THE UNIVERSITY of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

The biogeographic affinities of the Sri Lankan flora

A thesis submitted for the degree of Doctor of Philosophy

Lakmini Darshika Kumarage

Institute of Molecular Plant Sciences, University of Edinburgh

Royal Botanic Gardens, Edinburgh

January 2016

Abstract

The island of Sri Lanka's exceptional biodiversity and enigmatic biogeography begs investigation, as the island is key in understanding the evolution of the Asian tropical flora. Since the Jurassic, Sri Lanka has been subjected to remarkable tectonic changes, thus its flora could have been influenced by that of a number of nearby landmasses, as well giving Sri Lanka the potential to have played a wider role in the assemblage of floras elsewhere. Firstly, as Sri Lanka originated as a fragment of the supercontinent Gondwana, part of its flora may contain Gondwanan relict lineages. There is also the potential for immigration from Laurasia after the Deccan Plate collided with it 45-50 Mya. Further, Sri Lanka may harbour floristic elements from nearby land masses such as Africa and Southeast Asia as a result of long distance dispersals, and *in situ* speciation has the potential to have played an important role in enhancing the endemic Sri Lankan flora.

I tested the relative contributions of the above hypotheses for the possible origins of the Sri Lankan flora using three representative families, Begoniaceae, Sapotaceae and Zingiberaceae. These families represent both herbaceous and woody elements, and have high diversity across the tropics. Dated molecular phylogenies were constructed for each family. I used recent analytical developments in geographic range evolution modelling and ancestral area reconstruction, incorporating a parameter *J* to test for founder event speciation. A fine scale area coding was used in order to obtain a better picture of the biogeography of continental Asia.

Amongst all the models compared, a dispersal-extinction cladogenesis model incorporating founder event speciation proved to be the best fit for the data for all three families. The dates of origin for Sri Lankan lineages considerably post-date the Gondwanan break up, instead suggesting a geologically more recent entry followed by diversification of endemics within the island. The majority of Sri Lankan lineages have an origin in the Sunda Shelf (53%). Persistence of warm temperate and perhumid climate conditions in southwestern Sri Lanka resembling those of Peninsular Malaysia and Sumatra could have facilitated suitable habitats for these massive dispersals from the Sunda Shelf region. Some trans-oceanic long distance dispersals from Africa (11%) are also evidenced, again these are too young to accept a hypothesis of dispersal during

the Deccan Plate's migration close to the African coast during the late Cretaceous, but occurred later during the Miocene. Further, some lineages of Laurasian origin (20%) are evidenced in the Zingiberaceae with ancestral areas of China and Indochina, which is congruent with a post collision invasion. Among the families tested, dispersals have occurred stochastically, one during the Eocene, six during the Oligocene, seven during the Miocene, two during the Pliocene and one during the Pleistocene. The highest number of dispersals occurred during the Miocene when a warm climate was prevailing during the Miocene thermal maximum.

My results confirm that in situ speciation is an important contributor to the Sri Lankan flora. More rapid radiation of endemics has occurred during Pliocene-Pleistocene; two endemics in Begoniaceae, ten endemics in Sapotaceae and ten endemics in Zingiberaceae have evolved *in situ* during this period. Sri Lanka will have been subjected to expansion and contraction of climatic and vegetation zones within the island during glacial and interglacial periods, potentially resulting in allopatric speciation. As a conclusion, long distance dispersals have played a prominent role in the evolution of the Sri Lankan flora. The young ages challenge the vicariant paradigm for the origin and current disjunct distributions of the world's tropical lineages and provide strong evidence for a youthful tropics at the species level. The thesis contains six chapters; first two are introductory chapters, then there are three analytical chapters, one for each family, and finally a summary chapter is provided. Each analytical chapter is written as a stand-alone scientific publication, thus there is some repetition of relevant content in each.

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, as any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Lakmini Darshika Kumarage

January, 2016

Acknowledgements

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction to Biogeography

1.1 What is Biogeography?

The Earth is a diverse planet inhabited by millions of species found in many different habitats such as forests, deserts, deep seas and even some extreme environments like thermal vents. Since the beginning of natural philosophy, we have tried to understand and explain these different distributions of organisms over the earth's surface and the possible causes for such distributions. These endeavours developed into a new science - Biogeography. Biogeography is the study of the distribution of organisms in space and time. Some organisms inhabit wide ranges all over the world and are termed cosmopolitan, while others are restricted to a specific area and are termed endemic. Biogeography is a vast field of biology that requires information from other disciplines like systematics, molecular biology, evolutionary biology, ecology and geology. Recent advances in the field have developed new analytical methods to understand the causes of differing distributions of species over the earth surface.

1.2 The early history of biogeography as a discipline

Aristotle was the first scientist to ask central questions about biology, and also the first to present a view of a dynamic and changing world. However the consensus view until the mid eighteenth century was that the earth was largely static, in combination with a belief in God's creation of living organisms.

Carolus Linnaeus (1707-1778) considered living organisms to be immutable, and his explanation of their varying distributions involved a centre of origin on a paradisical mountain after the biblical deluge. He accepted that species originated on the slopes of Mountain Ararat and they were adapted to live on various habitats at different elevations. Once the flood receded they were capable of migrating downward and spreading throughout the continents.

Georges-Louis Buffon's (1707-1788) idea of the geographic origin of life was in contrast to that of Linnaeus, hypothesized to be near the North Pole during a period when the climate was much warmer. When the environment cooled, species migrated south and the species capable of inhabiting newly accessible environments survived,

while others died. His observation that environmentally similar but isolated regions have distinct assemblages of mammals and birds became called Buffon's law (Buffon, 1761, 1776) that is accepted as the first principle of biogeography (Lomolino et al. 1998).

Sir Joseph Banks (1743-1820) was a naturalist who accompanied Captain James Cook on his three year voyage around the world which permitted him to collect around 3600 plant specimens, including 1000 new to science. The discoveries by Banks confirmed Buffons Law of different regions having different biotas, but in contrast also found some cosmopolitan species showing wide distributions. His discoveries shed new light on the complexity of the living world (Lyte 1981).

Johann Reinhold Forster (1729-1798) came up with new ideas in the field of biogeography which lead to the application of Buffon's law to plants, birds and mammals not only in the tropics but also throughout the globe. He was the first scientist to publish systematic descriptions of the world's biotic regions based on plant assemblages and his explorations on patterns of species diversity are considered as the foundation of ecology. He observed that the species diversity decreases towards the poles from the equator and that the species diversity of an island is lower than the mainland and depends on available resources (Forster 1996).

The German botanist Karl Ludwig Willdenow (1765-1812) came up with contrasting ideas to the single-site origin of life. He proposed that there had been several refuges for plants on mountains, each with a distinct flora that eventually spread throughout the globe giving rise to new species communities (Willdenow 1811).

The 19th century is considered as a golden age for biogeography with an expansion of contributions from great naturalists and botanists. Alexander von Humboldt (1769-1859) is considered as the father of phytogeography, and one of the great naturalists of his time who had new views on floristic distributions. Through surveys conducted in the Andes he discovered distinct floristic elevational gradients that he named 'floristic belts' (von Humboldt 1805, 1808).

Later, Augustin P de Candolle (1778-1841) made an immense contribution to the field and made further explanations to Johann Reinhold Forster's views by incorporating

island age, volcanism, isolation and climate as factors determining an island's biodiversity. His view of species competing for natural resources in the environment for their survival and coexistence can be considered as the basis for the theory of evolution and ecology (de Candolle 1820, 1855).

Charles Lyell (1797-1875) was a great 19th paleobotanist who used the fossil record to draw inferences regarding ancient climates. His invaluable findings on the nature of earth's surface led him to be regarded as the father of geology. He was amongst the first to propose that the earth was ancient, and his findings revealed that the earth's surface and biotic component is not a static system, but dynamic (Lyell 1830).

The scene was set in the latter half of the 19th century for contributions from scientists such as Charles Darwin, Alfred Russel Wallace, Joseph Hooker and Philip Sclater to come up with novel interpretations of the living world and to provide the basis for a modern synthesis of biogeography.

1.3 A review of historical biogeographic methods

1.3.1 Paradigmshifts key to historical biogeographic analysis

Theory of evolution

Historically species were believed to be immutable and the continents and oceans were believed to be fixed and stable. Before biogeography could mature as a discipline, a modern, realistic view of geology and evolution was necessary.

The great naturalist Charles Darwin made an enormous contribution to the field of biogeography through his theory of evolution, based largely on observations made during his five year voyage on HMS Beagle. He reasoned that the occurrence of different forms of tortoises and finches on the different islands in the Galapagos archipelago were the result of geographic isolation and natural selection, leading him to propose the theory of evolution (Darwin 1859).

A British naturalist, Alfred Russel Wallace (1823-1913) came up independently with the theory of evolution through natural selection from the observations made during extensive field work done in Amazone river basin and Malay Archipelago (Wallace

1860, 1863, 1869, 1876). His ideas prompted Darwin to publish his ideas in "On the Origin of Species".

The laws of inheritance were synthesized by Gregor Johann Mendel (1822-1884) based on his experiments using pea plants, where he explained how traits could be passed from one generation to the other. Mendel's population genetic theory combined with Darwin's theory of natural selection lead to the modern synthesis of evolutionary theory, the current paradigm in evolutionary biology.

Plate tectonics

The true age of the earth and its tectonic history only became apparent and widely accepted during the latter half of the twentieth century. Sir Francis Bacon is accepted as the first person to have views of moving continents, but his ideas were never exposed for discussions (Lomolino et al. 1998).

The German scientist Alfred L. Wegener developed the theory of continental drift based on the observation of congruent patterns of coastlines and coal deposits across the Atlantic Ocean. According to his hypothesis, during the early history of earth, all the landmasses were connected to each other forming a super-continent called Pangaea that began to separate during Mesozoic. Wegener's hypothesis (Wegener 1912, 1966) matured into the theory of plate tectonics, further developed by Oliver (1968). The earth is now understood to be a complex, dynamic structure composed of continental crust, mantle and core. The forces generated from the intense heat produced beneath the earth's crust result in lateral plate movements and seafloor spreading. Plate tectonics revolutionised the field of biogeography through providing a potential mechanism for explaining disjunct distributions (Lomolino et al. 1998).

Phylogenetics

Phylogenetics is the representation of the relationships and history of evolution of organisms in the form of a bifurcating tree. The German biologist, Willi Hennig (1913-1976) is considered to be the founder of phylogenetic systematic and to begin with it was largely dependent on morphological characters (Hennig 1966). Cladistics revolutionaised phylogenetic analysis, where it builds on the fact that the members of a group who share a common evolutionary history, and are more "closely related" to

members of the same group than to other organisms. In cladistics, any group of organisms arerelated by descent from a common ancestor, and they exhibit a bifurcating pattern of cladogenesis. Common shared characters are called as Synplesiomorphic while changes in characters occurs in lineages results in shared derived characters so called Synapomorphies.

The advent of advanced technologies such as PCR and DNA sequencing allowed the development of molecular phylogenetics which revolutionised the field providing more strongly supported trees. Both nuclear and plastid DNA has been extensively utilized in the field of molecular phylogenetics to infer relationships between organisms at different taxonomic levels. Non coding region of the chloroplast genome are of interest since they evolve more rapidly than coding regions and mutations are accumulated exhibiting sufficient variation to be phylogenetically informative (Borsch & Quandt 2009). Chloroplast markers provide a useful tool at higher taxonomic levels, while internal transcribed spacer (ITS) regions of nuclear ribosomal DNA has been proved to be useful in inferring lower phylogenetic relationships such as interspecific variations due to its rapidly evolving nature (Kim et al. 1999; Baldwin et al. 1997) providing more phylogenetically informative characters. Further, ITS provide a better resolution for relatively young or rapidly radiated genera where chloroplast markers tend to fail to achieve resolution.

Phylogenetic trees constructed using sequence data can be either distance based or character based. Distance based methods utilize a distance matrix which is obtaining by calculating genetic distance between each pair of sequences. Character based methods are of three inference types, ie Maximum Parsimony (MP), Maximum Likelihood (ML) or Bayesian methods (BI). Both likelihood and Bayesian methods are model based while parsimony assumes all possible mutations are equally likely.

Parsimony has been widely used for phylogenetic studies due to its simplicity and the availability of software such as PAUP (Swofford 1993). The tree with minimum number of character state changes is considered the most parsimonious and the best hypothesis for the evolution of the data. When constructing deeper-time phylogenies, or data which presents a mixture of long and short branches, likelihood or Bayesian methods are preferable to avoid errors due to long branch attraction (Yang & Rannala

2012). In ML, a likelihood score is calculated under a substitution model and averaged over all possible states. ML calculations are computationally intensive (Guindon & Gascuel 2003; Yang & Rannala 2012), but software has been developed to cope with large datasets such as GARLI (Zwickl 2006) and RAxML (Stamatakis 2006).

In Bayesian inference, posterior probabilities are calculated depending on a chosen model of DNA sequence evolution. Due to calculation of posterior probability values, further analysis to calculate bootstrap values of support for monophyletic groups is not required, making the analysis more convenient (Yang & Rannala 2012).

Molecular dating

The mutations accumulated in organisms over time provide a useful tool in inferring divergence times with the use of fossil data, so-called Molecular dating. Molecular dating has been widely used in the field of biogeography to test biogeographic hypothesis and to determine the possible geological scenarios for plant distributions. The use of molecular clocks in phylogenetic studies of plants began in the 1990's (Renner 2005) and since then developments in theory and the availability of user friendly software has led to the publication of hundreds of plant phylogenetic studies dated using fossil data.

In order to date a phylogeny, the genetic distance between sequences or taxa is calculated and the substitution rate is calculated by dividing the genetic distance by the age of an appropriate fossil. Finally, the substitution rate is used to obtain the absolute ages of taxa. When a fossil record from the ingroup is unavailable for calibration, a fossil from a more distant outgroup is used in a first analysis to date a node and then secondary calibration is done to bridge them to obtain absolute ages of desired taxa (Renner 2005).

In an ideal case, a clocklike rate of molecular evolution is assumed, thus the substitution rate of all the branches of the phylogenetic tree assumes remains same (Rutschmann 2006; Drummond et al. 2006). However, in reality substitution rates vary over the branches because of varying generation times and mutation rates, resulting in departure from clock like evolution (Drummond et al. 2006).

Alternatively, divergence time is estimated by incorporating rate heterogeneity rather than trying to account for it, under the assumption of autocorrelation by estimating branch length information and modelling divergence times and rates (Rutschmann 2006). The penalised likelihood approach (PL) has been widely used, which assumes the substitution rates of two lineages either side of a node are auto correlated (Renner 2005; Rutschmann 2006). An alternative assumption is to use "local clocks", where different substitution rates are applied to different parts of the tree by assuming that the age of calibration nodes are known (Renner 2005). The Bayesian approach to autocorrelation uses MCMC procedure to obtain posterior distribution of rates and times (Rutschmann 2006).

Even though the relaxed molecular clock is useful in many cases, the need for a user-defined fully resolved tree topology in initially developed methods could be potentially disadvantageous since unresolved tree topologies could result in several equally plausible trees. Also due to alteration of posterior probabilities under relaxed clock assumption, the best tree of relaxed clock model might not be the same of strict molecular clock (Drummond et al. 2006). Errors in fossil ages, fossil misidentifications, sampling errors and tree topologies could result in erroneous divergence times. Use of large sequences, use of bootstrap values to calculate errors associated with branch lengths and assumption of strict clock model when data are approximately clock like could be used in order to obtain more reliable divergence time estimates (Renner 2005).

BEAST 1.7 (Drummond & Rambaut 2007) Bayesian Evolutionary Analysis by Sampling Trees is a widely used software which implements a Markov chain Monte Carlo (MCMC) algorithms for Bayesian phylogenetic inference to date phylogenies (Drummond et al. 2012). BEAUti is the graphical user interface for BEAST which allows beast to generate dated phylogenies under given models of sequence evolution and fossil calibrations (Drummond & Ho 2007). BEAST simultaneously generates both phylogeny and molecular dates, hence allowing each to influence the other during the analysis.

1.3.2 The beginnings of a modern synthesis of analytical historical biogeography

Dispersalism

Dispersalism was considered for a long time as the major driving force for disjunctions, which states that higher taxa originated at a centre of origin; organisms then disperse and colonize new areas depending on their capabilities for survival and the physical conditions of the new environment, with speciation resulting in new species. Dispersal is always a possibility and strong evidence for dispersal is normally only provided when a barrier is older than the disjunction (Morrone & Crisci 1995).

There are some striking examples of long dispersal events, such as the occurrence of*Andira, Drepanocarpus, Hernandia, Hymenaea, Sacoglottis* in both South America and Africa, which can only be explained by using dispersal since the lineages are not old enough to support vicariance (Givnish & Renner 2004). The floristic assembly of some lineages of families like Annonaceae, Bromeliaceae, Rapataeceae and Myristicaceae for example is also supported by long distance dispersal (Givnish & Renner 2004; Renner 2005).

Panbiogeography

Croizat (1952) prompted the concept of panbiogeography, a development from the ideas of Buffon, Candolle, Hooker and Wulff (Morrone & Crisci 1995). It challenged the dispersal paradigm and used vicariance to explain disjunct distributions prior to acceptance of plate tectonic theory. Vicariance is the breaking up of a once larger distribution by a barrier of some kind. Distinctly related organisms having similar disjunct distributions at present were believed to share connecting land areas in the past. Distribution maps were drawn and areas with closely related taxa were connected to obtain biogeographic homology using track analysis method (Morrone & Crisci 1995; Humphries & Parenti1999). Even though the distributions of immediate ancestors are obtained, the actual pathway of biotic migrations is not explained by a panbiogeography approach (Nelson 1973). The movements of continents and oceans were ignored when explaining the disjunctions and the direction of migrations could not be addressed. Phylogeny and systematic knowledge was not considered and all the organisms were

treated equally, irrespective to their dispersal capabilities and their divergence times (Lomolino et al. 1998).

Cladistic Biogeography

The combination of Croizats 'tracks', which are actually a network or graph of hypothetical historical area connections, and Hennigs new cladistic methodology of reconstructing evolution lead to the development of cladistic biogeography (Rosen et al. 1988). This was an improvement upon previous methods, and meant phylogenetic relationships between species could be analytically incorporated, resulting in a more rigorous way to explain disjunct distributions (Morrone & Crisci 1995; Lomolino et al. 1998; Humphries & Parenti 1999). A cladistic biogeographic analysis has two steps, firstly the terminal taxa in taxon cladogram are replaced by area component to construct taxon area cladograms. Several taxon area cladograms are then combined to produce general area cladograms by using methods such as component analysis, brooks parsimony analysis, three area statements or reconciled trees (Morrone & Crisci 1995). These were assumed to represent biogeographic relationships among taxa and regions thus reveal the possible ancestral areas for the taxa.

A drawback of cladistic biogeography is that incongruence between the area cladogram and the taxon cladogram is frequent due to dispersal, sympatric speciation and extinction. Even when multiple taxon cladograms have congruent topologies these may infact represent pseudo-congruence if splits occurred at different times (Donoghue & Moore 2003). All the taxa are treated equally irrespective to their age or dispersal ability (Humphries & Parenti 1999).

1.3.3   Approaches to analytical historical biogeography

The area of phylogenetic studies has been developed with the advent of new sophisticated parametric methods and software packages such as BEAST (Drummond & Rambaut 2007) which are used to date phylogenies and infer biogeographic scenarios. Ancestral area reconstruction methods are used to infer the hypothetical ancestral areas at nodes of phylogenies under maximum parsimony, maximum likelihood or Bayesian criteria, and sometimes in conjunction with a model of geographic range evolution. Age estimates are determined using relaxed molecular

clocks and the distribution history is being inferred in terms of vicariance, long distance dispersal, speciation and extinction events within each clade (Crisp et al. 2011).

Parsimony methods

Perhaps the most basic ancestral area reconstruction method is the generalized parsimony approach, by accepting the fewest number of character state changes across the tree, where the transformation of character states are weighted equally (Lamm & Redelings 2009). Geography is coded as presence or absence for each taxon and which minimizes the instances of range evolution across phylogeny is considered as most parsimonious reconstruction. In complicated analyses where frequent dispersal and extinction occur with complex distributions, parsimony reconstructions can result in many equally parsimonious reconstructions making interpretation difficult (Lamm & Redelings 2009). Drawbacks of parsimony based approaches such as lack of incorporation of branch length information and the lack of a geographic model, lead to the search for novel approaches.

Weighted ancestral area analysis (WAAA)

In weighted ancestral area analysis a range of taxa and their ancestral areas are calculated and more weight is given to the distribution of basal or ancestral lineages (Lamm & Redelings 2009; Hausdorf 1998). Areas are optimized on to a tree and the number of times that area is gained and lost in the tree is calculated separately. Weighted gain steps and weighted loss step values are computed and a probability index is calculated from the ratio of above two values. The area with a high probability index is considered more likely to be a part of an ancestral area than the other. Branch length information is not considered and the results are entirely dependent on choosing which sister clade is deemed ancestral, approach now seen as spurious (Lamm & Redelings 2009).

Dispersal-Vicariance analysis (DIVA)

Parsimony, event based, quantitative methods such as DIVA (Ronquist 1997) allowed scientists to investigate ancestral area relationships based on dispersal and extinction events (Ronquist 1997; Buerki et al. 2011). It is considered as one of the first model based historical biogeographic methods. Species distribution areas are identified and

cost of range evolution is given depending on dispersal or extinction, with vicariance given a zero cost. Branch length is not taken in to account (Clark et al. 2008) and optimal reconstruction is produced by a parsimony criterion that minimizes dispersal-extinction events (Ronquist 1997) and this can result in wider ancestral ranges (Kodandaramaiah 2010). Adding additional outgroups or restricting the number of possible ancestral ranges (Lamm & Redelings 2009) is used to overcome the problem, but this can result in discontinuous ancestral range reconstructions. Even in the absence of divergence times, general area relationships can be determined and prior knowledge of geological history of distributions are not needed. The rapid generation of results and user friendly software resulted in it being one of the most widely used methods in inferring ancestral area relationships.

Bayes DIVA

A weakness of DIVA was that a single tree topology was specified, with no account taken of other equally likely tree topologies. Bayes DIVA addresses phylogenetic uncertainty and improves upon DIVA by utilizing the posterior distribution of clades across multiple trees following Bayesian phylogenetic analysis. Ancestral area reconstructions are displayed as marginal distributions and the best reconstruction among all possible solutions is selected by integrating posterior distribution across all sampled trees (Nylander et al. 2008). This approach is implemented by the software package S-DIVA (Yu et al. 2010).

Dispersal Extinction Cladogensis

The parametric counterpart of DIVA, i.e. DEC model (Ree & Smith 2008) infers ancestral area reconstructions under a likelihood approach. It is a continuous time model, which treats local extinction and dispersal as stochastic processes and allows only single dispersal or extinction at any instant in time. This result in either range contraction or expansion (Ree & Sanmartin 2009; Lamm & Redelings 2009; Buerki et al. 2011) and for a given phylogeny, the likelihood of range inheritance of terminal taxa is calculated by integrating the rate of dispersal and extinction for each internal node.

It proceeds through a continuous time Markov process giving the probabilities of each ancestral range as the result (Lamm & Redelings 2009; Ree et al. 2005). Unlike DIVA,

the DEC model explains the ancestral range inheritance by daughter lineages with vicariance and dispersal both treated as equally important (Kodandaramaiah 2010). However, a mechanism of identical range inheritance when the ancestral range comprises multiple unit areas is not addressed (Lamm & Redelings 2009).

In contrast to DIVA, in DEC branch length information is incorporated and not weighting vicariance as the main factor in causing wide species ranges results in more realistic ancestral ranges. The ability for users to incorporate biological and physical parameters such as connections between areas at a given time and dispersal rates improves the flexibility and complexity (Clark et al. 2008; Kodandaramaiah 2010).

DEC is implemented in two software packages, Lagrange (Ree & Smith 2008) and RASP (Yu et al. 2011). A geography matrix coded as binary presence-absence values and a phylogenetic tree with branch lengths proportional to expected changes are taken as input. Stochastic dispersal and local extinction along branches are optimised resulting in possible range inheritance scenarios at nodes (Ree & Sanmartín 2009; Webb & Ree 2012).

Bayesian Binary MCMC (BBM)

BBM implemented in RASP utilizes a full hierarchical Bayesian approach for inferring ancestral states of a phylogeny. The code from the Bayesian phylogenetic reconstruction programme MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003) has been modified to perform geographic range reconstruction as a form of character evolution under one of four models (JC, JC+G, F81, F81+G) utilizing character state frequencies. Topological uncertainty and branch length are taken into account and the tree file and distribution file are taken as the input (Yu et al. 2011).

Figure 1.1 The processors assumed by different biogeographic analysis methods. Excerpted from Matzke 2013.

Simulated Historical Island Biogeography (SHIBA)

Simulated Historical Island Biogeography Analysis is a discrete temporal and spatial model based approach where a phylogeny is divided into time slicesand each time slice is integrated in a dynamic land area model (Webb & Ree 2012). Dispersal causes range expansion while extinction causes range contraction, with branching occurring within a time slice as a result of a speciation event. The ages of speciation events are incorporated into the land area model and probabilities of ancestral ranges are calculated (Webb & Ree 2012). SHIBA utilizes inputs as a distance matrix, a physical area matrix, a list of taxa, a chronogram and a presence absence matrix for taxa and has been used in few studies containing small number of taxa.

Bay Area

In the analysis where the number of area is large, a novel method, bay area is used where probablilities of biogeographic histories across the tree is calculated according to continuous time Markov Chain Monte Carlo process under Bayesian inference.

Resolution of geographic history is increased by increasing the number of areas which would reveals a better picture of events in earth's history (Landis et al. 2013) Furthermore, use of grid system when defining geographic area would help to infer the biogeographic history of an area which lacks well defined boundaries.

Recent advancement in biogeographic analysis allows likelihoods to be assigned to models of range evolution for a particular dataset. The package Biogeobears (Matzke 2014) allows choosing the best-fit model under a maximum likelihood framework by modeling transitions between discrete states (biogeographical ranges) along phylogenetic branches as a function of time. The DEC model calculates maximum likelihood ratios of ancestral states at speciation events, in a method similar to LAGRANGE (Ree et al. 2005). In addition to two free parameters; d (dispersal), e (extinction) included in the DEC model (Batalha-Filho et al. 2014; Landis et al. 2013) the new additional parameter *J* is added to model. The *J* parameter tests the contribution of founder event speciation to range expansion and controls the probability of two events during cladogenesis; founder event speciation versus sympatric and vicariant speciation (Matzke 2014).

Different ancestral area reconstruction methods and characteristics are summarized in the table 1.1.

Table 1.1 Characters of different ancestral area reconstruction methods.

Name	Optimality criterion	Method	Multiple trees	Incorpo-rates time	Software
DIVA	Parsimony approach	Cost based model	no	no	DIVA 1.1 (Ronquist 1997)
Bayes DIVA	Parsimony approach	Cost based model	yes	no	S-DIVA (Yu et al. 2010)
DEC model	Likelihood approach	Continuous-time model	No	Yes	Lagrange (Ree & Smith 2008)
SHIBA	Probability approach	Spatial-Temporal model	no	yes	SHIBA (Webb et al. 2012)
Bay area	Bayesian probabilistic approach	Spatial-Temporal model	no	yes	Bay area (Landis et al. 2013)

1.4 The current state of the field

Ancestral area reconstruction methods are a rapidly evolving field, addressing the most important questions in biogeography such as when, where and how species evolve. There has been a recent proliferation of methods with the development of various model-based approaches. There are advantages as well as disadvantages of each method over the others but incorporation of geological and biological parameters to biogeographic models is beginning to reveal a better understanding about the geographic history of biodiversity.

Due to the lack of testable hypotheses concerning long distance dispersal, and with the emergence of plate tectonic theory, vicariance was accepted as the major reason for disjunct plant distributions and long distance dispersal was considered as random noise thus taking less attention (Cowie & Holland 2006). However recent studies suggested that dispersal is a crucial factor in determining plant distributions due to the young ages of lineages which cannot be accepted as vicariance. For example, the highly endemic biodiversity in the Canary Islands provides no chances to integrate with a vicariance scenario since they never had continental connections, thus should be a result of long distance dispersal (Cowie & Holland 2006). Furthermore, there are wind current patterns, bird migratory routes, ocean currents in oceanic systems which support recurrent long distance dispersal patterns (Cowie & Holland 2006; Nathan 2006).

1.5 Summary

Biogeography has been often considered as a narrative addition to phylogenetic studies. However with the advent of new sophisticated models and accumulation of data from various other fields such as ecology, geology, molecular dating has provided opportunities to test biogeographic hypothesis and then to infer biogeographic scenarios. The selection of data in hypothesis testing is fundamental, and should be independent such as fossil data rather than geological data which could be subjective and biased (Crisp et al. 2011). More often vicariance hypotheses are testable due to the evidence provided, however more careful examination of independent evidences provide space to test dispersal hypotheses.

Selection of an appropriate model for hypothesis testing, accurate data and adequate taxon sampling are crucial factors in any analysis in order to obtain a more reliable picture of historical biogeographic scenarios (Webb & Ree 2012).

CHAPTER 2: Introduction to Sri Lankan biogeography

2.1 Current Geology and climate of Sri Lanka

The island of Sri Lanka is located south-east of the southernmost tip of peninsular India, with a total area of 65,610km^2 and known as Ceylon until 1972. It has a close proximity to India, separated by the shallow Palk Strait which is 32 km wide. It is positioned in the Indian Ocean, to the southwest of the Bay of Bengal, between latitude 5° and 10°N, and longitude 79° and 82°E.

Geologically Sri Lanka is overlain with younger rocks in places of metamorphic rocks of Precambrian origin. These rocks form three major lithotectonic units named the Vijayan Complex, the Highland Complex and the Vanni Complex. The Highland Complex is the largest, forming the backbone and covering 90% of the island. It is composed of supra crustal rocks, charnockites, enderbites, metabasits, metagabbro, metadiorites and some orthogneisses and is bounded by the Vanni Complex to the west and by the Vijayan Complex to the east. The origins of the formation of the different geological units dates back to the Precambrian. Jurassic formations include shales, carbonaceous shales and arkosic sandstone that are 180 million years old and are present in a small area of the Western coast. Miocene limestones underlie the north-western part of the country and extend south in a relatively narrow belt along the west coast. The island also contains laterites, gravels, red earths and limestone of Pleistocene origin, whereas the Holocene formations include alluvial deposits, coastal sandstone, beach mineral sands, peat and gem gravels (Herath 1984).

The narrow limestone shoal is found beneath the Palk Strait, also locally known as Rama's or Adam's Bridge based on the legend that it was built by Rama to rescue his consort Sitha, as mentioned in the Ramayana. The presence of sand belts beneath the coral have been used to controversially support that belief since corals usually form on rocky substrates. This reflects the complex geological history of the reefs, and there is some evidence to suggest they result from local uplift and this suggests that these land masses were in full connection until the last sea level rise 6000 years ago (Ashton& Gunatilleke 1987; McLoughlin 2001).

A recent study by Dissanayake et al. (2000) suggested closer geological affinities of Sri Lanka with Madagascar than with the Southern and Eastern parts of India, due to the presence of high grade basement rocks in Sri Lanka that are more similar to those of South-eastern Madagascar. Furthermore, evidence of a distinct mineral belt running from Antarctica through the highland series of Sri Lanka into Madagascar, Tanzania, and further north suggests close similarities between these lands dating back to Gondwanan times. Corundum, spinel, topaz, zircon, aquamarine, amethys are some common gems found in Sri Lanka and Madagascar and the presence of gems in eastern Antarctica confirms past links that are further supported by graphite formations which are common to all these land masses (Dissanayake & Chandrajith 1999).
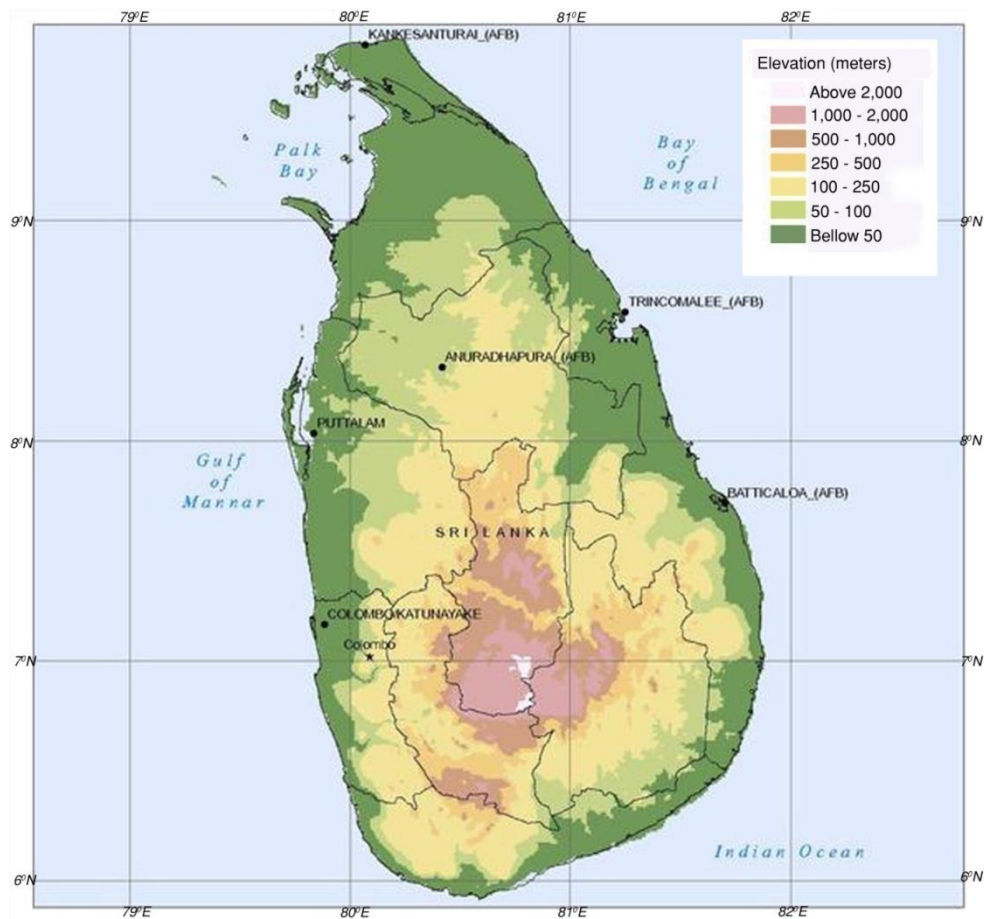
Figure 2.1 Geographic regions in Sri Lanka

The current climate of Sri Lanka can be described as tropical and warm. The average annual temperature for the country as a whole ranges from 28 to 30 °C. Diurnal

temperatures may vary by 4 to 7 °C. January is the coolest month, with the temperature dropping to 5°C (in the highlands) with May being the hottest month.

The island receives rain from a South-western monsoon from May to September and a North-eastern monsoon from December to February. South-western parts and windward slopes of the central highlands receive high rains during south western monsoons of around 2500mm and are classified as the wet zone. The North-eastern side receives high rains during December to February but the amount is not as much as in the South-western monsoons. Those areas that receive 1250mm-1750mm rainfall are known as the dry zone; this includes most of the North and North East of the island. However, North Western and South Eastern parts receive the lowest annual rain fall of less than 1250mm. An intermonsoonal period occurs from March until mid-May, with light, variable winds and evening thunder showers.

2.2 Palaeoclimate, geological changes

According to plate tectonics, the huge land mass of Gondwana began to breakup during the late Jurassic 180 million years ago (Figure 2.2). The Indian (or Deccan) Plate, composed of both India and Sri Lanka, rifted from East Africa 158-160 mya (Briggs 2003; Conti et al. 2002; Ashton & Gunatilleke 1987) along with the Mascarene plate comprised of Madagascar and the Seychelles Plateau. This block moved northward in close proximity to Africa allowing exchange of floristic elements between landmasses (Gunatilleke &Gunatilleke 1984; Ashton & Gunatilleke 1987).The split between the Madagascar and Indian plus Seychelles plates occurred about 95-85 million years ago (Ashton & Gunatilleke 1987; Conti et al. 2002).

About 65 million years ago the Seychelles Plateau separated from the Deccan Plate and remained fixed relative to Africa. Since then, the Deccan Plate drifted northward close to the African plate throughout the Cretaceous, with a 35° anti-clockwise rotation in the early Tertiary. The Deccan Plate moved north at about 20 centimetres (7.9 in) per year approaching equatorial latitudes during the Eocene(45 million years ago) and began colliding with Eurasia 55-65 Ma followed by a hard collision with the southern shore of Eurasia approximately 42-55 ma during the early Tertiary (Briggs 2003; Conti et al. 2002; Gunatillke & Gunatilleke 1984). It is believed that an aseasonal humid tropical

climate persisted on the Deccan Plate during its migration; this is supported by the presence of moribund laterites in the South-western area of the country, which can be formed only under wet seasonal climates (Ashton & Gunatilleke 1987). However, there is also evidence that the plate would have experienced aridification (Rutschmann & Eriksson 2004) as it moved north through 30-35 horse latitudess where patterns of atmospheric circulation result in very little precipitation and the formation of the desert belt (Lomolinoet al. 1998).
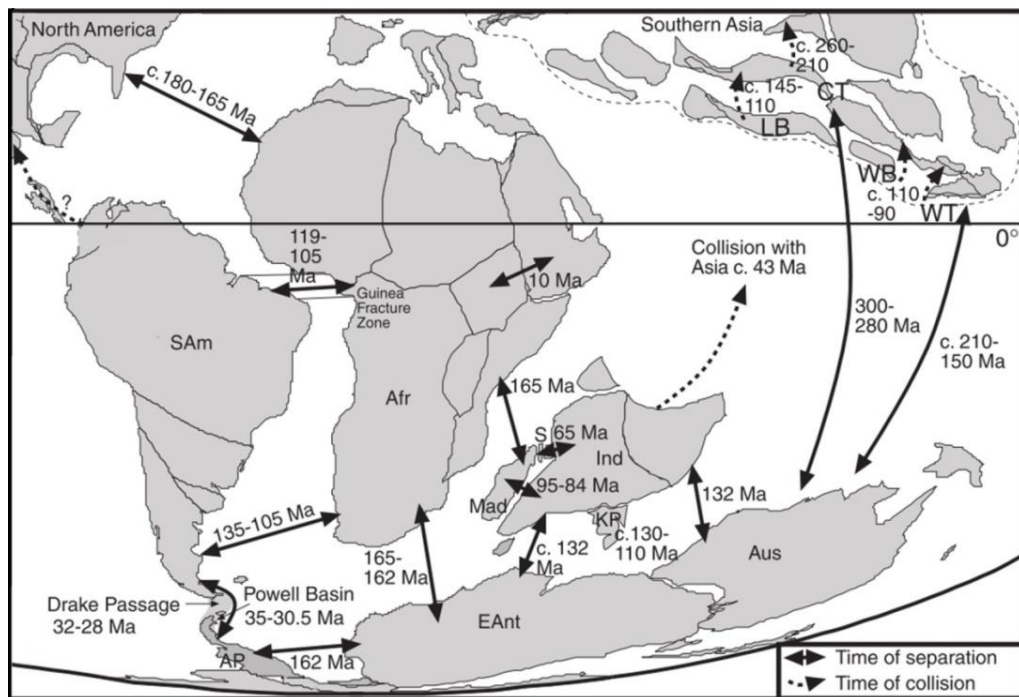
Figure 2.2 Break-up of Gondwana showing vicariance and collision times between continental fragments. Excerpted from McLoughlin 2001.

2.3 Flora of Sri Lanka

Sri Lanka has high floristic richness and species diversity, comprises 3900 species of land plants with the endemic elements heavily concentrated in the wet south-western quarter of the island. There are 2900 species of angiosperms on the island of which 830 are known to be endemic (Ashton & Gunatilleke 1987). Despite the fact that the flora of Sri Lanka has been extensively studied since the eighteenth century by various botanical experts, the biogeographic affinities of the flora remain poorly investigated (Ashton & Gunatilleke 1987).

Within the country eleven genera are known to be endemic, all of them showing restricted distributions in the wet zone including *Stemonoporus* (Dipterocarpaceae) 26 spp.and *Sonerila* (Melastomataceae) 22spp. An example of a more widespread genus that has high species endemism in Sri Lanka is *Memecylon* (Memecylaceae) which contains about 14 species distributed in lowland wet zone,eight species in the wet montane zone and eight species in the dry intermediate zone. Thirty of its species occur in Sri Lanka, of which 28 are considered to be endemics (Ashton & Gunatilleke 1987).

The forests of Sri Lanka can be classified into nine forest types:

1. Montane forest

2. Submontane forest

3. Lowland rain forest

4. Moist monsoon forest

5. Dry monsoon forest

6. Riverine dry forest

7. Mangroves

8. Sparse forest

9. Forest plantations (forest survey)

Among the forests distributed within the island, the lowland rain forest in the south western part of the country is considered very important, harbouring 90% of the endemics and high species diversity. The south western hills of Sri Lanka include most of the lowland and lower montane forests with biological significance as Sinharaja, Kanneliya, Nakiyadeniya, Dediyagala, Hiniduma etc. and the area is considered as the only aseasonal wet region in South Asia (Ashton & Gunatilleke 1987; Gunatilleke et al. 2005).

These forests are similar in floristic composition to the moist rain forests of the Western Ghats of Peninsular India suggesting past close affinities between the two land masses. These two may have acted as refugia for relicts of the Indian plate flora during changing climatic conditions. Due to their high endemicity and species diversity they have been declared as one of 25 biodiversity hotspots (Gunatilleke et al. 2005; Conti et al. 2002).

Several species are endemic to both the Agastyamalai-Nilgiri Hills of The Western Ghats and the Sri Lankan highlands, including *Abarema subcoriacea, Biophytum nudum, Chrysoglossum maculatum, Eugenia rotundata, Fahrenheitia zeylanica, Filicium decipens, Pavetta zeylanica,* and *Rubus micropetalus.*This high endemicity might be a result of the isolated migration of the Indian plate in the time between the Paleocene and Miocene epochs. This time period could have been sufficient for the evolution of new genera (Ashton & Gunatilleke 1987).

Another interesting feature is the absence of some Gondwanan plant and animal species in the Indian Peninsula despite their presence in Sri Lanka and Southeast Asia. Rapid latitudinal changes during the Late Cretaceous to Early Tertiary times caused climatic changes, resulting in the massive extinction of Gondwanan forms from India, but as an isolated island Sri Lanka was not affected as much (Rutschmann et al. 2004). Extensive volcanic eruptions 65 million years ago further impoverished the Indian biota and Sri Lanka may have acted as refuge harbouring Gondwanan Biota mainly in the South-western parts of the island (Gunatilleke et al. 2005; Conti et al. 2002).

One striking example of this is the presence of Gondwanan Crypteroniaceae such as *Axinandra zeylanica* endemic to Sri Lanka with a large number of con-generic species in Southeast Asia but none in India. This could be explained through extinction of the genus from India during the Tertiary (Conti et al. 2002; Rutschmann et al. 2004) according to phylogenetic evidence and molecular dating. Thus, Sri Lanka may have retained many of the Gondwanan elements that India lost (Karanth 2006).

There are not many angiosperm fossils in Sri Lanka, but fossils from India provide strong indications of past climatic history of these two landmasses. Late Cretaceous to early Tertiary fossils from India includes araucariaceous and podocarpaceous conifers and monocot families like Strelitziaceae, Zingiberaceae and Cyclanthaceae. Lakhanpal

(1970) identified 29 dicotyledon Cretaceous-early Tertiary fossils belonging to families including Clusiaceae, Tiliaceae, Combretaceae, Sonneratiaceae, and Anacardiaceae. Eocene deposits include macrofossils of Clusiaceae and Bombacaceae. This fossil content is similar to that of the Oligocene indicating the existence of a similar flora during those two time periods. Interestingly, lower Tertiary fossils include taxa from Juglandaceae and Myricaceae which might be a result of immigration from Laurasia (Lakhanpal 1970).

There are a few Oligocene deposits and fossils of Ebenaceae, Dipterocarpaceae, Ericaceae, Malvaceae etc. of the Miocene from Rajhastan supporting the presence of humid tropical conditions at that time in these land masses (Gunatilleke & Ashton 1987; Lakhanpal 1970).

Based on the fossil record Morley (2003) indicated that a substantial number of Indian sub-continental elements invaded Southeast Asia during the Oligocene when the Laurasian part of Southeast Asia and the Indian subcontinent were at similar latitudes and had similar climates. The distinct distribution of the genus *Cotylelobium* of Dipterocarpaceae in Sri Lanka and West Malesia argues for immigration from mainland Asia to Sri Lanka, or the reverse; the Indian subcontinent is therefore thought to be important in contributing to the biogeographic history of Southeast Asia (Ashton & Gunatilleke 1987).

2.4 Possible origins of Sri Lankan Flora

The Sri Lankan flora comprises elements that could have reached the island in several contrasting ways.

Firstly, as Sri Lanka and India originated as fragments of the southern supercontinent Gondwana, together with South America, Africa, Antarctica, Madagscar, New Guinea, Australia, New Zealand and New Caledonia. The breakup of the supercontinent Gondwana initiated during early Jurassic 180 million years ago (Mya), followed by subsequent continental drifts and sea floor spreading. The Deccan plate comprised of India and Sri Lanka was a part of Gondwana that initiated separation from Antarctica from 132 Mya followed by northward migration until collision with Asia c. 35 Mya. Being a part of Gondwana, part of Sri Lankan flora might be Gondwanan relicts.

However to be of Gondwana origin the disjunction should be reflected in phylogenetic splits occurring approximately 132 Mya-94 Mya concurrent with continental break-up. One good example is family Monimiaceae, which is represented by two species of *Hortonia* from Sri Lanka with one genus from West Africa and abundance in South America and Australia (Renner et al. 2010).

The chronogram in Figure 2.3 depicts times of lineage splits which would be concordant with Gondwanan vicariance.

Secondly, species could have arrived via the immigration of Laurasian lineages through Asia and India after the collision of the Deccan plate with Laurasia from c. 35 Mya. Thus, the species could have arrived via the immigration of Laurasian lineages through Asia and India which resulted in the mixing of the Deccan Gondwanan flora with the tropical flora occupying southern Laurasia. The collision of Deccan plate with southern Laurasia is accepted as the most marked possibility for interplate dispersal during the Tertiary (Morley 2000). However for this hypothesis to be accepted, we would expect that divergence times for lineages should not be older than 35 Mya and that Indo-Sri Lankan lineages would be nested within Laurasian ones.

Thirdly, long distance dispersal could have played a major role in the assemblage of the Sri Lanka flora from other land masses such as Africa and/or South East Asia. This could have occurred at any point in time, it is the only viable scenario for tropical disjunctions younger than ~33 Mya and Indo-Sri Lankan lineages would be nested within African and/or South East Asian ones.

Long distance dispersal has contributed to the sharing of floristic elements between tropical Asia and Africa, e.g. the appearance of *Begonia afromigrata* in Indochina which arrived from Africa during the Pliocene (de Wilde et al. 2011). There are other striking examples such as the occurrence of *Gleditsia* in South America with its closest relatives living in China. Also, long distance dispersal has been used to explain the colonization of some lineages within pantropical families such as Annonaceae, Myristicacae and Boraginaceae to their current distributions (Renner 2005).
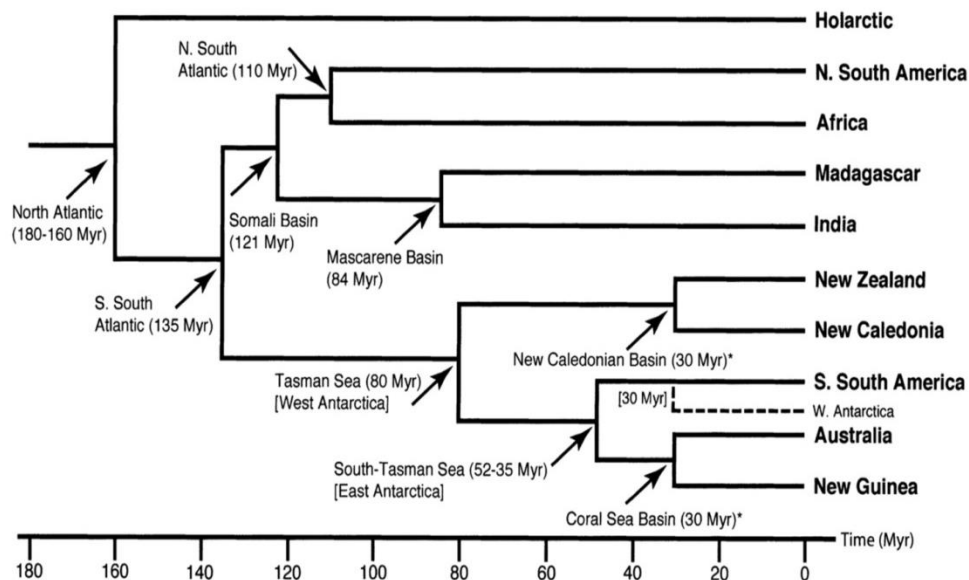
Figure 2.3 Geological area cladogram adapted from Sanmartin & Ronquist 2004 representing relationships between Gondwanan fragments and time of phylogenetic splits based on paleogeographic data. Vicariance is assumed to be at primary fragmentation. Asteriks (*) mark nodes, which are dated at 70-60 Ma in alternative reconstructions.

Another potentially important contributor to the Sri Lankan flora will have been diversification of lineages within the island (*in situ* speciation), resulting in the evolution of new species. Species that are endemic to Sri Lanka, and also have sister species in Sri Lanka, are likely to have arisen by *in situ* speciation, whereas endemics whose closest relatives occur elsewhere might be paleoendemics that once had wider distributions.

2.5 Study groups

Sri Lanka is a key location for understanding patterns of migration and biome assembly among tropical plants. To investigate its biogeographic affinities, study groups should be pantropical in distribution, well-sampled and thoroughly investigated in other regions. Sapotaceae, Begoniaceae and Zingiberaceae are ideal in this respect since all have dated phylogenies produced by previous studies (Bartish et al. 2011, Thomas 2012, Poulsen et al. unpublished) that are somewhat lacking in representatives from Sri Lanka.

2.5.1 FamilyBegoniaceae

Begoniaceae are a large pantropical family widely distributed in the tropics with greatest species diversity in Southeast Asia and South America. The largest genus

Begonia dominates the family with ca. 1800 species including 900 spp. in Asia with the bulk occurring in Southeast Asia (Hughes & Hollingsworth 2008, Rajbhandary et al. 2011). The genus *Begonia* has ten species present in Sri Lanka, of which two are known to be endemic (Dassanayake 1983), four are native and four are introduced. Within Sri Lanka, the two endemics are restricted to humid lowlands while the other native species are distributed in the hilly wet and intermediate zones.

Begonia has migrated from an origin in Africa during the Oligocene or late Eocene to China and Southeast Asia across India or dispersed via Asia (Thomas et al. 2012; Rajbhandary et al. 2011). The Sri Lankan species *Begonia malabarica* is sister to two species from the Socotran archipelago and with them forms the western limit of the Asian *Begonia* clade (Rajbhandary et al. 2011). It is suggested that continental Asia is the ancestral area for the Socotran *Begonia,* which resulted from long distance dispersal and diversifications of South Indian-Sri Lankan lineages to the Arabian-Socotran region during the Late Miocene. Thus, it would be interesting to determine the migration history of Sri Lankan *Begonia* species and their role in the diversification of the genus to the current hotspots with complete taxon sampling with the DNA regions used by Thomas et al. (2012).

2.5.2 Family Sapotaceae

Sapotaceae are a pantropical family of 53 genera and about 1100 species, mostly distributed in the tropical forests, but some extending to semi-arid and arid regions. The highest species richness is recorded in the tropical and subtropical regions of Asia and South America (Swenson & Anderberg 2005). In Sri Lanka, seven genera are native and all the nine species of *Palaquium*, three of *Isonandra* and four *Madhuca* are endemic (Dassanayake 1995). According to dated phylogenies, long distance transoceanic dispersal events from South America or Australasia across Atlantic or Indian oceans must have played a role in the diversification of Sapotaceae, rather than vicariance events such as the break-up of Gondwana (Bartish et al. 2011; Amstrong et al. 2014). Sri Lanka might have played an important role as a stepping stone in the movement of Sapotaceae between continents.

2.5.3 Family Zingiberaceae

Zingiberaceae, the largest family of Zingiberales, are pantropically distributed with a diverse set of genera in Africa, Southeast Asia and Pacific with only one neotropical genus *Renealmia* (Kress et al. 2005; Kress & Specht 2006). There are twelve genera found in Sri Lanka. Out of 36 species, 12 are known to be endemic although the status of some of the species is doubtful (Dassanayake 1983). The family is widely distributed in wet lowland and mid-montane primary forests but some of the endemic species are rare and have not been re-collected since the type gathering. In addition to understand the affinity of the Sri Lankan Zingiberaceae with those of other regions, molecular data could be used to investigate species limits within the family, and hence examine whether *in situ* speciation has occurred within Sri Lankan Zingiberaceae.

All species present in Sri Lanka belonging to each of the above families, their ecological status and distribution are summarized in the table 2.1.

Table 2.1 The species present in Sri Lanka for each study group, ecological status and their distribution

Family	Species	Ecological status	Distribution
Begoniaceae	*Begonia tenera*	Endemic	Sri Lanka
	Begonia thwaitesii	Endemic	Sri Lanka
	Begonia cordifolia	Native	Sri Lanka, India
	Begonia dipetala	Native	Sri Lanka, India
	Begonia malabarica	Native	Sri Lanka, India
Sapotaceae	*Palaquium pauciflorum*	Endemic	Sri Lanka
	Palaquium thwaitesii	Endemic	Sri Lanka
	Palaquium grande	Endemic	Sri Lanka
	Palaquium hinmolpedda	Endemic	Sri Lanka
	Palaquium zeylanicum	Endemic	Sri Lanka
	Palaquium rubiginosum	Endemic	Sri Lanka
	Palaquium petiolare	Endemic	Sri Lanka
	Palaquium laevifolium	Endemic	Sri Lanka
	Palaquium canaliculatum	Endemic	Sri Lanka
	Isonandra lanceolata	Native	Sri Lanka, India, Brunei, Borneo
	Isonandra compta	Endemic	Sri Lanka

	Isonandra montana	Endemic	Sri Lanka
	Isonandra zeylanica	Endemic	Sri Lanka
	Madhuca fulva	Endemic	Sri Lanka
	Madhuca moonii	Endemic	Sri Lanka
	Madhuca microphylla	Endemic	Sri Lanka
	Madhuca clavata	Endemic	Sri Lanka
	Madhuca neriifolia	Endemic	Sri Lanka
	Madhuca longifolia	Native	Sri Lanka, India
	Madhuca indica	Native	Sri Lanka, India
	Manilkara hexandra	Native	Sri Lanka, Thailand, Himalaya
	Mimusops elengi	Native	Sri Lanka, Malaysia. India, Thailand
	Xantolis tomentosa	Native	Sri Lanka, India, Burma
Zingiberaceae	*Zingiber cylindricum*	Endemic	Sri Lanka
	Zingiber wightianum	Native	Sri Lanka, India
	Amomum nemorale	Endemic	Sri Lanka
	Amomum trichostachyum	Endemic	Sri Lanka
	Amomum graminifolium	Endemic	Sri Lanka
	Amomum echinocarpum	Endemic	Sri Lanka
	Amomum hypoleucum	Endemic	Sri Lanka
	Amomum benthamianum	Endemic	Sri Lanka
	Amomum acuminatum	Endemic	Sri Lanka
	Amomum fulviceps	Native	Sri Lanka, India
	Amomum masticatorium	Native	Sri Lanka, India
	Amomum pterocarpum	Native	Sri Lanka, India
	Curcuma albiflora	Endemic	Sri Lanka
	Curcuma aromatica	Native	Sri Lanka, India
	Curcuma zedoaria	Native	Sri Lanka, India
	Curcuma oligantha	Native	Sri Lanka, India
	Alpinia fax	Endemic	Sri Lanka
	Alpinia rufecense	Endemic	Sri Lanka
	Alpinia abundiflora	Native	Sri Lanka, India
	Alpinia nigra	Native	Sri Lanka, Indo-Malaysia
	Alpinia malaccensis	Native	Sri Lanka, Malaysia
	Elettaria cardamomum	Native	Sri Lanka, India
	Hedichyum coronarium	Native	Sri Lanka, India, Malaysia

2.6 Distribution of Begoniaceae, Sapotaceae and Zingiberaceae in Sri Lanka
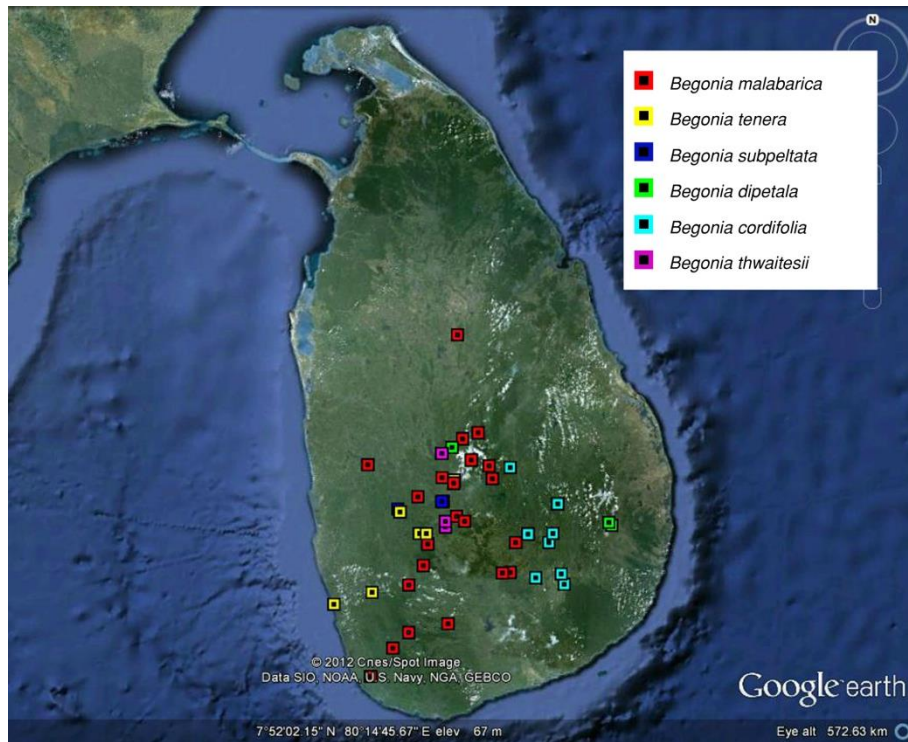
Figure 2.4 Species distribution of *Begonia* in Sri Lanka
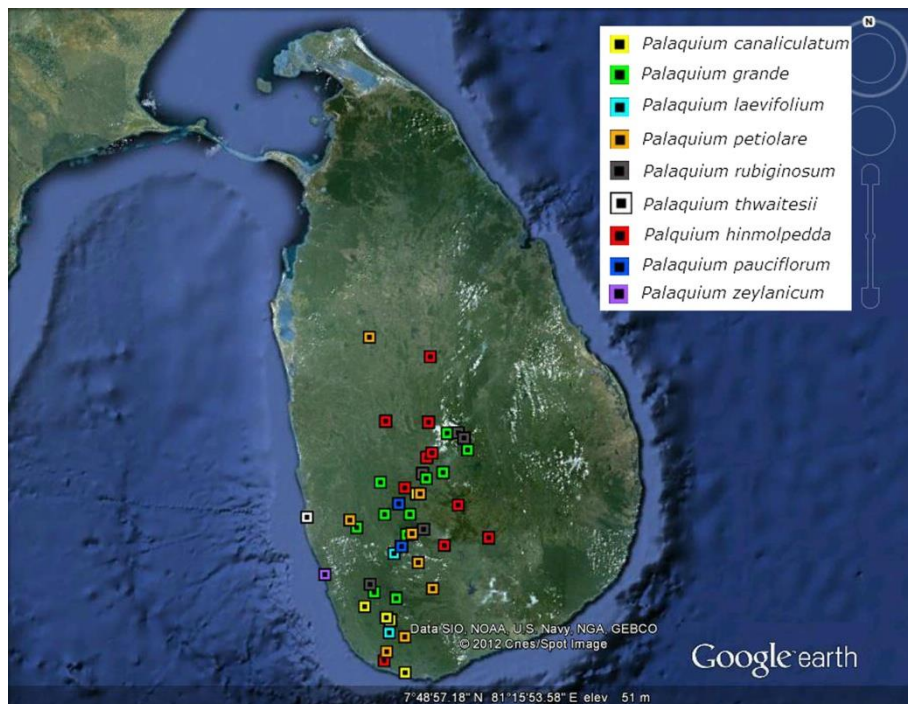
Figure 2.5 Distribution of *Palaquium* species in Sri Lanka
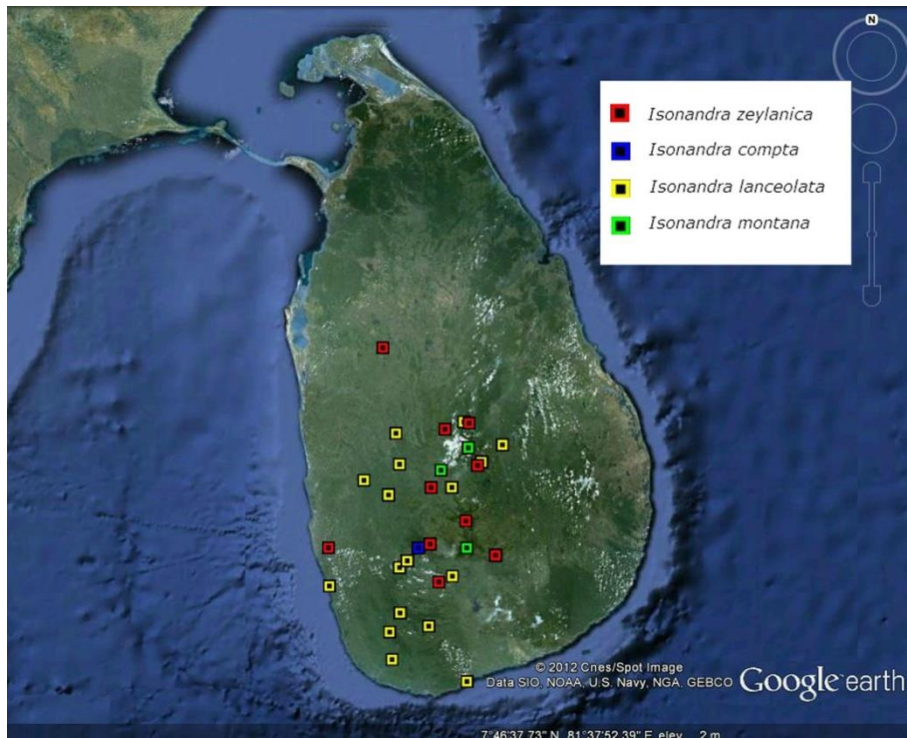
Figure 2.6 Distribution of *Isonandra* species in Sri Lanka
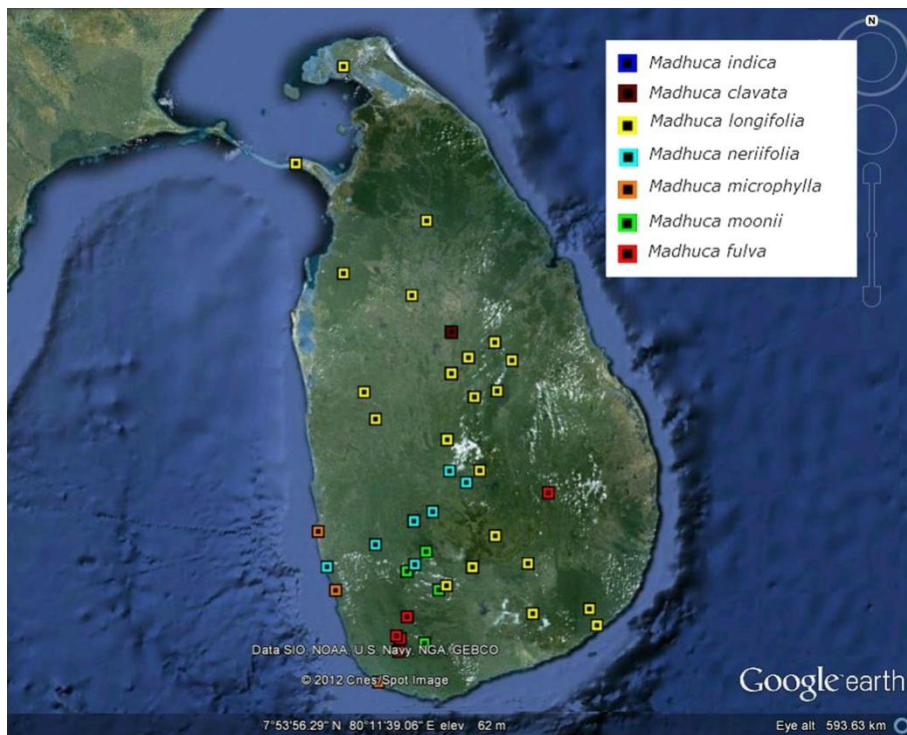
Figure 2.7 Distribution of *Madhuca* species in Sri Lanka
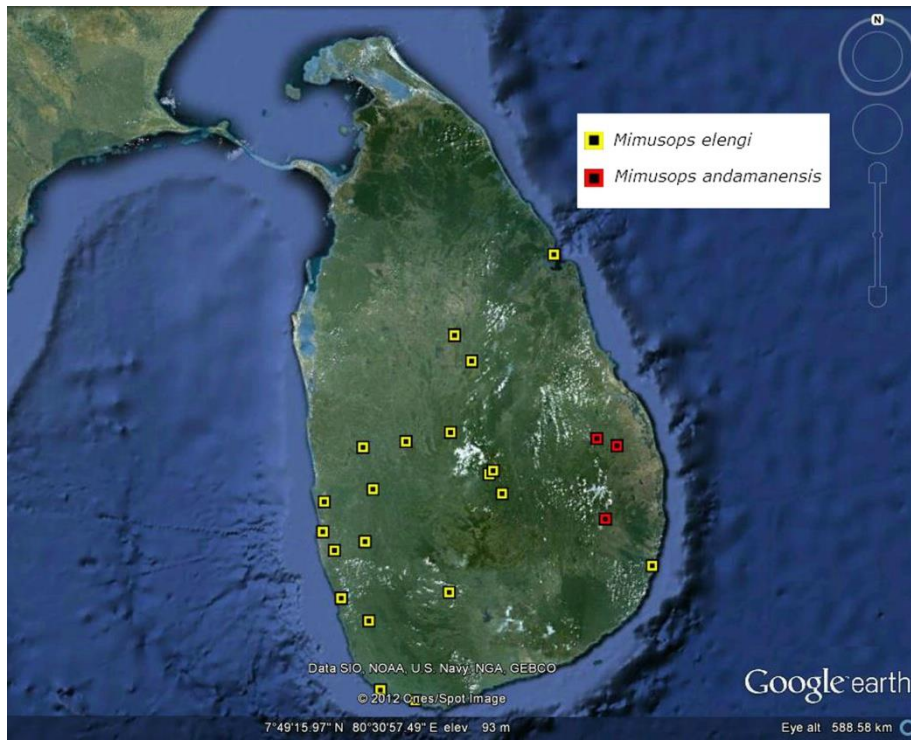
Figure 2.8 Distribution of *Mimusops* species in Sri Lanka
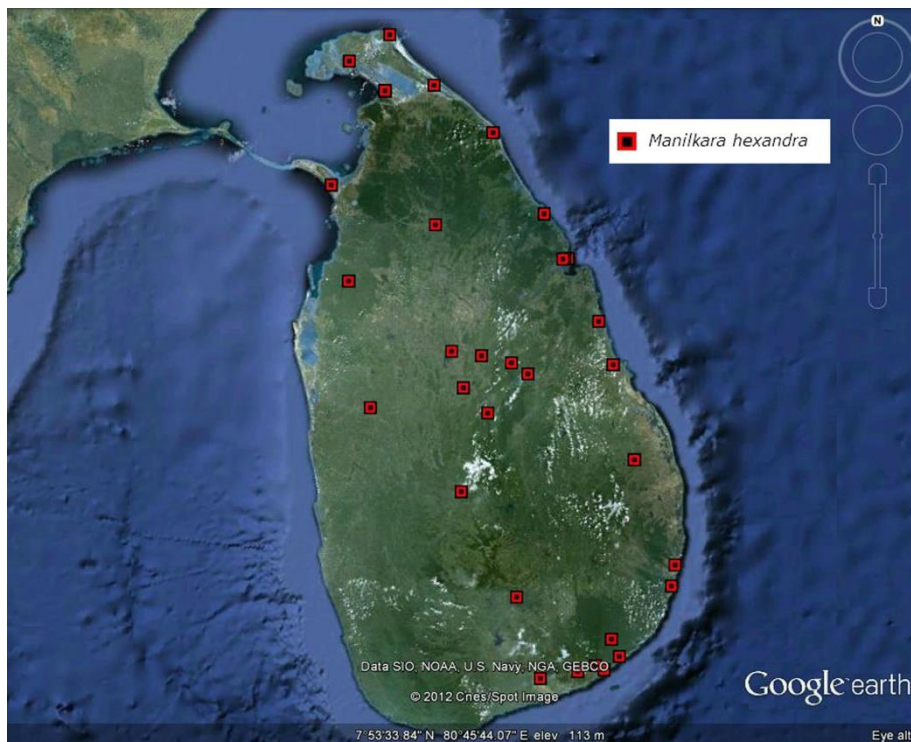
Figure 2.9 Distribution of *Manilkara hexandra* in Sri Lanka
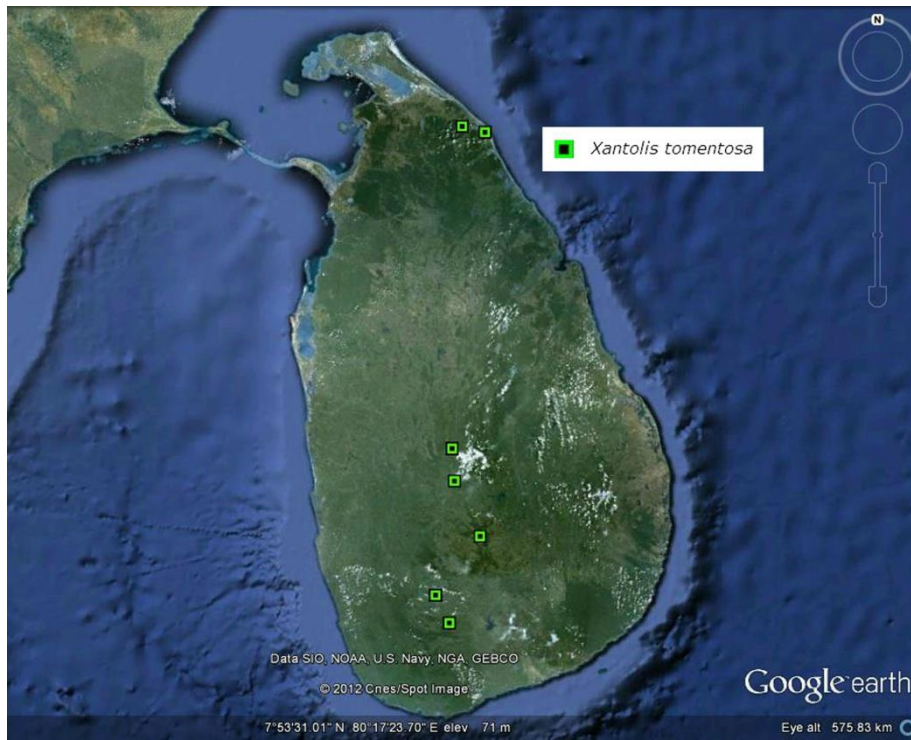
Figure 2.10 Distribution of *Xantolis tomentosa* in Sri Lanka
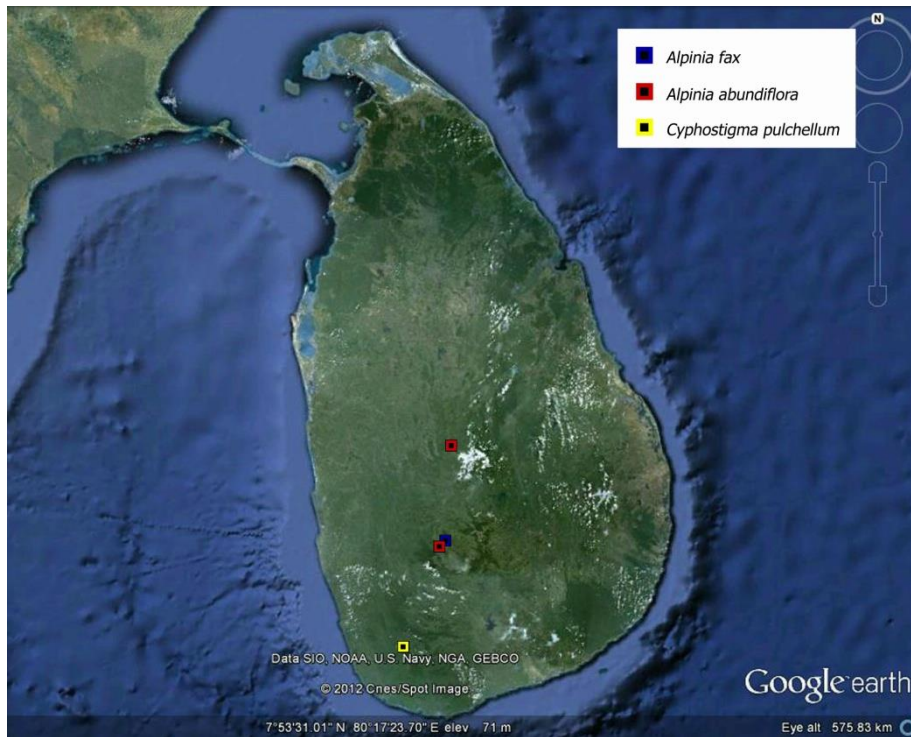
Figure 2.11 Distribution of *Alpnia fax, Alpinia abundiflora* and *Cyphostigma pulchellum* species in Sri Lanka
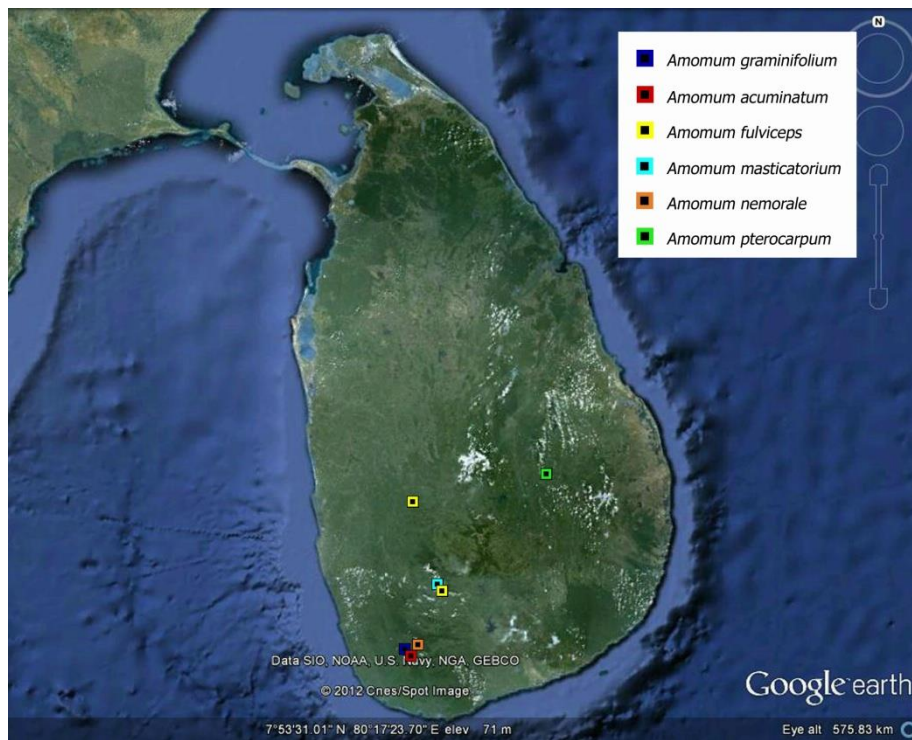
Figure 2.12 Distribution of *Amomum* species in Sri Lanka
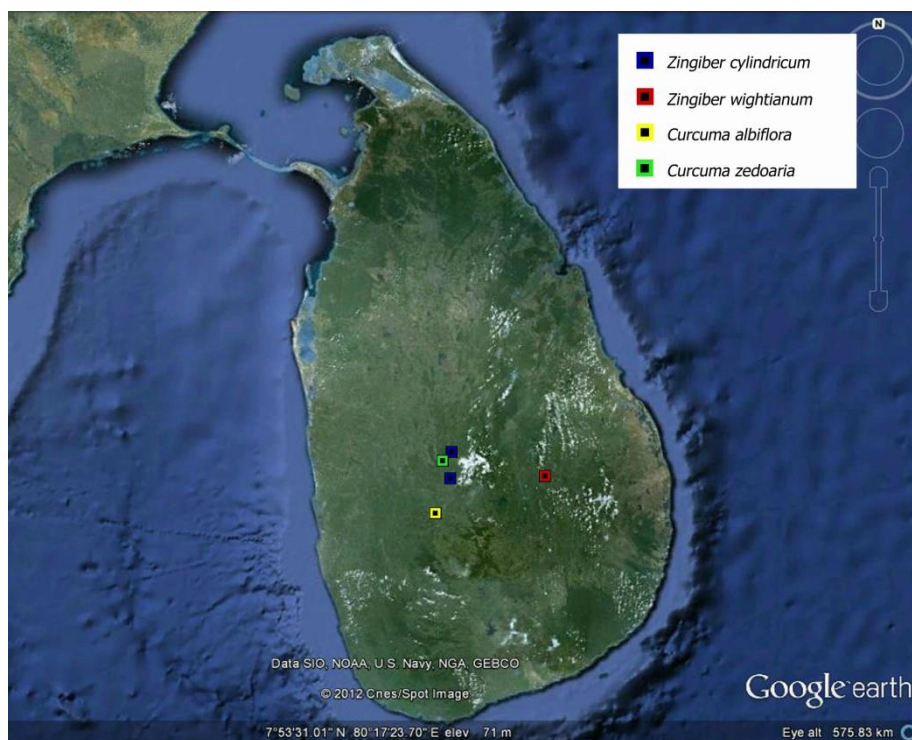
Figure 2.13 Distribution of *Curcuma* and *Zingiber* species in Sri Lanka

2.7 Field collections in Sri Lanka

Field collections were made at pre identified locations (list of locations is given below) defined based on the distribution of the species according to the flora of Ceylon and existing herbarium specimens in different herbaria. Transportation was made by a vehicle to the forests and inside the forests on foot. Field assistants were accompanied on all the excursions. For the DNA extractions leaf samples of each of the species were collected in silica. To prepare herbarium specimens a branch from each species was removed from the plant using secateurs and wrapped in newspaper. Then alcohol was added to prevent degradation of the specimens. All the information on the site and plant were recorded and the locations were recorded using a GPS. Plant habit, reproductive structures and any other important features were photographed. When collecting branches from trees tree climbers or poles were used.

All the local connections were made to get the collection permits and specimens were taken through proper channels to the Royal Botanic Gardens of Edinburgh, where I carried out my lab work. Silica dried specimens were used for DNA extractions and herbarium specimens were prepared and deposited at the Royal Botanic Gardens Edinburgh.
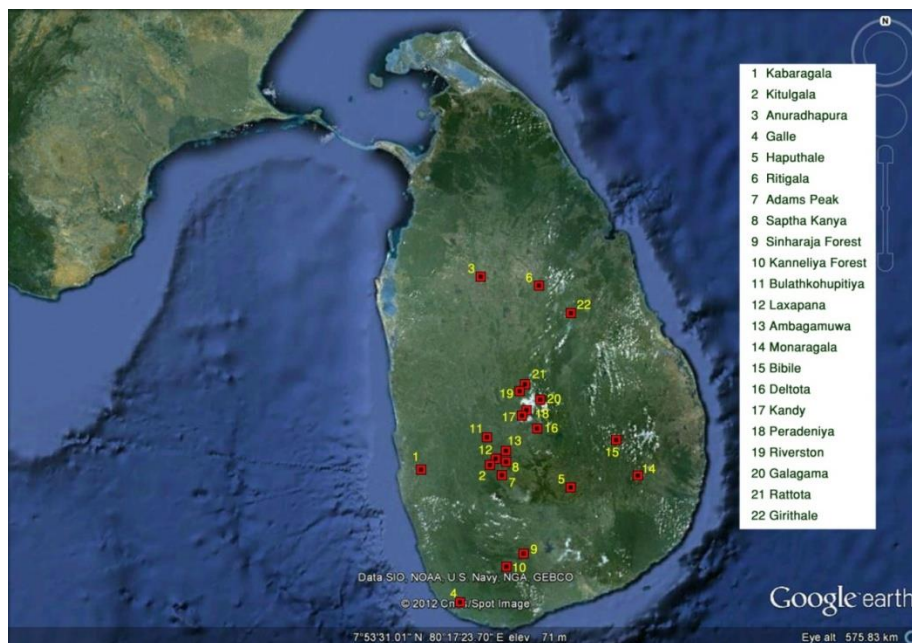
Figure 2.14 Localities of field collections in Sri Lanka

Table 2.2 Species collected in each family during the field work in Sri Lanka. Endemic species are indicated in bold.

Family	Genera	Species present	Species collected	% sampling
Begoniaceae	*Begonia*	***Begonia tenera*** ***Begonia thwaitesii*** *Begonia cordifolia* *Begonia dipetala* *Begonia malabarica* *Begonia subpeltata*	***Begonia tenera*** ***Begonia thwaitesii*** *Begonia cordifolia* *Begonia dipetala* *Begonia malabarica*	Endemic-100% Native- 75%
Sapotaceae	*Palaquium*	***Palaquium pauciflorum*** ***Palaquium thwaitesii*** ***Palaquium grande*** ***Palaquium hinmolpedda*** ***Palaquium zeylanicum*** ***Palaquium rubiginosum*** ***Palaquium petiolare*** ***Palaquium laevifolium*** ***Palaquium canaliculatum***	***Palaquium pauciflorum*** ***Palaquium thwaitesii*** ***Palaquium grande*** ***Palaquium hinmolpedda*** ***Palaquium zeylanicum*** ***Palaquium rubiginosum*** ***Palaquium petiolare*** ***Palaquium laevifolium*** ***Palaquium canaliculatum***	Endemic- 100%
	Isonandra	*Isonandra lanceolata* *Isonandra compta* ***Isonandra montana*** ***Isonandra zeylanica***	*Isonandra lanceolata* *Isonandra compta* ***Isonandra montana*** ***Isonandra zeylanica***	Endemic- 100% Native- 100%
	Madhuca	***Madhuca fulva*** ***Madhuca moonii*** ***Madhuca microphylla*** ***Madhuca clavata*** ***Madhuca neriifolia*** *Madhuca longifolia*	***Madhuca fulva*** ***Madhuca clavata*** ***Madhuca neriifolia*** *Madhuca longifolia*	Endemic- 60% Native- 100%
	Manilkara	*Manilkara hexandra*	*Manilkara hexandra*	Native- 100%
	Mimusops	*Mimusops elengi*	*Mimusops elengi*	Native – 100%
Zingiberaceae	*Zingiber*	***Zingiber cylindricum*** *Zingiber wightianum*	***Zingiber cylindricum*** *Zingiber wightianum*	Endemic- 100% Native - 100%
	Curcuma	***Curcuma albiflora*** *Curcuma aromatica* *Curcuma zedoaria* *Curcuma oligantha*	***Curcuma albiflora*** *Curcuma zedoaria* *Curcuma oligantha*	Endemic- 100% Native – 66%
	Ammomum	***Amomum nemorale*** ***Amomum graminifolium*** ***Amomum acuminatum*** ***Amomum echinocarpum*** *Amomum fulviceps* *Amomum masticatorium* *Amomum pterocarpum*	***Amomum nemorale*** ***Amomum graminifolium*** ***Amomum acuminatum*** *Amomum fulviceps* *Amomum masticatorium* *Amomum pterocarpum*	Endemic- 75% Native - 100%
	Alpinia	***Alpinia fax*** ***Alpinia abundiflora*** ***Alpinia rufescence*** *Alpinia nigra* *Alpinia malaccensis*	***Alpinia fax*** ***Alpinia abundiflora*** *Alpinia nigra* *Alpinia malaccensis*	Endemic - 66% Native - 100%
	Elettaria	*Elettaria cardamomum*	*Elettaria cardamomum*	Native 100%
	Hedichyum	*Hedichyum coronarium* *Hedichyum flavecens*	*Hedichyum coronarium*	Native 50%
	Cyphostigma	***Cyphostigma pulchellum***	***Cyphostigma pulchellum***	Endemic- 100%

2.8 Species portrait: Rediscovery of *Amomum nemorale* (endemic)

Many of the species collected for this project were found after not being seen for many decades. Some, such as *Begona thwaitesii*, were photographed for the first time. In particular, I highlight the rediscovery of *Amomum nemorale*, which was previously only known from type collection.

Amomum nemorale (Thw.) Trimen, Cat. 92. 1885; Trimen, Handb. Fl. Ceylon 4: 251. 1898; Baker in Hook. F., Fl. Br. Ind. 6: 233. 1892. Type: C.P. 3703 (PDA, K, BM)

*Elettaria nemoralis*Thw., Enum. Pl. Zeyl. 319. 1861.

Figure 2.15 Habit (left) and inflorescence (right) of *Amomum nemorale.*

"Leafy stem 60-120 cm; lamina shortly petiolate (c. 1 cm), 14-30 (?)x 3-5 cm, narrowly lanceolate, acuminate, glabrous, the margins callose-denticulate towards the tip; ligule short, rounded, entire, ciliate. Peduncle 5-7 cm, clothed with broad, loose scales. Inflorescence 2.5-3.5 cm long, subglobose to oblong. Bracts 0.7-1.5 cm long, orbicular or obovate, membranous, lightly pubescent towards the tip. Bracteoles lanceolate. Calyx 1.7 cm long, glabrous. Corolla greenish-white, tube 3.5 cm long; lobes up to 1.2 cm long, oblong, obtuse, subequal. Labellum veined with pink, reniform, 3- lobed, lateral lobes falcately recurved, modlobe smaller and rounded, 2-3 fid, Lateral staminodes? Anther c. 4 m, thecae glabrous, connective not prolonged into a crest. Ovary

pubescent.Capsule smooth, subglobose, c. 1.75 cm in diameter; seeds few, large, enclosed in a white spongy pulp" (Flora of Ceylon volume 4)

2.9 Aims of the doctoral research and structure of the thesis

Dated phylogenies are an important tool in historical biogeography because they provide a method for testing amongst alternative biogeographic hypotheses.

The objectives of this study are,

1. To incorporate Sri Lankan taxa into worldwide phylogenies for three families: Sapotaceae, Begoniaceae and Zingiberaceae.
2. To estimate the relative contributions to the Sri Lankan flora of Gondwanan relicts, immigration from nearby landmasses (Asia), long distance dispersal from Southeast Asia, and long distance dispersal from Africa.
3. To examine the contribution of *in-situ* speciation to biodiversity in Sri Lanka.
4. To investigate the contribution of Sri Lanka to the assembly of biotas elsewhere, by examining whether examples exist of:
    a. clades of Asian taxa whose most early diverging members occur in Sri Lanka, indicating a group whose ancestors rafted northwards with the subcontinent and then radiated outwards into Asia.
    b. long distance dispersal events out of Sri Lanka, as might be indicated for example by a clade of Southeast Asian species whose closest relative was Sri Lankan and whose next closest relatives were also Sri Lankan or occurred further west.

The thesis contains four empirical research chapters; The biogeography of Sri Lankan Begoniaceae (Chapter 3), The biogeography of Sri Lankan Sapotaceae (Chapter 4), The biogeography of Sri Lankan Zingiberaceae (Chapter 5) and the final chapter (Chapter 6) discussing main findings and conclusions. Each chapter is aimed at a scientific publication, thus there could be some content repetition.

References

Armstrong, K.E. et al., 2014. Patterns of diversification amongst tropical regions compared: a case study in Sapotaceae. *Frontiers in Genetics*, 5(December), pp.1–13.

Ashton, P.S. & Gunatilleke, C.V.S., 1987. New light on the plant geography of Ceylon I Historical plant geography. *Journal of Biogeography*, 14(3), pp.249–285.

Baldwin, B.G., 1992. Phylogenetic Utility of the Internal Transcribed Spacers of Nuclear Ribosomal DNA in Plants: An Example from the Compositae. *Molecular Phylogenetics and Evolution*, 1(1), pp. 3-16.

Batalha-Filho, H. et al., 2014. Phylogeny and historical biogeography of gnateaters (Passeriformes, Conopophagidae) in the South America forests. *Molecular Phylogenetics and Evolution*, 79, pp.422–432.

Bartish, I. V. et al., 2011. Vicariance or long-distance dispersal: Historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, 38(1), pp.177–190.

Borsch, T. & Quandt, D., 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution*, 282, pp.169–199.

Briggs, J.C., Collections, Z. & Museum, G., 2003. The biogeographic and tectonic history of India. *Journal of Biogeography*, 30, pp.381–388.

Buffon, G. L. L. & Compte De, 1761. Histoire naturelle generale. Paris: Imprimerie Royal.

Buffon, G. L. L. & Compte De, 1776. Servant de suite a l'histoire des animaux Quadrupedes (Supplement III) In: Histoire naturelle generale et particuliere. Paris: Imprimerie Royal.

Buerki S., et al,. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *J Biogeog*, 38, pp.531–550.

Clark, J.R. et al., 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Systematic biology*, 57(5), pp.693–707.

Cowie, R.H. & Holland, B.S., 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography*, 33, pp.193–198.

Conti, E. et al., 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution; international journal of organic evolution*, 56(10), pp.1931–1942.

Croizat, L., 1952. Manual of Phytogeography. The Hague: Dr. W. Junk.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray.

De Candolle, A. P., 1820. Essai elementaire de geographie botanique. De l'imprimerie de F.G. Levrault.

De Candolle, A. P., 1855. Geographie Botanique Raisonee (2 volumes). Paris: Masson.

De Wilde, J.J.F.E. et al., 2011. Pliocene intercontinental dispersal from africa to Southeast Asia highlighted by the new species begonia afromigrata (Begoniaceae). *Taxon*, 60(December), pp.1685–1692.

Dissanayake, C.B. & Chandrajith, R., 1999. Sri Lanka–Madagascar Gondwana Linkage: Evidence for a Pan-African Mineral Belt. *The Journal of Geology*, 107(2), pp.223–235.

Dissanayake, C.B., Chandrajith, R. & Tobschall, H.J., 2000. The geology, mineralogy and rare element geochemistry of the gem deposits of Sri Lanka. *Bulletin of the Geological Society of Finland*, 72(1-2), pp.5–20.

Donoghue, M.J. & Moore, B.R., 2003. Toward an integrative historical biogeography. *Integrative and comparative biology*, 43, pp.261–270.

Drummond, A.J. et al., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5), pp.699–710.

Drummond, A.J. et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29, pp.1969–1973.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.214.

Forster, J.R., 1996. Observations Made During A Voyage Round the World. University of Hawai'i Press.

Givnish, T.J. & Renner, S.S., 2004. Tropical Intercontinental Disjunctions: Gondwana Breakup, Immigration from the Boreotropics, and Transoceanic Dispersal. *International Journal of Plant Sciences*, 165, pp.S1–S6.

Guindon, S. & Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52, pp.696–704.

Gunatilleke, I. A. U.N. & Gunatilleke, C.V.S., 1984. Distribution of endemics in the tree flora of a lowland hill forest in Sri Lanka. *Biological Conservation*, 28(3), pp.275–285.

Gunatilleke, I., Gunatilleke, C. & Dilhan, M., 2005. Plant biogeography and conservation of the south-western hill forests of Sri Lanka. *Raffles Bulletin of Zoology Suppl*, 12(12), pp.9–22.

Hausdorf, B., 1998.Weighted ancestral area analysis and a solution of the redundant distribution problem. *Systematic Biology*, 47, pp. 445–456.

Hennig, W., 1966. Phylogenetic Systematics. 3rd edition. Translated by D.D. Davis and R. Zanderl. Urbana: University of Illinois Press.

Herath, W., 1984. Geology and Occurrence of Gems in Sri Lanka *. *J. Natn. Sci. Coun. Sri Lanka,* 12, pp.257–271.

Hughes, M. & Hollingsworth, P.M., 2008. Population genetic divergence corresponds with species-level biodiversity patterns in the large genus Begonia. *Molecular Ecology*, 17(11), pp.2643–2651.

Humphries, C.J. & Parenti, L., 1999. Cladistic Biogeography: Interpreting Patterns of Plant and Animal Distribution. New York: Oxford Univ. Press.

Jayasuriya, A.H.M., 1983. Begoniaceae. In: Dassanayake, M.D., Fosberg, F.R. (Eds.), Flora of Ceylon. Amerind Publishing, New Delhi, pp. 137–152.

Kim, S.C. et al., 1999. The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribeSonchinae (Asteraceae:Lactuceae). *Plant Systematics and Evolution*, 215, pp.85–99.

Kodandaramaiah, U., 2010. Use of dispersal-vicariance analysis in biogeography - A critique. *Journal of Biogeography*, 37, pp.3–11.

Kress, J.W. & Specht, C.D., 2006. The evolutionary and biogeographic origin and diversification of the tropical monocot Oder Zingiberales. *Aliso*, 22, pp.621–632.

Kress, W.J. et al., 2005. The molecular phylogeny of Alpinia (Zingiberaceae): A complex and polyphyletic genus of gingers. *American Journal of Botany*, 92(1), pp.167–178.

Lakhanpal, R. N., 1970. Tertiary floras of India and their bearing on the historical geology of the region. *Taxon,* 19(5), pp.675-694.

Lamm, K.S. & Redelings, B.D., 2009. Reconstructing ancestral ranges in historical biogeography: Properties and prospects. *Journal of Systematics and Evolution*, 47(June), pp.369–382.

Landis, M.J. et al., 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, 62(0), pp.789–804.

Lomolino, M.V., Riddle, B.R. & Brown, J.H., 1998. Biogeography, 3rd edn. Sinauer Associates, Sunderland, MA.

Lyell, C., 1830. Principles of Geology, Being an attempt to explain the former changes of the earth's surface, by reference to causes now in operation. London: John Murray.

Lyte, C., 1981. Sir Joseph Banks. 18th Cemtury Explorer, Botanist & Entrepreneur. *Med Hist*, 25(3), pp.329-330.

Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow im-proved accuracy and model-testing. *Frontiers of biogeography,* 5(4), pp.242-248.

Matzke, N.J., 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, 63(6), pp.951–970.

McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49(3), pp.271–300.

Morley, R. J., 2000. *Origin and evolution of tropical rain forests*. Chichester: John Wiley & Sons, Ltd.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(12), pp.5–20.

Morrone, J.J. & Crisci, J.V., 1995. Historical biogeography: introduction to methods. *Annual Review of Ecology and Systematics*, 26, pp.373-401.

Nathan, R., 2006. Long-distance dispersal of plants. *Science (New York, N.Y.)*, 313(August), pp.786–788.

Nelson, G. J. 1973. Comments on Hennig's "phylogenetic systematics" and its influence on ichthyology. *Syst. Zool,* 21, pp.364-374.

Nylander, J. a a et al., 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: Turdus). *Systematic biology*, 57(2), pp.257–268.

Praveen Karanth, K., 2006. Out-of-India Gondwanan origin of some tropical Asian biota. *Current Science*, 90(6), pp.789–792.

Rajbhandary, S. et al., 2011. Asian *Begonia*: out of Africa via the Himalayas? *Gardens' Bulletin Singapore*, 63, pp.277–286.

Ree, R.H. et al., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution; international journal of organic evolution*, 59(11), pp.2299–2311.

Ree, R.H. & Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, 36, pp.1211–1220.

Ree, R.H. & Smith, S. A, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57(1), pp.4–14.

Renner, S.S. et al., 2010. Biogeography of the Monimiaceae (Laurales): A role for East Gondwana and long-distance dispersal, but not West Gondwana. *Journal of Biogeography*, 37(7), pp.1227–1238.

Renner, S.S., 2005. Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science*, 10(11), pp.550–558.

Ronquist, F., 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(l), pp.195–203.

Ronquist, F. & Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), pp.1572–1574.

Rosen, B.R., 1988. Biogeographic patterns: a perceptual overview. In Analytical Biogeography (Myers, A.A. and Giller, P.S., eds), pp. 23–55.

Rutschmann, F. & Eriksson, T., 2004. Did Crypteroniaceae really disperse out of India? Molecular dating evidence from rbcL, ndhF, and rpl16 intron sequences. *International Journal of plant Science*, 165(S4), pp.S69–S89.

Rutschmann, F., 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions*, 12, pp.35–48.

Sanmartín, I. & Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic biology*, 53(2), pp.216–243.

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), pp.2688–2690.

Swenson, U. & Anderberg, A.A., 2005. Phylogeny, character evolution, and classification of Sapotaceae (Ericales). *Cladistics*, 21, pp.101–130.

Swofford, D.L., 1993. PAUP*: Phylogenetic Analysis Using Parsimony.

Thomas, D.C. et al., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, 39(1), pp.98–113.

Von Humboldt, A. 1805. Essai sur la geographie des plantes accompagne d'un tableau physique des regions equinoxiales, fonde sur les mesures executees, depuis le dixieme degre de latitude boreale j'usqu'au dixieme degre de latitude australe, pendent les annees 1799, 1800, 1801, 1802 et 1803. Paris: Levrault Schoell.

Von Humboldt, A., 1808. Ansichten der natur mit wissenschaftlichen erlauterungen. Tubingen, Germany: J.C. Cotta.

Wallace, A. R., 1860. On the Zoological Geography of the Malay Archipelago. Zoological Journal of the Linnaean Society, London 4: 172-184.

Wallace, A. R., 1863. On the physical geography of the Malay archipelago. Journal of the Royal Geographical Societyof London 33: 217-234.

Wallace, A. R., 1869. The Malay Archipelago. London: Macmillan.

Wallace, A. R., 1876. The Geographical Distribution of Animals. London: Macmillan.

Willdenow, K.L., 1811. The principles of Botany and of Vegetable Physiology. University of Edinburgh Press.

Wegener, A., 1912. Die Entstehung der Kontinente. Geologische Rundschau 3(4): 276-292.

Wegener, A., 1966. The Origin of Continents and Oceans. New York: Dover

Yang, Z. & Rannala, B., 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13(5), pp.303–314.

Yu, Y., Harris, a. J. & He, X., 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, 56(2), pp.848–850.

Zwickl, D.J., 2006. GARLI manual (version 0.95). *Software manual*, p.21.

CHAPTER 3: The early evolution and mode of range expansion in Asian *Begonia*

Introduction

Tropical areas contain a disproportionate amount of the world's biodiversity. This diversity is not spread evenly throughout the tropics, but concentrated in certain regions. Many of these regions with exceptionally high endemism and diversity are under threat, and are called Biodiversity Hotspots, representing 1.4% of earth's surface but containing 44% of flowering plant diversity (Myers et al. 2000). The origin of tropical hyperdiversity is only just beginning to be understood. The prevailing view of the importance of vicariance as a major influence on the evolution of the tropical flora has recently been challenged, as the arrival of molecular dating techniques and advancement of biogeographic models have revealed a much more recent origin for most Angiosperm families, and therefore more recent dispersal events provide much more plausible explanations for extant plant distribution patterns (Pennington & Dick 2004). Biodiversity hotspots provide an ideal opportunity to examine the origins and mode of range expansion of hyperdiverse tropical floras. Within continental Asia, three regions, Indo-Burma, Western Ghats and Sri Lanka, and South and central China have been declared as biodiversity hotspots with Indo-Burma and Western Ghats/Sri Lanka being placed amongst the hottest hotspots due to their rich endemic plant and vertebrate diversity (Myers et al. 2000).

The biogeographic history of Malesia has been well documented and extensively studied (Crayn et al. 2014), however the broad scale biogeographic patterns of Continental Asia remain comparatively unknown despite the high floristic richness of the region. The plant family Begoniaceae makes an excellent candidate to test biogeographic patterns within continental Asia due to its high levels of species richness and endemism in the above biodiversity hotspots. Begoniaceae are a large pantropical family widely distributed in the tropics with greatest species diversity in Southeast Asia and South America. The largest genus *Begonia* dominates the family with ca. 1800 species including ca. 700 species in the neotropics, 160 species in Africa and 900 spp. in Asia with the bulk occurring in Southeast Asia (Goodall-Copestake et al. 2009;

Rajbhandary et al. 2011; Thomas et al. 2012). The monotypic *Hillebrandia sandwicensisis* a Hawaiian endemic (Clement et al. 2004).

This morphologically diverse *Begonia* is characterised by asymmetrical leaves, unisexual flowers and monoecy, twisted, papillose stigmas and dry three winged capsules (Doorenbos et al. 1998, Judd et al. 2008). The Hawaiin endemic *Hillebrandia* deviates from *Begonia* by having more differentiated segments of the perianth semi inferior ovary and fruit dehiscence between styles (Clement et al. 2004, Forrest et al. 2005). Numerous phylogenetic, cytological and biogeographic studies have been carried out during past decades in order to resolve the relationships within this large, morphologically diverse family.

The genus was circumscribed in to 63 sections by Doorenbos et al. (1998) and currently is classified into 69 sections (Hughes et al. 2015). All other sections except section *Tetraphila* shows restricted distribution in single continent; either in Asia, Africa or neotropics however single lineage of *Tetraphila* has been recently discovered in continental South East Asia (De wilde et al. 2011).

Polyphyly of sections *Sphenanthera, Platycentrum* and *Leprosae* is evidenced (Forrest et al. 2005, Tebbit et al. 2006) based on nuclear ribosomal DNA. Several small sections were found to be nested within some large sections; section *Baryandra* within Philippine *Diploclinum*, sections *Alicida* and *Putzeysia* within Section *Diploclinum* and section *Monopteron* within *Platycentrum* (Rajbhandary et al. 2010; Thomas, 2010).

Several chloroplast and nuclear markers have being used in resolving sectional and species level relationships as well as biogeographic analysis in *Begonia* (Dewitte et al. 2008).

The phylogenetic relationships of African *Begonia* are well resolved by Plana (2003) by using 81 *Begonia* species covering 17 African sections with the use of chloroplast *trn*L intron. African lineages were found to be non-monophyletic and the study reveals the possibility of using *trn*L intron region in identifying major clades within the genera.

Goodall-Copestake,(2005) suggested an African ancestry for both Asian and American *Begonia* based on ca. 13kb of sequence data from eleven regions of nuclear,

mitochondrial and chloroplast DNA of 31 *Begonia* species. Forrest et al.(2005) also confirmed African paraphyly based on nuclear large subunit and internal transcribed spacer sequences along with morphological data. Neotropical, Asian, and Socotran *Begonia* were found nested within the African *Begonia*. This pattern was later confirmed by Thomas et al. (2012).

Within Asian *Begonia,* two well supported major clades were identified; one dominated by continental Asian taxa comprising species from sections *Parvibegonia, Diploclinum, Platycentrum* and *Sphenanthera*, and the second major clade comprise of *sections Rideleyella, Bracteibegonia, Petermannia* and polyphyletic Malesian section *Diploclinium* with Chinese section Coelocentrum being the sister.

Himalayan *Begonia* have been found nested within Asian *Begonia* grade which fall in to two groups; one with a grade of tuberous, deciduous species and the other with evergreen rhizomatous species (Rajhbandry et al. 2011). Two major sections of Himalaya *Begonia* are identified; one old of Miocene age (sect *Diploclinium*) and the other comprising of sections *Platycentrum*and *Sphenanthera* which originated in the Pliocene. The nuclear ITS region resulted in a poorly resolved back bone for the phyogeny, however the results hint at China or Indochina being the potential areas of origins for Asian *Begonia*.

The utility of chloroplast markers in resolving phylogenetic relationships were investigated by Thomas (2011), on Southeast Asian *Begonia* to test the monophyly of Asian sections and to build up the Asian *Begonia* phylogeny.

Among those regions, three fast evolving non-coding cpDNA regions: *ndh*A intron, *ndh*F-*rpl*32 spacer and the *rpl*32-*trn*L regions contained higher percentage of variable and parsimony informative sites than the other chloroplast regions tested: *mat*K, *pet*D gene and intron, *psb*B gene, *trn*L intron. Further, large global *Begonia* datasets comprising Neotropical, Southeastern Asian and Philippine *Begonia* were based on these three chloroplast regions (Moonlight et al. 2015). Thus, in order to complement the lacking South Asian and Himalayan element and to obtain a highly resolved phylogeny 20 sequences were newly generated for the current study and added to existing *Begonia* dataset comprising 159 species.

I aim to shed light on the biogeographic history of the Continental Asian hotspots, focussing on three main hypotheses. Asian *Begonia* are known to be monophyletic (Thomas et al. 2012), however the point of entry to Asia is still uncertain. Thus, first I aim to test the point of entry to Asia prior to initial diversification from an African ancestor. Secondly, I investigate the routes of migration across continental Asia which explains extant distributions, and the biogeographic connections of the hotspots. Finally I examine the mode of range expansion and speciation within continental Asia, and the balance between vicariance and dispersal.

Route of entry

The *Begonia* lineage first originated in Africa, based on plastid and mitochondrial DNA phylogenies sampling thirty *Begonia* species with representatives from Africa, America and Asia (Goodall-Copestake et al. 2010). Asian and Socotran *Begonia* were shown to be monophyletic, confirmed in a more densely sampled study of 112 species by Thomas et al. (2012) who also demonstrated the Asian species began to diversify during the early to mid Miocene and subsequently dispersed into Malesia. Both Asian *Begonia* monophyly and an African origin for the genus were confirmed by Moonlight et al. (2015) using 268 samples from across the tropics. A second African lineage was discovered in Asia by de Wilde et al. (2011). This was a single dispersal event in Pleistocene, resulting in the generation of a single new species which has not played a further role in the broader picture of Asian *Begonia* diversity. There are several hypotheses on the route of entry of *Begonia* into Asia. One possibility is a migration from Africa to Asia through an Arabian corridor (Goodall-Copestake et al. 2010) during favourable warmer and wetter climatic periods in Paleogene to Neogene (Zachos et al. 2001). Extensive aridification and global cooling might resulted in extinctions, leaving only *B. socotrana* on Socotra (Goodall-Copestake et al. 2010).

Long distance dispersal (LDD) is another possibility, although given the phylogenetic and population genetic patterns observed in *Begonia* LDD does not appear to be common (Hughes & Hollingsworth 2008). However the occurrence of *B. afromigrata*, an African element, in Laos and Thailand shows the feasibility of successful colonisation following LDD. The lack of evidence for a continuous wet forest corridor from Africa to Thailand during the Pleistocene favours long distance dispersal rather

than overland migration for this species (de Wilde et al. 2011).The route of entry for the monophyletic large radiation of Asian *Begonia* from Africa via LDD is perhaps more likely to be found in western or southern Asia given their proximity to the African area of origin for the genus. I aim to test the following hypotheses of routes of entry into Asia from Africa for the large genus *Begonia* (Figure 3.1).

1. Arabian corridor. During favourable climatic conditions *Begonia* could have been dispersed from Africa to Socotra and then migrated to Asia via a mesic Arabian corridor. If Socotra was a point on the route of entry to Asia, Socotran *Begonia* would form a sister clade to the rest of the Asian species.

2. Dispersal to Sri Lanka. If Sri Lanka was the route of entry, Sri Lankan *Begonia* would be the sister group to the rest of the Asian species. Long distance dispersal events could have taken place out of Africa to Sri Lanka during the time when the Deccan Plate containing Sri Lanka moved in a close proximity to Africa during its northward migration. Initial colonization in Sri Lanka could have been followed by dispersal to the Indian mainland followed by overland migration from West to East. Further, given the distribution of *Begonia* sect. *Reichenheimea*, found in Sri Lanka and with most diversity in Southeast Asia, there could have been some long dispersal from Sri Lanka to Southeast Asia.

3. Dispersal to India. If India acted as the route of entry via LDD, an Indian clade would take the sister position to the rest of the Asian species. Long distance dispersal events could have taken place out of Africa to the coastal mountains of western India during the northward migration of the Deccan Plate.

4. Dispersal to the Himalayas. Dispersal to the Himalayan range is another potential route of entry to Asia, and we would expect Himalayan *Begonia* to form a sister group to other Asian species, followed by migration to the rest of Asia, including to Sri Lanka via the Western Ghats.

Both Socotra and Sri Lanka are isolated and we might expect dispersal to and from these relatively small areas to be infrequent, leading to a simple sister relationship in the phylogeny if these areas harbour species from the initial immigration. Concerning the Western Ghats of India or the Himalaya, we might expect a more complicated scenario

in the hypothetical phylogeny, since the large area provides more potential for multiple lineages and dispersal events, resulting in species from the Western Ghats or the Himalaya being paraphyletic with other Asian species nested within.

Continental biogeography

Even though molecular phylogenetic studies on Begoniaceae have greatly advanced our knowledge of the evolution of the family, the relationships and biogeography of continental early diverging Asian *Begonia* lineages remain poorly understood (Goodall-Copestake et al. 2010; Thomas et al. 2012) and key samples from west and south Asia have not been included to date. Rajbhandary et al. (2011) increased the sampling of Himalayan species compared to earlier studies but failed to get an unequivocal biogeographic hypothesis as the phylogeny, based on ribosomal ITS sequences, was poorly supported.In this paper I have expanded the chloroplast sequence dataset from Thomas et al. (2012), which provides better phylogenetic support than ITS, by adding more samples from South Asia and the Himalayas. Thomas et al. (2012) coded continental Asia as a single geographic unit; my increased sampling allows me to expand this into five areas (India, Sri Lanka, Himalaya, China, Indo-China) to permit insight into development of the hotspot floras.
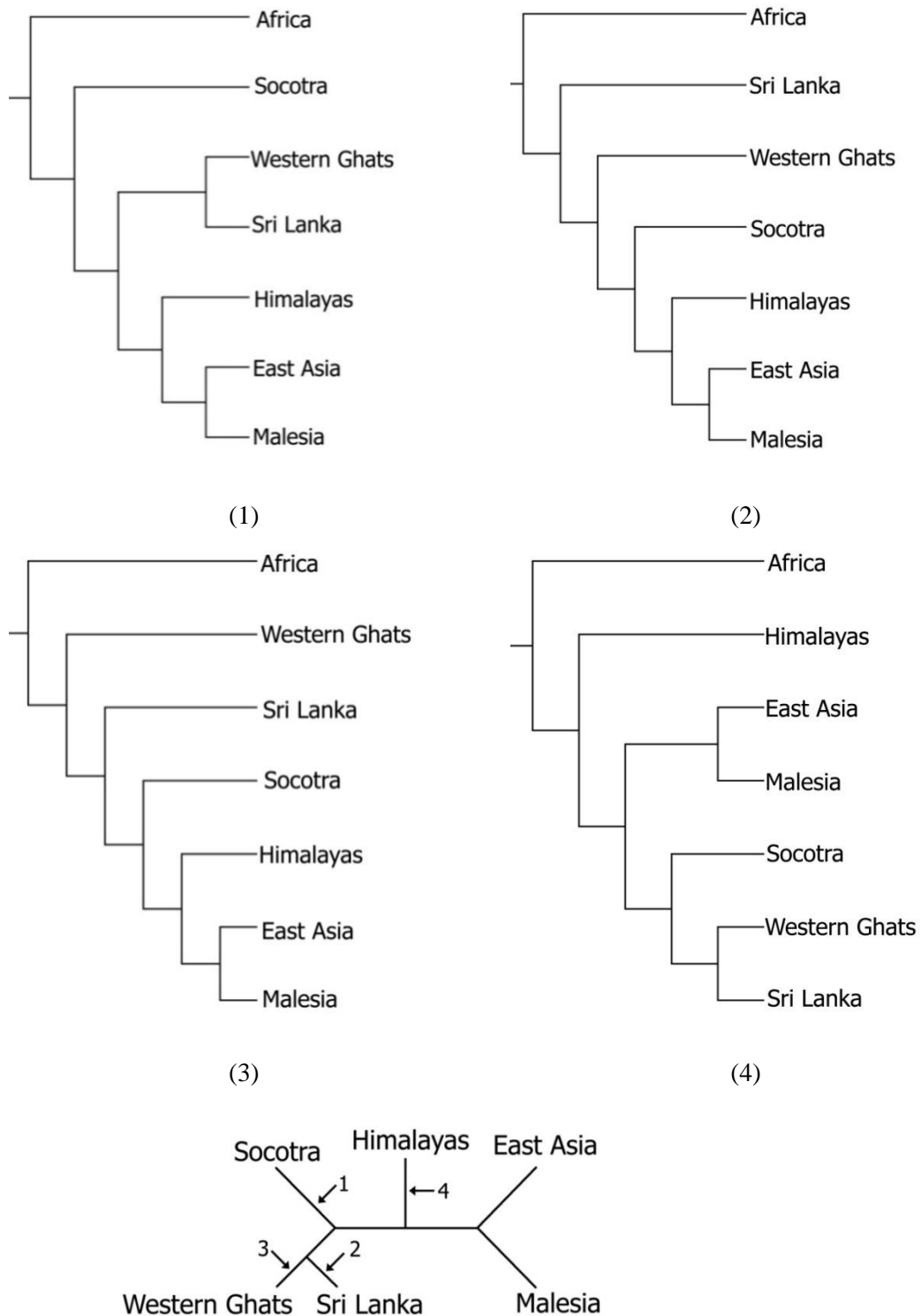
Figure 3.1 Phylogenetic representations showing the history of four hypothetical routes of entry of *Begonia* to Asia from Africa, and dispersal within Asia. The unrooted network at the bottom of the figure depicts hypothetical relationships among the major geographic areas with arrows indicating the placement of the root for each of the four colonization routes.

Mode of range expansion

In addition to testing these broad biogeographic hypotheses, I investigate the mode of range expansion which has led to the monophyletic radiation generating ca. 900 species and covering all of tropical Asia. In order to obtain a complete picture of Asian *Begonia* biogeography, we need to understand the contributions of vicariance, dispersal and founder event speciation to the radiation.

The parsimony, event based, quantitative method, Dispersal-vicariance analysis (DIVA) (Ronquist 1997) is a widely used method in inferring ancestral area relationships based on dispersal and extinction events (Ronquist 1997; Buerki et al. 2011) with vicariance given a zero cost. Phylogenetic branch length is not taken in to account (Clark et al. 2008) and optimal reconstruction is produced by a parsimony criterion that minimizes dispersal-extinction events (Ronquist 1997). DIVA can result in wide ancestral ranges (Kodandaramaiah 2010), and may not provide a good model for range evolution where it is not driven primarily by vicariance.

The parametric counterpart of DIVA, Dispersal-Extinction-Cladogenesis (DEC) (Ree & Smith 2008) infers ancestral area reconstructions under a likelihood approach. Even though local extinction and dispersal are treated as free parameters, DEC assigns a weight of 1 to each allowed cladogenesis event, and the probability of each event is obtained through dividing by the sum of the weights of all the allowed events (Matzke 2014). Thus the cladogenesis model is fixed and range inheritance by daughter lineages by different mechanisms are considered to have an equal probability. DEC also could result in numerous equally likely ancestral reconstructions at the base (Moonlight et al. 2015) and the mechanism of identical range inheritance when the ancestral range comprises multiple unit areas is not addressed (Lamm & Redelings 2009).

During cladogenesis when the ancestor is restricted to a single area, both DIVA and DEC models allow speciation within that area. However, when an ancestor occurs in more than one area, DEC assumes that one daughter lineage will always be restricted to one area and permits separation of the single area from the remaining ancestral range. However DIVA, on the other hand, permits classical vicariance where each daughter can occupy more than one area. When the ancestral area is wide, DIVA does not permit speciation within one area whereas DEC does (Ronquist & Sanmartín 2011).

Recent advancement in biogeographic analysis allows likelihoods to be assigned to models of range evolution for a particular dataset. The package Biogeobears (Matzke 2014) allows choosing the best-fit model under a maximum likelihood framework by modeling transitions between discrete states (biogeographical ranges) along phylogenetic branches as a function of time. The DEC model calculates maximum likelihood ratios of ancestral states at speciation events, in a method similar to LAGRANGE (Ree et al. 2005). In addition to two free parameters; d (dispersal), e (extinction) included in the DEC model (Batalha-Filho et al. 2014, Landis et al. 2013) the new additional parameter *J* is added to the model. The *J* parameter tests the contribution of founder event speciation to range expansion and controls the probability of two events during cladogenesis; founder event speciation versus sympatric and vicariant speciation (Matzke 2014). Founder event speciation is potentially important in lineage splitting especially in island systems, and here I use the functionality of Biogeobears to find the best fit model for the range expansion of Asian *Begonia*.

My study is the first to incorporate endemic Sri Lankan *Begonia* in a global Begoniaceae phylogeny, potentially permitting new insights into the early evolution and biogeography of Asian *Begonia*.

Materials and Methods

Taxon sampling

In order to represent potentially early divergent lineages in Asian *Begonia* the following samples (20 species; five Sri Lankan species, one Indian species, six Southeast Asian species and eight Himalayan species) were added to a previously published Asian *Begonia* DNA sequence alignment (Thomas et al. 2012) comprising 106 species: *Begonia tenera* and *B. thwaitesii* (both Sri Lankan endemics); *B. cordifiolia*, *B. malabarica* and *B. dipetala* (distributed in both in Sri Lanka and Western Ghats in India); *B. albo-coccinia* (Western Ghats). To shed more light on continental Asian Begonia biogeography, eight Himalayan species were added: *B. diocia*, *B. flagellaris*, *B. picta*, *B. hatacoa*, *B. tribenensis*, *B. bryophila*, *B. panchtharensis* and *B. rubella*. Voucher information for sequences used in the analysis is listed in table3.1. In addition, 53 samples of *Begonia* sect. *Baryandra* from the Philippines were added to complement the sampling in the Malesian region (Hughes et al. 2015). Asian *Begonia* have been

proved as monophyletic by previous studies (Goodall-Copestake et al. 2010; Rajbhandary et al. 2011; Thomas et al. 2012). Thus, I chose four African and five South American species as an out group for my study. The data set comprised 179 taxa representing 14 sections of *Begonia*. Voucher information for all the species used in the analysis are listed in appendix 1.

Table 3.1: Voucher information for newly generated sequences for the study

Species name	Voucher/living accession	EDNA number
B. cordifolia Thwaites	Kumarage 14 (E)	EDNA13-0032556
B. thwaitesii Hook	Kumarage 23 (E)	EDNA13-0032557
B. dipetala Graham	Kumarage 25 (E)	EDNA13-0032558
B. malabarica Lam.	Kumarage 28 (Photo voucher available)	EDNA13-0032559
B. tenera Dryand	Kumarage 68 (E)	EDNA13-0032560
B. albo-coccinea Hook	Photo voucher available	
B. puspitae Ardi	20111539	EDNA12-0029751
B. sublobata Jack	20101649	EDNA12-0025038
B. spec.	20100763	EDNA12-0025039
B. sect. Reichenheimea	20111543	EDNA12-0029747
B. sect. Reichenheimea	20111545	EDNA12-0029749
B. sect. Reichenheimea	20112191	EDNA12-0025037
B. picta Sm.	5993	EDNA14-0035336
B. tribenensis C.R.Rao	6043	EDNA14-0035338
B. bryophila Ined.	6100	EDNA14-0035340
B. rubella Ham. Ex D.Don	6000	EDNA14-0035341
B. hatacoa Buch.Ham.	5971	EDNA14-0035344
B. panchtharensis S.Rajbh.	5968	EDNA14-0035345
B. flagellaris Hara	6010	EDNA14-0035346
B. diocia Ham. Ex D.Don	13651	EDNA14-0035352

DNA Extraction, Sequencing and Alignment

Total genomic data was extracted from silica gel dried material using the DNeasy Plant Mini Kit (Qiagen, UK) according to the manufacturer's protocols. 25 μl PCR reactions were setup for amplification of both ITS and chloroplast regions.

Three non coding regions of plastid DNA (*ndh*A intron, *ndh*F-*rpl*32 spacer, *rpl*32-*trn*L spacer) were used in the current study to complement the previously available data. The primer sequences used in the study are listed in Table 3.2.

For the amplification of plastid DNA regions, each 25μl of PCR mixture contained 15.25μl of ddH$_2$O, 2.5μl of 10x reaction buffer, 1.25μl of 25mM MgCl$_2$, 2.5 μl of 2mM dNTPs, 0.75μl of 10μM forward primer, 0.75 μl of 10μM reverse primer, 0.8μl0.4% BSA, 0.2 μl of Biotaq DNA polymerase (Bioline, UK) and 1μl of DNA template. The temperature profile included an initial template denaturation step of 80°C for 5 minutes, followed by 30 cycles of denaturation at 95°C for 1 min, primer annealing at 50°C for 1 min, primer extension at 65°C for 4 min and a final extension step at 65°C for 5 min.

Amplified products were run on a 1% Agarose gel with Syber Safe as the staining agent and visualized in UV transiluminator. The PCR purification was done using EXOSAP-IT as 7 μl reaction mixtures. 5μl of PCR product was mixed with 2μl of EXOSAP and incubated for 37°C for 15 minutes followed by 80° C for 15 minutes.

Table 3.2: Primers used in the study for amplification of plastid DNA

DNA region	Primer	Primer sequence	References
*ndh*A intron	ndhAx1 ndhAx2	GCY CAA TCW ATT AGT TAT GAA ATA TTC GGT TGA CGC CAM ARA TTC CA	Shaw et al. 2007 Shaw et al. 2007
*ndh*F-*rpl*32	rpl32-R ndhF Beg1F Beg2R	CCA ATA TCC CTT YYT TTT CCA A GAA AGG TAT KAT CCA YGM ATA TT TGG ATG TGA AAG ACA TAT TTT GCT TTT GAA AAG GGT CAG TTA ATA ACA A	Shaw et al. 2007 Shaw et al. 2007 Thomas et al. 2014 Thomas et al.2014
*Rpl*32-*trn*L	rpl32-F trnL	CAG TTC CAA AAA AAC GTA CTT A CTG CTT CCT AAG AGC AGC GT	Shaw et al. 2007 Shaw et al. 2007

Sequencing PCR was done using purified PCR products as 10μl mixtures using 5.68μl of ddH$_2$O, 2μl of sequencing buffer, 0.32μl of primer, 1μl of Big dye and 1μl of template. The sequencing PCR protocol was, denaturation at 95°C for 30 sec, followed by 24 cycles of primer annealing at 50°C for 20sec, extension at 60°C for 4 min.

Separate forward and reverse sequencing PCR s were carried out and products were sent to the Genepool facility at the University of Edinburgh (Genepool, UK) for BigDye Terminator Cycle sequencing.

Sequences were manually aligned in Mesquite (3.03; Maddison & Maddison 2015). The following bases were excluded due to the ambiguities in the alignment or missing data at the region ends: 55-103 462-476 1092-1111 1230-1249 1263-1280 2046-2064 2136-2150 2258-2284 2368-2379 3370-3410 3571-3604 3971-4012 4485-4507.

Phylogenetic analysis and divergence time estimation

The plastid data set was analyzed separately under Bayesian inference (BI) using MrBayes 3.2.1 (Ronquist et al. 2012) in the CIPRES science gateway V. 3.3 (Miller et al. 2010) and treated as single partitions. MCMC runs were carried out for 10,000,000 generations and sampled every 1000 generations. A 25% burn-in was set to discard the first set of trees and the remaining trees were summarised as a 50% majority rule consensus tree, visualized in FigTree (Figure 3.2) (Rambaut 2009).

Models of sequence evolution were determined using jmodel test 2.1.3 (Posada et al. 2012). Twenty four models were tested and the optimal model was chosen using the Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC). The GTR+I+G model was the most probable model under both criteria for the 3-region dataset.

Bayesian divergence time estimation was performed using BEAST v.1.8.0 (Drummond & Rambaut 2007). The data set was treated as a single partition and run under a lognormal relaxed molecular clock.

The only *Begonia* macrofossils known (Sults & Axsmith 2011) is too young to provide a calibration point and lacks the synapomorphies for the placement of the family. Thus, the dates obtained from primary calibration of broadly sampled phylogeny of Cucurbitales-Fagales dataset by Thomas et al. 2012 were used as a secondary calibration. A similar calibration has been utilized in more recent studies on the divergence time estimates of Begoniaceae (Moonlight et al. 2015; Hughes et al. 2015; Rajbhandary et al. 2011).

Two calibration points were used for the divergence time estimates for the most recent common ancestor for *Begonia*. The crown group was given a mean age of 24 Ma and all Asian *Begonia* a mean age of 18.2 Ma. To account for the uncertainty of the age

estimates of primary analysis,the crown group was modelled with a normal distribution with a standard deviation of 3.57 Ma (Thomas et al. 2014).

Four separate Markov Chain Monte Carlo (MCMC) runs were carried out for 10000000 generations sampling every 1000 generations under Birth-Death model of speciation. Plots of the logged parameters for each run were visualised using Tracer v.1.5 (Drummond & Rambaut 2007) to confirm convergence between runs. Trees were combined in LOGCOMBINER (Drummond & Rambaut 2007) and burn-in was set for 25% for initial sample for each run and a single maximum clade credibility tree was obtained from Tree Annotator v.1.7.5 (Drummond & Rambaut 2007) and visualized in fig tree v.1.4.0 (Rambaut 2009).

Biogeographic analysis

Twenty-two geographic areas were coded based on extant distributions and areas of endemism, to permit a higher resolution of the range expansion in continental Asia and the mode of range expansion across the entire study area: (1) Americas; (2) Africa; (3) Socotra; (4) India; (5) Sri Lanka; (6) Himalaya; (7) China; (8) Indo-China; (9) Peninsular Malaysia; (10) Sumatra; (11) Borneo; (12) Sulawesi; (13) Papua New Guinea; (14) Negros; (15) Java-Lesser Sunda; (16) Luzon; (17) Lanyu&Batan; (18) Panay; (19) Palawan; (20)Mindanao; (21)Sibuyan; (22) Biliran. The data matrix was prepared by coding presence/ absence in each of the areas (Appendix 2) based on data in Hughes et al. (2015). The majority of *Begonia* species are narrow endemics (Hughes & Hollingsworth 2008), hence the maximum number of areas permitted in the ancestral range reconstructions was constrained to two in order to reflect biological reality and to constrain analysis time. Only the widespread *Begonia longifolia* and *Begonia palmata* are found in more than two geographic areas as coded in this study. They were coded as present in China and Indo-China by considering their current distribution hotspots and likely area of origin in Rajbhandary et al. (2011).

Ancestral areas at internal nodes were constructed using Biogeobears (Matzke 2013) under four models; DIVA-like (Ronquist 1997), DIVA-LIKE+J, DEC (Ree & Smith 2008) and DEC+J. Log likelihood values for each model were compared to identify the best model for inferring ancestral ranges at nodes.

Phylogenetic analyses - three-region matrix

Descriptive statistics for 179 taxa plastid dataset and nucleotide partitions including amplicon length, alignment length, number of included characters and number of variable characters are given in the table 3.3.

Trees derived from the non coding plastid DNA sequences of 179 taxa are presented in figures 3.2 & 3.3. Neotropical and Asian *Begonia* are nested within an African *Begonia* grade, with Asian+Socotran *Begonia* resolving as a strongly supported clade with PP=1.0 (Clade I). The Asian+Socotran *Begonia* clade has two major subclades, clades II and III (Figure 3.3).

Table 3.3 Dataset descriptive statistics

DNA region	Alignment length	Number of variable characters	Number of informative characters
Combined data set	4271	1537 (35.98%)	838 (19.62%)
*ndh*A	1419	467 (32.91%)	240 (16.91%)
*ndh*F-*rpl*32	1036	412 (39.76%)	240 (23.16%)
*rpl*32-*trn*L	1816	658 (36.23%)	358 (19.71%)

Clade II is a geographic mosaic and consists of species mainly from Socotra, Sri Lanka, India, Himalaya, and Eastern Asia; however it is not supported (PP=0.28). It contains the western limit of Asian *Begonia* and is the sister to clade III comprising largely eastern Asian *Begonia*. Clade III also does not show any support (PP=0.13) with the early diverging lineage of the Himalayan *Begonia dioica* being the sister to all other taxa. Hence clades II, III and *B. dioica* form three effectively unresolved lineages at the base of the Asian+Socotra clade. Within clade III three species from China and Indochina (*B. morsei, B. ningmingensis* and *B. masoniana*) form a strongly supported clade (PP=1.0) and *B. kingiana* from Peninsular Malaysia is sister to a largely Malesian clade comprising Philippine *Begonia* (Clade XIII) and other South East Asian *Begonia* (Clade XIV).

Two Socotran species, *B. socotrana* and *B. samahensis*, together with Indo-Sri Lankan species form clade IV with PP=0.92 with Socotran *Begonia* being the sister group to the Indo-Sri Lankan elements (clade V). Clade V is strongly supported (PP=1.0) comprising

a grade of Indian and Sri Lankan *Begonia* (*B. malabarica, B. dipetala, B. cordifolia*) with two Sri Lankan endemics*, B. tenera* and *B. thwaitesii,* nested within. The remaining Indian taxa form clade VIII (*B. floccifera*, *B. albo-coccinea*, *B. malabarica*) and are sister to everything else in the poorly supported Clade VII (PP= 0.59), making Indian *Begonia* paraphyletic. The Indo-Sri Lankan species, *B. malabarica* appears separately in clade V and Clade VIII. Sri Lankan *B. malabarica* grouped with *B. dipetala* and it is highly supported as monophyletic (PP=1.0) while Indian *B. malabarica* grouped with *B. albo-coccinea* with high support value (PP=1.0).

Clade IX (PP=0.99) is dominated by continental Asian (China+Indochina) *Begonia* with Himalayan species highly nested within. Clade XI (PP=1.0) consists of a paraphyletic group of Himalayan species intermixed with species from China, Indochina, Peninsular Malaysia, Sumatra and Sulawesi.

The largely Philippine *Begonia* sect.*Baryandra* (Clade XIV) resolves as monophyletic (PP=1.0) while other Southeast Asian *Begonia* form Clade XV, PP= 0.58.

Divergence time estimates

The mean divergence time estimate for Asian *Begonia* is 15.5 Mya.

The age of the Socotran/Indo-Sri Lankan clade (Clade IV) is 13.5 (8.0-19.2) Mya which is similar to age estimates by Thomas et al. (2012) of 13.6 (7.3–19.5) Mya. The Indo-Sri Lankan element (Clade V) shows a divergence time of 7.7 (3.9-12.6) Mya while the Sri Lankan endemics (Clade VI) show a more recent Pleistocene diversification of 0.8 (0.2-1.8) Mya.

The rest of the Indian taxa (Clade VIII) have a stem age of 15.1 (9.6-21) Mya and a crown age of 5.9 (3-9.4) Mya. Clade IX, dominated by species from Indo-China and Himalaya begins to diversify 14.4 (7.7-17.2) Mya. Himalayan *Begonia* (Clade XI) have an origin 7.9 (4.8-11.6) Mya and comprises two main sub clades. One is endemic to the Himalaya, and the other also contains species from Sumatra, Sulawesi, China and Indochina; clade XII which has a crown age of 5.2 (3.1-7.9) Mya.

Begonia diocia from Himalaya shows an early branching position in the phylogeny at 15.8 Mya and is weakly supported at the base of clade III. South East Asian *Begonia* (clades XIV and XV) has a crown age of (12.5 Mya).

Biogeographic analysis and ancestral area reconstructions

Amongst the models used in BIOGEOBEARS for ancestral area reconstructions, the models with the J parameter included gave results with higher likelihood values (Table 3.4), with DEC+J giving the highest likelihood score of -337.91, and was chosen as the best fitting model for the data.

Table 3.4 d (dispersal), e (extinction), j (j value, founder-event speciation) LnL (log likelihood) for each of the geographic range evolution models compared in Biogeobears.

Model	d	e	j	LnL
DEC	0.0045	0.0213	0	-417.47
DEC+J	7.00E-04	0.0013	0.0091	-337.91
DIVA LIKE	0.0043	0.009	0	-402.38
DIVA LIKE + J	8.00E-04	0	0.0089	-341.79
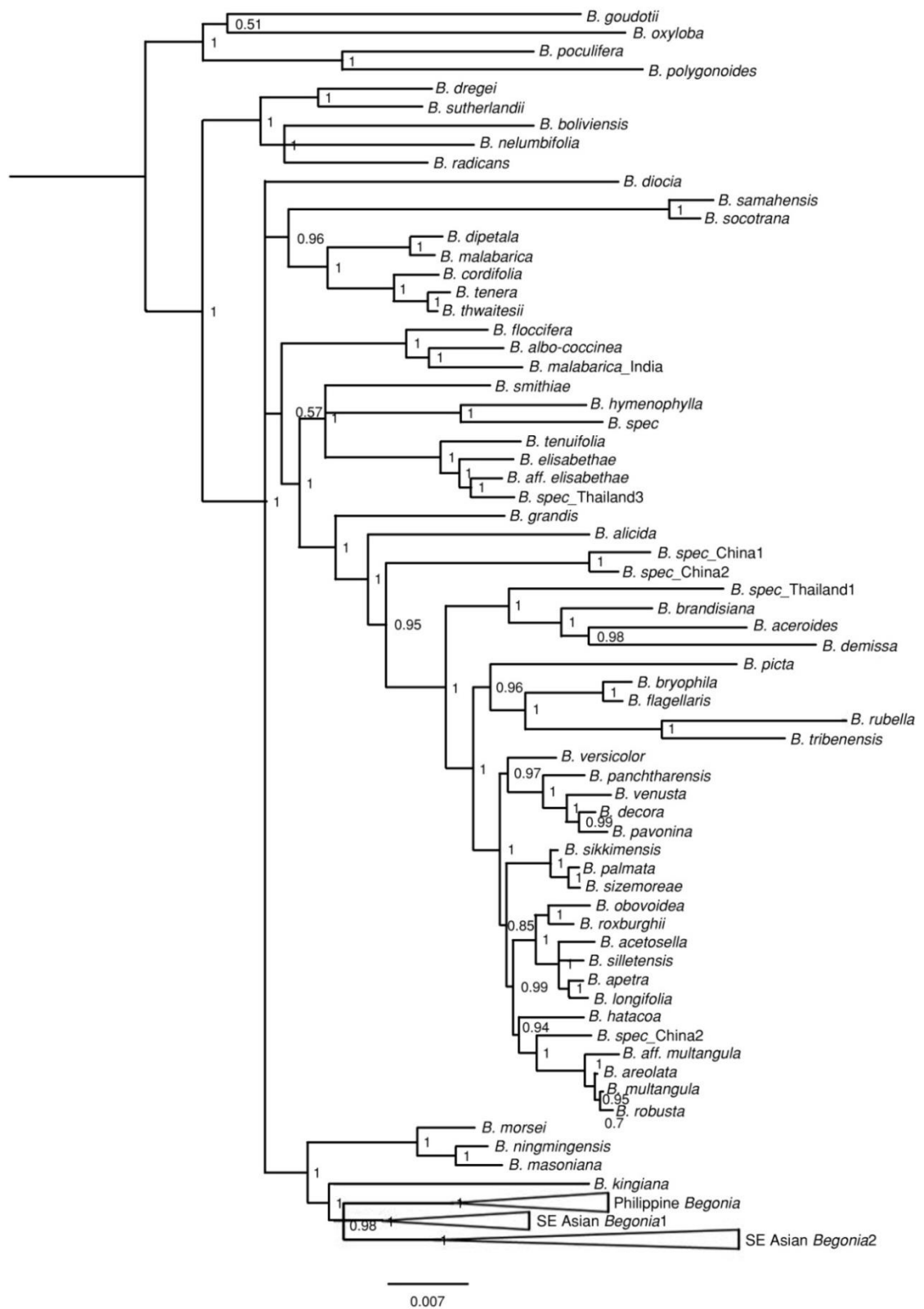
Figure 3.2 Bayesian majority rule consensus tree resulted from non coding region of plastid DNA. Bayesian posterior probability (PP) support values are indicated next to the nodes.
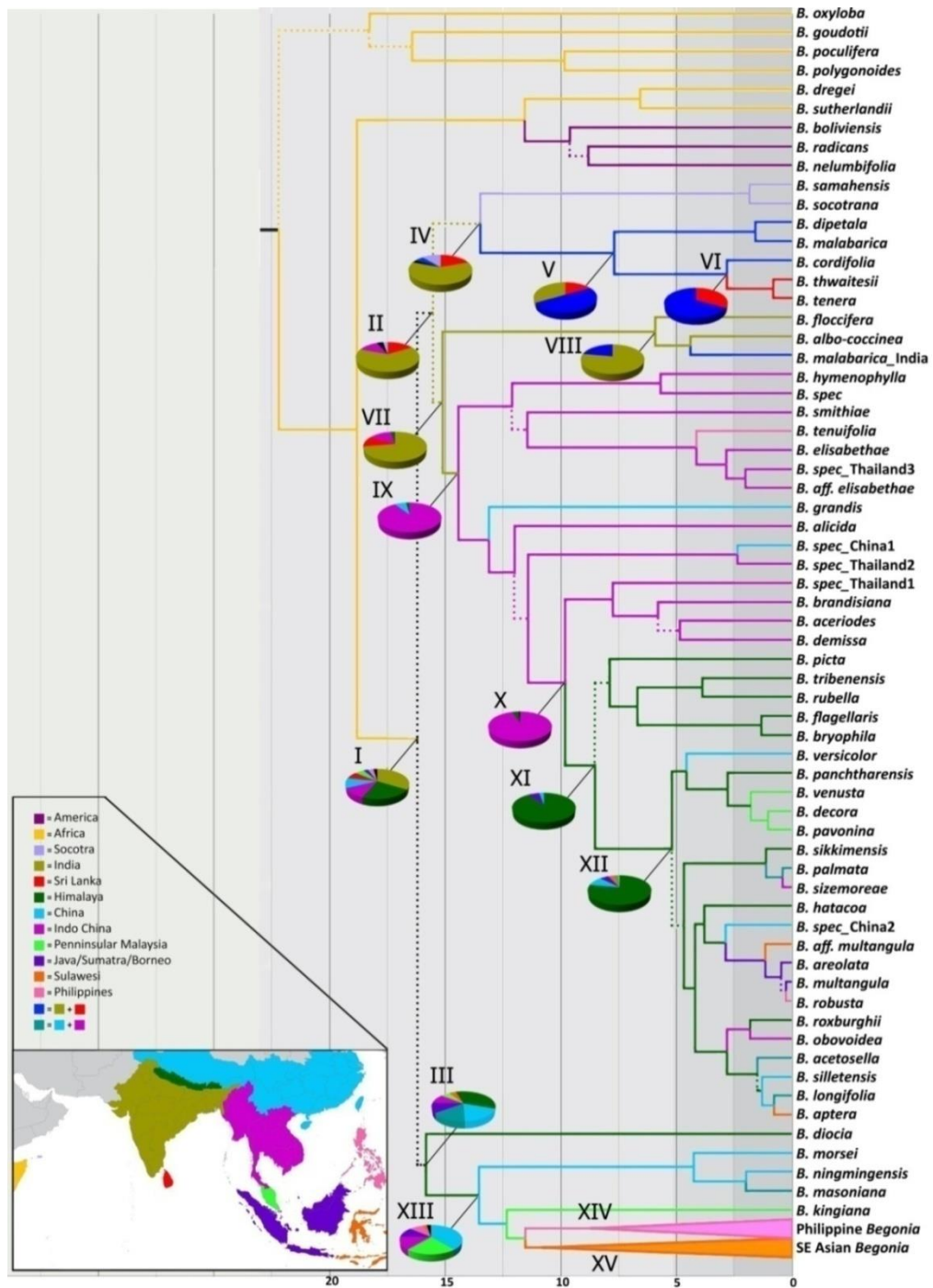
Figure 3.3 Maximum-clade-credibility chronogram of a Beast analysis of the three-region *Begonia* data set. Node heights indicate mean ages. Numbers at nodes represent clades in Table 3.2. Branches coloured according to their optimal range reconstructions under the DEC+J model in the package Biogeobears. Pie charts show the relative probability of ancestral state reconstructions at selected nodes. Geological epochs are indicated by background colour: light grey, Miocene (5.3–2.3 Mya); mid-grey, Pliocene (0.26–5.3 Mya); dark grey, Holocene and Pleistocene (0–0.26 Mya). Dotted lines indicate posterior clade probabilities less than 0.95.
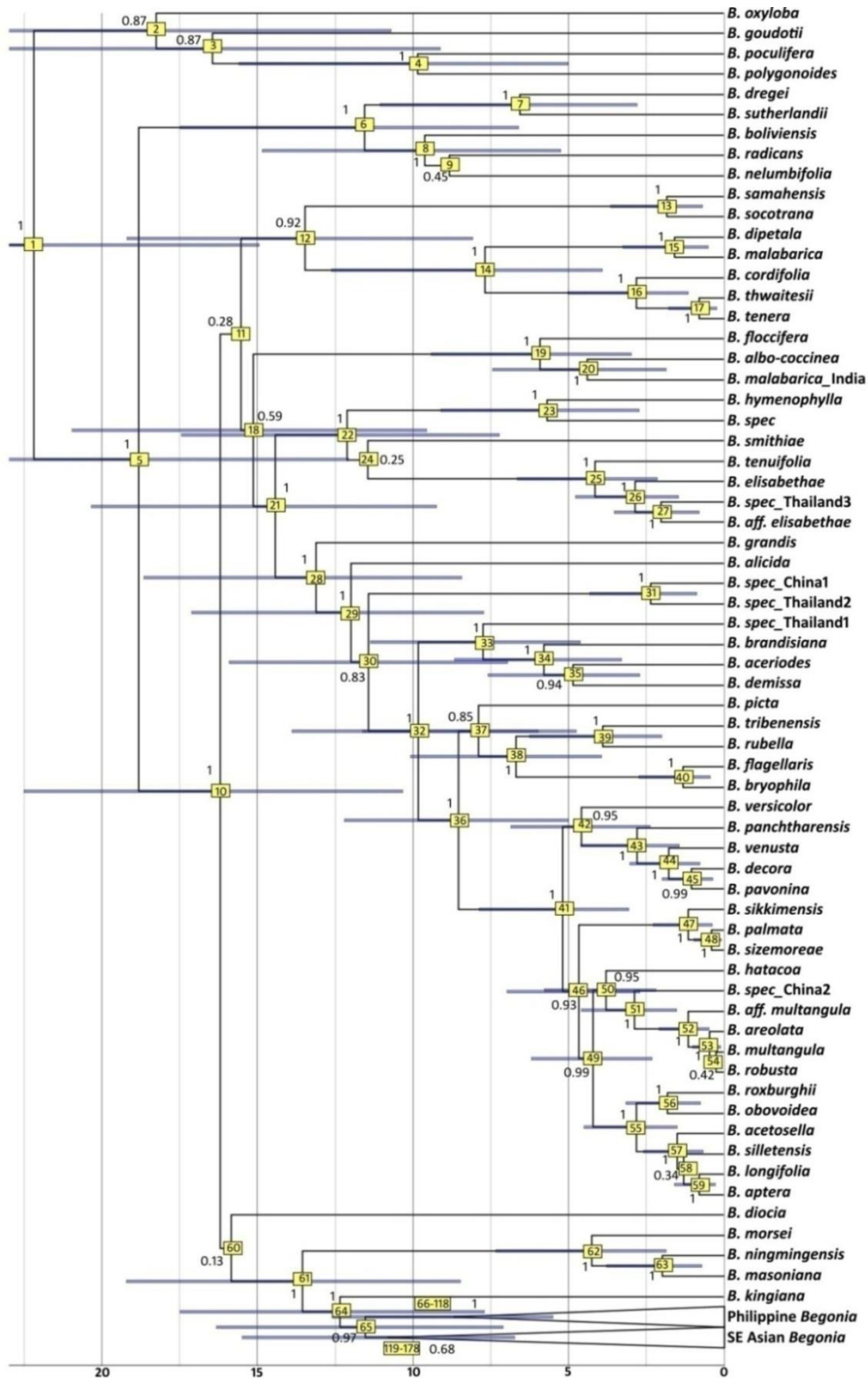
Figure 3.4 Maximum clade credibility chronogram of the three-region *Begonia* dataset. Node heights indicate mean ages. Node bars indicate 95% highest posterior density date ranges. Numbers inside boxes at each node represent node numbers and values next to nodes are the posterior probability values for each node.

The most probable ancestral areas for Asian *Begonia* clade II are India (D)=0.66, Sri Lanka (E)=0.14,Indochina (H)=0.1. My results in terms of tree topology best match my hypothesis of India being the point of entry to Asia. India *Begonia* did not resolve as a monophyletic sister group to rest of the Asian *Begonia*, but are paraphyletic at the base, thus potentially are the origin for all other Asian *Begonia*. However, these results depend on the rather poorly supported topology at the base of the clade, and should hence be interpreted as a hypothesis which needs further testing. This lack of resolution is probably be due to very rapid range expansion of Asian *Begonia* after initial colonization.

Clade III potentially has an Himalayan origin with probability values of Himalaya (F)=0.45, China (G)=0.2, Indochina (H)=0.11, however this is rather uncertain due to poor support at the base of the clade.

India is constructed as most probable ancestral area for clade IV containing Socotran and Indo Sri Lankan Begonia with probabilities of India (D)=0.62, Sri Lanka (E)=0.18, Socotra (C)=0.11. A dispersal to Socotra from India during the mid-Miocene is suggested, with the two Socotran endemics *B.socotrana* and *B. samahensis* arising from more recent Pleistocene speciation *in situ*.

The Indo Sri Lankan Clade V has an Indo Sri Lankan origin (India+Sri Lanka, DE=0.5) and two Sri Lankan endemics *B. tenera* and *B. thwaitesii* are the result of Pleistocene immigration from India followed by *in situ* speciation thereafter. The rest of the Indian species in Clade VIII have an Indian origin with area probabilities of India (D)=0.78, India+Sri Lanka(DE)=0.22.

The ancestral area for Continental Asian *Begonia* inClade IX is constructed as Indochina (H=0.97), which is also the ancestral area for Himalayan *Begonia* Clade X (Indochina, H=0.95). Within clade X, multiple dispersal events between China, Indochina, Peninsular Malaysia, Sumatra, Sulawesi and Himalaya during Miocene-Pliocene are inferred.

The geographic area of origin for Clade XIII is constructed as China during Mid-Miocene; China (G) = 0.38, Peninsular Malaysia (I) = 0.22, China+Sumatra (G+J) =

0.14 and the biogeographic history within this region is fully resolved by Thomas et al. (2012).

Table 3.5 Posterior probabilities, divergence ages and ancestral area probabilities; C= Socotra, D= India, E= Sri Lanka, F= Himalaya, G= China, H=Indochina, I= Peninsular Malaysia, J= Sumatra, S= Palawan, P= Luzon, R= Panay.

Clade No	PP	Divergence age	Ancestral Area Probability		
Clade I	1	16.2 (10.34-22.51)	D= 0.32	F=0.21	H= 0.1
Clade II	0.28	15.54	D= 0.66	E=0.14	H=0.1
Clade III	0.13	15.85	F= 0.45	G= 0.2	H= 0.11
Clade IV	0.92	13.48 (8.08-19.19)	D=0.62	E=0.18	C=0.11
Clade V	1	7.68 (3.91-12.62)	DE= 0.5	D= 0.3	E=0.2
Clade VI	1	2.82 (1.16-5.01)	DE= 0.67	E=0.31	
Clade VII	0.59	15.13 (9.57-20.95)	D= 0.69	H=0.15	E=0.12
Clade VIII	1	5.92 (2.98-9.42)	D= 0.78	DE= 0.22	
Clade IX	1	14.43 (9.24-20.34)	H= 0.97	G= 0.01	GH= 0.01
Clade X	1	9.83 (5.98-13.89)	H= 0.95	FH= 0.04	
Clade XI	1	8.54 (5-12.21)	F= 0.86	H= 0.05	FH= 0.07
Clade XII	1	5.19 (3.06-7.87)	F= 0.83	FH= 0.08	GH= 0.04
Clade XIII	1	13.55 (8.49-19.21)	G= 0.38	I= 0.22	GJ= 0.14
Clade XIV	1	8.72 (5.5-12.6)	S=0.36	P=0.34	R=0.19
Clade XV	0.58	10.82 (6.73-15.49)	J=0.99		

Discussion

Route of entry to Asia

The Afro-Arabian plate collided with Eurasia around 25 Mya during the late Oligocene (Yuan et al. 2005). The appearance of the Gulf of Aden rift caused the Arabian peninsula to become separated from continental Africa and Socotra ca. 10 Mya (Ghebreab 1998). My first hypothesis of Socotra being the point of route of entry to Asia, possibly facilitated by a mesic Arabian corridor, is rejected, as Socotran*Begonia* are not a sister group to all other Asian species, but nested within and likely to be of Indian origin. Further, evaporate and calcrete deposits suggest dry and warm conditions for areas of western Asia and the Arabian Peninsula during mid Miocene (Morley 2000;

Scotese 2003). Thus, overland migration of *Begonia* from Africa to Asia via dry corridor seems unlikely.

Since the separation of the Indian and African plates 132 mya (Ali & Aitchison 2008) Sri Lanka has remained static relative to India, together forming the Deccan plate. Sri Lanka is separated from India by the shallow, narrow Palk Strait. Although the Deccan plate could potentially have provided a refuge for African elements during its northward journey, in the case of *Begonia* these would have arrived by LDD given the age of the Asian clade. The hypothesis of Sri Lanka as the point of entry to Asia for *Begonia* is rejected, from the ancestral area reconstruction and also since Sri Lankan *Begonia* are not sister to other Asian species; indeed the Sri Lanka endemics are resolved as Pleistocene immigrants from India. There is no evidence for Himalaya to be the point of entry to Asia due to the young age of the Himalayan clade and the nesting of most Himalayan species within eastern Asian *Begonia*.

The point of entry of *Begonia* to Asia is obscure, since the topology at the base of the clade is poorly supported with PP=0.28. This lack of resolution is probably due to very rapid range expansion during the Miocene.

India harbours 61 native *Begonia* species (Rashid & Rahman 2012), however only 4 species (*B. albo coccinea, B. floccifera, B. malabarica, B. dipetala*) have been incorporated in the biogeograhic analysis due to practical difficulties in getting material from India. Indian *Begonia* belong to early diverging lineages, and lack of sampling may have contributed to the low back bone support in the phylogeny. If more Indian material could be incorporated in future analysis that would be beneficial in order to contribute towards a better resolved phylogeny due to breaking up some of the long early divergent branches in the tree.

Assuming the presented topology is correct, India is the most probable area for the route of entry to Asia and the initial colonizer could have undergone a very rapid radiation during favourable warm and moist conditions during Miocene in the Indian sub-continent (Yuan et al. 2005; Zachos et al. 2001). Subsequent dispersals then occurred to the islands Lanka and Socotra and to Southeast Asia via Continental Asia.

There is some evidence for a plant dispersal phase which occurred between Africa and India via Madagascar when the Deccan plate moved in close proximity to Africa during Cenomanian to Turonian times until Madagascar was separated from India and Seychelles about 90 Mya (Briggs et al. 2003; Conti et al. 2002 ; McLoughlin 2001). It is considered as one of the major inter plate dispersal paths for megathermal angiosperms and evidenced by plant families such as Sapindaceae, Palmae and Myrtaceae (Morley 2003; Morley 2000). This dispersal path was believed to cease when Madagascar was separated from India and Seychelles. Madagascar harbours 44 *Begonia* species, of which 42 are considered as endemic. Only one species from Madagascar, *B. goudotii* is incorporated in the analysis here. However, previous studies (Plana et al. 2003) have shown other Madagascan species to belong to a single clade. The Madagascan *Begonia*sampled here does not show a close affinity to Indo-Sri Lankan *Begonia* and the dates of entry are too young to support a dispersal pathway between Madagascar and Indo-Sri Lankan region for Begoniaceae.

The timing of entry of 15 Ma for Asian *Begonia* is highly congruent with the mid-Miocene climatic optimum peaking during 17-15 Mya when the global ice volume was low and bottom seawater temperatures were slightly higher (Zachos et al. 2001). Further, the moist and warm climates led to the expansion of megathermal vegetation in Asia up to Southern Japan in the north east and up to the northwest of the Indian subcontinent (Zachos et al. 2001; Morley 2000). The occurrence of rain forest taxa in the mid-Miocene fossil floras from Rajasthan in the northwest of the Indian peninsula confirms the existence of rain forests during that time (Lakhanpal 1970; Raven & Axelrod 1974). However, migration of *Begonia* across dry Arabian corridor is unlikely, thus long distance dispersals from Africa are more likely to have facilitated the middle Miocene arrival of *Begonia* into India and subsequent radiation thereafter.

Biogeography of Asian *Begonia*

Although the route of entry of *Begonia* into Asia remains somewhat speculative, a clearer picture of biogeographic range evolution of *Begonia* amongst the Asian hotspots has been resolved**.**

Early evolution

Based on the topology in Figure 3.3, the initial diversification of Asian *Begonia* occurred in India during the mid-Miocene followed by multiple dispersal events; two independent dispersals to Eastern continental Asia, a single dispersal to Sri Lanka and a single dispersal to Socotra. The warm phase during the mid-Miocene climatic optimum resulted in the expansion of mega thermal vegetation in Asia with spreading of rain forests throughout the continental Asia facilitate the east to west migration of *Begonia* to current hotspots under favourable climatic conditions. Further, dispersal from continental Asia to the Sunda Shelf region can potentially be explained by overland dispersal during the mid-Miocene when the moist and warm climates were predominant in South east Asia (Thomas et al. 2012).

Himalaya

The geological history of the Himalayas and the Tibetan Highlands is still is poorly understood and the timing of uplift and the evolution of regional climate remains enigmatic (Kutzbach et al. 1989, Molnar et al. 2010).

Two separate origins can be identified for Himalaya *Begonia*; firstly, clade XI where Himalayan species are highly nested within an eastern Asian *Begonia* grade with an area of origin optimised as Indo-China (F) (P=0.97) and the second consisting of the phylogenetically isolated *B. diocia* which is sister to Chinese and South East Asian *Begonia*, albeit with very low support.

The diversification of Himalayan *Begonia* sect *Diploclinium*, *B. picta, B. tribenensis, B. rubella, B. flagellaris* and *B. bryophila* (part of the 'Diploclinium Grade' of Rajbhandary et al. (2011) began 7.9 (4.8-11.4) Mya during late Miocene which is highly congruent with the onset of South Asian monsoons at 7.4 Mya (Copeland 1997). This is concurrent with the uplift of the Himalaya and the Tibetan plateau ca. 10 ma which reached sufficient altitudes to alter rain patterns in the Himalayan region (Guo et al. 2004; Molnar et al. 2010; Zheng et al. 2004). The *Diploclinium* grade is characterised by tuberous, seasonally adapted *Begonia* in which blooming is restricted to three months after the onset of monsoons followed by fruit dispersal during the following dry period (Rajbhandary et al. 2011).

The second major Himalayan grade composing *Begonia* sect *Platycentrum* and *Sphenanthera* intermixed with *Begonia* from China, Indochina, Peninsular Malaysia, Sumatra, Sulawesi, which diversified during 5.2 (3.1-7.9) Mya, indicating back and forth dispersal between the Himalaya and other regions. Continuing rapid uplift of the Tibetan plateau resulted in further intensification of East Asian monsoons. The species in this grade are evergreen, and have fruits adapted to rain-splash seed dispersal thus occupied in habitats where water is more constantly available (Rajbhandary et al. 2011). More recent speciation during the Pleistocene may have occurred in response to Pliocene-Pleistocene climatic cycles and changes in the monsoon intensity (Guo et al. 2004; Zheng et al. 2004).

The position of Himalayan *B. diocia* in the phylogeny is enigmatic and its placement at the base of the phylogeny is effectively unresolved (Figure 3.3). Further, the age of 15.8 Mya for the species predates the onset of South Asian monsoons during late Miocene, and whether the Himalaya had achieved sufficient height during that time to receive monsoon rain is uncertain. This relatively old lineage in Asian *Begonia* is further evidence for an area of origin in the region of the Indian Subcontinent.

Thus, most Himalayan *Begonia* are a result of back dispersal from Indo-china during late Miocene followed by further back and forth dispersal between the Himalaya and Indo-China during the Pliocene-Pleistocene which can be linked with climatic cycles during that time (Janssens et al. 2009). A similar scenario is observed in the genus *Impatiens* (Balsaminaceae), one of the most species rich genera of flowering plants with hotspots occurring in tropical Africa, the Himalayan region, Madagascar, South India and Sri Lanka and Southeast Asia. Himalayan species appear to have a Chinese origin with multiple separate dispersal events occurring during the late Miocene (Janssens et al. 2009).

Sri Lankan biogeography

The Deccan plate, comprising of India and Sri Lanka underwent remarkable climatic and vegetational changes during its northward journey to Eurasia (Conti 2002; McLoughlin 2001; Morley 2003). The plate would have experienced aridification as it moved north through the horse latitudes at 30-35 degrees where patterns of atmospheric circulation result in very little precipitation and the formation of the desert belt

(Lomolino et al. 1998), potentially causing a substantial amount of taxa to become extinct (Rutschmann & Eriksson 2004). The massive volcanism at the Cretaceous-Tertiary boundary during 65 Ma and extensive aridification during early Tertiary due to the uplift of Himalaya resulted in further impoverishment of allochthonous African elements from the Indian flora (Rutschmann & Eriksson 2004; Morley 2003; Conti 2002; McLoughlin 2001) leaving the Western Ghats and Sri Lanka as refugial areas for those flora.

The base of the Indo Sri Lankan clade V comprises of *B. dipetala* and *B. malabarica* and confirms a likely Indian origin for Sri Lankan *Begonia* with a more recent entry to Sri Lanka during the late Miocene ca. 7.7 Mya (3.9-12.6) which coincides with the development of South Asian monsoons during late Miocene 7.4 Mya (Copeland 1997). The intervening continental shelf between Sri Lanka and India provided an intermittent land connection until the last sea level rise 6000 years ago during the Holocene (McLoughlin 2001; Ashton & Gunatilleke 1987) and has potential to act as a migratory path for floristic exchange between the two land masses. However as the endemic Sri Lanka *Begonia*are restricted to altitudes of 1000-1200m, the overland migration via a lowland land bridge seems unlikely. The Western Ghats of India and the sub montane forests in Sri Lanka could have act as archipelago-like systems facilitating plant dispersals among them. The Sri Lankan endemics, *B. tenera* and *B. thwaitesii,* are the result of speciation *in situ*. Pleistocene deposits from the wet zone of Sri Lanka provides evidence for intermittent periods of seasonal tropical climate (Ashton & Gunatilleke 1987) which would favor in the recent radiation of tropical taxa in the island.

The monophyletic grouping of the Indian origin sample of *B. malabarica* with *B. albo-coccinea* is odd morphologically, since they have completely different life forms; *B. malabarica* is a woody cane-like species much more similar to *B. dipetala*, whilst *B. albo-coccinea* is a rhizomatous herb. This placement is best explained by hybridization resulting in chloroplast capture by the Indian *B. malabarica* lineage. The grouping of Sri Lankan *B. malabarica* with Sri Lankan *B. dipetala* is sensible and the preferred phylogenetic hypothesis for this species.

Incongruence between plastid and nuclear data sets can be mainly due to hybridization, however some factors such as recombination, gene paralogy and pseudogene formation

can result in incongruences between nuclear and plastid datasets (Feliner & Rosello 2007; Linder & Rieseberg 2004; Small et al. 2001). Further analysis with additional DNA markers and samples need to be carried out in order to resolve this uncertainty in the placement of *B. malabarica* from India and Sri Lanka.

The use of the nuclear ITS region might be a solution to examine this incongruence further, as it is bi- parentally inherited and can potentially track the paternal lineage, as shown in Hughes et al. (2015). Both nuclear and chloroplast genotypes can become introgressed in to other species resulting in interspecies hybridization and it is highly evidence in the species of *Begonia* sect *Baryandra.* This would result in a "plastid pool" where there is combination of different genotypes of neighbouring species would ultimately result in species with new genotypes (Hughes et al. 2015).

Socotra

The ancestral area reconstruction for Socotran *Begonia* and the lack of a sister relationship to other Asian *Begonia* favours an Indian origin (India (D)=0.62), thus long distance dispersal from India during the mid-Miocene is inferred. The conspicuous long branches suggest an extinction of taxa or long isolation of Socotran lineage during the Neogene drought towards the end of the Miocene (Yuan et al. 2005; Morley 2000) in the Arabian and Socotran region. These endemics are also the result of more recent Pliocene-Pleistocene speciation.

Exacum (Gentianaceae) which is distributed in Africa, Madagascar, Socotra, the Arabian peninsula, Sri Lanka, India, the Himalayas, mainland Southeast Asia and northern Australia, exhibits a similar biogeographic scenario to *Begonia* where long distance dispersal followed by extensive range expansion is found to be the most plausible reason for its extant distribution (Yuan et al. 2005). Sri Lanka and southern India were initially colonized by a Madagascan ancestor via long-distance dispersal and subsequently dispersed to Socotra-Arabia, northern India, and mainland Southeast Asia around the northern Indian Ocean Basin when the climate was warm and humid. The species were survived in isolation in refugia such as Socotra-Arabia, southern India–Sri Lanka, and perhaps mainland Southeast Asia during the Neogene drought and secondary diversification resulted in current distribution patterns (Yuan et al. 2005).

Pollination mechanisms, seed and fruit dispersal

Begonia species exhibits a low intraspecific gene flow, due to poor seed dispersal. Most *Begonia* are zoophilous, which are pollinated by generalized pollinators such as stingless bees (*Trigona* species), honey bees (*Apis cerana*), and bumble bees (*Bombus ephippiatus*) and flies (Hughes & Hollingsworth 2008; Kiew 2005; Dewitte et al. 2011).

Begonia female flowers do not offer rewards to pollinators, however they mimic the same colouration of yellow or orange colour of androecia of male flowers. Male flowers offer pollen for pollinators, and pollinators mistakenly visit female flowers and pollination can takes place (Renner 2006; Schemske et al. 1996). However some species in the neotropics such as *Begonia boliviensis* and *Begonia ferrugenia* are bird pollinationg species which possesses synpetalous, tubular perigones which is an adaptation for bird pollinating species (Hughes 2002). Zoochory is common in African *Begonia* and though has been directly observed in Asian *Begonia* (Tebbitt et al. 2006). However, some Asian sections like *Sphenanthera* and *Leprosa*possess indehiscent, thick, fleshy pericarps with bright colours which which is an indication for dispersal animals such as birds, bats and other vertebrates (Tebbitt et al. 2006).

However, majority of Asian species are anemochorous which produce large amount of minute seeds in large quantities. The capsules are dry with a membranous pericarp and the winged seeds are released through the slits when the capsules are shaken by the wind. The seeds produced are very tiny, size ranging from 300-600μm in length, which can be carried out by the wind effectively (de Lange & Bouman 1992; Kiew 2005). However, wind dispersal responsible only for short distance dispersals which is proven by limited gene flow and geographic isolation by most *Begonia* populations (Hughes &Hollingsworth 2008; Matolweni et al. 2000).

Further, some alternative dispersal mechanisms are seen in some Asian *Begonia* such as section *Platycentrum*. Their coriaceous pericarp, rain ballist capsule and unequal wings are characteriscs of a rain ballist syndrome. At the maturity pedicels get curved and the pericarp and smaller wings facing upwards while larger wing facing downwards. The rain drops cause the splash cups to move up and down thus allowing the seeds to get released from dehisced septa. The species in section *Parvibegonia* show bot different

mechanism, the capsules get dried and hand downward at maturity allowing the seeds to get dispersed by wind or rain (Tebbitt et al. 2006; Kiew 2005; Rajbhandry et al. 2011).

Thus most of the species are geographically isolated with narrow geographical ranges. Thus it is considered, long term isolation has led to more speciation than genetic drift between populations (Hughes & Hollingsworth 2008; Dewitte et al. 2011).

The genome of *Begonia* is highly diversed with large variation in chromosome length, width, total chromosome number and large secondary constrictions can be observed in those genotypes (Dewitte 2009a). Most of species possess chromosomes with 0.5-3.6μm in length and 0.3-1.5μm in width and the chromosomes of South American *Begonia* are smaller than that of Asian, African and Middle American *Begonia* (Dewitte 2009a).

A wide range of chromosome numbers has been reported, ranging from 2n=16 (*B. rex*) to 2n=156 (B *acutifolia*) (Doorenbos et al. 1998; Legro & Doorenbos 1969). It is suggested that X=13 (2n=26) is the basic chromosome number for *Begonia* across the world and others are a result of polyplodisation which should have been played a wide role in the diversification of lineages. Ployploides are either produced by means of somatic mutations in meristematic cells or unreduced gametes (Bretagnolle & Thompson 1995; Otto & Whitton 2000; Dewitte et al. 2011). Among 70 investigated genotypes, 10 produced unreduced gametes and this can be observed in both species and hybrids (Dewitte et al. 2009b). A high level of heterozygosity is transferred thorough 2n gametes and in some hybrids, viability only constrained within the 2n gametes. Thus, the polyploidy can be considered as a major factor contributing to diversification of lineages however it is considered that the frequency of occurrence is very low in the wild (Dewitte et al. 2011).

Mode of range expansion

My results confirm the importance of founder event speciation in *Begonia*, in which a daughter lineage jumps to an area completely outside the ancestral range immediately followed by speciation. The *J* parameter controls the probability of two events during cladogenesis; founder event speciation versus sympatric and vicariant speciation (Matzke 2014).

My present study confirms a recent origin (15.5 Mya) for Asian *Begonia* and a history dominated by a small number of long-distance dispersal events followed by extensive radiations in current hotspots. The youth of the Sri Lankan elements challenges the relict vicariant history for the evolution of the flora, and adds to a growing body of evidence supporting youthful tropics at the species level (Pennington et al. 2015). The early divergence of the Himalayan *B. dioica* is surprising and a hint at relictual elements in a relatively young montane landscape which begs further investigation.

References

Ashton, P.S. & Gunatilleke, C.V.S., 1987. New light on the plant geography of Ceylon I . Historical plant geography. *Journal of Biogeography*, 14(3), pp.249–285.

Ali, J.R. & Aitchison, J.C., 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166-35??Ma). *Earth-Science Reviews*, 88(3-4), pp.145–166.

Batalha-Filho, H. et al., 2014. Phylogeny and historical biogeography of gnateaters (Passeriformes, Conopophagidae) in the South America forests. *Molecular Phylogenetics and Evolution*, 79, pp.422–432.

Bretagnolle, F. & Thompson, J.D. (1995). Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autoploid plants. New Phytologist, Vol.129, No.1, (Januari 2005), pp. 1-22.

Briggs, J.C., Collections, Z. & Museum, G., 2003. The biogeographic and tectonic history of India.*Journal of Biogeography*, 30, pp.381–388.

Buerki S., et al,. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *J Biogeog*, 38, pp.531–550.

Burton, N., 2012. *Phylogenetics of Begonia sect. Gireoudia*. University of Edinburgh.

Copeland, P., 1997. The when and where of the growth of the Himalaya and the Tibetan Plateau. In: Ruddiman, W.F. (ed) Tectonic Uplift and Climate Change. New York: Plenum Press.

Clark, J.R. et al., 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Systematic biology*, 57(5), pp.693–707.

Clement, W.L. et al., 2004. Phylogenetic position and biogeography of Hillebrandia sandwicensis (Begoniaceae): A rare Hawaiian relict. *American Journal of Botany*, 91(6), pp.905–917.

Conti, E. et al., 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution; international journal of organic evolution*, 56(10), pp.1931–1942.

Crayn, D.M., Costion, C. & Harrington, M.G., 2014. The Sahul – Sunda floristic exchange : dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography*, 42, pp.11–24.

De Lange, A. & Bouman, F. 1992. Studies on BegoniaceaeIII. Seed micromorphology of the genus Begonia in Africa: Taxonomic and ecological implications. *Wageningen AgriculturalUniversity papers*: 1-82.

De Wilde, J.J.F.E. et al., 2011. Pliocene intercontinental dispersal from africa to Southeast Asia highlighted by the new species *Begoniaafromigrata* (Begoniaceae). *Taxon*, 60, pp.1685–1692.

Dewitte, A. et al., 2009a. Genome size variation in Begonia. *Genome*, 52 (10), pp. 829-838.

Dewitte, A. et al., 2009b. Occurrence of viable unreduced pollen in a *Begonia* collection. *Euphytica*, 168 (1), pp. 81-94.

Dewitte, A., Twyford, A.D., Thomas, D.C., Kidner, C.A. & Van Huylenbrock, J., 2011. Begonia: Genome Dynamism, Population Processes and Phylogenetic Patterns. *The Dynamical Processes of Biodiversity-Case studies of Evolution and Spatial Distribution*. pp 27-52.

Doorenbos, J., M. S. M. Sosef, and J. J. F. E. de Wilde . 1998 . *The sections of Begonia including descriptions, keys and species lists. Studies in Begoniaceae VI*. Wageningen Agricultural University Papers 98-2.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.214.

Feliner, G.N., & Rosello, J.A. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44:911-919.

Forrest, L.L., Hughes, M., and Hollingsworth, P. 2005. A phylo- geny of Begonia using nuclear ribosomal sequence data and morphological characters. Syst. Bot. 30(3): 671–682.

Ghebreab, W., 1998. Tectonics of the Red Sea region reassessed. *Earth Science Reviews*, 45, pp.1–44.

Goodall-Copestake, W.P. et al., 2010. The early evolution of the mega-diverse genus Begonia (Begoniaceae) inferred from organelle DNA phylogenies. *Biological Journal of the Linnean Society*, 101, pp.243–250.

Goodall-Copestake, W.P., Harris, D.J. & Hollingsworth, P.M., 2009. The origin of a mega-diverse genus: Dating Begonia (Begoniaceae) using alternative datasets, calibrations and relaxed clock methods. *Botanical Journal of the Linnean Society*, 159, pp.363–380.

Guo, Z. et al., 2004. Late Miocene-Pliocene development of Asian aridification as recorded in the Red-Earth Formation in northern China. *Global and Planetary Change*, 41(3-4), pp.135–145.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98.

Hughes, M. 2002. Population structure and speciation in *Begonia* L. PhD thesis. University of Glasgow, Glasgow.

Hughes, M. et al., 2015. Begonia Resource Centre.

Hughes, M. et al., 2015. The Miocene to Pleistocene colonization of the Philippine archipelago by Begonia sect. Baryandra (Begoniaceae). *American Journal of Botany*, 102, pp.695–706.

Hughes, M. & Hollingsworth, P.M., 2008. Population genetic divergence corresponds with species-level biodiversity patterns in the large genus Begonia. *Molecular Ecology*, 17(11), pp.2643–2651.

Janssens, S.B. et al., 2009. Rapid radiation of Impatiens (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. *Molecular Phylogenetics and Evolution*, 52(3), pp.806–824.

Judd, et al., 2008. *Plant Systematics: A phylogenetic approach*, 3rd ed. Sinauer Associates, Sunderland.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), pp.1647-1649.

Kiew, R. 2005. *Begonias of Peninsular Malaysia*. Natural History Publications (Borneo), Kota Kinabalu.

Kodandaramaiah, U., 2010. Use of dispersal-vicariance analysis in biogeography - A critique. *Journal of Biogeography*, 37, pp.3–11.

Kutzbach, J.E., et al., 1989. Sensitivity of climate to Late Cenozoic uplift in South-ern Asia and the American West: numerical experimentsJ. *Geophys. Res*, 94, pp.18 393–18 407.

Lakhanpal, R. N., 1970. Tertiary floras of India and their bearing on the historical geology of the region. *Taxon,* 19(5), pp.675-694.

Lamm, K.S. & Redelings, B.D., 2009. Reconstructing ancestral ranges in historical biogeography: Properties and prospects. *Journal of Systematics and Evolution*, 47(June), pp.369–382.

Landis, M.J. et al., 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, 62(0), pp.789–804.

Legro, RAH. & Doorenbos, J. (1969). Chromosome numbers in Begonia 1. *Netherlands Journal of Agricultural Sciences*, Vol.17, pp.

Linder, C.R., & Rieseberg, L.H. 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91: 1700-1708.

Lomolino, M.V., Riddle, B.R. & Brown, J.H., 1998. Biogeography, 3rd edn. Sinauer Associates, Sunderland, MA.

Maddison, W. P. & Maddison, D.R., 2015. Mesquite: a modular system for evolutionary analysis. Version 3.03 http://mesquiteproject.org.

Matolweni, L.O., Balkwill, K., & McLellam, T. 2000. Genetic diversity and gene flow in the morphology variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany,* 87,pp.431-439.

Matzke, N.J., 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, 63(6), pp.951–970.

McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49(3), pp.271–300.

Miller, M.A. et al., 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics,* 11, pp,43-48.

Molnar, P., Boos, W.R. & Battisti, D.S., 2010. Orographic Controls on Climate and Paleoclimate of Asia: Thermal and Mechanical Roles for the Tibetan Plateau. *Annual Review of Earth and Planetary Sciences*, 38(1), pp.77–102.

Moonlight, P.W. et al., 2015. Continental-scale diversification patterns in a megadiverse genus: the biogeography of Neotropical Begonia. *Journal of Biogeography*, 42(6), pp.1137–1149.

Morley, R. J., 2000. *Origin and evolution of tropical rain forests*. Chichester: John Wiley & Sons, Ltd.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(12), pp.5–20.

Myers, N. et al., 2000. Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), pp.853–8.

Otto, SP. & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics,* 34, pp. 401-437.

Pennington, R.T. & Dick, C.W., 2004. The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359(1450), pp.1611–1622.

Pennington, R.T., Hughes, M. & Moonlight, P.W., 2015. The Origins of Tropical Rainforest Hyperdiversity. *Trends in Plant Science*, 20(11), pp.693–695.

Plana, V. (2003) Phylogenetic relationships of the Afro-Mala- gasy members of the large genus Begonia inferred from trnL intron sequences. *Systematic Botany*, 28, pp.693–704.

Posada, D. et al., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, pp.772–772.

Rajbhandary, S., Hughes, M. & Shrestha, K.K., 2010. Distribution patterns of *Begonia* in the Nepal Himalaya. *Journal of Plant Science*, 7(0), pp. 73-78.

Rajbhandary, S. et al., 2011. Asian *Begonia*: out of Africa via the Himalayas? *Gardens' Bulletin Singapore*, 63, pp.277–286.

Rashid, M.E. & Rahman, M.A., 2012. Updated Nomenclature and taxonomic status of the plants of Bangladesh included in Hook. F., *THE FLORA OF BRITISH INDIA*: VOLUME-II. *Bangladesh J. Plant Taxon*, 19(2), pp.173-190.

Raven, P. H., & Axelrod, D.I., 1974. Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Gard*, 61,pp.539– 673.

Rambaut, A., 2009. FigTree, a graphical viewer of phylogenetic trees. *Institute of Evolutionary Biology University of Edinburgh*.

Ree, R.H. et al., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution; international journal of organic evolution*, 59(11), pp.2299–2311.

Ree, R.H. & Smith, S. A, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57(1), pp.4–14.

Renner, S.S. 2006. Rewardless flowers in the angiosperms and the role of insect cognition in their evolution, *Plant pollinator interactions: from specialization to generalization*. University of Chicago Press, Chicago.

Ronquist, F., 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(l), pp.195–203.

Ronquist, F. et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, pp.539–42.

Ronquist, F. & Sanmartín, I., 2011. Phylogenetic Methods in Biogeography. *Annual Review of Ecology, Evolution, and Systematics*, 42(1), pp.441–464.

Rutschmann, F. & Eriksson, T., 2004. Did Crypteroniaceae really disperse out of India? Molecular dating evidence from rbcL, ndhF, and rpl16 intron sequences. *International Journal of plant Science*, 165(S4), pp.S69–S89.

Schemske, D.W., et al. 1996. Decceit pollination in the monoecious, Neotropical herb *Begonia oaxacana* (Begoniaceae). *Floral biology: Studies on floral evolution in animal-pollinated plant*s: 292-318.

Scotese, C.R. 2003. PALAEOMAP Project. http://www.scotese.com/.

Small, R.L., Cronn, R.C., & Wendel, J.F. 2004. Use of nuclear genes for phylogeny reconstructions in plants. *Australian Systematic Botany* 17, 145-170.

Stults , D. Z. , &Axsmith, B.J., 2011. First macrofossil record of Begonia (Begoniaceae). *American Journal of Botany* 98 : 150 – 153 .

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), pp.2688–2690.

Tebbitt, MC.; Lowe-Forrest, LL.; Santoriello, A.; Clement, WL. & Swensen, SM. 2006. Phylogenetic relationships of Asian *Begonia*, with an emphasis on the evolution of rain-ballist and animal dispersal mechanisms in sections *Platycentrum, Sphenanthera* and *Leprosae*. *Systematic botany*, 31(2), pp. 327-336.

Thomas, D.C., 2010. Phylogenetics and historical biogeography of Southeast Asian *Begonia* L. (Begoniaceae). PhD thesis. University of Glasgow, Glasgow.

Thomas,D.C.,Hughes,M.,Phutthai, T.,Rajbhandary,S.,Rubite, R., Ardi, W.H. & Richardson, J.E., 2011a. A non-coding plastid DNA phylogeny of Asian *Begonia* (Begoniaceae): evidence for morphological homoplasy and sectional polyphyly. *Molecular Phylogenetics and Evolution*, 60, pp. 428–444.

Thomas, D.C. et al., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, 39(1), pp.98–113.

Yuan, Y.-M. et al., 2005. Phylogeny and biogeography of exacum (gentianaceae): a disjunctive distribution in the Indian ocean basin resulted from long distance dispersal and extensive radiation. *Systematic biology*, 54(1), pp.21–34.

Zachos, J. et al., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (New York, N.Y.)*, 292(April), pp.686–693.

Zheng, H. et al., 2004. Late Miocene and mid-Pliocene enhancement of the East Asian monsoon as viewed from the land and sea. *Global and Planetary Change*, 41(3-4), pp.147–155.

CHAPTER 4: Biogeographic history of Sri Lankan Sapotaceae

Introduction

Sapotaceae are a pantropical tree family, mostly distributed in tropical rain forests, but some extending to semi-arid and arid regions. The highest species richness is recorded in the tropical and subtropical regions of Asia and South America (Swenson et al. 2014). The family is composed of 53 genera and about 1100 species (Pennington 1991; Swenson et al. 2014; Anderberg 2005; Bartish et al. 2005; Bartish et al. 2011) It has a wide economic utility such as timber (*Manilakara bidentata*, *M. huberi*, *M. obovata* and *M. kauki*), chewing gum (*Manilkara zapota*), edible fruit (sapodilla plum, *Manilkara zapota*; star apple, *Chrysophyllum cainito*) and in the cosmetics industry; *Vitellaria paradoxa* (Pennington 1991; Govaerts et al.2001; Mathews 2009).

The most recent classification by Pennington (1991) divides the family into three subfamilies Chrysophylloideae, Sapotoideae, and Sarcospermatoideae and five tribes: Chrysophylleae, Isonandreae, Omphalocarpeae, Mimusopeae, and Sideroxyleae.

Sarcospermatoideae is a small subfamily with a narrow distribution in South and Southeast Asia. It comprises of *Sarcosperma* and possibly *Eberhardtia* (Smedmark et al. 2006), two small genera which are sisters to the large subfamilies Chrysophylloideae and Sapotoideae. Chrysophylloideae is composed of 600 species of shrubs and medium understory to giant canopy trees of rainforests in Africa, Australasia, and South America, although the numbers are likely to increase as a result of ongoing phylogenetic and taxonomic research (Swenson et al. 2008). Sapotoideae is a pantropical subfamily comprising 543 species with the highest species richness recorded in Indo Pacific region and the rest occurring in Africa and America (Smedmark et al. 2006).

Higher taxa within Sapotaceae have overlapping morphological variation making the generic delimitation within the problematic. The Tribe Mimusopeae is distinguished by having tripartite corolla lobes, petaloid staminodes and a basal-basiventral seed scar, which is further divided in to three subtribes; Mimusopinae, Manilkarinae and Glueminae (Pennington 1991).

Several chloroplast and nuclear markers have been widely utilised in the phylogeny and classification of Sapotaceae. The chloroplast region *ndh*F was utilized in the construction of the first Sapotaceae phylogeny; however the monophyly of the tribe Mimusopeae was not supported due to lack of variability with the use of a single chloroplast marker (Anderberg & Swenson 2003). However, later with the incorporation of morphological data a more resolved phylogeny was obtained and the family was divided into three subfamilies; Sarcospermatoideae, Sapotoideae and Chrysophylloideae. Sapotaceae were confirmed to have evolved in two separate lineages; Isonandrae-Mimusopeae-Syderoxyleae and Chrysophylleae-Omphalocarpeae. Among these tribes, Sideroxyleae are monophyletic, Isonandreae are polyphyletic while Mimusopeae are paraphyletic (Swenson & Anderberg 2005)

Later, additional chloroplast non-coding intergenic spacers; *trn*H-*psb*A*, trn*C-*Trn*D*, trn*C-*psb*M, *psb*M-*trn*D were utilized by Smedmark et al. (2006) in an analysis of subfamilial level relationships in Sapotoideae. *Eberhardtia aurata* is the sister for two major sub clades; Sideroxyleae and Sapoteae and within Sapotoideae, Sideroxyleae and Sapoteae, each were strongly supported as monophyletic. Polyphyly of the Isonandrae was confirmed; however it was poorly supported due to lack of sampling in the phylogeny (Smedmark et al. 2006).

A phylogeny of New Caledonian Sapotaceae was constructed by Bartish et al. (2005) using nuclear ribosomal DNA and the results conflicted with existing phylogenies. Some genera like *Pouteria, Niemeyera*were found to be non monophyletic while unrelated genera confined to Australia, New Caledonia and neighbouring islands were clustered in single clades (Bartish et al. 2005). This non monophyly was further confirmed by Swenson et al. 2007, with the addition morphological data and increased taxon sampling. Nuclear DNA provided useful information on species level relationships, however fail to obtain relationships at deeper nodes (Swenson et al. 2007).

Later, phylogenetic relationships within the*Niemeyera* complex and generic delimitations were investigated by Swenson et al. (2008) by utilizing both nuclear and chloroplast markers. Nuclear internal transcribed data (ITS) were combined with ETS

(External Transcribed Spacers) which resulted in better resolution (Swenson et al. 2008).

The intergenic spacers; trnD^{GUC}–trnT^{GGU}, trnS^{GCU}–trnG^{UUC}, rpoB–trnC^{GCA} and trnS^{UGA}–$trnf$M^{CAU} and trnH-psbA were utilized in the same study, however they resulted in a poorly resolved phylogeny. Among the chloroplast markers tested, trnH-psbA and trnS-trnG proven to be useful in resolving phylogenetic relationships at low levels whereas rpoB-trnC and trnD-trnT showed less variability, thus less informative. All the recent work based on morphological and different molecular markers (both nuclear and chloroplast data) reveal the need of a new classifications for Sapotaceae (Swenson et al. 2008)

Throughout the past decades, biogeographers have struggled to explain tropical intercontinental disjunctions and vicariance was accepted as the major driving force for many years with examples cited in families like Dipterocarpaceae (Ashton & Gunatilleke 1987) and Monimiaceae (Renner et al. 2010). However, with the development of dated phylogenies and ancestral area reconstruction methods, many tropical families such as Begoniaceae (Thomas et al. 2012) have been shown to have a more recent origin demonstrating the importance of long distance dispersal in determining distributions of plant genera. This has also been demonstrated in many other groups, such as *Araucaria* and *Nothofagus* (Setoguchi et al. 1998; Swenson et al. 2001; Cook & Crisp 2005).

The family Sapotaceae is an ideal candidate to study the biogeographic history of tropical families, with its pantropical distribution and the fact that it occupies various habitats. Several biogeographic analysis have been carried out for different groups within the family (Chrysophylloideae; Bartish et al. 2011, Isonandreae; Richardson et al. 2014, *Manilkara*; Armstrong et al. 2015), however a wider biogeographic analysis across the family has not been carried out so far.

The dated phylogeny based on previously published data by Swenson et al.(2008) demonstrated the an early diversification sub-family Chrysophylloideae in Africa during the late Cretaceous approximately 73-83 Ma (Bartish et al. 2011). A single vicariance event between South America and Australia is postulated. Multiple long distance

dispersals from Africa to the Neotropics, between Australia and New Caledonia and between Africa and Madagascar and a single dispersal back to Africa from the Neotropics are inferred, thus long dispersal events are considered as the major driving force for current disjunctions of Chrysophylloideae across the tropics (Bartish et al. 2011).

Smedmark & Anderberg (2007) suggest a northern hemisphere origin for Sideroxyleae in the early tertiary followed by extensive extinctions due to increasingly unfavourable climatic conditions, specifically decreases in temperature. The distributions of extant species in the new world may have been facilitated via the North Atlantic Land Bridge during the early Eocene and this is supported by the occurrence of Sapotaceae pollen in North America and Europe (Smedmark & Anderberg 2007; Bartish et al. 2011; Morley 2000). The pantropical distribution of *Manilkara* (Sapotoideae) has also been shown, through the generation of a dated phylogeny, to be the result of long distance dispersal events rather than vicariance caused by continental break-up (Armstrong et al. 2014). Nuclear ITS, plastid *rpl*32-*trn*L, *rpl*16-*trn*K and *trn*S-*trn*F were utilized in the study, however plastid data were not incorporated in the BEAST analysis due to less variability and less sampling (Armstrong et al 2014)

Sri Lanka is a key locality for understanding patterns in tropical plant biogeography. It is part of the Indian continental plate that was part of Gondwana. In Sri Lanka, Sapotaceae contains 27 species from all tribes and 25 species occur in the low or upland wet zone. The highest number of species is recorded in the tribe Isonandreae with species from *Palaquium, Madhuca, Isonandra, Payena* and *Diploknema*. All the nine species of *Palaquium*, five out of seven species of *Madhuca* and three out of four *Isonandra* present in the country are endemic thus making it an important family to look for phylogenetic and biogeographic relationships of the Sri Lankan flora. The Isonandreae are also an ideal group to study diversification patterns in continental Asia and South East Asia due to the diversity in India, Indochina, Sunda, Australia and Pacific islands (Richardson et al. 2014).

The purpose of this study is to determine the age and biogeographic affinities of Sri Lankan representatives of Sapotaceae by placing them in a global phylogenetic context.

I aimed to test the following hypothesis on origin and the diversification of Sri Lankan flora.

The breakup of the supercontinent Gondwana initiated during early Jurassic 180 million years ago (Mya), followed by subsequent continental drifts and sea floor spreading, was considered one of the major causes for tropical plant disjunctions we observe today. The Deccan plate comprised of India and Sri Lanka was a part of Gondwana that began to separate from Antarctica from 132 Mya followed by northward migration until collision with Asia c. 35 Mya. Being a part of Gondwana, the Sri Lankan flora might be composed of Gondwanan relicts (Ashton & Gunatilleke 1987) . However to be of Gondwana origin the disjunction should be reflected in phylogenetic splits occurring approximately 132 Mya-94 Mya, concurrent with continental break-up.

The collision of Deccan plate with southern Laurasia from c. 35 Mya is accepted as the most marked possibility for interplate dispersal during the Tertiary (Morley 2000). Thus, the Sri Lankan species could also have arrived via the immigration of Laurasian lineages through Asia and India, resulting in the mixing of the Deccan Gondwanan flora with Laurasian lineages. For this hypothesis to be accepted, we would expect that divergence times for lineages should not be older than 35 Mya and that Indo-Sri Lankan lineages would be nested within Laurasian ones.

Thirdly, long distance dispersal could have played a major role in the assemblage of the Sri Lanka flora from other land masses such as Africa and/or South East Asia. Long-distance dispersal could have occurred at any point in time, it is the only viable scenario for tropical disjunctions younger than ~33 Mya and Indo-Sri Lankan lineages would be nested within African and/or South East Asian ones.

Finally, a potentially important contributor to the Sri Lankan flora will have been diversification of lineages within the island (*in situ* speciation), resulting in the evolution of new species. Species endemic to Sri Lanka with sister species from Sri Lanka are considered as representing in situ speciation. More recent Pliocene-Pleistocene speciation may have resulted from changing climatic conditions during glacial/inter-glacial periods.

Thus, I aimed to estimate the relative contributions to the Sri Lankan flora of Gondwanan relicts, overland immigration from nearby landmasses (Asia), long distance trans-oceanic dispersal from Southeast Asia, and long distance dispersal from Africa and in situ speciation with the use of a representative plant family Sapotaceae.

Spatial relationships within Isonandreae from Sri Lanka were examined with a fine scale geographic coding for the Asian Sapotaceae in order to obtain a better understanding of biogeography within the continental Asia which was not addressed in previous Sapotaceae studies.

Materials and methods

Plant material and Taxon sampling

Both nuclear and chloroplast data were generated in my study. The ITS data set comprised 163 accessions and chloroplast data set comprised 136 accessions with species from all tribes of Sapotaceae. The sequences of a selection of taxa representing tribes Sideroxyleae and Mimusopeae were included in order to provide fossil calibration points within each tribe (Richardson et al. 2014; Armstrong et al. 2014).

Sarcosperma was used as the outgroup, as it is the sister to other lineages of Sapotaceae in previous studies (Anderberg & Swenson 2003). In order to complement South Asian elements lacking in previous studies, 24 species from Sri Lanka (100% Sri Lankan representatives) were added representing five genera *Palaquium, Isonandra, Madhuca, Manilkara, Mimusops*. Twenty four sequences were generated for the ITS matrix. The chloroplast matrix comprised 136 sequences and 86 of these sequences were newly generated (23 Sri Lankan representatives+ 63DNA extractions from EDNA bank) for *trn*H-*psb*A, *trn*C-D and *ndh*F. Voucher specimens of the newly generated ITS sequences and chloroplast sequences are listed in table 4.1 and table 4.2 respectively. Voucher information of all the samples used in the present study are listed in appendix 3.

Table 4.1 Voucher specimens of the newly generated ITS sequences

EDNA number	Species	Origin	Voucher
EDNA13-0032656	*Palaquium_laevifolium_*SL Engl.	Sri Lanka	Kumarage 60 (E)
EDNA13-0032663	*Isonandra_lanceolata_*SL Wight	Brunei, Indonesia, Borneo, Sri Lanka	Kumarage 47 (E)
EDNA13-0032654	*Madhuca_sp_*SL	Sri Lanka	Kumarage 58 (E)
EDNA13-0032657	*Madhuca_fulva_*SL J.F.Macber	Sri Lanka	Kumarage 62 (E)
EDNA13-0032658	*Madhuca_neriifolia_*SL H.J.Lam	Sri Lanka	Kumarage 63 (E)
EDNA13-0032661	*Madhuca_longifolia*2_SL J.F.Macber	Sri Lanka	Photo voucher available
EDNA13-0032648	*Palaquium_thwaitesii_*SL Trim.	Sri Lanka	Kumarage 43 (E)
EDNA13-0032650	*Palaquium_canaliculatum_*SL Engl.	Sri Lanka	Kumarage 45 (E)
EDNA13-0032653	*Palaquium_hinmolpedda_*SL P.Royen	Sri Lanka	Kumarage 64 (E)
EDNA13-0032655	*Palaquium_pauciflorum_*SL Engl.	Sri Lanka	Kumarage 59 (E)
EDNA13-0032660	*Palaquium_rubiginosum_*SLEngl.	Sri Lanka	Kumarage 65 (E)
EDNA14-003582	*Isonandra_montana_*SL Gamble	Sri Lanka	Kumarage 76 (E)
EDNA14-003579	*Isonandra_sp1_*SL	Sri Lanka	Kumarage 97 (E)
EDNA14-003580	*Palaquium_sp3_*SL	Sri Lanka	Kumarage 57 (E)
EDNA13-0032651	*Palaquium_grande_*SL Engl.	Sri Lanka	Kumarage 46 (E)
EDNA13-0032649	*Palaquium_petiolare_*SL Engl.	Sri Lanka	Kumarage 44 (E)
EDNA14-003581	*Palaquium_sp5_*SL	Sri Lanka	Kumarage 93 (E)
EDNA14-003578	*Isonandra_sp2_*SL	Sri Lanka	Kumarage 74 (E)
EDNA13-0032659	*Isonandra_compta_*SL Dubard	Sri Lanka	Emanuelsson 3039(S)
EDNA14-003583	*Isonandra_zeylanica_*SL	Sri Lanka	Kumarage 72 (E)
EDNA13-0032662	*Manilkara_hexandra_*SL Dubard	Sri Lanka	Photo voucher available

EDNA14-003584	*Mimusops_elengi_*SL L.	Sri Lanka	Kumarage 98 (E)
EDNA14-003585	*Madhuca_clavata_*SL Jayas.	Sri Lanka	Kumarage 99 (E)
EDNA14-003577	*Palaquium_sp*4_SL	Sri Lanka	Kumarage 75 (E)

Table 4.2 Voucher information for newly generated chloroplast sequences (*trn*H-*psb*A, *trn*C-*trn*D, *ndh*F)

EDNA number	Species name	Origin	Voucher
EDNA09-02183	*Madhuca elmeri* Merr. Ex. H.J.Lam	Indonesia, Kalimantan	Wilkie P1 347 (E)
EDNA09-02184	*Madhuca laurifolia* H.J.Lam	Peninsular Malaysia	Wilkie 843 (E)
EDNA09-02185	*Madhuca motleyana* Baehni	Peninsular Malaysia	Wilkie 837 (E)
EDNA09-02303	*Palaquium rostratum 2* Burck	Indonesia, Kalimantan	Slik CMF9452 (L)
EDNA09-02304	*Palaquium sericeum* H.J.Lam	Indonesia, Kalimantan	Slik CMF9737 (L)
EDNA09-02305	*Payena maingayi 2* C.B.Clarke	Peninsular Malaysia	Wilkie 841 (E)
EDNA09-02306	*Payena lucida 2*A.d.Candolle	Borneo	Ambri et al. AA1604 (L)
EDNA09-02307	*Madhuca sp.*	Peninsular Malaysia	Wilkie 834 (E)
EDNA09-02309	*Payena lucida 3*A.d.Candolle	Peninsular Malaysia	Wilkie 845 (E)
EDNA09-02312	*Palaquium microphyllum* King & Gamble	Malaysia	Pennington, Kochummen & Wong (K)
EDNA09-02313	*Payena obscura* Burck	Peninsular Malaysia	Wilkie 880 (E)
EDNA09-02316	*Madhuca kingiana 2* H.J.Lam	Malaysia, Sarawak	Wilkie 856 (E)
EDNA09-02317	*Payena leerii* Kurz	Malaysia	Wilkie 811 (E)
EDNA09-02318	*Palaquium obovatum* H.J.Lam	Thailand	Middleton 4387 (E)
EDNA09-02319	*Palaquium sp.*	Solomon Islands	Poulsen et al. 2488 (E)
EDNA09-02320	*Palaquium sp.2*	Indonesia, Sulawesi	Armstrong 370 (E)
EDNA09-02325	*Madhuca sp.*	Peninsular Malaysia	Wilkie 834 (E)
EDNA09-01452	*Palaquium sericeum* H.J.Lam	Indonesia, Kalimantan	Slik CMF9737 (L)

EDNA09-00716	*Palaquium clarkeanum* King & Gamble	Malaysia	Wilkie 501 (E)
EDNA09-00717	*Palaquium gutta* Baillon	Malaysia	Wilkie 504 (E)
EDNA09-00718	*Palaquium obovatum* Engler	Thailand	Middleton 4387 (E)
EDNA09-00719	*Palaquium oxleyanum* Pierre	Malaysia	Wilkie 527 (E)
EDNA09-00721	*Palaquium xanthochymum* Pierre ex Burck	Malaysia, Borneo	Wilkie 544 (E)
EDNA09-00722	*Palaquium formosanum* Hayata	Taiwan,Philippines	Chung & Anderberg 1421 (HAST)
EDNA09-00723	*Palaquium amboinense 3* Burck	Indonesia (native to New Guinea)	Wilkie 813 (E)
EDNA09-00952	*Palaquium beccarianum* van Royen	Malaysia,Papua, Borneo	Wijesundara s.n. (K)
EDNA09-00953	*Payena leerii* Kurz	Malaysia	Wilkie 811 (E)
EDNA09-00954	*Madhuca kingiana* 1 H.J.Lam	Malaysia, Sarawak	Wilkie 856 (E)
EDNA09-00955	*Palaquium sumatranum* Burck	Indonesia, Java	Wilkie 823 (E)
EDNA09-00956	*Palaquium rostratum 1* Burck	Indonesia, Kalimantan	Slik CMF9452 (L)
EDNA09-00957	*Palaquium maingayi* King & Gamble	Peninsular Malaysia	Wilkie 841 (E)
EDNA09-00958	*Madhuca malaccensis* H.J.Lam	Peninsular Malaysia	Wilkie 832 (E)
EDNA09-00959	*Payena obscura* Burck	Peninsular Malaysia	Wilkie 880 (E)
EDNA09-00987	*Palaquium galactoxylum* H.J.Lam	Australia	Bartish and Jessup 9 (S)
EDNA09-00990	*Pouteria maclayana* Baehni	Indonesia,Malaysia,PNG	Armstrong 316 (E)
EDNA09-00992	*Palaquium calophyllum* Pierre ex Burck	Indonesia, Kalimantan	Wilkie et al 1/477 (E)
EDNA09-00993	*Palaquium eriocalyx* H.J.Lam	Indonesia	Wilkie et al. 8/147 (E)
EDNA09-01129	*Palaquium rigidum* Pierre ex Dubard	Malaysia	Wilkie 878 (E)
EDNA09-01130	*Palaquium leiocarpum 1* Boeriage	Malaysia, Sarawak	Wilkie 870 (E)

EDNA09-01131	*Palaquium pseudorostratum* H.J.Lam	Malaysia, Sarawak	Wilkie 857 (E)
EDNA09-01132	*Palaquium ridleyi* King & Gamble	Malaysia, Sarawak	Wilkie 858 (E)
EDNA09-01133	*Palaquium rufolanigerum* P.Royen	Malaysia, Sarawak	Wilkie 859 (E)
EDNA09-01134	*Madhuca kuchingensis* Yii & P.Chai	Malaysia, Sarawak	Wilkie 860 (E)
EDNA09-01135	*Palaquium walsurifolium* Pierre ex Dubard	Malaysia, Sarawak	Wilkie 877 (E)
EDNA09-01136	*Madhuca sarawakensis*H.J.Lam	Sarawak, Malaysia	Wilkie 863 (E)
EDNA09-01137	*Madhuca erythrophylla* H.J.Lam	Sarawak, Malaysia	Wilkie 867 (E)
EDNA09-01138	*Madhuca kunstleri* H.J.Lam	Malaysia, Sarawak	Wilkie 868 (E)
EDNA09-01140	*Palaquium calophyllum* Pierre ex Burck	Indonesia, Kalimantan	Wilkie et al 1/477 (E)
EDNA09-01141	*Palaquium herveyi* King & Gamble	Malaysia, Sarawak	Wilkie 871 (E)
EDNA09-01142	*Palaquium hexandrum* Baillon	Malaysia, Sarawak	Wilkie 872 (E)
EDNA09-01143	*Madhuca barbata* T.D.Penn.	Sarawak, Malaysia	Wilkie 873 (E)
EDNA09-01144	*Palaquium cryptocariifolium* P.Royen	Malaysia, Sarawak	Wilkie 874 (E)
EDNA09-01378	*Madhuca sericea* H.J.Lam	Sarawak, Malaysia	Wilkie 879 (E)
EDNA09-01379	*Madhuca oblongifolia* Merrill	Sarawak, Malaysia	Wilkie 861 (E)
EDNA09-01380	*Madhuca korthalsii* H.J.Lam	Sarawak, Malaysia	Wilkie 876 (E)
EDNA09-01381	*Madhuca proxila* Yii & P.Chai	Sarawak, Malaysia	Wilkie 875 (E)
EDNA09-01382	*Madhuca lancifolia* H.J.Lam	Sarawak, Malaysia	Wilkie 853 (E)
EDNA09-01383	*Madhuca sp. nov*	Peninsular Malaysia	Wilkie 834 (E)
EDNA09-01384	*Palaquium amboinense 2* Burck	Indonesia (native to New Guinea)	Wilkie 813 (E)
EDNA09-01386	*Madhuca pachyphylla* (K.Krause) ined.	Indonesia, West Papua	Armstrong 313 (E)
EDNA09-01388	*Burckella polymera* P.Royen	Indonesia, West Papua	Armstrong 326 (E)

EDNA09-01389	*Pouteria firma* (Miq.) Baehni	Indonesia, West Papua	Armstrong 305 (E)
EDNA09-01451	*Palaquium quercifolium* Burck	Indonesia, Kalimantan	Slik CMF6780 (L)
EDNA13-0032656	*Palaquium_laevifolium_SL*	Sri Lanka	Kumarage 60 (E)
EDNA13-0032663	*Isonandra_lanceolata_SL* Wight	Brunei, Indonesia, borneo, Sri Lanka	Kumarage 47 (E)
EDNA13-0032654	*Madhuca_sp_SL*	Sri Lanka	Kumarage 58 (E)
EDNA13-0032658	*Madhuca_neriifolia_SL* H.J.Lam	Sri Lanka	Kumarage 63 (E)
EDNA13-0032661	*Madhuca_longifolia2_SL* J.F.macber	Sri Lanka	Photo voucher available
EDNA13-0032648	*Palaquium_thwaitesii_SL* Trim.	Sri Lanka	Kumarage 43 (E)
EDNA13-0032650	*Palaquium_canaliculatum_SL* Engl.	Sri Lanka	Kumarage 45 (E)
EDNA13-0032653	*Palaquium_hinmolpedda_SL* P.Royen	Sri Lanka	Kumarage 64 (E)
EDNA13-0032655	*Palaquium_pauciflorum_SL* Engl.	Sri Lanka	Kumarage 59 (E)
EDNA13-0032660	*Palaquium_rubiginosum_SL* Engl.	Sri Lanka	Kumarage 65 (E)
EDNA14-003582	*Isonandra_montana_SL* Gamble	Sri Lanka	Kumarage 76 (E)
EDNA14-003579	*Isonandra_sp1_SL*	Sri Lanka	Kumarage 97 (E)
EDNA14-003580	*Palaquium_sp3_SL*	Sri Lanka	Kumarage 57 (E)
EDNA13-0032651	*Palaquium_grande_SL* Engl.	Sri Lanka	Kumarage 46 (E)
EDNA13-0032649	*Palaquium_petiolare_SL* Engl.	Sri Lanka	Kumarage 44 (E)
EDNA14-003581	*Palaquium_sp5_SL*	Sri Lanka	Kumarage 93 (E)
EDNA14-003578	*Isonandra_sp2_SL*	Sri Lanka	Kumarage 74 (E)
EDNA13-0032659	*Isonandra_compta_SL* Dubard	Sri Lanka	Emanuelsson 3039 (S)
EDNA14-003583	*Isonandra_zeylanica_SL* Jeuken	Sri Lanka	Kumarage 72 (E)

EDNA13-0032662	*Manilkara_hexandra_SL* Dubard	Sri Lanka	Photo voucher available
EDNA14-003584	*Mimusops_elengi_SL* L.	Sri Lanka	Kumarage 98 (E)
EDNA14-003585	*Madhuca_clavata_SL* Jayas.	Sri Lanka	Kumarage 99 (E)
EDNA14-003577	*Palaquium_sp4_SL*	Sri Lanka	Kumarage 75 (E)

DNA extraction, amplification and sequencing

Total genomic data was extracted from silica gel dried material using the DNeasy Plant Mini Kit (Qiagen, UK) according to the manufacturer's protocols. 25 μl PCR reactions were setup for amplification of both ITS and chloroplast regions.

Nuclear ITS rDNA was amplified using ITS 5p and ITS 8p. Each 25μl of PCR mixture contained 5.75μl of ddH$_2$O, 2.5μl of 10x reaction buffer, 1.25μl of 25mM MgCl$_2$, 2.5 μl of 2mM dNTPs, 0.75μl of 10μM forward primer (ITS5p), 0.75μl of 10μM reverse primer (ITS8p), 10μl Betain, 0.25 μl of 0.4% BSA, 0.25μl of Biotaq DNA polymerase (Bioline, UK) and 1μl of DNA template.

The PCR temperature profile was, template denaturation at 95°C for 5 min followed by 34 cycles of denaturation at 95°C for 30 Sec, primer annealing at 50°C for 30 sec, primer extension at 72°C for 1.5 min followed by a final extension step at 72°C for 8 min.

Chloroplast data were obtained for three regions. The *trn*H-*psb*A spacer was amplified using the primers described by Hamilton (1999). The *trn*C-*trn*D region (consisting of the *trn*C- *pet*N spacer, the *pet*N gene, the *pet*N-*psb*M spacer, the *psb*M gene, and the *psb*M-*trn*D spacer) was amplified in two segments; the *trn*C-*psb*M region with the *trn*C (Demesure et al. 1995) and psbM2R (Lee & Wen 2004) primers, and the *psb*M-*trn*D spacer with the psbM1 (Lee & Wen 2004) and trnD (Demesure et al. 1995) primers. The 3´end of *ndh*F was amplified with primers ndhF5 and ndhF10 (Olmstead & Sweere 1994). The primer sequences are listed in Table 4.3.

For the *trn*H-*psb*A region, 25μl of PCR mixture contained 15.25μl of ddH$_2$O, 2.5μl of 10x reaction buffer, 1.25μl of 25mM MgCl$_2$, 2.5 μl of 2mM dNTPs, 0.75μl of 10μM forward primer, 0.75μl of 10μM reverse primer, 0.8μl of 0.4% BSA, 0.2μl of Biotaq DNA polymerase (Bioline, UK) and 1μl of DNA template. The temperature profile

included an initial template denaturation step of 96°C for 5 minutes, followed by 34cycles of denaturation at 96°C for 45 s, primer annealing at 53°C for 1 min, primer extension at 72°C for 30 sec and a final extension step at 72°C for 5 min

For the *ndh*F region, 25μl of PCR mixture contained 15.55μl of ddH$_2$O, 2.5μl of 10x reaction buffer, 1.25μl of 25mM MgCl$_2$, 1 μl of 2mM dNTPs, 1 μl of 10μM forward primer (ndhF5), 1μl of 10μM reverse primer (ndhF10), 1.5μl of 0.4% BSA, 0.2μl of Biotaq DNA polymerase (Bioline, UK) and 1μl of DNA template. The temperature profile included an initial template denaturation step of 95°C for 5 minutes, followed by 29 cycles of denaturation at 95°C for 45 s, primer annealing at 51°C for 45 sec, primer extension at 72°C for 3 min and a final extension step at 72°C for 10 min.

For the *trn*C-D region (consisting of the *trnC–petN* spacer, the *petN* gene, the *petN–psbM* spacer, the *psbM* gene, and the *psbM–trnD* spacer), 25μl of PCR mixture contained 15.25μl of ddH$_2$O, 2.5μl of 10x reaction buffer, 1.25μlof 25mM MgCl$_2$, 2.5μl of 2mM dNTPs, 0.75μl of 10μM forward primer, 0.75μl of 10μM reverse primer, 0.8μl of 0.4% BSA, 0.2 μl of Biotaq DNA polymerase (Bioline, UK) and 1μl of DNA template. The temperature profile included an initial template denaturation step of 94°C for 5 minutes, followed by 34 cycles of denaturation at 94°C for 1min, primer annealing at 52°C for 2 min, primer extension at 72°C for 2 min and a final extension step at 72°C for 2 min.

Table 4.3 Nuclear (ITS) and chloroplast primer sequences used in Sapotaceae DNA sequence generation.

Region	Primer name	Primer sequences	References
ITS	ITS5p	GGAAGGAGAAGTCGTAACAAG	Moeller & Cronk (1997)
	ITS8p	CACGCTTCTCCAGACTACA	Moeller & Cronk (1997)
*trn*H-*psb*A	trnH	ACTGCCTTGATCCACTTGGC	Hamilton (1999)
	psbA	CGAAGCTCCATCTACAAATGG	Hamilton (1999)
*trn*C-*psb*M	trnC	CCAGTTCAAATCTGGGTGTC	Demesure et al. (1995)
	psbM2R	TTCTTGCATTTATTGCTACTGC	Lee & Wen (2004)
*psb*M-*trn*D	psbM1	GCGGTAGGAACTAGAATAAATAG	Lee & Wen (2004)
	trnD	GGGATTGTAGTTCAATTGGT	Demesure et al. (1995)
*ndh*F	ndhF5	GTCTCAATTGGGTTATATG	Olmstead et al. (1994)
	ndhF10	CCCCCTA(CT)ATATTTGATACCTTCTC	Olmstead et al. (1994)

Amplified products were run on a 1% Agarose TBE gel with Syber safe as the staining agent and visualized in UV transiluminator. The PCR purification was done using EXOSAP IT as 7 μl reaction mixtures. 5μl of PCR product was mixed with 2μl of EXOSAP and incubated at 37°C for 15 minutes followed by 80° C for 15 minutes.

Sequencing PCR was done using purified PCR products as 10μlmixtures using 5.68μl of ddH$_2$O, 2μl of sequencing buffer, 0.32μl of primer, 1μl of Big dye and 1μl of template. The sequencing PCR protocol was, denaturation at 95°C for 30 sec, followed by 24 cycles of primer annealing at 50°C for 20sec, extension at 60°C for 4 min. Separate forward and reverse sequencing PCR were carried out and products were sent to the Genepool facility at the University of Edinburgh (Genepool, UK) for BigDye Terminator Cycle sequencing.

Sequence editing and alignment

Newly generated DNA sequences were edited in GeneiousR7 (7.1.4) (Kearse et al. 2012) and aligned manually in BioEdit 7.1.3 (Hall 1999) and checked for indels. Some regions were excluded from the ITS dataset due to ambiguous alignments or missing data at the region ends; 105-115, 156-163, 229-233, 335-344, 795-796, 803-807, and 1025-1036. Ambiguous alignment regions 1330-1350, 1770-1796, 4012-4026 and 5445-5463 were excluded from the chloroplast dataset.

Phylogenetic analysis

Both chloroplast and nuclear datasets were analyzed separately under Bayesian inference (BI) using MrBayes 3.2.1 (Ronquist et al. 2012) in the CIPRES science gateway V. 3.3 (Miller et al. 2010) and treated as single partitions. MCMC runs were carried out for 10,000,000 generations and sampled every 1000 generations. A 25% burn-in was set to discard the first set of trees and the remaining trees were summarised as a 50% majority rule consensus tree, visualized in FigTree (Rambaut 2009)and checked for hard incongruence between ITS and chloroplast trees. Plastid data did not provide a fully resolved tree and could not be combined with the ITS data due to well supported incongruence. Further, the ITS data gives more resolution than the chloroplast data and combination with exsisting ITS data comprising a high species number allowed better taxon sampling. Lower sample size in the plastid data would

reduce the reliability of biogeographic reconstructions; thus only the ITS data were used in the divergence time estimation and ancestral area reconstructions.

Models of sequence evolution were determined using jModeltest 2.1.3 (Posada et al. 2012). Maximum likelihood topologies were used to estimate the optimal evolutionary model and twenty four models were tested under the Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC).

Bayesian divergence time estimation

Bayesian divergence time estimation was performed for nuclear ITSDNA alignment using BEAST v.1.8.0 (Drummond & Rambaut 2007). GTR was selected as the nucleotide substitution model with a gamma distribution of rates among sites and a proportion of invariant sites. The data set was treated as a single partition and uncorrelated relaxed lognormal clock model was selected in order to relax the assumption of a molecular clock and allow for rate heterogeneity between lineages. The tree prior was set to random birth death speciation process, with a randomly generated starting tree.

Four separate Markov Chain Monte Carlo (MCMC) runs were carried out for 10,000,000 generations, sampling every 1000 generations, under a Birth-Death model of speciation. Plots of the logged parameters for each run were visualised using Tracer v.1.5 (Drummond & Rambaut 2007) to confirm convergence between runs. Time series plots of all parameters were analyzed in Tracer v.1.5 (Drummond & Rambaut 2007) to check for convergence and to confirm adequate effective sample sizes (ESSs). Trees were combined in LOGCOMBINER (Drummond & Rambaut 2007) with the burn-in set to 25% for initial sample for each run and a single maximum clade credibility (MCC) tree was obtained from Tree Annotator v.1.7.5 (Drummond & Rambaut 2007) and visualized in fig tree v.1.4.0 (Rambaut 2009).

The tree was calibrated using fossils at three nodes. Sideroxyleae pollen from the early Eocene of England dated at 47.8-56 Mya (Gruas-Cavagnetto 1976) was used to constrain the minimum age of the Sideroxyleae crown node. One of the criticisms of fossil based calibrations is that they only provide minimum age estimates, as the fossil could have been formed after the age for which the clade it represents was formed. By

constraining the crown node instead of the stem node I bias in favour of older age estimates. A log normal prior was used to constrain the age, with an offset of 52.2 Mya and a mean of 0.001. A Mid-Eocene (37.2–48.6Mya) *Tetracolporpollenites* pollen grain from the Isle of Wight was used to constrain the age of the crown node of the tribe Mimusopeae. This pollen grain closely resembles *Tieghemella heckelii*, a monotypic genus in Mimusopeae, and was used to constrain the age of the tribe Mimusopeae with an offset of 42.9 and mean of 0.095 so that the 95% probability limits lie within the midpoint of 42.9 Mya and the upper boundary of the mid Eocene (48.6 Mya).The final calibration point is based on a series of Oligocene (23–33.9 Mya) fossil leaves from Ethiopia (Jacobs et al. 2005) which was placed at the *Manilkara* crown node with an offset of 28 Mya and mean of 0.1.

Area delimitation and Ancestral Area Reconstructions

Eleven areas were coded based on the extant distribution, areas of endemism and geological history. In order to obtain a clear picture of South Asian biogeography, India, Sri Lanka and Himalaya were coded as separate geographic areas. East Asian elements were composed of an area East of Himalaya combined with China and Indochina. The Middle East element was composed of Iran, Turkey and the Arabian Peninsula which belongs to the Irano-Turanian flora (Takhtajan 1986).In Southeast Asia, the Sahul and Sunda Shelves were coded as separate states within the Malesian floristic region, which stretches from the Isthmus of Kra on the Malay Peninsula to Fiji (Takhtajan 1986; Van Welzen et al. 2005). Madagascan species were coded separately while species from Reunion, Comoros, Canary Islands and Cape Verde islands were coded as Africa. Seychelles and America were also assigned two geographic regions and all areas were coded according to their placement on different tectonic plates and the existence of distinct floras. Thus, the coded areas are (1) America; (2) Africa; (3) Madagascar; (4) Seychelles; (5) Middle East; (6) India; (7) Sri Lanka; (8) Himalaya; (9) East Asia; (10) Sunda Shelf; (11) Sahul Shelf.

The data matrix was prepared coding for presence/ absence in each of the areas depending on the collection details, voucher specimens and current distributions. The maximum number of areas in ancestral ranges was constrained to two to avoid wide ancestral ranges and excessive analysis time. Thus, all taxa were coded as present either

in a single area or two. *Mimusops elengi* which occurs in India, Sri Lanka and Burma and *Isonandra lanceolata* which occurs in India, Sri Lanka and Borneo were coded as present in India and Sri Lanka; the samples were collected from the latter and the region has the highest number of collections for the species.

Ancestral areas within internal nodes were constructed using Biogeobears (Landis 2013) in the R package under four models; DIVA like (Ronquist 1997), DIVA LIKE+J, DEC (Ree & Smith 2008), DEC+J. In addition to two free parameters; d (dispersal), e (extinction) included in DEC model (Batalha-Filho et al. 2014, Landis 2013), the new additional parameter "J" is added to the model to account for founder event speciation, which is not addressed by the models in DIVA and DEC. Log likelihood values for each model were compared and the model with the highest value was chosen as the best model for inferring ancestral ranges at nodes.

Results

Phylogenetic Analysis-ITS

The consensus tree resulting from chloroplast data has a very low clade support and the placement of some major taxa in the phylogeny is taxonomically doubtful. Thus, only nuclear ITS phylogeny is carried towards the results and discussion.

Table 4.4 Descriptive statistics of ITS and plastid data

DNA region	Amplicon length	Alignment length	Number of variable characters	Number of informative characters
ITS	1036	1024	562 (54.88%)	399 (38.96%)
Chloroplast	5463	5444	1082 (19.87%)	423 (7.77%)

The maximum clade credibility tree resulting from Bayesian analysis is shown in Figure 4.2. It comprises of a basal grade including the outgroup *Sarcosperma*, *Eberhardtia*, a grade of *Sideroxylon*, *Xantolis*, *Englerophytum*, *Omphalocarpum*, *Pouteria*, *Neolemonniera*, *Lecomtedoxa*, *Northia*, *Capurodendron*, *Inhambanella* within which is nested a large clade of all other species (clade III).

The Clade III is weakly supported as monophyletic (PP=0.23) and comprises tribes Mimusopeae and Isonandreae. An early diverging clade of *Baillonella, Vitellaria* and *Vitellariopsis* is resolved (clade IV; PP=1). Clade V, containing all Mimusopinae species examined, it is not supported as monophyletic (PP=0.47). The Manilkarinae clade (Clade VI) comprising the genera *Manilkara, Labramia, Faucherea* and *Labourdonnaisi*s strongly supported as monophyletic (PP=1.0). Clade VII containing the *Manilkara* grade is strongly supported as monophyletic (P=0.95) and *Letestua durissima* is nested within.

The Isonandreae (clade VIII) is highly supported as monophyletic (pp=1) and comprises of sub-clades *Palaquium* (Clade IX, PP=1), *Isonandra* (Clade XIII, PP=1) and clade XV (PP=1) containing *Madhuca, Burckella, Diploknema* and *Payena* species.

Palaquium (clade IX) is not monophyletic, with *Aulandra longifolia* and *Diploknema butyracea* nested in it. Within Clade IX, clade X is strongly supported as monophyletic (PP=1) with *P. impressionervium* from west of Sunda Shelf and *Palaquium petiolare* endemic to Sri Lanka being the sister clade to Sri Lankan endemic *Palaquium* clade. All other Sri Lanka *Palaquium* species are strongly supported as monophyletic (PP=1). The clade XI (PP=1) comprises of all other *Palaquium* species, dominated by species from Sunda Shelf, with *Palaquium* species from east of the line and China, and *Aulandra longifolia* and *Diploknema oligomera* nested within.

The genus *Isonandra* (Clade XIII) is not monophyletic and instead forms a grade within which *Madhuca utilis* and *Madhuca crassipes* are nested. The Sri Lankan endemic *Isonandra montana* is sister to the other species in the strongly supported clade XIII (PP=1) and with other Sri Lankan species nested within sub clade XIV with high support (PP=1).

Clade XVI is moderately supported (PP=0.87) with *Madhuca hainanensis* from China and Indo Sri Lankan *Madhuca longifolia* being the sister clade to other taxa. The Sri Lankan endemic *Madhuca clavata* and *Burckella* species are nested within it.

Within sub clade XVII, *Diploknema* is the sister to *Payena* clade (PP=0.52). All the *Payena* species (Payena clade) are strongly supported as monophyletic (PP=1).

The sub clade XVIII is strongly supported (PP=1) as monophyletic and is dominated by *Madhuca* species from Sunda Shelf with and contains clade XIX comprising *inter alia* Sri Lankan endemic *Madhuca* with a moderate support (PP=0.86).

Divergence time estimates

Mean ages with 95% HPD confidence intervals for key nodes and area probabilities are reported in Table 4.5. The MCC tree from the BEAST analysis (Figure 4.2) resolves the mean crown age of the tribe Mimusopeae as 43.8 (42.9-46.2) Mya in the Mid Eocene.

The mean age of sub tribe Mimusopinae (Clade V) is estimated to be 35.9 Mya and Manilkarinae (Clade VI) is 32.9 (29.4-58) Mya, both having originated during the Oligocene. The origin for genus *Mimusops* is 10 (5.5-16) Mya while *Mimusops elengi* show a more recent origin during the late Miocene 6.8 (3.5-11.1) Mya. The genus *Manilkara* (Clade VII) is resolved as 28.8 (28.0-31) Mya, with Indo Sri Lankan *Manilkara hexandra* diverging from19.0 (11.3-25.5) Mya during the early Miocene.

Table 4.5 Posterior probabilities, divergence ages and ancestral area probabilities; A= America, B= Africa, C= Madagascar, D= Seychelles, E= Middle East, F= India, G= Sri Lanka, H= Himalaya, I= East Asia, J= Sunda Shelf, K= Sahul Shelf.

Clade No	Clade Name	PP	Divergence time (Mya)	Area probabilities
I	Sapotaceae	1	92.49 (70.56-119.64)	BI=0.43, BJ=0.43, IJ=0.11
II	*Sideroxylon* clade	1	70.33 (60.11-82.75)	B=0.52, AB=0.45, A=0.03
III		1	43.5 (42.94-44.9)	B=0.95
IV	*Vitellaria* clade	1	34.47(23.54-43.12)	B=1.00
V	Mimusopinae	0.47	35.86	B=1.00
VI	Manilkarineae	1	32.9 (29.37-58)	B=0.86, BC=0.13
VII	*Manilkara*		28.81 (28.05-30.94)	B=0.99
VIII	Isonandreae	1	50.97 (43.54-59.76)	J=0.82, G=0.18
IX	*Palaquium*	1	42.86 (33.86-52.46)	J=0.71, GJ=0.16, G=0.13
X		1	31.93 (22.88-41.43)	G=0.57, J=0.28, GJ=0.15
XI		1	29.6 (20.82-40.88)	J=1.00
XII	*Madhuca/Isonandra /Payena*	1	46.31 (38.5-54.45)	J=0.78, G=0.16, GJ=0.06
XIII	*Isonandra*	1	36.91 (27.76-47.25)	J=0.42, G=0.38, GJ=0.20
XIV		0.36	28.49	G=0.61, J=0.34, GJ=0.03
XV	*Madhuca*	1	38.79 (31.77-46.09)	J=0.91, G=0.05, GJ=0.02

XVI		0.87	33.14 (25-40.02)	J=0.46, G=0.38, GJ=0.06
XVII	*Diploknema/Payena*	0.52	34.66 (26.76-42.25)	J=0.83, IJ=0.06, HJ=0.06
XVIII	*Madhuca*	1	27.47 (20.89-35.14)	J=1.00
XIX		0.86	8.37 (4.16-13.3)	J=0.97, GJ=0.01, JK=0.01

The mean crown age for Isonandreae (Clade VIII) is 51 (43.5-59.8) Mya and the divergence time estimate for *Palaquium* clade (Clade IX) is 42.9 (33.7-52.5) Mya, during the middle Eocene. The Sri Lankan *Palaquium* shows an initial origin 31.9 (22.9-41.4) Mya and diversified into two lineages, endemic *P. petiolare* during 27.1 (17.9-38) Mya and the second lineage giving rise to endemic taxa 16.4 (10.7-23.1) Mya during the mid-Miocene. The clade XI shows an initial origin during the Oligocene 29.6 Mya and further a rapid diversification during the mid-Miocene.

The basal taxon of clade XIII, Sri Lankan endemic *Isonandra montana* shows an early divergence 36.9(27.8-47.2) Mya while the second lineage (Clade XIV) diverged during 28.5 Mya followed by more recent speciation of endemics 5.3 (2.0-10.8) Mya.

The crown node age for clade XVI is 33.1 (25-40.0)Mya with Indo Sri Lankan *Madhuca longifolia* shows an origin 27.4 (17.3-36) Mya and endemic *Madhuca clavata* 31.5 Mya. The clade XIX comprises of Sri Lankan endemic *Madhuca* species that began to diversify from 11.5 (4.2-13.3) Mya followed by speciation of endemics more recently during the Pleistocene.

Ancestral area reconstruction and intercontinental dispersal events

Among the models compared in BIOGEOBEARS; DEC, DEC+J, DIVA LIKE, DIVA LIKE+J, the DEC+J model resulted in higher likelihood value, thus chosen as the best fit model for my data and the results are given in Table 4.6.

Table 4.6 d (dispersal), e (extinction), j (j value, founder-event speciation) LnL (log likelihood) for each of the geographic range evolution models compared in Biogeobears.

Model	d	e	j	LnL
DEC	0.0018	0.0034	0	-284.17
DEC+J	8e-04	0	0.0104	-264.49
DIVA LIKE	0.0019	0.0014	0	-290.50
DIVA LIKE+J	8e-04	0	0.0098	-271.96

The tribe Mimusopeae, sub-tribe Manilkarinae, Mimusopineae and the genera *Tieghemella, Autranella, Mimusops, Manilkara, Labramia* and *Faucherea/Labourdonnaisia* are all inferred to have African ancestry. *Mimusops* shows an initial origin in Africa during the late Eocene and long distance dispersals have occurred to Madagascar and Indo Sri Lankan region during the late Miocene. *Manilkara* shows a similar pattern of diversification which began in Africa during the late Eocene, and Indo Sri Lankan *Manilkara hexandra* and Eastern Asian *Manilkara kauki* shows an African origin, during the early Miocene. Neotropical *Manilkara* also have an African ancestry, which dispersed to South America during early Miocene 21.61 (12.97-27.68) Mya.
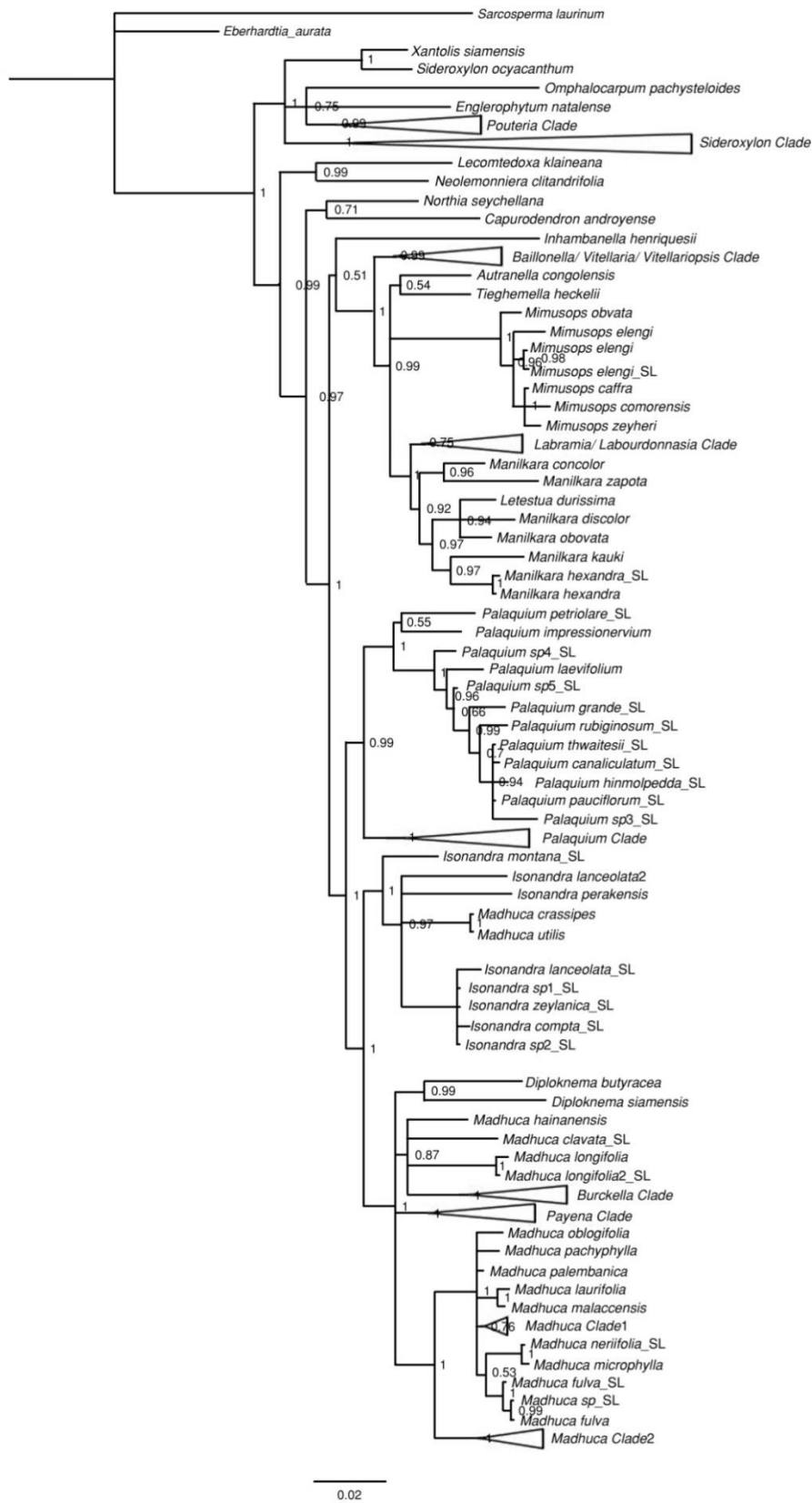
Figure 4.1. Bayesian majority rule consensus tree based on nuclear ITS data. Bayesian posterior probability (PP) support values are indicated next to the nodes.
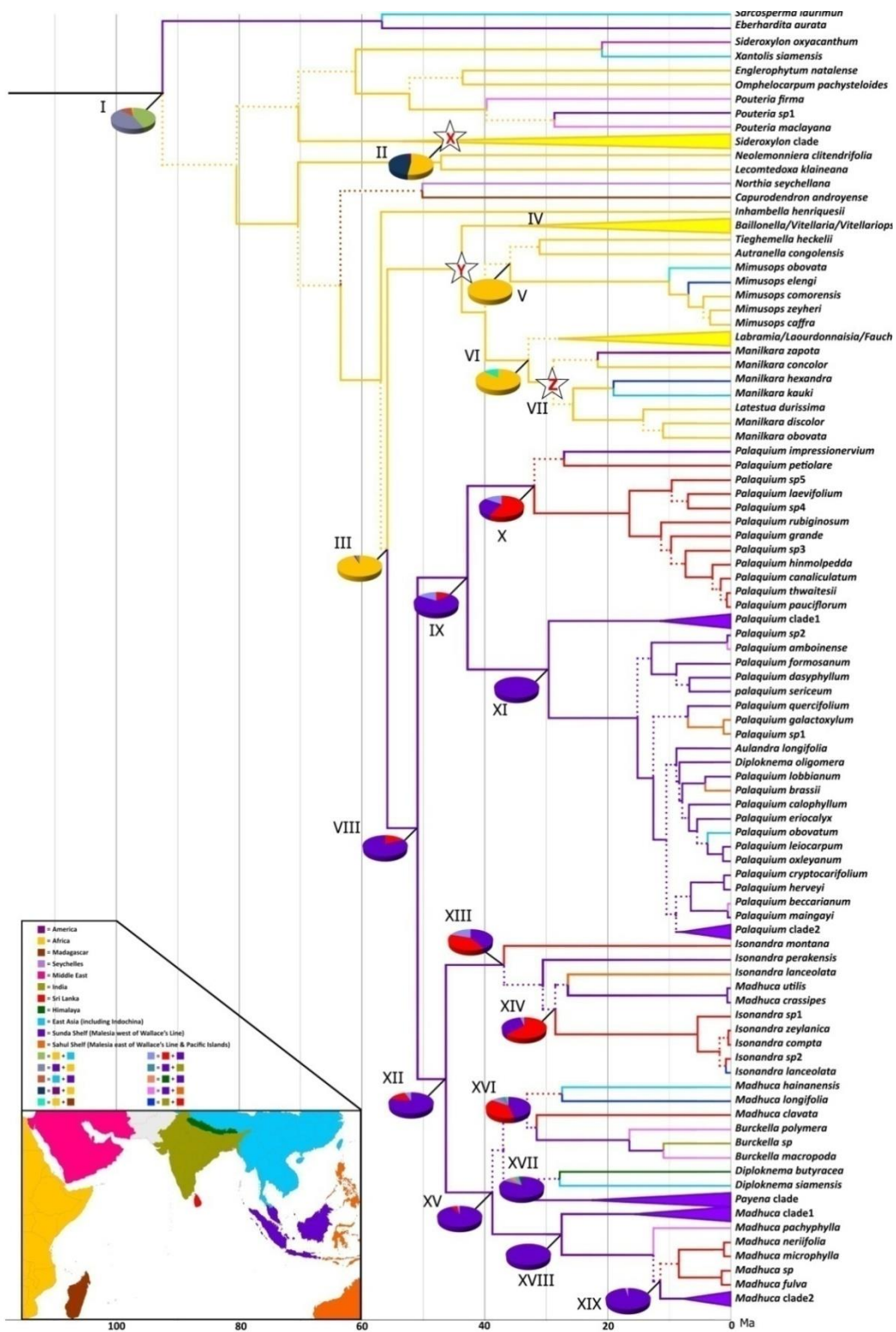
Figure 4.2 Maximum-clade-credibility chronogram of a Beast analysis of the ITS region Sapotaceaedata set. Node heights indicate mean ages. Numbers at nodes represent clades in Table 4.2. Branches coloured according to their optimal range reconstructions under the DEC+J model in the pacakge Biogeobears. Pie charts show the relative probability of ancestral state reconstructions at selected nodes. Dotted lines indicate posterior clade probabilities less than 0.95.
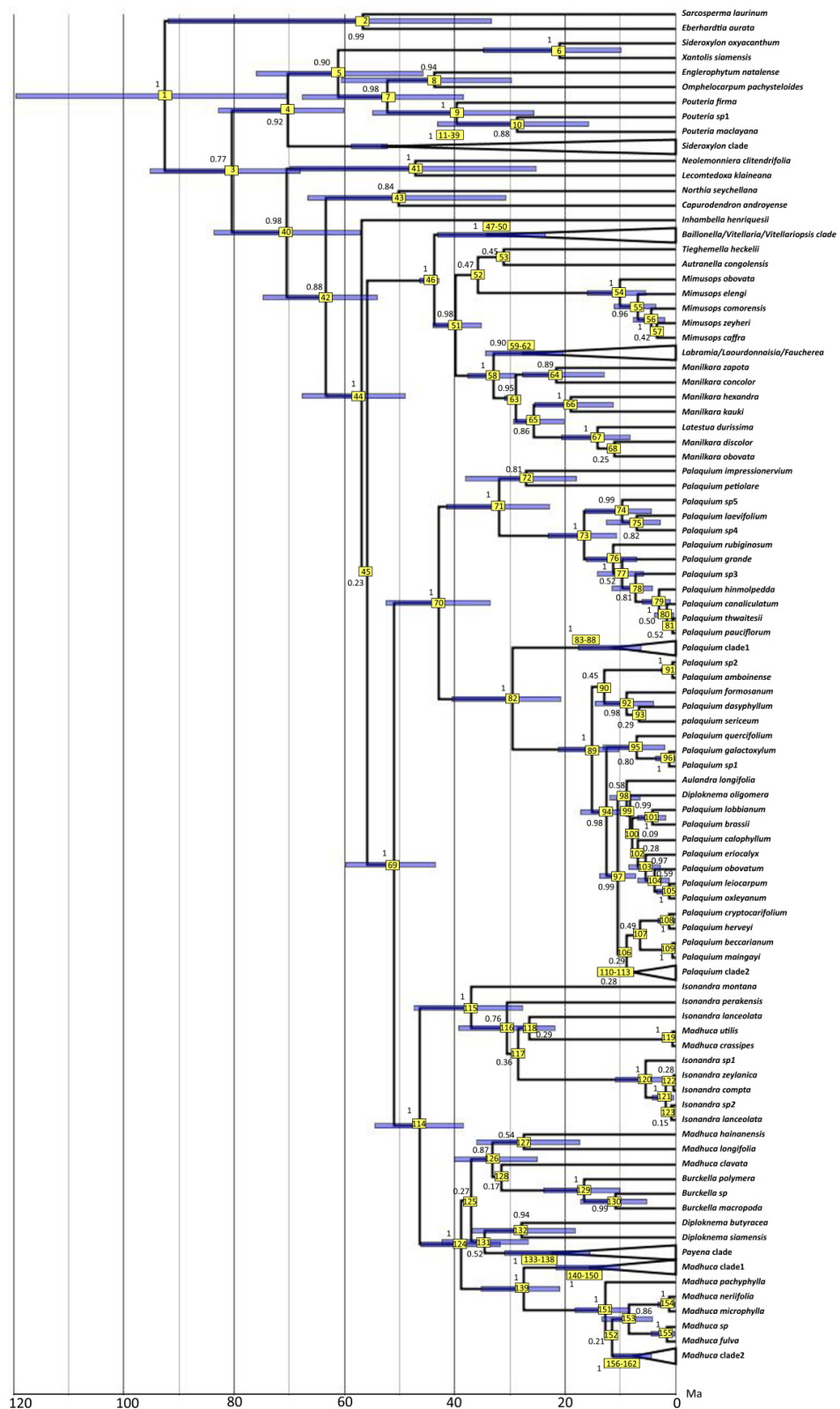
Figure 4.3 Maximum clade credibility chronogram of Sapotaceae ITS dataset. Node heights indicate mean ages. Node bars indicate 95% highest posterior density date ranges. Numbers inside boxes at each node represent node numbers and values next to nodes are the posterior probability values for each node.
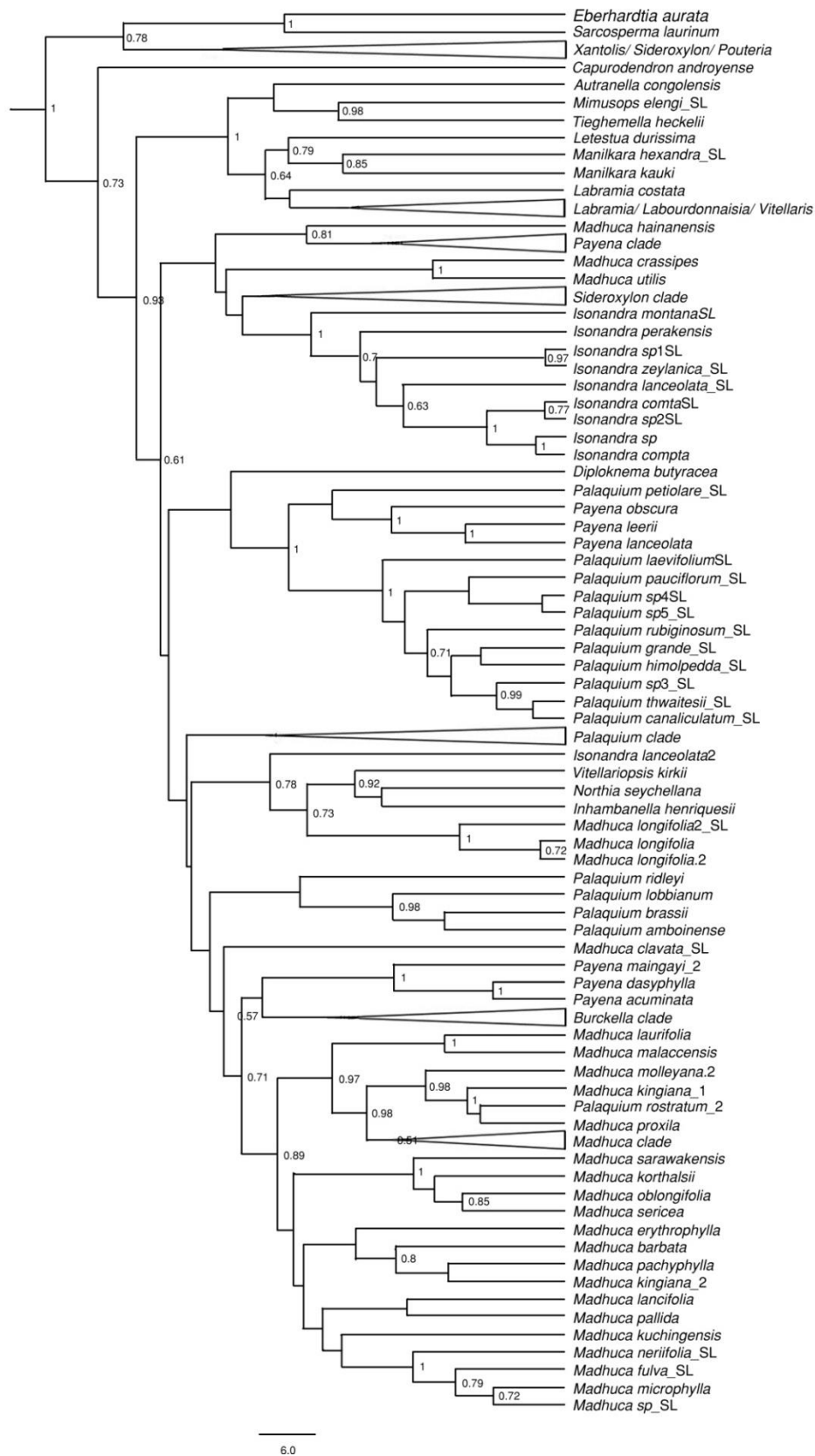
Figure 4.4 Maximum clade credibility chronogram of a Beast analysis of the Sapotaceae chloroplast dataset. Values next to nodes are the posterior probability values>0.5.

Isonandreae have originated from an African ancestor, and dispersed to Sunda Shelf followed by rapid diversification within and in surrounding areas. The *Palaquium* clade IX has an origin in Sunda shelf during the Eocene 42.9 (33.7-52.5) Mya, with diversification occurring during the Oligocene-Miocene with subsequent dispersal to the Sahul Shelf twice 1.1 (0.04-3.5) Mya and 4.1 (1.8-6.9) Mya, and Eastern Asia 3.7 (1.1-3.8) Mya during the late Miocene. Sri Lanka was reached by a single dispersal event and further diversification occurred in two lineages; one giving rise to *P. petiolare* during the Oligocene and the other lineage giving rise to ten endemic taxa during the Miocene 16.4 (10.7-23.1) Mya.

The genus *Isonandra* shows an initial origin on the Sunda Shelf 46.3 (38.5-54.4) Mya and subsequently spread to other areas; 2 separate dispersals to Sri Lanka and a single dispersal to the Sahul Shelf. The Sri Lankan endemic *Isonandra montana* has an origin during the late Eocene 36.9 (27.8-47.2) Mya and the rest of Sri Lankan endemics in the genus show an entrance to the island from Sunda shelf during Oligocene 28.5 Mya and further speciation of endemics thereafter. The ancestral area for Indo Sri Lankan *Isonandra lanceolata* is constructed as Sri Lanka, thus short distance dispersal from Sri Lanka to India has occurred more recently 0.76 Mya during the Pleistocene.

Sri Lankan *Madhuca* have three origins. The origin for Indo-Sri Lankan *Madhuca longifolia* and Chinese *Madhuca hainanensis* is constructed as Sri Lanka, thus short distance dispersal has occurred to the Indian continent and further dispersal to Eastern Asia has occurred during the Oligocene 27.4 (17.3-35.9) Mya. *Madhuca clavata* shows an early Oligocene entry to Sri Lanka during 31.5 Mya and further a more recent entry occurred during the mid-Miocene 11.5 Mya followed by in situ speciation of endemics thereafter. The *Burckella* clade nested within clade XVI shows an origin in Sunda Shelf and there are two dispersals to the Sahul Shelf and a single long distance dispersal to India which occurred during the Mid Miocene 10.9 (5.2-17.1) Mya.

Discussion

A non-Gondwanan origin of Sri Lankan Sapotaceae
The hypothesis of an ancient Gondwanan origin for the family in Sri Lanka is not supported by my data, as all lineages show origins that considerably post date the

breakup of Gondwana. All three larger genera, *Palaquium, Isonandra* and *Madhuca,* originated during the Eocene 31.9 (22.9-41.4) Mya, 36.9 (27.8-47.2) Mya and 33.1 (25-40.0) Mya respectively, followed by further diversification thereafter.

There is some evidence, based on other dated phylogenies, for Sri Lankan lineages being of Gondwanan origin such as in Crypteroniaceae. The family provides a classic example for overland migration referred to as the "out of India hypothesis", with arrival of some ancient Gondwanan lineages to Asia by rafting on the Indian plate (Conti et al. 2002; Rutschmann & Eriksson 2004; Moyle 2004; Karanth 2006). It comprises three genera, *Crypteronia* that is widely distributed in South East Asia, *Dactylocladus* with one species (*Dactylocladus stenostachys*) endemic to Borneo and *Axinandra zeylanica* endemic to Sri Lanka with three other *Axinandra* species occurring in the Malay Peninsula and the northern part of Borneo (Conti et al. 2002; Rutschmann & Eriksson 2004; Moyle 2004; Karanth 2006; Morley 2003; Renner et al. 2010). However the absence of these taxa in Peninsular India today is attributed to massive volcanism during the Cretaceous-Tertiary boundary and associated extensive aridification in India during the late Tertiary which resulted in substantial extinction (Morley 2003; Karanth 2006). Being an island perhaps Sri Lanka was less affected by these climatic changes and thus could have acted as a refugial area facilitating the persistence of these ancient Gondwanan lineages such as *Axinandra zeyalnica* (Ashton & Gunatilleke 1987; Morley 2003).

One striking example is the Sri Lankan endemic monotypic genus *Hortonia* in Monimiaceae, with other genera in the family occurring in America, Africa, Madagascar and The Mascarenes, New Caledonia, Australia, New Zealand and the Malesian region (Renner et al. 2010). The dated phylogeny of Monimiaceae suggests a Gondwanan origin for Sri Lankan *Hortonia* which dates back to 71 (57-84) Mya during the late Cretaceous. One possibility of this ancient lineage of *Hortonia* in Sri Lanka is explained as rafting on the Deccan plate to Asia (Ashton & Gunatilleke 1987). The other explanation is long distance dispersal of *Hortonia* from Antarctica to Sri Lanka, however the huge distance between two land masses ca 2100 Km (Ali & Aitchison 2008), raises the question of ability to cross such long distance over water according to Renner et al. (2010).

Other families have been speculated to be of Gondwanan origin but dated phylogenies need to be produced to confirm this. The family Dilleniaceae is one example of a classic Gondwanan distribution, with all Asian genera in the family found on the Indian Peninsula and Sri Lanka with their closest relatives in Madagascar and the Seychelles. Raven & Axelrod (1974) considered that *Schumacheria* in the family Dilleniaceae rafted on the Deccan plate to Asia based on Dickson's observations (1967-1969) that the Sri Lankan endemics share share many characters with the rare Bornean genus *Didesmandra*.

Dipoterocarpaceae, a family well represented in Sri Lanka and with a wider distribution in South America, Africa, Madagascar, India, Southeast Asia, and Malesia also exhibits a Gondwanan distribution, potentially facilitated by rafting on the Deccan plate (Ashton & Gunatilleke 1987; Givnish & Renner 2004; Ducousso et al. 2004; Renner et al. 2010). The endemic genus *Vateriopsis* in the Seychelles has its closest relative *Vateria* in India and Sri Lanka, and has large wingless fruits that lack dormancy. It has been argued that these features provide evidence for a vicariant origin as they are not adapted to any form of long distance dispersal (Ashton & Gunatilleke 1987).

Laurasian lineages

My second hypothesis was to test for any Laurasian lineages of Sapotaceae in Sri Lanka. The Deccan plate, composed of both India and Sri Lanka, collided with the Southern coast of Laurasia during the Eocene between 55-40 Mya (Briggs et al. 2003; Aitchison et al. 2008; Morley 2003), a profound tectonic event which led to the uplift of the Himalayas (White & Lister 2012; Ali & Aitchison 2008; Conti et al. 2002; Briggs et al. 2003). Since then, due to the connection with Asia, a vast immigration of Laurasian taxa and mixing of the floras from Deccan Gondwana and Laurasia was hypothesized (Ashton & Gunatilleke 1987, Morley 2000). However only a few migrants are evident; lower Tertiary fossils of Juglandaceae and Myricaceae in Assam provide evidence for the beginning of Laurasian immigrations (Ashton & Gunatilleke 1987). Families distributed in the Northern hemisphere with Laurasian origin such as *Pinus*, Hamamelidaceae, Juglandaceae, Myricaceae, Fagaceae and Clethraceae are completely lacking in South Asia (Ashton & Gunatilleke 1987). However, present study does not support any laurasian lineages in Sri Lanka for the family Sapotaceae and this may be

due to the migration difficulties caused by the Himalayan barrier and by the associated development of arid conditions after the montane uplift (Ashton & Gunatilleke 1987).

Long distance dispersals from Africa

A relatively close connection between Africa and the Deccan plate was facilitated via Madagascar, which may have allowed many plant taxa to disperse from Africa to the Deccan plate during its northward journey in the late Cretaceous (Morley 2003; Morley 1998; Aitchison et al. 2008). As the plate drifted along the west coast of Africa, small islands and land bridges may have facilitated short to medium distance over water dispersals during the early to late Cretaceous (Briggs et al. 2003). The distance between the African coast and the Deccan plate was ca. 420 km, which remained constant until c. 84 Mya and pollen records during that period provide strong evidence for dispersals (Morley 2003; Morley 1998; Rutschmann & Eriksson 2004). However, this dispersal path ceased when the Deccan plate separated from Madagascar Late Cretaceous (94 – 84 Mya); since then there have been few dispersal events between Africa and the Deccan plate during its northward journey to contact with Asia (Morley 2003; Morley 2000). There are two major dispersal phases from Africa to Asia evidenced based on macrofossil records, one occurring soon after the collision with Asia during the Eocene, 54-36 Mya and a further phase during the late Miocene 10-5 Mya (Morley 2000).

The tribe Mimusopeae evolved 52 Mya, followed by diversification 43 Mya during the Eocene when warm and wet climatic conditions were prevailing in the Northern hemisphere (Armstrong et al. 2014). Asian representatives of the genera *Manilkara* and *Mimusops* are nested within a grade of predominantly African genera (*Tieghemella, Autranella, Baillonella, Vitellaria*, and *Vitellariopsis*) suggesting an African origin that is highly supported (>95%) in ancestral area reconstructions. One possibility for the pantropical distribution of *Manilkara* and *Mimusops* could be attributed to migration through the boreotropics which could have occurred 65-45 Mya via the North Atlantic Land Bridge. However, the timing of pantropical distribution of the genera is too young to accept this hypothesis. However timings congruent with a boreotropical migration are consistent with several other families such as Burseraceae (Weeks et al. 2005), Malpighiaceae (Davis et al. 2002) and Meliaceae (Muellner et al. 2006). The expansion of tropical forests during the mid-Miocene climatic optimum, 17-15 Mya (Zachos et al.

2001) may also have facilitated the expansion of tropical taxa between Africa and Asia (Zhou et al. 2012). This pathway is followed by several animals such as primitive catarrhines, hominoids and lorisoids for the entry to Eurasia from Africa in the Early to Middle Miocene and some plant groups such as *Uvaria* (Annonaceae) (Zhou et al. 2012). Another possible explanation is migration from Africa to Asia after the collision of the Afro Arabian plate with Asia during the late Oligocene 25 Mya (Samuei et al. 1997). This could explain the migration of Asian *Manilkara* from Africa, as this pathway was available during the time frame of the split of this clade from its African relatives.

However, due to the young ages evident in the dated phylogeny of Sapotaceae, the most plausible explanation for the disjunct pantropical distribution of *Manilkara* and *Mimusops*is long distance dispersal from Africa to Madagascar, Asia and the Neotropics (this study; Armstrong et al. 2014). The recent origin of Sri Lankan *Manilkara hexandra* provides evidence for dispersal from Africa to the Indo-Sri Lankan region during the Miocene 19.04 (11.3-25.52) Mya, as does *Mimusops elengi* 6.81 Mya (3.47-11.12) which subsequently spread eastward into Malesia (Armstrong et al. 2014).

Long distance dispersals from Southeast Asia

My results confirm the Sundanian origin for Sri Lankan Isonandreae which is congruent with the results of Richardson et al. (2014). The Sri Lankan lineages are nested in Sundanian ones, indicating an origin in Sunda shelf followed by six independent long-distance dispersals; one during the Eocene 36.9 Mya, four during the Oligocene 33.1 Mya, 31.9 Mya, 31.5 Mya, 28.5 Mya and one during the mid-Miocene 11.5 Mya.

The Indian plate moved extremely close to the Malay Peninsula during its northward migration (Hall 2001) and the collision with Laurasian coast brought India and South East Asia to similar latitudes and within the same climatic belt. Further, the collision involved a glancing contact with Sumatra and then Burma in the late Paleocene onward 57 Mya facilitating more opportunities for floristic exchange (Ali & Aitchison 2008). However, the Indian flora is suggested to have been much more aggressive in migration, evidenced by the sudden appearance of Indian taxa in the Paleocene and Early Eocene in the palaeo floras of South East Asia and which resulted in the enrichment of the South East Asian flora (Morley 2003; Morley 2000).

The genus *Palaquium* has an ancestral area reconstructed as Sunda Shelf during the mid Eocene (Clade IX, Figure 4.2). It migrated to Sri Lanka giving rise to the endemic *Palaquium petiolare* during Oligocene and to a clade of endemic taxa (CladeX, Figure 4.2) during the Miocene. *Palaquium impressionervium* is likely a single back dispersal to Sunda shelf from Sri Lanka during the late Oligocene, but this supported by tree topology only.

Among *Isonandra* species, *Isonandra montana* was a much earlier arrival on Sri Lanka from Sunda Shelf during the Eocene (36.91 Mya, Clade XIII, Figure 4.2) followed by another dispersal phase during the Oligocene giving rise to other endemic *Isonandra* species (*Isonandra zeylanica, Isonandra compta, Isonandra* sp1, *Isonandra* sp 2) on the island (Clade XIV, Figure 4.2).

The ancestral area reconstruction for *Madhuca* (Clade XVIII) shows a Sundanian origin during the early Oligocene. The earliest arrival to the Indo Sri Lankan region occurred during the early Oligocene (33.14 Mya), giving rise to *Mudhuca longifolia*. Sri Lanka is constructed as the ancestral area for Indo Sri Lankan *Madhuca longifolia* and Chinese *Madhuca hainanensis,* consistent with a dispersal from Sri Lanka to India and then to Eastern Asia during the early-Mid Oligocene (27.4 Mya, Clade XVI, Figure 4.2), possibly by overland migration.

Sri Lanka is separated from India by a narrow sea barrier called Palk strait and there have been intermittent land connections between these two land masses until sea level rises 6000 years ago (Ashton & Gunatilleke 1987; McLoughlin 2001). During the period from the late Eocene (39-36 Mya) and Oligocene (36-25 Mya)warm pre humid conditions were prevailing across the Indian sub-continent (Morley 2000) and Pleistocene deposits from the wet zone of Sri Lanka suggest the existence of intermittent periods of seasonal tropical climate (Ashton & Gunatilleke 1987). Thus, relatively short distance dispersal over land or water would potentially facilitate this migration and further radiations during favourable climatic conditions. *Madhuca clavata*, a point endemic species in Sri Lanka, shows an origin in Sunda shelf (Clade XVI, Figure 4.2) during the Oligocene and another more recent dispersal event is evident within Sri Lankan *Madhuca*, during the Mid Miocene, which resulted in further evolution of endemics (Clade XIX, Figure 4.2).

The phylogeny presented here is consistent with relatively recent intercontinental dispersal being the major factor in disjunctions in the extant tropical flora. Sri Lankan Sapotaceae have a more recent origin, beginning in the Eocene ca. 36 Mya followed by speciation of endemics more recently during Pliocene-Pleistocene. The major direction of dispersal to Sri Lanka is from the east, predominantly from Sundania which has played a prominent role in enrichment of Sri Lankan flora. The warm temperate and perhumid climate in the south western part of Sri Lanka, with nearest analogues in Sumatra, perhaps may have facilitated these prevalent dispersals to Sri Lanka through providing favourable habitats (Ashton & Gunatilleke 1987). More intense sampling of Indian and Indochinese taxa is necessary to further test this hypothesis.

The Sapotaceae fruits possess different sizes ranging approximately from 1-10 cm, all with a fleshy, sweet colourful pericarps provide a hint for animal dispersals such as mammels and birds.

Mimosopeae fruits range from 1.5-10cm, with a fleshy, sweet pericarp which are dispersed by various mammals and birds. Asian and African *Manilkara* are dispersed by frugivorous birds such as doves and pigeons (Corlett 1998, snow 1981). It has been recorded that primates such as spider monkeys, howler monkeys, capuchins and tamarins (Chauvet et al. 2004; Chapman 1989; Oliviera & Ferrari 2000) and fruit bats (Uriarte et al. 2005) are responsible for fruit and seed dispersal of *Manilkara* in neotropics. *Palaquium* species possess yellowish green coloured fleshy fruits with average size ranging from 2-4 cm possibly a hint for animal dispersal such as monkeys and birds. Smooth, fleshy, red-orange coloured fruits of *Isonandra* species with average size (1.4 cm) possibly eaten by small birds and carry over long distances. Fruits of*Madhuca* species are large in size, about 3-4 cm long and the large size indicates the possible vectors as tropical bats, which are too large to be eaten and carried by birds. Thus, all Sapotaceae species possess a large fruits and seeds which are too bulky to be carried out and dispersed by wind. Thus, they possibly carried out by mammals, such as rodents and monkeys for short distances and in the gut content of birds across long distances.

Most of the trans-oceanic dispersals included in the study are dated back to Middle Miocene thermal maximum 23-12Mya) and warmer climates should have been

contributed to an increase in hurricanes and tornadoes which were capable of carrying large seeds and propagules over long distances (Graham 2006, Nathan et al. 2008). Phylogenetic studies have confirmed the capability of transoceanic dispersals of some other families with large fleshy fruits such as, Annonaceae (Su & Saunders 2009); *Adansonia*, Bombacaceae (Baum et al. 1998); *Atelia*, Leguminosae (Ireland et al. 2010); *Andira*, Leguminosae (Skema 2003); *Commiphora*, Burseraceae (Weeks et al. 2007); *Macherium*, Leguminosae (Lavin et al. 2000); *Symphonia*, Clusiaceae (Dick et al. 2003).

Incongruence between plastid and nuclear data

Other studies of Sapotaceae have relied on ITS datasets in ancestral area reconstructions due to the hard topological incongruence between chloroplast and nuclear datasets (Armstrong et al 2014; Richardson et al. 2015). One of the major reasons for utilizing ITS sequences is mainly due to their high variability since plastid data were not informative enough to test for alternative hypothesis. Chloroplast capture across long distances has been identified as the major cause for hard incongruences between plastid and nuclear data and it is demonstrated in several pant groups.

According to Armstrong et al. 2014, *M.hexandra* (Sri Lanka) and *M. littoralis* (Myanmar) are placed with other two Asian species in the nuclear phylogeny. However, those two species form a strongly supported clade (PP=1) in a plastid phylogeny with two African species, *M. mochisia* (Zambia) and *M. concolor* (South Africa), which is an indication of a chloroplast capture event (Armstrong et al. 2014). A possible hypothesis is hybridization between African and Asian lineages, with the African lineage donating the plastid DNA to the Asian species giving rise to *M. hexandra* and *M. littoralis* in Asia. That hypothesis could be confirmed by the origin of *Manilkara* in Africa and dispersal events from Africa to Asia and diversification thereafter.

Some other species in Sapotaceae such as *Chrysophyllum cuneifolium* and *Nesoluma* sps also show evidence of chloroplast capture, this time between South American and African lineages (Swenson et al. 2008; Smedmark & Anderberg 2007). Chloroplast capture has also been found in genera such as *Nothofagus* (Acosta & Premoli 2010), *Thuja* (Peng & Wang 2008), *Gossypium* (Wendel et al. 1995), *Heuchera* (Soltis et

al.1991; Soltis & Kuzoff 1995) and *Boykinia* (Soltis et al. 1996). However, these incongruences should be investigated with the use of additional markers and the exact mechanisms of underlying incongruences beg further explanations.

In situ speciation

My results confirm that *in situ* speciation is an important contributor to the richness of the Sri Lankan flora. Species that are endemic to Sri Lanka, and also have sister species in Sri Lanka, are likely to have arisen by *in situ* speciation, whereas endemics whose closest relatives occur elsewhere might be paleoendemics that once had wider distributions. Within Sri Lankan Sapotaceae, more rapid radiation of endemics occurred during the Pliocene-Pleistocene, with ten endemic species appearing during the Pleistocene, possibly as a result of the changing climatic conditions during inter glacial periods.

During the period of glacial maxima, the global temperature was cooler and drier, around 3-4°C in the equatorial regions and lowland rainforests in the equatorial zones trap species until favourable conditions (Bonnefille et al. 1992; Sosef 1994). The glacial cycles would have led to shifting and fragmentation of the climatic zones within Sri Lanka, where species would survive in isolation until environmental conditions become favourable. These contractions could have resulted in allopatric speciation giving rise to new species in isolated forest patches within the island and re-expansion of their ranges in the warm and wetter climate conditions during the Pleistocene cycles (Plana et al. 2004; Richardson et al. 2014).

References

Acosta, M. C. & Premoli, A. C., 2010.Evidence of chloroplast capture in South American Nothofagus (subgenus Nothofagus, Nothofagaceae). *Molecular Phylogenetics and Evolution* 54(1), pp. 235-242.

Aitchison, J.C., Ali, J.R. & Davis, A.M., 2008. When and where did India and Asia collide? *Geological Bulletin of China*, 27(9), pp.1351–1370.

Ali, J.R. & Aitchison, J.C., 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle

Jurassic through latest Eocene (166-35??Ma). *Earth-Science Reviews*, 88(3-4), pp.145–166.

Anderberg, A.A. & Swenson, U., 2003. Evolutionary lineages in Sapotaceae (Ericales): a cladistic analysis based on ndhF sequence data. *International Journal of Plant Sciences*, 164, pp.763–767.

Armstrong, K.E. et al., 2014. Patterns of diversification amongst tropical regions compared: a case study in Sapotaceae. *Frontiers in Genetics*, 5(December), pp.1–13.

Ashton, P.S. & Gunatilleke, C.V.S., 1987. New light on the plant geography of Ceylon I . Historical plant geography. *Journal of Biogeography*, 14(3), pp.249–285.

Bartish, I. V. et al., 2005. Phylogenetic relationships among Sapotaceae (Ericales) in Australia and Oceania: Molecular evidence for generic polyphyly and repeated dispersal. *Amer. J. Bot*. 92(4), pp.667–673.

Bartish, I. V. et al., 2011. Vicariance or long-distance dispersal: Historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, 38(1), pp.177–190.

Batalha-Filho, H. et al., 2014. Phylogeny and historical biogeography of gnateaters (Passeriformes, Conopophagidae) in the South America forests. *Molecular Phylogenetics and Evolution*, 79, pp.422–432.

Baum, D. A. et al., (1998). Biogeography and floral evolution of baobobs (Adansonia, Malvaceae) as inferred from multiple data sets. *Systematic Biology*, 47(2),pp.181-207.

Bonnefille, R.F. et al., 1992. Quantitative estimates of full glacial temperatures in equatorial Africa from palynological data. *Clim. Dyn*, 6, pp.251–257.

Briggs, J.C., Collections, Z. & Museum, G., 2003. The biogeographic and tectonic history of India. *Jounal of Biogeography*, 30, pp.381–388.

Chapman, C. A., 1989. Primate seed dispersal - the fate of dispersed seeds. *Biotropica*, 21(2), pp.148-154.

Chauvet, S. et al., 2004. Seed fate of two Sapotaceae species in a Guianan rain forest in the context of escape and satiation hypotheses. *Journal of Tropical Ecology,* 20, pp. 1-9.

Conti, E. et al., 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution; international journal of organic evolution*, 56(10), pp.1931–1942.

Cook, L.G. & Crisp, M.D., 2005. Not so ancient: the extant crown group of Nothofagus represents a post-Gondwanan radiation. *Proceedings of the Royal Society B: Biological Sciences*, 272, pp.2535–2544.

Corlett, R. T. (1998). Frugivory and seed dispersal by vertebrates in the Oriental (Indomalayan) Region. *Biological Reviews,* 73(4), pp.413-448.

Davis, C.C. et al., 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences USA*, 99, pp.6833–6837.

Demesure, B., Sodzi, N. & Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular ecology*, 4(1), pp.129–131.

Dick, C. W. et al., 2003. Molecular sytematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist*, 162(6),pp.691-703.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, pp.214.

Ducousso, M. et al., 2004. The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal before the India-Madagascar separation, about 88 million years ago. *Molecular Ecology*, 13(1), pp.231–236.

Givnish, T.J. & Renner, S.S., 2004. Tropical Intercontinental Disjunctions: Gondwana Breakup, Immigration from the Boreotropics, and Transoceanic Dispersal. *International Journal of Plant Sciences*, 165, pp.S1–S6.

Govaerts, R. et al., 2001. World checklist and bibliography of Sapotaceae London: Royal Botanic Gardens, Kew.

Graham, A., 2006. Modern processes and historical factors in the origin of the African element in Latin America. *Annals of the Missouri Botanical Garden,* 93(2),pp.335-339.

Gruas-Cavagnetto, C., 1976. Etude palynologique du Paleogene du sud de l'Angleterre. *Cah.Micropal*, 1, pp.1–49.

Hall, R., 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: Changing patterns of land and sea. *Faunal and Floral Migrations and Evolution in Se Asia-Australasia*, (Figure 1), pp.35–56.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98.

Hamilton, M.B., 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol*. 8, pp.521–523.

Ireland, H. E. et al., 2010. Biogeographical, ecological and morphological structure in a phylogenetic analysis of Atelia (Swartzieae, Fabaceae) derived from combined molecular, morphological and chemical data. *Botanical Journal of the Linnean Society,* 162, pp.39-53.

Jacobs, B. F. et al., 2005. Oligocene terrestrial strata of northwestern Ethiopia: a preliminary report on paleoenvironments and paleontology. *Palaeontol. Electron*, 8, pp.1–19.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), pp.1647-1649.

Landis, M.J. et al., 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, 62(0), pp.789–804.

Lavin, M. et al.,2000. Africa, the odd man out: Molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. *Systematic Botany,* 25(3),pp.449-467.

Lee, C. & Wen, J., 2004. Phylogeny of Panax using chloroplast trnC-trnD intergenic region and the utility of trnC-trnD in interspecific studies of plants. *Molecular Phylogenetics and Evolution*, 31(3), pp.894–903.

Mathews, J.P., 2009. Chicle: The chewing gum of the Americas, from the ancient Maya to William Wrigley. Tucson: University of Arizona Press.

McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49(3), pp.271–300.

Miller, M.A. et al., 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics,* 11, pp.43-48.

Moller, M., & Cronk, Q., 1997. Origin and relationships of Saintpaulia(Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany,* 84, pp.956 – 965.

Morley, R. J., 2000. *Origin and evolution of tropical rain forests*. Chichester: John Wiley & Sons, Ltd.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(12), pp.5–20.

Morley, R.J., 1998. Palynological evidence for Tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. *Biogeography and Geological Evolution of SE Asia*, pp.211–234.

Moyle, R.G., 2004. Calibration of molecular clocks and the biogeographic history of Crypteroniaceae. *Evolution; international journal of organic evolution*, 58(8), pp.1871–1873; discussion 1874–1876.

Muellner, A.N. et al., 2006. The mahogany family "out-of-Africa": divergence time estimation, global biogeographic patterns inferred from plastid rbcL DNA sequences, extant, and fossil distribution of diversity. *Molecular Phylogenetics and Evolution*, 40, pp.236–250.

Nathan, R., et al., 2008. Mechanisms of long-distance seed dispersal. *TRENDS in Ecology & Evolution*, 23(11),pp.638- 647.

Oliveira, A. C. M. & Ferrari, S. F., 2000. Seed dispersal by black-handed tamarins, Saguinus midas niger (Callitrichinae, Primates): implications for the regeneration of degraded forest habitats in eastern Amazonia. *Journal of Tropical Ecology*, 16,pp.709-716.

Olmstead, R.G., Sweere, J. a & Spangler, R.E., 1999. Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *Solanaceae IV: Advances in biology and utilization*, pp.111–137.

Peng, D. & Wang, X. Q., 2008. Reticulate evolution in Thuja inferred from multiple gene sequences: Implications for the study of biogeographical disjunction between eastern Asia and North America. *Molecular Phylogenetics and Evolution,* 47(3),pp.1190-1202.

Pennington, T. D., 1991. The genera of the Sapotaceae. Royal Botanic Gardens Kew & New York Botanical Garden.

Plana, V. et al., 2004. Pleistocene and pre-Pleistocene Begonia speciation in Africa. *Molecular Phylogenetics and Evolution*, 31, pp.449–461.

Posada, D. et al., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, pp.772–772.

Praveen Karanth, K., 2006. Out-of-India Gondwanan origin of some tropical Asian biota. *Current Science*, 90(6), pp.789–792.

Rambaut, A., 2009. FigTree, a graphical viewer of phylogenetic trees. *Institute of Evolutionary Biology University of Edinburgh*.

Raven, P. H., & Axelrod, D.I., 1974. Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Gard*, 61,pp.539– 673.

Ree, R.H. & Smith, S. a, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57(1), pp.4–14.

Renner, S.S. et al., 2010. Biogeography of the Monimiaceae (Laurales): A role for East Gondwana and long-distance dispersal, but not West Gondwana. *Journal of Biogeography*, 37(7), pp.1227–1238.

Richardson, J.E. et al., 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359(1450), pp.1495–1508.

Richardson, J.E. et al., 2014. The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). *Botanical Journal of the Linnean Society*, 174(1), pp.130–140.

Ronquist, F., 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(l), pp.195–203.

Ronquist, F. et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, pp.539–42.

Rutschmann, F. & Eriksson, T., 2004. Did Crypteroniaceae really disperse out of India? Molecular dating evidence from rbcL, ndhF, and rpl16 intron sequences. *International Journal of plant Science*, 165(S4), pp.S69–S89.

Samuei, M. A. et al., 1997. Field ob- servations from the Socotran platform: Their interpretation and cor- relation to Southern Oman. *Mar. Petr. Geol*, 14, pp.661–673.

Setoguchi, H. et al., 1998. Phylogenetic relationships within Araucariaceae based on rbcL gene sequences. *American Journal of Botany*, 85, pp.1507–1516.

Skema, C., 2003. Phylogeny and biogeography of Andira. A case study in Neotropical speciation. M.Sc. Thesis, Edinburgh: University of Edinburgh.

Smedmark, J.E.E. & Anderberg, A. a., 2007. Boreotropical migration explains hybridization between geographically distant lineages in the pantropical clade Sideroxyleae (Sapotaceae). *American Journal of Botany*, 94(9), pp.1491–1505.

Smedmark, J.E.E., Swenson, U. & Anderberg, A. a., 2006. Accounting for variation of substitution rates through time in Bayesian phylogeny reconstruction of Sapotoideae (Sapotaceae). *Molecular Phylogenetics and Evolution*, 39(3), pp.706–721.

Snow, D. W. (1981). Tropical frugivorous birds and their food plants - a world survey. *Biotropica,* 13(1),pp.1-14.

Soltis, D. E. et al., 1996. Discordance between ITS and chloroplast topologies in the Boykinia group (Saxifragaceae). *Systematic Botany*,21(2),pp.169-185.

Soltis, D. E. & Kuzoff, R. K., 1995. Discordance between nuclear and chloroplast phylogenies in the Heuchera group (Saxifragaceae). *Evolution*, 49(4),pp.727-742.

Soltis, D. E. et al., 1991. Chloroplast DNA variation within and among genera of the Heuchera group (Saxifragaceae) - Evidence for chloroplast transfer and paraphyly. *American Journal of Botany*, 78(8), pp.1091-1112.

Sosef, M.S.M., 1994. Refuge Begonias. Taxonomy, phylogeny and historical biogeography of Begonia sect. Loasibegonia and sect. Scutobegonia in relation to glacial forest refuges in Africa (Stud- ies in Begoniaceae 5). *Wageningen Agric. Univ. Pap*. 94, pp.1–306.

Su, Y. C. F. & Saunders, R. M. K., 2009. Evolutionary divergence times in the Annonaceae: evidence of a late Miocene origin of Pseuduvaria in Sundaland with subsequent diversification in New Guinea. *BMC Evolutionary Biology*, 9(153).

Swenson, U. & Anderberg, A.A., 2005. Phylogeny, character evolution, and classification of Sapotaceae (Ericales). *Cladistics*, 21, pp.101–130.

Swenson, U., Backlund, A., McLoughlin, S. & Hill, R.S., 2001. Nothofagus biogeography revisited with special emphasis on the enigmatic distribution of subgenus Brassospora in New Caledonia. *Cladistics*, 17, pp.28–47.

Swenson, U. et al., 2007. Phylogeny, diagnostic characters and generic limitation of Australasian Chrysophylloideae (Sapotaceae, Ericales): evidence from ITS sequence data and morphology. *Cladistics,* 23(3),pp.201-228.

Swenson, U., Nylinder, S. & Munzinger, J., 2014. Sapotaceae biogeography supports New Caledonia being an old Darwinian island. *Journal of Biogeography*, 41, pp.797–809.

Swenson, U., et al., 2008. Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): evidence of generic polyphyly and extensive morphological homoplasy. *Cladistics*, 24, pp.1006–1031.

Takhtajan, A., 1986. Floristic Regions of the World. Berkeley, CA: University of California Press.

Thomas, D.C. et al., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, 39(1), pp.98–113.

Uriarte, M. et al., 2005. Seedling recruitment in a hurricane-driven tropical forest: light limitation, density-dependence and the spatial distribution of parent trees. *Journal of Ecology,* 93(2),pp.291-304.

Van Welzen, P. C., Slik, J. W. F., & Alahuhta, J., 2005. Plant distribution patterns and plate tectonics in Malesia. *Kong. Dansk. Vidensk*, 55, pp.199–217.

Weeks, A., Daly, D.C. & Simpson, B.B., 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution*, 35, pp.85–101.

Weeks, A. & Simpson, B. B., 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution of historical biogeography of an "impossible" genus. *Molecular Phylogenetics and Evolution,* 42,pp.62-79.

Wendel, J. F. et al., 1995. An unusual ribosomal DNA sequence from Gossypium gossypioides reveals ancient, cryptic intergenomic introgression. *Molecular Phylogenetics and Evolution,*4(3),pp.298-313.

White, L.T. & Lister, G.S., 2012. The collision of India with Asia. *Journal of Geodynamics*, 56-57, pp.7–17.

Zhou, L. et al., 2012. "Out-of-Africa" dispersal of tropical floras during the Miocene climatic optimum: Evidence from Uvaria (Annonaceae). *Journal of Biogeography*, 39(2), pp.322–335.

Zachos, J. et al., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (New York, N.Y.)*, 292(April), pp.686–693.

CHAPTER 5: Pantropical biogeography of Zingiberaceae

Introduction

Zingiberaceae is the largest family of the order Zingiberales, with 53 genera and over 2000 species (Kress et al. 2005; Kress et al. 2002). It is pantropical in distribution, with one genus *Renealmia* L.f. in the Neotropics, four genera (*Aframomum* K.Schum., *Aulotandra* Gagnep., *Siphonochilus* J.M.Wood & Franks and *Renealmia* L.f.) occurring in Africa, with the bulk of the remaining genera occurring in South and South East Asia (Kress et al. 2002, Särkinen et al. 2007). The family is later divided into four sub families and six tribes based on the nuclear ITS and chloroplast *mat*K:Siphonochiloideae (Siphonochileae), Tamijioideae (Tamijieae), Alpinioideae (Alpinieae, Riedelieae), and Zingiberoideae (Zingibereae, Globbeae) (Kress et al. 2005). Costaceae, which was included in the family is now placed as the sister group to Zingiberaceae.*Siphonochilus* and *Tamijia* are the basal clade within Zingiberaceae (Kress et al. 2002).

Table 5.1 Division of Zingiberaceae according to Kress et al. (2002)

Subfamily	Tribe	Genera
Siphonochiliodeae	Siphonochileae	*Siphonochilus*
Tamijioideae	Tamijieae	*Tamijia*
Alpinioideae	Alpinieae Riedelieae	*Etlingera, Amomum, Alpinia, Hornstedia* *Riedelia, Burbidgea, Pleuranthodium*
Zingiberoideae	Zingibereae Globbeae	*Zingiber, Boesenbergia, Curcuma, Hedychium* *Globba, Hemiorchis, Gagnepainia, Mantisia*

The genus *Tamijia* shows a restricted distribution in Nothern Borneo. The monotypic genus *Siphonochilus* is found in Tropical Africa and all other genera are shared among Alpinioideaea and Zingiberoideae. The subfamily Zingiberoideae are found only in Asia where as Alpinioideae are widely distributed in Africa, Neotropics and Asia. Within Alpinioideae, *Renealmia* are distributed in Neotropics while *Aframomum* in both Africa

and Neotropics. The Alpinioideae have their main centre of diversity in Malesia, while Zingiberoideae in Indochina (Kress et al. 2002).

A phylogenetic analysis by Kress et al. (1995) on Zingiberales using both morphological and *rbc*L DNA sequence data could obtained a better resolution across the order by combining both data sets than either one alone. Later three molecular data sets, including two chloroplast genes (*rbc*L and *atp*B) and one nuclear gene (18S) were utilized in combination with morphological data, have further proven the effectiveness of combing both molecular and morphological data in phylogenetic studies (Kress et al. 2001).

Within Alpinoideaea, two genera, *Alpinia* and *Amomum* are two large genera with wider distributions across South Asia, Indo china to Southeast Asia (Xia et al. 2004). *Alpinia* is the largest and most taxonomically complex genus in Zingiberaceae and parsimony and Bayesian analysis using Internal Transcribed Spacer (ITS) and chloroplast *mat*K suggested six clades across the polyphyletic tribe Alpinieae (Kress et al. 2005). Phylogenetic analysis using same markers confirmed the polyphyly of the genus *Amoumum* with three well defined groups (Xia et al. 2004) and combined data sets resulted in more highly resolved phylogenies.

Further, chloroplast *mat*K have been used in various genera of the Zingiberaceae; *Aframomum, Alpinia, Curcuma, Globba, Hedychium* and *Zingiber* in order to find inter and intra-level relationships among them. Polymorphism assessment of *mat*K has proven the suitability of the region for DNA barcoding of Zingiberaceae (Selvaraj et al. 2008).

Studying the extant distribution patterns of tropical plant families and understanding the history of tropical floristic patterns is an emerging science. The Zingiberaceae is an ideal group for this purpose given their pantropical distribution and the availability of robust phylogenetic data (Kress & Specht 2006).

The order Zingiberales originated ca. 124 Mya, followed by divergence events leading to the formation of the family lineages by ca. 110 Mya (Kress & Specht 2006). The ancestral area for the entire Zingiberales was reconstructed as America and Southeast Asia (Kress & Specht 2006) using a Dispersal-Vicariance analysis (DIVA;Ronquist

1997). Even though the age estimates for the order are congruent with a Gondwanan vicariance hypothesis, the lack of tropical climate during that time in Gondwana suggests a Northern Laurasian distribution for the early diverging lineages. Fossil records in America, Eurasia and Greenland during upper Cretaceous to early Tertiary further favours the hypothesis of an ancestral Laurasian distribution(Kress & Specht 2006).

The Zingiberaceae-Costaceae group diversified and split into the currently recognised two families ca. 105 Mya. The DIVA analysis of Kress & Specht (2006) reconstructed a wide ancestral area hypothesis for the Zingiberaceae-Costaceae clade, effectively a Gondwanan distribution including Africa, America, Melanesia and Southeast Asia. The ancestral area for the Zingiberaceae was reconstructed as Africa and India. The Zingiberaceae is currently well represented in Africa, Madagascar, India and Sri Lanka. Fossil material found in the Intertrappean beds of the Deccan Plate which resembles *Amomum* (extant in Africa, tropical Asia and the Mascarenes) provides tantalising evidence for the migration of Zingiberaceae to tropical Asia via rafting on the Indian plate (Ashton & Gunatilleke 1987; Kress & Specht 2006).

The Deccan plate, composed of India and Sri Lanka, moved close to the Malay Peninsula during its northward migration in the mid Eocene, thus could have facilitated the dispersal of lineages to Southeast Asia (Kress & Specht 2006; Aitchison et al. 2008). Sri Lanka had an intermittent land connection with India since the separation of the Deccan plate from Gondwanaland until colliding with Eurasia in Miocene(Ashton & Gunatilleke 1987). Thus, Sri Lanka is a key location for examining the route of entry of Zingiberaceae into Tropical Asia and to interpret the timing and diversification of the family in Asia.

In Sri Lanka Zingiberaceae show an abundant distribution in the wet zone, mostly in lowland and mid montane primary forest, less so in secondary forests. Sri Lankan Zingiberaceae are composed of twelve genera, of which 12 out of 36 species are known to be endemic.

The Sri Lankan *Alpinia fax* clade is of particular importance since it is one of the potential sister groups to neotropical *Renealmia*, based on nuclear data (Särkinen et al.

2007).The members, *Alpinia fax* (endemic) & *Alpinia abundiflora* (native), of this small clade occur in Sri Lanka and a small part of South West India and their close affinity with African/neotropical clades suggest past dispersal paths across the Indian ocean following the breakup of Gondwana (Kress et al. 2005).

This study is the first to incorporate multiple Sri Lankan taxa in a global dated phylogeny and to infer molecular divergence age estimates for Sri Lankan Zingiberaceae. I aim to reconstruct the ancestral areas of distribution, focussing on Asian Zingiberaceae, using recent advances in biogeograhic analysis (BIOGEOBEARS, Matzke 2013), which allows us to compare models of goeographic range evolution. I used a finer scale area coding than in the studies of Särkinen et al. (2007) and Droop (2012) in order to obtain a more detailed picture of the origin and timing of diversification events during the early biogeographic evolution of Zingiberaceae.

Materials and methods

Taxon sampling

In order to complement lacking South Asian Zingiberaceae the following 21 samples were added to the DNA alignment generated by Droop (2012) comprising 127 taxa: *Zingiber cylindricum, Curcuma albiflora, Amomum graminiflolium, Amomum nemorale, Amomum acuminatum, Alpinia fax, Alpinia sp, Cyphostigma pulchellum, Amomum* sp 1, 2, 3, 5, 6, 7 (Sri Lankan endemics 80%), *Zingiber wightianum, Elettaria* sp, *Amomum pterocarpum*, *Amomum fulviceps*, *Amomum masticatorium*, *Alpinia abundiflora, Curcuma zedoaria* (distributed in both in Sri Lanka and India, natives 65%). In order to obtain a better understanding of biogeographic affinities with African and neotropical gingers, taxonsampling was increased by adding sequences of *Renealmia battenbergiana*, *Aframomum danielii*, *Aframomum sceptrum* (all native to Africa) which were downloaded from Genebank. The final data set comprised 151 taxa representing broadly all tribes of Zingiberaceae; Alpinieae, Zingibereae, Globbeae, Siphonochilieae and Tamijieae.

All the Asian species Zingiberaceae have been proved to be monophyletic and *Tamijia flagellaris, Siphonochilus kirkii, S. Decorus and S. Aethiopicus* (African) were identified as basally diverging lineages in the order Zingiberales in previous studies

(Särkinen et al. 2007; Kress et al. 2005).Thus the Asian species were assigned as monophyletic, allowing basally diverging African taxa to be the out group. The newly generated sequences are listed in table 5.2 and all the accessions used in the study are listed in Appendix 5.

Table 5.2 Voucher information for the newly generated sequences for the present study

Species name	Voucher	EDNA number
Zingiber_cylindricum Moon	Kumarage 7 (E)	EDNA13-0033869
Amomum_pterocarpum Thwaites	Kumarage 21 (E)	EDNA13-0033871
Zingiber_wightianum Thwaites	Kumarage 20 (E)	EDNA13-0033870
*Amomum_sp*1	Kumarage 34 (E)	EDNA13-0033867
Elettaria_sp	Kumarage 5 (E)	EDNA13-0033868
Curcuma_zedoaria Roxb.	Kumarage 24 (E)	EDNA13-0033872
Amomum_fulviceps Thwaites	Kumarage 33 (E)	EDNA13-0033873
*Amomum_sp*2	Kumarage 50 (E)	EDNA13-0033874
Curcuma_albiflora Thwaites	Kumarage 38 (E)	EDNA13-0033875
Amomum_graminifolium Thwaites	Kumarage 48 (E)	EDNA13-0033876
*Amomum_sp*3	Kumarage 82 (E)	EDNA13-0033878
Amomum_nemorale Benth. & Hook.f.	Kumarage 52 (E)	EDNA13-0033879
Cyphostigma_pulchellum Benth.	Kumarage 54 (E)	EDNA13-0033880
Amomum_acuminatum Thwaites	Kumarage 55 (E)	EDNA13-0033881
Amomum_masticatorium Thwaites	Kumarage67 (E)	EDNA13-0033882
*Alpinia_abundiflora*1 Burtt & R.M.Sm.	Kumarage 70 (E)	EDNA14-0035527
*Amomum_sp*5	Kumarage 86 (E)	EDNA14-0035531
*Amomum_sp*6	Kumarage 88 (E)	EDNA14-0035532
*Amomum_sp*7	Kumarage 89 (E)	EDNA14-0035533
Alpinia_sp	Kumarage 71 (E)	EDNA14-0035528
Alpinia_fax Burtt & R.M.Sm.	Kumarage 94 (E)	EDNA14-0035529

DNA region sampling

Many Studies on Zingiberaceae have used chloroplast markers such as *mat*K and *trn*L-F in combination with ITS in order to obtain phylogenetic resolution (Kress et al. 2005; Kress et al. 2002). The Internal Transcribed Spacer (ITS) is a useful marker in resolving species level relationships in Zingiberaceae and the high copy number makes it easy to amplify. *mat*K, which is located within the intron of *trn*K, is another marker with proven suitability in resolving species level relationships due to its high rate of substitution and low transition/transversion ratios. However, *trn*L-F does not provide highly resolved trees due to lack of variation, thus was not used in my study.

DNA extraction, amplification and sequencing

Total genomic data was extracted from silica gel dried material using DNeasy Plant Mini Kit (Qiagen, UK) according to the manufacturer's protocols. 25 µl PCR reactions were setup for amplification of both ITS and matK regions.

For ITS, each 25µl PCR reaction contained 16.125µl of ddH$_2$O, 2.5µl of 10x reaction buffer, 1.25µl 25mM MgCl$_2$, 2.5µl 2mM dNTPs, 0.75µl 10µM forward primer (ITS4), 0.75µl 10µM reverse primer (ITS5p), 0.125µl of Biotaq DNA polymerase (Bioline, UK) and 1µl of DNA of DNA template.

The PCRtemperature profile was template denaturation at 94°C for 3 min followed by 29 cycles of denaturation at 94°C for 1 min, primer annealing at 55°C for 1 min, primer extension at 72°C for 1.5 min followed by a final extension step at 72°C for 5 min.

The *mat*K region was amplified in four separate reactions using the primer pairs *trn*K1 and mIR; mSP2F and m5R; mIF and m8R; and m8Fa and *trn*K2R (Table 5.3).

Each 25µl PCR reactioncontained 15.25µl of ddH$_2$O, 2.5µl of 10x reaction buffer, 1.25µl 25mM MgCl$_2$, 2.5µl 2mM dNTPs, 0.75µl 10µM forward primer, 0.75µl 10µM reverse primer, 0.8µl 0.4% BSA, 0.2 µl of Biotaq DNA polymerase (Bioline, UK) and 1µl of DNA template. The temperature profile included an initial template denaturation step of 94°C for 5 minutes, followed by 35cycles of denaturation at 94°C for 30 s, primer annealing at 55°C for 45 Seconds, primer extension at 72°C for 1.5 minutes; and a final extension step at 72°C for 30 seconds.

Table 5.3 Nuclear (ITS) and chloroplast primer sequences used in Zingiberaceae DNA sequence generation.

Region	Primer name	Primer sequence	References
ITS	ITS4	TCCTCCGCTTATTGATATGC	White et al. 1990
	ITS5p	GGAAGGAGAAGTCGTAACAAG	White et al. 1990
*mat*K	*trnK*1F (F)	CTCAACGGTAGAGTACTCG	Manos & Steele 1997
	mIR (R)	CGT TTC ACA AGTACT GAA CTA	Kress et al. 2002
	mSP2F (F)	TGG GTT AGA GAC GAA TGT GT	Kress et al. 2002
	m5R (R)	AGG ATC CTT GAA AAT CCA TAG A	Kress et al. 2002
	mIF (F)	GTTCAG TAC TTG TGA AAC GTT	Kress et al. 2002
	m8R (R)	AGC ACA AGA AAG TCG AAG	Kress et al. 2002
	m8Fa (F)	TAC TTC GAC TTT CCT GTG CC	Kress et al. 2002
	*trnK*2R (R)	AACTAGTCGGATGGAGTAG	Steele & Vilgalys 1994

Amplified products were run on1% Agarose TBE gel with Syber safe as the staining agent and visualized in a UV transilumintor. PCR purification was carried outusing EXOSAP IT in7 μl reactions. 5μl of PCR product was mixed with 2μl of EXOSAP and incubated for 37°C at 15 minutes followed by 80° C for 15 minutes.

Sequencing PCR was carried out using purified PCR products in 10μl reactions using 5.68μl of ddH$_2$O, 2μl of sequencing buffer, 0.32μl of primer, 1μl of Big dye and 1–Xμl of template. The sequencing PCR protocol was denaturation at 95°C for 30 sec, followed by 24 cycles of primer annealing at 50°C for 20sec, extension at 60°C for 4 min. Separate forward and reverse sequencing PCR s were carried out and products were sent to the Genepool facility at the University of Edinburgh (Genepool, UK) for analysis.

Sequence editing and alignment

Newly generated DNA sequences were edited in Geneious (7.1.4; Kerase et al. 2012). *Amomum villosum* sequences of the relevant marker from Droop (2012) were used as the reference consensus sequence for assembling. Sequences were aligned manually in

Bioedit 7.1.3 (Hall 1999) and checked for indels. Bases were excluded due to uncertainties in the alignment or missing data at the region ends, ITS dataset: 108-113, 513-518, 696-776. *mat*K dataset: 3143-3149, 3171-3173.

Phylogenetic analysis and Bayesian divergence time estimation

Models of sequence evolution were determined using jmodeltest 2.1.3 (Posada et al. 2012). Maximum likelihood topologies were used to estimate the optimal evolutionary model and twenty four models were used under the Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC).

Both chloroplast and nuclear datasets were analyzed under Bayesian inference (BI) which was performed in Mrbayes 3.2.1 (Ronquist et al. 2012) in the CIPRES science gatewayV. 3.3 (Miller et al. 2010). The regions were treated as single partitions. MCMC runs were carried out for 10000000 generations and sampled every 1000 generations. 25% burn-in was set to discarded the initial trees and the remainder were summarised as a 50% majority rule consensus tree and were visualized in Fig tree (Rambaut 2009) and checked for congruence between regions. A combined *mat*K-ITS data matrix was not used due to hard incongruence between two tree topologies. The ITS region was chosen for molecular dating in BEAST and ancestral area reconstructions in BIOGEOBEARS, due to the more complete sampling and more resolved phylogeny.

Bayesian divergence time estimation was performed on the ITS alignment using BEAST v.1.8.0 (Drummond & Rambaut 2007). GTR was selected as nucleotide substitution model with Gamma + invariant site as the site heterogeneity model. The data set was treated as a single partition and uncorrelated relaxed lognormal clock model was selected in order to relax the assumption of a molecular clock and allow for rate heterogeneity between lineages. The tree prior was set to random birth death speciation process with a randomly started generating tree.

Four separate Markov Chain Monte Carlo (MCMC) runs were carried out for 10000000 generations sampling every 1000 generations. Plots of the logged parameters for each run were visualised using Tracer v.1.5 (Drummond & Rambaut 2007) to confirm convergence between runs. Time series plots of all parameters were analyzed in tracer

to check for convergence and to confirm adequate effective sample sizes. Trees were combined in LOGCOMBINER (Drummond & Rambaut 2007) and burn-in was set for 25% for initial sample for each run. A single maximum clade creditability tree was obtained from Tree Annotator v.1.7.5 (Drummond & Rambaut 2007) and visualized in fig tree v.1.4.0 (Rambaut 2009).

Fossil constraints and secondary calibrations

The fossil taxon *Zingiberopsis magnifolia* (Knowlton) Hickey from North America and Eurasia was used in the temporal calibration of the Zingiberaceae phylogeny (Peppe et al. 2007). On the basis of the leaf venation pattern, it has been clearly ascribed to Zingiberaceae and have been used in biogeographic studies in order to assign the age of the crown node of Zingiberaceae (Särkinen et al. 2007; Kress & Specht 2006).The age of the fossil is re-estimated and now it is accepted the earliest occurance was in the Maastrichtian, beginning 71 Ma.

The crown age of 69-71 Ma was used where the basal lineages *Siphonochilous* and *Aulotandra* splits from the remaining taxa in the phylogeny and a lognormal prior distribution was used in favour of older age estimates.

Biogeographic analysis

Geographical area delimitation and scoring

Twelve areas were coded based on the extant distribution, areas of endemism and geological history. In order to obtain a clear picture of the biogeographic history of South Asian Zingiberaceae, Sri Lanka and India were coded as separate geographic areas. The areas are (1) America; (2) Africa; (3) Madagascar; (4) India; (5) Sri Lanka; (6) China; (7) Indochina; (8) Sunda shelf; (9) Sahul shelf; (10) Papua New Guinea; (11) Philippines; (12) Australia. A data matrix was prepared coding presence/absence in each of the area depending on the collection details and voucher specimens (Zingiberaceae Resource Centre, RBGE). The maximum number of areas in ancestral ranges was constrained to two since all the species found in a single area except *Amomum petaloidium* and *Alpnia japonica* found in China-Indochina and *Amomum fulviceps, Amomum masticatorium, Zingiber wghitianum, Alpinia abundiflora* and *Elettaria* sp found both in India and Sri Lanka.

Ancestral Area Reconstructions

Ancestral areas within internal nodes were constructed using Biogeobears (Matzke 2013) in R package under four models; DIVA like (Ronquist 1997), DIVA LIKE+J, DEC (Ree & Smith 2008), DEC+J. DEC and DEC+J calculates maximum likelihood ratios of ancestral states at speciation events, in a method similar to LAGRANGE (Ree et al. 2005). In addition to two free parameters; d (dispersal), e (extinction) included in DEC model (Batalha-Filho et al. 2014; Matzke 2013) the additional parameter ''J'' is added to the model to account for founder event speciation. Founder event speciation is important in lineage splitting especially in island systems. The J parameter controls the probability of two events during cladogenesis; founder event speciation versus sympatric and vicariant speciation (Matzke 2013). Log likelihood values for each model were compared and the model with highest value was chosen as the best model for inferring ancestral ranges at nodes.

Results

Phylogenetic Analysis

Both ITS and *mat*K dating were carried out, however alignments were not combined in a single analysis due to the hard incongruence of the datasets. The ITS alignment was used in biogeographic analysis as it gives a better resolution across the family and is compatible with published phylogenies (Särkinen et al. 2007; Kress & Specht 2006; Kress et al. 2005). The resulted from Bayesian and beast analysis of nuclear ITS region are shown in figures 5.1 and 5.2 respectively.

Table 5.4 Descriptive statistics of nuclear ITS and *mat*K datasets

DNA region	Amplicon length	Alignment length	Number of variable characters	Number of informative characters
ITS	776	695	432 (62.15%)	360 (51.79%)
*mat*K	3173	3170	882 (27.82%)	540 (17.03%)

Within the Sri Lankan Zingiberaceae, the *Alpinia fax* clade which exhibits prominent incongruence between the trees derived from nuclear and chloroplast data (Figure 5.5). In the ITS tree, the Indo-Sri Lankan *Alpinia fax* clade is sister to *Aframomum* from

Africa, however it is not supported (PP=0.42). The Sri Lankan endemics *Cyphostigma pulchellum* and *Amomum nemorale* are sister to the Afro-American *Renealmia* with high support (PP=0.99). In the *mat*K tree the *Alpinia fax* clade (Clade A) is sister to *Cyphostigma pulchellum* and *Amomum nemorale* (although with low support; PP=0.83), and *Aframomum* is reconstructed as sister to *Renealmia* (although with low support; PP=0.84). However this topological incongruence does not contribute to any biogeographic incongruence since for both data sets Sri Lanka is constructed as the ancestral area for both African *Aframomum* and Neotropical *Renealmia*.

Considering the ITS trees, *Tamijia flagellaris* is constructed as the sister taxon to rest of the Zingiberaceae, which is strongly supported as monophyletic (Clade I, PP=1). The remaining taxa fall into clade II which comprises two strongly supported clades: the subfamilies Zingiberoideae (clade III, PP=1) and Alpiniodeae (Clade VI, PP=1).

Clade III consists of *Campandra*, *Gagnepainia*, *Hemiorchis*, *Curcuma*, *Distichichlamys*, *Hedychium* and *Zingiber*. Within clade III the Sri Lankan endemic *Curcuma albiflora* is nested within clade IV (PP=1) comprising species from China and Indochina, with the Indochinese *Curcuma roscoeana* as the sister taxon. The genus *Zingiber* is strongly supported as monophyletic (Clade V, PP=0.98) with the Sri Lankan endemic *Zingiber cylindricum* reconstructed as sister taxon to Indo-Sri Lankan and Eastern Asian (China and Indochina) *Zingiber*.

Clade VI comprises tribes Ridelieaeand Alpiniae. The Ridelieae are monophyletic with strong support (*Siliquamomum*/*Burbidgia*/*Ridelia* clade PP=1) while Alpinieae are non-monophyletic, being spread amongst six clades with varying levels of support.

Clade VIII is composed of Sri Lankan *Amomum nemorale, Cyphostigma pulchellum, Alpinia* species and *Elettaria* together with African *Aframomum* and Afro-American *Renealmia*, but is not supported (PP=0.27). This clade is nested within clade VII (also not supported; PP=0.42), with *Alpinia* from China, Indochina and Sunda Shelf constructed as the sister group. Sri Lankan *Amomum* are polyphyletic, with one species *Amomum nemorale* resolved in a clade with the monotypic *Cyphostigma pulchellum* (PP=1), sister to Afro-American *Renealmia* with a high support (PP=0.99). Other Sri

Lankan Amomums are nested within clades XII, XIII and XIV with varying statistical support.

Sri Lankan *Alpinia*, *Alpinia fax* clade (Clade A) are paraphyletic, with indo Sri Lankan *Elettaria* nested within (PP=1), forming a sister clade to *Aframomum* from Madagascar with no support (PP=0.42). The Indo-Sri Lankan *Amomum pterocarpum* is nested within the *Amomum/Eletariopsis* clade (collapsed in Fig. 3) and forms a clade with Australian *Amomum queenslandicum* with high support (PP=1.0). Clade IX contains the remaining *Alpinia* species of Eastern and Southeast Asian origin along with three Chinese *Amomum* species, *A. aff. paratsaoko, A. paratsaoko, A. coriandriodorum* nested within (PP=0.52).

The clade X (PP=1) is composed of species from Sunda shelf, Indo china, Papua New Guinea along withtwo strongly supported sub-clades of origins with Australian and Philippines. Indo Chinese *Hornstedtia hainanensis* and Sundanian *Amomum centrocephalum* are the sister group for rest of taxa. The Phillipine taxa form a monophyletic clade with strong support (P=0.96). The Australian *Alpinia* species are nested within clade X forming a monophyletic clade with high support value (PP=1).

The clade XI is composed of all remaining *Amomum* species and is strongly supported as monophyletic (PP=1). Sri Lankan endemic *Amomum* species are polyphyletic, highly nested within clade XI, as are the Indo-Sri Lankan taxa *A. masticatorium* and *A. fulviceps*.

Divergence time estimates

The outgroup *Siphonochilus* separated from the remaining Zingiberaceae 72.5 Mya, and *Tamijia* is basal to the remaining Zingiberaceae, separating from them about 57.6 Mya. The mean divergence time for remaining Zingiberaceae is 57.1 (45.5-67.3) Mya.

The divergence time for Zingiberoideae (Clade III) is 45.0 Mya and within clade III, Sri Lankan endemic *Curcuma albiflora* shows an origin dated back to 10.5 (5.0-16.8) Mya.

The mean divergence time estimate for the genus *Zingiber* (Clade V) is 36.7 (26.5-48.7) Mya with the Sri Lankan endemic *Zingiber cylindricum* being the early diverging lineage at 25.7 Mya (10.1-28.4). Indo Sri Lankan *Zingiber wightianum* is nested within

with a divergence time of 19.1 (10.1-28.4) Mya. The Alpiniodeae (Clade VI) shows an initial diversification beginning 46.3 (35.0-57.6) Mya.

The divergence time for clade VII is 35.3 Mya, and clade VIII consisting of Sri Lankan endemic and Afro-American taxa shows a divergence time of 27.3 Mya. Afro-American *Renealmia* separated from the Sri Lankan endemics *Amomum nemorale* and *Cyphostigma pulchellum* 21.5 (11.9-31.8) Mya.
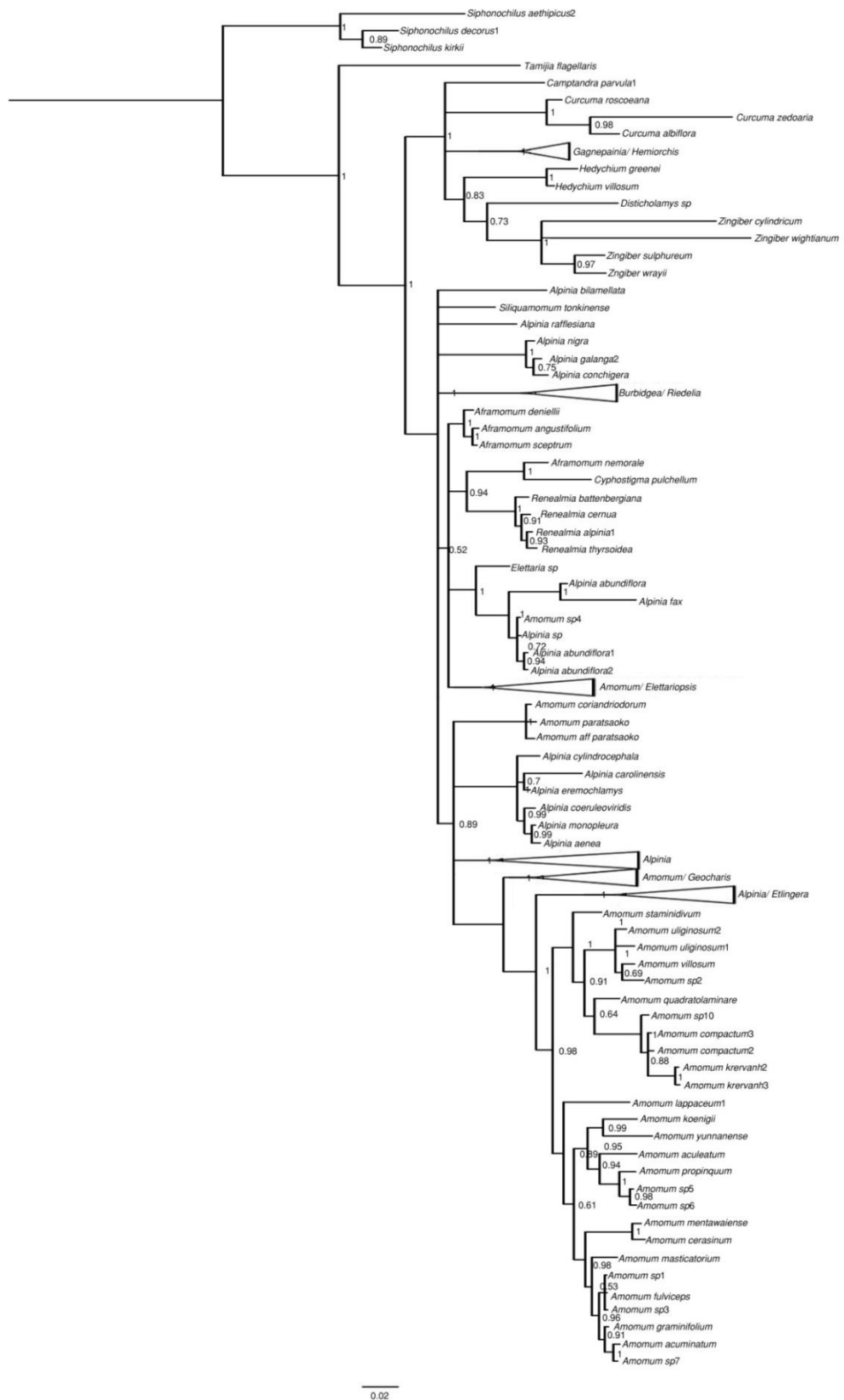
Figure 5.1. Bayesian majority rule consensus tree based on nuclear ITS data. Bayesian posterior probability (PP) support values are indicated next to the nodes.
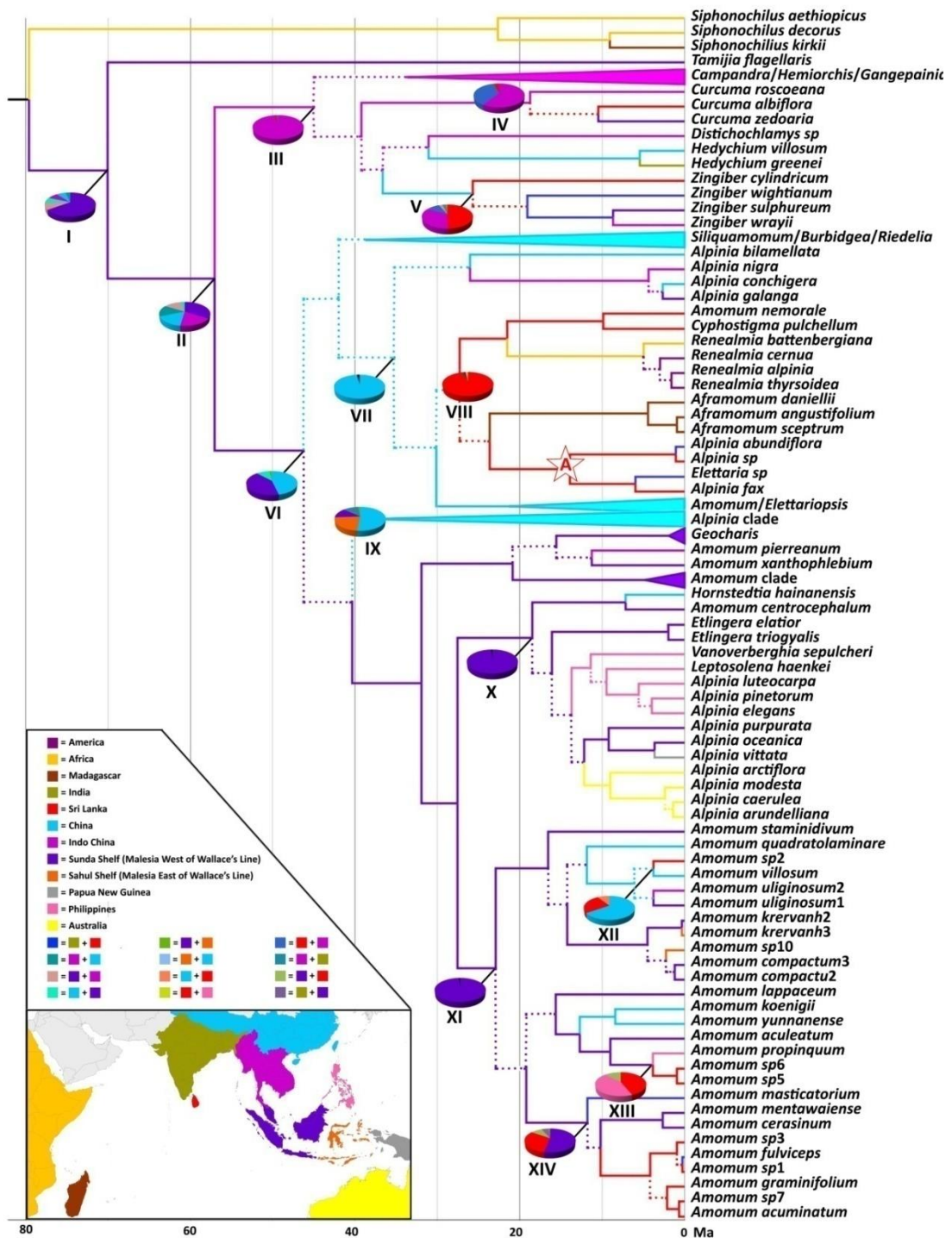
Figure 5.2 Maximum-clade-credibility chronogram of a relaxed molecular clock analysis of the ITS data set. Node heights indicate mean ages. Numbers at nodes represent clades in Table 5.2. Branches coloured according to their optimal range reconstructions under the DEC+J model in biogeobears. Pie charts show the relative probability of ancestral state reconstructions at selected nodes. Dotted lines indicate posterior clade probabilities less than 0.95.
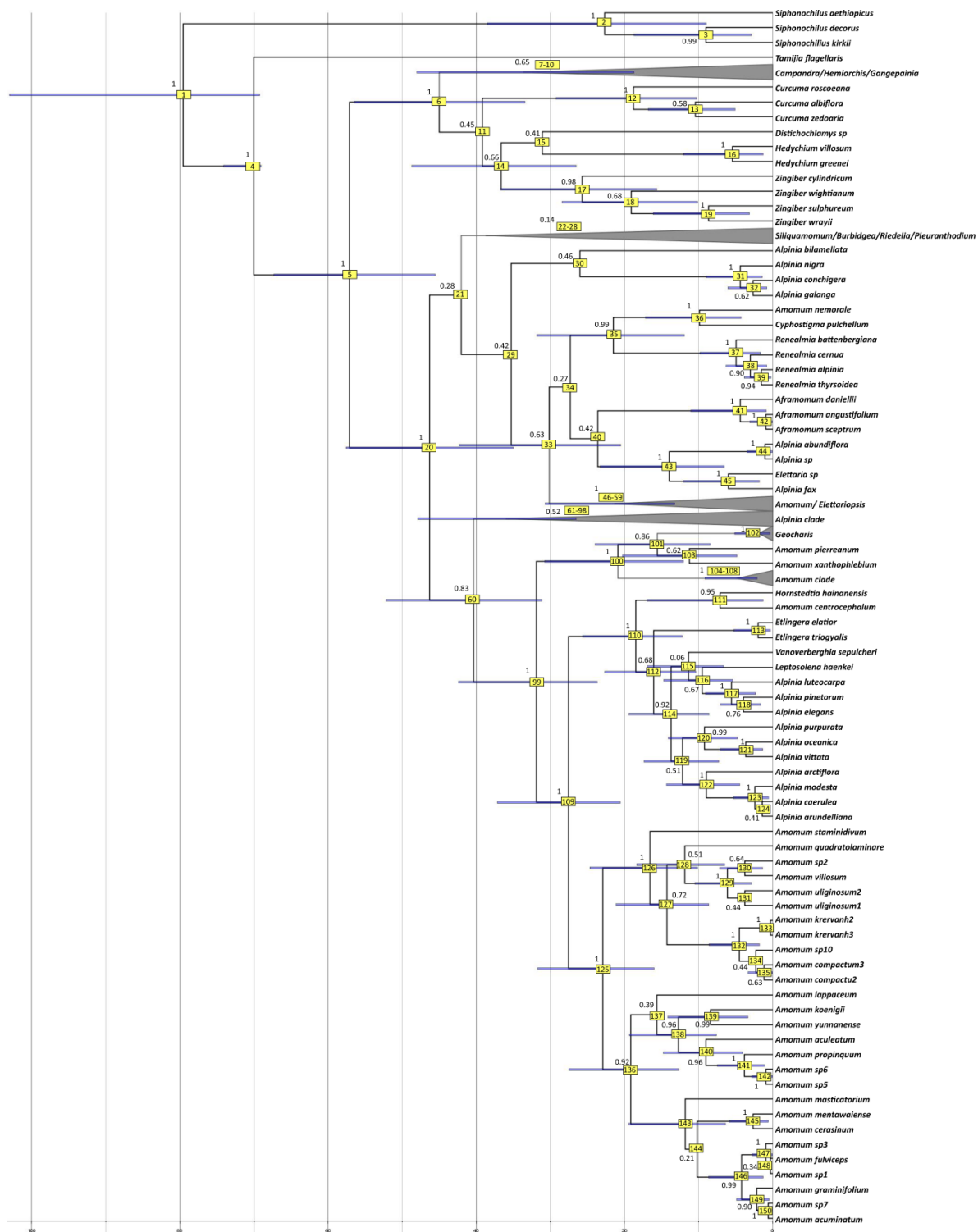
Figure 5.3 Maximum clade credibility chronogram of the ITS dataset. Node heights indicate mean ages. Node bars indicate 95% highest posterior density date ranges. Numbers inside boxes at each node represent node numbers and values next to nodes are the posterior probability values for each node.
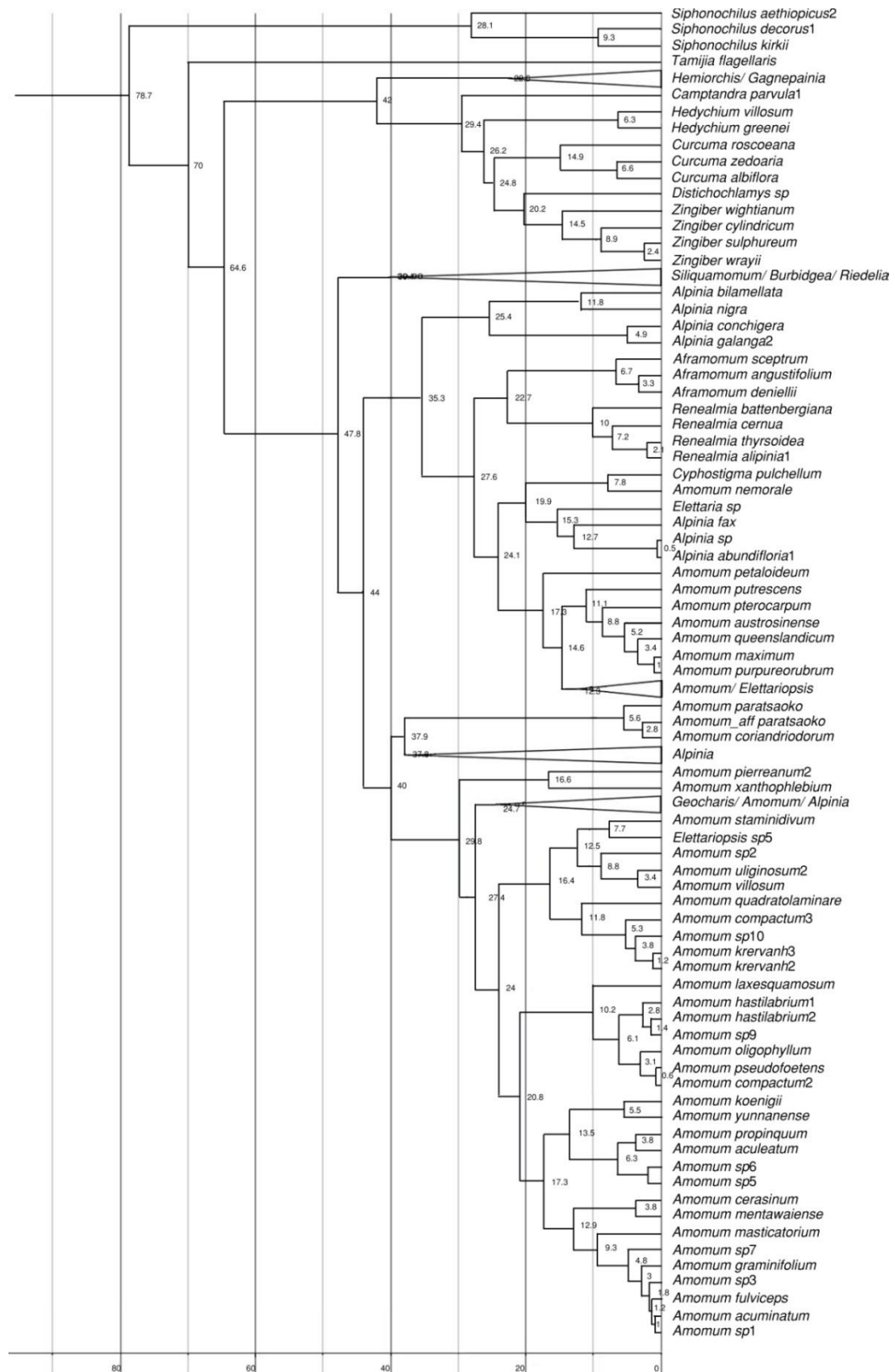
Figure 5.4 Maximum-clade-credibility chronogram of a relaxed molecular clock analysis of the *mat*K data set. Values next to nodes are node age values.

143
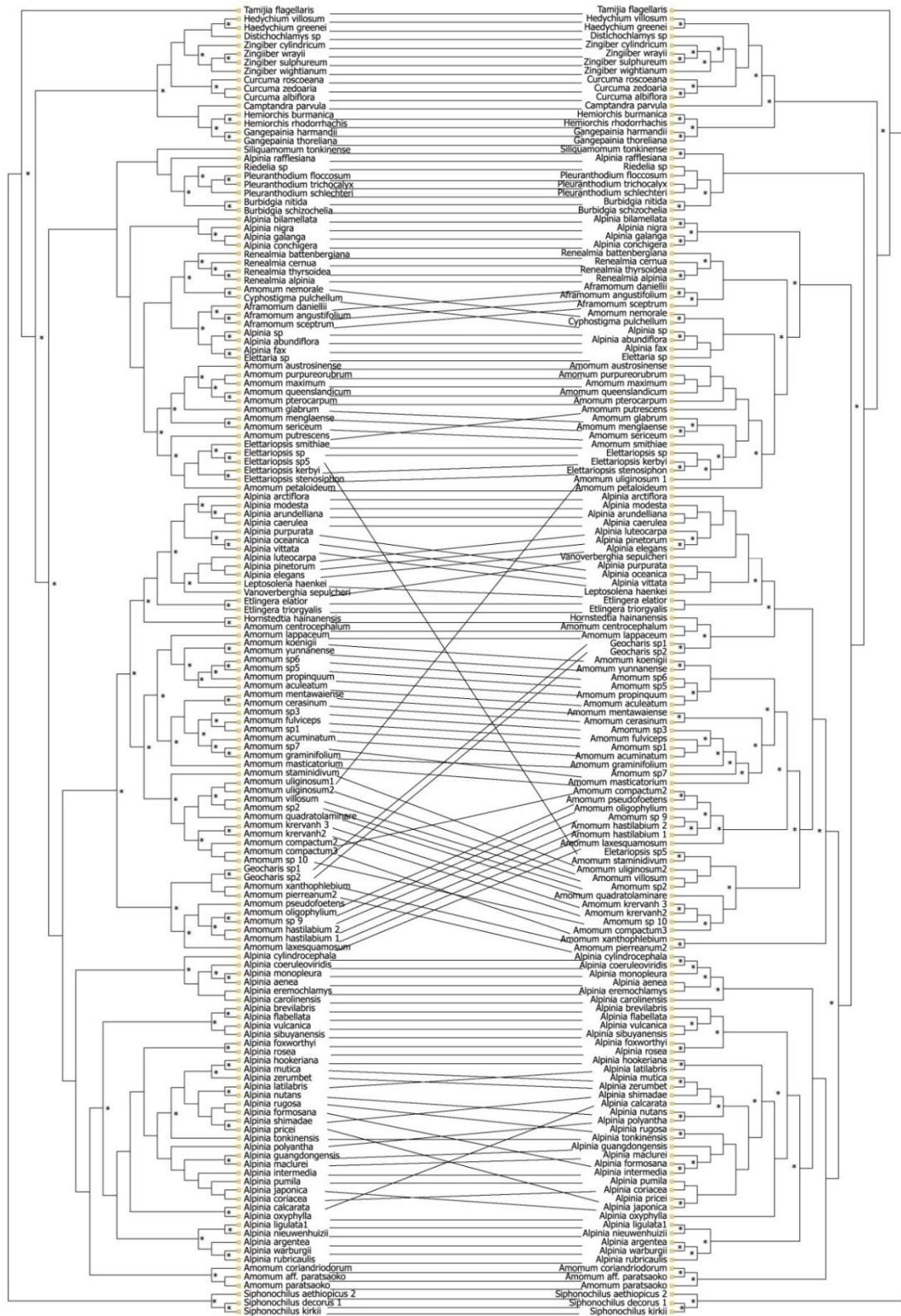
Figure 5.5 Histogram representing the incongruence between ITS and *mat*K phylogenies. Left: ITS right: *mat*K. * indicate posterior clade probabilities greater than 0.9.

Sri Lankan *Alpinia fax* clade (Clade A) separates from African *Aframomum* 23.6 Mya and the *Aframomum* clade shows a recent diversification beginning 4.4 (0.9-11.8) Mya. The Indo-Sri Lankan elements show an initial diversification 13.9 (6.5-23.3) Mya followed by more recent origin of the endemic *Alpinia fax* 6.0 (1.8-12) Mya.

The divergence time estimate for clade IX, consist of *Alpinia* species of Chinese origin, is 36.0 (26.5-47.9) Mya. Within clade X, which began to diversify 18.5 (12.2-25.7) Mya, Australian *Alpinia* date to 9.0 (4.4-14.3) Mya. Sri Lankan endemic *Amomum* are highly nested within clade XI, and show recent separate speciation events, such as *Amomum* sp 2 at 3.8 Ma and *Amomum* sp 5 & 6 more recently, 0.9 Ma. The Indo-Sri Lankan *Amomum masticatorium* is the sister lineage to the remainder of clade XIV which diversified 11.8 (6.4-19.5) Mya with a recent origin of 4.2 (1.2-8.6) Mya for 5 endemic *Amomum* species and the Indo Sri Lankan *Amomum fulviceps*.

According to *mat*K dating (Figure 5.4), African *Aframomum* and Neotropical *Renealmia* shows an initial diversification 22.67 (15.3-32.1) Mya.The divergence time for *Aframomum* is 6.7 (2.18-13.05), while for *Renaelmia* it is 10 (4.59-17.2) Mya.

Sri Lankan *Cyphostigma pulchellum*, *Alpinia* species, *Elettaria* and *Amomum* clade shows an initial diversification 19.94 (10.73-29.43) Mya and later *Cyphostigma* and *Amomum nemorale* has separated from others 7.82 (2.43-16.15) Mya. Indo Sri Lankan *Elettaria* is the basal taxon for *Alpinia* clade which began to diversify 15.27 (7.02-24.38) Mya.

Southeast Asian *Amomum cerasinum* and *Amomum mentawaiense* clade resolves as sister to Sri Lankan *Amomum* which began to diversify 12.92(7.43-19.62) Mya. Indo Sri Lankan *Amomum masticatorium* shows and initial diversification 9.3(4.73-14.88) Mya and the rest of *Amomum* endemics show more recent evolution of 4.7 (1.86-8.66)Mya.

Table 5.5 Posterioir probabilities, Divergance ages and Ancestral Area probabilities; E= Sri Lanka, F= China, G= Indo China, H= Sunda Shelf, I= Sahul Shelf, J= Papua New Guinea, K= Philippines.

Clade no	PP	divergance age	Area probabilities
Clade I	1	70.04 (69.04-74.1)	H= 0.64, GH=0.09, FH=0.08
Clade II	1	57.09 (45.52-67.31)	H=0.33, G=0.18, F=0.16

CladeIII	1	44.96 (33.41-56.5)	G= 0.97, E=0.02, EG= 0.01
CladeIV	1	18.74 (10.24-29.2)	G=0.58, EG=0.37, E=0.04
CladeV	0.98	25.7 (15.63-36.72)	F=0.46, H=0.41, FH=0.1
CladeVI	1	18.95 (10.75-28.85)	H=0.65, HJ=0.32, J=0.02
CladeVII	0.42	35.25	F=0.98
CladeVIII	0.27	27.31	E=0.92, C=0.03, A=0.02
CladeIX	0.52	36.02 (26-49-47.9)	F=0.50, I=0.22, H=0.13
CladeX	1	18.48 (12.19-25.65)	H=0.99
CladeXI	1	22.91 (15.98-31.67)	H=0.99
Clade XII	0.64	3.76 (1.34-7.13)	F=0.65, E=0.26, EF=0.09
Clade XIII	1	3.82 (1.09-7.46)	E=0.44, K=0.44, EK=0.13
Clade XIV	1	11.78 (6.35-19.46)	H=0.53, E=0.31, EH=0.09

Biogeographic analysis and ancestral area reconstructions

The DEC+J model resulted in a higher likelihood value compared to other models DEC, DIVA LIKE, DIVA LIKE+J tested in BGB, thus was chosen as the best fit model for my data (Table 5.6).

The basal node of the core Zingiberaceae in clade I is reconstructed as being Asian to the Sunda Shelf Sunda Shelf (H)= 0.64, Indo China+Sunda Shelf (GH)= 0.09, China+Sunda Shelf (FH)= 0.08 with an age of at least 71 Ma, having split from African and Madagascan lineages in the Siphonochilieae

Table 5.6 d (dispersal), e (extinction), j (J value, founder-event speciation) LnL (log likelihood) for each of the models compared in Biogeobears.

Model	d	e	j	LnL
DEC	0.0033	0.0048	0	-350.03
DEC+J	7e-04	0	0.0217	-297.63
DIVA LIKE	0.0031	4e-04	0	-332.48
DIVA LIKE+J	8e-04	0	0.0186	-298.87

The Zingiberoideae (clade III) have an Eastern Asian origin with Indochina constructed as the ancestral area with probabilities: Indo China (G)= 0.97, Sri Lanka (E)= 0.02, Sri Lanka+Indo China (EG)= 0.01.

The most likely geographic origin for clade VI is China or Sunda Shelf: China (F)= 0.46, Sunda Shelf (H)= 0.41, China+Sunda Shelf (F+H)= 0.1 and the ancestral area for clade VIII is constructed as Sri Lanka: Sri Lanka (E)= 0.92, Madagascar (C)= 0.03, America (A)= 0.02 with African *Aframomum* and Afro-American *Renealmia* nested within. The geographic origin for African *Renealmia battenbergiana* and other *Renealmia* from the Neotropics is Sri Lanka: Sri Lanka (E)=0.79, America+Sri Lanka (AE)=0.07, Africa+Sri Lanka (BE)= 0.07 during late Oligocene and the topology is highly supported. The African *Aframomum* have likely diverged from a Sri Lankan ancestor: Sri Lanka (E)= 0.84, Madagascar+Sri Lanka (CE)= 0.09, Madagascar (C)= 0.07 during the late Oligocene to early Miocene, although the topology at this node is not strongly supported. The ancestral area for the *Alpinia fax* clade is Sri Lanka (E=1.0).

The ancestral area for clade IX is constructed as China with multiple dispersal events between Eastern Asia (China & Indo China), Sunda-Sahul and the Philippines being inferred.

Clade X has an area of origin in Sunda shelf (H=0.99), a single dispersal to Australia during the mid Miocene followed by recent speciation *in situ*.

The ancestral area for clade XI is constructed as Sunda Shelf and within clade XI, four dispersal events have occurred to Sri Lanka (*Amomum* sp 2, *Amomum* sp 5 & 6, and 2 dispersals within sub clade XIV). The geographic origin for the clade XIV is Sunda Shelf with one dispersal to Sri Lanka giving rise to endemic taxa during Pliocene, and one taxon shared between India and Sri Lanka (*Amomum fulviceps*) with a Pleistocene origin in Sri Lanka.

Discussion

Ploidy level, Chromosome numbers in the family

Zingiberaceae shows a wide varation in chromosome numbers ranging from 2n=22 to 2n=96. Polyploidization and hybradization events have played a wide role evolution and diversification of ganera in the family Zingiberaceae. The small size and the large number of chromosome number arises in difficulties in carrying out karyological studies, however a recent study revelas the chromosome numbers of three species,

Alpinia zerumbet, Globba, and *Hedychium*. The somatic chromosome number of *Alpinia zerumbet* has been reported as 2n=52 and the size ranges fron 0.77-3.97μm while in *Gobba marantiana* 2n=52. The somatic chromosome number of *Hedychium spicatum* is reported as 2n=68 with length of chromosomes vary from 0.42-1.7μm (Bhadra & Bandyopadhyay 2016).

Within *Curcuma,* the basic chromosome number has been identified as 2n=40, 42 and the chromosome size ranges from 0.5-2.1μm. However, studies have shown that chromosomes size is correlated with climatic conditions they exist, mainly tropical or sub tropical genera contains smaller chromosomes while large chromosomes have been reported in genera in temperate climates. The genus *Curcuma* is widely spreaded in tropical and sub tropical regions in the world with high species diversity in the Eastern, South Eastern and South Asian region possess a small chromosome like the other genera; *Globba* and *Hedychium*. Triploids (3n) are more common in Chinese *Curcuma* and can be resultdue to the fusion of reduced (n) and unreduced (2n) gamates of diploids within or between species. The triploides are sterile and they might have some type of competitive advantage to distribute in a wide area (Chen & Xia 2013).

The basic chromosome number of *Curcuma* be x=7; some speices are polypolid such as *Curcuma longa* (a nonaploid (9x)2n=63) and *Curcuma amada* (a hexaploid (6x) with 2n=42 chromosomes. The genus *Zingiber* shows less variability in chromosome number and the basic chromosome number is reported as x=11. The diploid chromosome number of *Zingiber officinale* is 2n=22 while that of*Zingiber zerumet* is 2n=22 (Bhadra &Bandyopadhyay, 2016) within Indian *Curcuma* species. Variation of this basic chromosome number and resulting different ploidy levels may have contributed to species adapting to different ecological habitats they live in.

ITS and chloroplast incongruence

Chloroplast markers are usually single copies and rarely present problems of paralogy which can be problematic with the multi-copy genes such as ITS. However some chloroplast markers such as *trn*L-F are less informative in Zingiberaceae due slow rates of nucleotide substitution. Nuclear regions such as ITS aremore rapidly evolving in nature and have a proven wide utility in distinguishing between closely related taxa.

Hybridisation and uni-directionality of gene flow may result in incongruence between nuclear and plastid gene trees. It may also lead to extensive intra-individual ITS polymorphism (Zaveska et al. 2012; Rieseberg et al. 1996; Okuyama et al. 2005). In some genera of Zingiberaceae, hybridization along with the multi copy nature of the nuclear ITS region can result in paraloguous sequences being amplified in PCR and this may make phylogenetic interference difficult. Since direct sequencing may result in unreadable sequences, cloning is sometimes necessary for ITS region to reveal the broad range of polymorphisms within the individuals (Zaveska et al. 2012).

In some genera like *Curcuma*, polymorphism within individuals has been observed, commonly resulting in unreadable DNA sequences for the ITS region. It has been observed that the polyploid origin, homoploid hybridization and mode of reproductionhave a high impact on this gene paralogy in Zingiberaceae (Zaveska et al. 2012). According to their results, the basal lineages in the phylogeny of *Curcuma* have very low variation and are represented by single terminal sequences in theirphylogeny, which suggests a single ITS paralogue prevails in the genome. Hybrids and polyploid speciation is rare in these basal groups and they mainly reproduce asexually. However, some terminal taxa showed high levels of ITS polymorphism and they are mainly hybrids or allopolyploids. That is likely the cause of the high intra specific variation and ITS paralogues in those individuals, however it is independent from the ploidy level of a particular taxa (Zaveska et al. 2012).

In this study, the ITS sequences generated did not present any polymorphisms, and all were clear reads and did not highlight a need for cloning. It seems the problems of paralogy when using ITS in Zingiberaceae are mostly restricted to the *Curcuma* group.

Origin of the pantropical distribution of Zingiberaceae

The most basal lineages of Zingiberaceae are found in Africa (*Siphonochilus*) and Borneo (*Tamijia*) and *Siphonochilus* splits from the rest of Zingiberaceae 70.0 (69.0-74.1) Mya. Following the Cretaceous thermal maximum, global cooling 80-70 Mya resulted in the disappearance of megathermal taxa from the nothern hemisphere and subsequent dispersals of these taxa to the equatorial regions such as Southeast Asia. The greatest specific and generic diversity in the Zingiberaceae, and within the Alpinieae, is found in Continental Asia, to the west of Wallace's Line and to the north of the Isthmus

of Kra. The subfamily Alpinoideae has its centre of diversity in the Malesian region while that of Zingiberoideae lies in the monsoonal areas of Indochina and Thailand, which have been called the evolutionary centre of the Zingiberaceae due to its exceptional diversity in those areas (Larsen 2005).

Sunda-Sahul exchange

The more recently diverged Alpinieae (Clade IX, Fig. 5.2) show a migration from west to east across Wallace's Line, with many of these dispersals having occurred during the Miocene, however the topology at the base of the clade is poorly supported in both ITS and chloroplast trees. A similar pattern of dispersals can be seen in *Begonia* (Thomas et al. 2012), *Cyrtandra* (Gesneriaceae, Clark et al. 2008) and the Isonandreae (Sapotaceae, Richardson et al. 2014). Extensive land connections between Peninsular Malaysia and islands of Borneo, Java and Sumatra will likely have facilitated multiple transitions between these land masses (Voris 2000; Sathiamurthy & Voris 2006). Further, much of the Sunda Shelf and Penninsular Malaysia were covered by rain forests since the Miocene, thus should have provided suitable habitat for establishment and diversification of invading mega thermal taxa.

The timing of the split of *Alpinia oceanica* and *A. vittata* at 9.2 (4.8-14.0) Mya and their occurrence in New Guinea to the east of Wallace's line is consistent with a dispersal from Sunda shelf during the late Miocene, potentially facilitated by island hopping over the land masses that emerged in Sulawesi and New Guinea and the emergence of volcanic islands along the Sunda Arc, the Banda Arc and the Halmahera Arc (Hall 2001; Hall 2009). The Phillipine clade also shows Sunda Shelf origin 13.7 (8.6-19.4) Mya, and is best explained by long distance dispersal or island hopping via Palawan or the Sulu Archipelago between Borneo or Sangihe Arc between Sulawesi and the Philippines (Jones & Kennedy 2008).

Australian Zingiberaceae are monophyletic and are nested within clade X which has an origin in Sunda Shelf during the middle Miocene. The collision of the Australian plate with Pacific, Philippine and Sunda plates during the Oligo-Miocene boundary resulted in the formation of New Guinea and the islands of the Banda Arc in Indonesia (Morley 2000; Morley 2003) during the Mid Miocene. These islands provide a potential route for flora to disperse across Wallace's line, and the existence of flora with Australian

affinity in the Sunda shelf at present provide evidence for this dispersal route (Morley 2000). The majority of plant dispersals were into the Southeast Asian region from Australia rather than the opposite direction, such as *Phormium* or *Dianella* (Hemerocallidaceae) *Dacrydium* (Podocarpaceae) *Camptostemon* (Bombacaceae) *Dacrycarpus* (Podocarpaceae) (Morley 2003). However there are few examples of dispersals from the Southeast Asian area such as *Stenochlaena palustris* (Blechnaceae), Acacia, *Caesalpinia* and *Crudia* (Leguminosae) and *Merremia* (Convolvulaceae) observed during Miocene (Morley 2003).

Warmer and wetter conditions prevailed in the Sunda Shelf, resulting in latitudinal expansion of rain forests and the decline of the mesic biome and extinction of rain forest lineages during cenozoic in Australia, providing niches available for immigrants from Sunda Shelf (Kershaw et al. 1994; Greenwood & Christophel 2005; Crayn et al. 2014). Further, Sahul-Sunda collision during the Mid Miocene resulted in extension of lands east of Wallaces line, enhancing pathways for stepping-stone migration of plant lineages between Sahul and Sunda which is evidenced by an increase in floristic exchange rate from 12 Mya (Crayn et al. 2014). My results provide supportive evidence for the contributions of immigrations from Sunda shelf to Australian tropical flora which is in contrast to earlier views on Gondwanan origin for Australian flora.

According to the molecular age estimates, many of the Southeast Asian Alpinieae lineages in clades X and XI show a recent Plio-Pleistocene diversification associated with the climatic and sea level fluctuations and shifts in forest distribution (Cannon et al. 2009; Woodruff 2010).

Origin and diversification of Sri Lankan Zingiberaceae

My results strongly contradict a Gondwanan origin for Sri Lankan Zingiberaceae, which consist of much younger lineages ranging dating from 25.7 (15.6-36.7) Mya to 3.8 (1.3-7.1) Mya.

The Sri Lankan species of *Curcuma* and *Zingiber* are nested within species from China and Indochina and have an ancestral area reconstructed as Eastern Asia during the Oligo-Miocene boundary approximately 25 Ma. The collision of the Deccan plate, composed of both India and Sri Lanka, with the southern coast of Laurasia during the

Eocene between 55-40 Mya (Briggs et al. 2003; Aitchison et al. 2008; Morley 2003) resulted in the the mixing of floras from Deccan Gondwana and Laurasia (Ashton & Gunatilleke 1987; Morley 2000). A hypothesis of a Laurasian migration of these Zingiberaceae lineages is congruent with the timing and geographic origins observed in this study (Figure 5.2). More seasonal, monsoonal climates prevailed in nothern India during the early Miocene, becoming more ever-wet in the mid-Miocene (Morley 2000). This would have provided habitat for the immigrants from China-Indochina to diversify in evergreen rain forests in the Indo-Sri Lankan region.

My results confirm the Sundanian origin for Sri Lankan *Amomum*species. One hypothesis for this dispersals phase between Sunda Shelf and Indo-Sri Lankan region is believed to occur during Indian plate's northward migration. The Indian plate moved extremely close to the Malay Peninsula during its northward migration during the mid Eocene (Hall 2001) and the Indian plate and parts of Southeast Asia would have been at similar latitudes during the mid-Miocene. The collision with Laurasian coast 35 Mya brought the northeastern corner of the sub-continent into glancing contact with Sumatra and to a same climatic belt giving the potential for floristic exchange (Ali & Aitchison 2008). However, the Indian flora was much more aggressive in migration, thus a sudden appearance of Indian Paleocene and Early Eocene of taxa are evidenced in palaeo floras which resulted in the depletion of local Southeast Asian flora (Morley 2003; Morley 2000).

However, my date estimates do not support above dispersal phase, instead three independent long-distance dispersals occurred later are evidenced; ca 11.8 (6.3-19.4) Mya, 10.2 Mya, 3.8 (1-7.4) Mya and further diversification of endemics more recently associated with changing climatic conditions during glacial-interglacial periods during the Pliocene-Pleistocene. *Amomum* sp 2 shows a more recent (3.8 Mya) origin in China, thus a long distance dispersal or overland migration during Pliocene is inferred. *Amomum* species show a wide distribution in China- Indo China, India and Sri Lanka, thus, overland migration is a most plausible explanation for extant distributuion for the genus. The genus exhibits a diversity of fruits characters, smooth, echinate, winged, lobed and ridged. The method of dispersal of the fruits is not known for most *Amomum* species. However, the fleshy capsules and aromatic arils provide a hint for animal

dispersal, thus overland migration by animals may probably be the best explanation for extant distribution of the genus.

Pollination and seed dispersal mechanisms

Little is known about pollination of *Amomum* however, the majority of species possess characters which provide hints for animal pollination. Some trumpet shaped flowers with orange or yellow colouration like *Amomum aculeatum* are pollinated by bees. Some species such as the Sri Lankan endemic *Amomum nemorale* contains nectar guides for pollinators as sharp bright makings on labellum which facilitates bee landing. *Amomum subulatum* is a bumble bee pollinated species (Kishore et al. 2011) and *Amomum maximum* is pollinated by two species of honey bees, *Apis dorsata* and *Apis cerana* (Ren et al. 2007). Three pollination syndromes have been identified in Bornean Zingiberaceae; *Amegilla* bees are the only pollinators for *Amomum calyptrum, Amomum gyrolophos* and *Amomum oligophyllum* (Sakai et al. 1999). *Amomum roseisquamosum* has a long floral tube and is pollinated by spiderhunters, while *Amomum polycarpum* and other related species are pollinated by birds and various insects (Sakai et al. 1999). Large, dark red, showy bracts characterise the vertically standing inflorescences found in*Amomum apiculatum* and *Amomum centrocephalum,* which are pollinated by birds (Nagamasu & Sakai 1996). Many of the Sri Lankan *Amomum* species such as *Amomum masticatorium* and *Amomum fulviceps* possesses petaloid anther crests that might be involved in the attraction of as yet unknown pollinators.

Two types of flower colours can be seen in Neotropical *Renealmia;* one with bright red-pink flowers and the other group with white coloured flowers. Species such as *Renealmia alpinia* and *Renealmia cernua* have a tubular labellum, bright red-pink flowers and well developed nectary glands which are pollinated by hummingbirds (Maas, 1977) and it correlates with the high species diversity of hummingbirds in the Andes (Sarkinen et al. 2007). The other group of *Renealmia* with yellow/white flowers and poorly developed nectary glands are assumed to be pollinated by small insects such as bees (Maas 1977).

The majority of the species in the genus *Alpinia* are yellowish white in colour and those are pollinated by large bees, however some species are pollinated by bats and birds (Zhang et al. 2003; Kress &Spetch2006). Different mechanisms have evolved to reduce

self-pollination, such as different levels of styles and different timings of anther dehiscence have been observed (Zhang et al. 2003).

Since most of the species in Zingiberaceae flower near to the ground, animals are likely the most important vectors in fruit and seed dispersal (Pfeiffer et al. 2004). Fruits of *Amomum* are berries, size ranging from 4mm (*Amomum oligophyllum*) to 4.5 cm and covered with a fleshy aril. Some like *Amomum pterocarpum* have hairy fruits which can be trapped on bird's legs and travel long distances. Fruits of *Zingiber cylindricum* are brightly red coloured with a fleshy aril and borne at the base of the plants. Seeds and fruits are eaten by small mammals like rodents and monkeys and can be transported to long distances with their gut contents.

The genus *Aframomum* in African forests produces bright red, fleshy fruits with a sweet juicy pulp which is a reward for primates and other small mammals (Harris et al. 2000). Three different flower types: platform, tube and flag are present and pollination is mainly carried out by birds, bees and butterflies. An uncommon dispersal mechanism is exhibited by *Aframomum psuedostipulare*, which is dispersed by a fish that migrates to forest during flooding (Harris et al. 2000).

Interestingly, it has been observed that the major seed dispersal in *Globba* in Borneo is occurred with the aid of ants. The fleshy aril provides food for ants and allows easy handling. Dispersal by one species of ant (*Polyrhachis sp.*) has been observed to carry seeds about 800cm(Pfeiffer et al. 2004).

Sri Lanka's influence on African and Afro-American Zingiberaceae

Renealmia resolves as monophyletic (P=1) with the African species *R. battenbergiana* being the sister to other *Renealmia* species from the Neotropics. The genus forms a strongly supported clade (P=0.99) with the Sri Lankan endemics *Amomum nemorale* and *Cyphostigma pulchellum*. *Renealmia* was found by Sarkinen et al. (2007) to have the *Alpinia fax* clade as an unsupported sister group, with *Aframomum* or *Amomum-Elettariopsis* clades also as potential sister group to *Renealmia* due to poorly supported tree topology. This is the first instance that the endemics *Cyphostigma pulchellum* and *Amomum nemorale* have been incorporated in a phylogeny, and shows that the amphiatlantic *Renealmia* is nested within a clade with Sri Lanka as the ancestral area

(E=0.79, AE=0.07, BE= 0.07) The Afromadagascan genus *Aframomum*, the largest genus of African Zingiberaceae with approximately 80 species and one of the largest genera of African rainforest understory herbs (Harris et al. 2000), too exhibits a similar pattern, being nested within a clade which has Sri Lanka as the ancestral area.

The hypothesis for these Sri Lanka/Africa exchanges being caused by early dispersal during rafting of the Deccan plate is refuted as the divergence between the Sri Lankan and African lineages are too young. A possible scenario for tropical disjunctions between Africa and Asia is overland migration after the Afro-Arabian plate collided with Asia approximately 25 Mya (Kulju et al. 2007; Li et al. 2009; Yuan et al. 2005). This dispersal route out of Africa in to Southeast Asia is evidenced in some plants such as *Uvaria* (Annonaceae) (Zhou et al. 2012) and in some Primates (Stewart & Disotell 1998; Zhou et al. 2012). The closing of the Tethys Sea and the formation of the Gomphotherium Land Bridge occurred later during 19 – 16 Ma, during which time the mid-Miocene Thermal Maximum would have facilitated the movement of megathermal species from Asia across the Arabian Peninsula into Africa (Zhou et al. 2012). The entrance of *Renealmia* in to Africa 21.5 (11.9-31.84) during the late Oligocene is congruent with this migratory path. Global temperatures were high during the Eocene climatic optimum 55-50 Mya (Zachos et al. 2001), and then cooler until late Oligocene, when temperatures started rising again, peaking at the mid-Miocene Thermal Maximum (17–15 Ma; Zachos et al. 2001) which potentially facilitated tropical and subtropical vegetation becaming more widespread in Africa (Andrews & van Couvering 1975; Coetzee 1978) thus providing favourable habitats for immigrants.

However long distance dispersal from Sri Lanka directly to Africa is another potential explanation, which could occur at any point in geological time. Many African–Asian plant disjunctions have been explained by reference to long-distance dispersal, such as *Exacum* (Gentianaceae) (Yuan et al. 2005), *Bridelia* (Phyllanthaceae) (Li et al. 2009), *Macaranga* and *Mallotus* (Euphorbiaceae) (Kulju et al. 2007), *Osbeckia* (Melastomataceae) (Renner & Meyer 2001; Renner 2004).

The African genus *Renealmia* hasbeen shown to have very recent diversification in the Neotropics (Särkinen et al. 2007) which is congruent with the Pliocene date in this study for the split of African and Neotropical *Renealmia*. A similar, recent scenario applies to

the diversification of African *Aframomum*, which also has evolved from a Sri Lankan ancestor during the late Oligocene. *Aframomum* are dispersed by primates and mammals which feed on fruits and seeds, leading to them potentially being carried for long distances (Wrangham et al. 1994) and they are well adapted in exploiting new habitats and changing forest environments during Pleistocene (Harris et al. 2000). Thus the rapid diversification in these rain forest taxa should be a result of speciation in response to Pleistocene climatic fluctuations, which were severe in Africa, also evident in some other families like Begoniaceae (Plana et al. 2004). Short distance dispersals of these species is facillitated by ants, however birds such as toucans and also mammals can transport seeds over longer distances (García-Robledo & Kuprewicz 2009). Such LDD events have more influence on ranges of taxa than numerous small steps available through local dispersals or range expansions after climate shifts (Nathan 2006).

Renealmia is well represented both sides of the Atlantic with c. 15 species in Africa and 61 in the Neotropics (Maas & Maas 1987; Maas & Maas 1990). Due to the young ages of the lineages, its amphi Atlantic distribution does not favour either the Gondwanan vicariance, dispersal via amphi-Atlantic land-bridges or the boreotropical dispersal route which could have had occurred approximately 35 Mya (Särkinen et al. 2007). Thus, more recent oceanic long distance dispersal is the most plausible explanation for the dispersal from Africa to South America during the Miocene or Pliocene (15.8–2.7 Mya ago) and further speciation more recently during Pleistocene.

The origin of the Sri Lankan Zingiberaceae post dates a Gondwanan origin and shows a more recent entry to the island and diversification of lineages. My results further support the current view on long distance dispersals in being pivotal to the evolution of the world's tropical flora, in contrast to earlier views regarding the vicariance paradigm. Further it provides support for a growing body of evidence that the rainforests are youthful at the species level while they act as museums for higher level taxa to conserve traits over a long period of time.

References

Aitchison, J.C., Ali, J.R. & Davis, A.M., 2008. When and where did India and Asia collide? *Geological Bulletin of China*, 27(9), pp.1351–1370.

Ali, J.R. & Aitchison, J.C., 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166-35??Ma). *Earth-Science Reviews*, 88(3-4), pp.145–166.

Andrews, P. & van Couvering, J.A.H., 1975. Palaeoenviron ments in the East African Miocene. *Approaches to primate paleobiology* (ed. by X. Szalay), pp. 62–103. *Contribution to Primate Biology*, Vol. 5, Karger, Basel.

Ashton, P.S. & Gunatilleke, C.V.S., 1987. New light on the plant geography of Ceylon I . Historical plant geography. *Journal of Biogeography*, 14(3), pp.249–285.

Batalha-Filho, H. et al., 2014. Phylogeny and historical biogeography of gnateaters (Passeriformes, Conopophagidae) in the South America forests. *Molecular Phylogenetics and Evolution*, 79, pp.422–432.

Bhadra, S. & Bandyopadhyay, M., 2016. New chromosome number counts and karyotype analysis in three important genera of Zingiberaceae. *Nucleus*, 59(1), pp.35-40.

Briggs, J.C., Collections, Z. & Museum, G., 2003. The biogeographic and tectonic history of India. *Journal of Biogeography*, 30, pp.381–388.

Cannon, C.H., Morley, R.J. & Bush, A.B.G., 2009. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), pp.11188–11193.

Chen, X. & Xia, N., 2013. Chromosome Numbers and Ploidy Levels of Chinese *Curcuma* Species. *HORTSCIENCE,* 48(5), pp.525-530.

Clark, J.R. et al., 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Systematic biology*, 57(5), pp.693–707.

Coetzee, J.A., 1978. Climatic and biological changes in south- western Africa during the late Cainozoic. *Palaeoecology of Africa*, 10, pp.13–29.

Crayn, D.M., Costion, C. & Harrington, M.G., 2014. The Sahul – Sunda floristic exchange : dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography*, 42, pp.11–24.

Droop A.J. 2012. Systematic and biogeographic studies in the genus *Amomum* Roxb . (Zingiberaceae) in Sumatra. PhD thesis, University of Aberdeen and Royal Botanic Gardens Edinburgh, February 2012.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.214.

García-Robledo, C. & Kuprewicz, E.K., 2009. Vertebrate Fruit Removal and Ant Seed Dispersal in the Neotropical Ginger Renealmia alpinia (Zingiberaceae). *Biotropica*, 41(2), pp.209–214.

Greenwood, D.R. & Christophel, D.C., 2005. The origins and Tertiary history of Australian "tropical" rainforests. *Tropical rainforests: past, present, and future* (ed. by E. Ber- mingham, C.W. Dick and C. Moritz), pp. 336–373. Chi- cago University Press, Chicago.

Hall, R., 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: Changing patterns of land and sea. *Faunal and Floral Migrations and Evolution in Se Asia-Australasia*, (Figure 1), pp.35–56.

Hall, R., 2009. Southeast Asia's changing palaeogeography. *Blumea - Biodiversity, Evolution and Biogeography of Plants*, 54(1), pp.148–161.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98.

Harris, D. J., et al., 2000. Rapid radiation in Aframomum (Zingiberaceae): evidence from nuclear ribosomal DNA internal transcribed spacer (ITS) sequences. *Edinburgh Journal of Botany*, 57, pp.377–395.

Jones, A.W. & Kennedy, R.S., 2008. Evolution in a tropical archipelago: comparative phylogeography of Philippine fauna and flora reveals complex patterns of colonization and diversification. *Biological Journal of the Linnean Society*, 95, pp.620–639.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), pp.1647-1649.

Kershaw, A.P. et al., 1994. The Neogene: a period of transition. *History of the Australian vegetation: Cretaceous to Recent* (ed. by R.S. Hill), pp. 299–327. Cambridge University Press, Cambridge, UK.

Kishore, K. et al., 2011. Pollination studies in large cardamom (Amomum subulatum Roxb.) of Sikkim Himalayan region of India. *Scientia Horticulturae*, 129, pp.735–74

Kress, W. J., 1995. Phylogeny of the Zingiberanae: Morphology and molecules. Pages 443–460 inMono- cotyledons: Systematics and evolution (P. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries, eds.). Royal Botanic Gardens, Kew, U.K.

Kress, W.J. & Specht, C.D., 2006. The evolutionary and biogeographic origin and diversification of the tropical monocot Oder Zingiberales. *Aliso*, 22, pp.621–632.

Kress, W.J. et al., 2001. Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. Systematic Biology, 50(6):926–944.

Kress, W.J. et al., 2005. The molecular phylogeny of Alpinia (Zingiberaceae): A complex and polyphyletic genus of gingers. *American Journal of Botany*, 92(1), pp.167–178.

Kress, W.J., 2002. The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Amer. J. Bot*, 89, pp.1682–1696.

Kulju, K.K.M. et al., 2007. Molecular phylogeny of Macaranga, Mallotus, and related genera (Euphorbiaceae s.s.): Insights from plastid and nuclear DNA sequence data. *American Journal of Botany*, 94(10), pp.1726–1743.

Larsen, K., 2005. Distribution patterns and diversity centres of Zingiberaceae in SE Asia. *Biologiske Skrifter*, 55, pp.219–228, 2005.

Li, Y. et al., 2009. More Miocene Dispersal Between Africa and Asia—the Case of *Bridelia* (Phyllanthaceae). *Systematic Botany*, 34(3), pp.521–529.

Manos, P.S., & Steele, K.P., 1997. Phylogenetic analyses of ''higher'' Hamamelididae based on plastid sequence data. *American Journal of Botany,* 84, pp.1407–1419.

Maas, P.J.M., 1977. Flora Neotropica Monograph No. 18: Renealmia (Zingigeraceae–Zingiberoideae) Costoideae (Additions) (Zingibera- ceae). New York Botanical Garden, New York.

Maas, P.J.M., Maas, H., 1987. Notes on New World Zingiberaceae: III— Some new species in Renealmia. *Notes Roy. Bot. Gard. Edinburgh,* 44, pp.237–248.

Maas, P.J.M., Maas, H., 1990. Notes on New World Zingiberaceae: IV— Some new species of Costus and Renealmia. *Notes Roy. Bot. Gard. Edinburgh,* pp. 46, 307–320.

Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow im-proved accuracy and model-testing. *Frontiers of biogeography,* 5(4), pp.242-248.

Matzke, N.J., 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, 63(6), pp.951–970.

Miller, M.A. et al., 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics,* 11, pp,43-48.

Morley, R. J., 2000. *Origin and evolution of tropical rain forests*. Chichester: John Wiley & Sons, Ltd.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(12), pp.5–20.

Nagamasu, H. & Sakai, S., 1996. Amomum roseisquamosum (Zingiber- aceae), a new epiphytic ginger from Borneo. *Edinburgh Journal of Botany*, 53(1), pp.39–42.

Nathan, R., 2006. Long-distance dispersal of plants. *Science (New York, N.Y.)*, 313(August), pp.786–788.

Okuyama, Y., et al., 2005. Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian Mitella (Saxifragaceae). *Molec. Biol. Evol*. 22, pp. 285-296.

Peppe, D.J. et al., 2007. Fossil leaf species from the Fox Hills Formation (Upper Cretaceous: North Dakota, USA) and their paleogeographic significance. *Journal of Paleontology*, 81(3), pp.550–567.

Pfeiffer et al., 2004. Myrmecochory in the Zingiberaceae: seed removal of *Globba franciscii* and *G. propinqua* by ants (Hymenoptera-Formicidae) in rain forests on Borneo. *Journal of Tropical Ecology*, 20, pp.705-708.

Plana, V. et al., 2004. Pleistocene and pre-Pleistocene Begonia speciation in Africa. *Molecular Phylogenetics and Evolution*, 31, pp.449–461.

Posada, D. et al., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, pp.772–772.

Rambaut, A., 2009. FigTree, a graphical viewer of phylogenetic trees. *Institute of Evolutionary Biology University of Edinburgh*.

Ree, R.H. et al., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution; international journal of organic evolution*, 59(11), pp.2299–2311.

Ree, R.H. & Smith, S. a, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57(1), pp.4–14.

Ren, P.Y. et al., 2007. An example of flexistyly in a wild cardamom species (Amomum maximum (Zingiberaceae). *Plant Systematics and Evolution*, 267, pp.147–154.

Renner, S.S., 2004. Multiple Miocene Melastomataceae dispersal between Madagascar, Africa and India. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1450), pp.1485–1494.

Renner, S.S. & Meyer, K., 2001. Melastomeae come full circle: biogeographic reconstruction and molecular clock dating. *Evolution; international journal of organic evolution*, 55(7), pp.1315–1324.

Richardson, J.E. et al., 2014. The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). *Botanical Journal of the Linnean Society*, 174(1), pp.130–140.

Rieseberg, W. et al., 1996. Molecular marker incongruence in plant hybrid zomes and phylogenetic trees. *Acta Bot. Neerl*. 45, pp. 243-262.

Ronquist, F., 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(l), pp.195–203.

Ronquist, F. et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, pp.539–42.

Sakai, S. et al., 1999. Three pollination guilds and variation in floral characteristics of Bornean gingers (Zingiberaceae and Costaceae). *American Journal of Botany*, 86(5), pp.646–658.

Särkinen, T.E. et al., 2007. Recent oceanic long-distance dispersal and divergence in the amphi-Atlantic rain forest genus Renealmia L.f. (Zingiberaceae). *Molecular Phylogenetics and Evolution*, 44, pp.968–980.

Sathiamurthy, E. & Voris, K.H., 2006. Maps of Holocene Sea Level Transgression and Submerged Lakes on the Sunda Shelf. *Nat Hist J Chulalongkorn University*, Supplement, 2, pp.1–43. Maps available at http://fmnh.org/research_collections/zoology/zoo_sites/seamaps/.

Selvaraj et al., 2008. Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. *Bioinformation,* 3(1), pp. 24-27.

Steele & Vilgalys, 1994.Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the Plastid gene matK.*Systematic Botany,* 19, pp.126-142.

Stewart, C.B. & Disotell, T.R., 1998. Primate evolution - in and out of Africa. *Current Biology*, 8(16), pp.R582–588.

Thomas, D.C. et al., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, 39(1), pp.98–113.

Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, 27, pp.1153–1167.

White, T.J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR protocols: a guide to methods and applications.* London: Academic Press, pp.315-322.

Woodruff, D.S., 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation*, 19(4), pp.919–941.

Xia, Y. M. et al., 2004. A phylogenetic analysis of *Amomum* (Alpinioideae: Zingiberaceae) using ITS and matK DNA sequence data. *Syst. Bot*. 29, pp.334–344.

Yuan, Y.-M. et al., 2005. Phylogeny and biogeography of exacum (gentianaceae): a disjunctive distribution in the Indian ocean basin resulted from long distance dispersal and extensive radiation. *Systematic biology*, 54(1), pp.21–34.

Zachos, J. et al., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (New York, N.Y.)*, 292(April), pp.686–693.

Zaveska, E. et al., 2012. Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata. *Taxon,* 61(4), pp.747-763.

Zhang, L.B. & Renner, S., 2003. The deepest splits in Chloranthaceae as resolved by chloroplast sequences. *Int. J. Plant Sci*. 164 (5 Suppl.), S383–S392.

Zhou, L. et al., 2012. "Out-of-Africa" dispersal of tropical floras during the Miocene climatic optimum: Evidence from Uvaria (Annonaceae). *Journal of Biogeography*, 39(2), pp.322–335.

CHAPTER 6: Summary and Conclusions

During the present study 42 species (85.7% sampling from Sri Lankan representatives) were collected from three families from different pre identified localities in Sri Lanka. The collection list and the newly generated sequences are listed in tables 6.1 and 6.2 respectively.

Table 6.1 Species collected in Begoniaceae, Sapotaceae and Zingiberaceae during the field work in Sri Lanka. Endemic species are indicated in bold.

Family	Genera	Species present	Species collected	% sampling
Begoniaceae	*Begonia*	***Begonia tenera*** ***Begonia thwaitesii*** *Begonia cordifolia* *Begonia dipetala* *Begonia malabarica* *Begonia subpeltata*	***Begonia tenera*** ***Begonia thwaitesii*** *Begonia cordifolia* *Begonia dipetala* *Begonia malabarica*	Endemic-100% Native- 75%
Sapotaceae	*Palaquium*	***Palaquium pauciflorum*** ***Palaquium thwaitesii*** ***Palaquium grande*** ***Palaquium hinmolpedda*** ***Palaquium zeylanicum*** ***Palaquium rubiginosum*** ***Palaquium petiolare*** ***Palaquium laevifolium*** ***Palaquium canaliculatum***	***Palaquium pauciflorum*** ***Palaquium thwaitesii*** ***Palaquium grande*** ***Palaquium hinmolpedda*** ***Palaquium zeylanicum*** ***Palaquium rubiginosum*** ***Palaquium petiolare*** ***Palaquium laevifolium*** ***Palaquium canaliculatum***	Endemic-100%
	Isonandra	*Isonandra lanceolata* *Isonandra compta* ***Isonandra montana*** ***Isonandra zeylanica***	*Isonandra lanceolata* *Isonandra compta* ***Isonandra montana*** ***Isonandra zeylanica***	
	Madhuca	***Madhuca fulva*** ***Madhuca moonii***	***Madhuca fulva*** ***Madhuca clavata***	Endemic- 60% Native- 100%

		Madhuca microphylla	*Madhuca neriifolia*	
		Madhuca clavata	*Madhuca longifolia*	
		Madhuca neriifolia		
		Madhuca longifolia		
	Manilkara	*Manilkara hexandra*	*Manilkara hexandra*	Native- 100%
	Mimusops	*Mimusops elengi*	*Mimusops elengi*	Native – 100%
Zingiberaceae	*Zingiber*	*Zingiber cylindricum*	*Zingiber cylindricum*	Endemic-100%
		Zingiber wightianum	*Zingiber wightianum*	Native - 100%
	Curcuma	*Curcuma albiflora*	*Curcuma albiflora*	Endemic-100%
		Curcuma aromatica	*Curcuma zedoaria*	Native – 66%
		Curcuma zedoaria	*Curcuma oligantha*	
		Curcuma oligantha		
	Ammomum	*Amomum nemorale*	*Amomum nemorale*	Endemic- 75%
		Amomum graminifolium	*Amomum graminifolium*	Native - 100%
		Amomum acuminatum	*Amomum acuminatum*	
		Amomum echinocarpum	*Amomum fulviceps*	
		Amomum fulviceps	*Amomum masticatorium*	
		Amomum masticatorium	*Amomum pterocarpum*	
		Amomum pterocarpum		
	Alpinia	*Alpinia fax*	*Alpinia fax*	Endemic - 66%
		Alpinia abundiflora	*Alpinia abundiflora*	Native - 100%
		Alpinia rufescence	*Alpinia nigra*	
		Alpinia nigra	*Alpinia malaccensis*	
		Alpinia malaccensis		
	Elettaria	*Elettaria cardamomum*	*Elettaria cardamomum*	Native 100%
	Hedichyum	*Hedichyum coronarium*	*Hedichyum coronarium*	Native 50%
		Hedichyum flavecens		
	Cyphostigma	*Cyphostigma pulchellum*	*Cyphostigma pulchellum*	Endemic-100%

Table 6.2 Newly generated sequences for the present study for Begoniaceae, Sapotaceae and Zingiberaceae

Species name	*ndh*A intron	*ndh*F -*rpl*32	*Rpl*32 -*trn*L	ITS 5p-8p	*trn*H *psb*A	*trn*C *trn*D	*ndh* F	ITS	*mat* K
Family: Begoniaceae	✓	✓	✓						
B. cordifolia Thwaites	✓	✓	✓						
B. thwaitesii Hook	✓	✓	✓						
B. dipetala Graham	✓	✓	✓						
B. malabarica Lam.	✓	✓	✓						
B. tenera Dryand	✓	✓	✓						
B. albo-coccinea Hook	✓	✓	✓						
B. puspitae Ardi	✓	✓	✓						
B. sublobata Jack	✓	✓	✓						
B. spec.	✓	✓	✓						
B. sect. Reichenheimea	✓	✓	✓						
B. sect. Reichenheimea	✓	✓	✓						
B. sect. Reichenheimea	✓	✓	✓						
B. picta Sm.	✓	✓	✓						
B. tribenensis C.R.Rao	✓	✓	✓						
B. bryophila Ined.	✓	✓	✓						
B. rubella Ham. Ex D.Don	✓	✓	✓						
B. hatacoa Buch.Ham.	✓	✓	✓						
B. panchtharensis S.Rajbh.	✓	✓	✓						
B. flagellaris Hara	✓	✓	✓						
B. diocia Ham. Ex D.Don	✓	✓	✓						
Family: Sapotaceae									
Madhuca elmeri Merr. Ex. H.J.Lam					✓	✓	✓		
Madhuca laurifolia H.J.Lam					✓	✓	✓		
Madhuca motleyana Baehni					✓	✓	✓		

Palaquium rostratum 2 Burck					✓	✓	✓		
Palaquium sericeum H.J.Lam					✓	✓	✓		
Payena maingayi 2 C.B.Clarke					✓	✓	✓		
*Payena lucida 2*A.d.Candolle					✓	✓	✓		
Madhuca sp.					✓	✓	✓		
*Payena lucida 3*A.d.Candolle					✓	✓	✓		
Madhuca longifolia J.F.Macbr.					✓	✓	✓		
Palaquium microphyllum King & Gamble					✓	✓	✓		
Payena obscura Burck					✓	✓	✓		
Madhuca kingiana 2 H.J.Lam					✓	✓	✓		
Payena leerii Kurz					✓	✓	✓		
Palaquium obovatum H.J.Lam					✓	✓	✓		
Palaquium sp.					✓	✓	✓		
Palaquium sp. 2					✓	✓	✓		
Payena sp					✓	✓	✓		
Madhuca sp.					✓	✓	✓		
Palaquium sericeum H.J.Lam					✓	✓	✓		
Palaquium clarkeanum King & Gamble					✓	✓	✓		
Palaquium gutta Baillon					✓	✓	✓		
Palaquium obovatum Engler					✓	✓	✓		
Palaquium oxleyanum Pierre					✓	✓	✓		
Palaquium xanthochymum Pierre ex Burck					✓	✓	✓		
Palaquium formosanum Hayata					✓	✓	✓		
Palaquium amboinense 3 Burck					✓	✓	✓		
Palaquium beccarianum van Royen					✓	✓	✓		
Payena leerii Kurz					✓	✓	✓		

Madhuca kingiana 1 H.J.Lam					✓	✓	✓		
Palaquium sumatranum Burck					✓	✓	✓		
Palaquium rostratum 1 Burck					✓	✓	✓		
Palaquium maingayi King & Gamble					✓	✓	✓		
Madhuca malaccensis H.J.Lam					✓	✓	✓		
Payena obscura Burck					✓	✓	✓		
Madhuca fulva J.F. Macber					✓	✓	✓		
Palaquium galactoxylum H.J.Lam					✓	✓	✓		
Palaquium sp. 1					✓	✓	✓		
Pouteria maclayana Baehni					✓	✓	✓		
Palaquium calophyllum Pierre ex Burck					✓	✓	✓		
Palaquium eriocalyx H.J.Lam					✓	✓	✓		
Palaquium rigidum Pierre ex Dubard					✓	✓	✓		
Palaquium leiocarpum 1 Boeriage					✓	✓	✓		
Palaquium pseudorostratum H.J.Lam					✓	✓	✓		
Palaquium ridleyi King & Gamble					✓	✓	✓		
Palaquium rufolanigerum P.Royen					✓	✓	✓		
Madhuca kuchingensis Yii & P.Chai					✓	✓	✓		
Palaquium walsurifolium Pierre ex Dubard					✓	✓	✓		
Madhuca sarawakensis H.J.Lam					✓	✓	✓		
Madhuca erythrophylla H.J.Lam					✓	✓	✓		
Madhuca kunstleri H.J.Lam					✓	✓	✓		

Species									
Palaquium calophyllum Pierre ex Burck					✓	✓	✓		
Palaquium herveyi King & Gamble					✓	✓	✓		
Palaquium hexandrum Baillon					✓	✓	✓		
Madhuca barbata T.D.Penn.					✓	✓	✓		
Palaquium cryptocariifolium P.Royen					✓	✓	✓		
Madhuca sericea H.J.Lam					✓	✓	✓		
Madhuca oblongifolia Merrill					✓	✓	✓		
Madhuca korthalsii H.J.Lam					✓	✓	✓		
Madhuca proxila Yii & P.Chai					✓	✓	✓		
Madhuca lancifolia H.J.Lam					✓	✓	✓		
Madhuca sp. nov					✓	✓	✓		
Palaquium amboinense 2 Burck					✓	✓	✓		
Madhuca pachyphylla(K.Krause) ined.					✓	✓	✓		
Burckella polymera P.Royen					✓	✓	✓		
Pouteria firma(Miq.) Baehni					✓	✓	✓		
Palaquium quercifolium Burck					✓	✓	✓		
Palaquium_laevifolium_SL					✓	✓	✓		
Isonandra_lanceolata_SL Wight					✓	✓	✓		
Madhuca_sp_SL					✓	✓	✓		
Madhuca_neriifolia_SL H.J.Lam					✓	✓	✓		
Madhuca_longifolia2_SL J.F.macber					✓	✓	✓		
Palaquium_thwaitesii_SL Trim.					✓	✓	✓		
Palaquium_canaliculatum_SL Engl.					✓	✓	✓		
Palaquium_hinmolpedda_SL P.Royen				✓	✓	✓	✓		

*Palaquium_pauciflorum_*SL Engl.				✓	✓	✓	✓		
*Palaquium_rubiginosum_*SL Engl.				✓	✓	✓	✓		
*Isonandra_montana_*SL Gamble				✓	✓	✓	✓		
*Isonandra_sp1_*SL				✓	✓	✓	✓		
*Palaquium_sp3_*SL				✓	✓	✓	✓		
*Palaquium_grande_*SL Engl.				✓	✓	✓	✓		
*Palaquium_petiolare_*SL Engl.				✓	✓	✓	✓		
*Palaquium_sp5_*SL				✓	✓	✓	✓		
*Isonandra_sp2_*SL				✓	✓	✓	✓		
*Isonandra_compta_*SL Dubard				✓	✓	✓	✓		
*Isonandra_zeylanica_*SL Jeuken				✓	✓	✓	✓		
*Manilkara_hexandra_*SL Dubard				✓	✓	✓	✓		
*Mimusops_elengi_*SL L.				✓	✓	✓	✓		
*Madhuca_clavata_*SL Jayas.				✓	✓	✓	✓		
*Palaquium_sp4_*SL				✓	✓	✓	✓		
Isonandra_*sp2_*SL				✓	✓	✓	✓		
Family: Zingiberaceae									
Zingiber_cylindricum Moon								✓	✓
Amomum_pterocarpum Thwaites								✓	✓
Zingiber_wightianum Thwaites								✓	✓
*Amomum_sp*1								✓	✓
Elettaria_sp								✓	✓
Curcuma_zedoaria Roxb.								✓	✓
Amomum_fulviceps Thwaites								✓	✓
*Amomum_sp*2								✓	✓
Curcuma_albiflora Thwaites								✓	✓
Amomum_graminifolium								✓	✓

171

Thwaites									
*Amomum_sp*3								✓	✓
Amomum_nemorale Benth. & Hook.f.								✓	✓
Cyphostigma_pulchellum Benth.								✓	✓
Amomum_acuminatum Thwaites								✓	✓
Amomum_masticatorium Thwaites								✓	✓
*Alpinia_abundiflora*1 Burtt & R.M.Sm.								✓	✓
*Amomum_sp*5								✓	✓
*Amomum_sp*6								✓	✓
*Amomum_sp*7								✓	✓
Alpinia_sp								✓	✓
Alpinia_fax Burtt & R.M.Sm.								✓	✓

The present study supports the importance of LDD over vicariance as the major cause for extant tropical plant disjunctions. The dated phylogenies show that the LDD events are not as infrequent as once thought, and that many tropical lineages have evolved recently. The dated phylogenies and divergence time estimates of the present study suggest recent (Oligocene to Pleistocene) origins and diversification of Sri Lankan families tested; Begoniaceae, Sapotaceae and Zingiberaceae post dating a Gondwanan origin.

The family Begoniaceae shows an origin from an African ancestor and entry to Asia during the Miocene, and a point of entry tentatively reconstructed as India. The timing of entry to India via a dry Arabian corridor is not favored for *Begonia*, thus long distance dispersal is the acceptable hypothesis for the entry to Asia (Thomas et al. 2012). Overland migration explains dispersal from continental Asia to the South East Asian region and the predominant directional trend of dispersal is from west to East. However there is back dispersal from China-Indochina to Himalayan region which initiated during late Miocene with the onset of Asian monsoons at 7.4 Mya (Copeland

1997). Himalayan *Begonia* shows a further diversification event during 5.2 (3.1-7.9) Mya, and further back and forth dispersal could have occurred within the region. The position of Himalayan *B. diocia* in the phylogeny is enigmatic, it has a relatively old age (15.8 Mya) and potentially an area of origin in the region of the Indian Subcontinent. Indian ancestral lineages subsequently dispersed to Sri Lanka and Socotran islands, whose species form the western limits of Asian *Begonia*. Sri Lankan *Begonia* shows a more recent Pleistocene entry followed by more recent in-situ speciation of the endemics *B. tenera* and *B. thwaitesii*.

Within the family Sapotaceae, sub family Chrysophylloideae shows an ancient origin in Africa during the Cretaceous and a single vicariance event between South America and Australia is postulated early in the evolution of the group (Bartish et al. 2011). However Long distance dispersal has played a major role in the assembly of the current hotspots of Chrysophylloideae, between Africa to the Neotropics, between Australia and New Caledonia and between Africa and Madagascar and a single dispersal back to Africa from the Neotropics (Bartish et al. 2011). Within the sub family Sapotoideae, the biogeographic reconstructions of the genus *Sideroxylon* suggest a Northern hemisphere origin for tribe Sideroxyleae in the early Tertiary and subsequent spread to extant distributions via the North Atlantic Land Bridge during the early Eocene (Smedmark & Anderberg 2007). The present study focused on the biogeographic reconstruction of the Sri Lankan Sapotaceae with a high emphasis for tribe Isonandrae, which is well represented in Sri Lanka. My results confirm the origin of Indo Sri Lankan *Manilkara hexandra* as a long distance dispersal from an African ancestor during the Miocene and *Mimusops elengi* from Africa to Indo-Sri Lanka with subsequent spread eastward into Malesia.
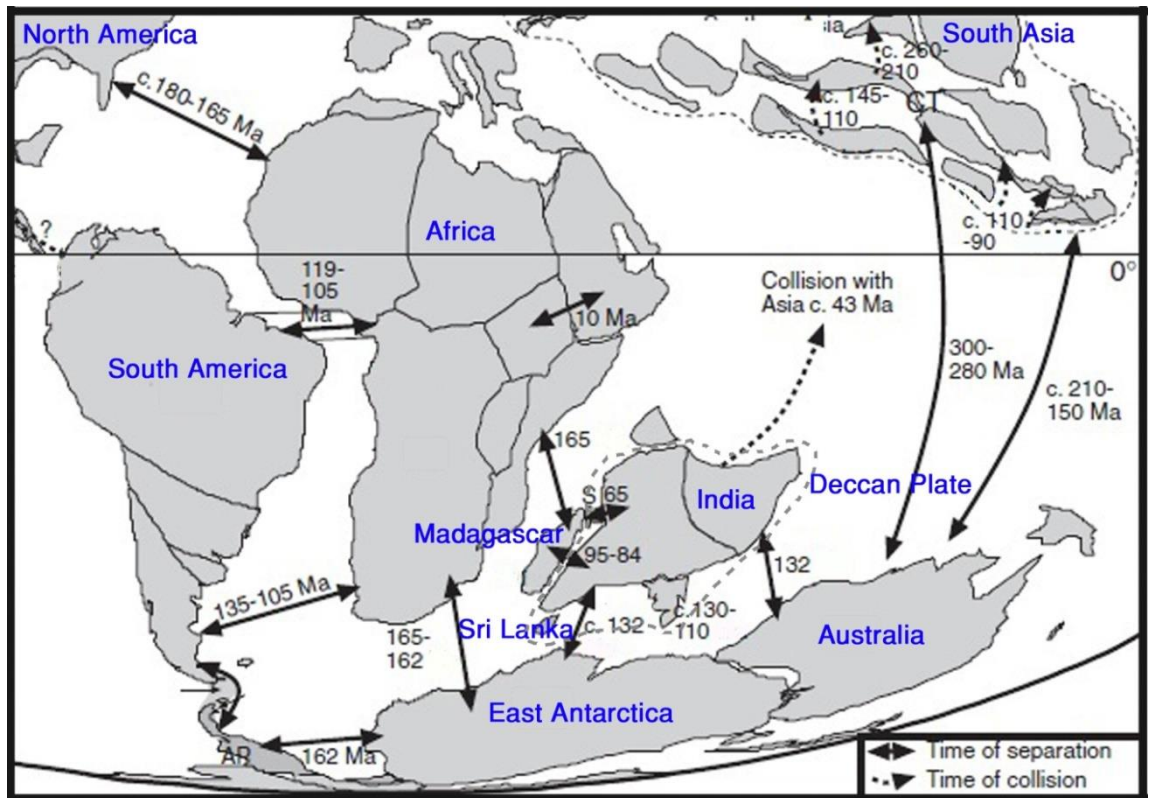
Figure 6.1Different geographic regions in world aand collision times between continental fragments. Excerpted from McLoughlin 2001.
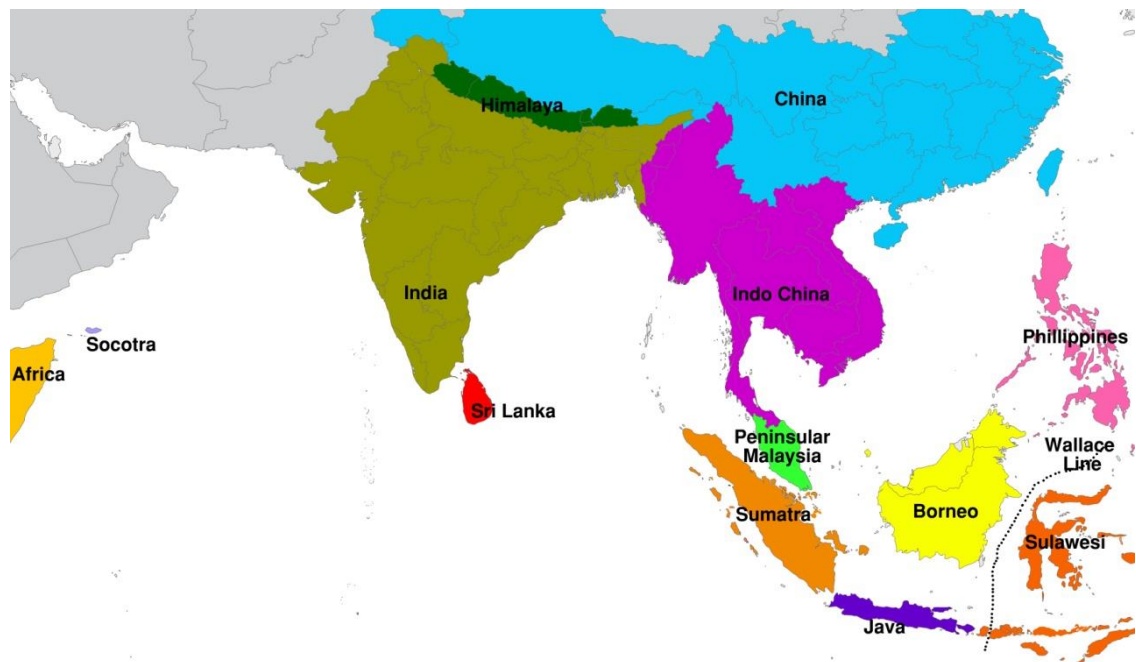
Figure 6.2 Present positions of continental plates and geographic boundaries.

Sri Lankan Isonandrae are of Sunda Shelf in origin, and have resulted from six independent long distance dispersals which occurred during the Eocene-Miocene period. The close proximity between the Malay Peninsula and the Deccan plate during its Northward migration provided many opportunities for floral exchange between two landmasses (Hall 2001). However the molecular age estimates post date this dispersal phase, thus probably have occurred after the Deccan plate collided with Laurasia during 45-50 Mya. The collision with Laurasian coast brought India and South East Asia to similar latitudes and within the same climatic belt, thus there is a sudden appearance of Indian taxa in South East Asian Palaeo floras (Morley 2003; Morley 2000). Further, there are two dispersals out of Sri Lanka evidenced, *Madhuca longifloia* to India and *Madhuca hainanensis* to Eastern Asia during early-Mid Oligocene. More rapid radiation of endemics have occurred as a result of the changing climatic conditions during glacial inter glacial periods during Pliocene-Pleistocene, giving rise to ten endemic species in Sri Lanka during the Pleistocene.

Within Zingiberaceae, *Siphonochilus* (J. M. Wood & Franks) found in Africa while *Tamijia* (S. Sakai & Nagam.) in Borneo. Zingiberaceae have dispersed into equatorial regions in South East Asia as results of global cooling following cretaceous thermal maximum (Kress & Specht 2006). The continental Asia harbors the highest diversity with Alpinoideae having its centre of diversification in the Malesian region while Zingiberoideae has itsin the monsoonal areas of Indochina and Thailand (Larsen 2005). Within Sri Lankan Zingiberaceae, the species in the genera *Curcuma, Zingiber* and *Alpinia* shows an Eastern Asian origin during Oligo-Miocene boundary, thus provide an evidence for overland migration into Sri Lanka via India. Sri Lankan *Amomum* has an origin in Sunda shelf by three independent dispersals during mid-miocene-pliocene and *Amomum* endemics are more recent Pleistocene speciation within the island.

Sri Lanka is constructed as the ancestral area for African/ Madagascan *Aframomum* andAfrican *Renealmia,* which shows an entry to Africa during Oligocene. The collision of the Afro-Arabian plate with Asia approximately 25 Mya (Kulju et al. 2007; Yuan et al. 2005) might have facilitated overland migration of taxa from Asia to Africa from a Sri Lankan ancestor; however long distance over water dispersal from Sri Lanka directly to Africa is another possibility which could have occurred at any point of time.

Neotropical *Renealmia* have originated from African ancestor via oceanic long distance dispersal during Miocene-Pleistocene and further speciation more recently during Pleistocene.

Table 6.3A summary for reconstructed geographic origins for the tested families, and the number of dispersals during different geological epochs.

Family	Origin	Era	No of dispersals	Age (Mya)
Begoniaceae	India	Late Miocene	1	7.68 (3.91-12.62)
Sapotaceae	Africa	Early Miocene	1	19.04 (11.3-25.52)
		Late Miocene	1	6.81 (3.47-11.12)
	Sunda Shelf	Eocene	1	36.91 (27.76-47.25)
		Oligocene	4	31.93 (22.88-41.43)
				28.49
				33.14 (25-40.02)
				31.53
		Mid Miocene	1	11.49
Zingiberaceae	Eastern Asia	Oligocene	2	25.7 (15.63-36.72)
				27.31
		Early Miocene	1	18.74 (10.24-29.2)
		Pliocene	1	3.76 (1.34-7.13)
	Sunda Shelf	Mid Miocene	1	11.8 (6.3-19.4)
		Late Miocene	1	10.2
		Pliocene	1	3.8 (1-7.4)

The dates of origin for Sri Lankan lineages considerably post-date the Gondwanan break up, instead suggesting the importance of more recent transoceanic long distance dispersal for their origin and diversification within the island. Among tested plant families, dispersals have occurred from Africa, India, Eastern Asia and Sunda Shelf into Sri Lanka.

Much of the Sri Lankan flora shows an origin in the Sunda Shelf (53%), especially genera *Madhuca, Palaquium*, *Isonandra* in the family Sapotaceae and the genus *Amomum* in the family Zingiberaceae. During the Deccan plate's northward migration, it moved very close to Peninsular Malaysia and had a glancing contact with Sumatra

during the late Paleocene. This could have facilitated plant dispersals between the Deccan plate and South East Asia. However Sri Lankan lineages post date this dispersal phase, as all have taken place after the Deccan plate collided with the Laurasian coast, which brought them to same altitudes and climatic belt.
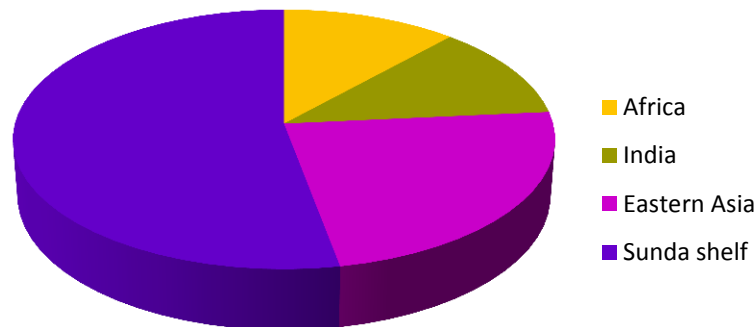
Figure 6.3 Reconstructed geographic origins for Sri Lankan flora and their relative contributions.

Sri Lankan flora has also been influenced by the Eastern Asian flora (20%), especially with regard to reconstructing as the ancestral area for some genera in family Zingiberaceae. After the Deccan plate collided with the Laurasian coast during the Eocene there would have been many opportunities for mixing of the Laurasian flora with the Deccan Gondwanan flora. Himalayan uplift resulted in numerous climatic and geological changes thus had a profound effect in the flora of that region.

Africa contributes 11% of the lineages observed to the Sri Lankan flora. My results post date the late Cretaceous dispersal phase, which is accepted as the major phase for Afro-Indo Sri Lankan disjunctions evidenced in some families like Crypterionaceae (Conti et al. 2002; Ashton & Gunatilleke 1987) and Dipterocarpaceae. Instead they are much younger indicating more recent trans-oceanic long distance dispersal events from Africa to the Indo Sri Lankan region.

India accounts for only 12% of the Sri Lankan families tested, despite the long proximities between the two land masses since Jurrasic times where they are separated by the narrow Palk Strait 32 km in length. The intervening continental shelf between Sri Lanka and India provided an intermittent land connection until the last sea level rise

6000 years ago during the Holocene (McLoughlin 2001; Ashton & Gunatilleke 1987) and has potential to act as a migratory path for floristic exchange between the two land masses. However as the endemic Sri Lanka *Begonia* are restricted to altitudes of 1000-1200m, the overland migration via a lowland land bridge seems unlikely. The Western Ghats of India and the sub montane forests in Sri Lanka could have act as archipelago-like systems, facilitating plant dispersals among them.

Of the four regions, the Sunda Shelf is overwhelmingly the main source for Sri Lankan floral elements, despite the geographic proximity of India. It would seen that climatic similarity between the perhumid forests of Sri Lanka and those of the Sunda Shelf has a lot of influence on successful immigration.
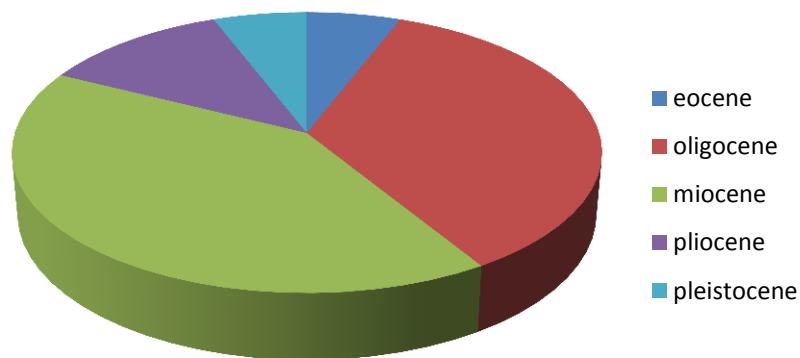
Figure 6.4 Percentage of dispersals in terms of geological epochs.

Among the families tested, dispersals have occurred stochastically, one during the Eocene, six during the Oligocene, seven during the Miocene, two during the Pliocene and one during the Pleistocene (Figure 6.4). The highest number of dispersals occurred during the Miocene when a warm climate was prevailing during the Miocene thermal maximum (Zachos et al. 2001).

The Deccan plate, composed of India and Sri Lanka, has probably undergone more changes than any other tropical region since it began to drift away from Antarctica during Cretaceous (Morley 2000). During its northward journey, as it entered the southern hemisphere high pressure zone, its Gondwanan flora was replaced by a tropical African flora. During the Paleocene, moist equatorial climate resulted in the evolution

of new lineages, with a rapidly diversifying mega thermal flora being dominant. The plate was at the equator during the early Eocene and extremely wet climates could have provided optimum climatic conditions for angiosperm proliferation, thus fully covered by multistoried rain forests. Warm, perhumid conditions during Late Eocene to Oligocene further nourished rain forest taxa. It is accepted that humid, tropical climate persisted in the Deccan plate since it collided with Eurasia, which is comparable to climate in India and Sri Lanka at present (Ashton & Gunatilleke 1987; Morley 2000). The massive volcanism at the Cretaceous-Tertiary boundary during 65 Ma and extensive aridification during early Tertiary due to the uplift of Himalaya resulted in further impoverishment of allochthonous African elements from the Indian flora (Rutschmann & Eriksson 2004; Morley 2003; Conti et al. 2002; McLoughlin 2001), leaving the Western Ghats and Sri Lanka as refugial areas for those flora.

Further, my results confirm that *in situ* speciation is an important contributor to the Sri Lankan flora. More rapid radiation of endemics has occurred during Pliocene-Pleistocene; two endemics in Begoniaceae, ten endemics in Sapotaceae and ten endemics in Zingiberaceae have evolved *in situ* during this period. Sri Lanka will have been subjected to expansion and contraction of climatic and vegetation zones within the island during glacial and interglacial periods, potentially resulting in allopatric speciation. The results provide strong evidence for youthful tropics which is congruent with the view of Koenen et al. (2015) on the evolution of rain forest species in the Neotropics. Rain forests have been acted as museums for higher level taxa, conserving diverse traits within them ultimately giving rise to new species (Pennington et al. 2015). A similar scenario of recent speciation also evident in Africa (Harris et al. 2000) and Asia (Thomas et al. 2012). Among birds and mammals, a recent Pleistocene speciation has been evidenced at higher altitudes; however it is throughout Pre-Pleistocene to Pleistocene towards the equator (Weir et al. 2007).
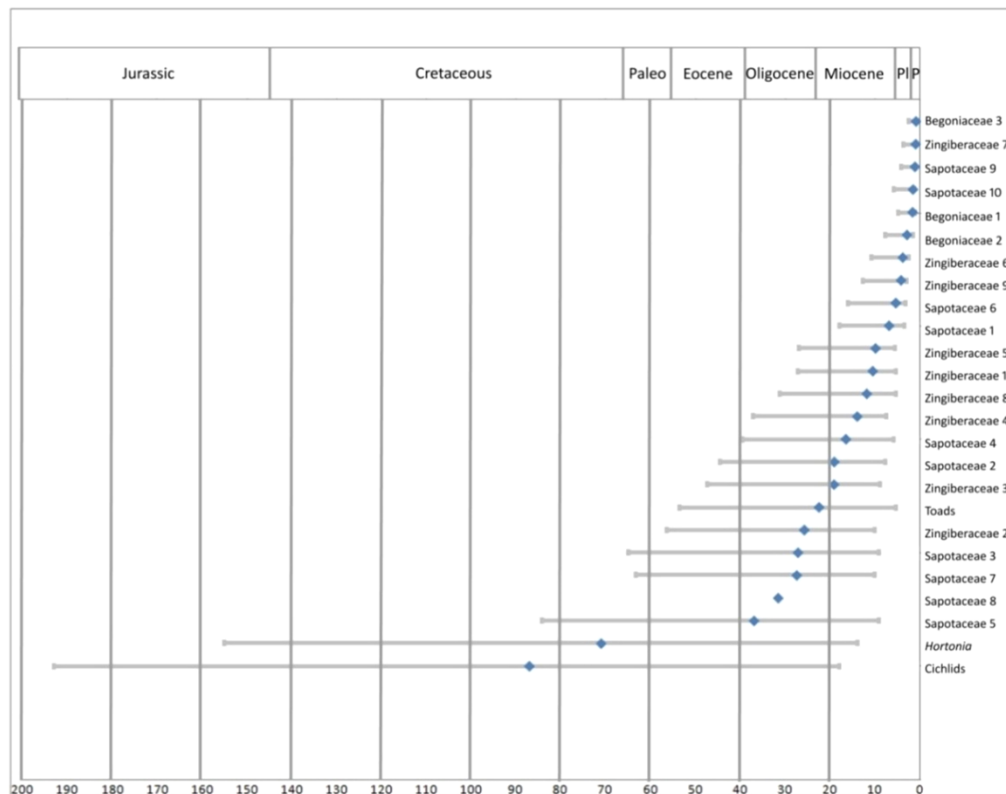
Figure 6.5 A plot of the minimum ages of Sri Lankan clades (25) from the present study and other published studies on plants (*Hortonia*) and animal groups (toads) arranged from oldest (left) to youngest. The error bars indicate the uncertainty around the minimum age estimates.

P., Pleistocene; Pl.,Pliocene; Paleo., Paleocene

The vicariance paradigm has long been accepted as the major cause for tropical disjunctions and assembly of flora land masses like Sri Lanka which was a part of Gondwana a long time ago. One striking example is the Sri Lankan endemic monotypic genus *Hortonia* in Monimiaceae, being the oldest lineage among Sri Lankan flora. The other genera in the family occur in America, Africa, Madagascar and The Mascarenes, New Caledonia, Australia, New Zealand and the Malesian region (Renner et al. 2010). The dated phylogeny of Monimiaceae suggests a Gondwanan origin for Sri Lankan *Hortonia* which dates back to 71 (57-84) Mya during the late Cretaceous. One possibility of this ancient lineage of *Hortonia* in Sri Lanka is explained as rafting on the Deccan plate to Asia (Ashton & Gunatilleke 1987). The other explanation is long distance dispersal of *Hortonia* from Antarctica to Sri Lanka, however the huge distance

between two land masses ca 2100 Km (Aitchison et al. 2008), raises the question of ability to cross such long distance over water according to Renner et al. (2010).

A classic Gondwanan distribution is observed among some animals, such as in the fresh water Cichlids (order Perciformes: family Cichlidae). They are mainly distributed Africa, South and Central America, Madagascar, and Indo/Sri Lanka. If they have originated in Africa and migrated into South America, Madagascar, and India via saltwater dispersal, the African clade should have been constructed as the sister clade for the rest. However, more recent dated phylogeny by Azuma et al. (2008) have found dates for major phylogenetic splits which are congruent with Gondwanan breakup of landmasses. The divergence time between Malagascan and Indo/Sri Lankan taxa is 87 Mya (69–106 MYA) which is congruent with time of separation between Madagascar and India during 90 Mya (Briggs et al. 2003; Conti et al. 2002; McLoughlin 2001). Further it supports the evidence to the time of separation between African and South American landmasses which probably occurred 100 Mya with a phylogenetic split occurring between African and neotropical clades around 89 Mya (72-108 Mya) (Azuma et al. 2008).

However, the present study largely contributes to the growing body of evidence in support of recent transoceanic dispersals as being a major factor in the tropical plant disjunctions seen today. More importantly it reveals the disproportionate importance of LDD in the development of the flora of small islands like Sri Lanka. LDD can happen by means of vectors such as wind, water, birds, mammals and extreme events like tropical cyclones, hurricanes and tornadoes (Nathan 2006), more importantly these climatic events have much influence on land clearing territory for new immigrants. In most tropical tree diaspores transport via ocean currents is much more feasible than dispersal by wind and Houle (1998) demonstrated that the trans-oceanic migration times are different in different water bodies which depend on paleocurrents and direction of wind (Table 6.2).

Trans-oceanic dispersal by birds is also possible, however the capability for dispersal over very long distances may probably be infrequent since they void gut contents frequently and their migratory route is dominantly north-south (Fukui 2003). Further,

for tiny seeds it is possible to get trapped on a bird's foot, however for tropical trees which produce large seeds, the scenario seems unlikely.

Table 6.4 Hypothesized rafting times across the Atlantic Ocean, Caribbean Sea and Southeast Indian Ocean during different periods throughout the Tertiary according to Houle (1998).

Ocean System	Rafting time 50 Mya	Rafting time 40 Mya	Rafting time 30 Mya
Atlantic Ocean	5.2-7.7 days	7.3-10.8 days	10-14.7 days
Caribbean Sea	11.2-18.2 days	10.2-16.6 days	9.3-15.1 days
SE Indian Ocean	24.2-25.6 days	18.4-19.5 days	11.5-12.2 days

Although such dispersals are considered as rare, long jumps available through LDD have much influence on ranges of taxa than numerous small steps available through local dispersals or range expansions after climate shifts.

Numerous recent dated phylogenies evoke the importance of relatively recent transoceanic dispersals on shaping the world's tropical flora such as: Dispersals between Asia and Africa: *Macaranga* and *Mallotus* (Euphorbiaceae) (Kulju et al. 2008), *Bridelia* (Phyllathaceae) (Li et al. 2009), *Gaertnera* (Rubiaceae) (Malcomber 2002)*Exacum* (Gentianaceae) (Yuan et al. 2005), Malphigiaceae (Davis 2002), Africa and Neotropics: Melastomataceae (Renner et al. 2001), *Annona* (Annonaceae) (Richardson et al. 2004), *Commiphora* (Burseraceae) (Weeks et al. 2007) Across pacific ocean Myrtaceae (Sytsma et al. 2004), *Piper* and *Peperomia*, *Hernandia* (Hernandeaceae) (Michalak et al. 2010). Some taxa like *Piper* and *Peperomia* (Piperaceae) have shown multiple dispersals between South America, Africa, Asia and the Pacific (Smith et al. 2008). Within Cucurbitaceae, forty-three successful long-distance dispersal events from Asia to Africa, North and South America and Australia have been reported over the last 60 Ma, an average of seven long-distance dispersals every 10 Ma (Schaefer et al. 2008). Thus, LDD's are possibly not so infrequent as once thought before and my results add to the growing body of importance of LDD over vicariance. Further, my study provides strong evidence for a directional trend of dispersals from west to east, from the climatically similar Sunda Shelf. Future work

should concentrate on putting further Sri Lankan elements into dated phylogenies, and also focus on more intense sampling covering the South and eastern Asian flora to provide better geographic coverage.

As a conclusion, long distance dispersals have played a prominent role in the evolution of the Sri Lankan flora. The young ages of Sri Lankan clades and also clades from other regions challenge the vicariant paradigm for their origin and provide further strong evidence for youthful tropics at the species level.

References

Aitchison, J.C., Ali, J.R. & Davis, A.M., 2008. When and where did India and Asia collide? *Geological Bulletin of China*, 27(9), pp.1351–1370.

Ashton, P.S. & Gunatilleke, C.V.S., 1987. New light on the plant geography of Ceylon I . Historical plant geography. *Journal of Biogeography*, 14(3), pp.249–285.

Azuma, Y. et al., 2008. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC evolutionary biology*, 8(1), p.215.

Bartish, I. V. et al., 2011. Vicariance or long-distance dispersal: Historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, 38(1), pp.177–190.

Briggs, J.C., Collections, Z. & Museum, G., 2003. The biogeographic and tectonic history of India.*Journal of Biogeography*, 30, pp.381–388.

Conti, E. et al., 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution; international journal of organic evolution*, 56(10), pp.1931–1942.

Copeland, P., 1997. The when and where of the growth of the Himalaya and the Tibetan Plateau. In: Ruddiman, W.F. (ed) Tectonic Uplift and Climate Change. New York: Plenum Press.

Davis, C.C. et al., 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences, USA*, 99, pp.6833–6837.

Fukui, A., 2003. Relationship between seed retention time in bird's gut and fruit characteristics. *Ornithological Science,* 2, pp.41-48.

Hall, R., 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: Changing patterns of land and sea. *Faunal and Floral Migrations and Evolution in Se Asia-Australasia*, (Figure 1), pp.35–56.

Harris, D. J. et al., 2000. Rapid radiation in Aframomum (Zin- giberaceae): evidence from nuclear ribosomal DNA internal transcribed spacer (ITS) sequences. *Edinburgh Journal of Botany*, 57, pp.377–395.

Houle, A., 1998. Floating islands: a mode of long distance dispersal for small and medium-sized terrestrial vertebrates. *Divers. Distrib*, 4, pp.201–216.

Koenen, E.J.M. et al., 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol*, 207, pp.327–339.

Kress, J.W. & Specht, C.D., 2006. The evolutionary and biogeographic origin and diversification of the tropical monocot Oder Zingiberales. *Aliso*, 22, pp.621–632.

Kulju, K.K.M. et al., 2007. Molecular phylogeny of Macaranga, Mallotus, and related genera (Euphorbiaceae s.s.): Insights from plastid and nuclear DNA sequence data. *American Journal of Botany*, 94(10), pp.1726–1743.

Larsen, K., 2005. Distribution patterns and diversity centres of Zingiberaceae in SE Asia. *Biologiske Skrifter*, 55, pp.219–228, 2005.

Li, Y. et al., 2009. More Miocene Dispersal Between Africa and Asia—the Case of *Bridelia* (Phyllanthaceae). *Systematic Botany*, 34(3), pp.521–529.

Malcomber, S.T., 2002. Phylogeny of Gaertnera Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution; international journal of organic evolution*, 56(1), pp.42–57.

McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49(3), pp.271–300.

Michalak, I.,2010. Trans-Atlantic, trans-Pacific and trans- Indian Ocean dispersal in the small Gondwanan Laurales family Hernandiaceae. *Journal of Biogeography,* 37(7),pp.1214-1226.

Morley, R. J., 2000. *Origin and evolution of tropical rain forests*. Chichester: John Wiley & Sons, Ltd.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(12), pp.5–20.

Nathan, R., 2006. Long-distance dispersal of plants. *Science (New York, N.Y.)*, 313(August), pp.786–788.

Pennington, R.T., Hughes, M. & Moonlight, P.W., 2015. The Origins of Tropical Rainforest Hyperdiversity. *Trends in Plant Science*, 20(11), pp.693–695.

Renner, S. S., and K. Meyer. 2001. Melastomeae come full circle: biogeographic reconstruction and molecular clock dating. *Evolution,* 55, pp.1315–1324.

Renner, S.S. et al., 2010. Biogeography of the Monimiaceae (Laurales): A role for East Gondwana and long-distance dispersal, but not West Gondwana. *Journal of Biogeography*, 37(7), pp.1227–1238.

Richardson, J.E. et al., 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359(1450), pp.1495–1508.

Rutschmann, F. & Eriksson, T., 2004. Did Crypteroniaceae really disperse out of India? Molecular dating evidence from rbcL, ndhF, and rpl16 intron sequences. *International Journal of plant Science*, 165(S4), pp.S69–S89.

Schaefer, H. et al., 2008. a dated phylogeny reveals an Asian origin of the gourd family (Curcubitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society Botany, Biological Sciences.* 276, pp.843-851.

Smedmark, J.E.E. & Anderberg, A. a., 2007. Boreotropical migration explains hybridization between geographically distant lineages in the pantropical clade Sideroxyleae (Sapotaceae). *American Journal of Botany*, 94(9), pp.1491–1505.

Smith, J.F. et al., 2008. Placing the origin of two species-rich genera in the late Cretaceous with later species divergence in the Tertiary: a phylogenetic, biogeographic and molecular dating analysis of Piper and Peperomia (Piperaceae). *Plant Systematics and Evolution*, 275, pp.9–30.

Sytsma, K.J. et al., 2004. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere. *International Journal of Plant Sciences*, 165(4 Suppl.):S85–S105.

Thomas, D.C. et al., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, 39(1), pp.98–113.

Weeks, A., Daly, D.C. & Simpson, B.B., 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution*, 35, pp.85–101.

Weir, J.T., Schluter, D. & Schl, D., 2007. The Latitudinal Gradient in Recent Rates of and Extinction Speciation Birds and Mammals. *Science*, 315(5818), pp.1574–1576.

Yuan, Y.-M. et al., 2005. Phylogeny and biogeography of *exacum* (gentianaceae): a disjunctive distribution in the Indian ocean basin resulted from long distance dispersal and extensive radiation. *Systematic biology*, 54(1), pp.21–34.

Zachos, J. et al., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (New York, N.Y.)*, 292(April), pp.686–693.

Appendices

Appendix1:Accessions used for the generation of DNA sequence data, Begoniaceae and EDNA numbers. Species indicated in bold are newly generated sequences for the present study.

Species name	Origin	Voucher
B. ningmingensis D.Fang, Y.G. Wei & C.I.Peng	Philippine	Peng **P20322** (HAST)
B. acuminatissima Merr.	Philippine	Rubite R321 (PNH)
B. anisoptera Merr.	Philippine	Rubite R479 (PNH)
B. biliranensis Merr.	Philippine	Rubite R311 (PNH)
B. blancii M.Hughes & C.I. Peng	Philippine	Peng P22545 (HAST)
B. culasiensis	Philippine	Rubite R234 (PNH)
B. calcicola Merr.	Philippine	Peng P20761 (HAST)
B. caminguinensis Elmer	Philippine	Rubite R506 (PNH)
B. castilloi Merr.	Philippine	Rubite R98 (PNH)
B. chingipengii Rubite	Philippine	Peng P23368 (HAST)
B. chloroneura P.Wilkie & Sands	Cultivated: RBGE (acc. num.: 19972555), Philippines, Luzon Island	Forrest, L.L. 128 (E)
*B. cleopatrae*_1 Coyle	Palawan	Wilkie et al., 25373 (E)
B. copelandii Merr.	Luzon	Adduru, M 259 (K)
B. dinglensis	Philippine	Peng P23859 (HAST)
B. elmeri Merr.	Luzon	Rubite R319 (PNH)
B. elnidoensis ined.	Philippine	Peng P25308 (HAST)
B. fenecis Merr.	Philippine	Peng P10794 (HAST)
B. gabaldonensis ined.	Philippine	Peng P23356 (HAST)
B. gitingensis Elmer	Philippine	Rubite R255 (PNH)
B. gueritziana Gibbs	Philippine	Peng P22311 (HAST)

B. gutierrezii Coyle	Philippine	Blanc s.n. (E)
*B. hernandioides*_1 Merr.	Cultivated: BGB (acc. num.: 00603589), Philippines	Forrest, L.L. 129 (E)
*B. hernandioides*_2 Merr.	Philippine	Peng P21006 (HAST)
*B. hughesii*_1 ined.	Philippine	Peng P23466 (HAST)
*B. hughesii*_2 ined.	Philippine	Peng P23475 (HAST)
B. klemmei Merr.	Philippine	Rubite R182 (PNH)
*B. longiscapa*_1 Warb.	Philippine	Rubite R298 (PNH)
*B. longiscapa*_2 Warb.	Philippine	Rubite R309 (PNH)
*B. luzonensis*_1 Warb.	Luzon	R316 (PNH)
*B. luzonensis*_2 Warb.	Luzon	R420 (PNH)
*B. luzonensis*_3 Warb.	Luzon	K030960 (HAST)
*B. manillensis*_1 A.DC.	Luzon	Rubite R256 (PNH)
B. merrilliana ined.	Philippine	Peng P23765 (HAST)
B. mindorensis Merr.	Palawan	Rubite R326 (PNH)
*B. nigritarum*_1 (Kamel) Steud.	Palawan	Rubite R419 (PNH)
*B. nigritarum*_2 (Kamel) Steud.	Palawan	Rubite R406 (PNH)
*B. nigritarum*_3 (Kamel) Steud.	Palawan	Peng P23855 (HAST)
B. obtusifolia Merr.	Panay	Peng P23828 (HAST)
B. oxysperma A.DC.	Luzon	Rubite R213 (PNH)
B. rhombicarpa A.DC.	Luzon	Cuming H. 510 (E)
B. rubitae M.Hughes	Philippines	Rubite R356 (PNH)
B. rufipila Merr.	Luzon	Rubite R256 (PNH)
*Begonia sp*_1	Luzon	Peng P23566 (HAST)
*Begonia sp*_2	Luzon	Kokubagata GK71 (HAST)

B. subnummarifolia Merr.	Sabah	No voucher
B. suborbiculata Merr.	Palawan	Rubite R353 (PNH)
B. sykakiengii ined.	Philippine	Peng P23856 (HAST)
B. tagbanua ined.	Philippine	Blanc s.n. (E)
B. taraw ined.	Philippine	Blanc s.n. taraw2 (E)
B. tayabensis Merr.	Luzon	Rubite R360 (PNH)
B. trichocheila Warb.	Luzon	Peng P20764 (HAST)
B. wadei Merr. & Quisumb.	Palawan	Rubite R699 (PNH)
B. woodii Merr.	Palawan	Peng P23479 (HAST)
B. dregei Otto & Dietr.	Cultivated: RBGE (acc. num.: 20000902), South Africa	McLellan, T. 415 (E)
B. goudotii A.DC.	Madagascar	Plana, V. 120 (E)
B. oxyloba Welw. ex. Hook.f.	Cultivated: RBGE (acc. num.: 19982761), Tanzania	Thomas, D.C. 08-141 (E)
B. poculifera Hook.f.	Cultivated: RBGE (acc. num.: 19923143), Cameroon	Forrest, L.L. 234 (E)
B. polygonoides Hook.f.	Ivory Coast	Van der Burg, W. J. 244 (WAG)
B. samahensis M.Hughes & A.G. Mill.	Cultivated: RBGE (acc. num.: 19990412), Yemen, Socotra	Thomas, D.C. 09-01 (E)
B. socotrana Hook.f.	Yemen, Socotra	Miller, A.G. 19210/10 (E)
B. sutherlandii Hook.f.	Cultivated: RBGE (acc, num.: 20010167), South Africa	Thomas, D.C. 08-140 (E)
B. boliviensis A.DC.	Cultivated: GBG (acc.num.: 00801998) Bolivia	No voucher available
B. nelumbifolia Cham. & Schlecht	Cultivated: RBGE (acc. num.: 19791888), Mexico	Hunt, D.R. 7516 (K)

B. radicans Vell.	Cultivated: GBG (acc. num.: 00908995), Brazil	No voucher available
B. tenera **Dryand**	**Sri Lanka**	**Kumarage 68 (E)**
B. thwaitesii **Hook**	**Sri Lanka**	**Kumarage 23 (E)**
B. cordifolia **Thwaites**	**Sri Lanka, India**	**Kumarage 14 (E)**
B. dipetala **Graham**	**Sri Lanka, India**	**Kumarage 25 (E)**
B. malabarica **Lam.**	**Sri Lanka, India**	**Kumarage 28 (E)**
*B. malabarica*_India Lam.	Cultivated: GBG (acc. num.: 00201896), India, Sri Lanka	Forrest, L.L. 288 (E)
B. albo-coccinea **Hook**	**India**	**Photo voucher available**
B. floccifera Bedd	Cultivated: GBG (acc. num.: 03009989), India	Forrest, L.L. 238 (E)
B. bryophila **Ined.**	**Nepal**	**6100**
B. diocia **Ham. Ex D.Don**	**Nepal**	**13651**
B. flagellaris **Hara**	**Nepal**	**6010**
B. hatacoa **Buch Ham.**	**Nepal**	**5971**
B. panchtharensis **S.Rajbh.**	**Nepal**	**5968**
B. picta **Sm.**	**Nepal**	**5993**
B. roxburghii A.DC.	Cultivated: GBG (acc. num.: 01100797), India	Thomas, D.C. 08-103 (E)
B. rubella **Ham. Ex D.Don**	**Nepal**	**6000**
B. sikkimensis **A.DC.**	**Cultivated: RBGE (acc, num.: 20051755), India**	**Thomas, D.C. 08-144 (E)**
B. tribenensis **C.R.Rao.**	**Nepal**	**6043**
B. aceroides Irmsch.	Thailand	Phutthai 243 (E)
B. acetosella Craib	Cultivated: GBG (acc, num.: 00107396), Vietnam	Thomas, D.C. & Ardi, W.H. 08-105 (E)
B. aff. elisabethae Kiew	Cultivated: RBGE (acc. Num.: 20020477), China	Moller, M. 01-156B (E)

B. alicida C.B.Clarke	Burma, Thailand	Phutthai 139 (E)
B. brandisiana Kurz	Burma, Laos	Brandis, D. 1327 (K)
B. demissa Craib	China	Phutthai 221 (E)
B. elisabethae Kiew	Penisular Malesia	Phutthai 239 (E)
B. grandis Dryand.	Cultivated: RBGE (acc. num.: 19521036), China	Thomas, D.C. 08-145 (E)
B. hymenophylla Gagnep.	Cambodia	Phutthai 232 (E)
B. masoniana Irmsch. Ex Ziesenh.	Cultivated: RBGE (acc. num.: 19980075), China	Thomas, D.C. 07-24 (E)
B. morsei Irmsch.	Cultivated: GBG (acc. num.: 19980076), China	Unknown s.n. (E)
B. obovoidea Craib.	Thailand	Phutthai 244 (E)
B. palmata D.Don	Cultivated: RBGE (acc. num.: 20020476), China	Moller, M. 01-127 (E)
B. silletensis A.DC.	Cultivated: GBG (acc, num.: 00115295), China	Thomas, D.C. 08-104 (E)
B. sizemoreae Kiew	Cultivated: GBG (acc, num.: 00101400), Vietnam	Thomas, D.C. 08-111 (E)
B. smithiae Geddes	Thailand	Chamchamroon 3662 (E)
B. spec	**Cambodgia**	**20100763 EDNA12-0025039**
*B. spec_*China1	China	Thomas 08-145 (E)
*B. spec_*China2	China	Forrest 31 (E)
*B. spec_*Thailand1	Thailand	Phutthai 195 (E)
*B. spec_*Thailand2	Thailand	Suddee 3375 (E)
*B. spec_*Thailand3	Thailand	Suddee 3371 (E)
B. versicolor Irmsch.	Cultivated: RBGE (acc, num.: 19980037), China	Forrest, L.L. 2 (E)
B. aff. congesta Ridl.	Cultivated: SBG, Malaysia, Borneo	Thomas, D.C. 09-05 (E)
B. areolata Miq.	Cultivated: BoBG,	Thomas, D.C. & Ardi, W.H. 09-

	Indonesia, Java	137 (E)
B. bracteata Jack	Indonesia, Java	Ardi, W.H. & Thomas, D.C. 25 (E)
B. chlorosticta Sands	Cultivated: SBG, Malaysia, Borneo	Thomas, D.C. 09-04 (E)
B. corrugata Kiew & S. Julia	Cultivated: SBG, Malaysia, Borneo	Thomas, D.C. 09-02 (E)
B. decora Stapf	Cultivated: RBGE (acc, num.: 20021608), Malaysia, Peninsul Malaysia	Neale, S. 8C (E)
B. fuscisetosa Sands	Peninsular Malaysia	20120800 (E)
B. goegoensis N.E.Br.	Cultivated: GBG (acc. num.: 01112557), Indonesia, Sumatra	Thomas, D.C. & Ardi, W.H. 08-107 (E)
B. harauensis Girm	Peninsular Malaysia	Thomas 09-134 (E)
B. kingiana Irmsch	Cultivated: GBG (acc. num.: 01807007), Malaysia, Peninsula Malaysia	Thomas, D.C. 08-102 (E)
B. laruei M.Hughes	Indonesia, Sumatra	Hughes, M. 1389 (E)
B. longifolia Blume	Cultivated: GBG (acc. num.: 20021848), Malaysia, Peninsula Malaysia	Neale, S. 11C (E)
B. multangula Blume	Indonesia, Bali	Thomas, D.C. & Ardi, W. H. 08-90 (E)
B. multijugata M. Hughes	Indonesia, Sumatra	Wilkie, P., Hughe, M., Sumadijaya, A., Rasnovi, S., Marlan & Suhardi 768 (E)
B. muricata Blume	Indonesia, Java	Ardi, W.H. & Thomas, D.C. 27 (E)
B. pavonina Ridl	Cultivated: RBGE (acc. num.: 20021611), Malaysia, Peninsula	Neale, S. 9C (E)

	Malaysia	
B. puspitae Ardi	Cultivated: RBGE Indonesia	20111539
B. resecta Miq. ex Koord.	Indonesia	Hughes 786
B. robusta Blume	Cultivated: BaBG, Indonesia, Java	Thomas, D.C. & Ardi, W.H. 08-133 (E)
*B. spec_*Borneo1	Cultivated: RBGE (acc. nom.: 20030131), Malaysia, Borneo	Thomas, D. C. 07-1 (E)
*B. spec_*Borneo2	Cultivated: BoBG, Indonesia, Borneo	Thomas, D. C. & Ardi, W. H. 09-136 (E)
*B. spec_*Sumatra	Cultivated: BoBG, Indonesia, Sumatra	Thomas, D. C. & Ardi, W. H. 08-132 (E)
*B. spec_*Sumbawa1	Cultivated: BoBG, Indonesia, Sumbawa	Thomas, D. C. & Ardi, W. H. 09-138 (E)
***B. spec sect Reichenheimea*1**	**Cultivated: RBGE, Indonesia, Java**	**20111543**
***B. spec sect Reichenheimea*3**	**Cultivated: RBGE, Indonesia, Java**	**20111545**
***B. spec sect Reichenheimea*4**	**Cultivated: RBGE, Indonesia, Java**	**20112191**
***B. sublobata* Jack**	**Cultivated: RBGE, Indonesia, Java**	**20101649**
B. sudjanae Jansson	Cultivated: GBG (acc, num.: 02605499), Indonesia, Sumatra	Thomas, D.C. & Ardi, W.H. 08-109 (E)
B. tenuifolia Dryand.	Indonesia, Bali	Thomas, D.C & Ardi, W.H. 08-86 (E)
B. venusta King	Cultivated: RBGE (acc, num.: 20021596), Malaysia, Peninsula Malaysia	Neale, Sophie 7 (E)
B. verecunda M.Hughes	Indonesia, Sumatra	Wilkie, P., Hughes, M., Sumadijaya, A., Rasnovi, S.,

		Marlan & Suhardi 618 (E)
B. aff. mekonggensis Gim & Wiriad	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-108 (E)
B. aff. multangula Blume	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 09-85 (E)
B. aptera Blume	Indonesia, Sulawesi	Smith, P. & Galloway, L. 67(E)
B. argenteomarginata Tebbitt	Cultivated: GBG (acc. Num.: 00803887), Papua New Guinea	Forrest, L.L. 145 (E)
B. bonthainensis Hemsl.	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 09-63 (E)
B. brevirimosa Irmsch	Cultivated: RBGE (acc. num.: 19821108, Papua New Guinea	Forrest, L.L. 137 (E)
B. capituliformis Irmsch	Indonesia, Sulawesi	Kinho, J. & Poulsen, A. 169 (E)
*B. chiasmogyna*M.Hughes	Cultivated: RBGE (acc. num.: 20021895) Indonesia, Sulawesi	Thomas, D.C. 07-29 (E)
*B. cleopatrae_*2 Coyle	Philippines, Palawan	Wilkie, P., Mendum, M., Argent, G.C.G., Cronk, Q., Middleton. D.J., Fuentes, R. & Chavez, R.V. 25373 (E)
B. comestibilis D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-62 (E)
B. didyma D.C. Thomas & Ardi	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 08-77 (E)
B. flacca Irmsch.	Cultivated: BaBG, Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-133 (E)
B. guttapila D.C. Thomas & Ardi	Cultivated: BaBG Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 08-81 (E)
B. hekensis D.C.Thomas	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 08-43 (E)
B. hispidissima Zipp. ex. Koord.	Indonesia, Sulawesi	Kinho, J. & Poulsen, A. 168 (E)

B. koordersii Warb.ex. L.B.Sm.et Wassh	Indonesia, Sulawesi	Koorders, S.H. 16246B
B. lasioura D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-110 (E)
B. macintyreana M.Hughes	Cultivated: RBGE (acc. num.: 20021848), Indonesia, Sulawesi	Thomas, D.C. 07-28 (E)
B. masarangensis Irmsch	Cultivated: PCHW Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 09-131 (E)
B. mendumiae M.Hughes	Cultivated: RBGE (acc. num.: 20021912), Indonesia, Sulawesi	Thomas, D.C. 07-27 (E)
B. negrosensis Elmer	Philippines, Negros	Wilkie, P. 76 (E)
B. nigritarum Steud.	Cultivated: RBGE (acc. num.: 19991994), Philippines, Luzon Island	Thomas, D.C. 07-25 (E)
B. nobiae spec	Indonesia, Sulawesi	Thomas 09-123 (E)
B. ozotothrix D.C.Thomas	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 08-58 (E)
B. poliloensis	Indonesia, Sulawesi	Forrest 139 (E)
B. prionota D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-97 (E)
B. pseudolateralis Warb.	Indonesia, Sulawesi	Thomas & Ardi 08-23 (E)
B. rantemarioensis D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-78 (E)
B. sanguineopilosa D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-125 (E)
B. serratipetala Imrsch.	Cultivated: RBGE (acc. num.: 19681637), Papua New Guinea	Forrest, L.L. 135 (E)
B. siccacaudata J.Door.	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 09-60 (E)
*B. spec_*New Guinea	Cultivated: BoBG,	Thomas, D. C. & Ardi, W. H.

	Indonesia, Papua	09-139 (E)
*B. spec_*Papua1	Papua	Armstrong, K. 351 (E)
*B. spec_*Philippines	Cultivated: RBGE (acc. nom.: 20080433), Philippines: Luzon Island	Thomas, D. C. 08-146 (E)
*B. spec_*Sumbawa2	Cultivated: BoBG, Indonesia, Sumbawa	Thomas, D. C. & Ardi, W. H. 09-138 (E)
*B. stevei*M.Hughes	Cultivated: RBGE (acc, num.: 20040642), Indonesia, Sulawesi	Thomas, D.C. 07-30 (E)
B. strigosa (Warb.) L.L.Forrest & Hollingsw.	Cultivated: GBG, Papua New Guinea	Forrest, L.L. 143 (E)
*B. symsanguinea*L.L. Forrest & Holingsw.	Cultivated: GBG (acc, num.: 00312793), Papua New Guinea	Forrest, L.L. 142 (E)
*B. torajana*D.C.Thomas & Ardi	Indonesia, Sulawesi	Thomas, D. C. & Ardi, W. H. 09-104 (E)
B. varipeltata D.C. Thomas	Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H. 08-51 (E)
B. vermeulenii D.C.Thomas & Ardi	Indonesia, Sulawesi	Vermeulen, J.J. 2301 (L)
B. watuwilensis Girmansyah	Cultivated: BaBG, Indonesia, Sulawesi	Thomas, D.C. & Ardi, W.H 09-55 (E)
B. weigallii Hemsl.	Solomon Islands	Pitisopa, F., Gardner, M. F., Herrington, S. 10 (E)

Appendix 2: Presence/ Absence coding of each taxa of Begoniaceae in the biogeographic analysis in Biogeobears. A= America, B= Africa, C= Socotra, D= India, E= Sri Lanka, F= Himalaya, G= China, H= Indo China, I= Penninsular Malasia, J= Sumatra, K= Borneo, L= Sulawesi, M= Papua New Guinea, N= Negros, O= Java, P= Luzon, Q= Lanyu_Batan, R= Panay, S= Palawan, T= Mindanao, U= Sibuyan, V= Biliran

Species Name	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
B. ningmingensis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. acuminatissima	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
B. anisoptera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
B. biliranensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
B. blancii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. calasiensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. calcicola	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. caminguinensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
B. castilloi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. chingipengii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. chloroneura	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. cleopatrae_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. copelandii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. dinglensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. elmeri	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
B. elnidoensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. fenecis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. gabaldonensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Species																						
B. gitingensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
B. gueritziana	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
B. gutierrezii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. hernandioides_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. hernandioides_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. hughesii_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. hughesii_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. klemmei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. longiscapa_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. longiscapa_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
B. luzonensis_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. luzonensis_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. luzonensis_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. manillensis_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. merrilliana	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. mindorensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. nigritarum_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. nigritarum_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. nigritarum_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. obtusifolia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. oxysperma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. rhombicarpa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. rubitae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
B. rufipila	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Begonia sp_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Begonia sp_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. subnummarifolia	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
B. suborbiculata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. sykakiengii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B. tagbanua	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. taraw	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. tayabensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. trichocheila	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. wadei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. woodii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. dregei	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. goudotii	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. oxyloba	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. poculifera	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. polygonoides	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. samahensis	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. socotrana	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. sutherlandii	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. boliviensis	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. nelumbifolia	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. radicans	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Species	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21
B. tenera	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. thwaitesii	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. cordifolia	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. dipetala	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. malabarica	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
*B. malabarica*_India	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. albo-coccinea	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. floccifera	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. bryophila	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. diocia	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. flagellaris	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. hatacoa	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. panchtharensis	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. picta	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. roxburghii	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. rubella	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. sikkimensis	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. tribenensis	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. aceroides	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. acetosella	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. aff. elisabethae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. alicida	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. brandisiana	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

B. demissa	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. elisabethae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. grandis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. hymenophylla	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. masoniana	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. morsei	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. obovoidea	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. palmata	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. silletensis	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. sizemoreae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. smithiae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec_China1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec_China2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec_Thailand1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec_Thailand2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. spec_Thailand3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. versicolor	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B. aff. Congesta	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
B areolata	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. bracteata	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. chlorosticta	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
B. corrugata	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

B. decora	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. fuscisetosa	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
B. goegoensis	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. harauensis	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. kingiana	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. laruei	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. longifolia	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
B. multangula	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. multijugata	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. muricata	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. pavonina	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. puspitae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. resecta	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. robusta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
*B. spec*_Borneo1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
*B. spec*_Kalimantan	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
*B. spec*_Sumatra	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
*B. spec*_Sumbawa1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
*B. spec sect Reichenheimea*1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
*B. spec sect Reichenheimea*3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
*B. spec sect Reichenheimea*4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. sublobata	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. sudjanae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
B. tenuifolia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. venusta	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B. verecunda	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B. aff. mekonggensis spec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. aff. multangula	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. aptera	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. argenteomarginata	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. bonthainensis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. brevirimosa	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. capituliformis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. chiasmogyna	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. cleopatrae_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B. comestibilis spec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. didyma	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. flacca	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. guttapila	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. hekensis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. hispidissima	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. koordersii	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. lasiouraspec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. macintyreana	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. masarangensis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B. mendumiae	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Species																					
B. negrosensis	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. nigritarum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. nobiaespec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. ozotothrix	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. poliloensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. prionotaspec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. pseudolateralis	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B. rantemarioensis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. sanguineopilosaspec	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. serratipetala	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
B. siccacaudata	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. spec_NewGuinea	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
B. spec_Papua1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
B. spec_Philippines	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B. spec_Sumbawa2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B. stevei	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. strigosa	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
B. symsanguinea	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
B. torajana	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. varipeltata	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. vermeulenii	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. watuwilensis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B. weigallii	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Appendix 3:List of taxa used in the biogeographic analysis of Sapotaceae and Voucher numbers. Species indicated in bold are newly generated sequences for the present study.

Species name	Origin	EDNA/voucher
Sarcosperma_laurinum Hook.f.	Hong Kong, China	Saunders 2000 (S)
Eberhardtia_aurata Lecomte	Malaysia	G. Hao 534 (S)
Xantolis_siamensis P.Royen	Thailand	Smitairi 1 (L)
Sideroxylon_oxyacanthum Baillon	Ethiopia, Arabia	Wood Y/75/388 (BM)
***Palaquium_laevifolium_SL* Engl.**	**Sri Lanka**	**Kumarage 60 (E)**
***Isonandra_lanceolata_SL* Wight**	**Brunei, Indonesia, borneo, Sri Lanka**	**Kumarage 47 (E)**
Nesoluma_polynesiacum Baillon	Hawaii	Degener 20770 (S)
Sideroxylon_americanum T.D.Penn.	Jamaica, Bahamas	Gillis 11576 (B)
*Sideroxylon_horridum*T.D.Pennington	Cuba	Gutiérrez & Nilsson 5 (S)
Sideroxylon_obovatum H.J.Lam	Venezuela, Aruba, virgin islands, antigua, Dutch west indies	García et al. 5586 (S)
Sideroxylon_obtusifolium T.D.Pennington	Brazil,Venezuela,Argentina, Trinidad, Paraguay, Guatemala	Alvarez et al. 28772 (B)
Sideroxylon_picardae T.D.Pennington	Dominican Republic, Haiti	Ekman 15576 (S)
Sideroxylon_occidentale baillon	Mexico	Carter & Sharsmith 4268 (B)
Sideroxylon_angustum T.D.Penn.	Cuba	Ekman 4034 (S)
Sideroxylon_lanuginosum Michx.	Mexico, USA,	Correll & Ogden 28456 (S)
Sideroxylon_lycioides L.	USA	Radford et al. 11453 (B)
Sideroxylon_persimile T.D.Pennington	Mexico,Nicaragua,Venezuela, Guatemala	Véliz 99.7038 (BM)

Sideroxylon_reclinatum Michx.	USA	Traverse 592 (GB)
Sideroxylon_tenax L.	USA	Radford & Leonard 11519 (B)
Sideroxylon_confertum C.Wright	Cuba	Ekman 17405 (S)
Argania_spinosa Skeels	Morocco	Nordenstam 9325 ?(S)
Sideroxylon_betsimisarakum Lecomte	Madagascar	Schönenberger et al. A-102 (UPS)
Sideroxylon_capiri Piettier	Guatemala, Honduras, Mexico, Nicaragua,Costa Rica, Trinidad	García 1848 (BM)
Sideroxylon_cubense T.D.Pennington	Cuba,Haiti,Dominican Republic	Beurton & Mory 927 (B)
Sideroxylon_floribundum Griseb.	Guatemala,Jamaica, Belize	Lundell 20263 (BM)
Sideroxylon_foetidissimum Jacq.	Dominican Republic, Jamaica,Virgin Islands,Guadeloupe	Lundin 638 (S)
Sideroxylon_inerme L.	Comoros, Tropical Africa, India, South Africa	Nielsen s.n. (S)
Sideroxylon_majus Baehni	Mauritius, Port Louis	Capuron 28185 (SF)
Sideroxylon_marginatum Decne.ex Webb	Cape Verde	Leyens CV-96-672 (B)
Sideroxylon_marmulano Banks ex Lowe	Madeira, Canaries, Cape Verde	Swenson & Fernandez 581 (S)
Sideroxylon_mascatense T.D.Penn.	Oman, Saudi Arabia,Ethiopia,Afganistan,Yemen	Thulin, Beier & Hussein 9774 (UPS)
Sideroxylon_repens T.D.Penn.	Dominican Republic	Greuter & Rankin 24954 (S)
Sideroxylon_salicifolium H.J.Lam	Jamaica,Mexico,Belize,Barbuda,USA,Bahamas	Gutiérrez & Nilsson 14 (S)

Sideroxylon_saxorum Lecomte	Madagascar	Jongkind 3500 (WAG)
Sideroxylon_tepicense T.D.Penn.	Mexico,Guatemala	Ctentry 2931 (S)
Sideroxylon_wightianum Hook.And Arn	Hong Kong, China	Hao 532 (S)
Omphalocarpum_pachysteloides Mildbr. Ex Hutch.& Dalziel	Ghana,Ivory coast, Congo	Jongkind 2351 (WAG)
Lecomtedoxa_klaineana Pierre ex Dubard	Tropical Africa	Veldhuizen 1509 (WAG)
Englerophytum_natalense T.D.Pennington	Tanzania	C. K. 3483 (S)
Poutera_firma (Miq.) Baehni	Indonesia, West Papua	Armstrong 305 (E)
Pouteria_maclayana Baehni	Indonesia,Malaysia,PNG	Armstrong 316 (E)
*Pouteria_sp*1	Malaysia	Armstrong 317 (E)
Neolemonniera_clitandrifolia Heine	Ghana	Jongkind, Schmidt & Abbiw 1777 (MO)
Northia_seychellana Hook.F.	Seychelles	Chong-Seng s. n. (S)
Capurodendron_androyense Aubreville	Madagascar	Humbert 28855 (B)
***Madhuca_sp*_SL**	**Sri Lanka**	**Kumarage 58 (E)**
***Madhuca_fulva*_SL J.F.Macber**	**Sri Lanka**	**Kumarage 62 (E)**
***Madhuca_neriifolia*_SL H.J.Lam**	**Sri Lanka**	**Kumarage 63 (E)**
***Madhuca_longifolia*2_SL J.F.Macbr.**	**Sri Lanka**	**Photo voucher available**
Madhuca_crassipes H.J.Lam	Borneo, Indonesia, malaysia	Jugah ak. Kudi 23757 (K)
Madhuca_utilis H.J.Lam	Malaysia	Pennington & Asri 10209 (K)
Aulandra_longifolia H.J.Lam	Malaysia, Sarawak	Christensen 1720 (K)
Palaquium_amboinense Burck	Indonesia (native to New Guinea)	Wilkie 813 (E)

Palaquium_beccarianum van Royen	Malaysia,Papua, Borneo	Wiijesundara s.n. (K) – native to West Papua and the Moluccas
Diploknema_oligomera H.J.Lam	Indonesia, Bogor	Chase 1360 (K)
Palaquium_brassii H.J.Lam	Indonesia, West Papua	Armstrong 311 (E)
Palaquium_clarkeanum King & Gamble	Malaysia	Wilkie 501 (E)
Palaquium_cryptocarifolium P.Royen	Malaysia, Sarawak	Wilkie 874 (E)
Palaquium_dasyphyllum Pierre	Malaysia	Slik 9592 (L)
Palaquium_eriocalyx H.J.Lam	Indonesia	Wilkie et al. 8/147 (E)
Palaquium_formosanum Hayata	Taiwan,Philippines	Chung & Anderberg 1421 (HAST)
Palaquium_galactoxylum H.J.Lam	Australia	Bartish and Jessup 9 (S)
Palaquium_gutta Baillon	Malaysia	Wilkie 504 (E)
Palaquium_herveyi King & Gamble	Malaysia, Sarawak	Wilkie 871 (E)
Palaquium_impressionervium Ng	Peninsular Malaysia	Wilkie FRI52851 (E)
Palaquium_leiocarpum Boeriage	Malaysia, Sarawak	Wilkie 870 (E)
Palaquium_lobbianum Kurck	Indonesia	Armstrong 331 (E)
Palaquium_hexandrum Baillon	Malaysia, Sarawak	Wilkie 872 (E)
Palaquium_maingayi King & Gamble	Peninsular Malaysia	Wilkie 846 (E)
Palaquium_microphyllum King & Gamble	Malaysia	Pennington, Kochummen & Wong (K)
Palaquium_obovatum Engler	Thailand	Middleton 4387 (E)
Palaquium_oxleyanum Pierre	Malaysia	Wilkie 527 (E)
Palaquium_pseudorostratum H.J.Lam	Malaysia, Sarawak	Wilkie 857 (E)
Palaquium_quercifolium Burck	Indonesia, Kalimantan	Slik CMF6780 (L)
Palaquium_ridleyi King & Gamble	Malaysia, Sarawak	Wilkie 858 (E)

Palaquium_rigidum Pierre ex Dubard	Malaysia	Wilkie 878 (E)
Palaquium_rostratum Burck	Indonesia, Kalimantan	Slik CMF9452 (L)
Palaquium_rufolanigerum P.Royen	Malaysia, Sarawak	Wilkie 859 (E)
Palaquium_sericeum H.J.Lam	Indonesia, Kalimantan	Slik CMF9737 (L)
*Palaquium_sp*1	Solomon Islands	Poulsen et al 2488 (E)
*Palaquium_sp*2	Indonesia, Sulawesi	Armstrong 370 (E)
Palaquium_sumatranum Burck	Indonesia, Java	Wilkie 823 (E)
Palaquium_walsurifolium Pierre ex Dubard	Malaysia, Sarawak	Wilkie 877 (E)
Palaquium_calophyllum Pierre ex Burck	Indonesia, Kalimantan	Wilkie et al 1/477 (E)
Palaquium_xanthochymum Pierre ex Burck	Malaysia, Borneo	Wilkie 544 (E)
***Palaquium_thwaitesii*_SL Trim.**	**Sri Lanka**	**Kumarage 43 (E)**
***Palaquium_canaliculatum*_SL Engl.**	**Sri Lanka**	**Kumarage 45 (E)**
***Palaquium_hinmolpedda*_SL P.Royen**	**Sri Lanka**	**Kumarage 64 (E)**
***Palaquium_pauciflorum*_SL Engl.**	**Sri Lanka**	**Kumarage 59 (E)**
***Palaquium_rubiginosum*_SL Engl.**	**Sri Lanka**	**Kumarage 65 (E)**
***Isonandra_montana*_SL Gamble**	**Sri Lanka**	**Kumarage 76 (E)**
***Isonandra_sp1*_SL**	**Sri Lanka**	**Kumarage 97 (E)**
***Palaquium_sp3*_SL**	**Sri Lanka**	**Kumarage 57 (E)**
***Palaquium_grande*_SL Engl.**	**Sri Lanka**	**Kumarage 46 (E)**
***Palaquium_petiolare*_SL Engl.**	**Sri Lanka**	**Kumarage 44 (E)**
***Palaquium_sp5*_SL**	**Sri Lanka**	**Kumarage 93 (E)**
***Isonandra_sp2*_SL**	**Sri Lanka**	**Kumarage 74 (E)**
***Isonandra_compta*_SL Dubard**	**Sri Lanka**	**Emanuelsson 3039 (S)**

*Isonandra_lanceolata*2 Wight	malaysia, Brunei, Borneo, Sri Lanka	Armstrong 314 (E)
Isonandra_perakensis King & Gamble	Malaysia	Pennington & Wong 10227 (K)
***Isonandra_zeylanica*_SL Jeuken**	**Sri Lanka**	**Kumarage 72 (E)**
Autranella_congolensis A.Chev.	Congo	Bokdam 4401 (WAG)
Tieghemella_heckelii Pierre ex Dubard	Ghana	Jongkind 3936 (WAG)
Vitellaria_paradoxa C.F.Gaertn.	Africa	Neumann 1512 (FR)
Vitellariopsis_cuneata Aubreville	Tropical Africa, Indonesia	Thomas 3662 (S)
Vitellariopsis_dispar Aubreville	South Africa	Balkwill & Balkwill (B)
Vitellariopsis_marginata Aubreville	Africa	Chase 1122 (S)
Faucherea_parvifolia Lecomte	Madagascar	Birkinshaw et al. 357(P)
Inhambanella_henriquesii Dubard	South Africa	de Winter & Vahrmeijer 8536 (S)
Baillonella_toxisperma Pierre	Africa	Cenarest (LBV)
Labourdonnaisia_calophylloides Bojer	Mauritius, Port Louis	Capuron 28241-SF (P)
Labourdonnaisia_revoluta Bojer	Mauritius	Bernardi 14717 (P)
Labramia_costata Aubreville	Madagascar	Randrianmanalinarivo 577 (UPS)
Labramia_mayottensis Labat, M.Pignal & O.Pascal	Madagascar	Labat et al. 3309 (S)
Letestua_durissima Lecomte	Congo	Normand s.n. (P)
***Manilkara_hexandra*_SL Dubard**	**Sri Lanka,Hawaii,Vietnam,India,Thailand, Himalaya**	**Photo voucher available**
Manilkara_concolor Gerstner	Africa	Labat et al 3309 (S)
Manilkara_discolor J.H.Hemsl.	Africa	K. Vollesen 2460 (S)
Manilkara_kauki Dubard	Thailand	Chantaranothai 2341

		(Khon Kaen Univ. Herbarium)
Manilkara_obovata J.H.Hemsl.	Africa	Schmidt et al 3274 (S)
Manilkara_zapota P.Royen	Colombia, Jamaica,Costa Rica,Mexico	J. Clayton 12 (S)
Mimusops_caffra E.Mey ex A.d.Candolle	Africa	Swenson & Karis 636 (S)
Mimusops_comorensis Engler	Comoros	Pignal & Ginguette 1065
***Mimusops_elengi_SL* L.**	**Sri Lanka**	**Kumarage 98 (E)**
Mimusops_obovata Sond.	Africa, Madagascar	Swenson & Karis 633 (S)
Mimusops_zeyheri Sond.	Africa	Dahlstrand 6386 (S)
Burckella_macropoda H.J.Lam	Indonesia (native to New Guinea)	Wilkie 818 (E)
Burckella_polymera P.Royen	Indonesia, West Papua	Armstrong 326 (E)
Burckella_sp	India	Armstrong 327 (E)
Diploknema_butyracea H.J.Lam	Nepal	Polunin, Sykes & Williams 3975 (UPS)
Diploknema_siamensis Fletcher	China	Middleton s.n.
Madhuca_barbata T.D.Penn.	Sarawak, Malaysia	Wilkie 873 (E)
Madhuca_curtisii Ridl.	Peninsular Malaysia	Wilkie 848 (E)
Madhuca_elmeri Merr. Ex. H.J.Lam	Indonesia, Kalimantan	Wilkie P1 347 (E)
Madhuca_erythrophylla H.J.Lam	Malaysia, Sarawak	Wilkie 867 (E)
Madhuca_hainanensis Chun & F.C.Chow	China	G. Hao 530 (S)
Madhuca_kingiana H.J.Lam	Malaysia, Sarawak	Wilkie 856 (E)
Madhuca_korthalsii H.J.Lam	Malaysia, Sarawak	Wilkie 876 (E)
Madhuca_kuchingensis Yii & P.Chai	Malaysia, Sarawak	Wilkie 860 (E)

Madhuca_kunstleri H.J.Lam	Malaysia, Sarawak	Wilkie 868 (E)
Madhuca_lancifolia H.J.Lam	Malaysia, Sarawak	Wilkie 853 (E)
Madhuca_laurifolia H.J.Lam	Peninsular Malaysia	Wilkie 843 (E)
Madhuca_leucodermis H.J.Lam	New Guinea	Takeuchi et al. 17858 (S)
Madhuca_malaccensis H.J.Lam	Peninsular Malaysia	Wilkie 832 (E)
Madhuca_microphylla Alston	Sri Lanka	Fagerlind 4790 (S)
Madhuca_motleyana Baehni	Peninsular Malaysia	Wilkie 837 (E)
Madhuca_oblongifolia Merrill	Sarawak, Malaysia	Wilkie 861 (E)
Madhuca_pachyphylla (K.Krause) ined.	Indonesia, West Papua	Armstrong 313 (E)
Madhuca_palembanica Forman	Indonesia	Triono, Saman & Victobery 11 (K)
Madhuca_palida Baehni	Sarawak, Malaysia	Wilkie 869 (E)
Madhuca_prolixa Yii & P.Chai	Sarawak, Malaysia	Wilkie 875 (E)
Madhuca_sarawakensis H.J.Lam	Sarawak, Malaysia	Wilkie 863 (E)
Madhuca_sericea H.J.Lam	Sarawak, Malaysia	Wilkie 879 (E)
Madhuca_sp	Peninsular Malaysia	Wilkie 834 (E)
Payena_acuminata Pierre	Indonesia, Java	Chase 1368 (K)
***Madhuca_clavata*_SL Jayas.**	**Sri Lanka**	**Kumarage 99 (E)**
***Palaquium_sp*4_SL**	**Sri Lanka**	**Kumarage 75 (E)**
*Payena_lucida*_1 A.de.Candolle	Peninsular Malaysia	Wilkie 845 (E)
Payena_maingayi C.B.Clarke	Peninsular Malaysia	Wilkie 841 (E)
Payena_leerii Kurz	Malaysia	Wilkie 811 (E)
Payena_ferruginea J.T.Pereira	Peninsular Malaysia	Wilkie 866 (E)
*Payena_lucida*_2 A.d.Candolle	Borneo	Ambri et al. AA1604 (L)
Payena_obscura Burck	Peninsular Malaysia	Wilkie 880 (E)

Appendix 4:Presence/ Absence coding of each taxa of Sapotaceae in the biogeographic analysis in Biogeobears: A= America, B= Africa, C= Madagascar, D= Seychelles, E= Middle East, F= India, G= Sri Lanka, H= Himalaya, I= East Asia, J= Sunda Shelf, K= Sahul Shelf

Species name	A	B	C	D	E	F	G	H	I	J	K
Sarcosperma_laurinum	0	0	0	0	0	0	0	0	1	0	0
Eberhardtia_aurata	0	0	0	0	0	0	0	0	0	1	0
Xantolis_siamensis	0	0	0	0	0	0	0	0	1	0	0
Sideroxylon_oxyacanthum	0	1	0	0	1	0	0	0	0	0	0
Palaquium_laevifolium_SL	0	0	0	0	0	0	1	0	0	0	0
Isonandra_lanceolata_SL	0	0	0	0	0	0	1	0	0	1	0
Nesoluma_polynesiacum	0	0	0	0	0	0	0	0	0	0	1
Sideroxylon_americanum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_horridum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_obovatum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_obtusifolium	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_picardae	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_occidentale	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_angustum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_lanuginosum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_lycioides	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_persimile	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_reclinatum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_tenax	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_confertum	1	0	0	0	0	0	0	0	0	0	0
Argania_spinosa	0	1	0	0	0	0	0	0	0	0	0
Sideroxylon_betsimisarakum	0	0	1	0	0	0	0	0	0	0	0
Sideroxylon_capiri	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_cubense	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_floribundum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_foetidissimum	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_inerme	0	1	0	0	0	1	0	0	0	0	0
Sideroxylon_majus	0	1	0	0	0	0	0	0	0	0	0

Species											
Sideroxylon_marginatum	0	1	0	0	0	0	0	0	0	0	0
Sideroxylon_marmulano	0	1	0	0	0	0	0	0	0	0	0
Sideroxylon_mascatense	0	1	0	0	1	0	0	0	0	0	0
Sideroxylon_repens	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_salicifolium	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_saxorum	0	0	1	0	0	0	0	0	0	0	0
Sideroxylon_tepicense	1	0	0	0	0	0	0	0	0	0	0
Sideroxylon_wightianum	0	0	0	0	0	0	0	0	1	0	0
Omphalocarpum_pachysteloides	0	1	0	0	0	0	0	0	0	0	0
Lecomtedoxa_klaineana	0	1	0	0	0	0	0	0	0	0	0
Englerophytum_natalense	0	1	0	0	0	0	0	0	0	0	0
Poutera_firma	0	0	0	0	0	0	0	0	0	1	1
Pouteria_maclayana	0	0	0	0	0	0	0	0	0	1	1
*Pouteria_sp*1	0	0	0	0	0	0	0	0	0	1	0
Neolemonniera_clitandrifolia	0	1	0	0	0	0	0	0	0	0	0
Northia_seychellana	0	0	0	1	0	0	0	0	0	0	0
Capurodendron_androyense	0	0	1	0	0	0	0	0	0	0	0
Madhuca_sp_SL	0	0	0	0	0	0	1	0	0	0	0
Madhuca_fulva_SL	0	0	0	0	0	0	1	0	0	0	0
Madhuca_neriifolia_SL	0	0	0	0	0	0	1	0	0	0	0
Madhuca_longifolia2_SL	0	0	0	0	0	0	1	0	0	0	0
Madhuca_crassipes	0	0	0	0	0	0	0	0	0	1	0
Madhuca_utilis	0	0	0	0	0	0	0	0	0	1	0
Aulandra_longifolia	0	0	0	0	0	0	0	0	0	1	0
Palaquium_amboinense	0	0	0	0	0	0	0	0	0	1	1
Palaquium_beccarianum	0	0	0	0	0	0	0	0	0	1	1
Diploknema_oligomera	0	0	0	0	0	0	0	0	0	1	0
Palaquium_brassii	0	0	0	0	0	0	0	0	0	0	1
Palaquium_clarkeanum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_cryptocarifolium	0	0	0	0	0	0	0	0	0	1	0
Palaquium_dasyphyllum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_eriocalyx	0	0	0	0	0	0	0	0	0	1	0
Palaquium_formosanum	0	0	0	0	0	0	0	0	0	1	0

Palaquium_galactoxylum	0	0	0	0	0	0	0	0	0	0	1
Palaquium_gutta	0	0	0	0	0	0	0	0	0	1	0
Palaquium_herveyi	0	0	0	0	0	0	0	0	0	1	0
Palaquium_impressionervium	0	0	0	0	0	0	0	0	0	1	0
Palaquium_leiocarpum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_lobbianum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_hexandrum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_maingayi	0	0	0	0	0	0	0	0	0	1	0
Palaquium_microphyllum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_obovatum	0	0	0	0	0	0	0	0	1	0	0
Palaquium_oxleyanum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_pseudorostratum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_quercifolium	0	0	0	0	0	0	0	0	0	1	0
Palaquium_ridleyi	0	0	0	0	0	0	0	0	0	1	0
Palaquium_rigidum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_rostratum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_rufolanigerum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_sericeum	0	0	0	0	0	0	0	0	0	1	0
*Palaquium_sp*1	0	0	0	0	0	0	0	0	0	0	1
*Palaquium_sp*2	0	0	0	0	0	0	0	0	0	1	0
Palaquium_sumatranum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_walsurifolium	0	0	0	0	0	0	0	0	0	1	0
Palaquium_calophyllum	0	0	0	0	0	0	0	0	0	1	0
Palaquium_xanthochymum	0	0	0	0	0	0	0	0	0	1	0
*Palaquium_thwaitesii*_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_canaliculatum*_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_hinmolpedda*_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_pauciflorum*_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_rubiginosum*_SL	0	0	0	0	0	0	1	0	0	0	0
*Isonandra_montana*_SL	0	0	0	0	0	0	1	0	0	0	0
*Isonandra_sp*1_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_sp*3_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_grande*_SL	0	0	0	0	0	0	1	0	0	0	0

Species											
*Palaquium_petiolare_*SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_sp5_*SL	0	0	0	0	0	0	1	0	0	0	0
*Isonandra_sp2_*SL	0	0	0	0	0	0	1	0	0	0	0
*Isonandra_compta_*SL	0	0	0	0	0	0	1	0	0	0	0
*Isonandra_lanceolata*2	0	0	0	0	0	0	0	0	0	0	1
Isonandra_perakensis	0	0	0	0	0	0	0	0	0	1	0
*Isonandra_zeylanica_*SL	0	0	0	0	0	0	1	0	0	0	0
Autranella_congolensis	0	1	0	0	0	0	0	0	0	0	0
Tieghemella_heckelii	0	1	0	0	0	0	0	0	0	0	0
Vitellaria_paradoxa	0	1	0	0	0	0	0	0	0	0	0
Vitellariopsis_cuneata	0	1	0	0	0	0	0	0	0	1	0
Vitellariopsis_dispar	0	1	0	0	0	0	0	0	0	0	0
Vitellariopsis_marginata	0	1	0	0	0	0	0	0	0	0	0
Faucherea_parvifolia	0	0	1	0	0	0	0	0	0	0	0
Inhambanella_henriquesii	0	1	0	0	0	0	0	0	0	0	0
Baillonella_toxisperma	0	1	0	0	0	0	0	0	0	0	0
Labourdonnaisia_calophylloides	0	1	0	0	0	0	0	0	0	0	0
Labourdonnaisia_revoluta	0	1	0	0	0	0	0	0	0	0	0
Labramia_costata	0	0	1	0	0	0	0	0	0	0	0
Labramia_mayottensis	0	0	1	0	0	0	0	0	0	0	0
Letestua_durissima	0	1	0	0	0	0	0	0	0	0	0
*Manilkara_hexandra_*SL	0	0	0	0	0	1	1	0	0	0	0
Manilkara_concolor	0	1	0	0	0	0	0	0	0	0	0
Manilkara_discolor	0	1	0	0	0	0	0	0	0	0	0
Manilkara_kauki	0	0	0	0	0	0	0	0	1	0	0
Manilkara_obovata	0	1	0	0	0	0	0	0	0	0	0
Manilkara_zapota	1	0	0	0	0	0	0	0	0	0	0
Mimusops_caffra	0	1	0	0	0	0	0	0	0	0	0
Mimusops_comorensis	0	1	0	0	0	0	0	0	0	0	0
*Mimusops_elengi_*SL	0	0	0	0	0	1	1	0	0	0	0
Mimusops_obovata	0	1	1	0	0	0	0	0	0	0	0
Mimusops_zeyheri	0	1	0	0	0	0	0	0	0	0	0
Burckella_macropoda	0	0	0	0	0	0	0	0	0	1	1

Burckella_polymera	0	0	0	0	0	0	0	0	0	1	1
Burckella_sp	0	0	0	0	0	1	0	0	0	0	0
Diploknema_butyracea	0	0	0	0	0	0	0	1	0	0	0
Diploknema_siamensis	0	0	0	0	0	0	0	0	1	0	0
Madhuca_barbata	0	0	0	0	0	0	0	0	0	1	0
Madhuca_curtisii	0	0	0	0	0	0	0	0	0	1	0
Madhuca_elmeri	0	0	0	0	0	0	0	0	0	1	0
Madhuca_erythrophylla	0	0	0	0	0	0	0	0	0	1	0
Madhuca_hainanensis	0	0	0	0	0	0	0	0	1	0	0
Madhuca_kingiana	0	0	0	0	0	0	0	0	0	1	0
Madhuca_korthalsii	0	0	0	0	0	0	0	0	0	1	0
Madhuca_kuchingensis	0	0	0	0	0	0	0	0	0	1	0
Madhuca_kunstleri	0	0	0	0	0	0	0	0	0	1	0
Madhuca_lancifolia	0	0	0	0	0	0	0	0	0	1	0
Madhuca_laurifolia	0	0	0	0	0	0	0	0	0	1	0
Madhuca_leucodermis	0	0	0	0	0	0	0	0	0	1	0
Madhuca_malaccensis	0	0	0	0	0	0	0	0	0	1	0
Madhuca_microphylla	0	0	0	0	0	0	1	0	0	0	0
Madhuca_motleyana	0	0	0	0	0	0	0	0	0	1	0
Madhuca_oblongifolia	0	0	0	0	0	0	0	0	0	1	0
Madhuca_pachyphylla	0	0	0	0	0	0	0	0	0	1	1
Madhuca_palembanica	0	0	0	0	0	0	0	0	0	1	0
Madhuca_palida	0	0	0	0	0	0	0	0	0	1	0
Madhuca_prolixa	0	0	0	0	0	0	0	0	0	1	0
Madhuca_sarawakensis	0	0	0	0	0	0	0	0	0	1	0
Madhuca_sericea	0	0	0	0	0	0	0	0	0	1	0
Madhuca_sp	0	0	0	0	0	0	0	0	0	1	0
Payena_acuminata	0	0	0	0	0	0	0	0	0	1	0
*Madhuca_clavata*_SL	0	0	0	0	0	0	1	0	0	0	0
*Palaquium_sp*4_SL	0	0	0	0	0	0	1	0	0	0	0
*Payena_lucida*_1	0	0	0	0	0	0	0	0	0	1	0
Payena_maingayi	0	0	0	0	0	0	0	0	0	1	0
Payena_leerii	0	0	0	0	0	0	0	0	0	1	0

Payena_ferruginea	0	0	0	0	0	0	0	0	0	1	0
Payena_lucida_2	0	0	0	0	0	0	0	0	0	1	0
Payena_obscura	0	0	0	0	0	0	0	0	0	1	0

Appendix 5: List of taxa used in the biogeographic analysis of Zingiberaceae, Voucher number and GenBank accession number of gene sequences. Species indicated in bold are newly generated sequences for the present study.

Species name	Voucher	ITS	*matK*
*Siphonochilus_aethiopicus*2 (Schweinf.) B.L.Burtt	GH 00-134 (US)	AF478792	AF478892
*Siphonochilus_decorus*1 (Drutten) Lock	GH 00-135 (US)	AF478793	AF478894
Siphonochilus_kirkii (Hook. F.) B.L. Burtt	Kress 92-3404	AF478794	AF478895
Tamijia_flagellaris S.Sakai & Nagam.	Kazuyuki S55 (KYO)	AF478797	AF478898
Distichochlamys_sp	Kress 01-6848 (E)	AF478745	AF478844
Alpinia_ligulata1 K.Schum.	L-98.0560 (LYON)	AY472361	AY742415
Alpinia_nieuwenhuizii Valeton	L-92.0474 (KES(LYON))	AY742366	AY742421
Alpinia_guangdongensis S.J.Chen & Z.Y.Chen	Kress 99-6372 (US)	AY742352	AY742408
Alpinia_maclurei Merr.	Kress 95-5540 (US)A	AY742362	AY742416
Alpinia_rubricaulis K.Schum.	Sul 02-251 (E)	AY742378	AY742432
Alpinia_zerumbet B.L.Burtt & R.M. Sm.	Liao 020704 (SCIB)	AY742389	AY742443
Alpinia_oxyphylla Miq.	Liao 020707 (SCIB)	AY742372	AY742425
Alpinia_brevilabris (Valeton) R.M.Sm.	Kress 94-5335 (US)	AY742341	AY742399
Alpinia_polyantha D. Fang	Kress 94-3744 (NMNH(SI))	AY745692	AY742427
Alpinia_nutans Roscoe	L-91.0066 (CS (LYON))	AY742369	AY742423
Alpinia_foxworthyi Ridl.	Kress 98-6293 (US)	AF478714	AF478814
Alpinia_rosea Elmer	Cronk 25436 (E)	AY742377	AY742431
Gagnepainia_harmandii K. Schum.	Williams 02-601 (DUKE)	AY339740	AY341103
Curcuma_roscoeana Wall.	Kress 98-6253 (US)	AF478739	AF478839
Hedychium_greenei W.W. Sm.	Kress 94-3721 (US)	AF478759	AF478858
Hedychium_villosum Wall.	Kress 00-6603 (US)	AF478762	AF478861

Zingiber_sulphureum I. Theilade	Kress 00-6719 (US)	AF478801	AF478904
Zingiber_wrayii Ridl	Kress 00-6721 (US)	AF478802	AF478905
Gagnepainia_thoreliana K. Schum	Williams 00-199 (DUKE)	AF478752	AF478851
Hemiorchis_burmanica Kurz	Kress 94-3721 (US)	AF478764	AF478862
Hemiorchis_rhodorrhachis K.Schum	Newman 861 (E)	AF478763	AF478863
Alpinia_sibuyanensis Elmer	L-99.0098 (LYON)	AY742381	AY742434
Alpinia_vulcanica Elmer	Wilkie 29129 (E)	AY742387	AY742441
Alpinia_flabellata Ridl.	None given	AY742349	AY742406
Alpinia_japonica (Thunb.)Miq.	Kress 99-6360 (HLA (LYON)	AY742357	AY742412
Alpinia_mutica Roxb.	L-94.0307 (CS (LYON))	AY742365	AY742420
Alpinia_coriacea T.L.Wu & S.J.Chen	Kress 95-5539 (US)	AY742344	AY742402
*Alpinia_rugosa*J.P.Liao,ined.	Liao 020702 (SCIB)	AY742379	AY742433
Alpinia_shimadae Hayata	Liu 1140-1 (TNU)	FJ496775	FJ496766
Alpinia_pumila Hook.f.	Kress 97-6119 (US)	AF478719	AF478819
Alpinia_formosana K.Schum	Kress 94-5336 (US)	AY742350	AY742407
Alpinia_intermedia Gagnep.	Kress 97-5780 (US)	AF478716	AF478816
Alpinia_coeruleoviridis K.Schum.	Johansson et al 240 (E)	AY742343	AY742401
Camptandra_parvula1 (Baker) Ridl.	Kress 99-6328 (US)	AF478730	AF478830
Alpinia_monopleura K.Schum.	SUL 143 (E)	AY742363	AY742419
Alpinia_cylindrocephala K.Schum.	SUL 02-60 (E)	AY742345	AY724203
Alpinia_aenea B.L.Burtt & R.M. Sm.Koidz.	Argent 0016 (E)	AY742351	AY742394
Alpinia_carolinensis Koidz.	Kress 94-3657 (US)	AF478711	AF478811
Alpinia_eremochlamys K.Schum	SUL 02-68 (E)	AY742346	AY742404
Renealmia_cernua (Roem. & Schult) Macbride	Kress 99-6414 (US)	AF478780	AF478881
*Renealmia_alpinia*1 (Rottb.) Maas.	Kress 99-6407 (US)	AF478778	AF478879
Renealmia_thyrsoidea (R.&P.) Poepp. & Endl.	Kress 99-6406 (LYON)	AF478783	AF478884
Burbidgea_nitida Hook f.	Kress 97-5781 (US)	AF478728	AF478828
*Burbidgea_schizocheila*1 Hackett	Kress 01-6867 (US)	AF478729	AF478829

Riedelia_sp	GH 96-281 (US)	AF478785	AF478886
*Pleuranthodium_trichocalyx*1 (Valeeton) R.M.Sm.	L-94.0315 (KMN)	AY742332	AY742392
Pleuranthodium_floccosum (Valeton) R.M.Sm.	L-78-0460 (KMN)	AY742333	AY742391
*Pleuranthodium_schlechteri*1 (K.Schum.)R.M.Sm.	Kress 00-6725 (US)	AF478775	AF478876
Amomum_mentawaiense Droop	ADP 2249 (E)		
Amomum_cerasinum Ridl.	AJD 160 (E)		
Amomum_aculeatum Roxb.	AJD 96 (E)		
*Amomum_pierreanum*2 Gagnep.	JKA 222 (E)		
Amomum_coriandriodorum S.Q.Tong et Y.M.Xia	Xia 721 (HITBC)	AY351987	AY352016
Amomum_paratsaoko S.Q.Tong et Y.M.Xia	Kress 00-6791	AY351997	AY352027
Amomum_aff._paratsaoko S.Q.Tong et Y.M.Xia	Kress 98-6197 (US)	AY351998	AY352028
*Amomum_xanthophlebium*Baker	AJD 81 (E)		
Alpinia_oceanica Burkill	L-94.0309 (LYON)	AY742370	AY742424
Amomum_staminidivum A.I.Gobilik,A.L. Lamb & A.D.Poulsen	ADP 2113		
Amomum_quadratolaminare S.Q.Tong	Xia 729 (HITBC)	AY352003	AY352033
*Amomum_krervanh*2 Pierre ex Gagnep.	AJD 10 (E)		
*Amomum_krervanh*3 Pierre ex Gagnep.	AJD 16 (E)		
*Amomum_villosum*Lour.	Kress 01-6978 (US)	AF478724	AF478824
*Amomum_sp*10	Kress 99-6373 (US)	AF478723	AF478823
*Siliquamomum_tonkinense*Baill.	Kress 00-6802 (US)	AF478791	AF478892
Alpinia_rafflesiana Baker	Rangsiruji 52 (E)	AY742376	AY742430
*Amomum_uliginosum*2 J.Konig.	99-0474 (LYON)	AY352008	AY352038
*Amomum_compactum*3 Sol.	Xia 720 (HITBC)	AY351986	AY352016

Ex.Maton			
Alpinia_purpurata (Viell.) K.Schum.	Rangsiruji 9 (LYON)	AY742375	AY742429
Alpinia_arctiflora (F.Muell.) Benth.	Rangsiruji 48 (E)	AY742336	AY742395
Alpinia_caerulea (R.Br.)Benth.	L-870285 (NSW)	AY742342	AY742400
Alpinia_vittata W.Bull.	Kress 99-6415 (US)	AF478720	AF478820
*Geocharis_sp*2	AJD 19 (E)		
Amomum_laxesquamosum K.Schum.	LYON 2000.0388b(HLA LYON)	AY351994	AY352024
Amomum_oligophyllum Droop	AJD 155 (E)		
*Amomum_hastilabium*2 Ridl.	ADP 2262 (E)		
*Amomum_hastilabium*1 Ridl.	AJD 76 (E)		
*Amomum_sp*9	AJD 92 (E)		
*Amomum_pseudofoetens*Valeton	ADP 2284 (E)		
Alpinia_modesta (F.mMuell.) K. Schum.	234226 (L-86.0459) (NSW)	AY742364	AY742418
Alpinia_arundelliana (F.M.Bailey)K.Schum.	L-95.00014 (NSW)	AY742338	AY742397
*Geocharis_sp*1	AJD 106 (E)		
Vanoverberghia_sepulchrei (Merr.)	Kress 95-5562 (US)	AF478798	478899
*Etlingera_elatior*1 (Jack) R.M.Sm.	Kress 94-4251 (US)	AF478749	AF478848
Etlingera_triorgyalis (Baker)R.M.Sm.	Ped. & Joh. 1065e (C)	AF414475	AF434864
Amomum_propinquum Ridl.	93-0558 (LYON)	AY351999	AY352029
Amomum_koenigii J.F.Gmel.	Xia 723 (HITBC)	AY351991	AY352021
Amomum_yunnanense S.Q.Tong.	Xia 737 (HITBC)	AY352012	AY352042
Amomum_sericeum Roxb.	Xia 730 (HITBC)	AY352005	AY352035
Elettariopsis_smithiae	Kress 99-6313 (US)	AY352013	AY352043
Amomum_menglaense S.Q.Tong	Xia 726 (HITBC)	AY351996	AY352026
Elettariopsis_sp	Kress 00-6720 (US)	AF478747	AF478846
Amomum_putrescens D.Fang.	Xia 728 (HITBC)	AY352002	AY352032
Amomum_queenslandicum R.M.Sm.	Kmn 1428 (HTBC)	AY352004	AY352034
*Amomum_austrosinense*D.Fang.	Xia 719 (HITBC)	AY351985	AY352015
Amomum_purpureorubrum S.Q.Tong	Xia 727 (HITBC)	AY352000	AY352030

et Y.M.Xia			
Amomum_maximum Roxb.	Xia 725 (HITBC)	AY351995	AY352025
*Amomum_glabrum*2 S.Q.Tong	Kress 00-6715 (US)	AF478721	AF478821
Aframomum_angustifolium (Sonn.) K.Schum.	Kress 92-3403 (US)	AF478704	AF478804
*Alpinia_galanga*2 (L.)Wild.	Kress 02-7213d (US)	AF478715	AF478815
Alpinia_conchigera Griff.	Kress 00-6706 (US)	AF478712	AF478812
Elettariopsis_kerbyi R.M. Sm.	Kress 96-5746 (US)	AF478746	AF478845
Elettariopsis_stenosiphon B.l.Burtt & R.M. Sm.	Kress 01-6847 (US)	AF478748	AF478847
Amomum_petaloideum (S.Q.Tong) T.L.Wu	Kress95-5508 (US)	AY769826	AY769789
Zingiber_cylindricum* Moon	**Kumarage 7 (E)**		
Amomum_pterocarpum* Thwaites	**Kumarage 21 (E)**		
Zingiber_wightianum* Thwaites	**Kumarage 20 (E)**		
Amomum_sp*1	**Kumarage 34 (E)**		
Elettaria_sp	**Kumarage 5 (E)**		
Curcuma_zedoaria* Roxb.	**Kumarage 24 (E)**		
Amomum_fulviceps* Thwaites	**Kumarage 33 (E)**		
Amomum_sp*2	**Kumarage 50 (E)**		
Curcuma_albiflora* Thwaites	**Kumarage 38 (E)**		
Amomum_graminifolium* Thwaites	**Kumarage 48 (E)**		
Amomum_sp*3	**Kumarage 82 (E)**		
Amomum_nemorale* Benth. & Hook.f.	**Kumarage 52 (E)**		
Cyphostigma_pulchellum* Benth.	**Kumarage 54 (E)**		
Amomum_acuminatum* Thwaites	**Kumarage 55 (E)**		
Amomum_masticatorium* Thwaites	**Kumarage67 (E)**		
Alpinia_abundiflora*1 Burtt & R.M.Sm.	**Kumarage 70 (E)**		
Amomum_sp*5	**Kumarage 86 (E)**		
Amomum_sp*6	**Kumarage 88 (E)**		
Amomum_sp*7	**Kumarage 89 (E)**		

Alpinia_sp	**Kumarage 71 (E)**		
Alpinia_fax Burtt & R.M.Sm.	**Kumarage 94 (E)**		
Alpinia_hookeriana Valeton	L-91.0064 (CS (LYON))	AY742356	AY742411
Alpinia_latilabris Ridl.	L-80.0710 (LYON)	AY742360	AY742414
Alpinia_argentea R.M.Sm.	L-95-0364 (US)	AY742337	AY742396
Alpinia_warburgii K.Schum.	Sul 02-169 (E)	AY742388	AY742442
Alpinia_tonkinensis Gagnep.	Liao 020709 (US)	AF478720	AF478820
Alpinia_pricei Hayata	Kress 01-6860 (LYON)	AY742374	AY742428
Alpinia_calcarata Roscoe	Kress 94-3657	AF478710	AF278810
Alpinia_bilamellata Makino	L-97-0268 (LYON)	AY742339	AY742398
Alpinia_nigra (Gaertn.) B.L.Burtt	Kress 01-6880	AY742444	AY742445
Leptosolena_haenkei C.Presl.	Funakoshi & Co 2006 (US)	AY742331	AY742390
*Amomum_lappaceum*1 Ridl.	ADP2353		
Alpinia_pinetorum(Ridl.) Loes.	L-87.0607 (LYON)	AY742373	AY742426
*Amomum_centrocephalum*1 A.D.Poulsen	AJD 29 (E)		
Hornstedtia_hainanensis T.L.Wu & S.J.Chen	Kress 97-5769 (US)	AF478766	AF478865
*Elettariopsis_sp*5	AJD 65 (E)		
Alpinia_luteocarpa Elmer	Kress 99-6403 (US)	AF478717	AF478817
*Amomum_compactum*2 Sol. Ex Maton	AJD 59		
Alpinia_elegans K.Schum.	Kress 99-6412 (US)	AF478713	AF478813
*Amomum_uliginosum*1 J.Konig.	99-0474 (LYON)	AY352008	AY352038
Aframomum_daniellii (Hook.f.)K.Schum.	Kress 99-6375 (US)	AF478705	AF478805
Aframomum_sceptrum K.Schum.	Carle 94-9097	AF478706	AF478806
Renealmia_battenbergiana Cummins ex Baker	Kress 94-5277(US)	AF478779	AF478880

Appendix 6: Presence/ Absence coding of each taxa of Zingiberaceae ceae in the biogeographic analysis in Biogeobears: A= America, B= Africa, C= Madagascar, D= India, E= Sri Lanka, F= China, G= Indo China, H= Sunda Shelf, I= Sahul Shelf, J= Papua New Guinea, K= Philippines, L= Australia

Species name	A	B	C	D	E	F	G	H	I	J	K	L
*Siphonochilus_aethiopicus*2	0	1	0	0	0	0	0	0	0	0	0	0
*Siphonochilus_decorus*1	0	1	0	0	0	0	0	0	0	0	0	0
Siphonochilus_kirkii	0	0	1	0	0	0	0	0	0	0	0	0
Tamijia_flagellaris	0	0	0	0	0	0	0	1	0	0	0	0
Distichochlamys_sp	0	0	0	0	0	0	1	0	0	0	0	0
Alpinia_ligulata1	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_nieuwenhuizii	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_hookeriana	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_latilabris	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_guangdongensis	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_maclurei	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_rubricaulis	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_argentea	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_warburgii	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_tonkinensis	0	0	0	0	0	0	1	0	0	0	0	0
Alpinia_zerumbet	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_oxyphylla	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_brevilabris	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_polyantha	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_nutans	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_foxworthyi	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_rosea	0	0	0	0	0	0	0	0	0	0	1	0
Gagnepainia_harmandii	0	0	0	0	0	1	1	0	0	0	0	0
Curcuma_roscoeana	0	0	0	0	0	0	1	0	0	0	0	0
Hedychium_greenei	0	0	0	1	0	0	0	0	0	0	0	0
Hedychium_villosum	0	0	0	0	0	1	0	0	0	0	0	0
Zingiber_sulphureum	0	0	0	0	0	0	0	1	0	0	0	0
Zingiber_wrayii	0	0	0	0	0	0	1	0	0	0	0	0

Species												
Gagnepainia_thoreliana	0	0	0	0	0	0	1	0	0	0	0	0
Hemiorchis_burmanica	0	0	0	0	0	0	1	0	0	0	0	0
Hemiorchis_rhodorrhachis	0	0	0	0	0	0	1	0	0	0	0	0
Alpinia_sibuyanensis	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_vulcanica	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_flabellata	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_japonica	0	0	0	0	0	1	1	0	0	0	0	0
Alpinia_pricei	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_mutica	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_coriacea	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_rugosa	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_calcarata	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_shimadae	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_pumila	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_formosana	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_intermedia	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_bilamellata	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_nigra	0	0	0	0	0	0	1	0	0	0	0	0
Alpinia_coeruleoviridis	0	0	0	0	0	0	0	0	1	0	0	0
*Camptandra_parvula*1	0	0	0	0	0	0	1	0	0	0	0	0
Alpinia_monopleura	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_cylindrocephala	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_aenea	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_carolinensis	0	0	0	0	0	0	0	0	1	0	0	0
Alpinia_eremochlamys	0	0	0	0	0	0	0	0	1	0	0	0
Renealmia_cernua	1	0	0	0	0	0	0	0	0	0	0	0
*Renealmia_alpinia*1	1	0	0	0	0	0	0	0	0	0	0	0
Renealmia_thyrsoidea	1	0	0	0	0	0	0	0	0	0	0	0
Burbidgea_nitida	0	0	0	0	0	0	0	1	0	0	0	0
*Burbidgea_schizocheila*1	0	0	0	0	0	0	0	1	0	0	0	0
Riedelia_sp	0	0	0	0	0	0	0	0	0	1	0	0
*Pleuranthodium_trichocalyx*1	0	0	0	0	0	0	0	0	0	1	0	0
Pleuranthodium_floccosum	0	0	0	0	0	0	0	0	0	1	0	0

Species												
*Pleuranthodium_schlechteri*1	0	0	0	0	0	0	0	0	0	1	0	0
Leptosolena_haenkei	0	0	0	0	0	0	0	0	0	0	1	0
Amomum_mentawaiense	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_cerasinum	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_aculeatum	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_pierreanum*2	0	0	0	0	0	0	1	0	0	0	0	0
Amomum_coriandriodorum	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_paratsaoko	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_aff._paratsaoko	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_xanthophlebium	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_lappaceum*1	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_pinetorum	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_oceanica	0	0	0	0	0	0	0	0	0	1	0	0
*Amomum_centrocephalum*1	0	0	0	0	0	0	0	1	0	0	0	0
Hornstedtia_hainanensis	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_staminidivum	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_quadratolaminare	0	0	0	0	0	1	0	0	0	0	0	0
*Elettariopsis_sp*5	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_krervanh*2	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_krervanh*3	0	0	0	0	0	0	0	0	1	0	0	0
Amomum_villosum	0	0	0	0	0	1	0	0	0	0	0	0
*Amomum_sp*10	0	0	0	0	0	0	0	0	1	0	0	0
Siliquamomum_tonkinense	0	0	0	0	0	1	0	0	0	0	0	0
Alpinia_rafflesiana	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_uliginosum*2	0	0	0	0	0	0	1	0	0	0	0	0
*Amomum_compactum*3	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_purpurata	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_arctiflora	0	0	0	0	0	0	0	0	0	0	0	1
Alpinia_caerulea	0	0	0	0	0	0	0	0	0	0	0	1
Alpinia_vittata	0	0	0	0	0	0	0	0	0	1	0	0
Alpinia_luteocarpa	0	0	0	0	0	0	0	0	0	0	1	0
Amomum_laxesquamosum	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_oligophyllum	0	0	0	0	0	0	0	1	0	0	0	0

Species												
*Amomum_hastilabium*2	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_hastilabium*1	0	0	0	0	0	0	0	0	1	0	0	0
*Amomum_sp*9	0	0	0	0	0	0	0	1	0	0	0	0
*Amomum_compactum*2	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_pseudofoetens	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_elegans	0	0	0	0	0	0	0	0	0	0	1	0
Alpinia_modesta	0	0	0	0	0	0	0	0	0	0	0	1
Alpinia_arundelliana	0	0	0	0	0	0	0	0	0	0	0	1
*Geocharis_sp*1	0	0	0	0	0	0	0	1	0	0	0	0
Vanoverberghia_sepulchrei	0	0	0	0	0	0	0	0	0	0	1	0
*Etlingera_elatior*1	0	0	0	0	0	0	0	1	0	0	0	0
Etlingera_triorgyalis	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_propinquum	0	0	0	0	0	0	0	0	0	0	1	0
Amomum_koenigii	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_yunnanense	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_sericeum	0	0	0	0	0	1	0	0	0	0	0	0
Elettariopsis_smithiae	0	0	0	0	0	0	1	0	0	0	0	0
Amomum_menglaense	0	0	0	0	0	1	0	0	0	0	0	0
*Amomum_uliginosum*1	0	0	0	0	0	0	0	1	0	0	0	0
Elettariopsis_sp	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_putrescens	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_queenslandicum	0	0	0	0	0	0	0	0	0	0	0	1
Amomum_austrosinense	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_purpureorubrum	0	0	0	0	0	1	0	0	0	0	0	0
Amomum_maximum	0	0	0	0	0	1	0	0	0	0	0	0
*Amomum_glabrum*2	0	0	0	0	0	1	0	0	0	0	0	0
Aframomum_angustifolium	0	0	1	0	0	0	0	0	0	0	0	0
*Alpinia_galanga*2	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_conchigera	0	0	0	0	0	1	0	0	0	0	0	0
Elettariopsis_kerbyi	0	0	0	0	0	0	0	1	0	0	0	0
Elettariopsis_stenosiphon	0	0	0	0	0	0	0	1	0	0	0	0
Amomum_petaloideum	0	0	0	0	0	1	1	0	0	0	0	0
Zingiber_cylindricum	0	0	0	0	1	0	0	0	0	0	0	0

Amomum_pterocarpum	0	0	0	1	1	0	0	0	0	0	0	0
Zingiber_wightianum	0	0	0	1	1	0	0	0	0	0	0	0
*Amomum_sp*1	0	0	0	0	1	0	0	0	0	0	0	0
Elettaria_sp	0	0	0	1	1	0	0	0	0	0	0	0
Curcuma_zedoaria	0	0	0	1	1	0	0	0	0	0	0	0
Amomum_fulviceps	0	0	0	1	1	0	0	0	0	0	0	0
*Amomum_sp*2	0	0	0	0	1	0	0	0	0	0	0	0
Curcuma_albiflora	0	0	0	0	1	0	0	0	0	0	0	0
Amomum_graminifolium	0	0	0	0	1	0	0	0	0	0	0	0
*Amomum_sp*3	0	0	0	0	1	0	0	0	0	0	0	0
Amomum_nemorale	0	0	0	0	1	0	0	0	0	0	0	0
Cyphostigma_pulchellum	0	0	0	0	1	0	0	0	0	0	0	0
Amomum_acuminatum	0	0	0	0	1	0	0	0	0	0	0	0
Amomum_masticatorium	0	0	0	1	1	0	0	0	0	0	0	0
*Alpinia_abundiflora*1	0	0	0	1	1	0	0	0	0	0	0	0
*Amomum_sp*5	0	0	0	0	1	0	0	0	0	0	0	0
*Amomum_sp*6	0	0	0	0	1	0	0	0	0	0	0	0
*Amomum_sp*7	0	0	0	0	1	0	0	0	0	0	0	0
Alpinia_sp	0	0	0	0	1	0	0	0	0	0	0	0
Aframomum_daniellii	0	0	1	0	0	0	0	0	0	0	0	0
Aframomum_sceptrum	0	0	1	0	0	0	0	0	0	0	0	0
Renealmia_battenbergiana	0	1	0	0	0	0	0	0	0	0	0	0
*Geocharis_sp*2	0	0	0	0	0	0	0	1	0	0	0	0
Alpinia_fax	0	0	0	0	1	0	0	0	0	0	0	0

Appendix 7:Example script used in Biogeobears in ancestral area reconstructions.

```
#########################################################
# This is an introductory example script for the
# R package "BioGeoBEARS" by Nick Matzke
#
# All scripts are copyright Nicholas J. Matzke,
# please cite if you use. License: GPL-3
# http://cran.r-project.org/web/licenses/GPL-3
#
# I am happy to answer questions at matzke@nimbios.org, but
# I am more happy to answer questions on the
# BioGeoBEARS google group
#
# The package is designed for ML and Bayesian inference
# of
#
# (a) ancestral geographic ranges, and
#
# (b) perhaps more importantly, models for the
#     evolution of geographic range across a phylogeny.
#
# The example below implements and compares:
#
# (1) The standard 2-parameter DEC model implemented in
#     the program LAGRANGE (Ree & Smith 2008); users will
#     notice that the ML parameter inference and log-
#     likelihoods are identical
#
# (2) A DEC+J model implemented in BioGeoBEARS, wherein
#     a third parameter, j, is added, representing the
#     relative per-event weight of founder-event / jump
#     speciation events at cladogenesis events.  The
#     higher j is, the more probability these events have,
#     and the less probability the standard LAGRANGE
```

```
#    cladogenesis events have.
#
# (3) Some standard model-testing (LRT and AIC) is
#    implemented at the end so that users may compare models
#
# (4) The script does similar tests of a DIVA-like model (Ronquist 1997)
#    and a BAYAREA-like model (Landis, Matzke, Moore, & Huelsenbeck, 2013)
#
#######################################################


#######################################################
# Installing BioGeoBEARS
#######################################################
# Uncomment this command to get everything
# Please use the "0-cloud" R repository at "http://cran.rstudio.com" as it is
# the only one that keeps download statistics
#######################################################
#
#
# # Install BioGeoBEARS from CRAN 0-cloud:
# install.packages("BioGeoBEARS", dependencies=TRUE, repos="http://cran.rstudio.com")
#
#######################################################


#######################################################
# SETUP -- libraries/BioGeoBEARS updates
#######################################################

# Load the package (after installation, see above).
library(optimx)        # You need to have some version of optimx available
                # as it is a BioGeoBEARS dependency; however, if you
                # don't want to use optimx, and use optim() (from R core)
                # you can set:
                # BioGeoBEARS_run_object$use_optimx = FALSE
                # ...everything should work either way -- NJM 2014-01-08
```

```
library(FD)        # for FD::maxent() (make sure this is up-to-date)
library(snow)      # (if you want to use multicore functionality; some systems/R versions prefer
library(parallel), try either)
library(parallel)
library(BioGeoBEARS)


##########################################################
# TO GET THE OPTIMX/OPTIM FIX, AND THE UPPASS FIX,
# SOURCE THE REVISED FUNCTIONS WITH THESE COMMANDS
#
# CRUCIAL CRUCIAL CRUCIAL:
# YOU HAVE TO RUN THE SOURCE COMMANDS AFTER
# *EVERY TIME* YOU DO library(BioGeoBEARS). THE CHANGES ARE NOT
"PERMANENT",
# THEY HAVE TO BE MADE EACH TIME.  IF YOU ARE GOING TO BE OFFLINE,
# YOU CAN DOWNLOAD EACH .R FILE TO YOUR HARD DRIVE AND REFER THE
source()
# COMMANDS TO THE FULL PATH AND FILENAME OF EACH FILE ON YOUR
# LOCAL SYSTEM INSTEAD.
##########################################################
library(BioGeoBEARS)
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_basics_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_generics_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_readwrite_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_classes_v1.R")
source("http://phylo.wdfiles.com/local--
files/biogeobears/BioGeoBEARS_calc_transition_matrices_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_detection_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_models_v1.R")
source("http://phylo.wdfiles.com/local--
files/biogeobears/BioGeoBEARS_extract_Qmat_COOmat_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_univ_model_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_stratified_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_plots_v1.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_simulate_v1.R")
```

```
source("http://phylo.wdfiles.com/local--files/biogeobears/calc_loglike_sp_v01.R")
source("http://phylo.wdfiles.com/local--files/biogeobears/calc_uppass_probs_v1.R")
source("http://phylo.wdfiles.com/local--
files/biogeobears/get_stratified_subbranch_top_downpass_likelihoods_v1.R")
calc_loglike_sp = compiler::cmpfun(calc_loglike_sp_prebyte)    # crucial to fix bug in uppass
calculations
calc_independent_likelihoods_on_each_branch =
compiler::cmpfun(calc_independent_likelihoods_on_each_branch_prebyte)
    # slight speedup hopefully


#######################################################
# SETUP: YOUR WORKING DIRECTORY
#######################################################
# You will need to set your working directory to match your local system

# Note these very handy functions!
# Command "setwd(x)" sets your working directory
# Command "getwd()" gets your working directory and tells you what it is.
# Command "list.files()" lists the files in your working directory
# To get help on any command, use "?".  E.g., "?list.files"

# Set your working directory for output files
# default here is your home directory ("~")
# Change this as you like
wd = np("C:/BGB ")
setwd(wd)

# Double-check your working directory with getwd()
getwd()


#######################################################
# SETUP: Extension data directory
#######################################################
# When R packages contain extra files, they are stored in the "extdata" directory
# inside the installed package.
```

```
#
# BioGeoBEARS contains various example files and scripts in its extdata directory.
#
# Each computer operating system might install BioGeoBEARS in a different place,
# depending on your OS and settings.
#
# However, you can find the extdata directory like this:
extdata_dir = np(system.file("extdata", package="BioGeoBEARS"))
extdata_dir
list.files(extdata_dir)

# "system.file" looks in the directory of a specified package (in this case BioGeoBEARS)
# The function "np" is just a shortcut for normalizePath(), which converts the
# path to the format appropriate for your system (e.g., Mac/Linux use "/", but
# Windows uses "\\", if memory serves).

# Even when using your own data files, you should KEEP these commands in your
# script, since the plot_BioGeoBEARS_results function needs a script from the
# extdata directory to calculate the positions of "corners" on the plot. This cannot
# be made into a straight up BioGeoBEARS function because it uses C routines
# from the package APE which do not pass R CMD check for some reason.

#########################################################
# SETUP: YOUR TREE FILE AND GEOGRAPHY FILE
#########################################################
# Example files are given below. To run your own data,
# make the below lines point to your own files, e.g.
# trfn = "/mydata/frogs/frogBGB/tree.newick"
# geogfn = "/mydata/frogs/frogBGB/geog.data"

#########################################################
# Phylogeny file
# Notes:
# 1. Must be binary/bifurcating: no polytomies
```

2. No negative branchlengths (e.g. BEAST MCC consensus trees sometimes have negative branchlengths)
3. Be careful of very short branches, as BioGeoBEARS will interpret ultrashort branches as direct ancestors
4. You can use non-ultrametric trees, but BioGeoBEARS will interpret any tips significantly below the
top of the tree as fossils! This is only a good idea if you actually do have fossils in your tree,
as in e.g. Wood, Matzke et al. (2013), Systematic Biology.
5. The default settings of BioGeoBEARS make sense for trees where the branchlengths are in units of
millions of years, and the tree is 1-1000 units tall. If you have a tree with a total height of
e.g. 0.00001, you will need to adjust e.g. the max values of d and e, or (simpler) multiply all
your branchlengths to get them into reasonable units.
6. DON'T USE SPACES IN SPECIES NAMES, USE E.G. "_"
###
# This is the example Newick file for Hawaiian Psychotria
# (from Ree & Smith 2008)
# "trfn" = "example.newick"
trfn = "example.newick"

# Look at the raw Newick file:
moref(trfn)

# Look at your phylogeny:
tr = read.tree(trfn)
tr
plot(tr)
title("Example Psychotria phylogeny from Ree & Smith (2008)")
axisPhylo() # plots timescale

###
Geography file
Notes:
1. This is a PHLYIP-formatted file. This means that in the

first line,
- the 1st number equals the number of rows (species)
- the 2nd number equals the number of columns (number of areas)
2. This is the same format used for C++ LAGRANGE geography files.
3. All names in the geography file must match names in the phylogeny file.
4. DON'T USE SPACES IN SPECIES NAMES, USE E.G. "_"
5. Operational taxonomic units (OTUs) should ideally be phylogenetic lineages,
i.e. genetically isolated populations. These may or may not be identical
with species. You would NOT want to just use specimens, as each specimen
automatically can only live in 1 area, which will typically favor DEC+J
models. This is fine if the species/lineages really do live in single areas,
but you wouldn't want to assume this without thinking about it at least.
In summary, you should collapse multiple specimens into species/lineages if
data indicates they are the same genetic population.
###

# This is the example geography file for Hawaiian Psychotria
# (from Ree & Smith 2008)
geogfn = "example.txt"

# Look at the raw geography text file:
moref(geogfn)

# Look at your geographic range data:
tipranges = getranges_from_LagrangePHYLIP(lgdata_fn=geogfn)
tipranges

# Set the maximum number of areas any species may occupy; this cannot be larger
# than the number of areas you set up, but it can be smaller.
max_range_size = 2

###
###
KEY HINT: The number of states (= number of different possible geographic ranges)
depends on (a) the number of areas and (b) max_range_size.

```
# If you have more than about 500-600 states, the calculations will get REALLY slow,
# since the program has to exponentiate a matrix of e.g. 600x600.  Often the computer
# will just sit there and crunch, and never get through the calculation of the first
# likelihood.
#
# (this is also what is usually happening when LAGRANGE hangs: you have too many states!)
#
# To check the number of states for a given number of ranges, try:
numstates_from_numareas(numareas=4, maxareas=4, include_null_range=TRUE)
numstates_from_numareas(numareas=4, maxareas=4, include_null_range=FALSE)
numstates_from_numareas(numareas=4, maxareas=3, include_null_range=TRUE)
numstates_from_numareas(numareas=4, maxareas=2, include_null_range=TRUE)

# Large numbers of areas have problems:
numstates_from_numareas(numareas=10, maxareas=10, include_null_range=TRUE)

# ...unless you limit the max_range_size:
numstates_from_numareas(numareas=10, maxareas=2, include_null_range=TRUE)
#######################################################
#######################################################


#########################################################
# DEC AND DEC+J ANALYSIS
#########################################################


#########################################################
# Run DEC
#########################################################

BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't
much faster at this scale
BioGeoBEARS_run_object$speedup=TRUE        # shorcuts to speed ML search; use FALSE
if worried (e.g. >3 params)
BioGeoBEARS_run_object$use_optimx = TRUE
```

BioGeoBEARS_run_object$calc_ancprobs=TRUE   # get ancestral states from optim run

Set up a time-stratified analysis
(un-comment to use; see example files in extdata_dir,
and BioGeoBEARS google group posts for further hints)
#BioGeoBEARS_run_object$timesfn = "timeperiods.txt"
#BioGeoBEARS_run_object$dispersal_multipliers_fn = "manual_dispersal_multipliers.txt"
#BioGeoBEARS_run_object$areas_allowed_fn = "areas_allowed.txt"
#BioGeoBEARS_run_object$areas_adjacency_fn = "areas_adjacency.txt"
#BioGeoBEARS_run_object$distsfn = "distances_matrix.txt"
See notes on the distances model on PhyloWiki's BioGeoBEARS updates page.

Input the maximum range size
BioGeoBEARS_run_object$max_range_size = max_range_size

Multicore processing if desired
BioGeoBEARS_run_object$num_cores_to_use=1
(use more cores to speed it up; this requires
library(parallel) and/or library(snow). Parellel,
is default on Macs in R 3.0+, but apparently still
has to be typed on Windows machines. Note: apparently
parallel works on Mac command-line R, but not R.app.
BioGeoBEARS checks for this and resets to 1
core with R.app)

Sparse matrix exponentiation is an option for huge numbers of ranges/states (600+)
I have experimented with sparse matrix exponentiation in EXPOKIT/rexpokit,
but the results are imprecise and so I haven't explored it further.
In a Bayesian analysis, it might work OK, but the ML point estimates are
not identical.
Also, I have not implemented all functions to work with force_sparse=TRUE.
Volunteers are welcome to work on it!!
BioGeoBEARS_run_object$force_sparse=FALSE

Give BioGeoBEARS the location of the geography text file

```
BioGeoBEARS_run_object$geogfn = geogfn

# Give BioGeoBEARS the location of the phylogeny Newick file
BioGeoBEARS_run_object$trfn = trfn

# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

# Divide the tree up by timeperiods/strata (uncomment this for stratified analysis)
#BioGeoBEARS_run_object = section_the_tree(inputs=BioGeoBEARS_run_object,
make_master_table=TRUE, plot_pieces=FALSE)
# The stratified tree is described in this table:
#BioGeoBEARS_run_object$master_table

# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE

# Set up DEC model
# (nothing to do; defaults)

# Look at the BioGeoBEARS_run_object; it's just a list of settings etc.
BioGeoBEARS_run_object

# This contains the model object
BioGeoBEARS_run_object$BioGeoBEARS_model_object

# This table contains the parameters of the model
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table

# Run this to check inputs. Read the error messages if you get them!
check_BioGeoBEARS_run(BioGeoBEARS_run_object)
```

```
# For a slow analysis, run once, then set runslow=FALSE to just
# load the saved result.
runslow = TRUE
resfn = "Psychotria_DEC_M0_unconstrained_v1.Rdata"
if (runslow)
    {
res = bears_optim_run(BioGeoBEARS_run_object)
res

save(res, file=resfn)
resDEC = res
    } else {
    # Loads to "res"
load(resfn)
resDEC = res
    }


#######################################################
# Run DEC+J
#######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size

# Set up the stratified part
#BioGeoBEARS_run_object$timesfn = "timeperiods.txt"
#BioGeoBEARS_run_object$dispersal_multipliers_fn = "manual_dispersal_multipliers.txt"
#BioGeoBEARS_run_object$areas_allowed_fn = "areas_allowed.txt"
#BioGeoBEARS_run_object$areas_adjacency_fn = "areas_adjacency.txt"
#BioGeoBEARS_run_object$distsfn = "distances_matrix.txt"
# See notes on the distances model on PhyloWiki's BioGeoBEARS updates page.
```

BioGeoBEARS_run_object$speedup=TRUE   # shorcuts to speed ML search; use FALSE if worried (e.g. >3 params)

BioGeoBEARS_run_object$use_optimx = TRUE

BioGeoBEARS_run_object$num_cores_to_use=1

BioGeoBEARS_run_object$force_sparse=FALSE   # sparse=FALSE causes pathology & isn't much faster at this scale

BioGeoBEARS_run_object$calc_ancprobs=TRUE   # get ancestral states from optim run

# This function loads the dispersal multiplier matrix etc. from the text files into the model object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)

Divide the tree up by timeperiods/strata (uncomment this for stratified analysis)
#BioGeoBEARS_run_object = section_the_tree(inputs=BioGeoBEARS_run_object, make_master_table=TRUE, plot_pieces=FALSE)
The stratified tree is described in this table:
#BioGeoBEARS_run_object$master_table

Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE

# Set up DEC+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resDEC$outputs@params_table["d","est"]
estart = resDEC$outputs@params_table["e","est"]
jstart = 0.0001

Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart

# Add j as a free parameter
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

resfn = "Psychotria_DEC+J_M0_unconstrained_v1.Rdata"
runslow = TRUE
if (runslow)
  {
  #sourceall("/Dropbox/_njm/__packages/BioGeoBEARS_setup/")

res = bears_optim_run(BioGeoBEARS_run_object)
res

save(res, file=resfn)

resDECj = res
  } else {
  # Loads to "res"
load(resfn)
resDECj = res
  }

#########################################################
# PDF plots
#########################################################
pdffn = "Psychotria_DEC_vs_DEC+J_M0_unconstrained_v1.pdf"
pdf(pdffn, width=34, height=44)

#########################################################
# Plot ancestral states - DEC
```

```
###########################################################
analysis_titletxt ="BioGeoBEARS DEC on Psychotria M0_unconstrained"

# Setup
results_object = resDEC
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res2 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.2, statecex=0.1, splitcex=0.1, titlecex=0.2,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)


###########################################################
# Plot ancestral states - DECJ
###########################################################
analysis_titletxt ="BioGeoBEARS DEC+J on Psychotria M0_unconstrained"

# Setup
results_object = resDECj
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

# Pie chart
```

plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.2, statecex=0.1, splitcex=0.1, titlecex=0.2,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

dev.off() # Turn off PDF
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr) # Plot it

###
DIVALIKE AND DIVALIKE+J ANALYSIS
###

NOTE: The BioGeoBEARS "DIVA" model is not identical with
Ronquist (1997)'s parsimony DIVA. It is a likelihood
interpretation of DIVA, constructed by modelling DIVA's
processes the way DEC does, but only allowing the
processes DIVA allows (widespread vicariance: yes; subset
sympatry: no; see Ronquist & Sanmartin 2011, Figure 4).
#
I thus now call the model "DIVALIKE", and you should also. ;-)
#

###
Run DIVALIKE
###
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size

Set up the stratified part
#BioGeoBEARS_run_object$timesfn = "timeperiods.txt"
#BioGeoBEARS_run_object$dispersal_multipliers_fn = "manual_dispersal_multipliers.txt"
#BioGeoBEARS_run_object$areas_allowed_fn = "areas_allowed.txt"

```
#BioGeoBEARS_run_object$areas_adjacency_fn = "areas_adjacency.txt"
#BioGeoBEARS_run_object$distsfn = "distances_matrix.txt"
# See notes on the distances model on PhyloWiki's BioGeoBEARS updates page.


BioGeoBEARS_run_object$use_optimx = TRUE
BioGeoBEARS_run_object$num_cores_to_use=1


BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't
much faster at this scale
BioGeoBEARS_run_object$speedup=TRUE        # shorcuts to speed ML search; use FALSE
if worried (e.g. >3 params)
BioGeoBEARS_run_object$calc_ancprobs=TRUE    # get ancestral states from optim run


# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)


# Divide the tree up by timeperiods/strata (uncomment this for stratified analysis)
#BioGeoBEARS_run_object = section_the_tree(inputs=BioGeoBEARS_run_object,
make_master_table=TRUE, plot_pieces=FALSE)
# The stratified tree is described in this table:
#BioGeoBEARS_run_object$master_table


# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE


# Set up DIVALIKE model
# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] =
"fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-
j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] =
"ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] =
"ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] =
"ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] =
"fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] =
0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] =
0.5

# No jump dispersal/founder-event speciation
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] =
"free"
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = 0.01
# BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = 0.01

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

runslow = TRUE
resfn = "Psychotria_DIVALIKE_M0_unconstrained_v1.Rdata"
if (runslow)
    {
res = bears_optim_run(BioGeoBEARS_run_object)
res

save(res, file=resfn)
resDIVALIKE = res
```

```r
    } else {
    # Loads to "res"
load(resfn)
resDIVALIKE = res
    }



#######################################################
# Run DIVALIKE+J
#######################################################
BioGeoBEARS_run_object = define_BioGeoBEARS_run()
BioGeoBEARS_run_object$trfn = trfn
BioGeoBEARS_run_object$geogfn = geogfn
BioGeoBEARS_run_object$max_range_size = max_range_size


# Set up the stratified part
#BioGeoBEARS_run_object$timesfn = "timeperiods.txt"
#BioGeoBEARS_run_object$dispersal_multipliers_fn = "manual_dispersal_multipliers.txt"
#BioGeoBEARS_run_object$areas_allowed_fn = "areas_allowed.txt"
#BioGeoBEARS_run_object$areas_adjacency_fn = "areas_adjacency.txt"
#BioGeoBEARS_run_object$distsfn = "distances_matrix.txt"
# See notes on the distances model on PhyloWiki's BioGeoBEARS updates page.


BioGeoBEARS_run_object$use_optimx = TRUE
BioGeoBEARS_run_object$num_cores_to_use=1


BioGeoBEARS_run_object$force_sparse=FALSE    # sparse=FALSE causes pathology & isn't
much faster at this scale
BioGeoBEARS_run_object$speedup=TRUE          # shorcuts to speed ML search; use FALSE
if worried (e.g. >3 params)
BioGeoBEARS_run_object$calc_ancprobs=TRUE    # get ancestral states from optim run


# This function loads the dispersal multiplier matrix etc. from the text files into the model
object. Required for these to work!
# (It also runs some checks on these inputs for certain errors.)
BioGeoBEARS_run_object = readfiles_BioGeoBEARS_run(BioGeoBEARS_run_object)
```

```
# Divide the tree up by timeperiods/strata (uncomment this for stratified analysis)
#BioGeoBEARS_run_object = section_the_tree(inputs=BioGeoBEARS_run_object,
make_master_table=TRUE, plot_pieces=FALSE)
# The stratified tree is described in this table:
#BioGeoBEARS_run_object$master_table


# Good default settings to get ancestral states
BioGeoBEARS_run_object$return_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_TTL_loglike_from_condlikes_table = TRUE
BioGeoBEARS_run_object$calc_ancprobs = TRUE


# Set up DIVALIKE+J model
# Get the ML parameter values from the 2-parameter nested model
# (this will ensure that the 3-parameter model always does at least as good)
dstart = resDIVALIKE$outputs@params_table["d","est"]
estart = resDIVALIKE$outputs@params_table["e","est"]
jstart = 0.0001


# Input starting values for d, e
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","init"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["d","est"] = dstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","init"] = estart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["e","est"] = estart


# Remove subset-sympatry
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","type"] =
"fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","init"] = 0.0
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["s","est"] = 0.0


BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ysv","type"] = "2-
j"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["ys","type"] =
"ysv*1/2"
```

```
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["y","type"] =
"ysv*1/2"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["v","type"] =
"ysv*1/2"

# Allow classic, widespread vicariance; all events equiprobable
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","type"] =
"fixed"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","init"] =
0.5
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["mx01v","est"] =
0.5

# Add jump dispersal/founder-event speciation
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","type"] = "free"
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","init"] = jstart
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","est"] = jstart

# Under DIVALIKE+J, the max of "j" should be 2, not 3 (as is default in DEC+J)
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","min"] =
0.00001
BioGeoBEARS_run_object$BioGeoBEARS_model_object@params_table["j","max"] =
1.99999

check_BioGeoBEARS_run(BioGeoBEARS_run_object)

resfn = "Psychotria_DIVALIKE+J_M0_unconstrained_v1.Rdata"
runslow = TRUE
if (runslow)
  {
  #sourceall("/Dropbox/_njm/__packages/BioGeoBEARS_setup/")

res = bears_optim_run(BioGeoBEARS_run_object)
res
```

```
save(res, file=resfn)

resDIVALIKEj = res
    } else {
    # Loads to "res"
load(resfn)
resDIVALIKEj = res
    }


pdffn = "Psychotria_DIVALIKE_vs_DIVALIKE+J_M0_unconstrained_v1.pdf"
pdf(pdffn, width=24, height=24)


#######################################################
# Plot ancestral states - DIVALIKE
#######################################################
analysis_titletxt ="BioGeoBEARS DIVALIKE on Psychotria M0_unconstrained"


# Setup
results_object = resDIVALIKE
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))


# States
res2 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)


# Pie chart
plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.2, statecex=0.1, splitcex=0.1, titlecex=0.2,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)


#######################################################
# Plot ancestral states - DIVALIKE+J
```

```
############################################################
analysis_titletxt ="BioGeoBEARS DIVALIKE+J on Psychotria M0_unconstrained"

# Setup
results_object = resDIVALIKEj
scriptdir = np(system.file("extdata/a_scripts", package="BioGeoBEARS"))

# States
res1 = plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="text", label.offset=0.45, tipcex=0.7, statecex=0.7, splitcex=0.6, titlecex=0.8,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

# Pie chart
plot_BioGeoBEARS_results(results_object, analysis_titletxt, addl_params=list("j"),
plotwhat="pie", label.offset=0.45, tipcex=0.2, statecex=0.1, splitcex=0.1, titlecex=0.2,
plotsplits=TRUE, cornercoords_loc=scriptdir, include_null_range=TRUE, tr=tr,
tipranges=tipranges)

dev.off()
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr)



############################################################
# SETUP: YOUR TREE FILE AND GEOGRAPHY FILE
############################################################
# Setup
library(ape)
library(BioGeoBEARS)    # for get_lagrange_nodenums(),
postorder_nodes_phylo4_return_table()
library(phylobase)      # required for postorder_nodes_phylo4_return_table()

# Fix to postorder_nodes_phylo4_return_table
source("http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_readwrite_v1.R")
```

```
# Set your working directory to put the PDF in (pick your own)
# Default
getwd()
# Example files are given below. To run your own data,
# make the below lines point to your own files, e.g.
trfn = "C:/BGB/example.newick"
geogfn = "C:/BGB/example.txt"


# Look at your phylogeny:
tr = read.tree(trfn)
tr
plot(tr, cex=0.5)
title("Example_BGB")
axisPhylo() # plots timescale


#######################################################
# Plot APE/BioGeoBEARS node numbers
#######################################################
ntips = length(tr$tip.label)
Rnodenums = (ntips+1):(ntips+tr$Nnode)
tipnums = 1:ntips
plot(tr, label.offset=0.25, cex=1.25)
axisPhylo()
tiplabels(cex=1.5)
nodelabels(text=Rnodenums, node=Rnodenums, cex=0.5)
title("APE/BioGeoBEARS node numbers")


# END PDF
dev.off()
cmdstr = paste("open ", pdffn, sep="")
system(cmdstr)


prt(tr)


names(res)
```

[1] "computed_likelihoods_at_each_node"

[2] "relative_probs_of_each_state_at_branch_top_AT_node_DOWNPASS"

 [3] "condlikes_of_each_state"

 [4] "relative_probs_of_each_state_at_branch_bottom_below_node_DOWNPASS"

[5] "relative_probs_of_each_state_at_branch_bottom_below_node_UPPASS"

 [6] "relative_probs_of_each_state_at_branch_top_AT_node_UPPASS"

 [7] "ML_marginal_prob_each_state_at_branch_bottom_below_node"

 [8] "ML_marginal_prob_each_state_at_branch_top_AT_node"

 [9] "relative_probs_of_each_state_at_bottom_of_root_branch"

 [10] "total_loglikelihood"

[11] "inputs"

[12] "outputs"

 [13] "optim_result"

res$computed_likelihoods_at_each_node
write.table(res$ computed_likelihoods_at_each_node, file="sap1.txt", sep="XTABX")

res$relative_probs_of_each_state_at_branch_top_AT_node_DOWNPASS
write.table(res$relative_probs_of_each_state_at_branch_top_AT_node_DOWNPASS
, file="sap2.txt", sep="XTABX")

res$condlikes_of_each_state
write.table(res$condlikes_of_each_state, file="sap3.txt", sep="XTABX")

res$relative_probs_of_each_state_at_branch_bottom_below_node_DOWNPASS
write.table(res$relative_probs_of_each_state_at_branch_bottom_below_node_DOWNPASS
, file="sap4.txt", sep="XTABX")

res$relative_probs_of_each_state_at_branch_bottom_below_node_UPPASS

```
write.table(res$relative_probs_of_each_state_at_branch_bottom_below_node_UPPASS,
file="sap5.txt", sep="XTABX")

res$relative_probs_of_each_state_at_branch_top_AT_node_UPPASS
write.table(res$relative_probs_of_each_state_at_branch_top_AT_node_UPPASS
, file="sap6.txt", sep="XTABX")

res$ML_marginal_prob_each_state_at_branch_bottom_below_node
write.table(res$ML_marginal_prob_each_state_at_branch_bottom_below_node, file="sap7.txt",
sep="XTABX")

res$ML_marginal_prob_each_state_at_branch_top_AT_node
write.table(res$ML_marginal_prob_each_state_at_branch_top_AT_node
, file="sap8.txt", sep="XTABX")

res$relative_probs_of_each_state_at_bottom_of_root_branch
write.table(res$relative_probs_of_each_state_at_bottom_of_root_branch
, file="sap9.txt", sep="XTABX")

res$total_loglikelihood
write.table(res$ML_marginal_prob_each_state_at_branch_top_AT_node, file="B.txt",
sep="XTABX")

res$inputs
write.table(res$ML_marginal_prob_each_state_at_branch_top_AT_node, file="B.txt",
sep="XTABX")

res$outputs
write.table(res$ML_marginal_prob_each_state_at_branch_top_AT_node, file="B.txt",
sep="XTABX")

res$optim_result
write.table(res$ML_marginal_prob_each_state_at_branch_top_AT_node, file="B.txt",
sep="XTABX")
```

Appendix 8: DNA sequence alignments and Biogeobears result output for Begoniaceae, Sapotaceae, Zingiberaceae

Files on the accompanying DVD comprise DNA sequence data and sequence alignments in FASTA format and Biogeobears ancestral area probabilities for each family in excel format

File	Contents
Chapter 3 Reference alignmet 179 taxa Begoniaceae_cpDNA.fas	Alignment of concatenated ndhA intron, ndhF-rpl32, rpl32-trnL sequences of 179 taxa in family Begoniaceae
Chapter 4 Reference alignmet 163 taxa Sapotaceae_ITS.fas	Alignment of ITS sequences of 163 taxa in family Sapotaceae
Chapter 4 Reference alignmet 136 taxa Sapotaceae_cpDNA.fas	Alignment of concatenated trnH-psbA spacer, trnC-trnD, trnC-psbM, psbM-trnD, ndhF sequences of 136 taxa in family Sapotaceae
Chapter 5 Reference alignmet 151 taxa Zingiberaceae_ITS.fas	Alignment of ITS sequences of 151 taxa in family Zingiberaceae
Chapter 5 Reference alignmet 151 taxa Zingiberaceae_cpDNA.fas	Alignment of ITS sequences of 151 taxa in family Zingiberaceae
Chapter 3 Ancestral area probabilities_Begoniaceae	Ancestral area probabilities for each state resulted in the biogeographic analysis of Begoniaceae in Biogeobears
Chapter 4 Ancestral area probabilities_Sapotaceae	Ancestral area probabilities for each state resulted in the biogeographic analysis of Sapotaceae in Biogeobears
Chapter 5 Ancestral area probabilities_Zingiberaceae	Ancestral area probabilities for each state resulted in the biogeographic analysis of Zingiberaceae in Biogeobears

.