coding and non-coding segments. An array of nucleotide diversity (*Pi*) values were unveiled across the protein-coding genes, spanning from a minimum of 0.0002, exemplified by *atp*I, to a maximum of 0.1042, represented by *rps*19. The collective nucleotide diversity, indicative of the average across protein-coding regions, was quantified at 0.0047. Among the pool of coding regions, a quintet of notably divergent segments emerged. Specifically, these encompassed *rps*19 (0.1042), *rbc*L (0.026), *rps*16 (0.0156), *rpl*16 (0.0086), and *pet*D (0.0056). Conversely, the coding regions exhibiting the least diversity were identified as *atp*I (0.0002), *psb*B (0.00033), *atp*A (0.00034), *ycf*2_copy (0.00039), and *ycf*2 (0.00049). These findings underscore the nuanced variability present across the coding domains of *Ficus* cp genomes, thereby enriching our comprehension of the genetic landscapes within the genus (**Fig. 5.27**)

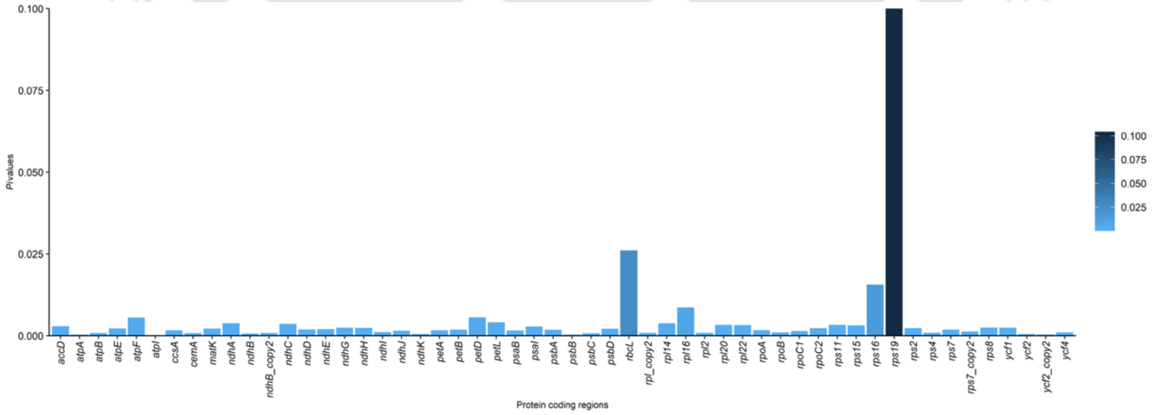


Figure 5.27: Nucleotide diversity (*Pi*) values of *Ficus* spp. among protein-coding regions

S. rutidosperma

In this study on Cleomaceae chloroplast genomes, hotspot regions were identified based on nucleotide diversity analysis of coding regions. The protein-coding genes exhibited a range of nucleotide diversity (*Pi*) values, from zero for *ycf*3 to 0.1452 for *ycf*1. The average diversity of protein-coding regions was calculated to be 0.0241. Among the five most divergent coding regions were *ycf*1 (0.1452), *clp*P (0.1214), *mat*K (0.0523), *rpl*32 (0.0476), and *ccs*A (0.0409). On the other hand, the least diverse coding regions were *ycf*3 (0), *rpl*2 (0.0029), *ndh*B (0.0034), *rps*7 (0.0061), and *rps*12 (0.0064) (**Fig. 5.28**).

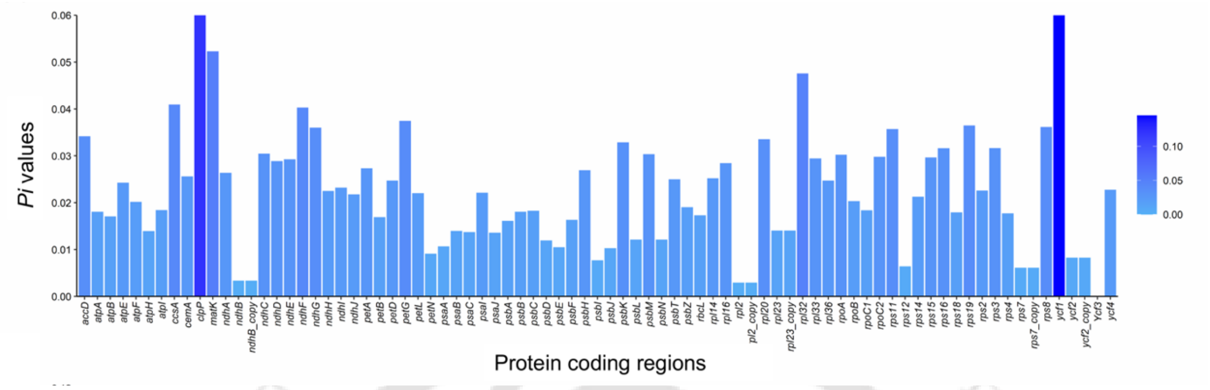


Figure 5.28: Nucleotide diversity (*Pi*) values of seven Cleomaceae spp. in protein-coding regions

The cp genomes of Cleomaceae species displayed significant sequence similarity. As is common in angiosperms, the IR regions being more conserved than the LSC and SSC regions (Shelke et al, 2022). This study identified the most divergent coding regions in the Cleomaceae cp genome. However, certain regions of the cp genome evolve at a much faster pace and, therefore, qualify as DNA barcodes (Guan et al. 2022). These highly diverse regions can be used as specific DNA barcodes to discriminate between Cleomaceae plants.

S. asper

Our investigation into the genetic diversity, particularly nucleotide diversity (*Pi*), across coding regions in seven *Streblus* spp. aimed to unravel the intricate evolutionary dynamics of their plastome. The analysis brought to light substantial variations in *Pi* values, indicative of diverse evolutionary trajectories within these genomes. In scrutinizing protein-coding genes, a spectrum of *Pi* values was observed. Certain genes, such as *psb*L, *pet*N, *rpl*23, and *psb*F, exhibited complete conservation (*Pi* = 0), underscoring the high degree of sequence uniformity. Conversely, genes like *rps*19 emerged as hotspots for genetic diversity, boasting the highest *Pi* value of 0.032. On average, the nucleotide diversity for protein-coding regions across these genomes was computed at 0.0123. The identification of the five most divergent coding regions- *rps*19 (0.032), *rps*8 (0.031), *rpl*22 (0.029), *mat*K (0.025), and *rps*11 (0.024), accentuates pockets of substantial genetic heterogeneity within the *Streblus* plastome. Conversely, certain protein-coding regions, including *rpl*2 (0.002), *rps*7 (0.003), *ndh*B (0.0032), *rps*18 (0.0045), and *atp*H (0.005), displayed low diversity, suggesting their preservation and conservation across these plastid genomes **(Fig. 5.29).**

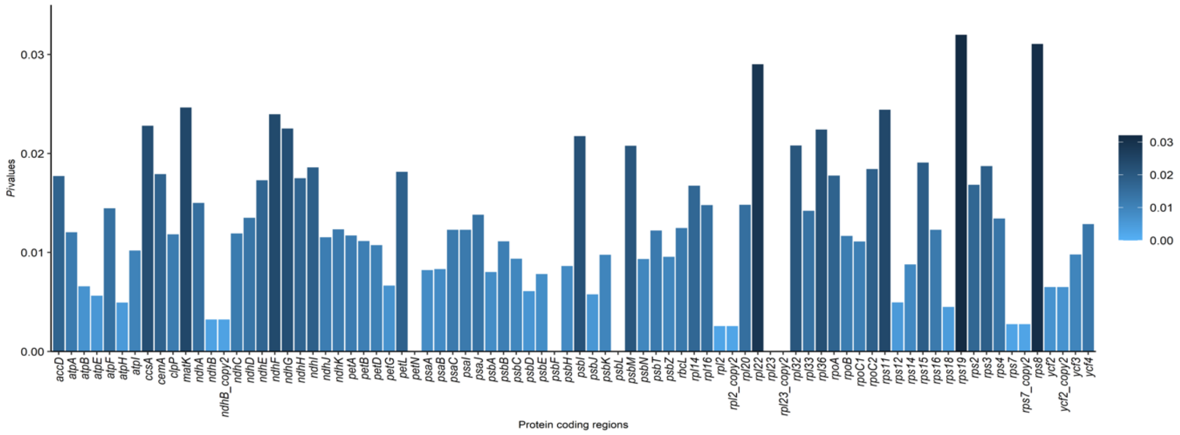


Figure 5.29: Nucleotide diversity (*Pi*) among protein coding regions in seven *Streblus* spp.

In considering the implications of these results, the observed genetic diversity signifies an intricate interplay of evolutionary forces shaping the plastome of *Streblus* spp. The genes with elevated *Pi* values may be indicative of regions undergoing adaptive evolution, responding to selective pressures in their respective ecological niches. Conversely, genes with low *Pi* values suggest regions under purifying selection, emphasizing their functional significance and conservation across the *Streblus* spp. similar to observations in *Broussonetia* spp. (Yang et al., 2022). The identification of highly conserved and divergent coding regions provides a nuanced perspective on the adaptive potential and genomic stability within *Streblus* cp genomes (Kong et al., 2017). These findings lay a foundation for further investigations into the functional implications of genetic diversity in the context of the species' ecological adaptations and evolutionary resilience.

T. peruviana

We delved into the genetic diversity, specifically nucleotide diversity (*Pi*), across coding regions in Plumerieae spp. to understand the evolutionary dynamics of their plastid genomes. Our analysis uncovered significant variations in *Pi* values, indicating the diverse evolutionary paths within these genomes. In protein-coding genes, we observed a wide range of *Pi* values. Some genes, like *psb*L, showed complete conservation (*Pi* = 0), while others, particularly *acc*D, exhibited the highest diversity with a *Pi* value of 0.130. On average, the nucleotide diversity for protein-coding regions in these genomes was calculated to be 0.019. Among the most divergent coding regions were *rps*18 (0.046), *clp*P1 (0.055), *ycf*1 (0.073), *rps*19 (0.075), and *acc*D (0.130), showcasing substantial genetic diversity. Conversely, certain protein-coding regions displayed low diversity, indicating their conservation across these plastid genomes. Notable examples include *psb*L (0), *psb*I (0.002), *rpl*23 (0.003), *rpl*23_copy (0.003), and *rpl*2 (0.003) (**Fig. 5.30**).

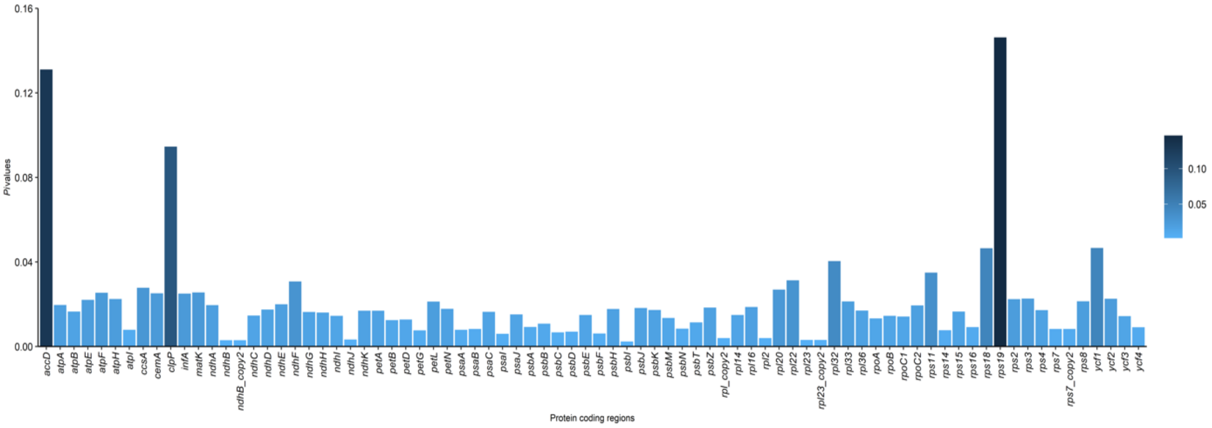


Figure 5.30: Nucleotide diversity (*Pi*) among protein coding regions in five Plumerieae spp.

In the thorough analysis of Plumerieae tribe cp genomes, we gained intriguing insights into the evolutionary dynamics, particularly in terms of nucleotide diversity across coding regions. Our results confirmed a wide spectrum of nucleotide diversity values within coding genes. Some regions, like *rps*19 (0.079) and *acc*D (0.130), exhibited significant diversity, emphasizing variable rates of evolution in these genes. On the other hand, conserved coding regions like *psb*L demonstrated low or zero *Pi* values. These diverse *Pi* values highlight the complex dynamics of plastid protein-coding gene evolution, aligning with previous research findings (Zhang et al., 2020).

5.3.2.5. *Nucleotide substitution rate (Ka/Ks)*

Investigating nucleotide substitution rates, represented by the Ka/Ks (Dn/Ds) ratio, within chloroplast genomes is crucial for understanding plant molecular evolution. This ratio compares the rates of non-synonymous (Ka) to synonymous (Ks) nucleotide substitutions, offering insights into the selective pressures on protein-coding genes (Xiong et al., 2020). Studying Ka/Ks in cp genomes reveals the nature and magnitude of natural selection, where a ratio greater than 1 (Ka/Ks > 1) indicates adaptive evolution, and a ratio less than 1 (Ka/Ks < 1) suggests purifying selection, reflecting functional conservation (Zhao et al., 2015).

Analysing these substitution rates allows to uncover the molecular mechanisms driving chloroplast genome evolution, identify genes under selective pressure, and understand the adaptive strategies of plant species over time (Zhu et al., 2021). High Ka/Ks ratios indicate regions of the genome where rapid evolutionary changes are occurring, potentially identifying genes critical for adapting to environmental factors like light, temperature, or pathogen resistance. Conversely, low Ka/Ks ratios are indicative of genes involved in essential functions such as photosynthesis, which are conserved to maintain their integrity.

The Ka/Ks analysis of five selected plant species demonstrated varied evolutionary patterns, highlighting both conserved and rapidly evolving regions within the chloroplast genome. This analysis significantly enhances our understanding of plant molecular evolution and aids in identifying genes essential for adaptation and survival. Such knowledge is vital for developing conservation strategies and improving agricultural practices, thereby enhancing crop resilience and productivity in response to environmental changes and global challenges.

C. quadrangularis

The assessment of the Ka/Ks ratio, reflecting the selection pressure on genes, was conducted for 82 common genes present in the cp genomes of six *Cissus* spp. Among these genes, 12 displayed a Ka/Ks ratio of 0 or had no result, while 70 exhibited a Ka/Ks ratio between 0 and 1. Three genes (*atp*F, *rpl*2, and *rpl*2_copy) showed more than one ratio. The majority, 67 genes, had a Ka/Ks ratio less than one (**Fig. 5.31**). The range of Ka/Ks ratios varied from 0.009 (*rps*11) to 1.077 (*rpl*2) in *Cissus* spp. Notably, most genes demonstrated a Ka/Ks ratio less than one, indicating purifying selection, where the selection works to preserve the function of these genes. A Ka/Ks ratio greater than 1 suggests positive selection, while a ratio equal to 1 implies neutral selection. Some genes exhibited zero or non-existent Ka/Ks ratios, often associated with extremely low Ks values or the absence of substitutions in matched sequences.

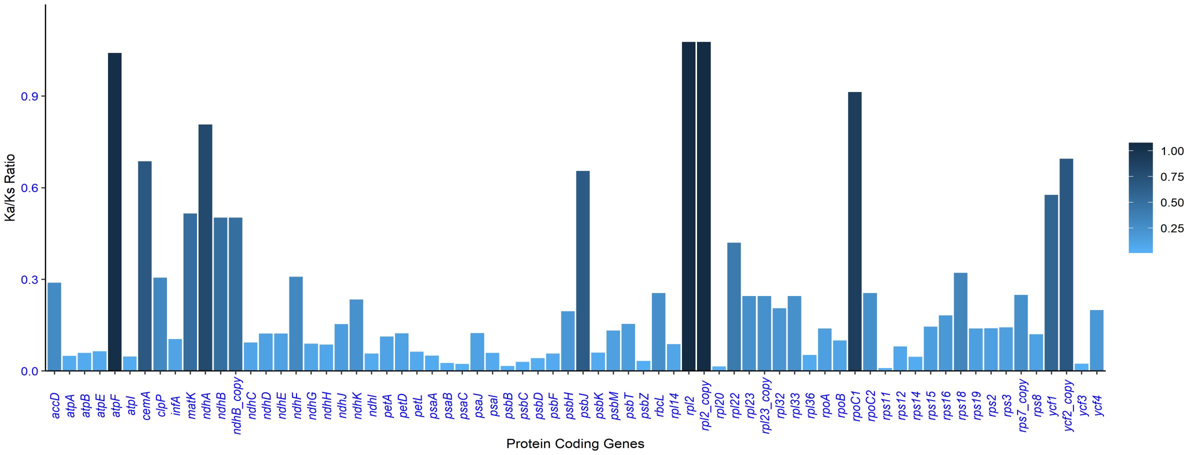


Figure 5.31: Comparison of Ka/Ks ratio generated in protein-coding genes aligned from six *Cissus* species cp genomes

The Ka/Ks ratio analysis of 82 common genes in the chloroplast genomes of six *Cissus* spp. reveals a strong trend towards purifying selection, with 67 genes exhibiting Ka/Ks ratios less than one. This indicates that these genes are under strong selective pressure to maintain their function, which is typical for essential genes in chloroplast genomes (Shidhi et al., 2021). The Ka/Ks ratio range observed in *Cissus* spp., aligns with findings in majority of the plant species where chloroplast genes generally show low Ka/Ks ratios due to the crucial roles they play in photosynthesis and other metabolic processes. For instance, similar studies in various angiosperms, such as Arabidopsis and rice, have reported Ka/Ks ratios predominantly below one for most chloroplast genes, further corroborating the prevalence of purifying selection in maintaining chloroplast function and stability (Chen et al., 2017).

Comparatively, in other plant species, the Ka/Ks ratio distribution often mirrors the pattern observed in *Cissus* spp. For example, research on the chloroplast genomes of several legume species and grasses has consistently shown that the majority of genes are subject to purifying selection, with Ka/Ks ratios significantly below one. However, the occasional presence of genes with Ka/Ks ratios greater than one, such as the *rpl*2 gene in *Cissus* spp., suggests episodes of positive selection that may be associated with adaptive changes in response to specific environmental pressures or evolutionary events. This pattern of selection is not unique to *Cissus* and has been documented in other taxa, including the adaptive evolution of certain chloroplast genes in response to high-altitude environments in Tibetan barley and alpine plants. These comparisons highlight the evolutionary dynamics within chloroplast genomes and the balance between conserving essential functions and adapting to ecological niches.

F. hispida

A comprehensive analysis of the 87 protein-coding genes was undertaken, yielding Ka/Ks ratio calculations for 49 of these genes. For the remaining genes, the computation was unfeasible due to a Ks value of zero. The resultant mean Ka/Ks ratio across the protein-coding genes within the *Ficus* cp genomes stood at 0.354. Notably, the scrutiny of these ratios unveiled notable patterns. Among the 49 genes examined, 30 manifested a Ka/Ks ratio below 1, signifying a robust influence of purifying selection upon these genes. Remarkably, a single gene, *rps*19, exhibited a Ka/Ks ratio surpassing unity at 1.044. This observation indicates a likelihood of positive selection acting upon this specific gene. Intriguingly, a Ka/Ks ratio of exactly 1 was not evident across any of the genes assessed. Such an observation is indicative of neutral evolution. These insights contribute substantively to our comprehension of the evolutionary dynamics influencing protein-coding genes within *Ficus* cp genomes (**Fig. 5.32**).

The comprehensive analysis of protein-coding genes within the *Ficus* chloroplast genomes revealed insightful evolutionary dynamics through the calculation of Ka/Ks ratios for 49 of these genes. Due to zero Ks values, the Ka/Ks ratio could not be computed for the remaining genes. The overall mean Ka/Ks ratio for the protein-coding genes was determined to be 0.354, reflecting a predominant influence of purifying selection aimed at preserving the essential functions of these genes. Among the 49 genes analyzed, 30 displayed Ka/Ks ratios below 1, underscoring the strong purifying selection pressure that acts to eliminate deleterious mutations and maintain the genetic integrity of these chloroplast genes.

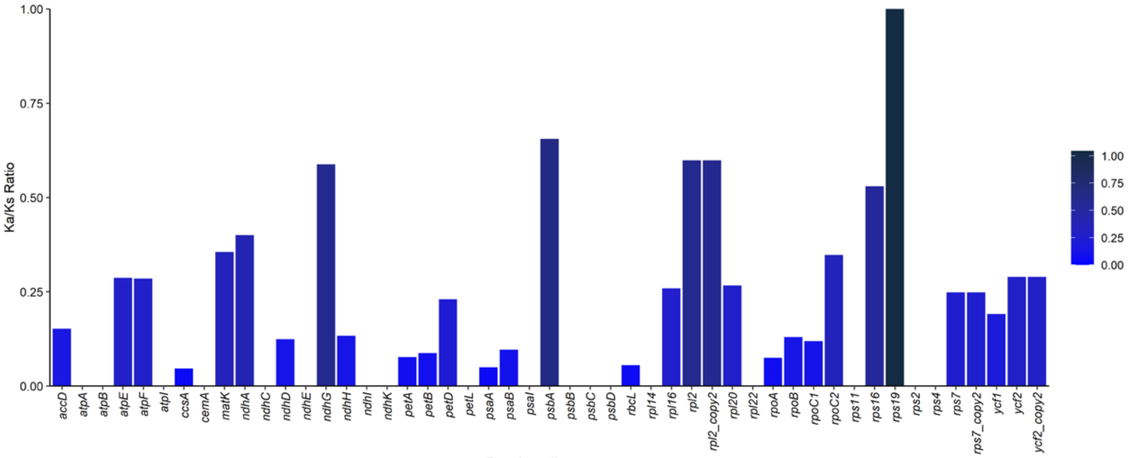


Figure 5.32: Comparison of Ka/Ks ratio generated in protein-coding genes from cp genome alignment of four *Ficus* spp.

Interestingly, the gene *rps*19 stood out with a Ka/Ks ratio of 1.044, suggesting a potential instance of positive selection. This ratio indicates that this gene may be undergoing adaptive changes, possibly in response to specific environmental pressures or unique functional requirements within the *Ficus* lineage. The absence of genes with a Ka/Ks ratio exactly equal to 1 suggests that none of the examined genes are evolving neutrally, highlighting the lack of genetic drift influence within these protein-coding sequences. These findings align with similar studies in other plant species, where purifying selection is the dominant force acting on chloroplast genes, but occasional positive selection events are observed, reflecting the complex evolutionary pressures shaping chloroplast genome evolution. This comprehensive analysis provides valuable insights into the selective forces acting on the chloroplast genomes of *Ficus*, contributing to a broader understanding of plant genomic evolution.

S. rutidosperma

The Ka/Ks ratio is an important metric for understanding the selective pressures acting on genes in a chloroplast genome. In this study, out of 82 protein-coding genes we calculated the Ka/Ks ratio for 45 protein-coding genes and for 37 genes could not be computed. The mean Ka/Ks ratio for 45 protein-coding genes in the *S. rutidosperma* versus the other six cp genomes (Cleomaceae) was 1.0999. Our analysis revealed that 28 genes out of the 45 genes had a Ka/Ks ratio of less than 1, indicating that these genes are under strong purifying selection. Additionally, we found that 17 genes had a Ka/Ks ratio of greater than 1, suggesting that these genes have experienced positive selection. A ratio (Ka/Ks) of exactly 1 was observed in case of no genes, but in case of *psa*J, *psb*Z, and *rpl*32 the ratio was close to 1 which is indicative of neutral evolution (**Fig. 5.33**).

These findings are consistent with previous studies (ALJuhani and Aljohani, 2022; Fan and Ma, 2022; Ren et al., 2022) on the evolution of chloroplast genomes, which have suggested that the majority of genes in these genomes are under strong purifying selection. However, our study also identified a number of genes that have experienced positive selection, which may indicate important adaptations to the environment or interactions with other organisms as reported in Saxifragaceae and Brassicaceae (Wu et al., 2020; Zhu et al., 2021). In conclusion, our analysis of the Ka/Ks ratio in the cp genome of seven Cleomaceae species highlights the importance of selective pressures on shaping evolution of these genes. Further research is needed to elucidate the specific mechanisms driving positive selection in some genes and explore functional implications of these adaptations.

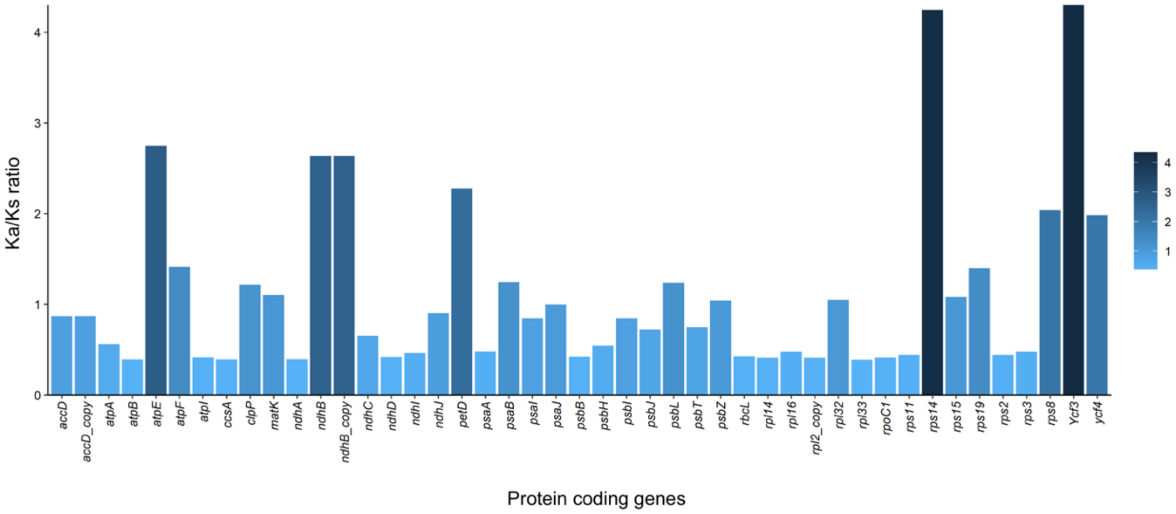


Figure 5.33: Comparison of Ka/Ks ratio generated for protein-coding genes from seven aligned Cleomaceae spp. cp genomes

S. asper

The investigation of protein-coding genes within the cp genomes of seven *Streblus* spp., we computed Ka/Ks ratios to elucidate the equilibrium between non-synonymous (Ka) and synonymous (Ks) substitutions, thereby shedding light on the selective pressures acting on individual genes. Noteworthy is the observation that certain genes, such as *atp*H, *pet*G, *psa*I, *psb*I, and *rpl*36, exhibit a Ka/Ks ratio of 0, indicative of neutral selection and evolution consistent with genetic drift. Conversely, only the gene *rps*14 displayed a Ka/Ks ratio greater than 1, suggesting positive selection and adaptive evolution. Our comprehensive analysis further unveiled that 64 out of 76 examined genes manifested Ka/Ks ratios ranging from zero to one, pointing to robust purifying selection influencing the evolution of these genes. Notably, the absence of genes with a Ka/Ks ratio exactly equal to one in *Streblus* spp. suggests the scarcity of instances of neutral evolution within this genomic context (**Fig. 5.34**).

In the broader scientific context, the predominance of genes under strong purifying selection underscores the significance of functional constraints in shaping the evolutionary dynamics of *Streblus* spp. cp genomes. The identification of genes undergoing positive selection, as exemplified by *rps*14, provides valuable insights into adaptive evolution; highlighting potential roles these genes may play in responding to environmental pressures (Chen et al., 2017). The absence of genes exhibiting exactly neutral evolution, coupled with the prevalence of purifying selection, accentuates the overall conservation of functional elements within the examined cp genomes (Zhu et al., 2021). These findings contribute to our understanding of the molecular evolution of *Streblus* spp., with implications for both evolutionary biology and plant adaptation strategies.

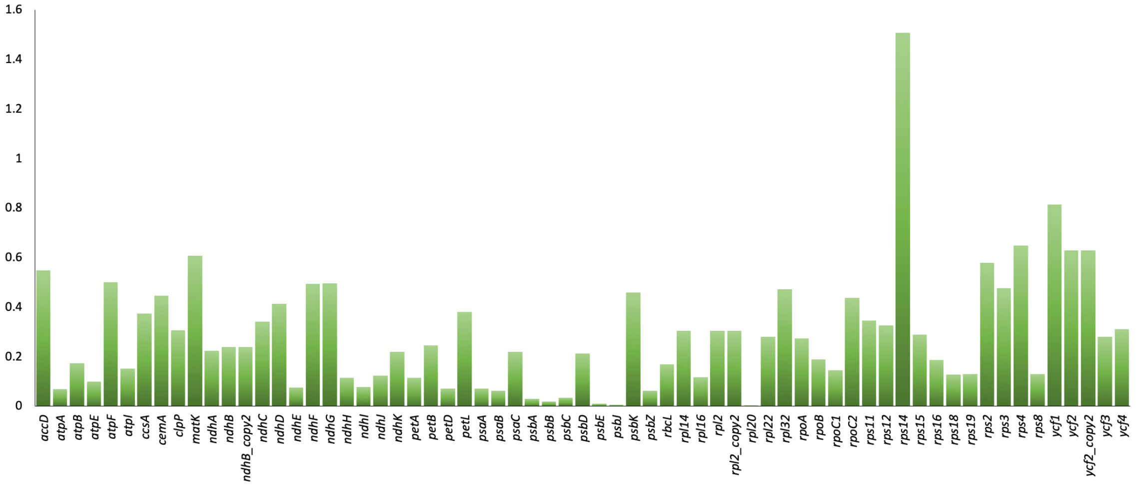


Figure 5.34: Nucleotide substitution rate (Ka/Ks value) in coding regions of seven *Streblus* spp.

T. peruviana

The Ka/Ks ratios for different protein-coding genes ware calculated among the cp genomes of five selected plants. These ratios help us to understand the balance between non-synonymous (Ka) and synonymous (Ks) substitutions and reveal the selective pressures on individual genes. Some genes showed a Ka/Ks ratio of 0, indicating neutral selection. This suggests that these genes evolve at a rate consistent with genetic drift. Examples of these genes include *pet*G, *pet*N, *psa*C, *psb*D etc. On the other hand, several genes had Ka/Ks ratios greater than 1, suggesting positive selection and adaptive evolution. Genes like *mat*K (1.479), *rps*18 (1.832), *ycf*2_copy (2.127), and *ycf*2 (2.200) are likely undergoing adaptive evolution due to selective pressures **(Fig. 5.35).**

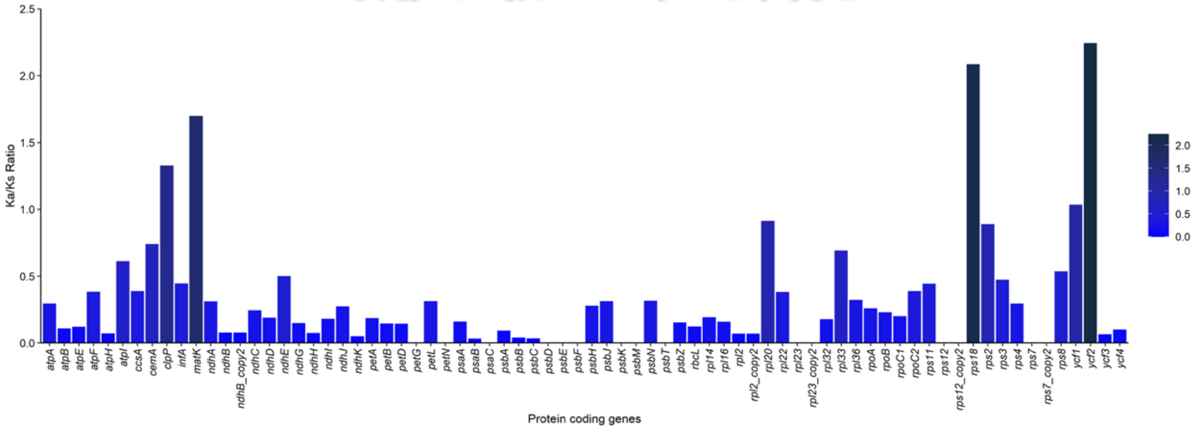


Figure 5.35: Nucleotide substitution rate (Ka/Ks value) in coding regions of five *Plumerieae* spp.

The analysis revealed that 53 out of 76 genes had Ka/Ks ratios between 0.030 and 0.914, indicating strong purifying selection acting on these genes. Genes with a Ka/Ks ratio of exactly one, such as *psb*N, *rps19*_copy, *ycf*3, and *ycf*4, suggest neutral evolution where synonymous and non-synonymous substitutions occur at an equal rate. The Ka/Ks ratios provide insights into the selective pressures on protein-coding genes in these cp genomes. Some genes evolve neutrally, while others undergo positive selection and adaptive evolution. Important genes like *mat*K, *rps*18, *ycf*2_copy, and *ycf*2 are actively adapting. Many genes are under purifying selection, indicating their crucial functions and evolutionary conservation.

5.3.2.6. *Codon usage bias*

The investigation of codon usage bias in cp genomes represents a crucial aspect of molecular evolution, shedding light on the intricate patterns of codon preferences within the protein-coding genes of these organelles. Codon usage bias refers to the uneven frequency with which synonymous codons are employed during protein translation (Tang et al., 2021). Understanding this bias in chloroplast genomes holds significant implications for unraveling the molecular mechanisms governing gene expression and evolution. The intricate interplay between codon usage patterns and various factors, including mutational pressures, selection, and genomic features, contributes to the fine-tuning of protein synthesis (Nie et al., 2014). Analysing codon usage bias provides valuable insights into the adaptive strategies employed by chloroplast genomes, aiding in the identification of genes crucial for environmental adaptation, evolutionary dynamics, and the optimization of protein production (Li et al., 2019). We studied codon usage bias in five small genome sized plants to understand the intricate molecular processes shaping cp genomes and their functional significance in plant biology below.

C. quadrangularis

The codon usage analysis revealed a consistent and pronounced preference for A/T(U) bases across all three positions in the codons studied. The GC content at each position was consistently below 40%, indicating a bias towards A/T richness. Leucine (L) emerged as the most frequently used amino acid, ranging from 9.01% to 10.85%, while tryptophan (W) (1.89% - 2.11%), and serine (S) (1.73% - 2.34%) were the least abundant. The total number of codons analysed amounted to 18,368 in *C. quadrangularis*.

The observed codon usage patterns, characterized by a strong preference for A/T(U) bases and low GC content, suggest a compositional bias in the chloroplast genomes of the studied species. This bias in base composition is a common feature in plant cp genomes, and it can be attributed to the high abundance of A/T-rich codons, reflecting the bias in synonymous codon usage (Qi et al., 2015). The over representation of leucine, a hydrophobic amino acid commonly found in cp proteins, may be associated with the specific functional and structural requirements of chloroplast-encoded proteins (Parvathy et al., 2022). Notably, differences in codon usage were observed among the studied species, where *C. discolour* utilized 28 codons, while *C. antarctica* and *C. tuberosa* employed 30 codons, with others using 29 codons in excess (RSCU > 1) (**Fig. 5.36**). This variation in codon usage among closely related species may be influenced by factors such as selection pressure, mutational biases, and the functional constraints of chloroplast-encoded genes (Gao et al., 2022). This could provide insights into the adaptive evolution and functional divergence within the chloroplast genomes of these related species as suggested by Zhang et al. (2021).

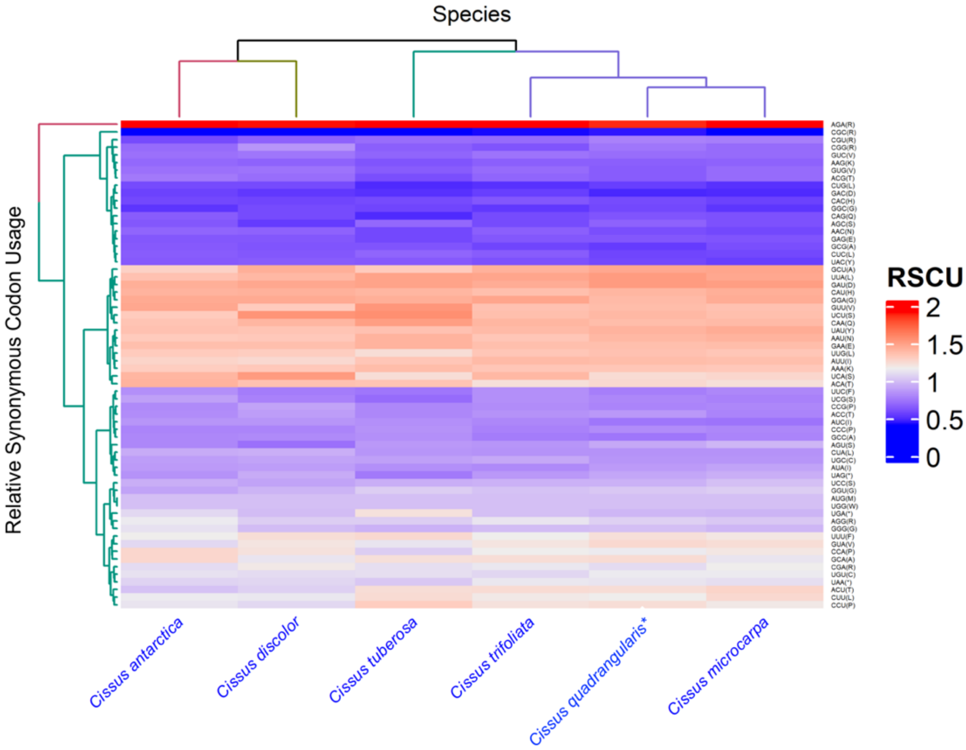


Figure 5.36: Codon usage bias of six *Cissus* spp. in terms of relative synonymous codon usage (RSCU)

F. hispida

The codon usage bias was investigated among the four *Ficus* spp. by analyzing the GC content and RSCU values. A similar observation to that of *Cissus* spp. revealed codons ending in A/T(U) were more frequent than those ending in GC at all three codon positions, resulting in an overall average GC content of less than 40% in all cases (Li et al. 2019). The GC1 ranged from 37.59% to 40.66% in *F. hispida* and *F. racemosa* respectively, whereas the GC2 was lowest in *F. squamosa* (35.83%) and the highest in *F. hispida* (40%). The highest GC content for the GC3 position was found in *F. squamosa* (38.22%), whereas the lowest concentration was found in *F. racemosa* (34.87%) (**Fig. 5.37**).

Highest number of total codons were observed in *F. auriculata* (19,968), followed by *F. squamosa* (19,584), and *F. racemosa* (18,944). The least number of total codons were found in *F. hispida* (18,688). Leucine (8.68–11.1%) was the most prevalent amino acid similar to previous analysis, followed by serine (7.9–9.6%). The least used codon was found to be tryptophan (1.85-2.09%).

Tryptophan and methionine had only one codon each and no preference for codon usage. Conversely, the remaining amino acids contained more than one synonymous codons. *F. racemosa* used 31 codons, whereas *F. auriculata* and *F. squamosa* used 30 codons, and *F. hispida* used 29 codons more frequently than expected at equilibrium (RSCU > 1). In all the species, the start codons AUG and UGG, which encode for the amino acids methionine and tryptophan, were utilized completely with an RSCU value of 1, but in case of *F. hispida* the codon GGU coding for guanine and in case of *F. squamosa* UCC coding for serine were also used as expected at equilibrium (RSCU=1).

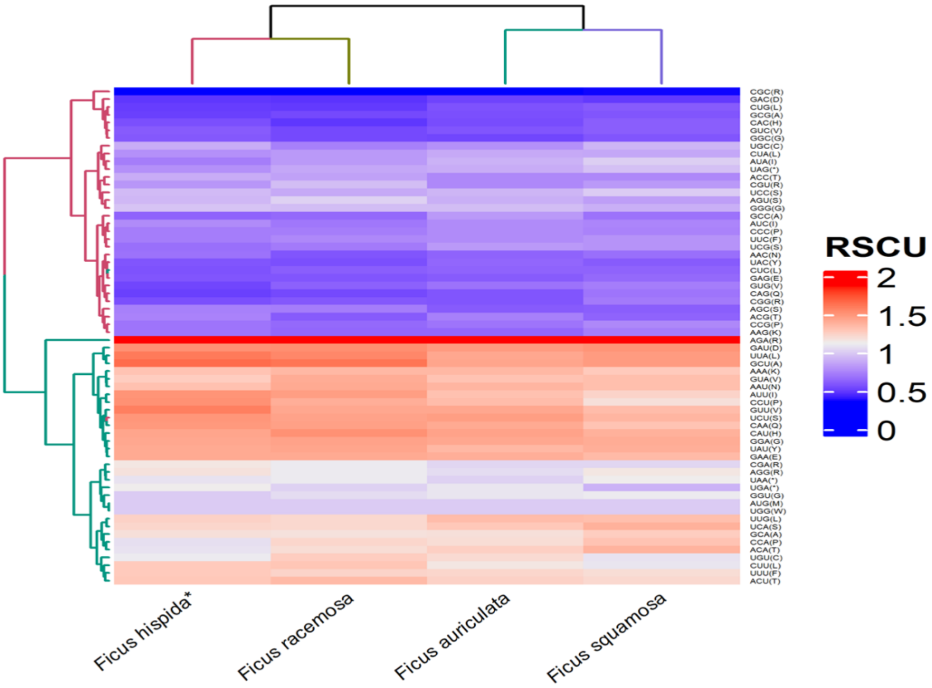


Figure 5.37: Codon usage bias in terms of relative synonymous codon usage (RSCU) for four species of *Ficus* spp.

S. rutidosperma

The codon usage bias was investigated in seven species by analyzing the GC content and RSCU values. Similar to previous findings A/T(U) ending codons were more frequent than those ending in GC and an overall average GC content of less than 40% was observed in all cases (Li et al. 2017). The GC1 ranged from 36.99% to 40.44%, whereas the GC2 was lowest in *T. chrysantha* (35.53%) and the highest in *G. gynandra* (39.08%). The highest GC content for the GC3 position was found in *T. chrysantha* (40.28%), whereas the lowest concentration was found in *G. gynandra* (34.04%). Leucine (7.99–8.01%) was the most prevalent amino acid similar to previous plants, followed by arginine and serine (5.99–6.01%). Tryptophan and methionine followed same pattern as in *Ficus* spp. The remaining amino acids used multiple synonymous codons where *C. pallida* used 29 codons, whereas *T. hassleriana*, *S. rutidosperma*, *G. gynandra*, and *C. lutea* used 30 codons more frequently (RSCU > 1). *C. paradoxa* and *T. chrysantha* used 31 and 33 codons more frequently (RSCU > 1). Similar to *Cissus* spp. and *Ficus* spp. Cleomaceae spp. used the start codons AUG and UGG completely without bias, which was confirmed by an RSCU value of 1 (**Fig. 5.38)** (AL-Juhani et al., 2022).

All seven Cleomaceae spp. used the same set of codons, with 61 amino acid codons, three stop codons, and one start codon. However, the number and type of codons that encode the 20 amino acids differ among species (Han et al., 2022). Similar findings were also reported in *Physalis* spp. (Feng et al., 2020). A preference for A or T at the third nucleotide position was observed for most codons encoding amino acids in all Cleomaceae spp., as well as in various other angiosperm families such as Cucurbitaceae, Fabaceae, Poaceae, Asteraceae, and Euphorbiaceae spp. (Nie et al. 2014; Somaratne et al., 2019; Ma et al. 2020; Sheng et al., 2021; Zhang et al., 2021). The frequency of codon usage varies in cp genomes of Cleomaceae, which may be attributed to gene selection, mutation, genetic drift, base substitution, and availability of tRNA (Somaratne et al., 2019; Feng et al., 2020; Han et al., 2022). This study also followed a similar pattern to previous studies suggesting GC-rich codons are preferred by monocots, while AT-rich codons are frequently used by dicots (Wang and Roossinck, 2006). Codon mutations in the third position are usually synonymous and do not alter the amino acid composition. However, mutations in the first and second positions are non-synonymous and can affect the encoded amino acid. The pattern of codon usage observed in Cleomaceae plants is in agreement with the codon usage of Asparagaceae*,* *Gynostemma* spp., and *Iris* spp. (Lee et al., 2019; Zhang et al., 2021; Feng et al., 2022)

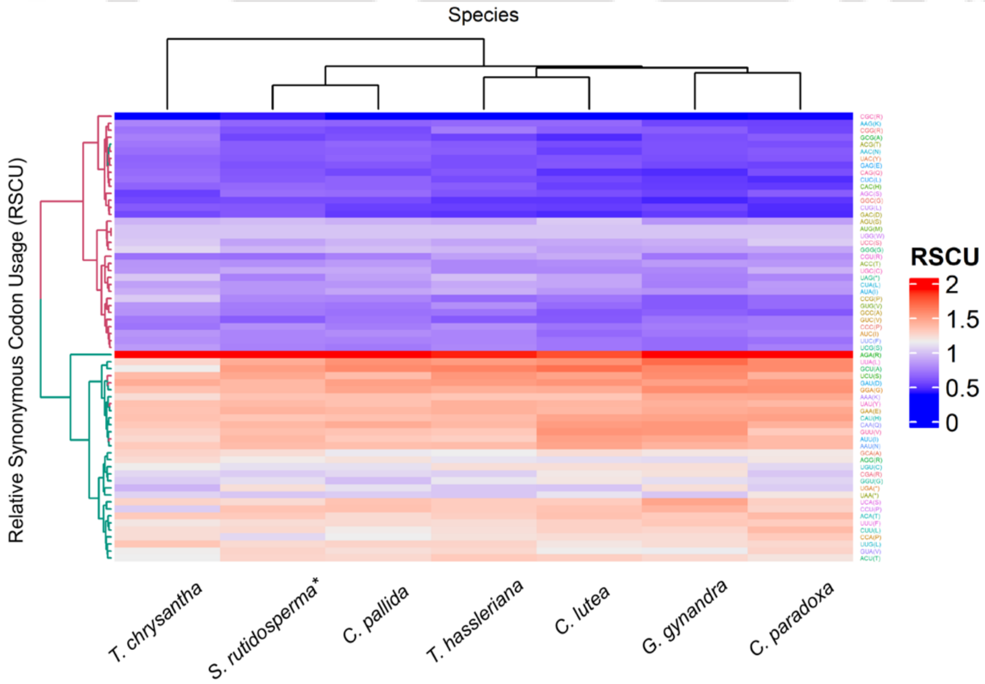


Figure 5.38: Codon usage bias in seven species of *Cleomaceae* in terms of relative synonymous codon usage

S. asper

Our investigation into the plastome of seven *Streblus* spp. involved a thorough analysis of 136,832 codons, revealing distinctive patterns in codon utilization. The GC% content across all codons exhibited variations, ranging from 37.15% in *S. asper* to 37.66% in *S. heterophyllus*. Notably, GC1, GC2, and GC3 values remained consistently below 40%, except for GC3 in *S. heterophyllus*, which registered at 40.87%. Singular codons, UGG and AUG, encoded tryptophan (W) and methionine (M), respectively, indicating an absence of codon preference for these specific amino acids. In contrast, a range of two to six codons encoded various amino acids, with particular emphasis on arginine (R), leucine (L), and serine (S) having six codons each. The discerned patterns in codon usage underscored the prevalence of certain amino acids, with R being notably prominent, and W identified as the least abundant. Codons encoding R consistently exhibited the highest RSCU values across species, with a peak of 2.24 (AGA) in *S. asper* and a nadir of 0.31 (CGC) in *S. banksii*. Furthermore, codons UUA (L), UCU (S), AUU (I), GAU (D), and GCU (A) consistently demonstrated RSCU values exceeding 1 across all species. Noteworthy is the equal usage of initiation codons AUG and UGG, with no preference of codon usage exhibited an RSCU value of 1 across all species **(Fig. 5.39).**

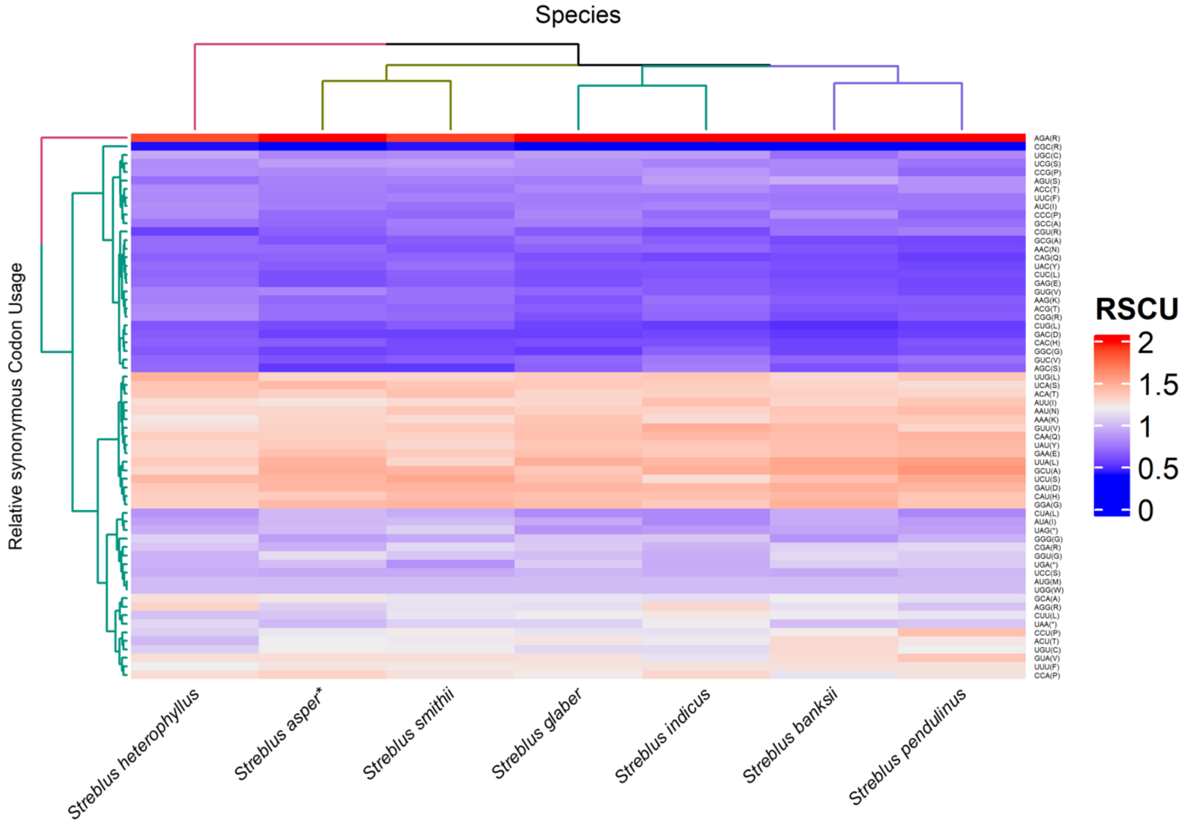


Figure 5.39: Codon usage bias among seven *Streblus* spp. through RSCU values

In the context of codon usage bias in *Streblus* spp., the observed patterns reflect distinctive preferences in the utilization of specific codons, particularly for amino acids such as R, L, and S (Tang et al., 2021). The consistency in RSCU values for initiation codons suggests a uniformity in the usage of M and W across the studied species suggesting a conserved translational initiation mechanisms as observed in most of the angiosperms (Li et al., 2018). The variations in codon usage patterns may be indicative of evolutionary forces shaping the plastome, potentially influenced by evolutionary pressures, mutation rates, and the need for efficient translation (Li et al., 2019; Parvathy et al., 2022). Further exploration into the functional implications of these codon usage patterns could provide valuable insights into the molecular adaptation and evolution of *Streblus* spp.

T. peruviana

The comprehensive examination of the plastome of *T. peruviana* involved the analysis of 51,716 codons, elucidating distinctive patterns in codon utilization. Remarkably, both tryptophan (W) and methionine (M) were encoded by singular codons, suggesting an absence of codon preference for these specific amino acids. In contrast, several other amino acids were encoded by a spectrum of two to six codons, with notable emphasis on arginine (R), leucine (L), and serine (S). The observed codon usage patterns underscored the prevalence of amino acids such as R, L, and S in *T. peruviana*, with tryptophan being identified as the least abundant among the amino acids. Significantly, the codon AGA, which codes for R exhibited the highest RSCU value across species, reaching 1.86 in *C. manghas*, while the codon CGC (encodes R) recorded the lowest RSCU value, a mere 0.36 in *P. obtusa*. Consistently, codons UUA (L), UUG (L), UCU (S), and GCU (A) demonstrated RSCU values exceeding 1 in all species. It is noteworthy that the initiation codons AUG and UGG, encoding methionine and tryptophan, respectively, exhibited equal usage across all species (RSCU = 1) (**Fig. 5.40**).

The analysis of the plastome in *T. peruviana*, has provided insights into distinctive patterns of codon usage with potential implications for the understanding of its genetic makeup (Deb et al., 2020). Notably, the single codons for W and M suggests an absence of codon preference (Parvathy et al., 2022). The observed variation in codon usage for different amino acids, ranging from two to six codons, emphasizes on the significance of amino acids (R, L and S) in *T. peruviana*. The prevalence of these amino acids is underscored by their diverse codon representations, with R exhibiting a noteworthy spectrum of usage across species as reported in *Arachis* sp. (Fabaceae) (Yang et al., 2023). The highest RSCU value for the AGA codon in *C. manghas*, contrasted with the lowest for the CGC codon in *P. obtusa*, both coding for R, indicates species-specific variations in the preference for certain codons, aligns with the findings in *Mesona chinensis*, *Gynostemma* spp. and in seed plants (Xu et al., 2011; Zhang et al., 2021; Tang et al., 2021). Further, the prominence of codons UUA, UUG, UCU, and GCU with RSCU values consistently exceeding 1 across all species suggests a shared preference for these codons in the plastome. Such uniformity in codon usage patterns may be indicative of evolutionary constraints or functional requirements related to the translation machinery (Xu et al., 2011).

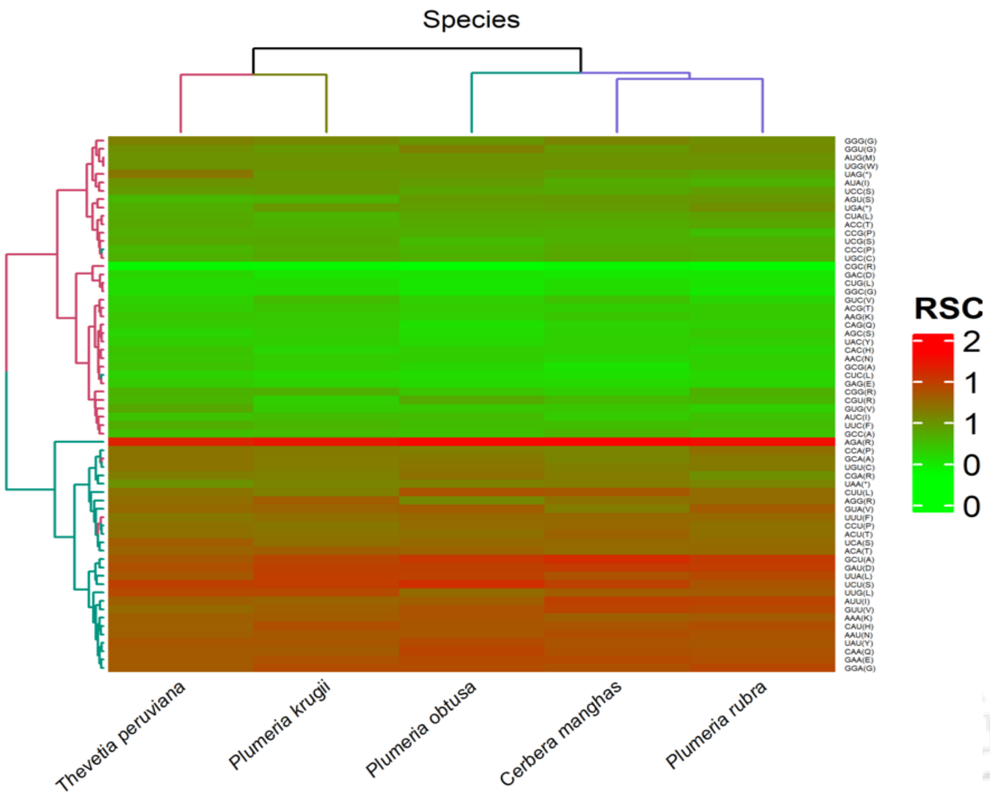


Figure 5.40: Codon usage bias among five Plumerieae species through RSCU values

It is intriguing to note that initiation codons AUG and UGG, encoding methionine and tryptophan, respectively, displayed equal usage across all species with an RSCU value of 1, suggesting a conserved translational initiation mechanism in the plastomes under investigation as reported in almost all plants (Li et al., 2017). The identification of tryptophan as the rarest among the amino acids in *T. peruviana* raises questions about the functional implications of its scarcity. Understanding the factors influencing codon usage bias, such as evolutionary pressures, mutation rates, and translational selection, will be crucial for unraveling the biological significance of these observations (Parvathy et al., 2022). The observed variations in codon preferences for different amino acids across species highlight the complex interplay of evolutionary forces and functional constraints shaping the plastome composition. Further investigations into the functional implications of these codon usage patterns may contribute to a deeper understanding of the molecular processes governing plastome evolution and gene expression in *T. peruviana*.

5.3.2.7. Phylogenetic analysis

Phylogenetic tree for *C. quadrangularis*

We constructed a maximum likelihood (ML) phylogenetic tree using the chloroplast genomes of 24 different species as a super barcode. The alignment consisted of 137,163 base pairs (bp) from the complete genomes of these species. The General Time Reversible model with gamma distribution and invariable sites (GTR+G+I) was identified as the best-fit model for our analysis. The ML tree was highly resolved, with the majority of nodes receiving maximum bootstrap support (**Fig. 5.41**). For our analysis, we used the Leeoideae subfamily of the Vitaceae family as the out-group. The phylogenetic tree successfully clustered the *Cissus* species into a single clade, which included *C. quadrangularis*, *C. discolor*, *C. tuberosa*, and *C. trifoliata*. This clade shared a common node with species from the *Tetrastigma* and *Cyphostemma* genera. Among the Cissus clade, *C. quadrangularis* was found to be closely related to *C. discolor*. In contrast, the genera *Vitis* and *Ampelopsis* formed a separate clade, yet shared a common node, indicating a close evolutionary relationship. Chloroplast genomes have proven to be invaluable for elucidating the evolutionary relationships among angiosperm plants (Shelke and Rangan, 2022). To clarify the evolutionary position of *Cissus* species, we constructed an ML tree using whole cp genome sequences from six *Cissus* species and 17 other species from the Vitoideae subfamily. Our findings demonstrate that chloroplast genomes are sufficient to establish robust phylogenetic relationships among *Cissus* species and other members of the Vitoideae subfamily, as evidenced by significant bootstrap support.

Within the Cissus clade, *C. quadrangularis* was found to be closely related to *C. discolor*. These results are consistent with previous studies that used various DNA barcodes and whole chloroplast genome sequences (Wen et al., 2018). The study also revealed that the *Cissus* genus is closely related to the Cayratieae tribe, which includes the *Tetrastigma* species. This relationship has been observed in prior phylogenetic studies of the Vitoideae subfamily (Wen et al., 2018; Zhang et al., 2016b). One notable challenge in our analysis was the placement of *C. antarctica*, which was excluded from the current phylogenetic tree due to unresolved positioning within the *Cissus* clade. Similar difficulties with resolving the placement of *C. antarctica* have been reported in previous studies, leading to the consideration of *Cissus* as a polyphyletic clade (Liu et al., 2013). Future research should aim to resolve these uncertainties by including a greater number of *Cissus* species and incorporating both organellar and nuclear markers to provide a more comprehensive view of their evolutionary relationships.

The phylogenetic analysis using chloroplast genome data has provided valuable insights into the evolutionary relationships within the *Cissus* genus and its position within the Vitoideae subfamily. Despite some unresolved issues, such as the placement of *C. antarctica* the study underscores the utility of cp genomes in plant phylogenetics. Further studies incorporating additional species and genetic markers will be essential for resolving remaining phylogenetic uncertainties and enhancing our understanding of the evolutionary dynamics within this group.

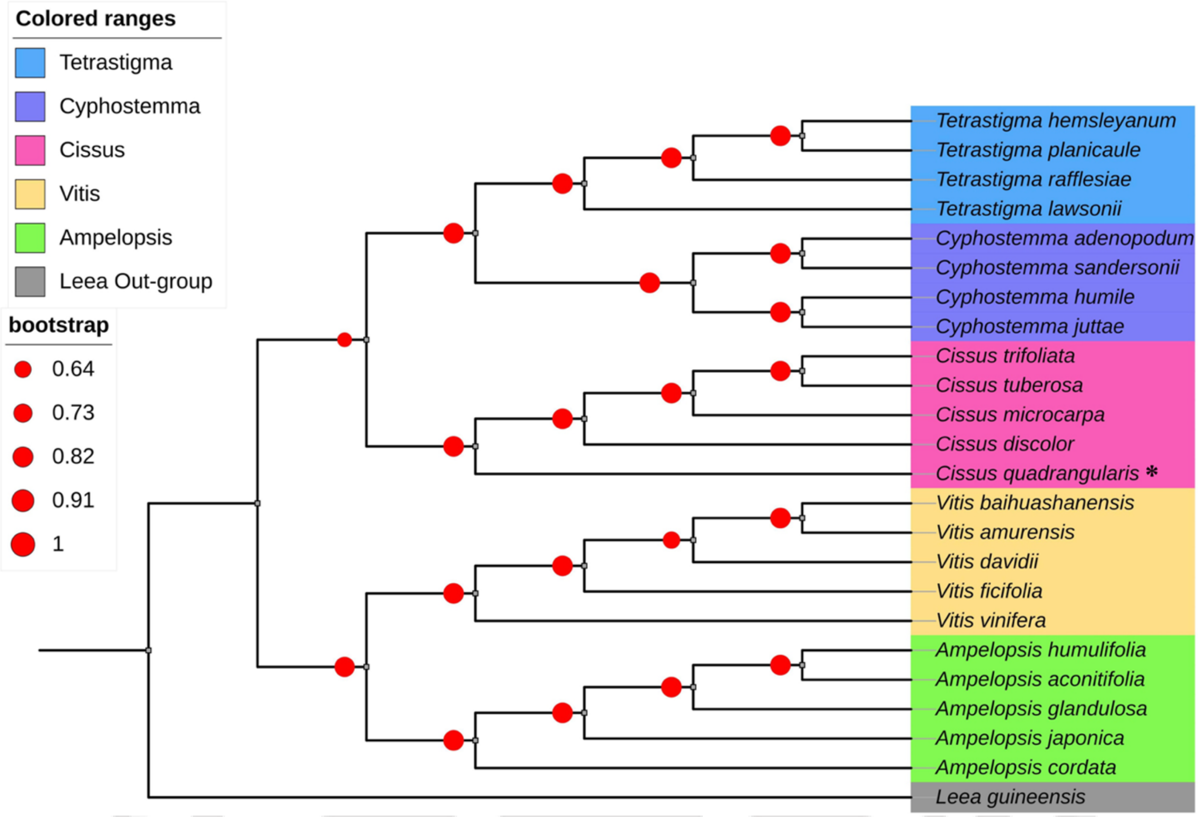


Fig. 5.41: Phylogenetic tree using whole chloroplast genomes from 24 Vitaceae species constructed through the maximum likelihood method.

Phylogenetic tree for *F. hispida*

A robust phylogenetic tree (**Fig. 5.42**) was constructed through ML methodology based on the complete chloroplast genomes of 21 species across six taxa. This phylogenetic analysis elucidates the interrelationships among Moraceae species. *Rosa rubiginosa* and *Prunus persicaria* from the Rosaceae family were included as out-group references. The tree topology reveals that all *Ficus* species form a cohesive cluster, with *Broussonetia* and *Artocarpus* species forming separate clusters. Within the *Ficus* cluster, *F. hispida* (Subgenus: Sycomorus) is adjacent to *F. squamosa*, and more distantly related to *F. virens* (Subgenus: Urostigma), highlighting phylogenetic relationships within the genus (Huang et al., 2022). All nodes were resolved with bootstrap values BV exceeding the threshold, affirming the tree's reliability and robustness. The phylogenetic tree showcases evolutionary relationships among several genera, including *Ficus*, *Broussonetia*, *Artocarpus*, *Morus*, and *Rosa*. *Ficus* is the most diverse genus, with multiple species branching from common nodes, followed by *Broussonetia* with fewer species, and *Artocarpus* displaying moderate diversity. This tree enhances our understanding of the intricate evolutionary connections within the Moraceae family.

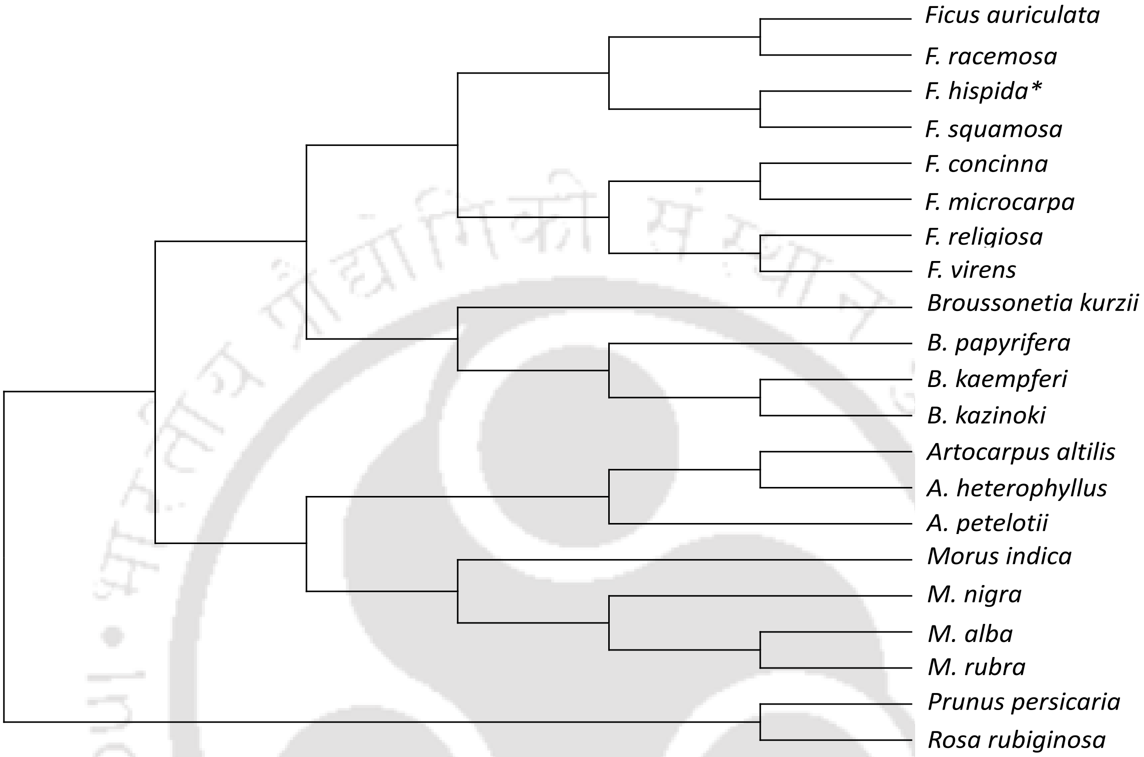


Fig. 5.42: Phylogenetic tree showcasing the relationship of *F. hispida* with 20 other species based on the whole chloroplast genomes.

Phylogenetic tree for *S. rutidosperma*

The ML phylogenetic tree (**Fig. 5.43**) based on the complete cp genomes of 17 species from 12 taxa, elucidating the relationships among Brassicales (Cleomaceae, Brassicaceae, Caricaceae). *Ceiba speciosa* from Malvaceae served as an out-group. Cleomaceae species formed a distinct cluster, followed by Brassicaceae and Caricaceae clusters. Within Cleomaceae, *S. rutidosperma* was adjacent to *T. hassleriana* but farthest from *C. lutea*. The branches of *S. rutidosperma*, *T. hassleriana, G. gynandra,* and *S. rutidosperma* had BV of 94 and 96, respectively. Brassicaceae and Caricaceae clusters had BVs of 100 for all branches. The taxonomic position of Cleomaceae has been debated, with previous classifications placing it in Capparaceae (Hall et al., 2002). Recent molecular studies suggest Cleome is more closely related to *Polanisia* and *Gynandropsis*, warranting its classification under Cleomaceae. Using cp genomes of *S. rutidosperma* and 16 other Cleomaceae species, the tree confirmed

Cleomaceae as a monophyletic clade closely related to Brassicaceae rather than Caricaceae. This aligns with previous studies based on nuclear ribosomal DNA (nrDNA) and cpDNA, as well as DNA sequencing. This study provides insights into the evolutionary relationships and correct taxonomic positions of these families, supported by molecular, morphological, and ecological evidence, and resolves relationships within Cleomaceae and Brassicaceae.

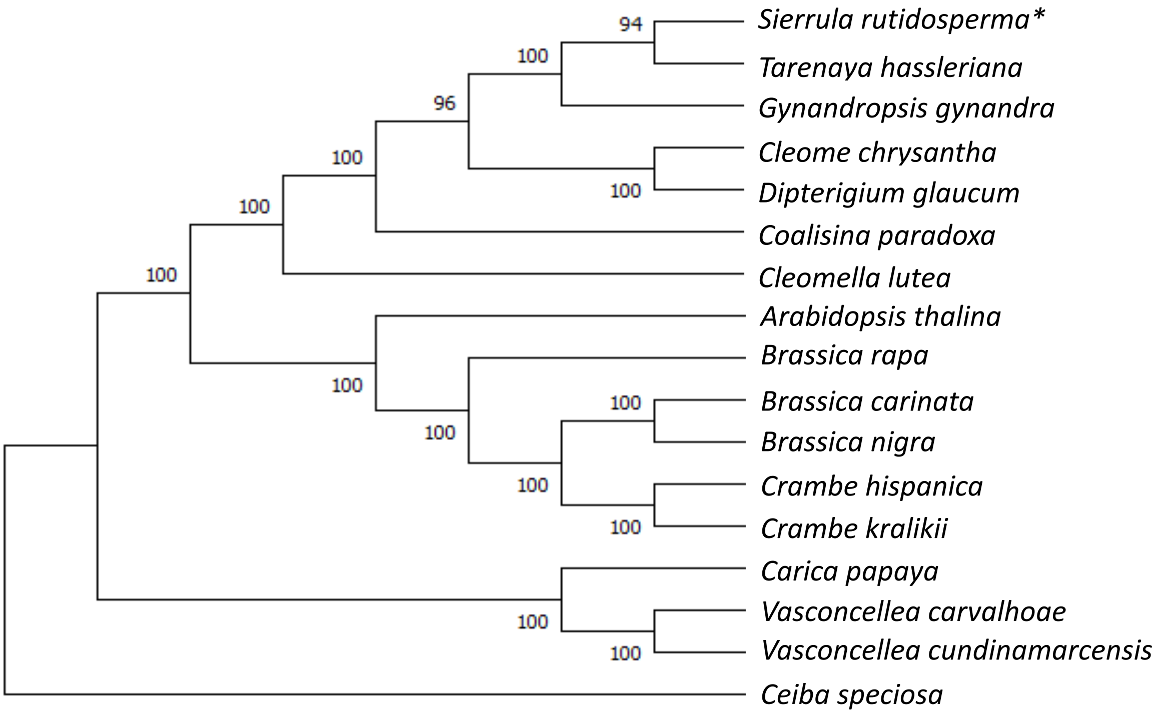


Fig. 5.43: Phylogenetic tree showcasing the taxonomic position of *S. rutidosperma* from 17 complete chloroplast genomes.

Phylogenetic tree for *T. peruviana*

Utilizing 18 complete plastid genomes and aligning 52,722 nucleotide base pairs (bp) from shared Locally Collinear Blocks (LCBs), a ML phylogenetic tree was constructed (**Fig. 5.44**). The GTR+G+I model, the best fit for our dataset, provided strong evidence for the monophyletic nature of Gentianales. Five well-supported families- Rubiaceae, Gelsemiaceae, Loganiaceae, Gentianaceae, and Apocynaceae- were identified, each with high bootstrap values. *Salvia campanulata* from Lamiaceae was used as the outgroup. Within Gentianaceae, Apocynaceae was identified as a sister group, sharing a common ancestral node with Loganiaceae. The analysis revealed a bipartite division within Apocynaceae. The first clade consisted of *T. peruviana* and *C. manghas*, diverging from other Plumerie species, indicating a close evolutionary relationship. The second clade included *P. rubra* and a subclade with *P. krugii* and *P. obtusa*. *T. peruviana* was closely clustered with *C. manghas* based on plastid markers like *mat*K and *rbc*L. This analysis reinforced the evolutionary relationship of *T. peruviana* within the lower taxonomical rank Plumerieae tribe of Apocynaceae, supporting its close relationship with *C. manghas*.

This phylogenetic investigation provided significant insights into the evolutionary relationships within Gentianales, strongly supporting its monophyletic nature and positioning it within the core asterid clade. This is consistent with previous research (Bremer et al., 2002). The analysis identified five families, each characterized by robust bootstrap values, confirming their distinct evolutionary lineages (Antonelli et al., 2021). Within Apocynaceae, the bipartite division highlighted *T. peruviana* and *C. manghas* as a distinct lineage from other *Plumeria* species (Tiernan et al., 2023; Wang et al., 2023). Plastid markers *mat*K and *rbc*L confirmed the close relationship between *T. peruviana* and *C. manghas*, consistent with plastid marker data (Wang et al., 2023). This study reaffirmed *T. peruviana*'s phylogenetic position within the paraphyletic tribe Plumerieae of Apocynaceae.

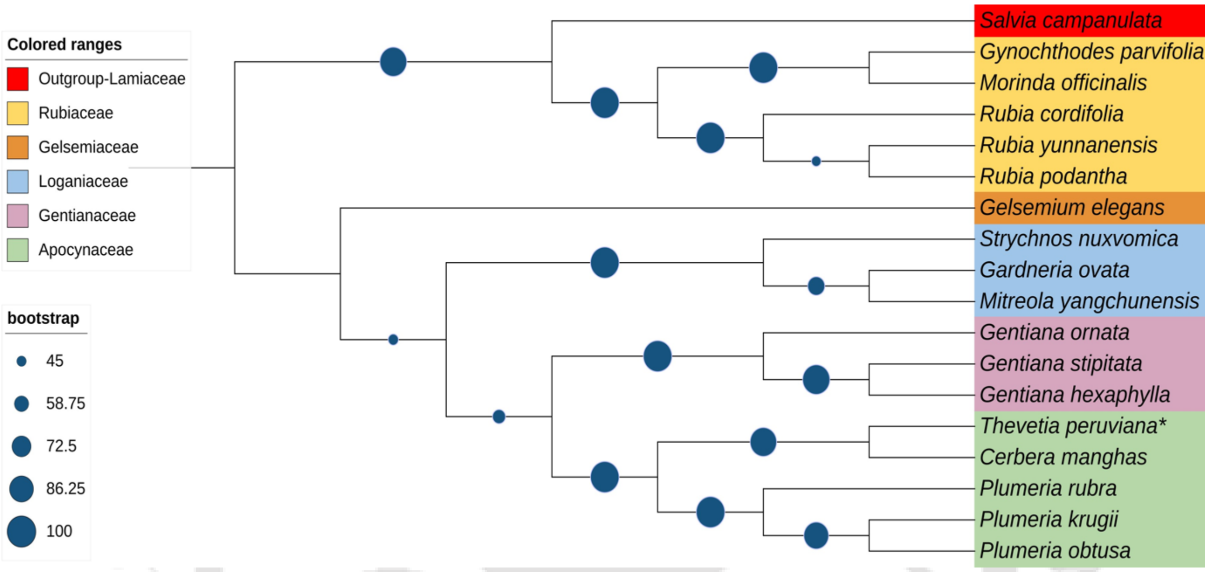


Fig. 5.44: Phylogenetic tree for *T. peruviana* based on the cp genomes from 17 Gentianales and one Lamiales constructed through ML method

5.4. Conclusion

The plastome mining of five small-genome-sized plants, has provided valuable insights into the structural and functional aspects of their circular cp genomes. The determined size of the circular cp genomes in all cases falls within the typical range observed in angiosperms (120-170 kb). The specific sizes for circular cp genome in *C. quadrangularis*, *F. hispida*, *S. rutidosperma*, *S. asper*, and *T. peruviana* were reported as 160,404 bp, 160,162 bp, 157,073 bp, 161,011 bp, and 155,148 bp, respectively, which fall within the typical range observed in angiosperms (120-170 kb). Across all five species, the structural comparison revealed consistency in the size ranges of the LSC, SSC, and IRs. Notably, the total number of genes varied, with *S. asper* exhibiting the lowest count (129 genes), while both *F. hispida* and *S. rutidosperma* had 132 genes each, and *C. quadrangularis* and *T. peruviana* possessed 133 genes each.

Tandem and SSR analysis uncovered variation in the number of repeats among the species. Nucleotide diversity was assessed, revealing notable variations among different coding regions. The highest diversity was observed in specific protein-coding regions, with *rps*19, *psa*I, and *ycf*1 exhibiting the most diversity in different species. Additionally, a limited number of protein-coding genes showed positive selection, with *C. quadrangularis, F. hispida, S. rutidosperma, S. asper*, and *T. peruviana* having 3, 1, 17, 1, and 5 genes with Ka/Ks >1, respectively. The analysis of RSCU highlighted distinct patterns among the five species, with R being the most prevalent amino acid in all cases, except for *C. quadrangularis*, where L took precedence. The phylogenetic analysis based on complete cp genomes for four of the species except *S. asper* revealed important evolutionary facts based on genomic information.

In conclusion, the plastome mining of these small GS plants has not only provided a comparative understanding of their cp genomes but has also revealed species-specific variations in structural organization, repeat content, nucleotide diversity, and codon usage. These findings contribute to the broader knowledge of plant plastomes and underscore the importance of detailed genomic analyses for a comprehensive understanding of plant biology.

Chapter 6

Summary and Future Prospects

Chapter 6: SUMMARY AND FUTURE PROSPECTS

6.1. Summary

An impressive array of 434 plant species across 101 families from IITG campus was documented. Notably, the prevalence of dicotyledonous families far surpassed that of monocots, standing at 80 and 21, respectively. The floristic analysis extends its significance by differentiating between exotic and indigenous plant species within specific families. Families like Proteaceae, Araliaceae, and Rhamnaceae are exclusively represented by exotic species, emphasizing the introduced nature of certain plant lineages. On the other hand, families such as Sapindaceae, Lythraceae, Rutaceae, and Magnoliaceae are exclusively represented by indigenous species, underscoring the ecological importance and natural occurrence of these families within the studied area. A total of 82 tree species, 98 shrub species, and 254 herbaceous species, including climbers and aquatics, were meticulously catalogued. The propagation strategies revealed that the majority (348 species) reproduce through seeds, emphasizing the importance of seed-based regeneration in maintaining plant populations. Additionally, the documentation identified species capable of regenerating through various mechanisms such as roots, rhizomes, tubers, stems, leaves, stolons, or suckers, showcasing the resilience and adaptability of the plant community.

The study further delves into the distribution of plant species across four zones of IIT Guwahati, with Zones 2 and 3 exhibiting higher average elevations and a unique composition of wild plants. This spatial distribution provides valuable insights into the ecological dynamics of the area. The dominance of specific ornamental plants, such as *Cassia fistula*, *C. javanica*, and *P. longifolia*, adds aesthetic value to the region, while the presence of sacred and medicinal plants like *Saraca asoka* underscores the ecological and cultural significance of the studied flora. The taxonomic analysis of the documented flora revealed interesting patterns, with Asteraceae emerging as the most speciose family, followed by Fabaceae, Poaceae, Solanaceae, Euphorbiaceae, and others. This information contributes significantly to our understanding of the plant community's structure and composition, laying a foundation for future research on ecological interactions, conservation strategies, and sustainable land management in the studied area.

In terms of diversity indices, Simpson's index (D) revealed that Z4 exhibits the highest diversity with a D value of 0.9774, closely followed by Z2 at 0.9769. Z1 and Z3 exhibited slightly lower diversity, with D values of 0.9767 and 0.9759, respectively. The Shannon diversity index (H′) revealed the highest species diversity, in Z1 with a value of 4.084 suggesting a rich ecological composition, potentially influenced by factors specific to that zone. Additionally, evenness (E) scores highlighted the even distribution of species in Z4 with the highest E score of 0.9047 and Z3 with the lowest at 0.8972, indicating distribution pattern of individuals among different zones. Jaccard's similarity index (J) further highlighted the similarity between Z1 and Z4, revealing an 80% similarity. This suggests that despite differences in diversity and evenness, Zones 1 and 4 share a significant proportion of plant species, potentially indicating ecological similarities or interconnectedness between these zones.

The genomic diversity of 110 vascular plants from the IIT Guwahati campus was assessed through flow cytometry, providing insights into the intricate dynamics of plant genomics. The study revealed a wide range of GS among the families, with highest no of species (18) from Fabaceae, followed by Amaranthaceae, Asteraceae, and Moraceae, each with seven species. The GS varied significantly, from 0.43 pg in *C. procera* to 42.5 pg in *S. sagittifolia*, underscoring the dynamic nature of plant genomics. The exploration extended to 47 herbaceous plants, representing 28 families, showcasing a spectrum of GS variation from 0.45 pg to 42.5 pg, with an average GS of 5.856 pg. Notably, GS of 29 herbaceous species was estimated for the first time, revealing new insights into the genomic dimensions of these plants. Among these, six species exhibited notable GS variation, while 12 species displayed remarkable stability in their genomic content. The study also encompassed GS quantification in 12 shrub and 51 tree species. Furthermore, GS estimation for 52 plant species demonstrated both consistency and variation compared to reported values, highlighting the importance of continued genomic research for a comprehensive understanding of plant adaptation and evolution within the campus ecosystem.

In addition to the genomic insights, the study highlighted the significance of using the modified HPI buffer, which proved instrumental in eliminating secondary metabolites and ensuring proper isolation of intact nuclei for more accurate flow cytometric analysis. This methodological improvement enhances the reliability of genomic data, laying the groundwork for future studies in plant genomics and contributing valuable information for biodiversity conservation and ecosystem management. The comprehensive approach of integrating genomic diversity analysis with methodological advancements positions this study as a significant contribution to the understanding of plant biology within the IIT Guwahati campus.

The cp genome analysis of five small-sized plants, namely *C. quadrangularis*, *F. hispida*, *S. asper*, *S. rutidosperma*, and *T. peruviana*, revealed the typical quadripartite arrangements with integrated IRs between LSC and SSC regions in all cases as found in Angiosperms. The genome lengths varied from 155,148 bp in *T. peruviana* to 161,011 bp in *S. asper*, with distinct differences in lengths of the LSC, SSC, and IR regions, reflecting diversity in the structural organization of cp genomes among these species. The examination of genetic features revealed variations in the number of total genes (129 to 133), tandem repeats (32 to 48), and SSRs (50 to 83), indicating the genomic diversity among these species. The presence of specific genetic markers like SSRs and tandem repeats offers potential utility for molecular identification through DNA barcoding and breeding programs. Nucleotide diversity (Pi values) differed among the plants, with *S. rutidosperm*a showing the highest value at 0.024, suggesting a comparatively higher level of sequence variation among the closely related compared species.

The Ka/Ks ratios, indicative of nucleotide substitution rate on protein-coding genes, revealed intriguing patterns, with *S. rutidosperma* displaying a notable ratio of 1.077, hinting at potential positive selection. Furthermore, the average codons and GC% content contributed to our understanding of the coding sequence characteristics and base composition. The AT-rich nature of cp genomes is highlighted by GC% values falling below 40% in all cases. This comprehensive analysis offers a nuanced view of the cp genomes of these five plant species, laying the groundwork for future investigations into their evolutionary relationships, adaptation strategies, and potential applications in genetic studies and conservation efforts.

6.2. Future prospects

This study establishes a robust foundation for future research in plant biology, genomics, and ecology and significantly contributes to unravelling floral diversity and genomic intricacies of flowering plants in the IIT Guwahati campus and its surroundings in following ways.

Ecological dynamics exploration:

- Incorporating ecological factors (soil composition, climate variations, microhabitats) can provide a nuanced understanding of plant-environment interactions.
- Study of ecological dynamics can enhance plant diversity by considering environmental influences.

Species conservation:

- Study on rare and endangered species can lead to formulation of effective conservation policies.
- Assessment of conservation status aids in maintaining biodiversity on the campus.
- Long term monitoring programs track trends and impacts of environmental changes on plant community over time.
- Observing spatial and temporal changes in plant diversity and genetic structures would lead to effective conservation strategies.

Community involvement and awareness:

- Involving the campus community in plant monitoring efforts fosters a sense of ownership, in turn raising awareness about biodiversity preservation and its benefits to society.

Genomic architecture understanding:

- Integration of epigenomics with GS can unfold three-dimensional arrangement and epigenetic modifications within plant genomes.
- Connecting GS to functional genomics, such as gene expression patterns and regulatory elements, can provide insights into the relationship between GS and organism complexity. Studying the

functional implications of GS variations can contribute to a deeper understanding of genomic architecture.

Genomic insights and comparative genomics:

- Genomic data from flow cytometry and plastome mining offer potential for comparative genomics studies.
- Collaboration with other institutions can foster a broader framework for understanding evolutionary relationships among plant species.

Functional genomics studies:

- Transcriptomic and proteomic analyses can complement genomic insights, unravelling molecular mechanisms of plant adaptation.
- Application of functional genomics enhances comprehension of key plant species in specific ecological conditions.

Marker development and crop improvement

- Identifications of conserved regions through cp genome mining can be developed as potential molecular markers with implications in DNA barcoding, plant breeding and genetic engineering.
- Development of molecular markers are particularly useful in understanding genetic diversity, evolutionary history, population genetics, marker-assisted selection (MAS) and construction of genetic maps providing plethora of information for crop improvement.

REFERENCES

Abbasi-Karin, S., Karimzadeh, G., Mohammadi-Bazargani, M. (2022). Interspecific chromosomal and genome size variations in *in vitro* propagated willow herb (*Epilobium* spp.) medicinal plant. *Cytologia*, *87*(2), 129-135.

Adan, A., Alizada, G., Kiraz, Y., Baran, Y., Nalbant, A. (2017). Flow cytometry: basic principles and applications. *Critical Reviews in Biotechnology*, *37*(2), 163-176.

Al-Juhani, W. S., Alharbi, S. A., Al Aboud, N. M., Aljohani, A. Y. (2022). Complete chloroplast genome of the desert date (*Balanites aegyptiaca* (L.) Del. comparative analysis, and phylogenetic relationships among the members of Zygophyllaceae. *BMC Genomics, 23*(1), 1-19.

Albach, D. C., Greilhuber, J. (2004). Genome size variation and evolution in Veronica. *Annals of Botany, 94*(6), 897-911.

ALJuhani, W. S., Aljohani, A. Y. (2022). Complete chloroplast genome of the medicinal plant *Cleome paradoxa* R. Br. Ex DC: Comparative analysis, and phylogenetic relationships among the members of Cleomaceae. *Gene, 845*, 146851.

Alshegaihi, R. M., Mansour, H., Alrobaish, S. A., Al Shaye, N. A., Abd El-Moneim, D. (2023). The first complete chloroplast genome of *Cordia monoica*: structure and comparative analysis. *Genes, 14*(5), 976.

Álvarez-Borrego, J., Gallardo-Escárate, C., Kober, V., López-Bonilla, O. (2007). Genome size estimation: a new methodology. In *Sixth Symposium Optics in Industry*, *6422*, 35-45).

Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K., Palmer, J. D. (2010). Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution, 27*(6), 1436-1448.

Alzahrani, D. A., Yaradua, S. S., Albokhari, E. J., Abba, A. (2020). Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics, 21*(1), 1-19.

Amiryousefi, A., Hyvönen, J., Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics, 34*(17), 3030-3031.

Andersson, L., Purugganan, M. (2022). Molecular genetic variation of animals and plants under domestication. *Proceedings of the National Academy of Sciences, 119*(30), e2122150119.

Animasaun, D. A., Morakinyo, J. A., Mustapha, O. T., Krishnamurthy, R. (2019). Genome size and ploidy variations in pearl millet (*Pennisetum glaucum*) and napier grass (*Pennisetum purpureum*) genotypes. *Acta Agronomica*, *68*(4).

Antonelli, A., Clarkson, J.J., et al. (2021). Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *American J Bot 108*, 1143–1165

APG, IV. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society 181*(1), 1-20. DOI: http://dx. doi. org/10.1111/boj, 12385

Ari, Ş., Arikan, M. (2016). Next-generation sequencing: advantages, disadvantages, and future. *Plant Omics: Trends and Applications*, 109-135.

Arisdason, W. et al. (2023). Status of plant diversity in India. *ENVIS, Centre on floral diversity*, India. https://bsienvis.nic.in/Content/StatusofPlantDiversityinIndia_17566.aspx?format=Print (Accessed on 9th Nov, 2023)

Aronson, M. F., La Sorte, F. A., Nilon, et al. (2014). A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the royal society B: Biological Sciences, 281*(1780), 20133330.

Arora, N. K. (2018). Biodiversity conservation for sustainable future. *Environmental Sustainability*, *1*(2), 109-111.

Arumuganathan, K., Earle, E.D. (1991). Estimation of nuclear DNA content of plants by flow cytometry. *Plant Molecular Biology Reporter, 9*, 229-241.

Bai, H. R., Oyebanji, O., Zhang, R., Yi, T. S. (2021). Plastid phylogenomic insights into the evolution.

Bansal, K. C., Saha, D. (2012). Chloroplast genomics and genetic engineering for crop improvement. *Agricultural Research, 1*, 53-66.

Baranyi, M., Greilhuber, J. (1995). Flow cytometric analysis of genome size variation in cultivated and wild *Pisum sativum* (Fabaceae). *Plant Systematics and Evolution 19*(4), 231-239.

Baranyi, M., Greilhuber, J. (1996). Flow cytometric and Feulgen densitometric analysis of genome size variation in *Pisum*. *Theoretical and Applied Genetics*, *92*, 297-307.

Barooah, C., Ahmed, I. (2014). Plant diversity of Assam: a checklist of angiosperms, gymnosperms. *Assam Sci. Technol. and Env. Counc*. ISBN: 978-81-9238-01-17

Barooah, C., Borthakur, S.K. (2003). Diversity and distribution of bamboos in Assam. *Bishen Singh and Mahendra Pal Singh (Eds.)*. Dehradun, India. ISBN: 8121102308

Barua, I. C., Deka, J., Devi, M. (2013). Invasive weeds and vegetation dynamics in Assam. In *Weed Science Society Conference 2013*, Bandung Indonesia.

Baruah, P. S., Borthakur, S. K., Tanti, B. (2020). Preventing extinction and improving conservation status of *Mesua assamica* (King Prain) Kosterm.-an endangered plant of Assam, India. *Acta Ecologica Sinica*, *40*(3), 185-189.

Baruah, S., Barman, P., Basumatary, S., Bhuyan, B. (2021). Diversity and ethnobotany of genus *Garcinia* L.(clusiaceae) in Assam, Eastern Himalaya. *Ethnobotany Research and Applications, 21*, 1-14.

Basak, S., Kesari , V., Ramesh, A. M., Rangan, L., Parida, A., Mitra, S. (2017). Assessment of genetic variation among nineteen turmeric cultivars of Northeast India: nuclear DNA content and molecular marker approach. *Acta Physiologiae Plantarum 39*, 45. https://doi.org/10.1007/s11738-016-2341-1

Basak, J. K., Qasim, W., Okyere, F. G., et al. (2019). Regression analysis to estimate morphology parameters of pepper plant in a controlled greenhouse system. *Journal of Biosystems Engineering*, *44*, 57-68.

Basak, S., Sarma, G. C., Rangan, L. (2010). Ethnomedical uses of Zingiberaceous plants of Northeast India. *Journal of Ethnopharmacology, 132*(1), 286-296.

Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist, 179*(4), 975-986.

Beaulieu, J. M., Moles, A. T., Leitch, I. J., Bennett, M. D., Dickie, J. B., Knight, C. A. (2007a). Correlated evolution of genome size and seed mass. *New Phytologist*, *173*(2), 422.

Beaulieu, J. M., Smith, S. A., Leitch, I. J. (2010). On the tempo of genome size evolution in angiosperms. *Journal of Botany*.

Begum, S. S., Nath, M., Roy, H., Borthakur, S. K. (2010). Flora of Nameri national park, Assam: I, rare, endangered and threatened plants. *Pleione*,172-79.

Bennett, M. D. (1976). DNA amount, latitude, and crop plant distribution. *Environmental and Experimental Botany, 16*(2-3), 93-108.

Bennett, M. D., Smith, J. B. (1976). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 274*(933), 227-274.

Bennett, M. D., Smith, J. B. (1991). Nuclear DNA amounts in angiosperms. *Philosophical Transactions: Biological Sciences,* 309-345.

Bennett, M. D., Leitch, I. J. (1997). Nuclear DNA amounts in angiosperms- 583 new estimates. *Annals of Botany*, *80*(2), 169-196.

Bennett, M. D., Leitch, I. J., Price, H. J., Johnston, J. S. (2003). Comparisons with Caenorhabditis (~ 100 Mb) and Drosophila (~ 175 Mb) using flow cytometry show genome size in Arabidopsis to be~ 157 Mb and thus~ 25% larger than the Arabidopsis genome initiative estimate of~ 125 Mb. *Annals of Botany, 91*(5), 547-557.

Bennett, M. D., Leitch, I. J. (2005a) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany, 95*(1): 99-113.

Bennett, M. D., Leitch, I. J. (2005b). Genome size evolution in plants. In *The Evolution of the Genome* (pp. 89-162). Academic Press.

Bennett, M. D., Leitch, I. J. (2005c). Plant genome size research: a field in focus. *Annals of Botany*, *95*(1), 1-6.

Bennetzen, J. L., Ma, J., Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of Botany, 95*(1), 127-132.

Bergounioux, C., Perennes, C., Miege, C., Gadal, P. (1986). The effect of male sterility on protoplast division in *Petunia hybrida*. Cell cycle comparison by flow cytometry. *Protoplasma, 130*, 138-144.

Bevan, M. W., Uauy, C., Wulff, B. B. et al. (2017). Genomic innovation for crop improvement. *Nature*, *543*(7645), 346-354.

Bhadra, S., Leitch, I. J., Onstein, R. E. (2023). From genome size to trait evolution during angiosperm radiation. *Trends in Genetics, 39*(10), 728-735.

Bhatnagar, A. K., Kapoor, R. (Eds.). (2018). Plant diversity in India. *IK International Pvt Ltd*.

Bhattacharjya, D. K., Sarma, S. K. (2016). Floristic composition and biological spectrum of weeds in agro-climatic zone of Nalbari district, Assam, India. *Tropicl. Plant Res. 3*(3): 573-585. https://doi.org/10.22271/TPR.2016.V3.I3.076

Bi, Y., Zhang, M. F., Xue, J. et al. (2018). Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Scientific Reports, 8*(1), 1184.

Bino, R. J., Lanteri, S., Verhoeven, H. A., Kraak, H. L. (1993). Flow cytometric determination of nuclear replication stage in seed tissues. *Annals of Botany, 72*(2), 181-187.

Birren, B., Lai, E. (2012). Pulsed field gel electrophoresis: a practical guide. *Academic Press.*

Blommaert, J. (2020). Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B, 287*(1933), 20201441.

Boardman, N. K., Linnane, A. W., Smillie, R. M. (Eds.). (1971). Autonomy and biogenesis of mitochondria and chloroplasts, 70. Amsterdam: North-Holland.

Bock, D. G., Andrew, R. L., Rieseberg, L. H. (2014). On the adaptive value of cytoplasmic genomes in plants. *Molecular Ecology*, *23*(20), 4899-4911.

Bock, R., Timmis, J. N. (2008). Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays, 30*(6), 556-566.

Bolger, M. E., Weisshaar, B., Scholz, U. et al. (2014). Plant genome sequencing—applications for crop improvement. *Current Opinion in Biotechnology*, *26*, 31-37.

Bora, P. J., Kumar, Y. (2003). Floristic diversity of Assam: study of Pabitora wildlife sanctuary. *Daya Books.* ISBN: 817-03-52703

Bordoloi, C., Kumar, S., Barbhuiya, A. M. et al. (2023). Herbal medicine used for wound healing by the tribes of the North Eastern states of India: a comprehensive review. *Journal of Herbal Medicine,* 100697.

Borsch, T., Berendsohn, W., Dalcin, E. et al. (2020). World Flora Online: placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. *Taxon, 69*(6), 1311-1341.

Bourge, M., Brown, S. C., Siljak-Yakovlev, S. (2018). Flow cytometry as tool in plant sciences, with emphasis on genome size and ploidy level assessment. *Genetics Applications*, *2*(2), 1-12.

Brandies, P., Peel, E., Hogg, C. J., Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11), 846.

Bremer et al., (2002). Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol. Phylo. and Evol., 24, 274–301.

Buragohain, J. (2011). Ethnomedicinal plants used by the ethnic communities of Tinsukia district of Assam, India. *Recent Res. in Sci. and Technol. 3*(9), 31- 42. ISSN (Online): 2076-5061

Burke, S. V., Lin, C. S., Wysocki, W. P. et al. (2016). Phylogenomics and plastome evolution of tropical forest grasses (*Leptaspis*, *Streptochaeta*: Poaceae). *Frontiers in Plant Science, 7*, 1993.

Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M., Olmo, E. (2016). Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research, 147*(4), 217-239.

Carta, A., Bedini, G., Peruzzi, L. (2020). A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytologist, 228*(3), 1097-1106.

Carvalho, C. R., Clarindo, W. R., Praça, M. M., Araujo, F. S., Carels, N. (2008). Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. *Plant Science, 174*(6), 613-617.

Casola, C., Li, J. (2022). Beyond RuBisCO: convergent molecular evolution of multiple chloroplast genes in C4 plants. *PeerJ, 10*, e12791.

Cavalier-Smith, T. (1980). r-and K-tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. *Biosystems, 12*(1-2), 43-59.

Cay, S. B., Cinar, Y. U., Kuralay, S. C. et al. (2022). Genome skimming approach reveals the gene arrangements in the chloroplast genomes of the highly endangered *Crocus* L. species: *Crocus istanbulensis* (B. Mathew) Rukšāns. *Plos One*, *17*(6), e0269747.

Čertnerová, D. (2022). Meet the challenges of analyzing small genomes using flow cytometry. *Cytometry Part A, 101*(9), 707-709.

Čertnerová, D., Galbraith, D. W. (2021). Best practices in the flow cytometry of microalgae. *Cytometry Part A, 99*(4), 359-364.

Chakravarty, S., Suresh, C. P., Puri, A., Shukla, G. (2012). North-east India, the geographical gateway of India's phytodiversity. *Indian Forester, 138*(8), 702.

Chan, A. P., Crabtree, J., Zhao, Q. et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology, 28*(9), 951-956.

Chatterjee, S., Saikia, A., Dutta, P., Ghosh, D., Worah, S. (2006). Review of biodiversity in North East India. Biodiversity significance of north east India view project. https://www.researchgate.net/publication/272364229.

Chen, H., Chen, H., Wang, B., Liu, C. (2023). Conserved chloroplast genome sequences of the genus *Clerodendrum* Linn.(Lamiaceae) as a super-barcode. *Plos One, 18*(2), e0277809.

Chen, J., Glémin, S., Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, *34*(6), 1417-1428.

Chen, J., Liu, X., Zhu, L., Wang, Y. (2013). Nuclear genome size estimation and karyotype analysis of *Lycium* species (Solanaceae). *Scientia Horticulturae*, *151*, 46-50.

Chen, S. C., Cannon, C. H., Kua, C. S., Liu, J. J., Galbraith, D. W. (2014). Genome size variation in the Fagaceae and its implications for trees. *Tree Genetics Genomes*, *10*, 977-988.

Chénais, B., Caruso, A., Hiard, S., Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*, *509*(1), 7-15.

Cheng, Y., Yang, P., Zhao, L. et al. (2019). Studies on genome size estimation, chromosome number, gametophyte development and plant morphology of salt-tolerant halophyte *Suaeda salsa*. *BMC Plant Biology*, *19*, 1-12.

Cho, M. S., Kim, J. H., Yamada, T., Maki, M., Kim, S. C. (2021). Plastome characterization and comparative analyses of wild crabapples (*Malus baccata* and *M. toringo*): insights into infraspecific plastome variation and phylogenetic relationships. *Tree Genetics Genomes, 17*(5), 41.

Choudhary, B. (2013). Biological diversity, environmental ethics and traditional ecological knowledge: concerns in Northeast India. In book: *Biodivers. and Ecol., Discovery Publishing House Pvt. Ltd.*

Choudhary, R. K., Srivastava, R. C., Das, A. K., Lee, J. G. (2012). Floristic diversity assessment and vegetation analysis of upper Siang district of eastern Himalaya in North East India. *Korean J. Pl. Taxon. 42*(3), 222-246. https://doi.org/10.11110/kjpt.2012.42.3.222

Choudhury, R. R., Basak, S., Ramesh, A. M., Rangan, L. (2014). Nuclear DNA content of *Pongamia pinnata* L. and genome size stability of *in vitro*-regenerated plantlets. *Protoplasma 251*, 703–709. https://doi.org/10.1007/s00709-013-0545-4

Choudhury, B., Khan, M. L. (2010). Conservation and management of endangered plant species: a case study from Northeast India. *Bioremediation, Biodiversity and Bioavailability, 4*(1), 47-53.

Choudhury, K., Kalita, S. R. (2015). Floristic diversity of Greater Sualkuchi area under Kamrup district of Assam. *Hexagon, 69*.

Christenhusz, M. J., Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa, 261*(3), 201-217.

Chung, H. J., Jung, J. D., Park, H. W. et al. (2006). The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Reports*, *25*, 1369-1379.

Chung, M. Y., Son, S., López-Pujol, J., Mao, K., Chung, M. G. (2021). Plant conservation practitioners can benefit from neutral genetic diversity. *Diversity, 13*(11), 552.

Claros, M. G., Bautista, R., Guerrero-Fernández, D. et al. (2012). Why assembling plant genome sequences is so challenging. *Biology, 1*(2), 439-459.

Convention on Biological Diversity (2012). The global strategy for plant conservation: 2011–2020. Richmond, U.K.: *Botanic Gardens Conservation International*. https://www.bgci.org/files/Plants2020/GSPCbrochure/gspc_english.pdf (Accessed on 2nd Nov, 2023)

Corlett, R. T. (2016). Plant diversity in a changing world: status, trends, and conservation needs. *Plant Diversity, 38*(1), 10-16.

Cregan, P. B., Akkaya, M. S., Bhagwat, A. A., Lavi, U., Rongwen, J. (2020). Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. In *Plant Genome Analysis*, 47-56, CRC Press.

Cui, H., Ding, Z., Zhu, Q., Wu, Y., Gao, P. (2020). Population structure and genetic diversity of watermelon (*Citrullus lanatus*) based on SNP of chloroplast genome. *3 Biotech*, *10*, 1-8.

Cullis, C. A., Vorster, B. J., Van Der Vyver, C., Kunert, K. J. (2009). Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants?. *Annals of Botany, 103*(4), 625-633.

D'hondt, L., Höfte, M., Van Bockstaele, E., Leus, L. (2011). Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status. *Molecular Plant Pathology*, *12*(8), 815-828.

Daniell, H., Lin, C. S., Yu, M., Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology, 17*, 1-29.

Dar, J. A., Sundarapandian, S. (2016). Patterns of plant diversity in seven temperate forest types of Western Himalaya, India. *Journal of Asia-Pacific Biodiversity, 9*(3), 280-292.

Daru, B. H., van der Bank, M., Davies, T. J. (2018). Unravelling the evolutionary origins of biogeographic assemblages. *Divers Distrib*. *24*, 313–324. https://doi.org/10.1111/ddi.12679

Darzynkiewicz, Z., Halicka, H. D., Zhao, H. (2010). Analysis of cellular DNA content by flow and laser scanning cytometry. *Polyploidization and Cancer*, 137-147.

Das, H., Kalita, D. (2016). Fibers and dye yielding plants of north East India. *Bioprospecting of Indigenous Bioresources of North-East India*, 77-99.

Das, K., Duarah, P. (2013). Invasive alien plant species in the roadside areas of Jorhat, Assam: their harmful effects and beneficial uses. *International Journal of Engineering Research and Applications, 3*(5), 353-358.

Das, R. (2016). Biodiversity of ethnomedicinal plants used by the ethnic tribal people of Barpeta district of Assam, North East India. *Asian Journal of Pharmaceutical Science and Technology, 6*(1), 27-32.

Das, R., Rangan, L. (2020). Correlation of genome size with cell phenotypic traits in non-edible oil crops. *Acta Physiologiae Plantarum, 42*, 1-10.

Das, R., Shelke, R. G., Rangan, L., Mitra, S. (2018). Estimation of nuclear genome size and characterization of Ty1-copia like LTR retrotransposon in *Mesua ferrea* L. *Journal of Plant Biochemistry and Biotechnology*, *27*, 478-487.

Dash, S. S., Kumar, S., Mao, A. A. (2022). Plant diversity at ecosystem level in India: dynamics and status. *In Biodiversity in India: Status, Issues and Challenges*, 431-461. Singapore: Springer Nature Singapore.

Davis, M. A., Hancock, D. D., Besser, T. E., Call, D. R. (2003). Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of Escherichia coli O157: H7. *Journal of Clinical Microbiology*, *41*(5), 1843-1849.

Dayton, P. K. (1972). Toward an understanding of community resilience and the potential effects of enrichments to the benthos at McMurdo Sound, Antarctica. *In Proceedings of the Colloquium on Conservation Problems in Antarctica*, 81-96. Lawrence, Kansas, USA: Allen Press.

De Laat, A. M. M., Blaas, J. (1984). Flow-cytometric characterization and sorting of plant chromosomes. *Theoretical and Applied Genetics, 67*, 463-467.

De, L. C., Medhi, R. P. (2014). Diversity and conservation of rare and endemic orchids of North East India-a review. *Indian Journal of Hill Farming, 27*(1), 138-153.

Deb, B., Uddin, A., Chakraborty, S. (2020). Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. *Archives of Virology, 165*, 557-570.

Deka, J., Kalita, J. C. (2013). Ethnobotanically important medicinal plants of Kamrup district, Assam, India used in fertility treatment. *Int Res J Pharm, 4*, 229-32.

Deka, K., Borthakur, S. K., Tanti, B. (2017). Distribution and population dynamics of *Calamus nambariensis* Becc. an endemic and threatened cane of Assam. *Annals of Plant Sciences, 6*(12), 1829-1834.

Dey, P. (2021). Flow Cytometry of body cavity fluid. *In Diagnostic Flow Cytometry in Cytology,* 153-168. Singapore: Springer Singapore.

Dhingra, A., Folta, K. M. (2005). ASAP: amplification, sequencing annotation of plastomes. *BMC Genomics, 6*, 1-13.

Dhyani, S., Fulwadhani, P. S., Gujre, N., Wate, S. R. (2020). A score-based method for assessing the importance of urban forest and its ecosystem services. *International Journal of Plant and Environment, 6*(01), 35-46.

Diallo, A. M., Nielsen, L. R., Kjær, E. D., Petersen, K. K., Ræbild, A. (2016). Polyploidy can confer superiority to West African *Acacia senegal* (L.) Willd. trees. *Frontiers in Plant Science, 7*, 821.

Díez, C. M., Gaut, B. S., Meca, E. et al. (2013). Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist, 199*(1), 264-276.

Dikshit, K. R., Dikshit, J. K. (2014). Weather and climate of North East India. In: North-East India: Land, People and Economy (2014th Eds.). *Advances in Asian Human Environmental Res*. Springer, Dordrecht.

Dobrogojski, J., Adamiec, M., Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiologiae Plantarum*, *42*, 1-13.

Dodsworth, S., Leitch, A. R., Leitch, I. J. (2015). Genome size diversity in angiosperms and its influence on gene space. *Current Opinion in Genetics Development, 35*, 73-78.

Dogan, Y., Nedelcheva, A. M., Yarci, C. (2008). Plant taxa used as brooms in several Southeast European and West Asian countries. *Natura Croatica: Periodicum Musei Historiae Naturalis Croatici, 17*(3), 193-206.

Doležel, J., Sgorbati, S., Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum, 85*(4), 625-631.

Doležel, J., Doleželová, M., Novák, F. J. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum, 36*(3), 351-357.

Doležel, J., Göhde, W. (1995). Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry: The Journal of the International Society for Analytical Cytology, 19*(2), 103-106.

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A. et al. (1998). Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Annals of Botany, 82*(suppl_1), 17-26.

Dolezel, J., Greilhuber, J., Suda, J. (2003). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols, 2*(9), 2233-2244. doi: 10.1038/nprot.2007.310

Doležel, J., Bartoš, J. A. N. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany, 95*(1), 99-110.

Doležel, J., Greilhuber, J., Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc 2*, 2233–2244 (2007). https://doi.org/10.1038/nprot.2007.310

Doležel, J., Greilhuber, J. (2010). Nuclear genome size: are we getting closer?. *Cytometry Part A, 77*(7), 635-642.

Doležel, J., Vrána, J., Šafář, J., Bartoš, J., Kubaláková, M., Šimková, H. (2012). Chromosomes in the flow to simplify genome analysis. *Functional Integrative Genomics*, *12*, 397-416.

Dong, W., Xu, C., Li, W. et al. (2017). Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA sequences. *Frontiers in plant science, 8*, 1148.

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, 105*(supplement_1), 11549-11555.

Downie, S. R., Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany, 40*(1), 336-351.

Dpooležel, J., Binarová, P., Lcretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum, 31*(2), 113-120.

Dutta, G., Devi, A. (2013). Plant diversity and community structure in tropical moist deciduous sal (*Shorea robusta* Gaertn.) forest of Assam, northeast India. *Journal of Environmental and Applied Bioresearch, 1*(3), 1- 4. ISSN, 2319, 8745.

Edwards, D., Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal, 8*(1), 2-9.

Ejaz, B., Mujib, A., Mamgain, J., Malik, M. Q., Syeed, R., Gulzar, B., Bansal, Y. (2021). Comprehensive in vitro regeneration study with SCoT marker assisted clonal stability assessment and flow cytometric genome size analysis of *Carthamus tinctorius* L.: an important medicinal plant. *Plant Cell, Tissue and Organ Culture (PCTOC),* 1-16.

El Zerey-Belaskri, A., Ribeiro, T., Alcaraz, M. L. et al. (2018). Molecular characterization of *Pistacia atlantica* Desf. subsp. atlantica (Anacardiaceae) in Algeria: genome size determination, chromosome count and genetic diversity analysis using SSR markers. *Scientia Horticulturae, 227,* 278-287.

Elfrida, E., Mubarak, A., Suwardi, A. B. (2020). The fruit plant species diversity in the home gardens and their contribution to the livelihood of communities in rural area. *Biodiversitas Journal of Biological Diversity, 21*(8).

Emerson, B. C. (2003). Genes, geology and biodiversity: faunal and floral diversity on the island of Gran Canaria. *Animal Biodiversity and Conservation, 26*(1), 9-20.

Faizullah, L., Morton, J. A., Hersch-Green, E. I. et al. (2021). Exploring environmental selection on genome size in angiosperms. *Trends in Plant Science*, *26*(10), 1039-1049.

Fan, Z. F., Ma, C. L. (2022). Comparative chloroplast genome and phylogenetic analyses of Chinese *Polyspora*. *Scientific Reports, 12*(1), 15984.

Farhat, P., Siljak-Yakovlev, S., Adams, R. P., Bou Dagher Kharrat, M., Robert, T. (2019). Genome size variation and polyploidy in the geographical range of *Juniperus sabina* L.(Cupressaceae). *Botany Letters, 166*(2), 134-143.

Fathima, F., Parveen, R., Khan, M. A. et al. (2023). *Saraca asoca*: from traditional herb to modern drug as a cure for various diseases. *Current Drug Therapy, 18*(4), 285-297.

Feng, J. L., Wu, L. W., Wang, Q. et al. (2022). Comparison analysis based on complete chloroplast genomes and insights into plastid phylogenomic of four *Iris* species. *BioMed Research International*.

Feng, S., Zheng, K., Jiao, K. et al. (2020). Complete chloroplast genomes of four *Physalis* species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. *BMC Plant Biology, 20*, 1-14.

Feuillet, C., Leach, J. E., Rogers, J., Schnable, P. S., Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends in Plant Science*, *16*(2), 77-88.

Fleischmann, A., Michael, T. P., Rivadavia, F. et al. (2014). Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Annals of Botany, 114*(8), 1651-1663.

Flinn, K. M., Lechowicz, M. J., Waterway, M. J. (2008). Plant species diversity and composition of wetlands within an upland forest. *American Journal of Botany, 95*(10), 1216-1224.

Fomicheva, M., Domblides, E. (2023). Mastering DNA content estimation by flow cytometry as an efficient tool for plant breeding and biodiversity research. *Methods and Protocols, 6*(1), 18.

Frydrych, I. (2021). Cell cycle analysis by flow cytometry. *Laboratory Techniques in Cellular and Molecular Medicine, 1st ed.*; Agrawal, K., Bouchal, J., Das, V., Drábek, J., Džubák, P., Hajdúch, M., Koberna, K., Ligasová, A., Mistrík, M., de Sanctis, JB, Eds, 193-200.

Galbraith, D. W. (2009). Simultaneous flow cytometric quantification of plant nuclear DNA contents over the full range of described angiosperm 2C values. *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry, 75*(8), 692-698.

Galbraith, D. W., Harkins, K. R., Maddox, J. M. et al. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science, 220*(4601), 1049-1051.

Galbraith, D., Loureiro, J., Antoniadi, I. et al. (2021). Best practices in plant cytometry. *Cytometry Part A, 99*(4), 311-317.

Gandhi, S. G., Awasthi, P., Bedi, Y. S. (2010). Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family. *Bioinformation, 5*(1), 16.

Gao, L. Z., Liu, Y. L., Zhang, D. et al. (2019). Evolution of Oryza chloroplast genomes promoted adaptation to diverse ecological habitats. *Communications Biology, 2*(1), 278.

Gao, L., Su, Y. J., Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *Journal of Systematics and Evolution, 48*(2), 77-93.

Gao, Y., Lu, Y., Song, Y., Jing, L. (2022). Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus*. *BMC Genomic Data, 23*(1), 46.

Gautam, A., Kushwaha, A., Rani, R. (2022). Reduction of hexavalent chromium [Cr (VI)] by heavy metal tolerant Bacterium *Alkalihalobacillus clausii* CRA1 and its toxicity assessment through flow cytometry. *Current Microbiology, 79*(1), 33.

Geng, X., Huang, N., Zhu, Y., Qin, L., Hui, L. (2022). Codon usage bias analysis of the chloroplast genome of cassava. *South African Journal of Botany, 151*, 970-975.

Ghorbanpour, M., Hadian, J., Nikabadi, S., Varma, A. (2017). Importance of medicinal and aromatic plants in human life. *In: Ghorbanpour, M., Varma, A. (eds) Medicinal Plants and Environmental Challenges.* Springer. https://doi.org/10.1007/978-3-319-68717-9_1

Gichuki, D. K., Ma, L., Zhu, Z. et al. (2019). Genome size, chromosome number determination, and analysis of the repetitive elements in *Cissus quadrangularis*. *PeerJ, 7*, e8201.

Gillette, D. P., Edds, D. R., Jha, B. R., Mishra, B. (2022). Thirty years of environmental change reduces local, but not regional, diversity of riverine fish assemblages in a Himalayan biodiversity hotspot. *Biological Conservation, 265*, 109427.

Giorgi, D., Farina, A., Grosso, V. et al. (2013). FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy. *PLoS One, 8*(2), e57994.

Gitzendanner, M. A., Soltis, P. S., Yi, T. S. et al. (2018). Plastome phylogenetics: 30 years of inferences into plant evolution. *In Advances in Botanical Research, 85*, 293-313. Academic Press.

Gnanesh, B. N., Mondal, R. et al. (2023). Genome size, genetic diversity, and phenotypic variability imply the effect of genetic variation instead of ploidy on trait plasticity in the cross-pollinated tree species of mulberry. *PloS One, 18*(8), e0289766. https://doi.org/10.1371/journal.pone.0289766

Godden, G. T., Jordon-Thaden, I. E., Chamala, S. et al. (2012). Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. *Plant Ecology Diversity*, *5*(4), 427-450.

Gogoi, L., Baruah, G., Borah, D., Deka, S. (2019). Biodiversity of a tribal village in Assam with special emphasis on the traditional knowledge of plants. *Plants and Environment, 1*(2), 83-88.

Gogoi, P., Sen, S. (2023). Ethnobotanical study of medicinal plants used traditionally for managing cuts and wounds by the rural people of Kailashpur, Assam, India. *Ind. J. Pharm. Edu. Res, 57*(3s), s678-s689.

Gogoi, P., Nath, N. (2021). Diversity and inventorization of angiospermic flora in Dibrugarh district, Assam, Northeast India. *Plant Sci. Today* *8*(3), 621-628. https://doi.org/10.14719/pst.2021.8.3.1118

Gómez-Díaz, J. A., Krömer, T., Carvajal-Hernandez, C. I., Gerold, G., Heitkamp, F. (2017). Richness and distribution of herbaceous angiosperms along gradients of elevation and forest disturbance in central Veracruz, Mexico. *Botanical Sciences, 95*(2), 307-328.

Gong, L., Ding, X., Guan, W. et al. (2022). Comparative chloroplast genome analyses of *Amomum*: insights into evolutionary history and species identification. *BMC Plant Biology, 22*(1), 1-17.

Goremykin, V. V., Holland, B., Hirsch-Ernst, K. I., Hellwig, F. H. (2005). Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molecular Biology and Evolution, 22*(9), 1813-1822.

Goswami, U. C., Basistha, S. K., Bora, D. et al. (2012). Fish diversity of North East India, inclusive of the Himalayan and Indo Burma biodiversity hotspots zones: a checklist on their taxonomic status, economic importance, geographical distribution, present status and prevailing threats. *International Journal of Biodiversity and Conservation, 4*(15), 592-613.

Govaerts, R. (2001), How many species of seed plants are there?. *Taxon, 50*, 1085-1090. https://doi.org/10.2307/1224723

Govaerts, R., Nic Lughadha, E., Black, N. et al. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci Data 8*, 215. https://doi.org/10.1038/s41597-021-00997-6

Granick, S. (1961). The chloroplasts: inheritance, structure, and function. *The cell*, 489-602. Academic Press.

Gregory, T. R (2002). Genome size and developmental complexity. *Genetica, 115*, 131-146.

Gregory, T. R., Nathwani, P., Bonnett, T. R., Huber, D. P. (2013). Sizing up arthropod genomes: an evaluation of the impact of environmental variation on genome size estimates by flow cytometry and the use of qPCR as a method of estimation. *Genome, 56*(9), 505-510.

Greilhuber, J., Obermayer, R. (1997). Genome size and maturity group in *Glycine max* (soybean). *Heredity, 78*(5), 547-551.

Greilhuber, J., Obermayer, R. (1998). Genome size variation in *Cajanus cajan* (Fabaceae): a reconsideration. *Plant Systematics and Evolution*, *212*, 135-141.

Greilhuber, J. (1998). Intraspecific variation in genome size: a critical reassessment. *Annals of Botany*, *82*, 27-35.

Greilhuber, J., Dolezel, J., Lysák, M. A., Bennett, M. D. (2005). The origin, evolution and proposed stabilization of the terms 'genome size'and 'C-value'to describe nuclear DNA contents. *Annals of botany, 95*(1), 255-260.

Greilhuber, J., Temsch, E. M., Loureiro, J. C. (2007). Nuclear DNA content measurement. Flow cytometry with plant cells: analysis of genes, chromosomes and genomes, 67-101.

Greilhuber, J., Dolezel, J., Leitch, I. J., Loureiro, J., Suda, J. (2010). Genome size. *Journal of Botany*, *4*.

Greilhuber, J., Leitch, I. J. (2012). Genome size and the phenotype. In *Plant Genome Diversity Volume 2: physical structure, behaviour and evolution of plant genomes*, 323-344. Vienna: Springer Vienna.

Greiner, S., Lehwark, P., Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research, 47*(W1), W59-W64.

Grothuesi, D., Tümmler, B. (1991). New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of Pseudomonas species. *Molecular Microbiology*, *5*(11), 2763-2776.

Grotkopp, E., Rejmánek, M., Sanderson, M. J., Rost, T. L. (2004). Evolution of genome size in pines (*Pinus*) and its life-history correlates: super tree analyses. *Evolution, 58*(8), 1705-1729.

Gruber, V., Zahaf, O., Diet, A. et al. (2010). Impact of the environment on root architecture in dicotyledoneous plants. In *Root Genomics*, 113-132. Berlin, Heidelberg: Springer Berlin Heidelberg.

Gu, C., Ma, L., Wu, Z., Chen, K., Wang, Y. (2019). Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales. *BMC Plant Biology, 19*(1), 1-19.

Guan, Y. H., Liu, W. W., Duan, B. Z. et al. (2022). The first complete chloroplast genome of *Vicatia thibetica* de Boiss.: genome features, comparative analysis, and phylogenetic relationships. *Physiology and Molecular Biology of Plants, 28*(2), 439-454.

Guerra, T. J., Galetto, L., Silva, W. R. (2014). Nectar secretion dynamic links pollinator behavior to consequences for plant reproductive success in the ornithophilous mistletoe *Psittacanthus robustus*. *Plant Biol. (Stuttg) 16*(5), 956-966. https://doi.org/10.1111/plb.12146

Guignard, M. S., Crawley, M. J., Kovalenko, D. et al. (2019). Interactions between plant genome size, nutrients and herbivory by rabbits, molluscs and insects on a temperate grassland. *Proceedings of the Royal Society B, 286*(1899), 20182619.

Guisinger, M. M., Kuehl, J. V., Boore, J. L., Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution, 28*(1), 583-600.

Guo, Y. Y., Yang, J. X., Bai, M. Z., Zhang, G. Q., Liu, Z. J. (2021). The chloroplast genome evolution of *Venus slipper* (Paphiopedilum): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biology*, *21*(1), 248.

Gurung, J., Chettri, N., Sharma, E. et al. (2019). Evolution of a transboundary landscape approach in the Hindu Kush Himalaya: Key learnings from the Kangchenjunga Landscape. *Global Ecology and Conservation, 17*, e00599.

Habibullah, M. S., Din, B. H., Tan, S. H., Zahid, H. (2022). Impact of climate change on biodiversity loss: global evidence. *Environmental Science and Pollution Research, 29*(1), 1073-1086.

Hall, J.C., Sytsma, K.J., Iltis, H.H. (2002). Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data†. *Am. J. Bot., 89*, 1826-1842. https://doi.org/10.3732/ajb.89.11.1826

Hamabata, T., Kinoshita, G., Kurita, K. et al. (2019). Endangered island endemic plants have vulnerable genomes. *Communications Biology, 2*(1), 244.

Hamilton, J. P., Robin Buell, C. (2012). Advances in plant genome sequencing. *The Plant Journal, 70*(1), 177-190.

Han, S., Bi, D., Yi, R. et al. (2022). Plastome evolution of *Aeonium* and *Monanthes* (Crassulaceae): insights into the variation of plastomic tRNAs, and the patterns of codon usage and aversion. *Planta, 256*(2), 35.

Hanson, A. C., Horner-Devine, M. C., Martiny, J. B. H., Fuhrman, J. A. (2013). Microbial biogeography. *Encyclopedia of Biodiversity (2nd Eds.),* Waltham: Academic Press, 271–79. ISBN: 978-01-23847-20-1

Hardie, D. C., Gregory, T. R., Hebert, P. D. (2002). From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *Journal of Histochemistry Cytochemistry*, *50*(6), 735-749.

Hardie, D. C., Hebert, P. D. (2003). The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome, 46*(4), 683-706.

Harris, E. H., Boynton, J. E., Gillham, N. W. (1994). Chloroplast ribosomes and protein synthesis. *Microbiological Reviews, 58*(4), 700-754.

Hawthorne, W., Lawrence, A. (2013). Plant identification: creating user-friendly field guides for biodiversity management. Routledge.

Hazarika, B. B., Hazarika, D., Saikia, D. D., Nath, A. K., Hazarika, M. (2021). Resource base of North East India; problems and prospects: a descriptive study. *Turkish Online Journal of Qualitative Inquiry, 12*(6).

Hazarika, T. K. (2012). Citrus genetic diversity of north-east India, their distribution, ecogeography and ecobiology. *Genetic Resources and Crop Evolution, 59*, 1267-1280.

Hazarika, T. K., Lalramchuana, Nautiyal, B. P. (2012). Studies on wild edible fruits of Mizoram, India used as ethno-medicine. *Genetic Resources and Crop Evolution, 59*, 1767-1776.

He, S., Yang, Y., Li, Z., Wang, X., Guo, Y., Wu, H. (2020). Comparative analysis of four *Zantedeschia* chloroplast genomes: expansion and contraction of the IR region, phylogenetic analyses and SSR genetic diversity assessment. *PeerJ, 8*, e9132.

Heidmann, I., Schade-Kampmann, G., Lambalk, J., Ottiger, M., Di Berardino, M. (2016). Impedance flow cytometry: a novel technique in pollen analysis. *PloS One*, *11*(11), e0165531.

Hesse, U. (2023). k-mer based genome size estimation in theory and practice. In *Plant Cytogenetics and Cytogenomics: Methods and Protocols*, 79-113. New York, NY: Springer US.

Hidalgo, O., Pellicer, J., Christenhusz, M. J. M. et al. (2017). Is there an upper limit to genome size? *Trends Plant Sci, 22*, 567–573. doi: 10.1016/j.tplants.2017.04.005.

Hillebrand, H., Bennett, D. M., Cadotte, M. W. (2008). Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology, 89*(6), 1510-1520.

Hinz, R., Sulser, T. B., Hüfner, R. et al. (2020). Agricultural development and land use change in India: a scenario analysis of trade-offs between UN Sustainable Development Goals (SDGs). *Earth's Future, 8*(2), e2019EF001287.

Hoang, N. V., Furtado, A., Botha, F. C., Simmons, B. A., Henry, R. J. (2015). Potential for genetic improvement of sugarcane as a source of biomass for biofuels. *Frontiers in Bioengineering and Biotechnology, 3*, 182.

Houben, A., Demidov, D., Gernand, D. et al. (2003). Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *The Plant Journal, 33*(6), 967-973.

Howe, C. J., Barbrook, A. C., Koumandou, V. L. et al. (2003). Evolution of the chloroplast genome. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1429), 99-107.

Howlett, B. J. (1996). Application of pulsed field gel electrophoresis to the study of genomes of fungal pathogens of plants. *Australasian Plant Pathology*, *25*, 219-225.

Huang, X., Coulibaly, D., Tan, W. et al. (2022). The analysis of genetic structure and characteristics of the chloroplast genome in different Japanese apricot germplasm populations. *BMC Plant Biology, 22*(1), 1-13.

Huang, Y. Y., Cho, S. T., Haryono, M., Kuo, C. H. (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PloS One, 12*(6), e0179055.

Huang, Y., Li, J., Yang, Z. et al. (2022). Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. *BMC Plant Biology, 22*(1), 1-15.

Huo, Y., Gao, L., Liu, B. et al. (2019). Complete chloroplast genome sequences of four *Allium* species: comparative and phylogenetic analyses. *Scientific Reports, 9*(1), 12250.

Islam, M. A., Meister, A., Schubert, V. et al. (2007). Genetic diversity and cytogenetic analyses in *Curcuma zedoaria* (Christm.) Roscoe from Bangladesh. *Genet Resour Crop Evol, 54*, 149–156. https://doi.org/10.1007/s10722-005-2639-0

Jaccard, P. (1912). Thed distribution of the flora in the alpine zone-1. *New Phytologist 11*(2), 37-50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

Jaiswal, V. (2010). Culture and ethnobotany of Jaintia tribal community of Meghalaya, Northeast India- a mini review.

Jander, K., Herre, E. (2010). Host sanctions and pollinator cheating in the fig tree-fig wasp mutualism. *Proce. of the Royal Soc. B: Biological Sci. 277*(1687), 1481-1488. https://doi.org/10.1098/rspb.2009.2157

Jansen, R. K., Raubeson, L. A., Boore, J. L. et al. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. In *Methods in Enzymology, 395*, 348-384. Academic Press.

Jensen, P. E., Leister, D. (2014). Chloroplast evolution, structure and functions. *F1000prime reports*, *6*.

Jian, H. Y., Zhang, Y. H., Yan, H. J. et al. (2018). The complete chloroplast genome of a key ancestor of modern roses, *Rosa chinensis* var. spontanea, and a comparison with congeneric species. *Molecules, 23*(2), 389.

Jiang, H., Tian, J., Yang, J. et al. (2022). Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. *BMC Plant Biology, 22*(1), 1-21.

Jiji, P. (2015). Endangered plants and their uses of Sivasagar District, Assam, India. *International Research Journal of Biological Sciences, 4*(7), 15-18.

Jin, J. J., Yu, W. B., Yang, J. B. et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology, 21*, 1-31.

Joe, A., Sabu, M. (2016). Wild ornamental bananas in India: an overview. *South Indian Journal of Biological Science, 2*, 213-221.

Jordan, G. J., Carpenter, R. J., Koutoulis, A., Price, A., Brodribb, T. J. (2015). Environmental adaptation in stomatal size independent of the effects of genome size. *New Phytologist, 205*(2), 608-617.

Jovtchev, G., Schubert, V., Meister, A., Barow, M., Schubert, I. (2006). Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. *Cytogenetic and Genome Research, 114*(1), 77-82.

Käffer, M. I., Ganade, G., Marcelli, M. P. (2009). Lichen diversity and composition in Araucaria forests and tree monocultures in southern Brazil. *Biodiversity and Conservation, 18*, 3543-3561.

Kahraman, K., Lucas, S. J. (2019). Comparison of different annotation tools for characterization of the complete chloroplast genome of *Corylus avellana* cv Tombul. *BMC Genomics, 20*, 1-12.

Kalita, D., Saikia, J., Mukherjee, A. K., Doley, R. (2014). An ethnomedicinal survey of traditionally used medicinal plants for the treatment of snakebite in Morigaon district of Assam, India. *International Journal of Medicinal and Aromatic Plants, 4*(2), 97-106.

Kandel, P., Gurung, J., Chettri, N., Ning, W., Sharma, E. (2016). Biodiversity research trends and gap analysis from a transboundary landscape, Eastern Himalayas. *Journal of Asia-Pacific Biodiversity, 9*(1), 1-10.

Kanjilal, U. N., Kanjilal, P. C., Das, A., De, R. N., Bor, N. L. (1934). Flora of Assam. Vol. 1-5. Government of Assam, Shillong.

Kar, A., Borthakur, S. K. (2007). Wild vegetables sold in local markets of Karbi Anglong, Assam.

Kazakoff, S. H., Imelfort, M., Edwards, D. et al. (2012). Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feedstock tree *Pongamia pinnata*. *PLoS One, 7*(12), e51687.

Kelly, L. J., Leitch, I. J. (2011). Exploring giant plant genomes with next-generation sequencing technology. *Chromosome Research, 19*, 939-953.

Kelly, S., Kramer, S., Schwede, A., et al. (2012). Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open Biology, 2*(4), 120033.

Khan, A., Asaf, S., Khan, A. L., et al. (2019). Complete chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from Lamiaceae. *PeerJ, 7*, e7260.

Khan, S., Nabi, G., Ullah, M. W., et al. (2016). Overview on the role of advance genomics in conservation biology of endangered species. *International journal of genomics*.

Khazaei, H., Monneveux, P., Hongbo, S., Mohammady, S. (2010). Variation for stomatal characteristics and water use efficiency among diploid, tetraploid and hexaploid Iranian wheat landraces. *Genetic Resources and Crop Evolution, 57*, 307-314.

Killeen, T. J., Jardim, A., Mamani, F., Rojas, N. (1998). Diversity, composition and structure of a tropical semideciduous forest in the Chiquitania region of Santa Cruz, Bolivia. *Journal of Tropical Ecology, 14*(6), 803-827.

Kim, H. T., Lim, K. B., Kim, J. S. (2019). New insights on *Lilium* phylogeny based on a comparative phylogenomic study using complete plastome sequences. *Plants, 8*(12), 547.

Kim, K. J., Choi, K. S., Jansen, R. K. (2005). Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution, 22*(9), 1783-1792.

Kim, K. R., Park, S. Y., Kim, H., et al. (2023). Complete chloroplast genome determination of *Ranunculus sceleratus* from Republic of Korea (Ranunculaceae) and comparative chloroplast genomes of the members of the Ranunculus Genus. *Genes, 14*(6), 1149.

Kim, Y. K., Jo, S., Cheon, S. H., et al. (2020). Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Frontiers in Plant Science, 11*, 22.

Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytologist*, *223*(2), 565-574.

Kirchoff, B. K., Delaney, P. F., Horton, M., Dellinger-Johnston, R. (2014). Optimizing learning of scientific category knowledge in the classroom: The case of plant identification. *CBE—Life Sciences Education, 13*(3), 425-436.

Knight, C. A., Ackerly, D. D. (2002). Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters, 5*(1), 66-76.

Knight, C. A., Beaulieu, J. M. (2008). Genome size scaling through phenotype space. *Annals of Botany*, *101*(6), 759-766.

Knight, C. A., Molinari, N. A., Petrov, D. A. (2005). The large genome constraint hypothesis: evolution, ecology and phenotype. *Annals of Botany, 95*(1), 177-190.

Kong, W. Q., Yang, J. H. (2017). The complete chloroplast genome sequence of *Morus cathayana* and *Morus multicaulis*, and comparative analysis within genus *Morus* L. *PeerJ*, *5*, e3037.

Kormelink, T. G., Arkesteijn, G. J., Nauwelaers, F. A., et al. (2016). Prerequisites for the analysis and sorting of extracellular vesicle subpopulations by high-resolution flow cytometry. *Cytometry Part A*, *89*(2), 135-147.

Koutecký, P., Smith, T., Loureiro, J., Kron, P. (2023). Best practices for instrument settings and raw data analysis in plant flow cytometry. *Cytometry Part A*.

Krahulcová, A., Trávníček, P., Krahulec, F., Rejmánek, M. (2017). Small genomes and large seeds: chromosome numbers, genome size and seed mass in diploid *Aesculus* species (Sapindaceae). *Annals of Botany, 119*(6), 957-964.

Krauss, S. L., Phillips, R. D., Karron, J. D., et al. (2017). Novel consequences of bird pollination for plant mating. *Trends in Plant Science, 22*(5), 395-410.

Kreft, H., Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences, 104*(14), 5925-5930.

Kreplak, J., Madoui, M. A., Cápal, P., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nature Genetics, 51*(9), 1411-1422.

Krishan, A. (1975). Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *The Journal of Cell Biology, 66*(1), 188-193.

Kron, P., Husband, B. C. (2012). Using flow cytometry to estimate pollen DNA content: improved methodology and applications. *Annals of Botany, 110*(5), 1067-1078.

Lane, I. G., Herron-Sweet, C. R., Portman, Z. M., Cariveau, D. P. (2020). Floral resource diversity drives bee community diversity in prairie restorations along an agricultural landscape gradient. *Journal of Applied Ecology, 57*(10), 2010-2018.

Lattier, J. D., Chen, H., Contreras, R. N. (2019). Variation in genome size, ploidy, stomata, and rDNA signals in Althea. *Journal of the American Society for Horticultural Science, 144*(2), 130-140.

Lavergne, S., Muenke, N. J., Molofsky, J. (2010). Genome size reduction can trigger rapid phenotypic evolution in invasive plants. *Annals of Botany, 105*(1), 109-116.

Leach, K., Montgomery, W. I., Reid, N. (2016). Modelling the influence of biotic factors on species distribution patterns. *Ecological Modelling, 337*, 96-106.

Lee, S. I., Kim, N. S. (2014). Transposable elements and genome size variations in plants. *Genomics Informatics, 12*(3), 87.

Lee, S. R., Kim, K., Lee, B. Y., Lim, C. E. (2019). Complete chloroplast genomes of all six *Hosta* species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. *BMC Genomics, 20*(1), 1-13.

Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O., Bennett, M. D. (2019). Angiosperm DNA C-values database (release 9.0, Apr 2019).

Leitch, I. J., Leitch, A. R. (2012). Genome size diversity and evolution in land plants. In *Plant genome diversity Volume 2*: *Physical structure, behaviour and evolution of plant genomes*, 307-322. Vienna: Springer Vienna.

Leitch, I. J., Kahandawala, I., Suda, J., et al. (2009). Genome size diversity in orchids: consequences and evolution. *Annals of Botany, 104*(3), 469-481.

Li, B., Zheng, Y. (2018). Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Scientific Reports, 8*(1), 9285.

Li, C., Zhou, L., Nie, J., et al. (2023). Codon usage bias and genetic diversity in chloroplast genomes of *Elaeagnus* species (Myrtiflorae: Elaeagnaceae). *Physiology and Molecular Biology of Plants, 29*(2), 239-251.

Li, F. W., Harkess, A. (2018). A guide to sequence your favorite plant genomes. *Applications in Plant Sciences*, *6*(3), e1030.

Li, G., Pan, Z., Gao, S., et al. (2019). Analysis of synonymous codon usage of chloroplast genome in *Porphyra umbilicalis*. *Genes Genomics, 41*, 1173-1181.

Li, G., Zhang, L., Xue, P. (2021). Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species. *Gene, 802*, 145866.

Li, G., Zhang, L., Xue, P., Zhu, M. (2023). Comparative analysis on the codon usage pattern of the chloroplast genomes in *Malus* species. *Biochemical Genetics, 61*(3), 1050-1064.

Li, W., Liu, Y., Yang, Y., et al. (2018). Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. *BMC Plant Biology, 18*(1), 1-11.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*(1), 157-166.

Li, Y., Zhang, L., Wang, T., et al. (2022). The complete chloroplast genome sequences of three lilies: Genome structure, comparative genomic and phylogenetic analyses. *Journal of Plant Research, 135*(6), 723-737.

Liang, C., Wang, L., Lei, J., et al. (2019). A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants. *Engineering, 5*(5), 907-915.

Liao, M., Wei, X. F., Ding, H. P., Tang, G. D. (2020). The complete chloroplast genome of the highly poisonous plant *Cerbera manghas* L.(Apocynaceae). *Mitochondrial DNA Part B, 5*(3), 3084-3085.

Lin, C. S., Chen, J. J., Huang, Y. T., et al. (2015). The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family. *Scientific Reports, 5*(1), 9040.

Lipnerová, I., Bureš, P., Horová, L., Šmarda, P. (2013). Evolution of genome size in *Carex* (Cyperaceae) in relation to chromosome number and genomic base composition. *Annals of Botany, 111*(1), 79-94.

Liu, B., Zhang, D., Gao, L. Z. (2016). The complete chloroplast genome sequence of *Cucumis sativus* var. Hardwickii, the wild progenitor of cultivated cucumber. *Mitochondrial DNA Part A, 27*(6), 4627-4628.

Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., Guan, X. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics, 13*, 1-7.

Liu, J., Niu, Y. F., Ni, S. B., et al. (2018). The complete chloroplast genome of *Artocarpus heterophyllus* (Moraceae). *Mitochondrial DNA Part B*, *3*(1), 13-14.

Liu, X. Q., Ickert-Bond, S. M., Chen, L. Q., Wen, J. (2013). Molecular phylogeny of *Cissus* L. of Vitaceae (the grape family) and evolution of its pantropical intercontinental disjunctions. *Mol. Phylogenet. Evol.*, *66*:43–53. doi:10.1016/j.ympev.2012.09.003

Liu, Y., Chen, W., Li, F., et al. (2019). The complete chloroplast genome sequence of *Ficus hirta* (Moraceae). *Mitochondrial DNA Part B*, *4*(2), 4041-4042.

Liu, Z. F., Ma, H., Ci, X. Q., et al. (2021). Can plastid genome sequencing be used for species identification in Lauraceae?. *Botanical Journal of the Linnean Society, 197*(1), 1-14.

Llamas, E., Pulido, P., Rodriguez-Concepcion, M. (2017). Interference with plastome gene expression and Clp protease activity in Arabidopsis triggers a chloroplast unfolded protein response to restore protein homeostasis. *PLoS Genetics, 13*(9), e1007022.

Lodh, S., Swamy, M. K. (2019). Phytochemical aspects of medicinal plants of Northeast India to improve the gynaecological disorders: an update. *Natural Bio-active Compounds: Volume 2: Chemistry, Pharmacology and Health Care Practices,* 353-367.

Loeuille, B., Thode, V., Siniscalchi, C., Andrade, S., Rossi, M., Pirani, J. R. (2021). Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *PeerJ, 9*, e10886.

Loureiro, J., Rodriguez, E., Doležel, J., Santos, C. (2006). Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Annals of Botany, 98*(3), 679-689.

Loureiro, J., Čertner, M., Lučanová, M., et al. (2023). The use of flow cytometry for estimating genome sizes and DNA ploidy levels in plants. In *Plant Cytogenetics and Cytogenomics: Methods and Protocols*, 25-64. New York, NY: Springer US.

Loureiro, J., Kron, P., Temsch, E. M., et al. (2021). Isolation of plant nuclei for estimation of nuclear DNA content: overview and best practices. *Cytometry Part A, 99*, 318 - 327.

Lysak, M. A., Dolezel, J. (1998). Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia, 51*(2), 123-132.

Lysak, M. A., Koch, M. A., Beaulieu, J. M., Meister, A., Leitch, I. J. (2009). The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution, 26*(1), 85-98.

Lyseng, M. P., Bork, E. W., Hewins, D. B., et al. (2018). Long-term grazing impacts on vegetation diversity, composition, and exotic species presence across an aridity gradient in northern temperate grasslands. *Plant Ecology, 219*, 649-663.

Ma, Y. P., Zhao, L., Zhang, W. J., et al. (2020). Origins of cultivars of *Chrysanthemum*—Evidence from the chloroplast genome and nuclear LFY gene. *Journal of Systematics and Evolution, 58*(6), 925-944.

Macas, J., Novák, P., Pellicer, J., et al. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PloS one*, *10*(11), e0143424.

Macey, M. G., Macey, M. G. (2007). *Flow cytometry*. Berlin, Germany, Springer.

Mahood, E. H., Kruse, L. H., Moghe, G. D. (2020). Machine learning: a powerful tool for gene function prediction in plants. *Applications in Plant Sciences, 8*(7), e11376.

Mani, M. S. (1974). Biogeographical evolution in India. In *Ecology and biogeography in India*, 698-724). Dordrecht: Springer Netherlands.

Manohar, S. M., Shah, P., Nair, A. (2021). Flow cytometry: Principles, applications and recent advances. *Bioanalysis*, *13*(3), 181-198.

Mao, A. A., Roy, D. K. (2016). Ethnobotanical studies in North East India: a review. *Indian ethnobotany: emerging trends. New Delhi: Scientific, 1*, 99-112.

Mao, A. A., Dash, S. S. (2019). Plant Discoveries. *BSI, Kolkata*, 1-163.

Marañón, T., Grubb, P. J. (1993). Physiological basis and ecological significance of the seed size and relative growth rate relationship in Mediterranean annuals. *Functional Ecology*, 591-599.

Margalef, R. (1968). Perspectives in ecological theory. *University of Chicago press, Chicago*, 111. https://doi.org/10.4319/lo.1969.14.2.0313

Marie, D., Brown, S. C. (1993). A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biology of the Cell, 78*(1-2), 41-51.

Markov, A. V., Anisimov, V. A., Korotayev, A. V. (2010). Relationship between genome size and organismal complexity in the lineage leading from prokaryotes to mammals. *Paleontological Journal, 44*, 363-373.

Masierowska, M., Stawiarz, E., Rozwałka, R. (2018). Perennial ground cover plants as floral resources for urban pollinators: A case of *Geranium* species. *Urban Forestry Urban Greening, 32*, 185-194.

Mazumder, M. K. (2014). Diversity, habitat preferences, and conservation of the primates of southern Assam, India: the story of a primate paradise. *Journal of Asia-Pacific Biodiversity, 7*(4), 347-354.

McKain, M. R., Hartsock, R. H., Wohl, M. M., Kellogg, E. A. (2017). Verdant: automated annotation, alignment and phylogenetic analysis of whole chloroplast genomes. *Bioinformatics, 33*(1), 130-132.

Mehmood, F., Ubaid, Z., Shahzadi, I., et al. (2020). Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). *PeerJ, 8*, e9552.

Mehra, P. N., Bawa, K. S. (1969). Chromosomal evolution in tropical hardwoods. *Evolution*, 466-481.

Mei, W., Stetter, M. G., Gates, D. J., Stitzer, M., Ross-Ibarra, J. (2017). Adaptation in plant genomes: bigger isn't better, but it's probably different. *bioRxiv*, 196501.

Michael, T. P. (2014). Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics, 13*(4), 308-317.

Midoko-Iponga, D., Krug, C. B., Milton, S. J. (2005). Competition and herbivory influence growth and survival of shrubs on old fields: implications for restoration of renosterveld shrubland. *Journal of Vegetation Science, 16*(6), 685-692.

Miflin, B. (2000). Crop improvement in the 21st century. *Journal of Experimental Botany*, *51*(342), 1-8.

Annual Report (2013-14). Ministry of Environment, forests and Climate Change (MoEF). http://moef.gov.in/wp-content/uploads/2018/04/AR-2013-14-Eng.pdf (Accessed on 5th Nov, 2023)

Mir, A. H., Upadhaya, K., Choudhury, H. (2014). Diversity of endemic and threatened ethnomedicinal plant species in Meghalaya, North-East India. *Int Res J Environ Sci, 3*(12), 64-78.

Mo, Z., Lou, W., Chen, Y., et al. (2020). The chloroplast genome of *Carya illinoinensis*: genome structure, adaptive evolution, and phylogenetic analysis. *Forests*, *11*(2), 207.

Moeglein, M. K., Chatelet, D. S., Donoghue, M. J., Edwards, E. J. (2020). Evolutionary dynamics of genome size in a radiation of woody plants. *American Journal of Botany, 107*(11), 1527-1541.

Mohanty, S., Das, A. B. (2006). Study of karyotype variability and genome size in 13 species of *Cassia* L. in interpreting interspecific genetic diversity. *Cytologia, 71*(3), 261-267.

Mu, J., Zhao, Y., He, Y., Sun, J., Yuan, Q. (2021). The complete chloroplast genome of *Prunus japonica* thunb.(Rosaceae), an ornamental and medicinal plant. *Mitochondrial DNA Part B, 6*(1), 112-114.

Mutke, J., Barthlott, W. (2005). Patterns of vascular plant diversity at continental to global scales. *Biologiske Skrifter, 55*(4), 521-531.

Naidu, M. T., Kumar, O. A. (2016). Tree diversity, stand structure, and community composition of tropical forests in Eastern Ghats of Andhra Pradesh, India. *J. of Asia-Pacific Biod. 9*(3), 328-334. https://doi.org/10.1016/j.japb.2016.03.019

Najafi, S., Ulker, M., Oral, E., et al. (2022). Estimation of nuclear DNA content in some *Aegilops* species: best analyzed using flow cytometry. *Genes, 13*(11), 1980.

Nath, S., Mallick, S. K., Jha, S. (2014). An improved method of genome size estimation by flow cytometry in five mucilaginous species of *Hyacinthaceae*. *Cytometry Part A*, *85*(10), 833-840.

Neale, D. B., Martinez-Garcia, P. J., De La Torre, A. R., Montanari, S., Wei, X. X. (2017). Novel insights into tree biology and genome evolution as revealed through genomics. *Annual Review of Plant Biology, 68*, 457-483.

Nie, X., Deng, P., Feng, K., et al. (2014). Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Molecular Biology Reporter, 32*, 828-840.

Ninawe, A. S., Swapna, T. S. (2017). Orchid diversity of Northeast India–traditional knowledge and strategic plan for conservation. *Journal of Orchid Society of India, 31*, 41-56.

Ning, C. U. I., Bao-Sheng, L. I. A. O., Liang, C. L., et al. (2020). Complete chloroplast genome of *Salvia plebeia*: organization, specific barcode and phylogenetic analysis. *Chinese journal of natural medicines*, *18*(8), 563-572.

Niu, Z., Xue, Q., Zhu, S., Sun, J., Liu, W., Ding, X. (2017). The complete plastome sequences of four orchid species: insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots. *Frontiers in Plant Science, 8*, 715.

Nock, C. J., Waters, D. L., Edwards, M. A., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, *9*(3), 328-333.

Ohno, S. (1972). So much "junk" DNA in our genome. In *Evolution of Genetic Systems. In Brookhaven Symposium in Biology,* 23, 366-370.

Ohri, D., Kumar, A. (1986). Nuclear DNA amounts in some tropical hardwoods. *Caryologia, 39*(3-4), 303-307.

Ohri, D. (1996). Genome size and polyploidy variation in the tropical hardwood genus *Terminalia* (Combretaceae). *Plant Systematics and Evolution, 200*, 225-232.

Ohri, D. (1998). Genome size variation and plant systematics. *Annals of botany, 82*, 75-83.

Ohri, D., Bhargava, A., Chatterjee, A. (2004). Nuclear DNA amounts in 112 species of tropical hardwoods- new estimates. *Plant Biology, 6*(5), 555-561.

Ohri, D. (2005). Climate and growth form: the consequences for genome size in plants. *Plant Biology, 7*(5), 449-458.

Oktaviani, I., Asril, M., Aryanti, Y., Leksikowati, S. S. (2019). A systematic survey of plant biodiversity study within the land of Institut Teknologi Sumatera (ITERA). *Journal of Science and Applicative Technology, 2*(1), 47-52.

Olejniczak, S. A., Łojewska, E., Kowalczyk, T., Sakowicz, T. (2016). Chloroplasts: state of research and practical applications of plastome sequencing. *Planta, 244*, 517-527.

Oliveira, W., Silva, J. L., Porto, R. G., et al. (2020). Plant and pollination blindness: risky business for human food security. *BioScience, 70*(2), 109-110.

Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M., Edwards, S. V. (2007). Origin of avian genome size and structure in non-avian dinosaurs. *Nature, 446*(7132), 180-184.

Otto, F. (1990). DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In *Methods in cell biology*, *33*, 105-110. Academic Press.

Oulo, M. A., Yang, J. X., Dong, X., et al. (2020). Complete chloroplast genome of *Rhipsalis baccifera*, the only cactus with natural distribution in the old world: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Plants*, *9*(8), 979.

Pal, R., Singh, D. R. (2016). Endemic orchids of Northeast India. In *Purkayastha, J. (eds) Bioprospecting of Indigenous Bioresources of North-East India*. Springer, Singapore. https://doi.org/10.1007/978-981-10-0620-3_3

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annual Review of Genetics, 19*(1), 325-354.

Panda, P. C., Mahapatra, A. K., Acharya, P. K., Debata, A. K. (2013). Plant diversity in tropical deciduous forests of Eastern Ghats, India: a landscape level assessment. *Int J Biodivers Conserv, 5*(10), 625-639.

Pandey, A., Negi, P. S. (2016). Traditional uses, phytochemistry and pharmacological properties of *Neolamarckia cadamba*: a review. *Journal of Ethnopharmacology, 181*, 118-135.

Park, I., Yang, S., Kim, W. J., et al. (2019). Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species identification. *Molecules, 24*(6), 1038.

Park, J., Oh, S. H. (2020). A second complete chloroplast genome sequence of *Fagus multinervis* Nakai (Fagaceae): intraspecific variations on chloroplast genome. *Mitochondrial DNA Part B*, *5*(2), 1868-1869.

Parvathy, S. T., Udayasuriyan, V., Bhadana, V. (2022). Codon usage bias. *Molecular Biology Reports*, *49*(1), 539-565.

Patil, A. E., Deosarkar, D. B., Khatri, N., Ubale, A. B. (2023). A comprehensive investigation of Genotype-Environment interaction effects on seed cotton yield contributing traits in *Gossypium hirsutum* L. Using multivariate analysis and artificial neural network. *Computers and Electronics in Agriculture*, *211*, 107966.

Pellicer, J., Leitch, I. J. (2014). The application of flow cytometry for estimating genome size and ploidy level in plants. *Molecular Plant Taxonomy: Methods and Protocols*, 279-307

Pellicer, J., Leitch, I. J. (2019). The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist*.

Pellicer, J., Fay, M. F., Leitch, I. J. (2010). The largest eukaryotic genome of them all?. *Botanical Journal of the Linnean Society, 164*(1), 10-15.

Pellicer, J., Hidalgo, O., Dodsworth, S., Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes, 9*(2), 88.

Pellicer, J., Powell, R. F., Leitch, I. J. (2021). The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants. *Molecular Plant Taxonomy: Methods and Protocols*, 325-361.

Pfosser, M., Heberle-Bors, E., Amon, A., Lelley, T. (1995). Evaluation of sensitivity of flow cytometry in detecting aneuploidy in wheat using disomic and ditelosomic wheat–rye addition lines. *Cytometry: The Journal of the International Society for Analytical Cytology, 21*(4), 387-393.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology, 13,* 131-144.

Ping, J., Feng, P., Li, J., Zhang, R., Su, Y., Wang, T. (2021). Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecology and Evolution, 11*(9), 4786-4802.

Powell, R. F., Pulido Suarez, L., Magee, A. R., et al. (2020). Genome size variation and endopolyploidy in the diverse succulent plant family Aizoaceae. *Botanical Journal of the Linnean Society, 194*(1), 47-68.

Pozarowski, P., Grabarek, J., Darzynkiewicz, Z. (2003). Flow cytometry of apoptosis. *Current Protocols in Cell Biology, 21*(1), 18-8.

Pradhan, N. S., Das, P. J., Gupta, N., Shrestha, A. B. (2021). Sustainable management options for healthy rivers in South Asia: the case of Brahmaputra. *Sustainability (Switzerland), 13*(3), 1087. https://doi.org/10.3390/su13031087

Prenna, G., Leiva, S., Mazzini, G. (1974). Quantitation of DNA by cytofluorometry of the conventional Feulgen reaction. *The Histochemical Journal*, *6*, 467-489.

Price, H. J., Johnston, J. S. (1996). Analysis of plant DNA content by Feulgen microspectrophotometry and flow cytometry. *Methods of Genome Analysis in Plants.–Boca Raton (CRC Press Inc) S*, 115-132.

Price, H. J., Hodnett, G., Johnston, J. S. (2000). Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. *Annals of Botany, 86*(5), 929-934.

Pucker, B. (2019). Mapping-based genome size estimation. *Biorxiv*, 607390.

Pucker, B., Irisarri, I., de Vries, J., Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quantitative Plant Biology, 3*, e5.

Pyšek, P., Lučanová, M., Dawson, W., et al. (2023). Small genome size and variation in ploidy levels support the naturalization of vascular plants but constrain their invasive spread. *New Phytologist*, *239*(6), 2389-2403.

Qi, Y., Xu, W., Xing, T., et al. (2015). Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity. *Evolutionary Bioinformatics*, *11*, EBO-S22566.

Qiu, Y., Hamernick, S., Ortiz, J. B., Watkins, E. (2020). DNA content and ploidy estimation of *Festuca ovina* accessions by flow cytometry. *Crop Science, 60*(5), 2757-2767.

Qiu, Y., Hirsch, C. D., Yang, Y., Watkins, E. (2019). Towards improved molecular identification tools in fine fescue (*Festuca* L., Poaceae) turfgrasses: nuclear genome size, ploidy, and chloroplast genome sequencing. *Frontiers in Genetics, 10*, 1223.

Qu, X. J., Zou, D., Zhang, R. Y., Stull, G. W., Yi, T. S. (2023). Progress, challenge and prospect of plant plastome annotation. *Frontiers in Plant Science, 14*, 1166140.

Quraishi, H. A., Islam, N., Iqbal, A., et al. (2018). Therapeutical and medicinal properties of Neem (*Azadirachta indica*) in context of Unani system of medicine: a review study. *Journal of Drug Delivery and Therapeutics, 8*(6-s), 394-399.

Rabah, S. O., Lee, C., Hajrah, N. H., et al. (2017). Plastome sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. *The Plant Genome, 10*(3).

Raghu, K., Naidoo, Y., Dewir, Y. H. (2019). Secretory structures in the leaves of *Hibiscus sabdariffa* L. *South African Journal of Botany, 121*, 16-25.

Rahman, M. H., Khan, M. A. S. A., Roy, B., Fardusi, M. J. (2011). Assessment of natural regeneration status and diversity of tree species in the biodiversity conservation areas of Northeastern Bangladesh. *Journal of Forestry Research, 22*, 551-559.

Rai, P. K., Lalramnghinglova, H. (2011). Ethnomedicinal plants of India with special reference to an Indo-Burma hotspot region: an overview. *Ethnobotany Research and Applications, 9,* 379-420.

Raju, V. S., Reddy, C. S., Suthari, S. (2010). Flowering plant diversity and endemism in India: an overview. *Algae, 2*(40,000), 6-25.

Ramesh, A. M., Basak, S., Choudhury, R. R., Rangan, L. (2014). Development of flow cytometric protocol for nuclear DNA content estimation and determination of chromosome number in *Pongamia pinnata* L., a valuable biodiesel plant. *Applied Biochemistry and Biotechnology, 172*, 533–548. https://doi.org/10.1007/s12010-013-0553-z

Rani, C., Chawla, S., Mangal, M., et al. (2012). *Nyctanthes arbor-tristis* Linn.(Night Jasmine): a sacred ornamental plant with immense medicinal potentials.

Rasch, E. M. (2004). Feulgen-DNA cytophotometry for estimating C values. *Drosophila Cytogenetics Protocols*, 163-201.

Rayburn, A. L., Auger, J. A., Benzinger, E. A., Hepburn, A. G. (1989). Detection of intraspecific DNA content variation in *Zea mays* L. by flow cytometry. *Journal of Experimental Botany, 40*(11), 1179-1183.

Ren, W., Jiang, Z., Zhang, M., et al. (2022). The chloroplast genome of *Salix floderusii* and characterization of chloroplast regulatory elements. *Frontiers in Plant Science, 13*, 987443.

Řepková, J. (2010). Potential of chloroplast genome in plant breeding. *Czech Journal of Genetics and Plant Breeding, 46*(3), 103-113.

Rion, V., Gallandat, J. D., Gobat, J. M., Vittoz, P. (2018). Recent changes in the plant composition of wetlands in the Jura Mountains. *Applied Vegetation Science, 21*(1), 121-131.

Rogalski, M., do Nascimento Vieira, L., Fraga, H. P., Guerra, M. P. (2015). Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant Science, 6,* 586.

Rossini, B. C., de Moraes, M. L. T., Marino, C. L. (2021). Complete chloroplast genome of *Myracrodruon urundeuva* and its phylogenetics relationships in Anacardiaceae family. *Physiology and Molecular Biology of Plants, 27*, 801-814.

Roswell, M., Dushoff, J., Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos, 130*(3), 321-338.

Roy, A. B., Purohit, R. (2018). Indian shield: precambrian evolution and Phanerozoic reconstitution. Elsevier.

Roy, A., Das, S. K., Tripathi, A. K., Singh, N. U., Barman, H. K. (2015). Biodiversity in North East India and their conservation. *Progressive Agriculture, 15*(2), 182-189.

Roy, P. S., Saran, S. (2004). Biodiversity Information System for North East India. *Geocarto International, 19*(3), 73-80.

Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., Burleigh, J. G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol*, *14*.

Ruhlman, T. A., Jansen, R. K. (2014). The plastid genomes of flowering plants. *Chloroplast Biotechnology: Methods and Protocols*, 3-38.

Saadi, S. M. A. I., Mondal, I., Sarkar, S., Mondal, A. K. (2020). Medicinal plants diversity modelling using remote sensing GIS technology of Chilkigarh, West Bengal, India. *Trop. Plant Res, 7*(2), 440-451.

Sadhu, A., Bhadra, S., Bandyopadhyay, M. (2016). Novel nuclei isolation buffer for flow cytometric genome size estimation of Zingiberaceae: a comparison with common isolation buffers. *Annals of Botany, 118*(6), 1057-1070.

Saikia, A. P., Ryakala, V. K., Sharma, P., Goswami, P., Bora, U. (2006). Ethno-botany of medicinal plants used by Assamese people for various skin ailments and cosmetics. *J. of Ethnopharmacol. 106*(2), 149-57. https://doi.org/10.1016/j.jep.2005.11.033

Saikia, P., Choudhury, B., Khan, M. L. (2012). Floristic composition and plant utilization pattern in homegardens of Upper Assam, India. *Trop. Ecol. 53*(1), 105-118. ISSN (online): 0564-3295

Sajem, A. L., Rout, J., Nath, M. (2008). Traditional tribal knowledge and status of some rare and endemic medicinal plants of North Cachar Hills district of Assam, Northeast India. *Ethnobotanical Leaflets, 2008*(1), 31.

Salehi, B., Konovalov, D. A., Fru, P., et al. (2020). *Areca catechu*—From farm to food and biomedical applications. *Phytotherapy Research, 34*(9), 2140-2158.

Salse, J., Feuillet, C. (2007). Comparative genomics of cereals. *Genomics-Assisted Crop Improvement: Vol. 1: Genomics Approaches and Platforms*, 177-205.

Santos, V., Silva, E. F. D., Almeida, C. (2020). Genome size and identification of repetitive DNA sequences using low coverage sequencing in *Hancornia speciosa* Gomes (Apocynaceae: Gentianales). *Genetics and Molecular Biology*, *43*.

Sarkar, D., Kundu, A., Saha, A., et al. (2011). First nuclear DNA amounts in diploid (2 n= 2 x= 14) *Corchorus* spp. by flow cytometry: genome sizes in the cultivated jute species (*C. capsularis* L. and *C. olitorius* L.) are~ 300% smaller than the reported estimate of 1100–1350 Mb. *Caryologia, 64*(2), 147-153.

Sarma, N., Shankar, U., Mao, A. A. (2019). An assessment of the invasive flora of Amchang Wildlife Sanctuary in the Kamrup District of Assam, India.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research, 6*(5), 283-290.

Scheunert, A., Dorfner, M., Lingl, T., Oberprieler, C. (2020). Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *PLoS One, 15*(3), e0226234.

Schmitz-Linneweber, C., Tillich, M., Herrmann, R. G., Maier, R. M. (2001). Heterologous, splicing-dependent RNA editing in chloroplasts: allotetraploidy provides trans-factors. *The EMBO Journal, 20*(17), 4874-4883.

Scholes, D. R., Paige, K. N. (2015). Plasticity in ploidy: a generalized response to stress. *Trends in Plant Science*, *20*(3), 165-175.

Schubert, I., Vu, G. T. (2016). Genome stability and evolution: attempting a holistic view. *Trends in Plant Science, 21*(9), 749-757.

Schuldt, A., Ebeling, A., Kunz, M., et al. (2019). Multiple plant diversity components drive consumer communities across ecosystems. *Nature Communications, 10*(1), 1460.

Scobeyeva, V. A., Artyushin, I. V., Krinitsina, A. A., et al. (2021). Gene loss, pseudogenization in plastomes of genus *Allium* (Amaryllidaceae), and putative selection for adaptation to environmental conditions. *Frontiers in Genetics, 12*, 674783.

Seddon, N., Smith, A., Smith, P., et al. (2021). Getting the message right on nature-based solutions to climate change. *Global Change Biology, 27*(8), 1518-1546.

Sedgwick, P. (2012). Pearson’s correlation coefficient. *BMJ*, *345*.

Sekar, K. C., Thapliyal, N., Pandey, A., et al. (2023). Plant species diversity and density patterns along altitude gradient covering high-altitude alpine regions of west Himalaya, India. *Geology, Ecology, and Landscapes*, 1-15.

Semeraro, T., Turco, A., Arzeni, S., et al. (2021). Habitat restoration: an applicative approach to “biodiversity heritage relicts” in social-ecological systems. *Land, 10*(9), 898.

Senapati, A., Basak, S., Rangan, L. (2022). A review on application of DNA barcoding technology for rapid molecular diagnostics of adulterants in herbal medicine. *Drug Safety, 45*(3), 193-213.

Senapati, A., Chetri, B. K., Mitra, S., Shelke, R. G., Rangan, L. (2023). Decoding the complete chloroplast genome of *Cissus quadrangularis*: insights into molecular structure, comparative genome analysis and mining of mutational hotspot regions. *Physiology and Molecular Biology of Plants*, 1-16.

Servet, C., Aktağ, A. (2019). The composition of woody plants in the cemeteries of various religious communities in Istanbul-Turkey. *Urban Forestry Urban Greening, 43*, 126350.

Shang, C., Li, E., Yu, Z., et al. (2022). Chloroplast genomic resources and genetic divergence of endangered species *Bretschneidera sinensis* (Bretschneideraceae). *Frontiers in Ecology and Evolution, 10*, 873100.

Shannon, C. E., Weaver, W. W. (1963). The mathematical theory of communication. *The University of Illinois press, Urbana*, 117

Sharma, N., Behera, M. D., Das, A. P., Panda, R. M. (2019). Plant richness pattern in an elevation gradient in the Eastern Himalaya. *Biodiversity and Conservation, 28*, 2085-2104.

Sharma, S., Kaushik, S., Raina, S. N. (2019). Estimation of nuclear DNA content and its variation among Indian Tea accessions by flow cytometry. *Physiology and Molecular Biology of Plants, 25,* 339-346.

Shelke, R. G., Rangan, L. (2020). The role of transposable elements in *Pongamia* unigenes and protein diversity. *Molecular Biotechnology, 62*(1), 31-42.

Shelke, R. G., Rangan, L. (2022). The whole chloroplast genome of *Mesua ferrea*: Insight into the dynamic pattern of evolution and its comparison with species from recently diverged families. *Gene, 846*, 146866.

Shelke, R. G., Banerjee, R. P., Joshi, B., et al. (2022). Chloroplast genome of *Lithocarpus dealbatus* (Hook. f. Thomson ex Miq.) Rehder establishes monophyletic origin of the species and reveals mutational hotspots with taxon delimitation potential. *Life, 12*(6), 828.

Shelke, R. G., Basak, S., Rangan, L. (2020). Development of EST-SSR markers for *Pongamia pinnata* by transcriptome database mining: cross-species amplification and genetic diversity. *Physiology and Molecular Biology of Plants, 26*, 2225-2241.

Shelke, S., Das, B. (2015). Dose response and adaptive response of non-homologous end joining repair genes and proteins in resting human peripheral blood mononuclear cells exposed to γ radiation. *Mutagenesis, 30*(3), 365-379.

Shen, J., Zhang, X., Landis, J. B., et al. (2020). Plastome evolution in Dolomiaea (Asteraceae, Cardueae) using phylogenomic and comparative analyses. *Frontiers in Plant Science, 11*, 376.

Sheng, J., She, X., Liu, X., Wang, J., Hu, Z. (2021). Comparative analysis of codon usage patterns in chloroplast genomes of five *Miscanthus* species and related species. *PeerJ, 9*, e12173.

Shidhi, P. R., Nadiya, F., Biju, V. C., et al. (2021). Complete chloroplast genome of the medicinal plant *Evolvulus alsinoides*: comparative analysis, identification of mutational hotspots and evolutionary dynamics with species of Solanales. *Physiology and Molecular Biology of Plants, 27*(8), 1867-1884.

Shneyer, V. S., Rodionov, A. V. (2019). Plant DNA barcodes. *Biology Bulletin Reviews*, *9*(4), 295-300.

Simeone, M. C., Grimm, G. W., Papini, A., et al. (2016). Plastome data reveal multiple geographic origins of Quercus Group Ilex. *PeerJ, 4*, e1897.

Simonin, K. A., Roddy, A. B. (2018). Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biology, 16*(1), e2003706.

Simpson, E. H. (1949). Measurement of diversity. *Nature 163*, 688. https://doi.org/10.1038/163688a0

Singh, A. P. (2017). Butterflies associated with major forest types in Arunachal Pradesh (Eastern Himalaya), India: implications for ecotourism and conservation planning. *Journal of Threatened Taxa, 9*(4), 10047-10075.

Singh, J. S., Chaturvedi, R. K. (2017). Diversity of ecosystem types in India: Aa review. *Proceedings of the Indian National Science Academy, 83*(1), 569–594. https://doi.org/10.16943/ptinsa/2017/41287

Singh, R. K., Pretty, J., Pilgrim, S. (2010). Traditional knowledge and biocultural diversity: learning from tribal communities for sustainable development in northeast India. *Journal of Environmental Planning and Management, 53*(4), 511-533.

Sinha, R. K., Dubey, M., Tripathi, R. D., et al. (2010). India as a megadiversity nation. *Int. Society of Environ. Botanists, 16*. https://isebindia.com/09-12/10-11-3.html

Siwakoti, M. (2006). An overview of floral diversity in wetlands of Terai region of Nepal. *Our Nature, 4*(1), 83-90.

Skaptsov, M. V., Lomonosova, M. N., Kutsev, M. G., et al. (2017). The phenomenon of endopolyploidy in some species of the Chenopodioideae (Amaranthaceae). *Botany Letters, 164*(1), 47-53.

Sliwinska, E. (2018). Flow cytometry-a modern method for exploring genome size and nuclear DNA synthesis in horticultural and medicinal plant species. *Folia Horticulturae, 30*(1), 103.

Sliwinska, E., Loureiro, J., Leitch, I. J., et al. (2022). Application-based guidelines for best practices in plant flow cytometry. *Cytometry Part A*, *101*(9), 749-781.

Sliwinska, E., Zielinska, E., Jedrzejczyk, I. (2005). Are seeds suitable for flow cytometric estimation of plant genome size?. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, *64*(2), 72-79.

Sobral, B. W., Honeycutt, R. J., Atherly, A. G., McClelland, M. (1990). Analysis of rice (*Oryza sativa* L.) genome using pulsed-field gel electrophoresis and rare-cutting restriction endonucleases. *Plant Molecular Biology Reporter*, *8*, 253-275.

Somaratne, Y., Guan, D. L., Wang, W. Q., Zhao, L., Xu, S. Q. (2019). The complete chloroplast genomes of two *Lespedeza* species: insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants, 9*(1), 51.

Song, F., Li, T., Burgess, K. S., Feng, Y., Ge, X. J. (2020). Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L.(Polygonaceae) in China. *BMC Plant Biology, 20*(1), 1-15.

Song, W., Chen, Z., Shi, W., et al. (2022). Comparative analysis of complete chloroplast genomes of nine species of *Litsea* (Lauraceae): Hypervariable regions, positive selection, and phylogenetic relationships. *Genes, 13*(9), 1550.

Souza, U. J. B. D., Vitorino, L. C., Bessa, L. A., Silva, F. G. (2020). The complete plastid genome of *Artocarpus camansi*: a high degree of conservation of the plastome structure in the family Moraceae. *Forests*, *11*(11), 1179.

Spiers, J. A., Oatham, M. P., Rostant, L. V., Farrell, A. D. (2018). Applying species distribution modelling to improving conservation based decisions: a gap analysis of Trinidad and Tobago's endemic vascular plants. *Biodiversity and Conservation, 27*, 2931-2949.

Srivastava, V., Lafond, V., Griess, V. C. (2019). Species distribution models (SDM): applications, benefits and challenges in invasive species management. *CABI Reviews*, 1-13.

Stecher, G., Tamura, K., Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution, 37*(4), 1237-1239.

Strasberg, D. (1996). Diversity, size composition and spatial aggregation among trees on a 1-ha rain forest plot at La Réunion. *Biodiversity Conservation, 5*, 825-840.

Straub, S. C., Parks, M., Weitemier, K., et al. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany, 99*(2), 349-364.

Suda, J., Leitch, I. J. (2010). The quest for suitable reference standards in genome size research. *Cytometry part A, 77*(8), 717-720.

Suda, J., Krahulcová, A., Trávníček, P., et al. (2007). Genome size variation and species relationships in *Hieracium* sub-genus Pilosella (Asteraceae) as inferred by flow cytometry. *Annals of Botany, 100*(6), 1323-1335.

Suda, J., Meyerson, L. A., Leitch, I. J., Pyšek, P. (2015). The hidden side of plant invasions: the role of genome size. *New Phytologist*, *205*(3), 994-1007.

Sun, C., Li, J., Dai, X., Chen, Y. (2018). Analysis and characterization of the *Salix suchowensis* chloroplast genome. *Journal of Forestry Research*, *29*(4), 1003-1011.

Sussman, R. W., Rakotozafy, A. (1994). Plant diversity and structural analysis of a tropical dry forest in southwestern Madagascar. *Biotropica*, 241-254.

Sutter, L., Jeanneret, P., Bartual, A. M., et al. (2017). Enhancing plant diversity in agricultural landscapes promotes both rare bees and dominant crop-pollinating bees through complementary increase in key floral resources. *Journal of Applied Ecology, 54*(6), 1856-1864.

Syme, A. E., McLay, T. G., Udovicic, F., et al. (2021). Long-read assemblies reveal structural diversity in genomes of organelles–an example with Acacia pycnantha. *Gigabyte*.

Takamatsu, T., Baslam, M., Inomata, T., et al. (2018). Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and de novo assembly. *Frontiers in Plant Science, 9*, 266.

Talukdar, R., Wary, S., Mili, C., Roy, S., Tayung, K. (2020). Antimicrobial secondary metabolites obtained from endophytic fungi inhabiting healthy leaf tissues of *Houttuynia cordata* Thunb., an ethnomedicinal plant of Northeast India. *J Appl Pharm Sci, 10*(09):099-106.

Tang, D., Wei, F., Cai, Z., et al. (2021). Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. *Development Genes and Evolution*, *231*, 1-9.

Tang, J., Xia, H. A., Cao, M., et al. (2004). A comparison of rice chloroplast genomes. *Plant Physiology*, *135*(1), 412-420.

Taram, M., Borah, D., Tanaka, N. (2020). *Peliosanthes ligniradicis*, a new species (Asparagaceae) from Arunachal Pradesh, NE India. *Phytotaxa, 438*(1), 43-48.

Temsch, E. M., Koutecký, P., Urfus, T., Šmarda, P., Doležel, J. (2022). Reference standards for flow cytometric estimation of absolute nuclear DNA content in plants. *Cytometry Part A, 101*(9), 710-724.

Thornthwaite, J. T., Sugarbaker, E. V., Temple, W. J. (1980). Preparation of tissues for DNA flow cytometric analysis. *Cytometry: The Journal of the International Society for Analytical Cytology, 1*(3), 229-237.

Tian, S., Lu, P., Zhang, Z., Wu, J. Q., Zhang, H., Shen, H. (2021). Chloroplast genome sequence of Chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume chloroplast genomes. *BMC Genomics, 22*(1), 1-14.

Tian, Y., Wu, H., Zhang, G., Wang, L., Zheng, D., Li, S. (2020). Perceptions of ecosystem services, disservices and willingness-to-pay for urban green space conservation. *Journal of Environmental Management, 260*, 110140.

Tiernan, N., Nakamura, K., Burns, C., et al. (2023). Phylogenetic relationships of Cuban and Caribbean *Plumeria* (Apocynaceae) based on the plastid genome. *Biological Journal of the Linnean Society, blad042*.

Tillich, M., Lehwark, P., Pellizzer, T., et al. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Research, 45*(W1), W6-W11.

Trávníček, P., Ponert, J., Urfus, T., et al. (2015). Challenges of flow-cytometric estimation of nuclear genome size in orchids, a plant group with both whole-genome and progressively partial endoreplication. *Cytometry Part A, 87*(10), 958-966.

Tripathi, S. K., Roy, A., Kushwaha, D., et al. (2016). Perspectives of forest biodiversity conservation in Northeast India. *J. of Biodivers., Bioprospecting and Development 3*(02), 1-9. https://doi.org/10.4172/2367-0214.1000157.

Twyford, A. D., Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources, 17*(5), 858-868.

Tyagi, S., Jung, J. A., Kim, J. S., Won, S. Y. (2020). A comparative analysis of the complete chloroplast genomes of three *Chrysanthemum boreale* strains. *PeerJ, 8,* e9448.

Tynsong, H., Dkhar, M., Tiwari, B. K. (2020). Review: traditional ecological knowledge of tribal communities of North East India. *Biodiversitas, 21*, 3209-3224.

Valles, J., Canela, M. Á., Garcia, S., et al. (2013). Genome size variation and evolution in the family Asteraceae. *Caryologia: International Journal of Cytology, Cytosystematics and Cytogenetics*, *66*(3), 221-235.

Varshney, R. K., Dubey, A. (2009). Novel genomic tools and modern genetic and breeding approaches for crop improvement. *Journal of Plant Biochemistry and Biotechnology, 18*, 127-138.

Vera-Paz, S. I., Díaz Contreras Díaz, D. D., Jost, M., et al. (2022). New plastome structural rearrangements discovered in core Tillandsioideae (Bromeliaceae) support recently adopted taxonomy. *Frontiers in Plant Science, 13*, 924922.

Vilhar, B., Greilhuber, J., Koce, J. D., Temsch, E. M., Dermastia, M. (2001). Plant genome size measurement with DNA image cytometry. *Annals of Botany*, *87*(6), 719-728.

Vitales, D., Fernández, P., Garnatje, T., Garcia, S. (2019). Progress in the study of genome size evolution in Asteraceae: analysis of the last update. Database, 2019, baz098.

Vrána, J., Kopecký, D., Kubaláková, M., et al. (2021). Single-cell genomics reveals hidden genome variations in plants. *Nature, 590*(7844), 427-432.

Vrána, J., Cápal, P., Bednářová, M., Doležel, J. (2014). Flow cytometry in plant research: a success story. *Applied Plant Cell Biology: Cellular tools and approaches for plant biotechnology*, 395-430.

Wakamiya, I., Messina, M. G., Newton, R. J., Price, H. J. (1973). Genome size, tracheid volume, and environmental factors in the genus *Pinus*. In *Proceedings*, *22*, 169. Louisiana State University, Division of Continuing Education.

Walter, J. A., Shoemaker, L. G., Lany, N. K., et al. (2021). The spatial synchrony of species richness and its relationship to ecosystem stability. *Ecology, 102*(11), e03486.

Wang, G., Kong, F., Zhang, S., et al. (2015). A tomato chloroplast-targeted DnaJ protein protects Rubisco activity under heat stress. *Journal of Experimental Botany, 66*(11), 3027-3040.

Wang, J. H., Moore, M. J., Wang, H., Zhu, Z. X., Wang, H. F. (2021). Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. *Gene, 765*, 145103.

Wang, J., Liu, J., Kang, M. (2015). Quantitative testing of the methodology for genome size estimation in plants using flow cytometry: a case study of the *Primulina* genus. *Frontiers in Plant Science*, *6*, 354.

Wang, L., Roossinck, M. J. (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Molecular Biology, 61*, 699-710.

Wang, R. J., Cheng, C. L., Chang, C. C., et al. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC evolutionary biology*, *8*(1), 1-14.

Wang, R., Wang, X., Liu, Z., Han, Y., Xie, Q. (2023). Comparative chloroplast genome analysis of four *Hippophae rhamnoides* subspecies and its phylogenetic analysis. *Genetic Resources and Crop Evolution*, 1-15.

Wang, W., Lanfear, R. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biology and Evolution, 11*(12), 3372-3381.

Wang, Y., Zhang, C.F., Odago, W.O., Jiang, H., Yang, J.X., Hu, G.W., Wang, Q.F. (2023). Evolution of 101 Apocynaceae plastomes and phylogenetic implications. *Mol Phylogen Evol 180*, 107688

Wang, Z., Cai, Q., Wang, Y., et al. (2022). Comparative analysis of codon bias in the chloroplast genomes of Theaceae species. *Frontiers in Genetics, 13*, 824610.

Wang, Z., Xu, B., Li, B., Zhou, Q., Wang, G., Jiang, X., ... Xu, Z. (2020). Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *PeerJ, 8,* e8251.

Wanichthanarak, K., Nookaew, I., Pasookhush, P., et al. (2023). Revisiting chloroplast genomic landscape and annotation towards comparative chloroplast genomes of Rhamnaceae. *BMC Plant Biology*, *23*(1), 59.

Waquar, A. A., Ram, K., Mohammad, A. A., Lee, J. (2023). Genome size estimation of false daisy, cheek weed, pot marigold and marigold. *Bioinformation*, *19*(9), 976.

Waters, M. T., Langdale, J. A. (2009). The making of a chloroplast. *The EMBO journal, 28*(19), 2861-2873.

Weiss-Schneeweiss, H., Schneeweiss, G. M. (2012). Karyotype diversity and evolutionary trends in angiosperms. In *Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes*, 209-230. Vienna: Springer Vienna.

Wen, J., Lu, L. M., Nie, Z. L. et al. (2018). A new phylogenetic tribal classification of the grape family (Vitaceae). *J Syst Evol, 56*:262–272. doi:10.1111/jse.12427

Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics, 18*, 1-12.

Wilson, J. D., Morris, A. J., Arroyo, B. E., et al. (1999). A review of the abundance and diversity of invertebrate and plant foods of granivorous birds in northern Europe in relation to agricultural change. *Agriculture Ecosystems Environment, 75*(1-2), 13-30.

Wu, H. Y., Wong, K. H., Kong, B. L. H., et al. (2022). Comparative analysis of chloroplast genomes of *Dalbergia* species for identification and phylogenetic analysis. *Plants, 11*(9), 1109.

Wu, S., Chen, J., Li, Y., et al. (2021). Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. *BMC Plant Biology, 21*, 1-16.

Wu, Y., Liu, F., Yang, D. G., et al. (2018). Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Frontiers in Plant Science, 9*, 376.

Wu, Z., Liao, R., Yang, T., et al. (2020). Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics, 21*, 1-14.

Wyman, S. K., Jansen, R. K., Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics, 20*(17), 3252-3255.

Wysocki, W. P., Clark, L. G., Kelchner, S. A., et al. (2014). A multi–step comparison of short–read full plastome sequence assembly methods in grasses. *Taxon, 63*(4), 899-910.

Xia, X., Peng, J., Yang, L., Zhao, X., Duan, A., Wang, D. (2022). Comparative analysis of the complete chloroplast genomes of eight *Ficus* species and insights into the phylogenetic relationships of *Ficus*. *Life*, *12*(6), 848.

Xiang, K. L., Mao, W., Peng, H. W., et al. (2022). Organization, phylogenetic marker exploitation, and gene evolution in the plastome of *Thalictrum* (Ranunculaceae). *Frontiers in Plant Science, 13*, 897843.

Xiao-Ming, Z., Junrui, W., Li, F., et al. (2017). Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Scientific Reports*, *7*(1), 1555.

Xiao Tao, L., Jiang Xia, Y., Jian Wei, T. (2011). Diversity and composition of understory vegetation in the tropical seasonal rain forest of Xishuangbanna, SW China. *Revista de Biología Tropical, 59*(1), 455-463.

Xiao, Y., Li, X., Cao, Y., Dong, M. (2016). The diverse effects of habitat fragmentation on plant–pollinator interactions. *Plant Ecology, 217*, 857-868.

Xiong, Q., Hu, Y., Lv, W., Wang, Q., Liu, G., Hu, Z. (2021). Chloroplast genomes of five *Oedogonium* species: genome structure, phylogenetic analysis and adaptive evolution. *BMC genomics*, *22*(1), 1-14.

Xiong, Y., Xiong, Y., He, J., et al. (2020). The complete chloroplast genome of two important annual clover species, *Trifolium alexandrinum* and *T. resupinatum*: genome structure, comparative analyses and phylogenetic relationships with relatives in Leguminosae. *Plants*, *9*(4), 478.

Xu, C., Cai, X., Chen, Q., et al. (2011). Factors affecting synonymous codon usage bias in chloroplast genome of Oncidium gower ramsey. *Evolutionary Bioinformatics, 7*, EBO-S8092.

Xu, K., Lin, C., Lee, S. Y., Mao, L., Meng, K. (2022). Comparative analysis of complete *Ilex* (Aquifoliaceae) chloroplast genomes: insights into evolutionary dynamics and phylogenetic relationships. *BMC Genomics, 23*(1), 203.

Xu, X., Shen, Y., Zhang, Y., et al. (2023). A comparison of 25 complete chloroplast genomes between sister mangrove species *Kandelia obovata* and *Kandelia candel* geographically separated by the South China Sea. *Frontiers in Plant Science, 13*, 1075353.

Yang, J., Chu, Q., Meng, G., Kong, W. (2022). The complete chloroplast genome sequences of three *Broussonetia* species and comparative analysis within the Moraceae. *PeerJ*, *10*, e14293.

Yang, L., Wang, Z., Hua, J. (2019). Measuring cell ploidy level in *Arabidopsis thaliana* by flow cytometry. *Plant Innate Immunity: Methods and Protocols*, 101-106.

Yang, S., Li, G., Li, H. (2023). Molecular characterizations of genes in chloroplast genomes of the genus *Arachis* L.(Fabaceae) based on the codon usage divergence. *Plos One, 18*(3), e0281843.

Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in plant science*, *7*, 959.

Yao, X., Tang, P., Li, Z., Li, D., Liu, Y., Huang, H. (2015). The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PloS One, 10*(6), e0129347.

Yarra, R. (2020). Plastome engineering in vegetable crops: current status and future prospects. *Molecular Biology Reports, 47*(10), 8061-8074.

Yengkhom, S., Uddin, A., Chakraborty, S. (2019). Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. assamica and *Camellia sinensis* var. sinensis in comparison to *Camellia pubicosta*. *Journal of Integrative Agriculture, 18*(12), 2771-2785.

Yue, F., Cui, L., Depamphilis, C. W., Moret, B. M., Tang, J. (2008). Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. *BMC Genomics, 9*(1), 1-9.

Zhai, W., Duan, X., Zhang, R., et al. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Molecular Phylogenetics and Evolution*, *135*, 12-21.

Zhang, Y., Ma, L., Guo, Y., et al. (2021). High-throughput flow cytometry analysis of plant genome size using propidium iodide staining. *Methods in Ecology and Evolution, 12*(4), 621-629.

Zhang, P., Xu, W., Lu, X., Wang, L. (2021). Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species. *Physiology and Molecular Biology of Plants*, 1-11.

Zhang, Y., Du, L., Liu, A., et al. (2016). The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Frontiers in plant science*, *7*, 306.

Zhang, Y., Tian, L., Lu, C. (2023). Chloroplast gene expression: recent advances and perspectives. *Plant Communications*.

Zhang, Y., Zhang, A., Li, X., Lu, C. (2020). The role of chloroplast gene expression in plant responses to environmental stress. *International Journal of Molecular Sciences, 21*(17), 6082.

Zhang, Z. (2022). KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics, Proteomics and Bioinformatics, 20*(3), 536-540.

Zhang, Z. C., Hu, S. H., Peng, Y. Q., et al. (2019). The complete chloroplast genome of Mexican marigold (*Tagetes erecta* L., Asteraceae). *Mitochondrial DNA Part B, 4*(2), 3587-3588.

Zhang, Z., Zhang, D. S. (2022). The complete chloroplast genome sequence of *Ficus sarmentosa* (Moraceae, Rosales), a widely distributed fig tree in East Asia. *Mitochondrial DNA Part B*, *7*(9), 1597-1598.

Zhao, F., Chen, Y. P., Salmaki, Y., et al. (2021). An updated tribal classification of Lamiaceae based on plastome phylogenomics. *BMC Biology, 19*, 1-27.

Zhao, K., Li, L., Quan, H., et al. (2021). Comparative analyses of chloroplast genomes from 14 *Zanthoxylum* species: identification of variable DNA markers and phylogenetic relationships within the genus. *Frontiers in Plant Science, 11*, 605793.

Zhao, P., Zhou, H. J., Potter, D., et al. (2018). Population genetics, phylogenomics and hybrid speciation of Juglans in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Molecular Phylogenetics and Evolution*, *126*, 250-265.

Zhao, Y., Yin, J., Guo, H., et al. (2015). The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Frontiers in Plant Science, 5*, 696.

Zheng, G., Wei, L., Ma, L., Wu, Z., Gu, C., Chen, K. (2020). Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships. *Plant Molecular Biology, 102*, 659-676.

Zhou, P., Zhang, Q., Li, J., Li, F., Huang, J., Zhang, M. (2023). A first insight into the genomic background of *Ilex pubescens* (Aquifoliaceae) by flow cytometry and genome survey sequencing. *BMC Genomics, 24*(1), 1-11.

Zhou, W., Armijos, C. E., Lee, C., et al. (2023). Plastid genome assembly using long-read data. *Molecular Ecology Resources*.

Zhu, B., Qian, F., Hou, Y., Yang, W., Cai, M., Wu, X. (2021). Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PLoS One*, *16*(3), e0248556.

Zonneveld, B. J. M. (2021). Selected perennial plants do provide convenient standards for the determination of genome sizes with flow cytometry. *Plant Systematics and Evolution, 307*(2), 28.

Zonneveld, B. J. M., Leitch, I. J., Bennett, M. D. (2005). First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany, 96*(2), 229-244.

LIST OF PUBLICATIONS

Published

1. **Senapati A**, Gujre N, Mitra S, Rangan L (2019). Flora of IIT Guwahati, Vol: 1. IIT Guwahati, pp212
2. **Senapati A,** Basak S, Rangan L (2022). A Review on Application of DNA Barcoding Technology for Rapid Molecular Diagnostics of Adulterants in Herbal Medicine. ***Drug Safety***. 45(3):193-213. doi: 10.1007/s40264-021-01133-4
3. **Senapati A,** Chetri BK, Mitra S, Shelke RG, Rangan L (2023). Decoding the complete chloroplast genome of *Cissus quadrangularis*: insights into molecular structure, comparative genome analysis and mining of mutational hotspot regions. ***Physiology & Molecular Biology of Plants*** 30(4), 470-486. https://doi.org/10.1007/s12298-023-01312-w
4. Rangan L, Chetri BK, **Senapati A**, Basak S, Shelke RG (2023). Plant DNA Analysis: Estimation of Nuclear DNA Content in Plant Homogenates. Book Chapter in Springer. (Accepted)
5. Chetri BK, **Senapati A**, et al. (2024). Untangling the evolutionary tapestry: Plastome insights unveiled from Sieruela rutidosperma to Cleomaceae. Biology Bulletin. (Accepted)

Submitted

1. **Senapati A**, Chetri BK, Shelke RG, Rangan L (2023). Phylogenetic Insights through Morphological, Taxonomical, and the Complete Plastome Genome Analysis of *Thevetia peruviana* (Pers.) K.Schum. Journal of Plant Research. (Under review)
2. Chetri BK, **Senapati A**, Shelke RG, Rangan L (2023). Stitching plastome evolution: Genome assembly insights from *Sieruela rutidosperma* to Cleomaceae. Functional & Integrative Genomics**.** (Under review)

Conferences

1. **Senapati A**, Boro H, Rangan L (2019). Genome size estimation of herbaceous plants of Assam. Research conclave, IIT Guwahati, Assam-781039
2. **Senapati A**, Kharchandy VF, Rangan L (2021). Genome size estimation of woody tree species from Noth-Guwahati: A flow-cytometric approach. International conference on biotechnology for sustainable agriculture, environment and health (BSAEH-2021). Jaipur, India
3. **Senapati A**, Rangan L (2022). Floral diversity of North Guwahati region with special reference to traditionally used medicinal plants. NERC-2022

Seminars and Workshops

1. Participated in a 2 days' workshop titled "Online workshop on flow cytometry techniques and applications" organized by NECBH: 21-22 Dec, 2020.
2. Successfully participated in 22nd Indo-US flow cytometry workshop on "Flow cytometry and its applications in biological research and clinical diagnostics" organized by TETC 22nd-28th Feb, 2021
3. Participated in the workshop on "Python training for scientific computing and data sciences". SAB, IIT Guwahati, April 9-24, 2022
4. A part of the International seminar on medicinal plant production and advancement in herbal medicine, nutraceuticals, cosmetics and other herbal product formulation technology (MedProTech). School of Agro and Rural Technology, IIT Guwahati, October 28-29, 2022

GeneBank (NCBI) Submissions

1. *Cissus quadrangularis* Accession No. **OP414588**
2. *Ficus hispida* Accession No. **OQ220336**
3. *Streblus asper* Accession No. **OQ220335**
4. *Sierrula rutidosperma* Accession No. **OQ281590**
5. *Thevetia peruviana* Accession No. **OQ376289**

Drug Safety
https://doi.org/10.1007/s40264-021-01133-4

REVIEW ARTICLE

A Review on Application of DNA Barcoding Technology for Rapid Molecular Diagnostics of Adulterants in Herbal Medicine

Alok Senapati[1] · **Supriyo Basak**[2] · **Latha Rangan**[1]

Accepted: 21 October 2021

Abstract

The rapid molecular diagnostics of adulterants in herbal medicine using DNA barcoding forms the core of this meticulously detailed review, based on two decades of data. With 80% of the world's population using some form of herbal medicine, authentication, quality control, and detection of adulterants warrant DNA barcoding. A combined group of keywords were used for literature review using the PubMed, the ISI Web of Knowledge, Web of Science (WoS), and Google Scholar databases. All the papers ($N = 210$) returned by the search engines were downloaded and systematically analyzed. Detailed analysis of conventional DNA barcodes were based on retrieved sequences for internal transcribed spacer (ITS) (412,189), *rbc*L (251,598), *mat*K (210,835), and *trn*H-*psb*A (141,846). The utility of databases such as The Barcode of Life Data System (BOLD), NCBI, GenBank, and Medicinal Materials DNA Barcode Database (MMDBD) has been critically examined for the identification of unknown species from known databases. The current review gives an overview of the ratio of adulterated to authentic drugs for some countries along with the state of the art technology currently being used in the identification of adulterated medicines. In this review, efforts were made to systematically analyze and arrange the research and reviews on the basis of technical progress. The review concludes with the future of DNA-based herbal medicine adulteration detection, forecasting the reliance on the metabarcoding technology.

Key Points

This study warrants against the indiscriminate usage of herbal medicine with no prior DNA authentication.

DNA authentication comparing different countries and continents showed that Bhutan had the lowest proportion of adulterated herbal drugs, whereas Australia had the highest proportion.

The DNA barcoding technology fails in extremely processed plant materials and the best solution is to reduce the product length.

1 Introduction

Plants are vital to human health and well being through their roles in medicine, nutrition, and the ecosystem. Herbal medicine sustained human civilization until the 19th century, when the modern synthetic pharmaceutical industry created a whole new molecule-based prophylactic world. Yet, herbal medicines continue to underpin important and alternative medicine systems that are widely subscribed the world over. Herbal medicines lose their competitive edge due to the lack of authentication and quality control of herbal products at the industrial level, with a well-established protocol of analytical methods combined with DNA barcoding. Equally relevant is the application of DNA barcoding technology for rapid molecular diagnostics of adulterants in herbal medicine. This review critically brings to the fore the latter aspect, meticulously detailed and systematically developed. Therefore, using DNA barcoding as a base, we examine its methods, current statistics, success rates, limitations, and the latest technological developments. Given the wide divergence in terminology and systems of alternative medicine that use herbal medications, a short survey at this point would bring in a certain degree of clarity.

Alok Senapati, Supriyo Basak have equal contribution.

✉ Latha Rangan
lrangan@iitg.ac.in; latha_rangan@yahoo.com

[1] Applied Biodiversity Laboratory, O Block, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

[2] Department of Bioscience and Biotechnology, Banasthali Vidyapith, Tonk, Rajasthan 304 022, India

Adis

Physiol Mol Biol Plants
https://doi.org/10.1007/s12298-023-01312-w

RESEARCH ARTICLE

Decoding the complete chloroplast genome of *Cissus quadrangularis*: insights into molecular structure, comparative genome analysis and mining of mutational hotspot regions

Alok Senapati[1] · Bimal K. Chetri[2] · Sudip Mitra[2] · Rahul G. Shelke[1] · Latha Rangan[1]

Received: 14 December 2022 / Revised: 25 March 2023 / Accepted: 24 April 2023

Abstract *Cissus quadrangularis* L., a member of the Vitaceae family, is an important medicinal plant with widespread application in Indian traditional medicines. *C. quadrangularis* L. whole chloroplast genome of 160,404 bp was assembled using a genome skimming approach from the whole genome library. The assembled chloroplast genome contained a large single-copy region (88,987 bp), a small single-copy region (18,621 bp), and pairs of inverted repeat regions (26,398 bp). It also comprised 133 genes, including 37 tRNAs, eight rRNAs, and 88 protein-coding genes. Aside from that, we annotated three genes *atp*H, *pet*B, and *psb*L, as well as one duplicated copy of the *ycf*1 gene in *C. quadrangularis* L. that had previously been missing from the annotation of compared *Cissus* chloroplast genomes. Five divergent hotspot regions such as *pet*A_*psb*J (0.1237), *rps*16_*trn*Q-UUG (0.0913), *psb*C_*trn*S-UGA (0.0847), *rps*15_*ycf*1 (0.0788), and *rps*2_*rpo*C2 (0.0788) were identified in the investigation that could aid in future species discrimination. Surprisingly, we found the overlapping genes *ycf*1 and *ndh*F on the IRb/SSC junction, rarely seen in angiosperms. The results of the phylogenetic study showed that the genomes of the *Cissus* species under study formed a single distinct clade. The detailed annotations given in this study could be useful in the future for genome annotations of *Cissus* species. The current findings of the study have the potential to serve as a useful resource for future research in the field of population genetics and the evolutionary relationships in the *Cissus* genus.

Keywords *Cissus quadrangularis* L. · Inverted repeats · Phylogeny · Simple sequence repeats · Tandem repeats

Introduction

Cissus quadrangularis L. is a perennial evergreen fleshy climber that belongs to the family Vitaceae. In common parlance, it is referred to as veldt grape, adamant creeper, and hadjod etc. The genus *Cissus*, formerly placed in the family Ampelidaceae, is thought to contain around 350 species that can be found in various habitats around the world (Onyeweaku et al. 2020). The family Vitaceae is comprised of 14 different genera, with *Cissus* being the most abundant genus (Onyeweaku et al. 2020). Members of this genus are found in South-East Asia, the Arabian Peninsula, tropical and subtropical Africa, and the Indian subcontinent. The plant grows in plain coastal regions, woodlands, and wastelands up to 500 m in elevation. It generates adventitious roots at nodes, which makes the process of vegetative propagation quite simple.

Since antiquity, people have used *C. quadrangularis* L. for therapeutic purposes in a number of traditional Ayurvedic medications. Its name, asthisamharaka, translates to "that which avoids the deterioration of bones" since it is supposed to help cure shattered bones and is used as a tonic and painkiller in Siddha medicine (Sivarajan and Balachandran 1994). Some compounds in *Cissus* act as glucocorticoid

---

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-023-01312-w.

✉ Rahul G. Shelke
rahul.sg98@gmail.com

✉ Latha Rangan
latha_rangan@yahoo.com; lrangan@iitg.ac.in

[1] Applied Biodiversity Laboratory, O Block, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

[2] School of Agro and Rural Technology, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

FLORA
of
IIT Guwahati

Alok Senapati | Nihal Gujre
Sudip Mitra | Latha Rangan

Vol-I

FLORA of IIT Guwahati

Alok Senapati | Nihal Gujre
Sudip Mitra | Latha Rangan

Its gives us immense pleasure in presenting the floral wealth of our own campus (IIT Guwahati) and witnessing our dream becoming a reality. In a time when climate change is no more a bad dream but is a harsh reality, presence of plants is absolutely essential for all living organisms to live sustainably. Exploring and documenting the floral wealth of IIT Guwahati campus neither can be a onetime exercise nor it can be documented in one volume of a book such this one. We shall continue our effort and explore more and try our best to bring the hidden floral wealth of this campus.

- Authors

ISBN

Price : ₹ 400/-

Annexure I

Dicotyledons

Acanthaceae

Hygrophila polysperma (Rox.) Andn.

Justicia adhatoda L.

Justicia gendarussa (L)Burm.

Phlogacanthus thyrsiflorus (Roxb.) Nees

Thunbergia grandiflora (Roxb. ex Rottl.) Roxb.

Aizoaceae

Coldenia procumbens L.

Amaranthaceae

Achyranthes aspera L.

Alternanthera philoxeroides (Mart.) Griseb.

Alternanthera sessilis (L.) Roxb.ex. DC.

Alternanthera triandra Lam.

Amaranthus blitum L.

Amaranthus spinosus L.

Amaranthus viridis L.

Chenopodium ficifolium Sm.

Anacardiaceae

Anacardium occidentale L.

Mangifera indica L.

Semecarpus anacardium L.f.

Spondias pinnata (L.f.)Kurz

Annonaceae

Polyalthia longifolia Sonn.

Aparagaceae

Yucca gigantea Lem.

Apiaceae

Centella asiatica (L.) Urban

Hydrocotyle javanica Thunb.

Apocynaceae

Allamanda cathartica L.

Alstonia scholaris (L.) R.Br.

Calotropis gigantea (L.) Dryand.

Catharanthus roseus (L.) G.Don

Cascabela thevetia (L.) Lippold

Holarrhena antidysenterica (L.) Wall. ex A. DC

Ichnocarpus frutescens (L.) W.T.Aiton

Nerium oleander L.

Pachypodium lamerei Drake

Plumeria rubra L.

Rauvolfia canescens L.

Tabernaemontana divaricata (L.) R.Br. ex Roem. & Schult.

Wrightia tomentosa (Roxb.) Roem. & Schult.

Araliaceae

Hydrocotyle sibthorpioides L.

Ascelpediaceae

Asclepias curassavica L.

Calotropis procera (Ait.) Roxb.

Cryptolepis sinensis (Lour.) Merr.

Asteracaeae

Acmella repens (Walter) Rich.

Ageratum conyzoides (L.) L.

Artemisia dracunculus L.

Bidens pilosa L.

Blumea lacera (Burm.f.) DC.

Blumea sp.

Callistephus chinensis (L.) Nees

Centipeda minima (L.) A.Br. & Aschers

Chrysanthemum indicum L.

Cyanthillium cinereum (L.) H.Roxb.

Dahlia pinnata Cav.

Dichrocephala integrifolia (L.f.) O.Kuntze

Eclipta prostrata (L.) L.

Elephantopus scaber L.

Enhydra fluctuans Lour.

Erigeron Canadensis L.

Eupatorium cannabinum L.

Eupatorium odoratum L.

Eupatorium perfoliatum L.

Gnaphalium luteo-album L.

Gnaphalium uliginosum L.

Helianthus annuus L.

Lactuca sativa L.

Leontopodium alpinum Colm. ex Cass.

Mikania micrantha Kunth.

Parthenium hysterophorus L.

Senecio vulgaris L.

Sonchus arvensis L.

Sonchus wightianus DC.

Spilanthes acmella Murr

Spilanthes paniculata Wallich ex DC.

Synedrella nodiflora (L.) Gaertn.

Tanacetum coccineum (Willd.)Grierson.

Tridax procumbens L.

Vernonia cinerea (L.) Less.

Wedelia calendulacea L.

Xanthium indicum Koenig

Zinnia angustifolia Kunth

Balsaminaceae

Hydrocera triflora (L.) Wight & Arn.

Impatiens balsamina L.

Bignoniaceae

Jacaranda mimosifolia D.Don

Stenolobium stans (L.) Juss. ex Kunth

Boraginaceae

Cynoglossum lanceolatum Forssk.

Heliotropium indicum L.

Brassicaceae

Brassica campestris L.

Cardamine hirsuta L.

Rorippa benghalensis (DC.) Hara

Caesalpiniaceae

Caesalpinia bonduc Flem.

Cassia alata L.

Cassia nodosa Buch.-Ham.

Cassia occidentalis L.

Cassia sophera L.

Cassia tora L.

Bauhinia racemosa Lam.

Calophyllaceae

Mesua ferrea L.

Cannabaceae

Cannabis sativa L.

Capparaceae

Capparis zeylanica L.

Crateva magna (Lour.)DC.

Caricaceae

Carica papaya L.

Caryophyllaceae

Drymaria diandra Blume

Polycarpon prostratum (Forssk.)Ash.

Ceratophylaceae

Ceratophyllum demersum L.

Cleomaceae

Cleome gynandra L.

Cleome rutidosperma DC.

Cleome viscosa L.

Combretaceae

Terminalia arjuna (Roxb. ex DC.) Wight & Arn.

Terminalia chebula Retz.

Convolvulaceae

Argyreia nervosa (Burm.f.)Boj.

Cuscuta campestris Yunck.

Evolvulus nummularius (L.) L.

Ipomoea aquatica Forssk.

Ipomoea carnea Jacq.

Ipomoea nil (L.) Roth.

Merremia vitifolia (Burm.f.) Hall.f.

Crassulaceae

Bryophyllum pinnatum (Lam.) Oken

Cucurbitaceae

Coccinia grandis (L.)Voigt.
Momordica charantia L.var.
Dilleniaceae
Dillenia indica L.
Elatinaceae
Bergia ammannioides Roxb.
Euphorbiaceae
Acalypha hispida Burm.f.
Acalypha indica L.
Acalypha wilkesiana Müll.Arg.
Codiaeum variegatum (L.) A.Juss.
Croton tiglium L.
Croton bonplandianum Baill.
Croton caudatus Geist.
Croton sparsiflorus Morong
Euphorbia hirta L.
Euphorbia milii Des Moul.
Euphorbia neriifolia L.
Euphorbia pulcherrima Willd.
Hevea brasiliensis Müll.Arg.
Jatropha curcas L.
Jatropha gossypifolia L.
Macaranga denticulata (Blume) Müll.Arg.
Mallotus philippensis (Lam.) Muell.
Manihot esculenta Crantz
Phyllanthus emblica L.
Ricinus communis L.
Triadica sebifera (L.) Small
Vernicia fordii (Hemsl.) Airy Shaw
Fabaceae
Acacia auriculiformis A.Cunn. ex Benth.
Aeschynomene aspera L.
Aeschynomene indica L.
Albizia lebbeck (L.) Benth.
Albizia saman (Jacq.) Merr.
Alysicarpus vaginalis (L.) DC.
Bauhinia acuminata L.
Butea monosperma (Lam) Kuntz.
Caesalpinia pulcherrima (L.) Sw.
Calliandra surinamensis Benth.
Cassia fistula L.
Cassia javanica L.
Centrosema pubescens (DC.) Benth.
Clitoria ternatea L.
Crotalaria pallida Aiton
Dalbergia sissoo Roxb.
Delonix regia (Hook.) Raf.
Desmodium gangeticum (L.) DC.
Desmodium heterocarpon (L.)DC.
Desmodium racemosum (Thunb.)DC.
Desmodium triflorum (L.)DC.
Erythrina stricta Roxb.
Indigofera tinctoria L.
Leucaena leucocephala (Lam.) de Wit
Millettia pachycarpa Benth.
Mimosa diplotricha C. Wright ex Sauv.
Mucuna pruriens (L.)DC.
Peltophorum pterocarpum (DC.) K.Heyne.
Pongamia pinnata (L.) Pierre
Pueraria tuberosa (Willd.) DC.
Saraca asoca (Roxb.) Willd.
Senna occidentalis (L.) Link
Sesbania grandiflora (L.) Poiret
Flacourtiaceae
Casearia vareca Roxb.
Flacourtia jangomas (Lour.)Raeush
Heliconiaceae
Heliconia rostrata Ruiz & Pav.
Hypericaceae
Hypericum japonicum Thunb.
Lamiaceae
Clerodendrum thomsoniae Balf.f.
Clerodendrum trichotomum Thunb.
Leonurus sibiricus L.

Leucas aspera (Willd.) Link

Ocimum basilicum L.

Ocimum sanctum L.

Rotheca serrata (L.) Steane & Mabb.

Stachys oblongifolia Benth

Tectona grandis L.f.

Teucrium viscidum Bl.

Lauraceae

Cinnamomum tamala Nees & Eberm

Litsea glutinosa (Lour.) CB Robinson

Lecythidaceae

Barringtonia acutangula (L.)Gaertner

Leeaceae

Leea guineensis G.Don

Loranthaceae

Dendrophthoe falcate (L.f.) Ettingsh

Lythraceae

Cuphea hyssopifolia Kunth

Lagerstroemia indica L.

Lagerstroemia speciosa (L.)Pers.

Magnoliaceae

Magnolia champaca (L.) Baill. ex Pierre

Malvaceae

Abelmoschus esculentus (L.) Moench

Abelmoschus manihot (L.) Medik

Abelmoschus moschatus Medik.

Abroma augustum (L.) L.f.

Abutilon indicum (L.) Sweet

Abutilon theophrasti Medik.

Bombax ceiba L.

Corchorus olitorius L.

Hibiscus rosa-sinensis L.

Malvastrum coromandelianum (L.) Garcke

Sida acuta Burm.f.

Sida cordifolia L.

Sida rhombifolia L.

Sterculia villosa Roxb. ex Sm.

Thespesia lampas (Cav.) Dalzell & A. Gib.

Triumfetta rhomboidea Jacq.

Urena lobata L.

Melastomataceae

Melastoma malabathricum L.

Meliaceae

Azadirachta indica A.Juss.

Dysoxylum binectariferum Hook.f.

Melia azedarach L.

Toona ciliata Roem

Menispermaceae

Cissampelos pareira L.

Stephania japonica (Thunb.)Miers

Tinospora cordifolia (Willd.)Hook.f. & Thomson

Menyanthaceae

Nymphoides indicum (L.) O.Ktze.

Nymphoides hydrophylla (Lour) O.Ktze.

Mimosaceae

Mimosa pudica L.

Molluginaceae

Glinus lotoides L.

Moraceae

Artocarpus heterophyllus Lam.

Broussonetia papyrifera (L.) Vent

Ficus elastica Roxb. ex Hornem.

Ficus hispida L.f.

Ficus infectoria Roxb.

Ficus racemosa L.

Ficus religiosa L.

Streblus asper Lour.

Musacae

Musa paradisiaca L.

Myrtaceae

Callistemon viminalis G.Don

Eucalyptus obliqua L'Hér

Psidium guajava L.

Syzygium cumini (L.) Skeels

Nelumbonaceae

Nelumbo nucifera Gaertn.

Nyctaginaceae

Boerhavia diffusa L.

Bougainvillea glabra Choisy

Nyctanthes arbor- tristis L.

Nymphaeaceae

Euryale ferox Salisb.

Nymphaea pubescens Willd.

Onagraceae

Ludwigia hyssopifolia (G.Don.)Excell

Ludwigia octovalvis (Jacq.) Raven

Oxalidaceae

Averrhoa carambola L.

Oxalis corniculata L.

Papaveraceae

Argemone mexicana L.

Passifloraceae

Passiflora foetida L.

Phyllanthaceae

Antidesma accuminatum Wight

Antidesma bunius (L.) Spreng.

Baccaurea ramiflora Lour.

Bridelia retusa (L.) Spreng.

Phyllanthus acidus Skeels

Phyllanthus amarus Schum.

Phyllanthus pulcher Wall. ex Müll.Arg.

Phyllanthus urinaria L.

Piperaceae

Piper thomsonii Hook.f

Plantaginaceae

Mecardonia procumbens (Mill.) Small

Polygalaceae

Polygala arvensis Willd.

Salomonia cantoniensis Lour.

Polygonaceae

Persicaria barbata (L.) Hara

Persicaria hydropiper (L.) Delarbre

Persicaria lapathifolia (L.) Delarbre

Polygonum plebeium Roxb.

Portulacaceae

Portulaca oleracea L.

Portulaca quadrifida L.

Proteaceae

Grevillea robusta A.Cunn. ex R.Br.

Ranunculaceae

Ranunculus sceleratus L.

Rhamnaceae

Ziziphus jujuba Mill.

Rosaceae

Rosa alba L.

Rubus rugosus Smith

Rubiaceae

Anthocephalus chinensis (Lamk.)A.Rich ex Walp.

Coffea arabica L.

Coffea benghalensis Heyne ex M. Roemer

Dentella repens (L.)JR.Forst & G.Forst

Hedyotis diffusa Willd.

Ixora coccinea L.

Neolamarckia cadamba (Roxb.) Bosser

Paederia scandens (Lour.)Merr.

Rutaceae

Aegle marmelos (L.) Correa

Citrus aurantium L.

Citrus grandis (L.) Osbeck.

Citrus limon (L.) Osbeck

Glycosmis pentaphylla (Retz.) DC.

Micromelum integerrimum (Roxb. ex DC)M.Roemer

Murraya koenigii (L.) Sprengel

Murraya paniculata (L.) Jack

Sapindaceae

Cardiospermum halicacabum L.

Litchi chinensis Sonn.

Sapindus mukorossi Gaertn.

Sapotaceae

Mimusops elengi L.

Scrophulariaceae

Limnophila heterophylla Benth.

Lindernia ciliata (Colsm.) Penn.

Lindernia crustacean (L.) Fl.

Scoparia dulcis L.

Solanaceae

Atropa belladonna L.

Capsicum annuum L.

Capsicum chinense Jacq

Capsicum frutescens (Linn.)

Cestrum nocturnum L.

Datura metel L.

Datura stramonium L.

Nicotiana plumbaginifolia Viv.

Nicotina tobacum L.

Physalis angulata L.

Physalis minima L.

Solanum indicum L.

Solanum khasianum C. B. Cl.

Solanum lycopersicum L.

Solanum melongena L.

Solanum nigrum L.

Solanum rostratum Dunal

Solanum sisymbriifolium Lam.

Solanum surattense Burm.f.

Solanum torvum Sw.

Solanum tuberosum L.

Solanum xanthocarpum Wendl.

Theaceae

Eurya japonica Thunb.

Schima wallichii (DC.) Korth.

Tiliaceae

Corchorus aestuans L.

Grewia villosa Willd.

Trapaceae

Trapa natans L.

Tropaeolaceae

Tropaeolum majus L.

Urticaceae

Boehmeria macrophylla Hornem

Pouzolzia hirta (Bl.) Hassk.

Sarcochlamys pulcherrima Gaud.

Verbenaceae

Callicarpa rubra Vahl.

Clerodendrum glandulosum Lindl.

Clerodendrum viscosum Vent.

Lantana camara L.

Lippia alba (Mill.) ex. Britton & Wilson

Phyla nodiflora (L.) Greene

Stachytarpheta jamaicensis (L.) Vahl

Vitex negundo L.

Vitaceae

Cayratia trifolia (L.) Domin

Cissus quadrangularis Wall.

Monocotyledons

Alismataceae

Sagittaria sagittifolia L.

Amaryllidaceae

Crinum amoenum Roxb.

Araceae

Amorphophallus bulbifer (Roxb.) Bl.

Arum Sp.

Colocasia esculenta (L.) Schott

Lasia spinosa (L.) Thw

Pistia stratiotes L.

Pothos cathcartii Schott

Syngonium podophyllum Schott

Arecaceae

Cocos nucifera L.

Dypsis lutescens Beentje & J.Dransf.

Phoenix dactylifera L.

Wodyetia bifurcata A.K.Irvine

Asparagaceae

Furcraea foetida(L.) Haw.

Cannaceae

Canna indica L.

Commelinaceae

Commelina attenuata Koenig ex. Vahl.

Commelina benghalensis L.

Murdannia nudiflora (L.) Brenan

Tradescantia fluminensis Vell.

Costaceae

Cheilocostus speciosus (J.Konig) C.Specht

Costus igneus Nak

Cyperaceae

Bulbostylis barbata (Rottb.) C.B. Clark

Carex rostrata Stokes

Cyperus cephalotes Vahl.

Cyperus compressus L.

Cyperus difformis L.

Cyperus digitatus Roxb.

Cyperus haspan L.

Cyperus iria L.

Cyperus pilosus Vahl.

Cyperus rotundus L.

Cyperus strigosus L.

Cyperus tenuispica Steud.

Eleocharis congesta Don.

Fimbristylis miliacea (L.) Vahl

Kyllinga brevifolia Rottb.

Pycreus polystachyos (Rottb.) Beauv.

Dioscoreaceae

Dioscorea bulbifera L.

Hydrochaaritaceae

Hydrilla verticillata (L.f.) Royle

Nechamandra alternifolia (Roxb.) Thw.

Lemnaceae

Lemna perpusilla Torr.

Liliaceae

Gloriosa superba L.

Marantaceae

Schumannianthus dichotomus (Roxb.) Gagnep

Najadaceae

Najas graminea Del.

Orchidaceae

Rhynchostylis retusa (L.) Blume

Vanda roxburghii R.Br.

Poaceae

Agrostis capillaris L.

Axonopus compressus (Sw.) P.Beauv.

Bambusa vulgaris Schrad. Ex J.C.Wendl.

Brachiaria mutica (Forssk.) Stapf.

Chrysopogon gryllus (L.) Trin.

Coix lachryma-jobi L.

Cynodon dactylon (L.) Pers.

Dactyloctenium aegyptium (L.) Willd.

Digitaria ciliaris L.

Echinochloa colona (L.) Link

Echinochloa crus-galli (L.)P.Beauv.

Echinochloa cruspavonis (H.B.K.)Schutt.

Echinochloa stagnina (Retz.) P.Beauv.

Eleusine indica (L.) Gaertn.

Eragrostis atrovirens (desf.)Trin

Eragrostis tenella (L.)P.Beauv.

Hygroryza aristata (Retz.) Nees ex Wight.

Hymenachne acutigluma (Steud.) Gill.

Imperata cylindrica (L)P.Beauv.

Panicum repens L.

Paspalidium flavidum (Retz.) A.Camus

Paspalum conjugatum Berg.

Paspalum longifolium Roxb.

Paspalum paspaloides (Much.) Screen.

Pennisetum pedicellatum Trin.

Phragmites karka (Retz.) Trin.

Pseudoraphis spinescens (Roxb.) Vickery
Saccharum arundinaceum Retz.
Sacciolepis interrupta (Willd.) Stapf.
Setaria pumila (Poir.) Roem. & Schult.

Pontedariaceae

Eichhornia crassipes (Mart.) Solms
Monochoria hastata (L.) Solms

Potamogetonacea

Potamogeton nodosus Poir.

Smilacaceae

Smilax anceps Willd.

Zingiberaceae

Alpinia nigra (Gaertn.)Burt.
Amomum aromaticum Roxb.
Costus speciosus (Koen.) Smith
Curcuma amada Roxb.
Curcuma aromatica Salisb.
Curcuma caesia Roxb.
Curcuma longa L.
Curcuma zedoaria (Christm.) Roscoe
Kaempferia rotunda L.
Mollugo pentaphylla L.
Zingiber officinale Roscoe
Zingiber rubens Roxb.
Zingiber zerumbet (L.) Roscoe ex Sm.

Annexure II

(a) GS estimation of 47 herbaceous plants (new estimates are highlighted)

S. No	**Species**	**Family**	**Reference Standard**	**Estimated GS (2C in pg)**	**Reported GS (2C in pg)**
1	*Hygrophila polysperma*	Acanthaceae	*P. pinnata*	4.6 ± 0.014	
2	*Sagittaria sagittifolia*	Alismataceae	*Z. mays*	42.5 ± 0.006	42.5
3	*Amaranthus spinosus*	Amaranthaceae	*R. communis*	1.08 ± 0.015	2
4	*Amaranthus viridis*	Amaranthaceae	*R. communis*	1.41 ± 0.002	
5	*Amaranthus blitum*	Amaranthaceae	*P. pinnata*	1.75 ± 0.006	
6	*Alternanthera sessilis*	Amaranthaceae	*R. communis*	3.29 ± 0.32	
7	*Achrynthes bidenta*	Amaranthaceae	*R. communis*	3.41 ± 0.021	
8	*Achrynthes aspera*	Amaranthaceae	*P. pinnata*	3.52 ± 0.009	
9	*Alternanthera philoxeroides*	Amaranthaceae	*R. communis*	5.76 ± 0.048	
10	*Centella asiatica*	Apiaceae	*R. communis*	1.3 ± 0.003	
11	*Colocasia esculenta*	Araceae	*Z. mays*	8.2 ± 0.012	8.2
12	*Syngonium podophyllum*	Araceae	*Z. mays*	9.7 ± 0.02	9.7
13	*Amorphophallus bulbifer*	Araceae	*Z. mays*	18.6 ± 0.009	18.6
14	*Hydrocotyl sibthorpioides*	Araliaceae	*Z. mays*	2.36 ± 0.02	
15	*Parthenium hysterophorus*	Asteraceae	*Z. mays*	2.53 ± 0.017	
16	*Mikania micrantha*	Asteraceae	*Z. mays*	3.04 ± 0.038	

17	*Acmella oleracia*	Asteraceae	*P. pinnata*	6.66 ± 0.015	
18	*Dahlia pinnata*	Asteraceae	*P. pinnata*	8.62 ± 0.014	8.62
19	*Ageratum conyzoides*	Asteraceae	*Z. mays*	9.50 ± 0.012	4
20	*Tanacetum coccineum*	Asteraceae	*R. communis*	11.6 ± 0.031	
21	*Acmella uliginosa*	Asteraceae	*P. pinnata*	15.72 ± 0.087	
22	*Cannabis indica*	Cannabaceae	*R. communis*	2.05 ± 0.013	
23	*Cleome rutindospermum*	Cleomaceae	*S. lycopersicum*	1 ± 0.003	
24	*Commelina benghalensis*	Commelinaceae	*Z. mays*	15 ± 0.015	15
25	*Merremia vitifolia*	Convolvulaceae	*R. communis*	2.06 ± 0.01	
26	*Cheilocostus speciosus*	Costaceae	*S. lycopersicum*	3.90 ± 0.01	
27	*Bryophyllum pinnatum*	Crassulaceae	*S. lycopersicum*	0.91 ± 0.01	
28	*Kalanchoe pinnata*	Crassulaceae	*S. lycopersicum*	0.99 ± 0.018	
29	*Clitoria ternatea*	Fabaceae	*Z. mays*	11.08 ± 0.032	11.08
30	*Ocimum basilicum*	Lamiaceae	*P. pinnata*	7.44 ± 0.017	7.44
31	*Sida acuta*	Malvaceae	*O. sativa*	1.23 ± 0.02	
32	*Oxalis violacea*	Oxalidaceae	*R communis*	2.8 ± 0.048	
33	*Argemon mexicana*	Papaveraceae	*S. lycopersicum*	1.16 ± 0.044	
34	*Phyllanthus urinaria*	Phyllanthaceae	*P. pinnata*	1 ± 0.024	
35	*Veronica cinerea*	Plantaginaceae	*R. communis*	2.623 ± 0.028	
36	*Pennisetum purpureum*	Poaceae	*P. pinnata*	1.6 ± 0.001	5.8

37	*Imperata cylindrica*	Poaceae	*Z. mays*	10.8 ± 0.009	10.8
38	*Ranunculus sceleratus*	Ranunculaceae	*Z. mays*	8 ± 0.018	8
39	*Salvinia minima*	Salviniaceae	*R. communis*	0.5 ± 0.003	
40	*Cardiospermum halicacabum*	Sapindaceae	*S. lycopersicum*	0.85 ± 0.01	
41	*Scoparia daulcis*	Scrophulariaceae	*P. pinnata*	1.567 ± 0.12	
42	*Solanum xanthocarpum*	Solanaceae	*Z. mays*	5.29 ± 0.01	
43	*Capsicum frutescens*	Solanaceae	*Z. mays*	7.42 ± 0.016	6.8
44	*Nicotiana plumbaginifolia*	Solanaceae	*P. pinnata*	8.2 ± 0.021	8.2
45	*Nicotiana tabacum*	Solanaceae	*P. pinnata*	10.36 ± 0.008	10.36
46	*Typha angustifolia*	Typhaceae	*R. communis*	0.45 ± 0.015	0.46
47	*Curcuma zedoaria*	Zingiberaceae	*P. pinnata*	1.82 ± 0.011	3.32

(b) GS estimation of 12 shrubs (new estimates are highlighted)

S. No	Species	Family	Reference Standard	Estimated GS (2C)	Reported GS (2C)
1	*Yucca gigantea*	Aparagaceae	*P. pinata*	12 ± 0.002	12
2	*Cascabela thevetia*	Apocynaceae	*S. lycopersicum*	0.63 ± 0.004	
3	*Nerium oleander*	Apocynaceae	*S. lycopersicum*	0.765 ± 0.013	
4	*Tabernaemontana divaricata*	Apocynaceae	*V. radiata*	2.32 ± 0.009	
5	*Calotropis procera*	Ascelpediaceae	*V. radiata*	0.43 ± 0.002	
6	*Codiaenum sp.*	Euphorbiaceae	*P. pinata*	15.41 ± 0.021	
7	*Rotheca serrata*	Lamiaceae	*S. lycopersicum*	11.42 ± 0.012	
8	*Gloriosa superba*	Liliaceae	*P. pinata*	8.96 ± 0.013	8.96
9	*Urena lobata*	Malvaceae	*P. pinata*	3.98 ± 0.046	
10	*Bougainvillea glabra*	Nyctaginaceae	*Z. mays*	8.2 ± 0.008	8.2
11	*Ixora coccinea*	Rubiaceae	*O. sativa*	1.93 ± 0.024	
12	*Clerodendrum glandulosum*	Verbenaceae	*P. pinata*	1.68 ± 0.05	1

(c) GS estimation of 51 trees (new estimates are highlighted)

S. No	Species	Family	Reference standard	Estimated GS (2C)	Reported GS (2C)
1	*Anacardium occidentale*	Anacardiaceae	*S. lycopersicum*	0.948 ± 0.006	1.11
2	*Mangifera indica*	Anacardiaceae	*S. lycopersicum*	1.031 ± 0.002	0.91
3	*Polyalthia longifolia*	Annonaceae	*O. sativa*	1.987 ± 0.008	2.24
4	*Alstonia scholaris*	Apocynaceae	*S. lycopersicum*	1.190 ± 0.008	
5	*Plumeria rubra*	Apocynaceae	*V. radiata*	3.321 ± 0.014	
6	*Dypsis lutescens*	Arecaceae	*S. lycopersicum*	3.449 ± 0.048	
7	*Stenolobium stans*	Bignoniaceae	*S. lycopersicum*	0.783 ± 0.007	1.6
8	*Trema orientalis*	Cannabaceae	*S. lycopersicum*	1.116 ± 0.014	
9	*Terminalia arjuna*	Combretaceae	*S. lycopersicum*	7.104 ± 0.010	7.1
10	*Terminalia chebula*	Combretaceae	*S. lycopersicum*	7.4 ± 0.012	
11	*Butea monosperma*	Fabaceae	*O. sativa*	2.54 ± 0.014	1.12
12	*Acacia mangium*	Fabaceae	*P. pinnata*	1.632 ± 0.028	1.3
13	*Adenanthera pavonina*	Fabaceae	*S. lycopersicum*	4.578 ± 0.019	1.36
14	*Acacia auriculiformis*	Fabaceae	*S. lycopersicum*	1.600 ± 0.003	1.37
15	*Cassia fistula*	Fabaceae	*S. lycopersicum*	1.202 ± 0.008	1.4
16	*Senna siamea*	Fabaceae	*O. sativa*	1.246 ± 0.011	1.4
17	*Erythrina stricta*	Fabaceae	*O. sativa*	2.28 ± 0.048	2.2
18	*Leucaena leucocephala*	Fabaceae	*V. radiata*	3.584 ± 0.013	2.6

19	*Dalbergia sissoo*	Fabaceae	*O. sativa*	1.788 ± 0.006	3.24
20	*Cassia javanica*	Fabaceae	*S. lycopersicum*	1.215 ± 0.002	7.52
21	*Albizia lebbeck*	Fabaceae	*S. lycopersicum*	1.745 ± 0.002	
22	*Albizia saman*	Fabaceae	*S. lycopersicum*	1.277 ± 0.030	
23	*Bauhinia racemosa*	Fabaceae	*S. lycopersicum*	0.846 ± 0.008	
24	*Bauhinia variegata*	Fabaceae	*V. radiata*	0.772 ± 0.003	
25	*Calliandra hematocephala*	Fabaceae	*S. lycopersicum*	1.343 ± 0.008	
26	*Delonix regia*	Fabaceae	*S. lycopersicum*	1.518 ± 0.010	
27	*Parkia timoriana*	Fabaceae	*S. lycopersicum*	1.466 ± 0.020	
28	*Cinnamomum tamala*	Lauraceae	*S. lycopersicum*	3.078 ± 0.009	
29	*Lagerstroemia indica*	Lythraceae	*S. lycopersicum*	0.905 ± 0.011	
30	*Lagerstroemia speciosa*	Lythraceae	*S. lycopersicum*	0.874 ± 0.018	
31	*Magnolia champaca*	Magnoliaceae	*S. lycopersicum*	5.362 ± 0.046	
32	*Ficus religiosa*	Moraceae	*S. lycopersicum*	0.984 ± 0.001	1.41
33	*Ficus elastic*	Moraceae	*S. lycopersicum*	1.714 ± 0.017	1.44
34	*Ficus racemosa*	Moraceae	*O. sativa*	0.875 ± 0.002	1.44
35	*Ficus benghalensis*	Moraceae	*S. lycopersicum*	1.132 ± 0.002	1.45
36	*Artocarpus heterophyllus*	Moraceae	*S. lycopersicum*	3.788 ± 0.045	2.3
37	*Ficus hispida*	Moraceae	*S. lycopersicum*	0.903 ± 0.003	
38	*Streblus asper*	Moraceae	*S. lycopersicum*	0.970 ± 0.004	

39	*Moringa oleifera*	Moringaceae	*O. sativa*	0.676 ± 0.008	1.22
40	*Psidium guajava*	Myrtaceae	*O. sativa*	1.135 ± 0.011	0.56
41	*Callistemon viminalis*	Myrtaceae	*S. lycopersicum*	1.552 ± 0.017	
42	*Syzyzium jambos*	Myrtaceae	*O. sativa*	2.297 ± 0.012	
43	*Nyctanthes arbor-tristis*	Oleaceae	*P. pinnata*	1.984 ± 0.014	2.46
44	*Phyllanthus emblica*	Phyllanthaceae	*O. sativa*	1.986 ± 0.023	2.9
45	*Glochidion sps.*	Phyllanthaceae	*V. radiata*	1.487 ± 0.002	
46	*Grevillea robusta*	Proteaceae	*O. sativa*	2.072 ± 0.020	1.6
47	*Zizyphus mauritiana*	Rhamnaceae	*O. sativa*	2.034 ± 0.014	2.35-2.44
48	*Neolamarckia cadamba*	Rubiaceae	*O. sativa*	1.712 ± 0.015	1.39
49	*Aegle marmelos*	Rutaceae	*S. lycopersicum*	0.672 ± 0.001	1.1
50	*Mimusops elengi*	Sapotaceae	*O. sativa*	1.883 ± 0.005	0.56
51	*Ailanthus altissima*	Simaroubaceae	*S. lycopersicum*	1.269 ± 0.010	2.08

Annexure III

(a) Annotated genes present in *C. quadrangularis* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replication	**Ribosomal RNA genes**	*rrn16* ^{D}	*rrn23* ^{D}	*rrn4.5* ^{D}	*rrn5* ^{D}	
	Transfer RNA genes	*trnA-UGC* ^{D*}	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*	*trnF-GAA*
		trnfM-CAU	*trnG-GCC*	*trnG-UCC*	*trnH-GUG*	*trnI-CAU* ^{D}
		trnI-GAU ^{D*}	*trnK-UUU*	*trnL-CAA* ^{D}	*trnL-UAA*	*trnL-UAG*
		trnM-CAU	*trnN-GUU* ^{D}	*trnP-UGG*	*trnQ-UUG*	*trnR-ACG* ^{D}
		trnR-UCU	*trnS-GCU*	*trnS-GGA*	*trnS-UGA*	*trnT-GGU*
		trnT-UGU	*trnV-GAC* ^{D}	*trnV-UAC* *	*trnW-CCA*	*trnY-GUA*
	Large subunit of ribosome (LSU)	*rpl2* ^{D*}	*rpl14*	*rpl16* *	*rpl20*	*rpl22*
		rpl23 ^{D}	*rpl33*	*rpl36*	*rpl32*	
	Small subunit of ribosome (SSU)	*rps2*	*rps3*	*rps4*	*rps7* ^{D}	*rps8*
		rps11	*rps12* ^{D*}	*rps14*	*rps15*	*rps16* *
		rps18	*rps19* ^{D}			
	RNA polymerase	*rpoA*	*rpoB*	*rpoC1* *	*rpoC2*	
Genes for photosynthesis	**Photosystem I**	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*

		psbL	_psbM_	_psbN_	_psbT_	_psbZ_
	Cytochrome b/f complex	_petA_	_petB_ *	_petD_ *	_petG_	_petL_
		petN				
	ATP synthase	_atpA_	_atpB_	_atpE_	_atpF_ *	_atpH_
		atpI				
	ATP-dependent protease subunit p	_clpP_ **				
	Large subunit of rubisco	_rbcL_				
	NADH dehydrogenase	_ndhA_ *	_ndhB_ ^{D*}	_ndhC_	_ndhD_	_ndhE_
		ndhF	_ndhG_	_ndhH_	_ndhI_	_ndhJ_
		ndhK				
Other genes	**Maturase**	_matK_				
	Envelop membrane protein	_cemA_				
	Subunit of acetyl-CoA-carboxylase	_accD_				
	C-type cytochrome synthesis gene	_ccsA_				
	Prolamin-box binding factor	_pbf1_				
Unknown function	**Hypothetical chloroplast reading frames**	_ycf_1	_ycf_2 ^{D}	_ycf_3 **	_ycf_4	_ycf_15 ^{D}

$*$ Genes having single intron.

$**$ Gene having two introns.

^{D} Gene having duplicated copies.

(b) Annotated genes present in *F. hispida* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replicating genes	rRNA genes	*rrn4.5*^{D}	*rrn5*^{D}	*rrn16*^{D}	*rrn23*^{D}	
	tRNA genes	*trnA-UGC*^{D*}	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*	*trnF-GAA*
		trnfM-CAU	*trnG-GCC*	*trnG-UCC**	*trnH-GUG*	*trnI-CAU*
		trnI-GAU^{D*}	*trnK-UUU**	*trnL-CAA*^{D}	*trnL-UAA**	*trnL-UAG*
		trnM-CAU^{D}	*trnN-GUU*^{D}	*trnP-UGG*	*trnQ-UUG*	*trnR-ACG*^{D}
		trnR-UCU	*trnS-GCU*	*trnS-GGA*	*trnS-UGA*	*trnT-GGU*
		trnT-UGU	*trnV-GAC*^{D}	*trnV-UAC**	*trnW-CCA*	*trnY-GUA*
	Large subunit of ribosome (LSU)	*rpl2*^{D*}	*rpl14*	*rpl16**	*rpl20*	*rpl22*
		rpl23^{D}	*rpl32*	*rpl33*	*rpl36*	
	Small subunit of ribosome (SSU)	*rps2*	*rps3*	*rps4*	*rps7*^{D}	*rps8*
		rps11	*rps12*^{D*}	*rps14*	*rps15*	*rps16**
		rps18	*rps19*^{D}			
	RNA polymerase	*rpoA*	*rpoB*	*rpoC1**	*rpoC2*	
Photosynthetic genes	Photosystem I	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbN*	*psbT*	*psbZ*

	Cytochrome b/f complex	*petA*	*petB**	*petD**	*petG*	*petL*
		petN				
	ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**	*atpH*
		atpI				
	ATP-dependent protease (p subunit)	*clpP***				
	RubisCO large subunit	*rbcL*				
	NADH dehydrogenase	*ndhA**	*ndhB*^{D*}	*ndhC*	*ndhD*	*ndhE*
		ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK				
Additional genes	**Maturase**	*matK*				
	Protein in envelop membrane	*cemA*				
	Acetyl-CoA-carboxylase subunit	*accD*				
	Cytochrome synthesis gene (C-type)	*ccsA*				
	Translation initiation factor	*inf*A				
Genes with unknown function	**Hypothetical reading frames for chloroplast**	*ycf*1^{D}	*ycf*2^{D}	*ycf*3**	*ycf*4	

$*$ Genes with one intron.

$**$ Gene having two introns.

^{D} Gene having duplicated copies.

(c) Annotated genes present in *S. rutidosperma* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replicating genes	**rRNA genes**	*rrn4.5* ^{D}	*rrn5* ^{D}	*rrn16* ^{D}	*rrn23* ^{D}	
	tRNA genes	*trnA-Ile* *	*trnA-UGC* *	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*
		trnF-GAA	*trnfM-CAU*	*trnG-GCC*	*trnG-UCC*	*trnH-CAC*
		trnH-GUG	*trnI-CAU* ^{D}	*trnI-GAU* *	*trnK-UUU*	*trnL-CAA* ^{D}
		trnL-UAA *	*trnL-UAG*	*trnM-CAU*	*trnN-GUU* ^{D}	*trnP-UGG*
		trnQ-UUG	*trnR-ACG* ^{D}	*trnR-UCU*	*trnS-AGC*	*trnS-GGA*
		trnS-UCA	*trnT-GGU*	*trnT-UGU*	*trnV-GAC* ^{D}	*trnV-UAC*
		trnW-CCA	*trnY-GUA*			
	Large subunit of ribosome (LSU)	*rpl2* ^{D*}	*rpl14*	*rpl16*	*rpl20*	*rpl22*
		rpl23 ^{D}	*rpl32*	*rpl33*	*rpl36*	
	Small subunit of ribosome (SSU)	*rps2*	*rps3*	*rps4*	*rps7* ^{D}	*rps8*
		rps11	*rps12* ^{D*}	*rps14*	*rps15*	*rps16* *
		rps18	*rps19* ^{D}			
	RNA polymerase	*rpoA*	*rpoB*	*rpoC1* *	*rpoC2*	
Photosynthetic genes	**Photosystem I**	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbN*	*psbT*	*psbZ*

	Cytochrome b/f complex	*petA*	*petB**	*petD**	*petG*	*petL*
		petN				
	ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**	*atpH*
		atpI				
	ATP-dependent protease (p subunit)	*clpP***				
	RubisCO large subunit	*rbcL*				
	NADH dehydrogenase	*ndhA**	*ndhB*^{D*}	*ndhC*	*ndhD*	*ndhE*
		ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK				
Additional genes	**Maturase**	*matK*				
	Protein in envelop membrane	*cemA*				
	Acetyl-CoA-carboxylase subunit	*accD*				
	Cytochrome synthesis gene (C-type)	*ccsA*				
	Prolamin-box binding factor	*pbf1*				
Genes with unknown function	**Hypothetical reading frames for chloroplast**	*ycf1*^{D}	*ycf2*^{D}	*ycf3***	*ycf4*	

$*$ Genes having one intron.

$**$ Gene having two introns.

^{D} Gene having duplicated copies.

(d) Annotated genes present in *S. asper* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replicating genes	rRNA genes	*rrn4.5*^{D}	*rrn5*^{D}	*rrn16*^{D}	*rrn23*^{D}	
	tRNA genes	*trnA-UGC*^{D*}	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*	*trnF-GAA*
		trnfM-CAU	*trnG-GCC*	*trnH-GUG*	*trnI-GAU*^{D*}	*trnK-UUU**
		trnL-CAA^{D}	*trnL-UAA**	*trnL-UAG*	*trnM-CAU*^{D}	*trnN-GUU*^{D}
		trnP-UGG	*trnQ-UUG*	*trnR-ACG*^{D}	*trnR-UCU*	*trnS-GCU*
		trnS-GGA	*trnS-UGA*	*trnT-GGU*	*trnT-UGU*	*trnV-GAC*^{D}
		trnV-UAC*	*trnW-CCA*	*trnY-GUA*		
	Large subunit of ribosome (LSU)	*rpl2*^{D*}	*rpl14*	*rpl16*	*rpl20*	*rpl22*
		rpl23^{D}	*rpl32*	*rpl33*	*rpl36*	
	Small subunit of ribosome (SSU)	*rps2*	*rps3*	*rps4*	*rps7*^{D}	*rps8*
		rps11	*rps12*^{D*}	*rps14*	*rps15*	*rps16**
		rps18	*rps19*			
	RNA polymerase	*rpoA*	*rpoB*	*rpoC1**	*rpoC2*	
Photosynthetic genes	Photosystem I	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbT*	*psbZ*	
	Cytochrome b/f complex	*petA*	*petB**	*petD**	*petG*	*petL*

		petN				
	ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**	*atpH*
		atpI				
	ATP-dependent protease (p subunit)	*clpP***				
	RubisCO large subunit	*rbcL*				
	NADH dehydrogenase	*ndhA**	*ndhB*^{D*}	*ndhC*	*ndhD*	*ndhE*
		ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK				
Additional genes	**Maturase**	*matK*				
	Protein in envelop membrane	*cemA*				
	Acetyl-CoA-carboxylase subunit	*accD*				
	Cytochrome synthesis gene (C-type)	*ccsA*				
	Prolamin-box binding factor	*pbf1*				
Genes with unknown function	**Hypothetical reading frames for chloroplast**	*ycf*1^{D}	*ycf*2^{D}	*ycf*3**	*ycf*4	

$*$ Genes having one intron.

$**$ Gene having two introns.

^{D} Gene having duplicated copies.

(e) Annotated genes present in *T. peruviana* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replicating genes	**rRNA genes**	*rrn4.5*^{D}	*rrn5*^{D}	*rrn16*^{D}	*rrn23*^{D}	
	tRNA genes	*trnA-UGC*^{D*}	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*	*trnF-GAA*
		trnfM-CAU	*trnG-GCC*^{D*}	*trnH-GUG*	*trnI-GAU*^{D*}	*trnK-UUU**
		trnL-CAA^{D}	*trnL-UAA**	*trnL-UAG*	*trnM-CAU*^{D}	*trnN-GUU*^{D}
		trnP-UGG	*trnQ-UUG*	*trnR-ACG*^{D}	*trnR-UCU*	*trnS-GCU*
		trnS-GGA	*trnS-UGA*	*trnT-GGU*	*trnT-UGU*	*trnV-GAC*^{D}
		trnV-UAC*	*trnW-CCA*	*trnY-GUA*		
	Large subunit of ribosome (LSU)	*rpl2*^{D*}	*rpl14*	*rpl16*	*rpl20*	*rpl22*
		rpl23^{D}	*rpl32*	*rpl33*	*rpl36*	
	Small subunit of ribosome (SSU)	*rps2*	*rps3*	*rps4*	*rps7*^{D}	*rps8*
		rps11	*rps12*^{D*}	*rps14*	*rps15*	*rps16**
		rps18	*rps19*^{D}			
	RNA polymerase	*rpoA*	*rpoB*	*rpoC1**	*rpoC2*	
Photosynthetic genes	**Photosystem I**	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbN*	*psbT*	*psbZ*

	Cytochrome b/f complex	*petA*	*petB*	*petD*	*petG*	*petL*
		petN				
	ATP synthase	*atpA*	*atpB*	*atpE*	*atpF*	*atpH*
		atpI				
	ATP-dependent protease (p subunit)	*clpP* [**]				
	RubisCO large subunit	*rbcL*				
	NADH dehydrogenase	*ndhA*[*]	*ndhB* [D*]	*ndhC*	*ndhD*	*ndhE*
		ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK				
Additional genes	**Maturase**	*matK*				
	Protein in envelop membrane	*cemA*				
	Acetyl-CoA-carboxylase subunit	*accD*				
	Cytochrome synthesis gene (C-type)	*ccsA*				
	Translation initiation factor	*infA*				
Genes with unknown function	**Hypothetical reading frames for chloroplast**	*ycf*1[D*]	*ycf*2 [D]	*ycf*3[**]	*ycf*4	

* Genes having one intron.

** Gene having two introns.

[D] Gene having duplicated copies.