AN ABSTRACT OF THE THESIS OF

Sushovita Pal for the degree of Master of Science in Integrative Biology

presented on December 8, 2022.

Title: Seeking sacred sequences: A comparative population-genetic analysis of *Ficus religiosa* chloroplast and mitochondrial DNA.

Abstract approved: _____

Dee R. Denver

Chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA) are often used for the study of plant genetics and evolution. Plastid and mitochondrial genome sequences have allowed us to investigate plant evolution on a genetic level to infer molecular evolutionary rates, population-genetic processes, co-evolutionary phenomena, and numerous evolutionary questions and hypotheses. *Ficus religiosa*, a fig species commonly referred to as Bodhi trees and sacred peepal trees, has broad ecological and cultural relevance along with growing scientific and medicinal features of interest. However, evolutionary, and genetic analyses of this species are

lacking, despite this growing interest and cultural ties to Buddhism and other spiritual traditions of Asia. This thesis aims to investigate evolutionary processes in *F. religiosa* organelle DNA through a population-genetic lens and interrogate various historical and apocryphal stories surrounding claims of descent from the historical Bodhi tree in Bodhgaya, India under which Siddhartha Gautama, the Buddha, sat during his enlightenment in ~500-600 BCE. This thesis research relies on a collection of 61 different *F. religiosa* samples from across the globe, including representatives from two sacred Bodhi tree lineages: the Bodhi tree currently at the Buddha's seat of enlightenment in Bodhgaya, India and Sri Maha Bodhi in Anuradhapura, Sri Lanka. The first major objective was to examine the relative levels of DNA sequence polymorphism in *F. religiosa* cpDNA and mtDNA using population genetic metrics such as nucleotide diversity (π). A second objective was to investigate the population-genetic structure in the 61 *F. religiosa* samples to investigate the genetic relationships of the plants examined, and to assess claims of descent from the original sacred Bodhi tree. The thesis results revealed higher genetic diversity in cpDNA to mtDNA by a factor of ~15, consistent with patterns observed in most other plant species. Population-genetic structure analyses of the 61 *F. religiosa* samples for cpDNA showed four distinct clusters; 5/5 samples

known to be derived from the Bodhgaya Bodhi tree and 4/4 samples from the Sri Maha Bodhi tree all showed near 1.0 probability of membership to the same cluster. Four trees with uncertain apocryphal stories were analyzed; 3/4 showed high probability of membership to the same cluster as the Bodhgaya Bodhi and Sri Maha Bodhi. This thesis provided new insights into *F. religiosa* organellar genome evolution and an important historical-genetic information for a tree lineage held sacred by hundreds of millions of people around the globe, and a novel genetic approach to assessing claims of ancestry to the sacred Bodhi tree lineage.

Seeking sacred sequences: A comparative population-genetic analysis of *Ficus religiosa* chloroplast and mitochondrial DNA

by

Sushovita Pal

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented December 8, 2022
Commencement June, 2023

Master of Science thesis of Sushovita Pal presented on December 8, 2022.

APPROVED:

Major Professor, representing Integrative Biology

Head of the Department of Integrative Biology

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes the release of my thesis to any reader upon request.

Sushovita Pal, Author

ACKNOWLEDGEMENTS

This thesis is a labor of love and would not have been possible without the love and support of people around me. I wish to express my deepest gratitude to my advisor Dr. Dee Denver for the support, kindness, and guidance he has inspired me with. He believed in me, understood me, and trusted me as I navigated through life situations and mentored me to pursue what truly matters to me. His constant support and thoughtful presence have helped me push through the stumbling blocks I came across in the thesis and life in general. No amount of thank you will ever be enough.

Also, I would like to express my gratitude towards the graduate committee: Professors Jerri Bartholomew, Jamie Cornelius, and David Maddison, each of whom provided constructive feedback and support throughout my degree. I owe my sincere gratitude to Dana Howe, a mentor, a woman in STEM, whose sheer presence and competence has always inspired me. Her support and expertise in the lab and in scientific writing has been paramount to me being equipped with skills in lab and whose feedback with writing has only helped me hone in those writing skills for the thesis. I would like to extend my gratitude to Dr. Sulochana Wasala whose presence in the lab was very inspiring and whose collaboration during Dr.

Denver's travel and sample collection was paramount to my thesis. I would also convey sincere gratitude to Amila Liyanage for the travel support during sample collections.

I would like to thank the undergraduates before me who were instrumental to this thesis. Their support with the sample collection, cpDNA sequencing is what has made the thesis what it is today. I thank Beth Davis, the first Bodhi Tree Project undergraduate student, who developed the lab's *F. religiosa* DNA extraction protocol. I also thank Rebekah Cleary, Jeilene Hambly, Christina Lew, and Steffani Myers who contributed to cpDNA and mtDNA sequencing. Special thanks to Triet Ho and Hannah LaVeine for contributing to the sample collection and mtDNA sequencing as well as for their warm presence in the lab for the past year. The samples for my thesis were collected from diverse and global locations that required important partnerships. I thank Gerald Hirata (Kaua'i Soto Zen temple), Kim Starr (Starr Environmental, Mau'i), Paramacharya Sadasivanathaswami (Kaua'i Himalaya Academy), Glenn Coryell, Reverend Tennes (Kahalui Buddhist Temple), Kerry Kiyohara (Makawai Hongwanji Mission), Thilak Rodrigo (Samadhi Buddhist Vihara), Ajahn Chandako (Vimutti Buddhist Monastery) for the sample collection during travel. Adding to my gratitude for the lab, I am very grateful for

the presence and company of two wonderful humans: Brenna Prevelige and Jazlee Crowley who in a very short period have managed to empower me and instill strength in me to move ahead with my research direction. I express my sincere gratitude to the 4 T's and 1 M: Tresa, Tara, Traci, and Trudy, and Melissa as well as the Department of Integrative Biology, who were integral in helping me navigate the academic life.

I would like to thank my entire family especially my parents, my uncle, aunt, and cousins for their support and love, my sister Awantika for the strength with which she exists in my life, my husband Aayam for his proficiency in R, python, and everything digital. Finally, I would like to thank my guru; Sadhguru - a phenomenal being whose tools for wellbeing I shamelessly use and bow down to.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Chapter 1: Introduction

The "tree of life" paradigm has come to symbolize the interconnectedness and diversity of all life on the planet. Bacteria, Archaea, and Eukaryota are the three domains of life that contain all organismal lifeforms (Letunic and Bork, 2019). Bacteria and Archaea are the first two domains exclusive of minuscule single-celled organisms. Eukaryota, the third domain, encompasses many microscopic species and well-known groupings like animals, plants, and fungi.

Among the Eukaryota, plants hold a crucial link to understanding the evolution of life as the terrestrial biosphere, geosphere, and global temperatures were all modified by plant colonization of land (Beraldi-Campesi, 2013; Strother et al., 2011; Padian et al., 1992). We cannot claim to have a complete understanding of evolution unless we understand plant biology, given that plant evolution has shaped so much change of our planet. Important metabolic systems arose during the evolution of plants, changing both the physical and biological worlds of our planet, such as the development of photosynthesis. Chloroplasts and mitochondria collaborate to provide energy to plants; together, they form the powerhouses of the plant cell. Chloroplasts and mitochondria possess their own genomes. This is due

to the evolutionary origins of both plastids (chloroplasts) and mitochondria by endosymbiosis, which stand as two of the most critical events in the history of eukaryotic life (Sagan, 1967). We now know that a cyanobacterial endosymbiont's genetic, biochemical, and cell biological integration into a heterotrophic host eukaryote approximately a billion years ago (Sagan, 1967) paved the way for the evolution of diverse algal groups in a wide range of aquatic and eventually, terrestrial environments.

Mitochondria and chloroplast are double membrane-bound semiautonomous organelles that harbor genetic material: mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA), respectively. These organelles are equipped with the associated molecular machinery for the regulation of gene expression (Gutman and Niyogi, 2009; Smith and Keeling, 2015; Peralta Castro et al., 2020). Hence, plants possess two small and cytoplasmic extrachromosomal genomes outside the nuclear genome, which reside in mitochondria and chloroplast. Both cpDNA and mtDNA evolved from endosymbiont bacterial genomes from cyanobacteria and α-proteobacteria, respectively (Sagan, 1967; Chevigny et al., 2019). However, over long evolutionary time, a vast majority of the genes of ancestral prokaryotes were transferred to the host cell's nucleus (Martin, 2015). Consequently, present-day

plant chloroplast and mitochondrial genomes contain approximately 100 - 200 genes that encode proteins required for indispensable components for oxidative phosphorylation and photosynthesis. Nevertheless, chloroplasts and mitochondria harbor several thousands of proteins, and most are encoded by the nucleus, translated into the cytoplasm, and transported to their target organelle. (Saki and Prakash, 2017; Sakamoto and Takami, 2018).

The chloroplast and mitochondrial genomes are often used for the study of plant evolution. (Qiu et al., 2010; Olmstead and Palmer,1994). cpDNA has been a focus of plant molecular evolution and systematics research. Several features of this genome have facilitated molecular evolutionary analyses, such as its small size, extensive molecular characterization, and lower rates of nucleotide substitution, which provide features that facilitate the study of plant phylogeny. (Chevigny et al., 2019, Clegg et al., 1994). On the other hand, plant mtDNA shows a remarkable variation in size and is either linear or circular in structure, depending on the species (Stewart, 2017). Unlike the situations in animals where mtDNA usually evolves at very high rates, plant mtDNA often shows extremely low rates of nucleotide substitution (Palmer, Jeffrey D. et al.). This has led to mtDNA often not being the focus of many plant molecular phylogenetic studies. A recent assessment of

genomic resources revealed 7,367 chloroplast and 441 mitochondrial genome sequences in the NCBI GenBank Organelle Genome Resources of land plants (Yang et al., 2022).

There has been evidence that the synonymous substitution rates in mitochondrial genes are typically several times slower than in plastid genes (Sloan et al., 2012) Recent discoveries, however, suggest that there is more variation in plant organelle genome evolution than previously thought. For example, it has been widely generalized that plastid genes evolve faster than mitochondrial genes in plants. However, Zhu et al., 2014 showed that intragenomic rate heterogeneity exists by uncovering a 340-fold range of synonymous substitution rate variation among the mitochondrial and plastid genomes from bugleweed, *Ajuga reptans*. This is by far the largest amount of synonymous rate heterogeneity ever reported for a genome. However, the evolutionary forces driving this phenomenon are unclear and require further investigation.

Other reports also point to a potential case for lineage-specific accelerated mtDNA sequence evolution in angiosperms; case in point Parkinson et al., 2005, a study that provided evidence for reversible underlying changes in the mitochondrial mutation rate in the Pelargonium family Geraniaceae. These findings suggest that

the pace of mitochondrial mutation in plants is more malleable than previously thought. Nevertheless, our ability to sequence and study these plastid and mitochondrial genomes has allowed us to investigate this for an ecologically and culturally important species of interest, such as figs.

The genus Ficus (Moraceae) constitutes one of the largest genera of angiosperms, with more than 800 species of trees, epiphytes, and shrubs in the family Moraceae, many of which are commonly known as figs (Britannica, 2010). Native primarily to tropical areas of East Asia, many fig species are now distributed throughout the tropical and subtropical regions worldwide. Many are tall forest trees buttressed by great spreading roots; others are ornamental. *F. religiosa* has piqued widespread interest due to its religious and cultural relevance along with interesting scientific and medicinal attributes. *F. religiosa* is well known for its long lifespan, cultural relevance, and medicinal properties (Singh, Singh, and Goel, 2011). More recently, it has also been studied for its antioxidant properties (Shankar et al., 2021; Karunanidhi et al., 2021; Agarwala et al., 2018), along with its potential applications for applicability in emulsions (Aslam and Akhtar, 2021) and potential for diverse pharmacological activities (Saida, Tulasi, and Narasu, 2021). These and other interesting biological properties of this species include antimicrobial

functioning (Bhavyashree and Xavier, 2021), anti-diabetic potential (Jayant and Vijayakumar, 2021; Pandit, Phadke, and Jagtap, 2010; Senthilkumar and Manimekalai, 2020), interaction with nitrogen-fixing bacteria (Lin et al., 2020), and root fiber properties (Moshi et al., 2020).

Siddhartha Gautama, the historical Buddha, is said to have experienced enlightenment under a *F. religiosa* tree, widely known as the Bodhi tree, approximately 2,500 years ago. This lineage of *F. religiosa* holds great significance to Buddhist traditions: that of the sacred Bodhi tree under which the Buddha sat during his enlightenment in ~500-600 BCE, in what is today called Bodhgaya, India. After this experience, the Buddha traveled the Ganges basin for the next 45 years of his life, teaching his newfound wisdom and path to enlightenment to an ever-increasing group of followers until his death at the age of 80. After his death, Buddhist teachings spread across the Asian continent in the two millennia that followed and further into other continents in recent times. As Buddhism spread, the original Bodhi tree saplings would often be ceremonially shared with newly established temples and monasteries. Trees derived from this original Bodhi tree are widely considered to be sacred in Buddhist traditions.

Today, two living Bodhi trees of special significance to Buddhist communities exist. First, the Bodhi tree at Anuradhapura, Sri Lanka - known as 'Sri Maha Bodhi' - arrived and was planted in 288 BCE. A sapling of the original Bodhi tree in Bodhgaya was brought to Sri Lanka by Sanghamitta, the daughter of the Buddhist Indian king Ashoka, and ceremoniously planted in Anuradhapura. This tree continues to live today, is widely recognized as the oldest historical tree on the planet and is a UNESCO World Heritage Site (Britannica 2019). Saplings derived from this tree have, in turn, been ceremoniously dedicated to other Buddhist communities around the world over the last two millennia, including the Bodhi tree at Foster Botanical Garden in Honolulu, Hawai'i, that is included in the present study (sample O15). Second, the tree currently growing at the site of the Buddha's enlightenment in Bodhgaya, India is of special significance to Buddhist communities – this tree is visited by many thousands of Buddhist pilgrims each year. The genetic legacy of this tree, however, is unclear. Over the last twenty-five centuries, the tree at this site has experienced many assaults and periods of neglect as Buddhism's prominence in India declined over the two millennia. For example, in 254 BC, Tissarakkha, Ashoka's queen, destroyed the original Bodhi tree in Bodhgaya, as she did not favor Ashoka embracing Buddhism (D Krishnan,

2013). However, a second tree grew shortly from the roots of the destroyed tree, replacing the original. Despite this and other stories of tree destruction followed by regeneration, the Muslim invasion of India during the 13th-17th centuries CE dealt Indian Buddhism and the Bodhi tree unrecoverable damage.

Between 1862-1878, Alexander Cunningham, a British archaeologist, excavated the Mahabodhi Temple at Bodhgaya and discovered a much decayed and dead Bodhi tree (D Krishnan, 2013). Cunningham planted a new Bodhi tree, from a nearby *F. religiosa* sapling, in 1881. Although this tree has inspired many thousands of Buddhists who have visited this site over the last century, its genetic relationship to the 'original' Bodhi tree under which the Buddha sat remains unknown. Further, many trees around the world claim ancestry to the original Bodhi tree, with various degrees of historical or other information backing up these claims. The stories accompanying the heritage of various *F. religiosa* – often with claims of ancestry to the original Bodhi tree – are often described as apocryphal in scholarly communities. The research described here provides the opportunity to shed light onto the genetic interrelationships of Bodhi trees around the planet and interrogate some apocryphal claims.

Owing to its broad ecological and cultural importance, many aspects of *F. religiosa* biology have been studied. However, evolutionary, and genetic analyses of this species are lacking. A 2016 study examined the matK cpDNA gene as a species barcoding tool for *F. religiosa* but did not reveal substantial insights into within-species diversity (Sivalingam, D. 2016). A 2020 study provided a phylogenomic analysis of *F. religiosa* and other fig species but did not analyze within-species evolutionary patterns (Wang et. al. 2020). To provide insights into within species *F. religiosa* evolution, this thesis investigated cpDNA and mtDNA variation among many trees from diverse worldwide locations and commercial sources. Two objectives were investigated:

For the first objective, I examined the relative levels of DNA sequence polymorphism in *F. religiosa* cpDNA and mtDNA. Based on previous observations in other plant systems (Bryan,1999), the level of *F. religiosa* sequence polymorphism was hypothesized to be lower in mtDNA than in cpDNA. This was investigated by analyzing *F. religiosa* cpDNA and mtDNA using a DNA sequencing and population genetic analysis strategy. Population-genetic measures such as nucleotide diversity ($\boldsymbol{\pi}$) were estimated and compared between *F. religiosa* cpDNA and mtDNA.

For the second objective, I investigated population genetic structure in the *F. religiosa* samples analyzed, using both cpDNA and mtDNA. These analyses provided an opportunity to investigate the potential genetic relationships of *F. religiosa* with varying claims of descent from the original sacred Bodhi tree. Patterns of population genetic clustering in *F. religiosa* cpDNA and mtDNA structure were analyzed using STRUCTURE and a principal coordinate analysis approach. This thesis provided new insights into organellar genome evolution within a plant lineage and important historical genetic insights into a tree lineage held sacred by millions of people on the planet.

Chapter 2: Methods.

2.1 Leaf Sample Collection:

F. religiosa leaf samples were obtained from diverse locations around the world for this analysis. Many Denver lab research team members contributed to this effort over the last decade. The samples were collected through a mixture of international research travel, sourcing through the mail, and procurement of *F. religiosa* plants and seeds from commercial sources. While many of these samples had strong historical claims of descent from one of the two famous sacred trees (either the Sri Maha Bodhi tree or the Bodhgaya Bodhi tree), some came accompanied by various types of apocryphal stories, and others had no claims of sacred descent. All leaves collected from sacred temples or monasteries were done with permission from resident monks and/or administrators. The details are presented in Table 2.1.

Table 2.1: List of *F. religiosa* leaves collected by the Denver lab for DNA extraction.

DNA Sample	**Source** [a]	**Ancestry** [b]
BG1A	Bodhi tree currently at Buddha's seat of enlightenment, Maha Bodhi temple, Bodhgaya, India	BGB
BG1B	Bodhi tree currently at Buddha's seat of enlightenment, Maha Bodhi temple, Bodhgaya, India	BGB
BG1C	Bodhi tree currently at Buddha's seat of enlightenment, Maha Bodhi temple, Bodhgaya, India	BGB
BG2	Inside Maha Bodhi Complex, Bodhgaya, India	UNK
BG3	Inside Maha Bodhi Complex, Bodhgaya, India	UNK
BG4	Shechen Tennyi Dargyeling Nyingma Buddhist monastery, Bodhgaya, India	UNK
BG6	Mahabodhi complex walkway entrance, Bodhgaya, India	UNK
BG7	Maha Bodhi Society building, Bodhgaya, India	UNK
BG8	In front of Hari Om Cafe, Bodhgaya, India	UNK
BG9	In front of Bangladesh Buddhist monastery, Bodhgaya, India	UNK
BG10	In front of Wat Thai temple, Bodhgaya, India	UNK
BG11	Inside Maha Bodhi Complex, Bodhgaya, India	UNK
BG12	On road past Cambodian temple, Bodhgaya, India	UNK
BTN	commercially sourced, Bodhi Tree Nursery, Dallas, TX, USA	APO-BGB
CG14	commercially sourced, CiboGro, Gansevoort, NY, USA	UNK
JCU	James Cook University campus, Douglas, Queensland, Australia	APO-BGB
K1	Kaua'i Hindu Monastery, Kaua'i, HI, USA	APO-BGB
K2	County Building, Lihue, Kaua'i, HI, USA	UNK
K3	Kaua'i Soto Zen Temple, Kaua'i, HI, USA	BGB
M1	Mau'i Arts and Cultural Center, Mau'i, HI, USA	UNK

DNA Sample	Source [a]	Ancestry [b]
M2	U. of Hawai'i, Mau'i, Mau'i, HI, USA (large tree)	UNK
M3	U. of Hawai'i, Mau'i, Mau'i, HI, USA (seedling)	UNK
M5	U. of Hawai'i, Mau'i, Mau'i, HI, USA (seedling)	SMB
M7	Paia Rinzai Zen Buddhist Temple, Mau'i, HI, USA	UNK
M9	Lahaina Hongwanji Buddhist Mission, Mau'i, HI, USA	UNK
M11	Lahaina Jodo Buddhist Mission, Mau'i, HI, USA	BGB
ML1	Mountain Lamp Meditation Retreat Center, Denning, WA, USA	APO-BGB
NZ1	Vimutti Theravada Buddhist Monastery, Auckland, New Zealand	UNK
NZ2	Samadhi Buddhist Vihara, Rolleston, New Zealand	SMB
O4	U. of Hawai'i, Mau'i, Mau'i, HI, USA (derived from Foster Garden tree)	SMB
O5	Ventura Street, little sapling by driveway, Oahu, HI, USA	UNK
O6	Ventura Street, little sapling by driveway, Oahu, HI, USA	UNK
O7	Ventura Street, little sapling by driveway, Oahu, HI, USA	UNK
O12	Pacific National Cemetery Oahu, HI, USA	UNK
O14	Pacific National Cemetery Oahu, HI, USA	UNK
O15	Foster Botanical Garden, Oahu, HI, USA	SMB
O16	Soto Mission of Hawaii Shojobi, Honolulu, HI, USA	UNK
O17	Nichiren Mission of Hawaii, Honolulu, HI, USA	UNK
O18	Wahiawa Hongwanji Mission, Oahu, HI, USA	UNK
SL1	Gangaramaya Temple, Colombo, Sri Lanka	UNK
SL2	Kandy, Sri Lanka	UNK
SL4	Outside Temple of the Tooth, Kandy, Sri Lanka	UNK
SL5	Thuparama from ground, 3 trees inside wall, Anuradhapura, Sri Lanka	UNK

DNA Sample	Source [a]	Ancestry [b]
SL10	Tivakka, Sri Lanka	UNK
SL12	Sri Vajiragnana, Anuradhapura, Sri Lanka	UNK
SL14	Tree growing out of the wall of SMB complex, Anuradhapura, Sri Lanka	UNK
SL15	SMB Complex Parking Lot, Anuradhapura, Sri Lanka	UNK
SL19	Nilketha Villa Eco-Hotel, Anuradhapura, Sri Lanka	UNK
SL20	Abhayasekarama Temple, Negombo, Sri Lanka	UNK
SL21	Angurukaramulla Temple, Negombo, Sri Lanka	UNK
SL22	Weuda, Kurunegala-Kandy Road, Kurunegala, Sri Lanka	UNK
SL23	Weuda, Kurunegala-Kandy Road, Kurunegala, Sri Lanka	UNK
SM1	Commercially sourced, Strictly Medicinal, Williams, OR, USA	UNK
SS2	Commercially sourced, Sacred Seeds, USA	UNK
TS8	Commercially sourced, Sacred Seeds, Winooski, VT, USA	UNK
TW1	Commercially sourced, Sacred Seeds, Santa Rosa CA, USA	UNK

[a] commercially sourced samples were ordered over the internet
[b] BGB = Bodhgaya Bodhi tree; SMB = Sri Maha Bodhi tree; APO-BGB = apocryphal claim of sacred Bodhi tree ancestry from Bodhgaya; UNK = Unknown

2.2 DNA Extraction, PCR, DNA Sequencing

All collected leaves were dried and preserved at 4°C prior to the DNA extraction process. Total DNA extraction was performed following the Qiagen DNeasy Plant Kit protocol described in Costa and Roberts (2014), with the slight modification

that all samples (~20 mg) were grounded with a frozen mortar and pestle instead of a mini bead beater. PCR amplification was achieved using the established protocol (Denver et al., 2003), with the slight modification that the starting genomic DNA was diluted 1:20 in water. After PCR, the samples were run on an agarose gel. Successful PCR samples were purified following the solid phase reversible immobilization protocol (Elkin et al., 2001) using magnetic beads to isolate amplicons. After the purification, direct-end sequencing reactions were performed using ABI big dye; these reactions were then sent to the Center for Quantitative Life Sciences (CQLS) for Sanger sequencing using the ABI 3730xl DNA Analyzer.

For this analysis, long homopolymeric nucleotide runs within the reference *F. religiosa* chloroplast genome (NC_033979.1; Brun et al., 2017) and mitochondrial genomes (Wang et al., 2021) were targeted due to their high mutation rates and associated expectation of within-species natural variation. Homopolymer regions were identified using DNAsp (Rozas, 2009) for cpDNA. Primer3 was used to design species-specific primers to target and amplify the homopolymers (Ye et al., 2012; Untergasser et al., 2012). PCR amplification and direct-end sequencing followed standard Denver lab protocols (Denver, 2003).

The reference *F. religiosa* mitochondrial genome reported by Wang et al. (2021) was obtained through a direct request to the corresponding author. However, unlike the chloroplast genome, which was assembled into a complete singular genome (Brun et al., 2017), the mitochondrial genome was assembled into 15 contigs. The complete structure of *F. religiosa* mtDNA is unknown; it might be that these contigs represent 15 distinct linear DNA molecules or are subcomponents of a single large circular mitochondrial genome. Bio python was used to examine mtDNA contig length, %AT, and homopolymer stretch (Cock et al., 2009). The exact methodology is described in source code A.1 (see Appendix). which parses and iterates through each contig to find the respective quantities of interest, such as lengths and %AT, and source code A.2, which parses and creates a list of all homopolymers, filtering them as per length requirement (see Appendix).

Using this information, Primer3 was used to design primers to target the longest mtDNA homopolymer stretches, 12bp or greater, with the expectation of the 10bp C homopolymer on contig 15 (Ye et al., 2012; Untergasser et al., 2012). A total of 16 primer sets were designed and tested using BTN genomic DNA (gDNA). Eight primer sets were successfully amplified and sequenced with BTN.

These eight primer sets were then used to generate cpDNA sequence data for all the extracted gDNA samples. Table 2.2 includes the total number of loci we worked with from both cpDNA and mtDNA.

Table 2.2 *F. religiosa* loci details for cpDNA and mtDNA.

Organelle DNA	**Locus Name**	**Reference Position**	**Homopolymer Nucleotide**	**Homopolymer Length**
cpDNA	FrCp14723	14723	A	12
cpDNA	FrCp60605	60605	G	9
cpDNA	FrCp58543_5'	58543	T	10
cpDNA	FrCp58543_3'	58543	A	10
cpDNA	FrCp118538	118538	A	16
cpDNA	FrCp82452	82452	A	12
cpDNA	FrCp85060	85060	T	12
mtDNA	C9-7	2527	A	12
mtDNA	C10-17	938	T	12
mtDNA	C10-20	23366	T	12
mtDNA	C11-22	5061	A	12
mtDNA	C11-24	8930	T	12
mtDNA	C12-29	34182	T	12
mtDNA	C13-34	32757	A	13
mtDNA	C15-48	695	C	10

2.3 Comparative DNA sequence analysis.

2.3.1 DNA sequence quality control and alignment

MEGA software, version 10, was used to examine and perform quality control on all the electropherograms (Tamura et al., 2013). First, fluorescent DNA sequence data were visually inspected to ensure that the peaks were clear and distinct from one another and that there was little to no background signal in the electropherograms. Second, for homopolymer regions where variation was expected to predominate, the number of peaks was carefully manually counted to double-check the presence of the same number of nucleotides in both the forward and reverse primer. For example, if there were 12 A's in the reference sequence, it was checked to see if the forward sequence had 12 A’s too, and the reverse had 12 T’s. After DNA sequence quality control, MEGA was used to perform multiple sequence alignments using the embedded application MUSCLE (Kumar et. al. 2018). DNA sequences were then aligned to the reference mtDNA sequence (Wang. et al., 2021), and forward and reverse DNA reads were combined to yield full-amplicon sequences. These completed amplicons were used to identify

homopolymer variations, SNPs, and any other forms of variation within the sequence alignment.

2.3.2 Comparative nucleotide diversity

The relative levels of DNA sequence polymorphism in *F. religiosa* cpDNA and mtDNA were investigated for the first objective. To do this, nucleotide diversity analyses were performed, which offer one measure of the degree of DNA sequence polymorphism within a population, as first introduced by Nei and Li (1979). Nucleotide diversity is the average number of nucleotide differences per site between two DNA sequences chosen randomly from a sample population. It is mathematically defined as

$$\text{Diversity} = \sum_{ij} x_i x_j \pi_{ij}$$

Where i, j is the proportion of different nucleotides between the i^{th} and j^{th} types of DNA sequences and are the respective frequencies of these sequences.

After DNA sequence alignment, estimated nucleotide diversity (π) for *F. religiosa* cpDNA and mtDNA using two different software packages: MEGA, version 10.0 (Kumar et. al. 2018) and DnaSP (Librado & Rozas, 2009) were calculated. DnaSP calculates the average number of nucleotide differences per site

between two sequences, or nucleotide diversity, Pi (Nei 1987, equations 10.5 or 10.6), as well as its sampling variance and standard error (Nei 1987, equation 10.7). While DnaSP allows only one method for calculating nucleotide diversity, it provides auxiliary statistics such as site variation details, haplotype diversity, and standard error. Mega, on the other hand, allows us to get nucleotide diversity using an ensemble of models such as Maximum Composite Likelihood and p-distance; however, its result only includes the nucleotide diversity and lacks auxiliary details such as site variation or standard errors.

2.4 Population genetic analysis:

The population genetics program STRUCTURE was used for comparative analysis. It is an open-source program for population analysis developed by Pritchard et al. (2000a). STRUCTURE analyzes differences in the distribution of genetic variants amongst populations with a Bayesian iterative algorithm by placing samples into groups whose members share similar patterns of variation. STRUCTURE identifies populations from the data and assigns individuals to that population representing the best fit for the variation patterns. Typically, STRUCTURE is the first step in

examining population structures that emerge from the sample set to provide a preamble to further genetic analysis or to infer the origins of individuals with unknown population characteristics.' (Pritchard et al., 2000)

More specifically, STRUCTURE version 2.3.4 (Raj, Anil et al. 2013) was used to investigate the genetic cluster in the *F. religiosa*. STRUCTURE applies a model to the data of K assumed populations or genetic groups, each characterized by a subset of allele frequencies identified in the data. As we do not know the number of clusters K beforehand, we calculate the likelihood of the data for a range of K values by creating posterior probabilities for the chosen cluster size K. Here, a range of predefined K settings is run to obtain their posterior probabilities. These values will increase as K increases will be the largest for the most appropriate K, beyond which probabilities tend to be very similar for higher K values. Hence, the smallest stable K value represents the optimum value. I analyzed all 61 mtDNA and cpDNA sequences from across the globe, respectively, for (number of target clusters) K = 1 to K = 8. For each K, three independent runs with a burn-in of 1,000,000 and 1,000,000 iterations were performed, and all other parameters were set to the default values. To determine the best K value, ΔK (Evanno et al. 2005) in the R package Pop Helper was used (Francis, Rob. 2017). The STRUCTURE plots were

then visualized for the best K in the stacked bar plot formats. The goal here was to analyze the results for the K that best supports the data to study the influence of sacred/non-sacred lineage on the clusters.

Using homopolymer and SNP data from each sample, principal coordinate analysis (PCoA) was done using the sci-kit learn python package (Pedregosa, Fabian et al. 2011). Homopolymer and SNP features were generated for each sample and concatenated to form a single feature. Here the homopolymer features comprised the homopolymer length at different homopolymer sites, whereas the SNP features comprised one hot encoding for the SNP nucleotide for each SNP site. Scaling transformation was applied to these features before PCoA as the homopolymer features, and SNP features had different scales, which is known to introduce errors during PCoA. The Seaborn package was used for the visualization of the principal components.

Chapter 3: Results

3.1 Analysis of Genetic Diversity in *F. religiosa* cpDNA and mtDNA

Seven regions of cpDNA and eight regions of mtDNA were analyzed in 61 *F. religiosa* samples. The seven sets of cpDNA PCR primers were designed using the NCBI reference sequence, with each locus centered on a homopolymeric nucleotide run. Each of these seven loci was PCR-amplified and direct-end sequenced from our 61 different *F. religiosa* samples (Table 3.1). A total of approximately 177,100 bp of cpDNA was sequenced from the 61 samples, with an average of 2,900 bp per sample. Similarly, the eight sets of mtDNA PCR primers were designed using the reference provided by (Wang et al.). Approximately 232,350 bp of mtDNA was sequenced across the 61 samples, with an average of 3,800 bp per sample. The summary statistics and loci details for both chloroplast and mitochondrial data are detailed in table 3.1.

Table 3.1: *F. religiosa* cpDNA and mtDNA loci examined. Here we show the number of homopolymer variants observed across different samples along with the number of samples each variant was observed for. For example. 2: G8 (49), G9 (11) entails that overall, 2 variants were observed were G8 variant and G9 variant was seen in 49 and 11 samples respectively.

Organelle DNA	**Locus Name**	**Reference Homopolymer**	**No. homopolymer allele variants observed**	**No. SNP variant sites**
cpDNA	FrCp14723	A12	5: A9 (6), A10 (2), A11 (30), A12 (22), A15 (1)	2
cpDNA	FrCp60605	G9	2: G8 (49), G9 (11)	3
cpDNA	FrCp58543_5'	T10	4: T9 (1), T10 (41), T11 (18), T12 (1)	4
cpDNA	FrCp58543_3'	A10	3: A9 (1), A10 (48), A11 (12)	2
cpDNA	FrCp118538	A16	5: A13 (6), A14 (2), A15 (3), A16 (48), A18 (2)	1
cpDNA	FrCp82452	A12	4: A10 (1), A11 (15), A12 (39), A13 (6)	2
cpDNA	FrCp85060	T14	4: T12 (1), T13 (12), T14 (47), T15 (1),	2
mtDNA	C9-7	A12	2: A12 (59), A14 (2)	1
mtDNA	C10-17	T12	2: T12 (60), T13 (1)	2
mtDNA	C10-20	T12	1: T12 (61)	1
mtDNA	C11-22	A12	1: A12 (61)	0
mtDNA	C11-24	T12	1: T12 (61)	0
mtDNA	C12-29	T12	1: T12 (61)	0
mtDNA	C13-34	A13	2: A13 (60), A14 (1)	0
mtDNA	C15-48	C10	2: C9 (1), C10 (60)	2

Table 3.1 lists the number of variations seen in cpDNA and mtDNA across different loci. There are higher number of variations in homopolymer regions as compared to non-homopolymer region across for both cpDNA and mtDNA. This is expected as homopolymers typically experience higher insertion-deletion mutation rates due to slip-strand mispairing during DNA replication. Moreover, more variation is observed in the cpDNA data than in the mtDNA data. A total of 27 homopolymer allele variants was observed cpDNA as compared to 12 homopolymer allele variants in mtDNA. Similarly, 16 SNP variant sites were observed in cpDNA as compared to six such sites for mtDNA. Most of the mtDNA variance, both homopolymer and SNA were unique to sample SL10.

Initial assessments of nucleotide diversity (π), which considers only base-substitution variants, and insertion-deletion (InDel) variation were performed for both cpDNA and mtDNA, using the software application DnaSP (Rozas, 2009). The analyses of the 61 samples included 2,903 chloroplast sites. π for cpDNA was calculated based on eleven segregating sites in the data, resulting in a value of 0.00057. The InDel analysis included 46 sites, resulting in an InDel diversity of 0.00156 for cpDNA. The DnaSP analysis of our 61 samples included 3,809 mitochondrial sites. The mtDNA π was calculated as 0.00002, based on two

segregating sites. The InDel analysis included 14 sites, resulting in an InDel diversity of 0.000017 for mtDNA.

Genetic diversity was further analyzed using nine different models of nucleotide substitution for π estimation in MEGA X (Kumar et. al. 2018); note that DnaSP offers only one model (Nei's original model). A high degree of consistency was observed across all models for both the cpDNA and mtDNA datasets, and the variation is even lower than chloroplast (Table 3.2).

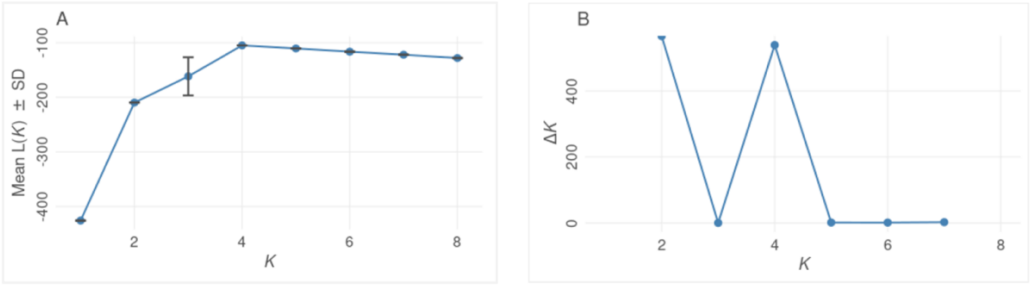
Table 3.2 : A comparison of π estimates in cpDNA and mtDNA, based on nine different nucleotide substitution models.

Model Name	**cpDNA**	**mtDNA**
Maximum Composite Likelihood	0.00074639	0.00005
p-distance	0.00074401	0.00005
Jukes-Cantor	0.00074505	0.00005
Kimura 2-parameter	0.0007452	0.00005
Tajima-Nei	0.0007457	0.00005
Tamura 3-parameter	0.0007453	0.00005
Tamura Nei	0.0007459	0.00005
LogDet (Tamura Kumar)	0.0008321	0.0001
No. of Differences	1.98852	0.1858

3.2 Analysis of Population-Genetic Structure in *F. religiosa* cpDNA:

Patterns of population-genetic structure among the *F. religiosa* samples studied here were analyzed using STRUCTURE (Raj, Anil et al. 2013) and other applications (see Methods). Bayesian maximum likelihood estimation (MLE) was performed for different values of K (number of clusters). For the cpDNA data, the MLE analysis was done for K ranging from 1 to 8. The respective likelihood estimations for each of these choices for the number of clusters (prior) are shown

Figure 3.1: Analysis of K for cpDNA. (A) shows how the likelihood of the cpDNA dataset for a given K changes for different numbers of K ranging from 1 to 8. Similarly, (B) shows how the rate of change of this likelihood for a certain K when K was increased by 1 for the same dataset. Here we can see that the log-likelihood increment is high when K increases from 1 to 2 and 3 to 4. The error bars show the standard error between values across 3 independent replicate runs.

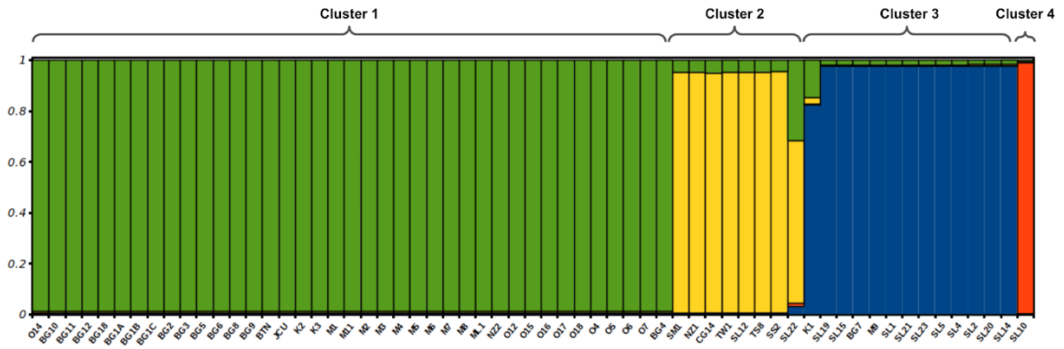


in Fig 3.1(A). Likelihood values increased with increasing K values analyzed, up until a peak value at K=4. The likelihood values then decreased more and more for levels of K beyond 4. Hence the optimal *K* for our analysis here was concluded to be four and was used for subsequent analyses.

Figure 3.1(B) shows the change of maximum likelihood values (ΔK) with increasing single increments of cluster size K. The sharp increase in the likelihood estimation for K 1->2 and 3->4 points to the higher likelihood of the cpDNA dataset for increasing cluster sizes. Here the dip of ΔK for 2->3 can be attributed to the large standard error of MLE for K =3 as the ΔK is calculated with the lower bounds of the MLE estimations. The MLE plateaus for K greater than 4, as depicted by small values of ΔK for K greater than 4. Hence the best conclusion is that 4 ancestral groups were best supported by the dataset.

Figure 3.2 is a stacked bar chart that visualizes the probabilities of samples belonging to each cluster across all 61 samples for a cluster size of K=4. Here the probabilities for each cluster are stacked and colored as per Table 3.3. The 61 samples plotted across Y-axis are sorted in descending order of their respective probabilities for Cluster 1 (the cluster labeling is also detailed in Table 3.3). The X-axis stacks the probability for each cluster and sums up to 1 for each individual -

Figure 3.2: STRUCTURE analysis for cpDNA of F. religiosa which shows the probability of each sample being classified to each of the four clusters.

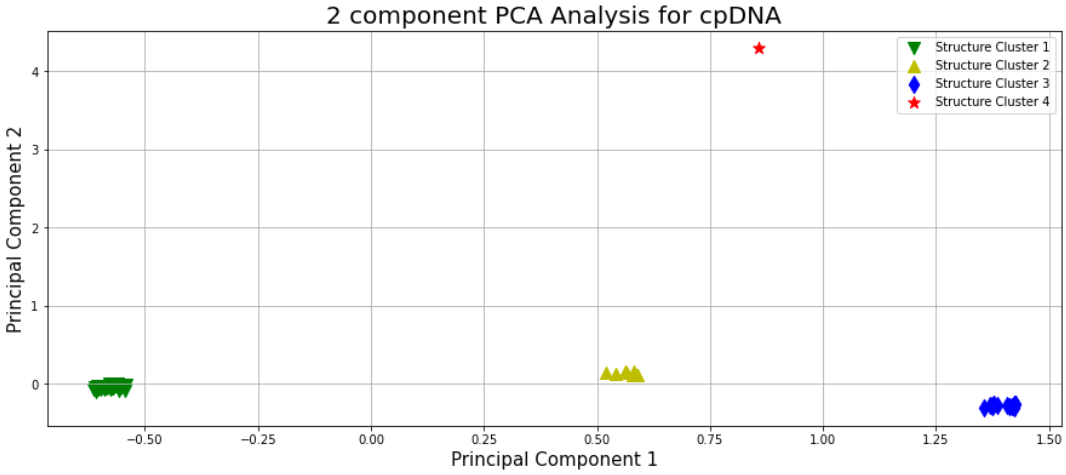


sample. All the samples with known historical descent from the Bodhgaya Bodhi tree and the Sri Maha Bodhi tree had near 1.0 probabilities for Cluster 1 (green). Most of the samples with the highest probability for membership in Cluster 2 (yellow) were from commercial sources. Most the samples with the highest probability for membership in Cluster 3 (blue) were from Sri Lanka. The lone representative showing the highest probability for membership in Cluster 4 (red) is SL10 from Sri Lanka.

Table 3.3: cpDNA cluster details. No. Samples indicates the number of *F. religiosa* samples analyzed that received the highest probability score for that cluster.

Cluster ID	**Cluster Color**	**No. Samples**	**Cluster Details**
Cluster 1	Green	39 / 61	Most samples from Bodhgaya and Hawai'i, one BTN, one New Zealand sample Sacred lineage: BG1A, BG1B, BG1C, and O15
Cluster 2	Yellow	8 / 61	primarily commercial samples except for SL22 and NZ1
Cluster 3	Blue	13 / 61	samples: Mostly Sri Lanka samples except for K1, M9, BG7, CG14, NZ1
Cluster 4	Red	1 / 61	SL10, from Sri Lanka

Figure 3.3: Two component Principal Component Analysis (PCA) plot for chloroplast genome (cpDNA)



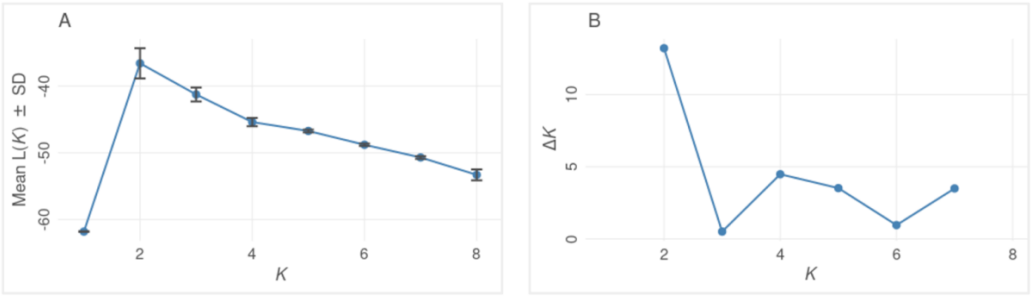
The Principal Component Analysis for cpDNA reveals four distinct clusters. Here, component 1 explains ~58% of the variance, whereas the explained variance for component 2 is ~28% for a total explained variance of ~86% for these top two components. There is strong correspondence between the clusters in the structure and in the PCA plot. Here Clusters 1,2,3, and 4 are color coded using the same details from Table 3.3, and patterns are mapped as delta, triangle, diamond, and star for Clusters 1,2,3, and 4, respectively.

3.3 Analysis of population-genetic structure in *F. religiosa* mtDNA:

Patterns in mtDNA population-genetic structure were analyzed among the 61 *F. religiosa* samples using STRUCTURE (Raj, Anil et al. 2013) and other applications (see Methods), following the same basic approach previously described for cpDNA.

Bayesian maximum likelihood estimation (MLE) was performed for different values of K (number of clusters) using STRUCTURE. The MLE analysis for mtDNA was done for K ranging from 1 to 8. The respective likelihood estimations for each of these choices for the number of clusters (prior) are shown in Fig 3.4(A). Here, the likelihood value for K reaches a peak at a value of two.

Figure 3.4: (A) shows how the likelihood of the mtDNA dataset for a given K changes for different numbers of K ranging from 1 to 8. Similarly, (B) shows how the rate of change of this likelihood for a certain K when K was increased by 1 for the same dataset. Here we can see that the log-likelihood increment is high when K increases from 1 to 2. The error bars show the standard error between values across 3 independent replicate runs.



The likelihood then progresses to a plateau for levels of K beyond four. Hence the optimal *K* for our analysis here is two, as it is the smallest number of clusters with the maximum likelihood estimation.

Figure 3.4(B) tracks ΔK, the change of maximum likelihood for a single increment for different numbers of cluster size K. Here, the ΔK plot begins with a high initial value for K 1->2. The MLE relatively plateaus for K transitions greater than two, as depicted by small values of ΔK for K greater than 2. Hence, two ancestral groups were best supported by the mtDNA dataset.

Figure 3.5: STRUCTURE analysis for mtDNA of *F. religiosa*

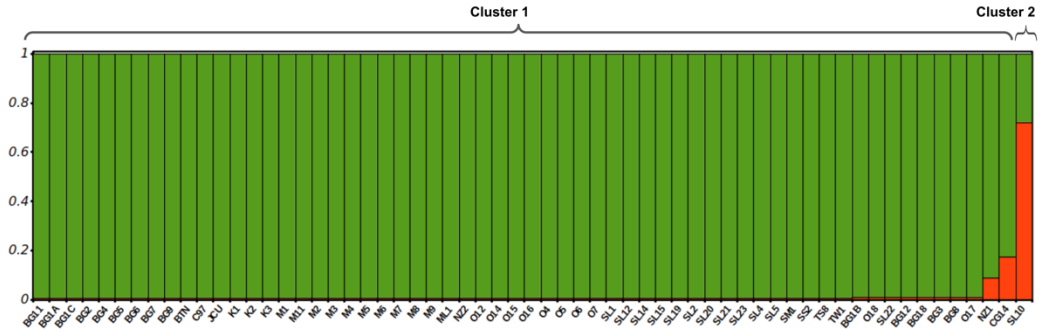
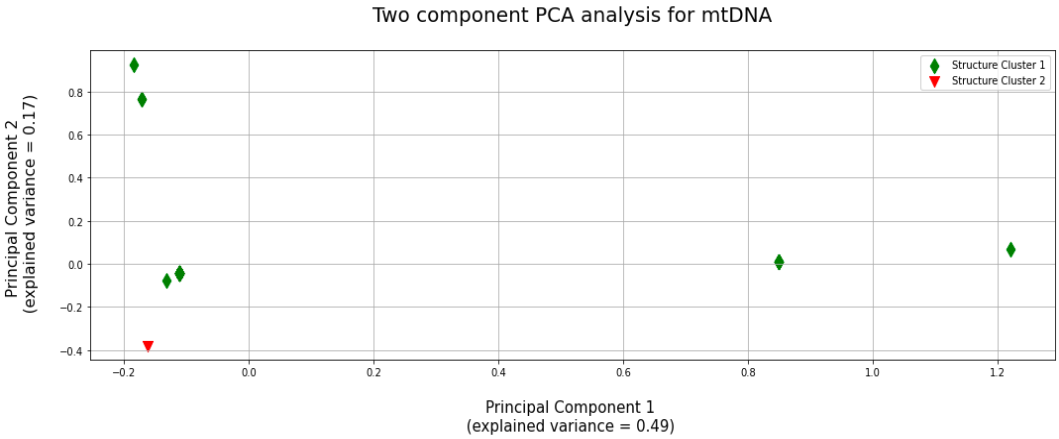


Figure 3.6: Two component Principal Component Analysis (PCA) for mtDNA.



In Figure 3.5, on the X axis, 61 samples from the mtDNA dataset are shown, and on the Y axis, the probability of those samples falling into the particular cluster is plotted. The cluster labeling is detailed in Table 3.3. The green cluster received

the highest probability scores for 60/61 samples analyzed. SL10 was the only sampled receiving a higher probability score for membership in Cluster 2.

As shown in Figure 3.6 the Principal Component Analysis for mtDNA does not reveal strong evidence for distinct clusters. Here explained variance for Component 1 is ~49%, whereas the explained variance for Component 2 is ~17% explaining a combined ~66% of the total variance Here Clusters 1 and 2 are color coded using the exact details from Table 3.3, and patterns are mapped as diamond and inverted triangles for Clusters 1 and 2, respectively.

Chapter 4: Discussion

This thesis analyzes *Ficus religiosa* cpDNA and mtDNA sequence across 61 samples representing diverse global locations. As shown in table 2.1, some samples have historical ancestry to one of two sacred trees (Bodhgaya Bodhi tree, Sri Maha Bodhi tree), others have accompanying apocryphal stores, and some have unknown ancestry. As observed in most other plant systems (Makarenko et al., 2021, Chevigny et al., 2019, Clegg et al., 1994) our results in *F. religiosa* show that the genetic diversity was higher in cpDNA as compared to the mtDNA by a factor of ~15 (table 3.2). Most variations were observed in the homopolymer regions as expected (table 3.1).

Population-genetic structure for cpDNA among the 61 *F. religiosa* samples revealed four prominent clusters as outlined in table 3.3. It is interesting to note that all the trees with strong historical evidence of ancestry from the Bodhgaya Bodhi tree or Sri Maha Bodhi tree had very high probabilities (near 1.0) of membership in Cluster 1, which suggests that a genetic signal is associated with the historical sacred Bodhi tree lineage. Out of the total 61 samples, 39 of them showed highest probability for membership in Cluster 1; most of them were from either Bodhgaya or Hawai'i, except for one sample from New Zealand (NZ2), one that was

commercially sourced (BTN). The BTN sample was accompanied by an apocryphal story of descent from the Bodhgaya Bodhi tree, said to have come to the Americas by way of migrant workers from India working on construction of the Panama Canal during the early 1900s. The cpDNA population-genetic data reported here supports this claim. Two other apocryphal stories of descent from the Bodhgaya Bodhi tree were supported by this analysis: the ML sample is said by members of the Mountain Lamp meditation community to be derived from this sacred tree, and the JCU (James Cook University, Australia) sample is said to be derived from this tree as well.

No samples with strong historical ancestry to the Bodhgaya Bodhi tree or Sri Maha Bodhi (D Krishnan, 2013) tree received the highest probability scores for the other three clusters. Samples receiving the highest probability scores for Cluster 2 mostly derived from commercial sources (6/8). The remaining two samples, NZ1 and SL22, were from two very different locations (New Zealand, Sri Lanka) and do not have any documented claims of ancestry to sacred trees. Samples receiving the highest probability scores for Cluster 3 were mostly from Sri Lanka. There were 13 such samples, out of which only three (K1, BG7, M9) were not collected from

Sri Lanka. Finally, a single sample, SL10, was the sole tree with the highest probability score for Cluster 4.

The level of genetic variation in *F. religiosa* mtDNA was much lower than in cpDNA, limiting ability to infer patterns of population-genetic structure from this data. The mtDNA analysis supports only two major clusters, with 60/61 samples receiving highest probability scores for Cluster 1 and only 1/61 samples (SL10) receiving the highest score for Cluster 2. While very limited variation is seen among the 61 *F. religiosa* samples for the mtDNA cluster analysis, the observation that SL10 stands out is in alignment with the cpDNA genetic structure where this sample was also an outlier.

Observations in the cpDNA cluster analysis provide evidence for a genetic signal associated with the sacred Bodhi tree lineage, represented by sample O15 which is known to derive from Sri Maha Bodhi, which in turn is known to derive from the original Bodhi tree under which the Buddha sat during his enlightenment ~2,500 years ago. The presence of the BG1 samples in the same cluster as the O15 also supports the hypothesis that Sri Maha Bodhi tree and the Bodhgaya Bodhi tree share genetic ancestry. This observation supports the hypothesis that the tree presently at the seat of enlightenment in Bodhgaya, planted by Alexander

Cunningham in the late 1800s, also shares ancestry with the original Bodhi tree. The cpDNA results from STRUCTURE were further supported by the results from the PCA analysis (Fig. 3.3). The mtDNA results, however, showed too little variation to derive meaningful insights.

Regarding future directions, more samples of *F. religiosa* should be included to paint a more comprehensive picture of genetic diversity in this species, and to evaluate a broader range of Bodhi tree apocryphal stories. Better representation of trees from the native range of *F. religiosa* (Southeast Asia, from Pakistan to Vietnam), and inclusion of members of other closely related species, would be especially valuable in discriminating among lineages. While the current study has been primarily regarding cpDNA and mtDNA, nuclear DNA analysis is missing and would add further knowledge into the relative levels of genetic diversity among these three genomes in *F. religiosa*. This system also offers opportunities to investigate a great diversity of evolutionary and biological phenomena, including fig-wasp co-evolution and ethnobiology. Moreover, the analysis may also be extended to samples collected from other species like “Bar” (*Ficus Benghalensi*s) as they are of great significance culturally to “peepal” (*F. religiosa*) (Chakraborty, Abhisek et al 2021, Murugesu, Suganya et al. 2021). “Bar”

is generally believed to be coupled with “peepal” and are generally planted together in many religious sites in different parts of Asia. Such a project done through the lens of ethnobiology might provide important knowledge on the broader bio-cultural relationships between humans and plants.

Bibliography

Agarwala, Richa et al. (2018). “Database resources of the National Center for Biotechnology Information”. In: *Nucleic Acids Research* 46, pp. D8 –D13.

Aslam, Irfan, and Naveed Akhtar (2021). “Development of Ficus religiosa extract-loaded emulsion system for topical application: Characterization and stability evaluation”. In: *Tropical Journal of Pharmaceutical Research.*

Beraldi-Campesi, Hugo (2013). “Early life on land and the first terrestrial ecosystems”. In: Ecological Processes 2, pp. 1–17.

BhavyasreeP, G and S XavierT (2021). “A critical green biosynthesis of novel CuO/C porous nanocomposite via the aqueous leaf extract of Ficus religiosa and their antimicrobial, antioxidant, and adsorption properties”.

Boratyn, Grzegorz M. et al. (2013). “BLAST: a more efficient report with usability improvements”. In: Nucleic Acids Research 41, W29 –W33.

Britannica, T. Editors of Encyclopedia (2019, June 27). Ficus. Encyclopedia Britannica. https://www.britannica.com/plant/Ficus.

Chakraborty, Abhisek et al. “Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa species reveal evolutionary mechanisms of longevity.” iScience 25 (2021): n. pag.

Chevigny, Nicolas et al. (2019). “RADA is a main branch migration factor in plant mitochondrial recombination and its defect leads to mtDNA instability and cell cycle arrest”. In: *bioRxiv.*

Clegg, Michael T. et al. (1994). “Rates and patterns of chloroplast DNA evolution.” In: Proceedings of the National Academy of Sciences of the United States of America 9115, pp. 6795–801.

Cock, Peter J. A. et al. (2009). “Biopython: freely available Python tools for

computational molecular biology and bioinformatics". In: Bioinformatics 25, pp. 1422–1423.

Flood, Pádraic J. et al. (2018). "Reciprocal cybrids reveal how organellar genomes affect plant phenotypes". In: Nature Plants 6, pp. 13–21.

Francis, Rob. "pophelper: an R package and web app to analyze and visualize population structure." Molecular Ecology Resources 17 (2017): n. pag.

Guo, Wenhu et al. (2016). "Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution." In: Molecular biology and evolution 33 6, pp. 1448–60.

Gutman, Benjamin L. and Krishna K. Niyogi (2009). "Evidence for Base Excision Repair of Oxidative DNA Damage in Chloroplasts of Arabidopsis thaliana*". In: The Journal of Biological Chemistry 284, pp. 17006 –17012.

Jayant, Kottapalli Kameswara and B. S. Vijayakumar (2021). "In-vitro anti-oxidant and anti-diabetic potential of endophytic fungi associated with Ficus religiosa". In: Johnson, Mark et al. (2008). "NCBI BLAST: a better web interface". In: *Nucleic Acids Research* 36, W5 –W9.

Karunanidhi, Priyanka et al. (2021). "Triphenylphosphonium functionalized Ficus religiosa L. extract loaded nanoparticles improve the mitochondrial function in oxidative stress induced diabetes." In: AAPS PharmSciTech 22 5, p. 158.

Kumar, Sudhir, Glen Stecher, and Koichiro Tamura (2016). "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." In: *Molecular biology and evolution* 33 7, pp. 1870–4.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution, 35 6, 1547-1549.

Letunic, Ivica and Peer Bork (2019). "Interactive Tree of Life (iTOL) v4: recent updates and new developments". In: *Nucleic Acids Research* 47, W256 –W259.

Librado, Pablo and Julio Rozas (2009). “DnaSP v5: a software for comprehensive analysis of DNA polymorphism data”. In: Bioinformatics 25 11, pp. 1451–2.

Lin, Shih-Yao et al. (2020). “Description of Azoarcus nasutitermitis sp. nov. and Azoarcus rhizosphaerae sp. nov., two nitrogen-fixing species isolated from termite nest and rhizosphere of Ficus religiosa”. In: *Antonie van Leeuwenhoek* 113, pp. 933 –946.

Makarenko, Maksim S. et al. (2021). “The Insights into Mitochondrial Genomes of Sunflowers”. In: *Plants* 10.

Moshi, A. Arul Marcel et al. (2020). “Characterization of surface-modified natural cellulosic fiber extracted from the root of Ficus religiosa tree.” In: *International journal of biological macromolecules.*

Murugesu, Suganya et al. “Phytochemistry, Pharmacological Properties, and Recent Applications of Ficus benghalensis and Ficus religiosa.” Plants 10 (2021): n. pag.

Olmstead, Richard G. and Jeffrey D. Palmer (1994). “Chloroplast DNA systematics: a review of methods and data analysis”. In: *American Journal of Botany* 81, pp. 1205–1224.

Padian, Kevin et al. (1992). “Terrestrial Ecosystems Through Time: Evolutionary Paleoecology of Terrestrial Plants and Animals”. In: *BioScience*.

Palmer, Jeffrey D. et al. “Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates.” Proceedings of the National Academy of Sciences of the United States of America 97 13 (2000): 6960-6 .

Pandit, Rucha, Ashish Phadke, and Aarti G. Jagtap (2010). “Antidiabetic effect of Ficus religiosa extract in streptozotocin-induced diabetic rats.” In: Journal of ethnopharmacology 128 2, pp. 462–6.

Parkinson, Christopher L. et al. (2005). “Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae”. In: BMC Evolutionary Biology 5, pp. 73 –73.

Pedregosa, Fabian et al. “Scikit-learn: Machine Learning in Python.” J. Mach. Learn. Res. 12 (2011): 2825-2830.

Peralta-Castro, Antolín et al. (2020). “Plant Organellar DNA Polymerases Evolved Multifunctionality through the Acquisition of Novel Amino Acid Insertions”. In: *Genes 11*.

QIU, Yin-Long et al. (2010). “Angiosperm phylogeny inferred from sequences of four mitochondrial genes”. In: Journal of Systematics and Evolution 48.6, pp. 391–425.

Raj, Anil et al. “Variational Inference of Population Structure in Large SNP Datasets.” bioRxiv (2013): n. pag.

Rozas, Julio (2009). “DNA sequence polymorphism analysis using DnaSP.” In: Methods in molecular biology 537, pp. 337–50.

Sagan, L. G. (1967). “On the origin of mitosing cells”. In: Journal of theoretical biology 143, pp. 255–74.

Saida, Lavudi, Cdsln Tulasi, and Mangamoori Lakshmi Narasu (2021). “Evaluation of chemo-preventive efficacy of Ficus religiosa latex extract by flow cytometry analysis and gene expression studies performed by RT-PCR in various cell lines”. In: *Future Journal of Pharmaceutical Sciences 7.*

Sakamoto, Wataru and Tsuneaki Takami (2018). “Chloroplast DNA Dynamics: Copy Number, Quality Control and Degradation.” In: *Plant & cell physiology 596,* pp. 1120–1127.

Saki, Mohammad, and Aishwarya Prakash (2017). “DNA damage related crosstalk between the nucleus and mitochondria.” In: *Free radical biology & medicine 107*, pp. 216–227.

Senthilkumar, Subramanian and Rakkiyasamy Manimekalai (2020). “Impact of bioorganic additive chlorophyll-b of Ficus religiosa, on optical, thermal, mechanical behaviors of KDP optical single crystals: A novel NLO material”. In: *Optik 204*, p. 164045.

Shankar, Preethi et al. (2021). “Ficus religiosa (Linn.) bark extract secondary metabolites bestow antioxidant property inducing cell cytotoxicity to human breast cancer cells, MDA-MB-231 by apoptosis involving apoptosis-related proteins, Bax, Bcl-2 and PARP”.

Singh, Damanpreet, Bikram Singh, and Rajesh Kumar Goel (2011). “Traditional uses, phytochemistry and pharmacology of Ficus religiosa: a review.” In: Journal of ethnopharmacology 134 3, pp. 565–83.

Sivalingam, D., Rajendran, R., & Anbarasan, K. (2016). Studies on DNA barcoding of sacred plant - Ficus religiosa L.

Sloan, Daniel B. et al. (2012). “Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates”. In: *PLoS Biology* 10.

Smith, David Roy and Patrick J. Keeling (2015). “Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes”. In: *Proceedings of the National Academy of Sciences 112*, pp. 10177 – 10184.

Strother, Paul K. et al. (2011). “Earth’s earliest non-marine eukaryotes”. In: Nature 473, pp. 505–509.

Tamura, Koichiro et al. (2013). “MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.” In: *Molecular biology and evolution* 3012, pp. 2725–9.

Untergasser, Andreas et al. (2012). “Primer3—new capabilities and interfaces”. In: Nucleic Acids Research 40, e115 –e115.

Varré, Jean-Stéphane et al. (2019). “Complete Sequence, Multichromosomal Architecture and Transcriptome Analysis of the Solanum tuberosum Mitochondrial Genome”. In: International Journal of Molecular Sciences 20.

Wang, Gang et al. (2021). “Genomic evidence of prevalent hybridization throughout the evolutionary history of the fig-wasp pollination mutualism.” In: Nature communications 12 1, p. 718.

Ye, Jian et al. (2012). “Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction”. In: BMC Bioinformatics 13, pp. 134 –134.

Zhu, Andan et al. (2014). “Unprecedented heterogeneity in the synonymous substitution rate within a plant genome.” In: Molecular biology and evolution 315, pp. 1228–36.

Appendices

Appendix A: Sequencing details

TABLE A.1: *F. religiosa* mtDNA homopolymers identified
Contig Id: Conjugate id formed after concatenating the contig id from reference genome and site-id from homopolymers identified. i.e., C1-1 refers to the site 1 which was found in contig 1 of reference genome.
Contig Site: Exact location of the homopolymer in the corresponding mtDNA contig.

Contig ID	Contig Site	Repeating Nucleotide	Homopolymer Length
C1-1	12190	A	10
C3-3	26602	T	10
C5-5	25882	T	12
C9-7	2527	A	12
C9-9	19257	A	15
C9-11	29945	T	10
C9-13	41859	T	10
C9-15	48233	A	12
C10-17	938	T	12
C10-19	22932	T	11
C11-21	3675	A	10
C11-23	5640	A	11
C12-25	1387	A	10
C12-27	25758	T	13
C12-29	34182	T	12
C13-31	9775	T	11
C13-33	26744	A	11
C13-35	35896	A	10
C13-37	69145	A	10
C13-39	75067	A	11
C14-41	20599	A	10
C14-43	27261	T	10
C14-45	38351	A	10
C14-47	46474	A	11
C3-2	22848	A	11
C5-4	749	T	13
C6-6	2764	A	14
C9-8	7554	A	10
C9-10	28481	A	11
C9-12	40928	T	11
C9-14	46272	T	11
C9-16	48613	T	10

Contig ID	Contig Site	Repeating Nucleotide	Homopolymer Length
C10-18	21904	T	15
C10-20	23366	T	12
C11-22	5061	A	12
C11-24	8930	T	12
C12-26	25480	A	10
C12-28	31248	A	11
C13-30	1751	T	10
C13-32	26535	A	14
C13-34	32757	A	13
C13-36	63935	T	10
C13-38	72576	A	10
C14-40	3249	A	10
C14-42	24162	T	10
C14-44	29830	T	10
C14-46	44234	A	11
C15-48	695	C	10

SOURCE CODE A.1: Finding Contig Lengths and %AT

```
1   import sys
2   from Bio import SeqIO
3   from Bio.SeqUtils import GC
4
5   # Read File
6   FastaFile=open("embplant_mt.K115.contigs.graph1.1.path_sequence.fasta", 'r')
7
9   # Calculate and print lengths and AT%
10  print("N.".ljust(4),"Name".ljust(100),"Length".ljust(10)," AT".ljust(10))%
11  for i, rec in enumerate (SeqIO.parse(FastaFile, 'fasta')):
12  name, seq = rec.id, rec.seq
13  base_counts = {base: rec.seq.count(base) for base in ['A', 'T', 'G', 'C']}
14  total=len(rec.seq)
15  print(str(i+1).ljust(4), rec.id.ljust(100),str(len(rec.seq)).ljust(10),
16  f"{100 - GC(rec.seq) :.3f}".ljust(10))
```

SOURCE CODE A.2: Finding Homo Polymers

```
1       import sys
2       from Bio import SeqIO
3       from Bio.SeqUtils import GC
4       from munch import Munch
5
7       # Read File
8       FastaFile=open("embplant_mt. K115.contigs.graph1.1.path_sequence.fasta",     'r')
9
10      # Setup for tracing homopolymers
11      i=0
12
13      for idx, rec in enumerate (SeqIO.parse(FastaFile, 'fasta')):
14      contig_name, seq=rec.id, rec.seq
15      curr_base=str (0)
16      homopolymer_objects=Munch ({curr_base: Munch({"site":0,"base":curr_base,"count":1})})
17
19      # Calculate Homo Polymers
20      for site_idx, base in enumerate(seq [1:]):
21      if base==homopolymer_objects[curr_base].base:
22          homopolymer_objects[curr_base]. count+=1
23      else:
24          homopolymer_objects[str(site_idx+1)] =Munch({"site":site_idx+1,
25          "base":base,"count":1})
26          curr_base=str(site_idx+1)
27
29      # Print Result
30      print("\n\n\n\n","C"+str(idx+1).ljust(4),",", f"Id:{contig_name}")
31      print("N.". ljust(4),",","Site". ljust(10),",","Base".ljust(10),",","BP #".ljust(10))
32      for v in homopolymer_objects.values():
33      if v.count >=10:
34      i=i+1
35      print(str(i).ljust(4),",",str(v.site).ljust(10),",",
36      str(v.base).ljust(10),",",str(v.count).ljust(10))
```

Appendix B: STRUCTURE analysis details

Table B.1 Homopolymer Variations in cpDNA sample.

DNA sample	Location	FrCp14723	FrCp60605	FrCp58543_5'	FrCp58543_3'	FrCp118538	FrCp82452	FrCp85060
BTN/BTN2	lab	11	8	10	10	16	12	14
SM1	lab	9	8		10	13	13	14
TS8	lab	9	8	11	10	13	13	13
CG14	lab	9	8	11	10	13	13	14
TW1	lab	9	8	11	10	13	13	13
JCU	Australia	11	8	10	10	15	12	13
ML1	Washington	11	8	10	10	18	12	14
SL1	Sri Lanka	12	9	11	11	16	11	14
SL2	Sri Lanka	12	9	11	11	16	11	14
SL4	Sri Lanka	12	9	11	11	16	11	14
SL5	Sri Lanka	12	9	11	11	16	11	14
SL6	Sri Lanka	0	0	0	0	16	0	0
SL7	Sri Lanka	0	0	0	0	0	0	0
SL10	Sri Lanka	15		9	9		11	13
SL11	Sri Lanka	0	0	0	0	0	0	0
SL12	Sri Lanka	9	8	11	10	14	13	14
SL13	Sri Lanka	0	0	0	0	0	0	0
SL14	Sri Lanka	12	9	11	11	16	11	13
SL15	Sri Lanka	12	9	11	11	16	11	13
SL19	Sri Lanka	12	9	11	11	16	11	13
SL20	Sri Lanka	12	9	11	11	16	11	13
SL21	Sri Lanka	12	9	11	11	16	11	13
K1	Kauai	11	9	12	11	16	12	13
K2	Kauai	11	8	10	10	16	12	13
K3	Kauai	11	8	10	10	16	12	13
O1	Oahu	0	8	0	0	0	0	0
O2	Oahu	0	0	0	0	0	0	0

DNA sample	Location	FrCp14723	FrCp60605	FrCp58543_5'	FrCp58543_3'	FrCp118538	FrCp82452	FrCp85060
O3	Oahu	0	0	0	0	0	0	0
O4	Oahu	11	8	10	10	16	12	14
O5	Oahu	12	8	10	10	16	12	14
O6	Oahu	12	8	10	10	16	12	14
O7	Oahu	12	8	10	10	16	12	14
O8	Oahu	11	0	0	0	0	0	0
O9	Oahu	11	0	0	0	0	0	0
O10	Oahu	0	8	0	0	0	0	0
O11	Oahu	0	8	0	0	0	0	0
O12	Oahu	11	8	10	10	15	12	14
O13	Oahu	11	0	0	0	0	0	0
O14	Oahu	11	8	10	10	15	12	14
O15	Oahu	12	8	10	10	16	12	14
O16	Oahu	12	8	10	10	16	12	14
O17	Oahu	12	8	10	10	16	12	14
O18	Oahu	12	8	10	10	16	12	14
M1	Maui	10	8	10	10	18	11	14
M2	Maui	12	8	10	10	16	12	14
M3	Maui	12	8	10	10	16	12	14
M4	Maui	12	8	10	10	16	12	14
M5	Maui	12	8	10	10	16	12	14
M6	Maui	12	8	10	10	16	12	14
M7	Maui	12	8	10	10	16	12	14
M8	Maui	11	8	10	10	16	12	14
M9	Maui	12	9	11	10	16	11	14
M11	Maui	11	8	10	10	16	12	14
SL22	Sri Lanka	10	8	11	10	13	10	12
SL23	Sri Lanka	12	9	11	11	16	11	14
BG3	Bodhgaya	11	8	10	10	16	12	14

DNA sample	Location	FrCp14723	FrCp60605	FrCp58543_5'	FrCp58543_3'	FrCp118538	FrCp82452	FrCp85060
BG4	Bodhgaya	11	8	10	10	16	12	14
BG8	Bodhgaya	11	8	10	10	16	12	14
BG12	Bodhgaya	11	8	10	10	16	12	14
BG18	Bodhgaya	11	8	10	10	16	12	14
SS2	lab	9	8	11	10	13	13	15
BG1A	Bodhgaya	11	8	10	10	16	12	14
BG1B	Bodhgaya	11	8	10	10	16	12	14
BG1C	Bodhgaya	11	8	10	10	16	12	14
BG2	Bodhgaya	11	8	10	10	16	12	14
BG5	Bodhgaya	11	8	10	10	16	12	14
BG6	Bodhgaya	11	8	10	10	16	12	14
BG7	Bodhgaya	12	9	10	11	16	11	14
BG9	Bodhgaya	11	8	10	10	16	12	14
BG10	Bodhgaya	11	8	10	10	16	12	14
BG11	Bodhgaya	11	8	10	10	16	12	14
NZ1	New Zealand	11	9	11	10	14	11	14
NZ2	New Zealand	11	8	10	10	16	12	14

Table B.2 STRUCTURE cluster analysis, maximum likelihood estimation probabilities for K = 2.

Id	Sample Name	Cluster 1 Probability	Cluster 2, Probability
1	BG11	0.994	0.006
2	BG12	0.991	0.009
3	BG18	0.991	0.009
4	BG1A	0.994	0.006
5	BG1B	0.993	0.007
6	BG1C	0.994	0.006
7	BG2	0.994	0.006

Id	Sample Name	Cluster 1 Probability	Cluster 2, Probability
8	BG3	0.991	0.009
9	BG4	0.994	0.006
10	BG5	0.994	0.006
11	BG6	0.994	0.006
12	BG7	0.994	0.006
13	BG8	0.991	0.009
14	BG9	0.994	0.006
15	BTN	0.994	0.006
16	C97	0.994	0.006
17	CG14	0.828	0.172
18	JCU	0.994	0.006
19	K1	0.994	0.006
20	K2	0.994	0.006
21	K3	0.994	0.006
22	M1	0.994	0.006
23	M11	0.994	0.006
24	M2	0.994	0.006
25	M3	0.994	0.006
26	M4	0.994	0.006
27	M5	0.994	0.006
28	M6	0.994	0.006
29	M7	0.994	0.006
30	M8	0.994	0.006
31	M9	0.994	0.006
32	ML1	0.994	0.006
33	NZ1	0.914	0.086
34	NZ2	0.994	0.006
35	O12	0.994	0.006
36	O14	0.994	0.006
37	O15	0.994	0.006
38	O16	0.994	0.006
39	O17	0.991	0.009

40	O18	0.993	0.007
41	O4	0.994	0.006
42	O5	0.994	0.006
43	O6	0.994	0.006
44	O7	0.994	0.006
45	SL1	0.994	0.006
46	SL10	0.279	0.721
47	SL12	0.994	0.006
48	SL14	0.994	0.006
49	SL15	0.994	0.006
50	SL19	0.994	0.006
51	SL2	0.994	0.006
52	SL20	0.994	0.006
53	SL21	0.994	0.006
54	SL22	0.993	0.007
55	SL23	0.994	0.006
56	SL4	0.994	0.006
57	SL5	0.994	0.006
58	SM1	0.994	0.006
59	SS2	0.994	0.006
60	TS8	0.994	0.006
61	TW1	0.994	0.006

Table B.3 STRUCTURE cluster analysis maximum likelihood probabilities for K = 4.

Id	**Sample Name**	**Cluster 1 probability**	**Cluster 2 Probability**	**Cluster 3, probability**	**Cluster 4 probability**
1	BG10	0.988	0.006	0.005	0.001
2	BG11	0.988	0.006	0.005	0.001
3	BG12	0.988	0.006	0.005	0.001
4	BG18	0.988	0.006	0.005	0.001
5	BG1A	0.988	0.006	0.005	0.001

Id	Sample Name	Cluster 1 probability	Cluster 2 Probability	Cluster 3, probability	Cluster 4 probability
6	BG1B	0.988	0.006	0.005	0.001
7	BG1C	0.988	0.006	0.005	0.001
8	BG2	0.988	0.006	0.005	0.001
9	BG3	0.988	0.006	0.005	0.001
10	BG4	0.989	0.006	0.005	0.001
11	BG5	0.988	0.006	0.005	0.001
12	BG6	0.988	0.006	0.005	0.001
13	BG7	0.018	0.004	0.977	0.001
14	BG8	0.988	0.006	0.005	0.001
15	BG9	0.988	0.006	0.005	0.001
16	BTN	0.988	0.006	0.005	0.001
17	CG14	0.05	0.943	0.006	0.001
18	JCU	0.988	0.006	0.005	0.001
19	K1	0.148	0.027	0.824	0.002
20	K2	0.988	0.006	0.005	0.001
21	K3	0.988	0.006	0.005	0.001
22	M1	0.988	0.006	0.005	0.001
23	M11	0.988	0.006	0.005	0.001
24	M2	0.988	0.006	0.005	0.001
25	M3	0.988	0.006	0.005	0.001
26	M4	0.988	0.006	0.005	0.001
27	M5	0.988	0.006	0.005	0.001
28	M6	0.988	0.006	0.005	0.001
29	M7	0.988	0.006	0.005	0.001
30	M8	0.988	0.006	0.005	0.001
31	M9	0.018	0.004	0.977	0.001
32	ML1	0.988	0.006	0.005	0.001
34	NZ1	0.047	0.947	0.005	0.001
35	NZ2	0.988	0.006	0.005	0.001
36	O12	0.988	0.006	0.005	0.001
37	O14	0.989	0.006	0.004	0.001

Id	Sample Name	Cluster 1 probability	Cluster 2 Probability	Cluster 3, probability	Cluster 4 probability
38	O15	0.988	0.006	0.005	0.001
39	O16	0.988	0.006	0.005	0.001
40	O17	0.988	0.006	0.005	0.001
41	O18	0.988	0.006	0.005	0.001
42	O4	0.988	0.006	0.005	0.001
43	O5	0.988	0.006	0.005	0.001
44	O6	0.988	0.006	0.005	0.001
45	O7	0.988	0.006	0.005	0.001
46	SL1	0.018	0.004	0.977	0.001
47	SL10	0.007	0.002	0.002	0.989
48	SL12	0.047	0.946	0.006	0.001
49	SL14	0.018	0.004	0.978	0.001
50	SL15	0.019	0.003	0.977	0.001
51	SL19	0.018	0.003	0.977	0.001
52	SL2	0.017	0.004	0.978	0.001
53	SL20	0.017	0.004	0.978	0.001
54	SL21	0.018	0.004	0.977	0.001
55	SL22	0.316	0.64	0.032	0.012
56	SL23	0.018	0.004	0.977	0.001
57	SL4	0.017	0.003	0.978	0.001
58	SL5	0.018	0.004	0.977	0.001
59	SM1	0.048	0.945	0.005	0.001
60	SS2	0.046	0.947	0.006	0.001
61	TS8	0.047	0.946	0.006	0.001
62	TW1	0.049	0.945	0.006	0.001