DNA Barcoding of Flowering Plants in Jambi, Indonesia

Dissertation

Submitted in partial fulfillment of requirements for the degree of
Doctor of Philosophy (Ph.D.)
at Forest Genetics and Forest Tree Breeding Department
Faculty of Forest Sciences and Forest Ecology
Geörg-August Universität of Göttingen

By
Fitri Yola Amandita
Born in Jakarta, Indonesia

Göttingen 2015

Supervisor	: Prof. Dr. Reiner Finkeldey
Referee	: Prof. Dr. Reiner Finkeldey
Co-referee	: Prof. Dr. Holger Kreft
Date of disputation	: 26 February 2016

Acknowledgements

I would first like to express my deepest gratitude to Prof. Dr. Reiner Finkeldey for accepting me as his Ph.D. student, for his support, helpful advice and patient guidance throughout my study. I have been extremely lucky to have a supervisor who always cared about my problems and made easy for me to finish my study. I would also thank Prof. Dr. Holger Kreft for his co-supervision and reviews on my thesis. Special thanks to Prof. Dr. Elvira Hörandl, my third supervisor, for her kindly support. Sincere thanks to Prof. Martin Ziehe, my fourth examiner, for his constant support during my study.

I would present my deep appreciation to Dr. Katja Rembold for her friendship, continuous advice and support throughout my research. My great thankfulness goes to Dr. Barbara Vornam for her guidance on the data analyzes. I would also acknowledge Prof. Dr. Konstantin Krutovsky, Prof. Dr. Oliver Gailing, and Prof. Dr. Elizabeth Gillet for their comments on my project presentations to improve my understanding of my research.

My sincere thanks are to Alexandra Dolynska and Melanie Schmitt for their enormous help on the DNA analyzes. Many thanks to Regina Berkeley for her kind support in administration matters whenever I need. I would also like to acknowledge Gerold Dinkel and Marco Winkler for their help and other supports.

I am very grateful to Natalie Braidenbach for her warm friendship and sharing that make my study life memorable. Special thanks to Johanna Ropertz for the valuable help in translating the summary of this thesis. Many thanks to Dr. Markus Müller and Dr. Oleksandra Kuchma for their kindness, helps and supports during my study at the Institute of Forest Genetics and Forest Tree Breeding. Completing this work would have been more difficult without the support and friendship provided by the other members of the Institute: Nurlan Torokeldiev, Laura Cuervo, Devrim Cuming, and Sinan Hagenah.

I also gratefully acknowledge the CRC 990 (EFForTS) team members in Germany and Indonesia for making my research possible. I would also like to acknowledge The German Research

Foundation (DFG) for funding my research. I particularly thank Ministry of Environment and Forestry in Indonesia for their constant supports during my study. Very special thanks to Hardianto Mangopo and Fajar Adityarama for their assistances during the field work. I would also need to express my gratitude to Prof. Dr. Iskandar Z. Siregar and Dr. Sri Rahayu as counterparts in my project. Special thanks to Dr. Nicole Brinkmann for her prompt coordination. I thank Dr. Essy Harnelly who helped me understanding my research topic at the beginning of my study.

My sincerest thanks to Dr. Camille Couralet for her kind friendship and willingness to proof-read this dissertation. I also appreciate the Indonesian community in Göttingen for accepting me as a part of the big family. Many thanks are to Herr and Frau Adams for their hospitalities during my stay in Göttingen.

My countless gratitude to my mother for her endless love, and to my brothers and in-law sisters for their caring. Finally, I would like to express my profound gratitude to my greatest motivators: Dede and Sam.

Zusammenfassung

Der Verlust der Wälder und seine damit verbundene Degradierung ist ein schnell fortschreitender Prozess, der vor allen in den tropischen Regionen stattfindet. Sumatra als Beispiel, hat einen Großteil seiner Waldflächen vor allem durch Palmöl- und Kautschukplantagen umgewandelt.

Die Diversität von tropischen Blütenpflanzen ist auf Sumatra sehr hoch und noch nicht vollständig erforscht. Im Rahmen des Naturschutzes und zum Erhalt der Biodiversität ,sollte die Erforschung von tropischen Tieflandregenwäldern, wie sie in Sumatra vorkommen, prioritär sein. Jedoch ist die Bestimmung von vielen verschiedenen Pflanzenarten eine sehr große Herausforderung und erfordert neue Methoden der Artenerkennung.

Die erste erfolgreiche Anwendung an Blütenpflanzen mittels DNA Barcoding in 2003 ermögliche erstmals eine schnelle Erkennung von Arten. DNA Barcoding entwickelte sich dann recht schnell zu einer neuen Methode in der Wissenschaft der Biologie. Die Anwendung dieser Methode, einschließlich in der Erforschung der Pflanzenwelt, weit verbreitet. In der Praxis zeigte sich jedoch, dass DNA Barcoding von Pflanzen, an seine Grenzen kam, da das Fehlen von sogenannten universal Genen, die Bestimmung eines großen Artspektrum, nicht möglich machte. Ungeachtet der Diskussionen darüber welche geeigneten universal Gene, wurden matK und rbcL als Haupt Barcodes für die Erkennung der diverser Pflanzen, ausgesucht.

Die vorliegende Studie setzte sich zum Ziel, DNA Barcodes zu entwickeln, die die Untersuchung von Blütenpflanzen besser ermöglicht. Für die Entwicklung der DNA Barcodes wurden sogenannte Core Barcodes von Pflanzen (matK und rbcL) benutzt. Diese beiden Barcodes wurden aufgrund ihrer universellen Fähigkeit Pflanzen verschiedener Taxa zu identifizieren, ausgesucht und ausgewertet. Die Untersuchung dieser Marker wird zu der weiteren Entwicklung von nützlichen Informationen von Barcodes führen, um tropische Pflanzen besser zu identifizieren zu können.

Die Untersuchungsflächen liegen in der Provinz von Jambi, im Zentrum von Sumatra. Verschiedene Exemplare tropischer Blütenpflanzen wurden gesammelt. Die insgesamt 32 Studienflächen liegen in unterschiedlichen Landschaften. Zum einen, Bukit Duabelas National Park und zum anderen, der Harapan Regenwald. Getrocknete Pflanzenteile und Herbariumbelege wurden von 5.100 Proben gesammelt und angelegt. Morphologische Merkmale wurden gemeinsam mit Taxonomen untersucht. Insgesamt wurden mehr als 1.100 Arten von 436 Gattungen, 136 Familien und 49 Ordnungen, identifiziert.

Aufgrund verschiedener Limitierungen, wurden nur die Hälfte der Proben analysiert. Die DNA Amplifikationen und Sequenzierungen der untersuchten Proben, wurden auf Basis der matK and rbcL marker durchgeführt. Im Ergebnis, wurden 3.500 Barcodes von 500 tropischen Blütenpflanzen erzeugt.

Von jedem erfolgreich erzeugten Barcode wurden dann die Sequenzen bearbeitet um die maximale Qualität zu garantieren. Die molekulare Identifizierung wurde durch gezieltes nachforschen der Nukleotidsequenzinformation, auf Datenbanken wie GenBank and BOLD, ergänzt. Die Ergebnisse der morphologischen und molekularen Identifikationen wurden miteinander verglichen. Des Weiteren, wurden die Sequenzen bei denen die Identifizierung bestätigt wurde, in weitere Analysen unter folgenden Gesichtspunkten, einbezogen: (1) Der Identifizierungserfolg basiert auf der Methode der „Best- Close- Match“ Analyse von Meier (2006), welche zwei Sequenzen pro Art involviert, (2) die Analyse der sogenannten „Barcoding gaps“ involviert zwei Sequenzen pro Art, (3) die Analyse der taxonomischen Auflösung involviert eine Sequenz pro Art, für Familie- und Gattungsniveau und zwei Sequenzen pro Art für das Artniveau. Alle Analysen basieren auf den Ergebnissen von dem ClustalIW Programm, in dem Sequenzen abgeglichen (Thompson et al 1994) und in MEGA6 integriert wurden (Tamura et al 2013).

Die Ergebnisse zeigen, dass rbcL eine viel höhere Sequenz Serialiserbarkeit zeigt (95 %) als matK (66%). Obwohl zwei Primer benutzt wurden, war die Amplifikation von dem matK Gen sehr kompliziert. Eine hohe Anzahl der matK Sequenzen scheiterten und konnten den Standard Qualitätskriterien der CBOL Plant Working Group nicht erfüllen (2009)

Der Vergleich zwischen morphologischen und der molekularen Identifikation deckten auf, dass matK und rbcL am besten funktionierten, wenn Pflanzenarten in Gattungen zugeordnet wurden. Ungefähr 3% der gesamten Anzahl erzeugter Barcodes wurden falsch identifiziert (ungültige Referenz Sequenz, nicht korrekte morphologische Bestimmung, falsche Markierung und Kontamination der Proben während der Feld- und Laborarbeit). Schätzungen des Identifizierungserfolgs basieren auf der Analyse von Meier (2006), zeigten, das mehr als 70% der Arten, die in dem Datensatz (161 Arten) einbezogen wurden, korrekt identifiziert wurden, wenn einzelne Barcodes benutzt wurden. Die zwei- Loci Barcodes waren in der Lage den Identifizierungserfolg auf 80% zu erhöhen.

Die „Barcode-gap" Analyse ergab, dass weder matK noch rbcL erfolgreich zwischen der intra-spezifischen und der inter- spezifische Divergenz unterscheiden konnten. Allerdings waren die zwei Barcodes in der Lage mindestens 70% der Arten zu unterscheiden, die im Datensatz voneinander einbezogen wurden. Elf Arten wurden nicht unterschieden, trotz der Zwei- Loci Barcode Methode. Diese gehörten zu einer Artenreichen Gruppe der *Ficus, Santiria,* und *Litsea*.

Neun phylogenetische Bäume wurden mit drei Methoden in dieser Studie rekonstruiert (Neighbor Joining Algorithmus, Maximale Sparsamkeit, Maximum Likelihood Methode). Die Berechnungen der monophyletischen Prozentsatzes zeigten, dass die Kombination von matK und rbcL als zwei- Loci Barcodes höhere taxonomische Auflösungen zeigten, als wenn sie einzeln angewendet wurden. Es zeigte jedoch, dass 15 Gattungen und 21 Arten, nicht monophylitisch waren, trotz aller Marker. Die Topologie von 30 Ordnungen wurden rekonstruiert und nach APG III (2009) verglichen. Verschiedene Ordnungen wurden nicht korrekt bestimmt, wenn matK und rbcL angewendet wurden. Aber die Kombination beider Marker war in der Lage die Ordnungen richtig zuzuordnen.

Abschließend kann man sagen, dass matK und rbcL waren für die Studie nicht zufriedenstellend waren. Als Kern Barcodes, waren beide Marker effektiv, um Pflanzenarten bis zum Gattungsniveau zu identifizieren. Diese plastid Marker waren aber nicht ausreichend Variable, um nah-verwandte Taxa zu unterscheiden, in der Kombination aber in der Lage sind, genauere Unterschiede im Gattungsniveau zu machen.

Summary

Deforestation and forest degradation in tropical regions are progressing. As one of tropical forest areas, Sumatra has lost most of its natural forests which were converted into the plantation, such as oil-palm plantations and rubber plantations.

Biodiversity of tropical flowering plants in Sumatra is extraordinarily high, yet it is not fully explored. Biodiversity exploration should be put as a priority of the conservation efforts in threatened area such as Sumatran tropical lowland forests. However, limiting factors, particularly in work of species identification, make the species exploration become more challenging. A new method of species identification is highly demanded.

The first successful application of DNA barcoding in 2003 has answered the demand for a rapid method for species identification. Immediately afterward, DNA barcoding became a new trend in the world of biology sciences. The application of this method has been widely spread, including flora exploration. However, the practices of plant DNA barcoding have been confronted by the unavailability of universal genes that can work throughout the whole plant taxa. Despite the debates that are going on around which are the most suitable gene, matK and rbcL are selected as the core barcodes for plants.

This study was carried out aiming to generate DNA barcodes over flowering plant species in Sumatra. The DNA barcodes were generated using core barcodes for the plant (matK and rbcL). These two barcodes were evaluated based on their performance in identifying species, as well as their universality to be utilized over a wide range of plant taxa. The investigation of these markers will contribute to the development of useful barcode information for the identification of plants in tropical forests.

The study sites were located in Jambi province in the center of Sumatra. Specimens of flowering plants were collected from thirty-two study plots distributed in two landscapes, Bukit Duabelas National Park and Harapan Rainforest. Dried-leaf specimens and herbarium vouchers were

collected from 5.100 samples. Morphological identification was conducted by collaborated taxonomist reporting more than 1.100 species of 436 genera, 136 families, and 49 orders.

Due to time limitation, only half of the total samples were analyzed. DNA amplification and sequencing were conducted using matK and rbcL to each of the investigated specimens. As a result, 3.500 barcodes were generated over 500 flowering plant species.

Sequencing editing was done to each successfully generated barcode to maximize the quality and readability of the sequences. The molecular identification was conducted by inquiring the generated barcodes to the nucleotide databases, such as GenBank and BOLD. The results from morphological and molecular identification were compared to each other. Furthermore, all the sequences that the identification has been confirmed, were included in the analysis, as follows: (1) identification success using best-close match analysis (Meier et al 2006) involving two sequences per species, (2) barcoding gap analysis involving two sequences per species, (3) taxonomical resolution analysis involving one sequence per species for family and genus level and two sequences per species for species level. All of these analyzes were based on the result of sequence alignment conducted by using ClustalW program (Thompson et al 1994) embedded in MEGA6 (Tamura et al 2013).

The results show that rbcL has much higher level of sequence recoverability than matK (95% and 66%, respectively). Despite using two primers, amplifying and sequencing matK gene was very difficult and a great number of sequences were failed to fulfill the criteria of good-quality sequence according to CBOL Plant Working Group standard (2009).

The comparison between morphological and molecular identifications revealed that matK and rbcL worked best assigning a plant specimen to a certain genus. About 3% of the total numbers of generated barcodes were misidentified due to several reasons (invalid reference sequences, incorrect morphological identification, specimen mislabeling and contamination during field work and lab work). Estimation of identification success using best-close match analysis showed that more than 70% of species included in the dataset (161 species) were correctly identified when using single barcode. The two-loci barcodes were able to increase the identification success up to 80%.

The barcoding gap analysis revealed that neither matK nor rbcL was succeeded to create a clear gap between the intra-specific and inter-specific divergences. However, these two barcodes were able to discriminate at least 70% of the species included in the dataset from each other. Eleven species were not discriminated even with the two-loci barcode, mostly belong to species-rich groups such as Ficus, Santiria, and Litsea.

Nine phylogenetic trees were reconstructed in this study using three different methods (Neighbor Joining, Maximum Parsimony, Maximum Likelihood). Calculation of monophyletic percentages showed that combining matK and rbcL to be a two-loci barcode resulted in higher taxonomical resolution than if using single barcode. However, fifteen genera and twenty-one species were found to be non-monophyly with all of the markers. Topologies of 30 orders were reconstructed and compared to the one according to APG III (2009). Several orders were misplaced when using matK and rbcL, but the combination of these markers was able to put these orders into the correct positions.

In conclusion, matK and rbcL were not satisfying in all manners. As the core barcodes, these two markers were effective to be used in plant species identification at least up to genus level. These plastid markers were not sufficiently variable to discriminate closely-related taxa. The combination of matK and rbcL, however, was proven to have a higher level of discriminatory power.

At the end of this study, all of the barcodes that were successfully generated will be uploaded to BOLD in hope to provide a considerably large number of new molecular information of yet undocumented plant species in Sumatra. This will add to the rapidly growing molecular information on plant diversity in the tropics.

Contents

List of tables

List of figures

List of appendixes

1 INTRODUCTION

1.1 Flora diversity in Sumatra

Indonesia is home to some of the most magnificent tropical forests in the world. Although Indonesia comprises only 1,3 percent of the Earth's land surface, it is a mega biodiversity hotspot, including 11 percent of the world's plant species (FWI/GFW, 2002). The plant diversity in Indonesia ranks fifth in the world, with more than 38.000 species (55% of which are endemic). Palm diversity in Indonesia ranks first in the world, with 477 species (225 of which are endemic). More than half of the total timber producing trees species with economic value, mostly members of the Dipterocarpaceae family, are found in this country of which 155 are endemic (BAPPENAS 2003).

Sumatra, stretching from latitude 6° South to 6° North and from longitude 95° to 106° East, is one of the main islands of Indonesia that contains some of the most important tropical rainforests in the world which is rich in flora and fauna. In term of flora diversity, the Sumatran forests are comparable to the Borneo forests and are richer than those found in Java and Sulawesi (Meijer 1981 in Whitten et al 2000). Sumatra is reported as one of the global centers of vascular plant diversity with 3.000 to 5.000 species per 10.000 km^2 (Barthlott et al 2005).

Whitten et al (2000) reported that Sumatra has one of the largest tropical lowland forest areas in the world. The diversity of tree species in Sumatran lowland forest is extremely high. A study by Roos et al (2004) shows that plant species richness in Sumatra is estimated up to 10.600 species with more than 300 endemic species. More information about Sumatra' flowering plant diversity is presented in sub-chapter 2.3.

Despite being one of the biodiversity hotspots in Indonesia, Sumatra has been neglected in comparison to other islands in the Malayan Archipelago. Many scientists mistakenly consider that the flora of Sumatra is sufficiently well known since it is similar to that of the Malaysian peninsula. Many parts, especially the central portion of the island, are almost unexplored territory from the floristic standpoint (Laumonier 1997).

1.2 Forest degradation and deforestation in Sumatra

The World Bank reported in 2001 that the total forest area in Sumatra has decreased from over 23 million hectares to probably less than 16 million hectares between 1985 and 1997. A remote-sensing study by Margono et al (2012) quantified 7,54 million hectares of primary forest loss in Sumatra during 1990 – 2010 and an additional 2,31 million hectares of primary forest were degraded.

The central region of Sumatra, such as the South Sumatra, Jambi, North Sumatra and Riau provinces, suffer the highest rate of forest degradation (Holmes 2002). From 2000 to 2010, the deforestation rate was estimated to be more than 5% in the eastern lowlands of Sumatra (Miettinen 2011). Lambert and Collar (2002) indicated that the southern provinces of Sumatra have lost most of their lowland forests, including those in protected areas.

The causes of this massive deforestation and forest degradation are a large-scale conversion for timber or estate crop plantations (in particular, oil palm), small-holder conversion, and unsustainable and illegal logging. Nevertheless, deforestation in Indonesia is largely the result of a corrupt political and economic system that regarded natural resources, especially forests, as a source of revenue to be exploited for political and personal interests (FWI/GFW 2002).

Oil-palm is one of the world's most rapidly increasing crops. Gautam et al (2000) mentioned Sumatra as an excellent target for oil-palm plantation due to its fertile soil and good accessibility. This is why large investors came into Sumatra to establish plantation crops which have been acknowledged to be responsible for 40 percent of the deforested areas each year. Data derived from a study by Koh et al (2011) showed that 3,9 million hectares of lowland forests in Sumatra were converted into oil-palm plantations. Moreover, small-holder plantations of *Hevea* (rubber), coconut-palm, coffee, fruit, clove and cinnamon trees cannot be dismissed as causes of deforestation as Laumonier (1997) mentioned that these agricultural systems covered over three million hectares of the island.

1.3 DNA barcoding to accelerate species exploration

Due to the extensive loss of natural habitat caused by deforestation and forest degradation, a great number of species has become endangered. Many of these species may not even be discovered before they are finally extinct. Several studies reported that deforestation caused the biodiversity loss of tropical fauna, such as forest-dwelling birds (Koh et al 2011), mammals (Maddox et al 2007), orangutan (Gaveau et al 2009). Undoubtedly, the destruction affected the flora diversity as well. Kiew (2001, as cited in Kiew 2002) estimated at least 200 species of flowering plants in Malaysia peninsula has been extinct, which might reflect the extinction rate of flowering plants in Sumatra.

Species exploration in tropical forests seems to be slower than the species loss because of several limiting factors, such as the number of taxonomic specialists working in this region, inadequate herbarium collections, and inaccessible taxonomic literature (Kiew 2002, Tautz et al 2003, Meyer and Paulay 2005). Species exploration becomes more challenging when the species cannot be identified morphologically. Identification keys based upon morphology characteristics could be difficult to use when some features are not visible, as happens when specimens are not well developed, or outside specific life stages (e.g. flowering period).

With the increasing availability of molecular data, overcoming the limitations of morphological characters is possible by using molecular traits to help species identification. In principle, DNA variation can be used as a character to differentiate species. Based on this principle, DNA barcoding enables specimen identification of any part or developmental stage of the species (Hebert et al 2003).

In order to speed up the species exploration in tropical forests, reliable and efficient methods of species identification using molecular traits are highly demanded (Finkeldey et al 2009). The use of DNA barcoding for species identification is of particular relevance for extremely species-rich ecosystems such as tropical forests. The capacity of DNA barcoding to identify species rapidly and accurately is an important advantage in the effort to explore and to document genetic information of highly diverse but endangered populations such as those occurring in Sumatran tropical lowland forest.

1.4 The objectives and general information of the study

Deforestation and forest degradation have destroyed forest areas all over Sumatra, including Jambi. Located on the east coast of central Sumatra, Jambi has suffered severe forest loss, especially in lowland areas. Remaining forests are confined to few regions including Bukit Duabelas National Park and Harapan Rainforest. These areas consist of secondary forests, mixed with agroforests, mainly rubber and oil-palm plantations.

This study took place in Jambi and focused on four different land-use systems: secondary forest, jungle rubber, oil-palm plantation, and rubber plantation. The aims of this study are to generate DNA barcodes of flowering plant species in Jambi using two DNA chloroplast markers (matK and rbcL) and to evaluate the effectiveness of these two markers as DNA barcodes for flowering plants. Crucial characteristics for evaluating the performance of DNA barcode include universal applicability, ease of data retrieval and sufficient variability of the marker used (Kress and Erickson 2007, Fazekas et al 2008). The investigation of these markers will contribute to the development of useful barcode information for the identification of plants in tropical forests.

In order to accomplish the aims, several works has been carried out: a) specimen collection of all flowering plant species that had been found inside the sampling plots distributed in Bukit Duabelas National Park and Harapan rainforest; b) DNA extraction, amplification, and sequencing of two samples per species; and c) DNA sequence analysis to generate accurate DNA barcodes and to evaluate the identification success, discriminatory power, and taxonomical resolution level of matK, rbcL, and the combination of these two markers. The work was conducted from January 2013 to July 2015.

Plant species identification is expected to be challenging using DNA barcoding due to the high diversification of plants (Linder 2008, Richardson et al 2001). Moreover, tropical plant species tends to have lots of close relatives, which at the end will make the effort to distinguish these species more difficult (Hollingsworth 2009a, Couvreur et al 2008).

The applications of DNA barcoding rely on the existence of a robust reference library that will allow DNA sequences to be assigned to known taxonomic groups. Currently, DNA sequence

information of many plant species is still lacking. For instance, there are an estimated 350,000 species of angiosperms (flowering plants), but as of September 2015, only about 88.000 species have DNA sequence information stored in one of the most prominent nucleotide database, Barcode of Life Database (BOLD). This means there are opportunities to submit novel sequences and to contribute to the global plant species documenting effort.

In order to enrich the digital library of plant biodiversity, all the DNA barcodes that have been generated through this study will be uploaded to BOLD. Thus, the DNA sequence information of all the species that have been investigated in this study can be freely accessed for everyone who may be interested and for further review.

2 LITERATURE REVIEW

2.1 A brief history of DNA barcoding

In 2003, Paul Herbert and colleagues published a paper about species identification using DNA sequences (Hebert et al 2003a). The paper was considered as the first example of animal identification method using a part of mitochondrial DNA region cytochrome oxidase subunit 1 (CO1). The term "DNA barcode" was introduced as a short gene sequence from a standardized region of DNA that is utilized as a marker for species identification (Hebert et al 2003a, 2003b). This successful study was soon followed by similar studies using CO1 for species identification of other animal groups such as birds (e.g., Hebert et al 2004, Waugh 2011), fishes (e.g., Ward et al 2005, Wong 2011), mammals (e.g., Lim 2012, Echi et al 2013), amphibians and reptiles (e.g., Crawford et al 2010, Bina Perl et al 2014, Nagy et al 2012).

As a relatively new method, DNA barcoding has been applied to taxa across the tree of life. This has been made possible because DNA barcoding can identify organisms at any stage of development (e.g., Ko et al 2013, Heimeier et al 2010, Barber and Boyce 2006, Hausmann et al 2009), or at particular gender (e.g., Elsasser et al 2009), or specimens isolated from small and incomplete tissue, whether it is fresh, broken or old (e.g., Valentini et al 2008, Hajibabaei et al 2006). This method also helps to discover new species and to identify cryptic species (e.g., Hebert et al 2004b, Pauls et al 2010, Ward et al 2008).

One of the most significant contributions of DNA barcoding is as an environmental law enforcement tool. DNA barcoding has been used actively to prevent biodiversity illegal trading (e.g., Asis et al 2014, Eaton et al 2009) and illegal hunting (e.g., Sanches et al 2005, Bitanyi et al 2012). In regards to commercial trading, DNA barcoding has proven its ability to uncover the mislabeling of food products (e.g., Cawthorn et al 2012, Baker et al 2010, Nicole et al 2012).

As the potential of DNA barcoding has been recognized, several awards and meetings have been conducted since then to promote the implementation of this new approach in species identification practices (Savolainen et al 2005). In 2004, the Consortium for the Barcode of Life (CBOL) was established as an international initiative focusing on the development of DNA

barcoding as a global standard for the identification of biological species. The first international conference of DNA barcoding, held on February 2005 in London, has been the starting point of comprehensive discussion on the DNA barcoding studies and implementations (BOLI 2010-2015).

One of the main topics discussed at the first DNA barcoding international conference was to find a particular DNA region as the global DNA barcode. Ideally, DNA barcodes should be short enough (400 – 800 bp) to be easily recovered from target specimens and have sufficient sequence variation to discriminate among species (Hebert et al 2003b, Savolainen et al 2005). The sequence variation in this barcode region should be high enough between species (inter-specific), but low within species (intra-specific). Therefore, the selection of a barcode region is complicated by the trade-off between the need for universal application and optimal rates of sequence variation (Kress et al 2005).

In principle, the work of DNA barcoding is divided into two parts: (1) building the barcode library of known species and (2) matching or assigning the barcode sequence against the barcode library for identification (Kress and Erickson 2012). In order to fulfill the need for an extensive barcode library, CBOL established the Barcode of Life Database (BOLD) in 2005. BOLD is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records. The key feature of this database is a persistent linkage between barcode sequences and their source specimens (Ratnasingham and Hebert 2007). As a free-accessed database, BOLD is targeted to provide a standardized identification tool for the largest possible community of end-users. Up to date, more than four million barcode sequences have been uploaded to BOLD which covered more than 238.000 species (BOLD 2014).

Although DNA barcoding has been admitted as a cost-efficient and rapid method of species identification (e.g., Rach et al 2008, Dentinger et al 2010), debates over its potentials are ongoing (Moritz and Cicero 2004, Hebert and Gregory 2005, DeSalle 2006). Prendini (2005) suggested that DNA barcoding alone cannot describe a new species, so that it should be used in association with traditional taxonomic methods in order to adequately describe species. DNA

barcoding is most successful in species description when it is used in complement with other sources of information, such as morphological data (Goldstein and DeSalle 2011).

As a species identification method, there are at least two weaknesses of DNA barcoding. First, there is no agreement on a consensual divergence threshold for species delimitation. Recent studies proposed different thresholds across taxonomic groups (e.g., Ratnasingham and Hebert 2007, Hebert et al 2003) but neither has proved useful for all taxa (e.g., Renaud et al 2012, Hickerson et al 2006). Second, DNA barcoding has limited success when identifying new species in unstudied groups (Meyer and Paulay 2005). Despite all the debates and controversies over its feasibility and utility, DNA barcoding has grown exponentially in terms of the number of sequences generated as barcodes as well as its applications.

2.2 DNA barcoding of plants

Several studies have evaluated the utility of plant DNA barcoding (e.g., Kress et al 2009, de Vere et al 2011, Lahaye et al 2008, Gonzales et al 2009, Wang et al 2010, Burgess et al 2011, de Groot et al 2011, Li et al 2011) and a variety of applications have been developed that showed the wide potential of plant DNA barcoding. For instance, DNA barcoding has been used for the verification of plant products such as medicinal plants (e.g., Chen et al 2010, Mankga et al 2013), kitchen spices (e.g., De Mattia et al 2011), and tea plants (e.g., Stoeckle et al 2011). Equally diverse ecological applications have been implemented, including the identification of invasive species (e.g., Bleeker et al 2008) and characterization of below-ground plant diversity (e.g., Kesanakurti et al 2011). Some study cases of plant DNA barcoding are presented with more details in sub-chapter 2.2.3.

Even though plant DNA barcoding has been widely used as mentioned above, there is no single gene that can satisfy a global DNA barcode for plants. In many animal groups, the mitochondrial gene CO1 has been proven as an effective barcode for species identification (Hebert et al 2003a). Due to low substitution rate of mitochondrial DNA in plants (Chase et al 2005), CO1 is considered to be insufficiently variable to distinguish species across a wide range of plant taxa (Cowan and Fay 2012).

2.2.1 Universal barcodes for plants

Finding a universal DNA barcode for plants presents a number of challenges compared to animals (Cowan and Fay 2012). In animals, there is a high base-substitution rate, but the gene content and order are highly conserved. Whereas in plants, with a few exceptions in specific taxa, base substitution rates are much lower, and there are frequent genome rearrangements and transfers of genes between different genomes and across species (Palmer et al 2000).

According to Chase et al (2007), the most important characteristics of a universal barcode are to be able to be amplified across all taxa using standardized primers and ease to be sequenced. In addition, the barcode needs to exhibit sufficient variability for species-level identification, ideally with high inter-specific and low intra-specific sequence divergence. To facilitate analysis, the barcode should be easily aligned, and should contain few insertions and deletions, as these complicate the comparison and can be difficult to interpret. Also, for many potential applications, the barcode needs to be recoverable from herbarium samples and other degraded samples. Finding a barcode that completely fulfills the desired criteria is not entirely achievable. This calls for decisions based on a trade-off between sequence quality, power to discriminate among species, and cost (CBOL Plant Working Group 2009).

In many cases, the level of species discrimination in plants with standard DNA barcoding markers are lower than those obtained by CO1 in many animal groups (Fazekas et al 2009). However, this is a matter of the taxa group in question; the species level might be reasonably good in some groups and quite poor in others. When the specimen sampling is limited geographically which then restricts the number of closely related species, rates of species discrimination are expected to be greater (e.g., Burgess et al 2011).

A number of candidate gene regions were suggested as potential barcodes for plants including coding genes and non-coding genes in the nuclear and plastid genomes (e.g., Kress et al 2005, Kress and Erickson 2007, Taberlet et al 2007, Chase et al 2007). In 2009, the Plant Working Group under CBOL recommended a two-loci standard barcode (matK + rbcL) for plant DNA barcoding (CBOL Plant Working Group 2009, Hollingsworth et al 2009). In this next section, some widely-used plant barcodes will be described in more details.

2.2.1.1 matK gene

The chloroplast matK marker consists of a ca. 841 base pairs (bp) at the center of the gene, located between bp 205–1046 (including primer sites) in the complete *Arabidopsis thaliana* plastid genome sequence (Hollingsworth et al 2011). This gene, which encodes a maturase enzyme, evolves rapidly (Hilu et al 1997) and is considered to be one of the most informative loci for determining phylogenetic relationships (Hilu et al 2003).

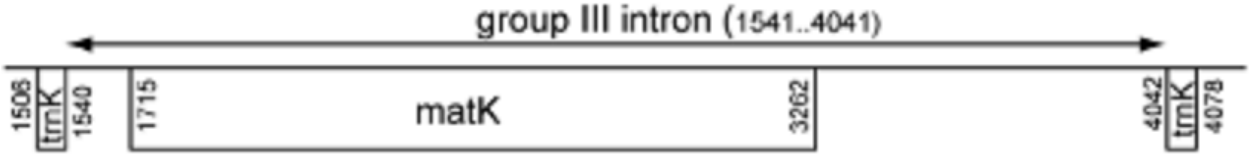


Figure 2.1 Genomic view of matK gene

There are mixed reports in the literature about amplification and sequencing success using matK (e.g., de Vere et al 2012, Kress et al 2009, Roy et al 2010). Although matK has been found to be an informative region in several plant DNA barcoding studies on various floristic or biodiversity hotspots (e.g., Chase et al 2007, Lahaye et al 2008, Soltis et al 2001) and to be quite variable in numerous plant groups (e.g., Hilu et al 2003, Hidayat et al 2011), this locus was not found to be useful in several other studies (e.g., Kress & Erickson 2007, Chen et al 2010, Zhang et al 2015). As a region with the highest species discrimination among the other coding regions ever been studied, matK was not easily retrievable using universal primers, particularly in non-angiosperm samples (Cowan and Fay 2012). As a consequence, primers of matK still need to be optimized, probably to be adapted to specific taxonomic groups.

Using the best currently available universal primer pair (3F/1R; Kim unpublished) on diverse sample sets typically results in PCR and sequencing success of ca. 70% in angiosperms. Using secondary primer pair (390F/1326R; Cuenoud et al 2002) can increase amplification and sequencing success by another ca. 10% (Hollingsworth et al 2011).

2.2.1.2 rbcL gene

The chloroplast rbcL marker consists of a 599 bp region at the 5' end of the gene, located at bp 1–599 (including primer sites) in the complete *Arabidopsis thaliana* plastid genome sequence (Hollingsworth et al 2011). This gene is the first gene to be sequenced in plants, exists as a single copy and contains no introns (Zurawski et al 1981).

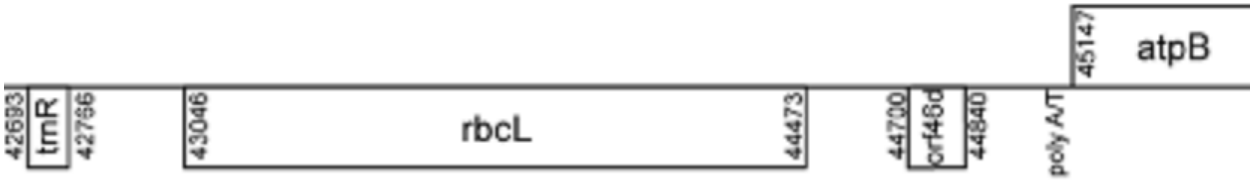


Figure 2.2 Genomic view of rbcL gene

Since it is one of the most conserved genes in the chloroplast genome, this gene has been widely used as a tool to retrace the evolutionary relationships of plant groups that diverged over historical time. It is easily amplified and sequenced in most land plants, but showed too little variation to enable identifying all plant species (e.g., Hollingsworth et al 2009), thus it has been discounted as a species-level discriminator (Renner 1999, Salazar et al 2003). To increase the power of this gene for phylogenetic purposes, it should be combined with more variable regions (Vijayan and Tsou 2010).

2.2.1.3 Chloroplast non-coding genes

The chloroplast genome can be divided into three functional categories: exons, introns, and intergenic spacers, the latter two do not encode proteins and are referred to as non-coding regions (Shaw et al 2005). The non-coding genes are supposed to evolve more rapidly than the coding genes, enabling it to serve as primary source of data for molecular analysis. Because of this reason, non-coding sequences of chloroplast DNA became important tools in the phylogenetic analysis of a broad range of plant groups at a variety of taxonomic levels (Kelchner

2000). Nevertheless, the non-coding regions usually exhibited poor sequence quality, mainly caused by microsatellite regions and stutters (Devey et al 2009).

One of the most widely-used genes as DNA barcode, trnL, is a chloroplast non-coding gene which has conserved priming sites (Taberlet et al 2006) and has a relatively slow rate of evolution (Shaw et al 2005). The limited inter-specific divergence of this gene makes it an unlikely universal marker for species-level identification. This gene is in general simple to sequence, although mononucleotide repeats can impact on sequencing reads in some taxa (Hollingsworth et al 2011).

The non-coding spacer, trnH-psbA, has been used successfully in a range of plant DNA barcoding studies (e.g., Gonzales et al 2009, Kress et al 2009, Kress et al 2010) and is an obvious choice of a supplementary barcode (Hollingsworth et al 2011). The trnH-psbA region is straightforward to be amplified across land plants and is one of the most variable inter-genic spacers in plants (Shaw et al 2007). A study by Kress and Erickson (2007) showed that trnH-psbA has dramatically higher sequence variability than the coding genes because it has a higher number of single nucleotide polymorphisms (SNPs). In a recent study by Ghahramanzadeh et al (2013), it is proposed that trnH-psbA is more informative than rbcL.

2.2.1.4 Nuclear genes

Nuclear genes are bi-parentally inherited, therefore they are less conserved compared to chloroplast genes which are maternally inherited. It is also more difficult to recover nuclear DNA from herbarium samples and other degraded samples. This is because there is less copy per cell of this genome compared to the plastid genome.

The nuclear internal transcribed spacer (ITS) region is identified as a potential barcode marker by Kress et al (2005). This non-coding region generally shows high levels of inter-specific sequence variability (Cowan and Fay 2012), but is difficult to be amplified and sequenced in diverse sample sets and it is prone to fungal contamination (Alvarez and Wendel 2003,

Hollingsworth et al 2011). In cases where ITS is difficult to be amplified and performing unsatisfactorily, ITS2 represents a useful alternative (Yao et al 2010).

The ITS region has been used successfully to classify angiosperms (Li et al 2011) and the ITS2 region has been useful in identifying a range of medicinal plants compared to other genetic markers (Chen et al 2010). A study by CBOL Plant Working Group China (2011) revealed that the discriminatory power of ITS2 is higher than that of plastid markers.

2.2.2 The success of multi-loci DNA barcode for plants

In a recent review of the most optimal barcode for plants, Hollingsworth et al (2011) indicated that none of the barcodes proposed is perfect in every respect. Some studies suggest strategies based on a single chloroplast region (e.g., Lahaye et al 2008) or a combination of different regions (e.g., Kress and Erickson 2007, Chase et al 2007, Holingsworth et al 2009). A study by Kress and Erickson (2007) showed that the various combinations of two loci were all more powerful at differentiating between species than either locus individually. In overall, it appears that different genetic markers can be applied to plants with different degrees of success.

CBOL Plant Working Group (2009) suggested that there are no other two-loci or multi-loci barcode provided appreciably greater species resolution than the matK+rbcL combination. However, in some complex groups, such as in the genus Berberis (Roy et al 2010), the combination of matK with rbcL is not sufficient to distinguish all species.

There are many studies showing that the species discriminatory level will be increased by combining the core barcodes with supplementary genes such as trnL, trnH-psbA, and ITS or ITS2 (e.g., Li et al 2011, Jeanson et al 2011, Armenise et al 2012). Taberlet et al (2007) and Janzen (2005) indicated the use of multi-loci barcode will give much higher species discrimination level for studies within a limited geographical area.

2.2.3 Case studies of plant DNA barcoding

This sub-chapter is dedicated to exhibiting three case studies of the application of DNA barcoding on plant species identification and phylogenetic analysis.

2.2.3.1 DNA Barcoding the native flowering plants and conifers of Wales (De Vere et al 2011)

This study is claimed to be the first national DNA barcode resource that covers the native flowering plants and conifers for the nation of Wales (1.143 species). This study did not only investigate DNA extracted from fresh materials but from herbarium specimens as well.

Recoverability of DNA barcodes was lower using herbarium specimens, compared to the freshly collected material, mostly due to lower amplification success, but this is balanced by the increased sampling efficiency of species that have already been collected, identified, and verified by taxonomic experts. Using rbcL and matK, this study has resulted in a very high rate of recovery for rbcL but lower for matK. The effectiveness of the DNA barcodes for identification was assessed using three approaches: the presence of a barcode gap (using pairwise and multiple alignments), formation of monophyletic groups using Neighbour-Joining trees, and sequence similarity in BLAST searches. These approaches yielded similar results, providing relative discrimination levels of 69.4 to 74.9% of all species and 98.6 to 99.8% of genera using both markers. A database of DNA barcodes for Welsh native flowering plants and conifers was established as the output of the study, offering a valuable and complete platform for a wide range of applications that require accurate species identification.

2.2.3.2 Identification of Amazonian trees with DNA barcodes (Gonzales et al 2009)

This study was conducted in two hectares of a tropical forest in French Guiana, investigating all trees with a diameter at the breast height of more than 10 cm. A total of 1.073 trees were sampled, encompassing 301 tree morpho-species, 143 genera, and 54 angiosperm families. These samples were barcoded using eight different markers: rbcLa, rpoC1, rpoB, matK, ycf5, trnL, psbA-trnH, and ITS.

As the result, psbA-trnH had the best performance as a DNA barcode, also being universally amplifiable. The core barcode, matK, was sequenced with a relatively low level of success compared to other markers even after using two different pairs of primers. However, this gene turned out to have an adequate rate of variation and successfully revealed cryptic species. Furthermore, ITS did not seem promising as a universal DNA barcode for tropical forest plant species as limited sequencing success was observed in this study. In conclusion, this study indicates that the combination of two or more markers did not greatly improve the overall performance in comparison with single markers.

2.2.3.3 DNA barcoding the floras of biodiversity hotspots (Lahaye et al 2006)

In this study, an intensive specimen collection in two biodiversity hotspots (Meso-America and Southern Africa) was conducted. More than 1.600 samples were collected and investigated to compare seven potential barcodes: matK, rbcL, trnH-psbA, trnL, rpoC1, rpoB, and ycf5.

This study indicates that either matK or trnH-psbA are the most suitable region for plant DNA barcoding. Using matK alone or in combination with trnH-psbA resulted in 90% of correct species identification. By combining the molecular and morphological data, this study has successfully assigned almost 60% of the sampled tree saplings to known species and the remaining to the genus and the family.

2.3 Flowering plants in Sumatra

Sumatra is divided into two major areas: highland areas in northern and central Sumatra and lowland and coastal areas in the western, eastern and southern parts of the island. In most of the literature, the vegetation of Sumatra was described according to these area divisions.

The highland areas of Sumatra comprises of sub-montane forest, montane forest, and cloud forest. These types of forest are dominated by vegetation of Lauraceae, Fagaceae, and conifers such as *Pinus merkusii* and *Agathis dammara* (de Wilde 1989). Davis et al (1995) reported that

endemism is relatively high in some areas of Sumatra's highland, including the mountains of Aceh (Gunung Lauser), the Indragiri foothills, and the Tigapuluh mountains.

The lowland areas of Sumatra were originally covered by evergreen forest. Some areas of lowland forest were dominated by commercially important vegetation, such as timber trees of the dipterocarps and *Eusideroxylon zwageri* (Lauraceae) which are included in the IUCN Red List of Threatened Species. Whitten (2000) wrote that Sumatra's lowland forests are characterized by the presence of tall trees with large-buttresses and smooth-barked trunks. Lianas and climbing palms are also abundant in this type of forest. Meanwhile, the under-storey vegetation is composed of Marantaceae, Zingiberaceae, ferns, many Araceae and a few isolated Poaceae (Laumonier 1997).

Estimations about the plant diversity in Sumatra are uncertain. Whitmore and Tantra (1986) reported 364 genera of 86 families in Sumatra which have at least one big tree species. A very optimistic estimation of de Wilde (1989) suggested the total number of 8.000-10.000 plant species in Sumatra. In 1990, Laumonier recorded 2.500 tree species in Sumatra which were assumed to represent only 70-80% of the total tree flora. Furthermore, Davis et al (1995) estimated that the Sumatra flora comprises about 10.000 species, which is much higher than the number of plant species in Malaysia peninsula. However, the endemism in Sumatra is generally considered to be lower than the Peninsula. Laumonier (1997) suggested that the lower endemism in Sumatra could be due to the fact that endemic species have become extinct.

Table 2.1 Species number and endemicity of some taxa in Sumatra compared to Malaysia peninsula (Kiew 2002)

	Sumatra		Malaysia Peninsula	
	No. of species	Endemicity (%)	No. of species	Endemicity (%)
Dipterocarpaceae	95	10	156	16
Elaeocarpus	34	38	42	47
Ericaceae	76	59	47	44
Impatiens	30	96	11	72
Monophyllaea	5	80	7	71
Nepenthes	29	76	10	50
Orchidaceae	1.118	41	850	27
Paraboea	6	80	18	77
Rhaphidophora	15	26	15	13
Schismatoglottis	16	56	7	14

This study took place in the Jambi province, in which remaining forests are scarce. In 1997, Laumonier described the lowland forest areas in the center of Jambi as covered with some species of Dipterocarpaceae, such as *Anisoptera megistocarpa, Dipterocarpus lowii, D. crinitus, Hopea ferruginea, Shorea macroptera, S. pauciflora, S. singkawang*. Meanwhile, Davis et al (1997) reported that the mountain region of this province (the Kerinci mountain) contains 2.000 to 3.000 vascular plant species, most of which were good-quality timber trees, both dipterocarps and non-dipterocarps. Besides the commercial species, wild fruit trees including *Mangifera* (mangoes), *Durio* (durians), *Lansium* (langsat), and *Nephelium* (rambutan); and medicinal plants such as *Parashorea lucida, Sindora spp., Lansium domesticum, Aglaia argentea, Eurycoma longifolia* and *Monophyllaee horsfieldii*; were considered to be locally important commodities.

Figure 2.3 (left to right) *Shorea singkawang* in Bukit Duabelas National Park; upper-canopy tree of *Kompassia sp.*, a climber strangling the host tree

Table 2.2 Species of the upper canopy (35-45 m) in the eastern lowland forests of Jambi (Davis et al 1995)

ANACARDIACEAE	BURSERACEAE	OLACACEAE
Mangifera rigida	*Santiria conferta*	*Scorodocarpus borneensis*
Mangifera torquenda	*S. rubiginosa*	*Ochanostachys amentacea*
Pentaspadon velutinus	*S. tomentosa*	*Strombosia ceylanica*
	S. griffithii	
APOCYNACEAE	*Dacryodes incurvata*	CELASTRACEAE
Dyera costulata		*Kokoona reflexa*
	EUPHORBIACEAE	*Lophopetalum beccarianum*
DIPTEROCARPACEAE	*Aporusa nervosa*	
Anisoptera costata	*Baccaurea costulata*	ARECACEAE
Anisoptera laevis	*Blumeodendron tokbrai*	*Oncosperma horridum*
Anisoptera marginata	*Blumeodendron calophyllum*	*Livingstonia kingiana*
Anisoptera megistocarpa	*Pimelodendron griffithianum*	
Dipterocarpus crinitus	*Ptychopyxis sp.*	POLYGALACEAE
Dipterocarpus lowii	*Trigonopleura malayana*	*Xanthophyllum amoenum*
Hopea dryobalanoides		*Xanthophyllum rufum*
Parashorea lucida	ANNONACEAE	*Xanthophyllum sulphureum*
Shorea acuminata	*Cyathocalyx ramuliflorus*	*Xanthophyllum vitellium*
Shorea dasyphylla	*Monocarpia marginalis*	
Shorea hopeifolia	*Polyalthia sumatrana*	MYRISTICACEAE
Shorea lumutensis	*Polyalthia hypoleuca*	*Myristica gigantea*
Shorea macroptera	*Xylopia malayana*	*Gymnacranthera bancana*
Shorea ovalis	*Xylopia ferruginea*	*Gymnacranthera farquhariana*
Shorea parvifolia		*Horsfieldia pulcherrima*
Shorea gibbosa	SAPOTACEAE	*Horsfieldia triandra*
Shorea johorensis	*Palaquium oxleyanum*	
Vatica stapfiana	*Palaquium cryptocariifolium*	SAPINDACEAE
	Payena acuminata	*Nephelium lappaceum*
CAESALPINIACEAE	*Payena endertii*	*Xerospermum norhonianum*
Dialium laurinum	*Pouteria malaccensis*	
Dialium maingayi		STERCULIACEAE
Koompasia malaccensis	CLUSIACEAE	*Heritiera sumatrana*
	Calophyllum depressinervosum	*Scaphium macropodum*
LOGANIACEAE	*Calophyllum pulcherrimum*	
Fragraea gigantea	*Calaphyllum calaba*	
MELIACEAE		
Dysoxylum acutangulum		

3 MATERIALS AND METHODS

3.1 Study sites

This study took place in two landscapes in Jambi Province, Sumatra, Indonesia: Bukit Duabelas National Park and Harapan Rainforest. These two study sites are representing the remaining tropical rainforests in Sumatra. Bukit Duabelas National Park (1°51'S 102°39'E) covers 605 km^2 of area consists of primary forest and logged-forest. Meanwhile, Harapan Rainforest (2°14'S 103°19'E) was used to be a logging-concession area which was converted into flora and fauna sanctuary managed by local and international NGOs.

The study sites comprise of 32 plots sized 50 m x 50 m distributed around the landscapes. These plots were established on 4 kinds of land-use: secondary forest, jungle rubber, oil-palm plantation and rubber plantation; each was represented by 8 plots.

3.2 Specimen collection

Specimen collection was conducted in all of the study plots. Specimens of big trees (diameter at breast height $\geq$ 30 cm) were collected from the whole area of each plot and specimens of under-storey vegetation (shrubs, lianas, seedlings) were collected from five sub-plots sized 5 m x 5 m. Each species found in the study site was sampled at least 3 times. From each sample, leaf tissues (approximately 2 cm^2) were collected fresh and dried in silica-gel for DNA analysis; herbarium vouchers were prepared and stored in Herbarium Bogoriensis and BIOTROP Herbarium, Bogor, Indonesia; several high-quality photographs were taken for further identification and to be uploaded along the DNA barcodes to DNA barcoding database.

3.3 Morphological species identification

Each collected specimen was morphologically identified by collaborating taxonomists by matching the correspondent herbarium vouchers against the reference vouchers available at the Herbarium Bogoriensis and BIOTROP Herbarium, Bogor, Indonesia. The results of the

morphological identification then were compared to the molecular identification results (see sub-chapter 3.5.1).

3.4 DNA analysis

The laboratory work, that has been carried out in this study, was taking place in Forest Genetics and Forest Tree Breeding Department, Georg-August-University Göttingen.

3.4.1 DNA extraction

Based on the preliminary result of morphological species identification, the selection of which samples for analysis that would be analyzed was done by choosing 2 specimens per species. DNA extractions then were performed on healthy dried leaf tissue from all of the selected samples using the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany), following manufacturer's protocols. The concentration and quality of the extracted DNA were checked by 0,8-1% agarose gel electrophoresis with a Lambda DNA size marker (Roche), visualized by UV illumination using a polaroid camera.

3.4.2 DNA amplification

Each extracted DNA then was amplified by performing Polymerase Chain Reaction (PCR) using universal primers listed in Table 3.1. For rbcL, the amplification was straight-forward. But for matK, two stages amplification was performed. At the first stage, all investigated samples were included using universal primer 1RKIM and 3FKIM (Table 3.1). The second stage, using primer 390f and 990R (Table 3.1), included only samples that was not amplified or that produced multiple PCR products at the first stage.

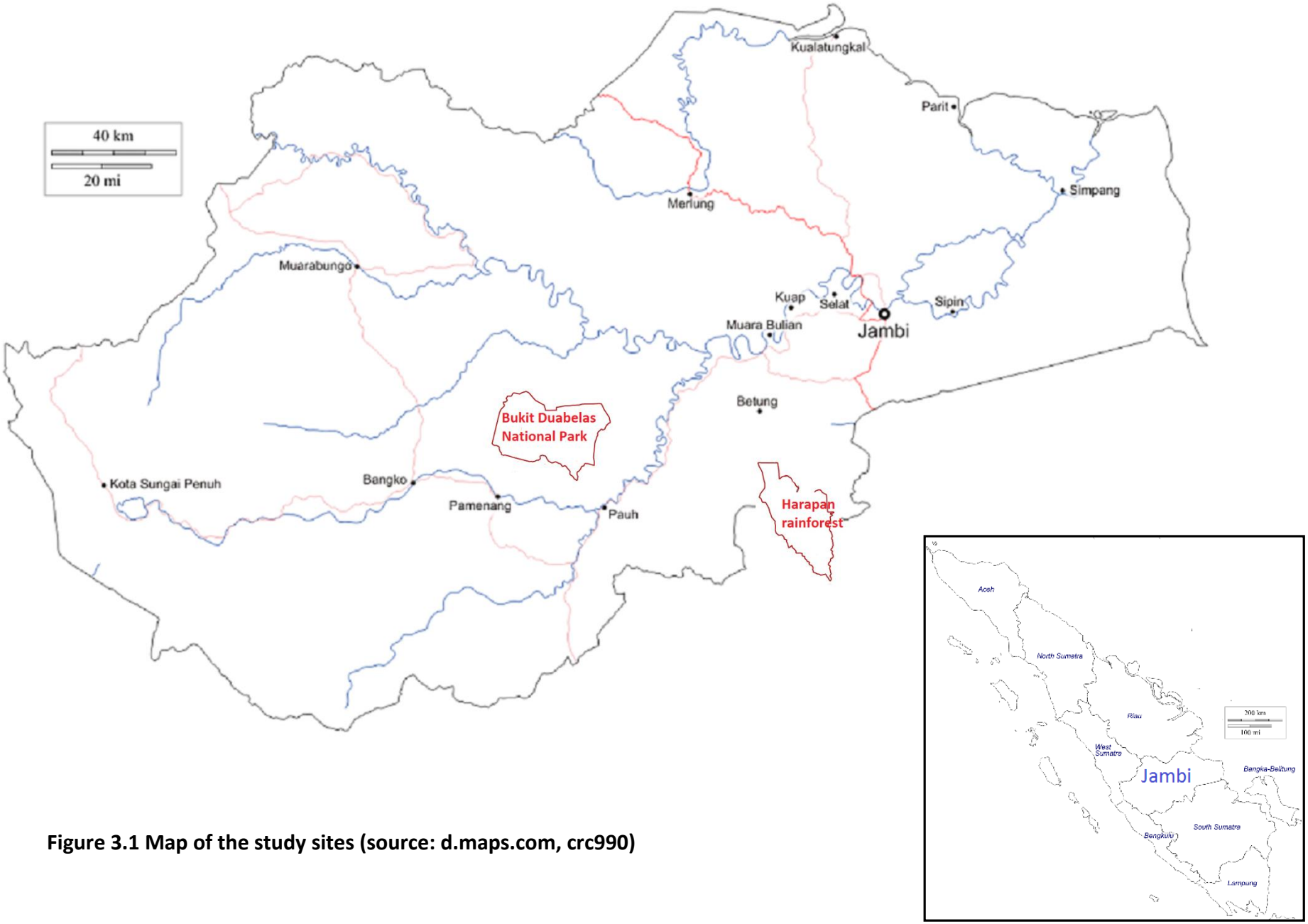


Figure 3.1 Map of the study sites (source: d.maps.com, crc990)

Table 3.1 List of utilized primers

No.	Region	Name of primer	Primer sequence (5′ → 3′)	Reference
1	matK	3F_KIM_f	CGTACAGTACTTTTGTGTTTACGAG	Ki-Joong Kim (unpublished)
		1R_KIM_r	ACCCAGTCCATCTGGAAATCTTGGTTC	Ki-Joong Kim (unpublished)
		390f	CGATCTATTCATTCAATATTTC	Cuenoud et al (2002)
		990R	GGACAATGATCCAATCAAGGC	Dayananda et al (2006)
2	rbcL	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	Krees and Erickson (2007)
		rbcLa_r	GAAACGGTCTCTCCAACGCAT	Fazekas et al (2008)

The PCR was performed in a Peltier Thermal Cyler PTC-200 (MJ Research Inc.) with a volume of 15µl reaction mixture (Table 3.2). PCR temperature profiles for the amplification reaction are shown in Table 3.3.

Table 3.2 Reaction mixture of PCR amplification reagent

Reagent	Volume (µl)
PCR buffer	1,5
$MgCl_2$	1,5
Forward primer (5pmol/ml)	1
Reverse primer (5pmol/ml)	1
dNTPs	1
Taq	0,2
H_2O	6,8
Template DNA (5-10 ng)	2

Table 3.3 Temperature profiles for PCR amplification reaction

Step	Condition
Step 1	Initial denaturation at 95° C for 15 minutes
Step 2	35 cycles of: 1. Denaturation at 94° C for 1 minute 2. Annealing at 50° C for 1 minute 3. Elongation at 72° C for 1,5 minutes
Step 3	Final extension at 72° C for 20 minutes

Amplification success rates were calculated for each marker as the ratio of the number of successfully amplified samples with the total number of PCR using the corresponding marker regardless the number of PCR repetition in order to obtain successful amplification.

3.4.3 DNA sequencing

To obtain purified DNA for sequencing, the PCR products were separated in agarose gels by electrophoresis. The DNA fragments in the agarose gel were sliced with a razor and then purified using the GENECLEAN® Kit (MP Biomedicals, Illkirch, France).

Table 3.4 Reaction mixture of PCR sequencing reagent

Reagent	Volume (µl)
Big Dye	1
5X buffer	1,5
Forward/reverse primer (5pmol/ml)	1
H2O	4,5
Template DNA (5-10 ng)	2

Table 3.5 Temperature profiles for PCR sequencing reaction

Step	Condition
Step 1	Initial denaturation at 96^{o} C for 1 minutes
Step 2	35 cycles of: 1. Denaturation at 96^{o} C for 10 seconds 2. Annealing at 45^{o} C for 10 seconds 3. Elongation at 60^{o} C for 4 minutes
Step 3	Final extension at 72^{o} C for 20 minutes

The sequencing reactions were performed using the ABI Prism™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit v1.1 (Applied Biosystems), based on the principles described by Sanger et al (1977). Data were collected from capillary electrophoresis on an ABI Prism 3100® Genetic Analyzer with the Sequence Analysis Software v3.1 (Applied Biosystems). The sequencing was performed with the same primers used for amplification in both directions. The

sequencing reaction mixture is shown in Table 3.4, while the temperature profiles of the PCR for sequencing are shown in Table 3. 5.

Sequencing success rates were calculated for each marker as the ratio of the number of bi-directional consensus sequences that were successfully obtained with the total number of successfully amplified samples regardless the number of repetition in order to obtain successful sequencing.

3.5 Sequence analysis

3.5.1 Sequence editing

Sequence editing was done to ensure the DNA sequences are as accurate as possible. CodonCode Aligner™ software was used to trim the sequence ends (the first and last 20 bp should contain less than 2 nucleotides showing quality value (QV) less than 20) and to assemble the forward and reverse sequences from each investigated samples. Sequences that failed to be assembled into bi-directional consensus sequences were excluded from the data set. Every successfully assembled contig then was checked for base call disagreements and ambiguities and manually edited where necessary.

Each of these edited sequences then was assigned to a particular taxon by comparing it with the nucleotide sequences in GenBank database and Barcode of Life Database (BOLD) using Basic Local Alignment Search Tool (BLAST, Altschul et al 1990). When BLAST result showed high E-value, the corresponding sequence was removed from the data set. The lower the E-value, the more similar the query sequence to the reference sequence in the database (Madden, 2002).

The results of sequence identification then were cross-checked with the morphological identification results. The match between morphology and molecular identification results were counted into three levels: species, genus, and family. When the species name from the molecular identification matched the species name from the morphological identification, then

it was counted as correct species. When the assignment result only matched the genus or family, then it was counted as correct genus or family. When the results between morphological and molecular identification did not match, it was counted as incorrect identification if the assignment results with both matK and rbcL were similar (at least on family level), or it was counted as mislabeling/contamination if the results matK and rbcL were different. Herbarium specimens were double-checked in cases of incorrect identification.

The quality of contig was determined according to the CBOL (CBOL Plant Working Group 2009). Quality parameters including mean of contig length, mean QV and forward and reverse sequences overlap. In addition, the percentage of high quality contig (mean QV $\geq$ 30, post-trim length >100 bp for both forward and reverse sequences, minimum 50% forward-reverse sequences overlap) and low-quality contig (mean QV < 20) were determined.

3.5.2 Sequence alignment

Sequence alignment was carried out independently for each marker in two stages. First, multiple sequences were aligned according to their families using the ClustalW program (Thompson et al 1994) embedded in MEGA6 (Tamura et al 2013). Reference sequences were downloaded from BOLD/GenBank and included in the alignment for those species represented with only one sample. The alignment results were subsequently checked for the occurrence of ambiguities caused by the presence of indels and/or substitutions and edited where necessary. In the second stage, all aligned sequences from each family were manually aligned with sequences from other families. Gaps were added where necessary and the final alignment were trimmed at both ends. The aligned sequences of rbcL and matK were combined to obtain two-loci DNA barcodes using SequenceMatrix software (Vaidya et al 2011) and then the concatenated alignment was exported as NEXUS files.

3.5.3 Calculation of identification success

Identification success was calculated by best-close match method as implemented in TaxonDNA (Meier et al 2006). A threshold value T was determined for each dataset as a divergences

percentage below which 95% of all intra-specific distances were found. In this method, all recovered sequences at each of the three markers (matK, rbcL, matK+rbcL) were formatted as both database and query; a query can only be identified if the corresponding sequence has a match in the dataset that falls into the 0% to T% interval. All queries without such a match would remain unidentified. If the name was identical, the query was considered an identification success. The identification was considered incorrect when the name was mismatched and considered ambiguous when several equally good best matches were found that belonged to different species.

3.5.4 DNA barcode gaps analysis

This analysis only included the species with at least two representatives. DNA barcode gaps were evaluated by comparing the distribution of intra- versus inter-specific divergences (Meyer and Paulay 2005). Matrices of pair-wise distances were created using MEGA6 (Tamura et al 2013) based on the single-locus alignment and two-loci alignment. The distance calculation was following Tamura-Nei model (1993) assuming the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies with Gamma-distributed rates between sites and the pattern between lineages were assumed to be heterogeneous. The calculation results of intra-specific divergence and inter-specific divergence in these matrices then were separated using ExcaliBAR (Aliabadian et al 2014) to facilitate the measures of distance range and distance mean of each type of divergence. Frequency (%) distribution of intra-specific and inter-specific divergences of each marker was calculated and depicted in graphics using Excel to see if there were barcode gaps existed between the intra-specific and inter-specific divergences.

Furthermore, one-way ANOVA was conducted to estimate the significance of mean difference of the intra-specific and inter-specific divergence between markers and unpaired t-test to estimate the significance of mean difference between intra-specific and inter-specific divergences of each marker.

Based on the genetic distance, discriminatory power was calculated as the proportion of discriminated sequences/species with the total number of sequences/species included in the analysis. A sequence was considered as discriminated when the genetic distance of this sequence with the other sequence of the same species was lower than the genetic distance with other sequences of the other species.

3.5.4 Phylogenetic tree reconstruction and analysis

Based on the aligned sequences, phylogenetic trees were reconstructed using MEGA6 (Tamura et al 2013) with three different algorithms: maximum parsimony (MP), maximum likelihood (ML), and neighbor joining (NJ). Phylogenetic trees were reconstructed based on this two-loci alignment in the same way as the phylogenetic trees reconstruction of single-locus alignment. Percentages of species, genus, and family monophyletic clade were calculated from each reconstructed tree.

Furthermore, ordinal phylogenies were reconstructed based on Maximum Likelihood trees of each used marker and then were compared to APG III (The APG III 2009) ordinal phylogenies to see if there were inconsistencies between these two topologies.

4 RESULT

4.1 Morphologically verified samples

The morphological identification of the samples was done as this thesis was being written. Up to July 2015, approximately 4.160 samples (80% of total samples) has been verified as 1.100 morphologically defined species spanning 436 genera, 136 families, and 49 orders. Families with highest sample numbers are Rubiaceae, Annonaceae, Fabaceae, Phyllanthaceae, Myrtaceae, Euphorbiaceae, Lauraceae, Moraceae, and Burseraceae.

Table 4.1 Sample composition of the dominant families in the study sites

Family	Number of genera	Number of species	Number of samples
Anacardiaceae	8	15	49
Annonaceae	22	53	276
Apocynaceae	19	26	105
Araceae	8	11	85
Arecaceae	4	5	165
Asteraceae	12	12	48
Burseraceae	4	20	181
Cannabaceae	2	5	33
Connaraceae	5	11	39
Dipterocarpaceae	5	12	69
Euphorbiaceae	21	45	188
Fabaceae	27	48	250
Lamiaceae	8	17	74
Lauraceae	13	47	186
Malvaceae	17	33	155
Meliaceae	9	26	104
Menispermaceae	15	17	64
Moraceae	6	33	186
Myristicaceae	4	21	112
Myrtaceae	4	27	213
Phyllanthaceae	10	53	243
Piperaceae	1	3	38
Poaceae	9	13	107
Rubiaceae	29	65	360
Rutaceae	11	18	68
Sapindaceae	12	25	83
Sapotaceae	6	15	54
Vitaceae	7	12	62

4.2 DNA sequence recoverability and quality

From all 5.128 samples, only 3.932 samples were selected for investigation. Due to time restriction, only 2.590 samples were finally included in the study.

DNA materials were extracted from dried-leaf specimens with no noticeable difficulties. The amplification and sequencing, however, turned to be more problematic especially when using matK primers. Recoverability of DNA sequences for rbcL was high in overall (amplification and sequencing success were 96,91% and 94,66%, respectively). Meanwhile, the amplification and sequencing results using the primer of matK was only moderately successful (79,05% and 65,81%, respectively).

Table 4.2 Amplification and sequencing successfulness of two regions

Number of samples for investigation	5.238	
Number of extracted samples (up to April 2015)	3.932	
	matK	rbcL
Successful amplification	1.834	2.510
Unsuccesful amplification	486	80
Successful sequencing	1.207	2.376
Unsuccesful sequencing	627	140

The amplification and sequencing success rate of 28 dominant families are shown in Figure 4.1 and Figure 4.2. All of these families were amplified successfully more than 90% with rbcL. The same level of amplification success rate was obtained with matK for only four families, namely Sapotaceae, Sapindaceae, Connaraceae, and Dipterocarpaceae. Meanwhile, the amplification success rate for the other families ranged between 60 to 80 percent with Rubiaceae, Arecaceae, and Poaceae at the lowest. The sequencing success hardly reached 90% except for Dipterocarpaceae with rbcL. Some families, namely Rubiaceae, Poaceae, Arecaceae and Araceae have an extremely low level of sequencing success rate (<20%) using matK.

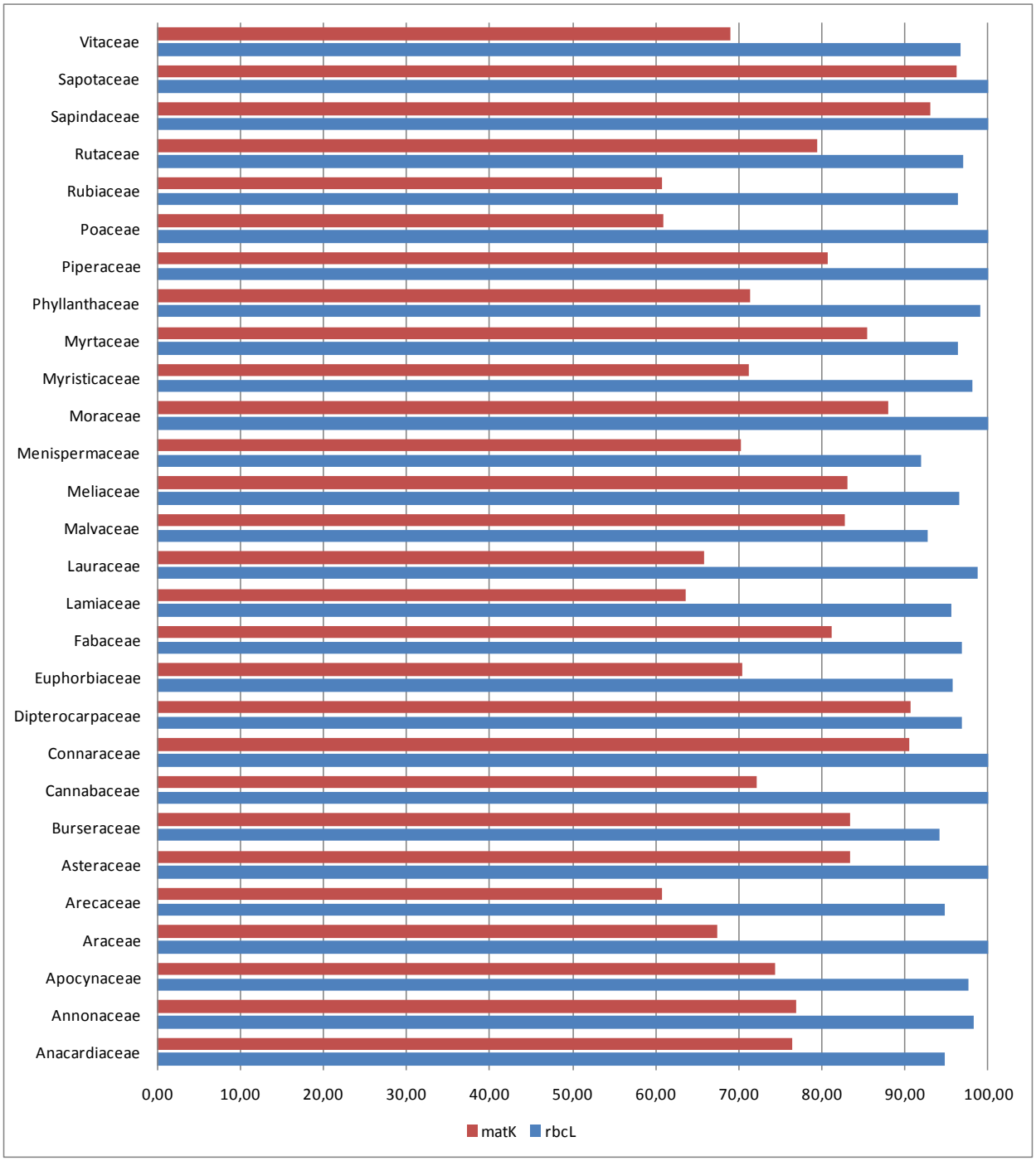


Figure 4.1 Amplification success rate (%) of the dominant families

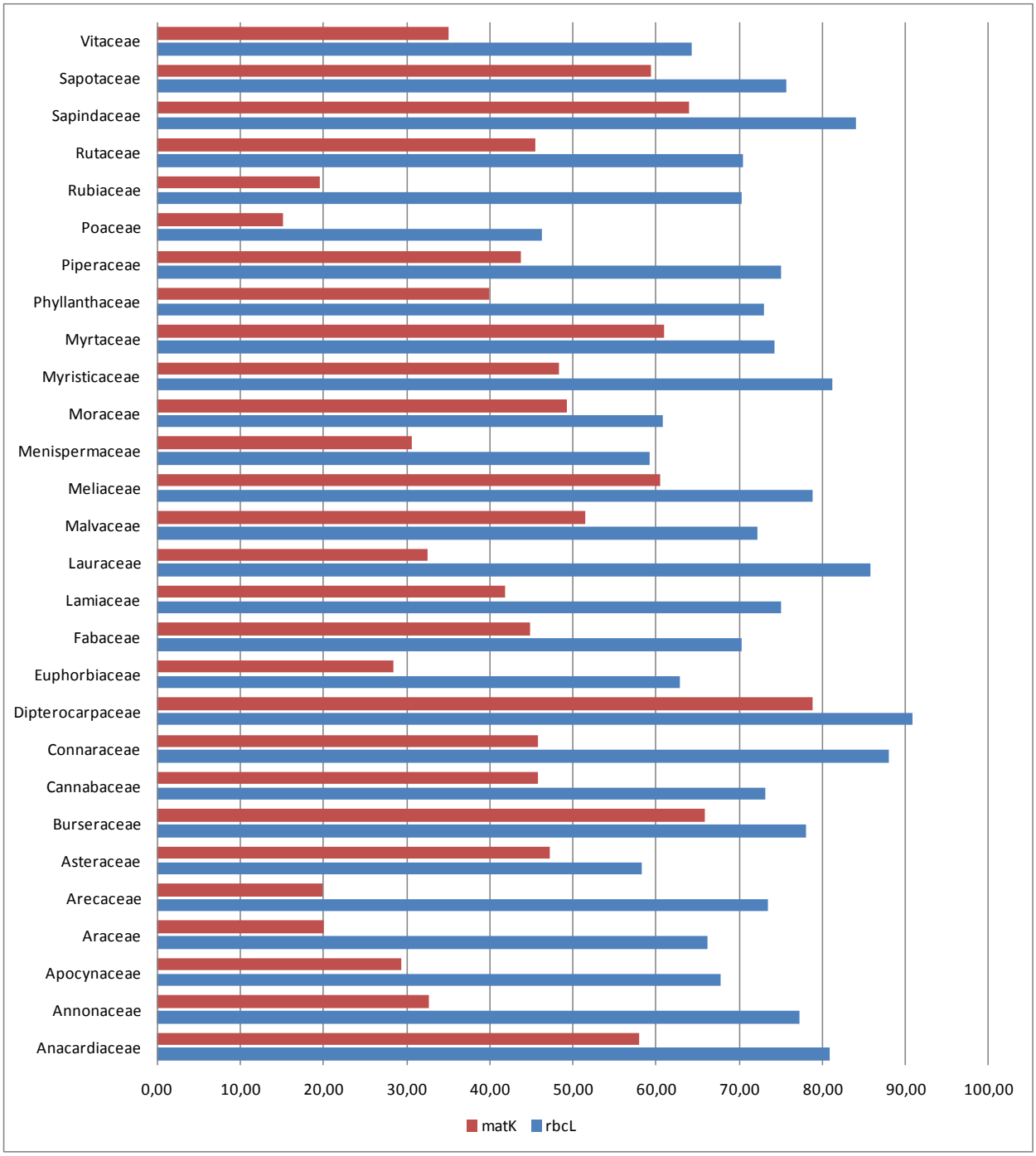


Figure 4.2 Sequencing success rate (%) of the dominant families

A total of 1.207 barcode sequences were generated from matK representing 424 species of 97 families of 40 orders, and 2.376 barcodes from rbcL representing 539 species of 126 families of 44 orders. When the number of successfully generated sequences compared to the number of amplification has been made for each region, then the recovery rate was 52% for matK and 92% for rbcL.

Overall sequence quality for the rbcL region was high: 98,95% of the successfully sequenced samples had greater than 50% contig overlap, and 41,21% were evaluated as high-quality sequences following the criteria from CBOL Plant Working Group (2009). Using the same criteria, the sequencing success was lower for matK: 81,08% for sequences with contig overlap >50%, and only 25,23% were high-quality sequences. Nevertheless, the mean of sequence quality value of both regions was in the same level.

Table 4.3 Summary of the sequence quality of matK and rbcL regions

	matK	rbcL
Mean of sequences length (bp) [SD]	696 [73]	599 [50]
Percentage of contigs with overlap >50%	81,08	98,95
Mean of sequences quality value [SD]	37 [14]	37 [11]
Percentage of sequences with low quality bases <1%	74,32	82,64
Percentage of sequences with substitutions and internal gaps <1%	50	74,28
Percentage of high quality sequences	25,23	41,21

4.3 Comparison between morphological and molecular identification

Molecular identification was conducted for all samples that were successfully barcoded, but only samples that have been morphologically identified were included in the further analysis, thus these two identification methods were comparable. Since the morphological identification was on progress as this thesis was being completed, there were a considerable amount of sequences that the identification was not yet confirmed morphologically (200 sequences of matK and 529 sequences of rbcL). Therefore, these sequences were excluded from the dataset.

For both regions, the highest match between morphological and molecular identification was on genus level. The matched identification on species-level was higher with matK than with rbcL. Incorrect identification was relatively low for both regions.

Table 4.4 Comparison between morphological identification and molecular identification

Percentage of matched identification	matK	rbcL
Species level	30,15	22,44
Genus level	46,64	51,28
Family level	12,06	17,76
Incorrect identification	3,64	3,33
Mislabeling/contamination	7,51	5,19

Table 4.4 shows there a relatively high incidence of mislabeling/contamination which might happen during the fieldwork and laboratory work. In the case of incorrect identification and mislabeling/contamination, sequences were again excluded from the dataset. And since the study aims to compare the performance of matK and rbcL and to generate two-loci barcodes, only samples from which both matK and rbcL sequences were successfully recovered were included in the further analysis. Consequently, only 322 samples from 161 species (each was represented by two samples) were included in best-close match and barcode-gap analysis and 334 samples from 334 species (one representation for each species) were included in phylogenetic analysis.

Figure 4.3 shows the matched results between morphological and molecular identification of samples from dominant families. The highest species-matched identification was obtained in family Cannabaceae using matK (>70%). The molecular identification using matK did not always result in a higher species-level match compared to rbcL, as shown here by family Arecaceae, Asteraceae, Menispermaceae, Poaceae, Rutaceae, and Vitaceae.

In the further analysis, each sequence included in the dataset was named according to the morphological identification result with the assumption that this identification was more accurate since the herbarium vouchers were cautiously compared with reliable reference vouchers.

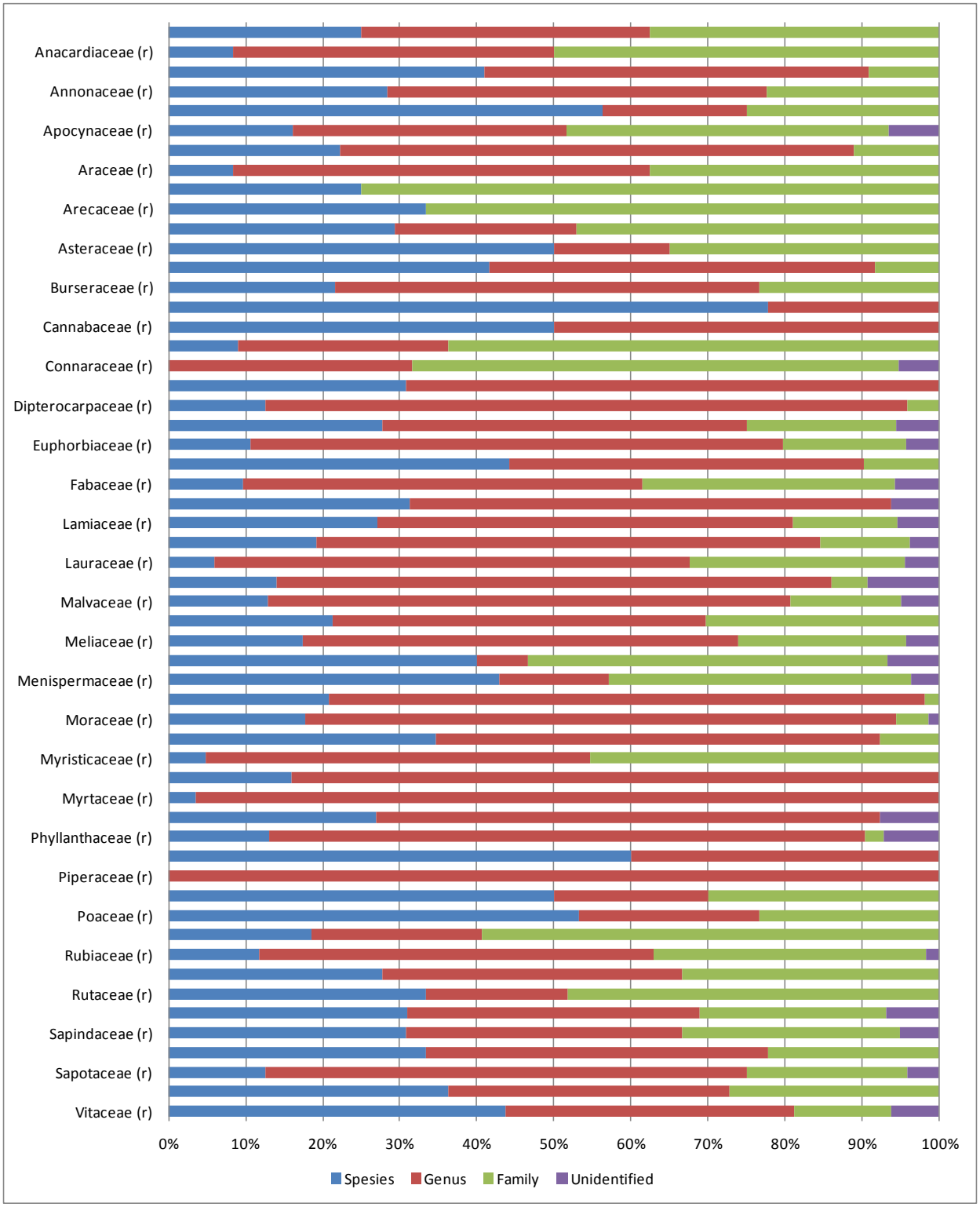


Figure 4.3 Matched identification of the dominant families with matK (m) and rbcL (r)

4.4 Identification success assessment using best-close match method

According to the distribution of the intra-specific divergences of the dataset, the threshold values for matK, rbcL, and the combination of both markers were estimated to be equal, that was 0,6% in which below this value 95% of all intra-specific divergences were found. The best-close match test was conducted using this threshold which results were shown in Table 4.5.

Table 4.5 The best-close match test result with 0,6% threshold

Identification results	Number of sequences		
	matK	rbcL	matK+rbcL
Correct	252 (78,25%)	230 (71,42%)	261 (81,05%)
Ambiguous	50 (15,52%)	76 (23,60%)	35 (10,86%)
Incorrect	9 (2,79%)	11 (3,41%)	14 (4,34%)
No match within threshold	11 (3,41%)	5 (1,55%)	12 (3,72%)

According to the best-close match results, matK has higher species identification success in overall compared to rbcL, but the highest correct species identification was obtained by the combination of both markers. Ambiguous results were relatively high for all of the markers and the combination.

Figure 4.4 shows that the species identification success was different depending on which taxa were in question. All of the sequences from the species of Asteraceae, Euphorbiaceae, Malvaceae, and Primulaceae were correctly identified using both matK and rbcL and the combined marker. Burseraceae had the lowest identification success with less than 40% for all of the markers. rbcL showed the lower performance in almost all of the given families compared to matK and matK+rbcL. Moreover, the two-loci barcode did not always improve the identification success, as happened in Annonaceae and Fabaceae.

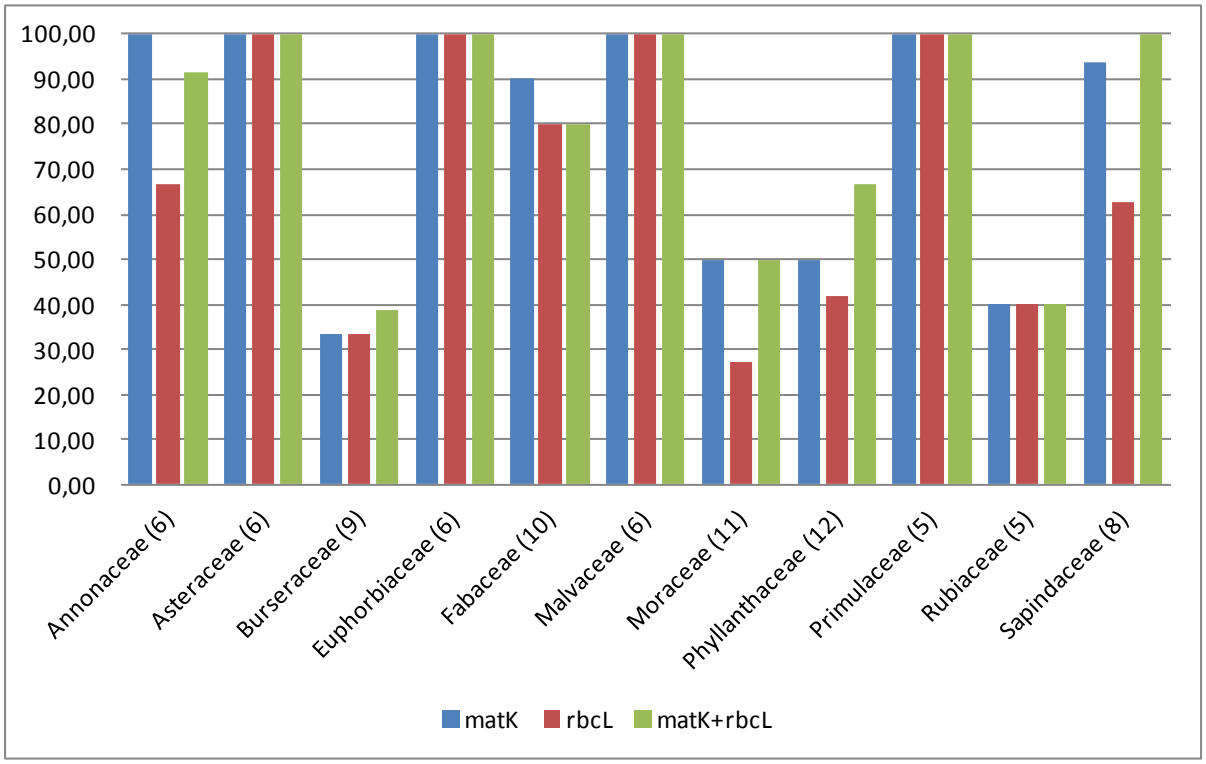


Figure 4.4 Percentage of correct identification of families with the largest dataset (number of species included in the dataset)

Table 4.6 shows 52 species from different families which failed to be correctly identified by at least one of the marker. Most of these species were identified ambiguously or incorrectly with rbcL but were identified correctly with matK and/or the combined marker. There were 22 species which were unable to be identified correctly even with all of the markers. The combined matK+rbcL was only resulting in correct identification when at least one of the markers showed a similar result.

Table 4.6 List of species with ambiguous/incorrect identification results

Family	Species	matK	rbcL	matK+rbcL
Annonaceae	*Uvaria hirsuta*	Correct	Ambiguous/Incorrect	Correct
	Uvaria lobbiana	Correct	Ambiguous/Incorrect	Correct
Apocynaceae	*Hunteria zeylanica*	Unidentified	Correct	Unidentified
Burseraceae	*Canarium gracile*	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Canarium littorale	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Santiria apiculata	Correct	Ambiguous/Incorrect	Correct
	Santiria laevigata	Ambiguous/Incorrect	Correct	Correct
	Santiria oblongifolia	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Santiria rubiginosa	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Santiria tomentosa	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Dioscoreaceae	*Dioscorea salicifolia*	Correct	Unidentified	Correct
Fabaceae	*Fordia nivea*	Correct	Unidentified	Unidentified
	Paraderris elliptica	Unidentified	Correct	Unidentified
Lamiaceae	*Clerodendrum deflexum*	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Clerodendrum disparifolium	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Lauraceae	*Litsea oppositifolia*	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Litsea umbellata	Correct	Ambiguous/Incorrect	Correct
Meliaceae	*Aglaia argentea*	Correct	Ambiguous/Incorrect	Correct
	Aglaia spectabilis	Correct	Ambiguous/Incorrect	Correct
Moraceae	*Artocarpus elasticus*	Correct	Ambiguous/Incorrect	Correct
	Ficus aurata	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Ficus grassularioides	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Ficus ribes	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Ficus sagittata	Correct	Ambiguous/Incorrect	Correct
	Ficus schwarzii	Correct	Ambiguous/Incorrect	Correct
	Ficus variegata	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Myristicaceae	*Gymnacranthera farquhariana*	Ambiguous/Incorrect	Correct	Correct
	Horsfieldia pulcherrima	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Myristica maxima	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Myrtaceae	*Syzygium borneense*	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Syzygium lineatum	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Pandaceae	*Galearia fulva*	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Galearia aristifera	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Phyllanthaceae	*Aporosa octandra*	Correct	Ambiguous/Incorrect	Correct
	Aporosa subcaudata	Correct	Ambiguous/Incorrect	Correct
	Baccaurea dulcis	Ambiguous/Incorrect	Correct	Correct
	Baccaurea pyriformis	Ambiguous/Incorrect	Correct	Correct
	Breynia racemosa	Unidentified	Ambiguous/Incorrect	Ambiguous/Incorrect
	Bridelia tomentosa	Unidentified	Correct	Unidentified
	Glochidion sericeum	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Glochidion superbum	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Phyllanthus oxyphyllus	Correct	Ambiguous/Incorrect	Correct
	Phyllanthus urinaria	Correct	Ambiguous/Incorrect	Correct
Polygalaceae	*Xanthophyllum eurhynchum*	Correct	Ambiguous/Incorrect	Correct
	Xanthophyllum wrayi	Correct	Ambiguous/Incorrect	Correct
Rhizophoraceae	*Gynotroches axillaris*	Ambiguous/Incorrect	Correct	Correct
Rubiaceae	*Spermacoce exilis*	Unidentified	Unidentified	Unidentified
	Timonius flavescens	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
	Timonius wallichianus	Ambiguous/Incorrect	Ambiguous/Incorrect	Ambiguous/Incorrect
Sapindaceae	*Dimocarpus longan*	Correct	Ambiguous/Incorrect	Correct
	Nephelium mutabile	Correct	Ambiguous/Incorrect	Correct
	Nephelium subfalcatum	Correct	Ambiguous/Incorrect	Correct

4.5 Barcode gap analysis

Intra-specific and inter-specific divergences were calculated as pair-wise distances between aligned sequences using MEGA6 (Tamura et al 2013) following Tamura-Nei model (1993). Table 4.7 shows that matK gained higher levels of intra-specific divergence and inter-specific divergence compared to rbcL. The combination of matK and rbcL has the median divergence between the two single markers.

Table 4.7 Intra-specific and inter-specific divergences of matK, rbcL and matK+rbcL

DNA barcodes	Intra-specific divergences		Inter-specific divergences	
	Range	Mean (SD)	Range	Mean (SD)
matK	0,0000 - 0,0660	0,0014 (0,0061)	0,0000 - 0,9130	0,3054 (0,0979)
rbcL	0,0000 - 0,0320	0,0008 (0,0035)	0,0000 - 0,2030	0,0964 (0,0276)
matK+rbcL	0,0000 - 0,0480	0,0011 (0,0043)	0,0000 - 0,4030	0,1836 (0,0502)

However, according to the unpaired t-test result of each marker (Table 4.8), the mean values of intra-specific divergence were extremely different with the mean values of inter-specific divergence.

Table 4.8 Unpaired t-test between intra-specific and inter-specific divergences of matK, rbcL, and matK+rbcL

Marker	Intra-specific divergence vs inter-specific divergence	
	Mean difference	p value
matK	0,3040	<0,0001***
rbcL	0,0956	<0,0001***
matK+rbcL	0,1825	<0,0001***

The one-way ANOVA result in Table 4.9 shows that when each marker was compared to each other, there were no significant different between the intra-specific divergences, but the inter-specific divergences were extremely different. The highest different was found between matK and rbcL.

Table 4.9 One-way ANOVA of intra-specific and inter-specific divergences of matK, rbcL, and matK+rbcL

Comparison	Intra-specific divergence		Inter-specific divergence	
	Mean difference	p value	Mean difference	p value
matK vs rbcL	0,0006	>0,05ns	0,2090	<0,0001***
matK vs matK+rbcL	0,0003	>0,05ns	0,1218	<0,0001***
rbcL vs matK+rbcL	-0,0003	>0,05ns	-0,0872	<0,0001***

The frequency (%) distribution of intra-specific and inter-specific divergence was using three markers (Figure 4.5) showed that there was no clear barcode gaps existed as the intra-specific divergences overlapped with inter-specific divergences. Barcode gap would only be clearly defined when the minimum value of inter-specific divergence is higher than the maximum value of the intra-specific divergence which in this study was not the case.

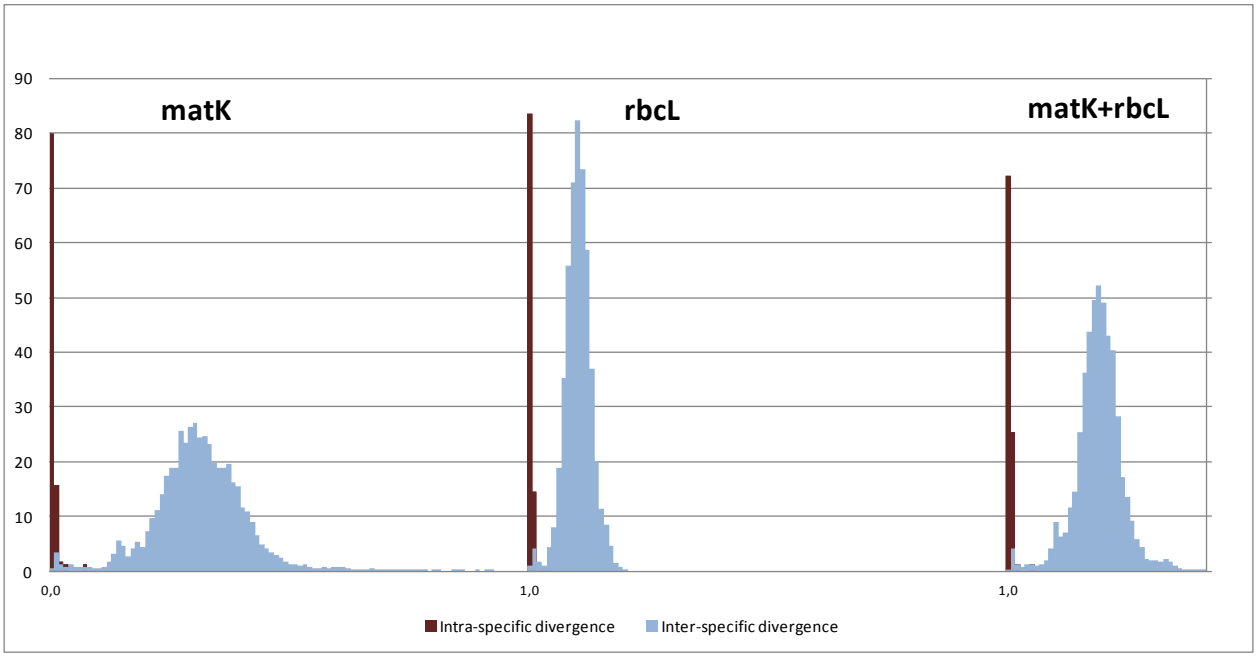


Figure 4.5 The frequency (%) distribution of intra-specific divergence and inter-specific divergence

Moreover, based on the intra-specific and inter-specific divergences of all the sequences included in the dataset, the discriminatory power of each marker were calculated; results are shown in Table 4.10. The discriminatory powers of all markers were considered high as more

than 70% of included species were successfully differentiated from each other. The matK+rbcL marker has the highest number of discriminated sequences and species compared to matK or rbcL alone. Most of the species were discriminated by the two sequences included in the dataset.

Table 4.10 Discriminatory power of matK, rbcL, and matK+rbcL

	matK	rbcL	matK+rbcL
No. of disriminated sequences (%)	257 (79,81)	234 (72,67)	287 (89,13)
No. of discriminated species (%)	132 (81,99)	117 (72,67)	147 (91,31)
No. of species discriminated by 2 sequences (%)	125 (77,64)	117 (72,67)	140 (86,96)
No. of species discriminated by 1 sequence (%)	7 (4,35)	0 (0,00)	7 (4,35)

The one-way ANOVA in Table 4.10 shows that the discriminatory power of matK was not significantly different, but the discriminatory power of the combined marker was significantly different from the single markers.

Table 4.11 One-way ANOVA of the discriminatory power of matK, rbcL, and matK+rbcL

Comparison	Discriminatory power	
	Mean difference	p value
matK vs rbcL	0,0714	>0,05ns
matK vs matK+rbcL	-0,0932	<0,01**
rbcL vs matK+rbcL	-0,1646	<0,001***

Forty four out of 161 species (Table 4.12) were unable to be discriminated by rbcL and eleven of them were not able to be discriminated by any of the markers included the two-loci barcode. These species were mostly from species-rich genera, such as *Ficus, Santiria, Litsea*.

Table 4.12 List of species which were not able to be discriminated by at least one marker (yes=discriminated, no = not discriminated)

Family	Species name	matK	rbcL	matK+rbcL
Annonaceae	*Uvaria hirsuta*	Yes	No	Yes
	Uvaria lobbiana	No	No	Yes
Burseraceae	*Canarium gracile*	No	No	No
	Canarium littorale	Yes	No	Yes
	Santiria laevigata	No	Yes	Yes
	Santiria oblongifolia	No	No	No
	Santiria rubiginosa	No	No	No
	Santiria tomentosa	No	No	No
	Santiria apiculata	Yes	No	Yes
Dipterocarpaceae	*Shorea parvifolia*	No	Yes	Yes
	Shorea acuminata	No	Yes	Yes
Fabaceae	*Fordia nivea*	Yes	No	Yes
	Kunstleria ridleyi	Yes	No	Yes
Lamiaceae	*Clerodendrum deflexum*	No	No	Yes
	Clerodendrum disparifolium	No	No	Yes
Lauraceae	*Litsea oppositifolia*	No	No	Yes
	Litsea umbellata	Yes	No	Yes
Meliaceae	*Aglaia argentea*	Yes	No	Yes
	Aglaia spectabilis	Yes	No	Yes
Moraceae	*Artocarpus elasticus*	Yes	No	Yes
	Artocarpus rigidus	No	No	Yes
	Ficus aurata	No	No	No
	Ficus grossularioides	No	No	Yes
	Ficus ribes	No	No	Yes
	Ficus variegata	No	No	No
	Ficus schwarzii	Yes	No	No
	Ficus sagittata	Yes	No	Yes
Myristicaceae	*Gymnacranthera farquhariana*	No	Yes	Yes
	Horsfieldia pulcherrima	No	No	Yes
	Myristica maxima	No	No	Yes
Myrtaceae	*Syzygium borneense*	No	No	No
	Syzygium lineatum	No	No	Yes
Pandanaceae	*Galearia aristifera*	No	No	Yes
	Galearia fulva	No	No	No
Phyllanthaceae	*Baccaurea dulcis*	No	Yes	Yes
	Baccaurea pyriformis	No	Yes	Yes
	Breynia racemosa	Yes	No	Yes
	Glochidion superbum	No	No	No
	Glochidion sericeum	No	No	Yes
	Aporosa octandra	Yes	No	Yes
	Aporosa subcaudata	Yes	No	Yes
	Phyllanthus oxyphyllus	Yes	No	Yes
	Phyllanthus urinaria	Yes	No	Yes
Polygalaceae	*Xanthophyllum eurhynchum*	Yes	No	Yes
	Xanthophyllum wrayi	Yes	No	Yes
Rubiaceae	*Timonius flavescens*	No	No	No
	Timonius wallichianus	No	No	No
Sapindaceae	*Nephelium cuspidatum*	Yes	No	Yes
	Dimocarpus longan	Yes	No	Yes
	Nephelium subfalcatum	Yes	No	Yes

4.6 Phylogenetic analysis

Nine phylogenetic trees (see Appendix 4.1 - 4.9) were reconstructed based on multiple sequence alignments of matK, rbcL, and matK+rbcL using three different methods: Maximum Parsimony (MP), Neighbor Joining (NJ), and Maximum Likelihood (MP). Each tree was observed and similar topologies were found amongst these trees.

Table 4.13 Percentage of monophyletic clades recovered in nine reconstructed phylogenetic trees

Region	Monophyletic with support value >70%								
	Maximum Parsimony (MP)			Neighbor Joining (NJ)			Maximum Likelihood (ML)		
	Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
matK	95,92	68,42	73,91	93,88	66,67	69,57	97,96	64,91	68,94
rbcL	95,92	63,16	60,25	93,88	63,16	63,98	89,90	63,16	55,90
matK+rbcL	100,00	71,93	73,29	100,00	64,91	73,91	100,00	70,18	75,16

Monophyletic percentages were calculated in family level, genus level, and species level (Table 4.13). Seventeen families were not included in the calculation of family-level monophyletic percentage as these families were presented with only one taxon. The monophyletic percentage was highest in family level and lowest in genus level, regardless the marker used. The only exception was found in MP and ML trees using rbcL where the monophyletic percentage in genus level was slightly higher than in species level. The two-loci marker provided 100% taxonomic resolution in family-level with all three different methods.

Species-level monophyletic percentages were calculated from phylogenetic trees reconstructed based on two-sequences-per-sample dataset as explained in chapter Materials and Methods (trees are not shown). Trees with matK and matK+rbcL have a higher monophyletic percentage in species level compared to rbcL.

Table 4.14 shows the genera that were failed to form monophyletic clades. There were fifteen non-monophyletic genera that were found in all nine trees, namely: *Dacryodes, Canarium*, and *Santiria* from Burseraceae, *Horsfeldia* and *Myristica* from Myristicaceae, *Croton* from Euphorbiaceae, *Artocarpus* from Moraceae, *Sterculia* from Malvaceae, *Shorea* from Dipterocarpaceae, *Nephelium* from Sapindaceae, *Litsea* from Lauraceae, *Fordia* from Fabaceae,

Alstonia from Apocynaceae, *Panicum* from Poaceae, and *Rourea* from Connaraceae. The highest number of non-monophyletic genus was found in NJ phylogenetic tree using matK, meanwhile the lowest number was found in NJ tree of matK+rbcL.

Table 4.14 Non-monophyletic genera (v) found in nine phylogenetic trees

Family	Genus	matK			rbcL			matK+rbcL		
		ML	MP	NJ	ML	MP	NJ	ML	MP	NJ
Burseraceae	Dacryodes	v	v	v	v	v	v	v	v	v
	Canarium	v	v	v	v	v	v	v	v	v
	Santiria	v	v	v	v	v	v	v	v	v
Myristicaceae	Horsfeldia	v	v	v	v	v	v	v	v	v
	Myristica	v	v	v	v	v	v	v	v	v
Euphorbiaceae	Croton	v	v	v	v	v	v	v	v	v
Moraceae	Artocarpus	v	v	v	v	v	v	v	v	v
Malvaceae	Sterculia	v	v	v	v	v	v	v	v	v
Dipterocarpaceae	Shorea	v	v	v	v	v	v	v	v	v
Sapindaceae	Nephelium	v	v	v	v	v	v	v	v	v
Lauraceae	Litsea	v	v	v	v	v	v	v	v	v
Fabaceae	Fordia	v	v	v	v	v	v	v	v	v
Apocynaceae	Alstonia	v	v	v	v	v	v	v	v	v
Poaceae	Panicum	v	v	v	v	v	v	v	v	v
Connaraceae	Rourea	v	v	v	v	v	v	v	v	v
Rubiaceae	Spermacoce	v	v		v	v	v	v	v	
	Nauclea				v	v	v			
Phyllanthaceae	Aporosa	v		v						
Rhizophoraceae	Pellacalyx			v						
Primulaceae	Ardisia			v						

The non-monophyletic species (Table 4.15) mostly came from dominant families, such as Burseraceae, Myristicaceae, Moraceae, Phyllanthaceae, Lauraceae, Sapindaceae, and Annonaceae. Most of these species were found to be non-monophyletic in phylogenetic trees based on rbcL. Twenty-one species were non-monophyletic in all of the phylogenetic trees.

Table 4.15 Non-monophyletic species (v) found in nine phylogenetic trees

Family	Species	matK			rbcL			matK+rbcL		
		ML	MP	NJ	ML	MP	NJ	ML	MP	NJ
Lamiaceae	*Clerodendrum deflexum*	v	v	v	v	v	v	v	v	v
	Clerodendrum disparifolium	v	v	v	v	v	v	v	v	v
Lauraceae	*Litsea oppositifolia*	v	v	v	v	v	v	v	v	v
	Litsea robusta	v	v	v					v	v
	Litsea umbellata				v	v				
Myristicaceae	*Horsfieldia pulcherrima*	v	v	v	v	v	v	v	v	v
	Myristica maxima	v	v	v	v	v	v	v	v	v
	Gymnacranthera farquhariana	v	v	v	v	v		v	v	v
Phyllanthaceae	*Glochidion sericeum*	v	v	v	v	v	v	v	v	v
	Glochidion superbum	v	v	v	v	v	v	v	v	v
	Aporosa octandra				v	v	v			
	Aporosa subcaudata				v	v	v			
	Baccaurea dulcis	v	v	v	v	v		v	v	
	Baccaurea mollis				v	v				
	Baccaurea pyriformis	v	v	v						
	Phyllanthus oxyphyllus	v			v	v	v			
	Phyllanthus urinaria				v	v	v			
Dipterocarpaceae	*Shorea acuminata*	v	v	v	v	v	v	v	v	v
Myrtaceae	*Syzygium cf. borneense*	v	v	v	v	v	v	v	v	v
	Syzygium lineatum	v	v		v	v	v	v	v	v
Pandaceae	*Galearia aristifera*	v	v	v	v	v	v	v	v	v
	Galearia fulva	v	v	v	v			v	v	v
Moraceae	*Artocarpus elasticus*	v	v	v	v	v	v	v	v	v
	Artocarpus rigidus	v	v	v	v	v	v	v	v	v
	Artocarpus integer				v	v	v			
	Ficus grossularioides	v	v	v	v	v	v	v	v	v
	Ficus sagittata	v	v	v	v	v	v	v	v	v
	Ficus variegata	v	v	v	v	v	v	v	v	v
	Ficus aurata	v	v	v	v	v	v	v	v	
	Ficus ribes	v	v		v	v	v	v	v	v
	Ficus schwarzii				v	v	v			
Burseraceae	*Canarium cf. gracile*	v	v	v	v	v	v	v	v	v
	Canarium littorale				v	v	v	v	v	v
	Santiria laevigata	v	v	v	v	v	v	v	v	v
	Santiria oblongifolia	v	v	v	v	v	v	v	v	v
	Santiria rubiginosa	v	v	v	v	v	v	v	v	v
	Santiria tomentosa	v	v	v	v	v	v	v	v	v
	Santiria apiculata				v	v	v			
	Dacryodes rostrata	v								
Rubiaceae	*Timonius cf. flavescens*	v	v	v	v	v	v	v	v	v
	Timonius wallichianus	v	v	v	v	v	v	v	v	v
Stemonuraceae	*Gomphandra quadrifida*					v				
Ebenaceae	*Diospyros truncata*		v							
Primulaceae	*Ardisia korthalsiana*	v	v	v						
	Ardisia pterocaulis				v	v	v			
Sapotaceae	*Palaquium obovatum*				v	v				
Fabaceae	*Fordia nivea*				v	v	v			
	Kunstleria ridleyi				v	v	v			

Table 4.15 (continued)

Family	Species	matK			rbcL			matK+rbcL		
		ML	MP	NJ	ML	MP	NJ	ML	MP	NJ
Polygalaceae	*Xanthophyllum eurhynchum*				v	v	v			
	Xanthophyllum wrayi				v	v	v			
Verbenaceae	*Stachytarpheta indica*				v	v	v			
	Stachytarpheta jamaicensis	v	v	v						
Annonaceae	*Fissistigma manubriatum*				v	v				
	Uvaria hirsuta				v	v	v			
	Uvaria lobbiana	v	v		v	v	v	v	v	
Euphorbiaceae	*Croton oblongus*				v	v				
Rhizophoraceae	*Gynotroches axillaris*				v	v	v			
	Pellacalyx lobbii	v	v							
Malvaceae	*Leptonychia caudata*				v	v				
	Scaphium affine				v	v				
Dilleniaceae	*Tetracera indica*				v	v				
Meliaceae	*Aglaia argentea*				v	v	v			
	Aglaia spectabilis				v	v	v			
	Dysoxylum excelsum				v	v				
Rutaceae	*Luvunga eleutherandra*				v	v				
Sapindaceae	*Dimocarpus longan*					v	v			
	Nephelium cuspidatum	v	v	v						
	Nephelium laurinum				v	v	v			
	Nephelium rubescens				v	v	v			
	Pometia pinnata	v	v	v						

The ordinal topologies of flowering plants are presented in Figure 4.6 – 4.8, showing the relationship between orders of flowering plants and the grouping of these orders. These ordinal topologies were reconstructed based on ML phylogenetic trees of each marker and were compared to the ordinal topologies according to APG III (The APG III 2009).

The matK marker misplaced Myrtales and failed to separate Laurales from Magnoliales. Meanwhile, the rbcL marker misplaced Aquifoliales and grouped Malpighiales and Brassicales into one monophyletic clade. This marker also failed to make Santalales a monophyletic clade. However, this marker was successful in separating Laurales from Magnoliales. Finally, the combination of matK and rbcL improved the topologies of the tree and succeeded to put all the orders into the right position.

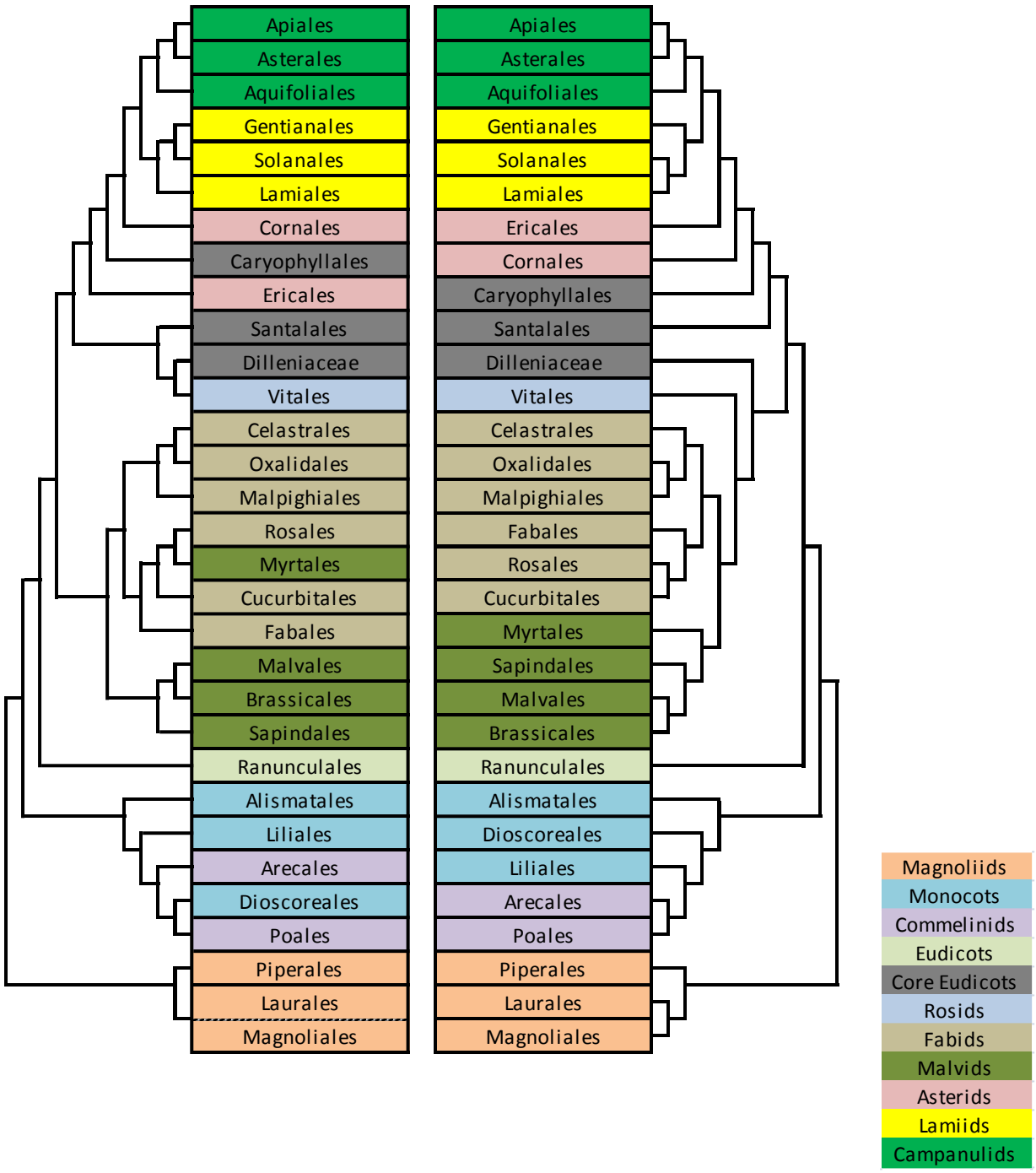


Figure 4.6 Comparison of ordinal topologies of the APG III phylogenetic tree (right) with ML phylogenetic tree generated based on matK marker (left)

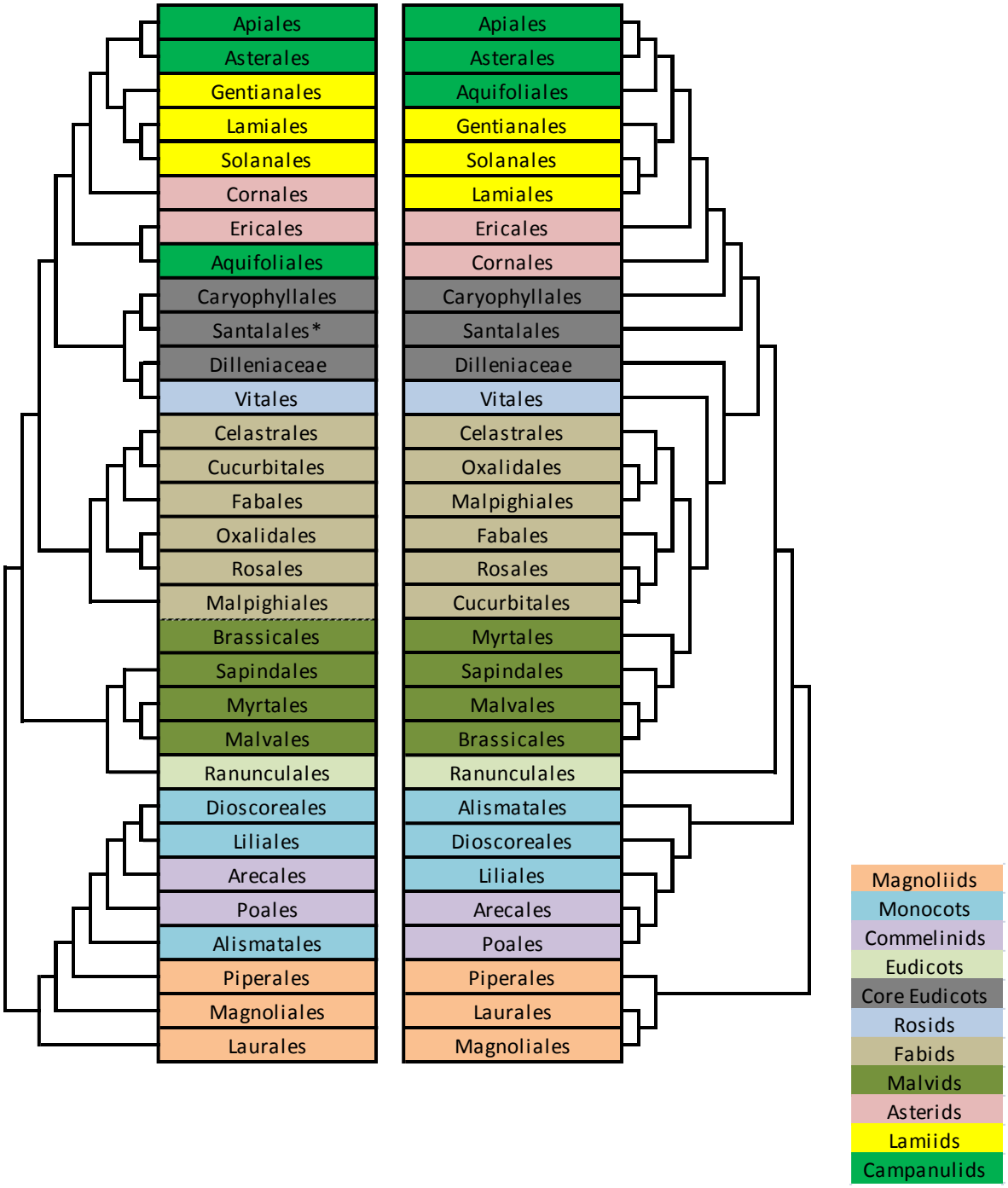


Figure 4.7 Comparison of ordinal topologies of the APG III phylogenetic tree (right) with ML phylogenetic tree generated based on rbcL marker (left; *Santalales was non-monophyletic clade)

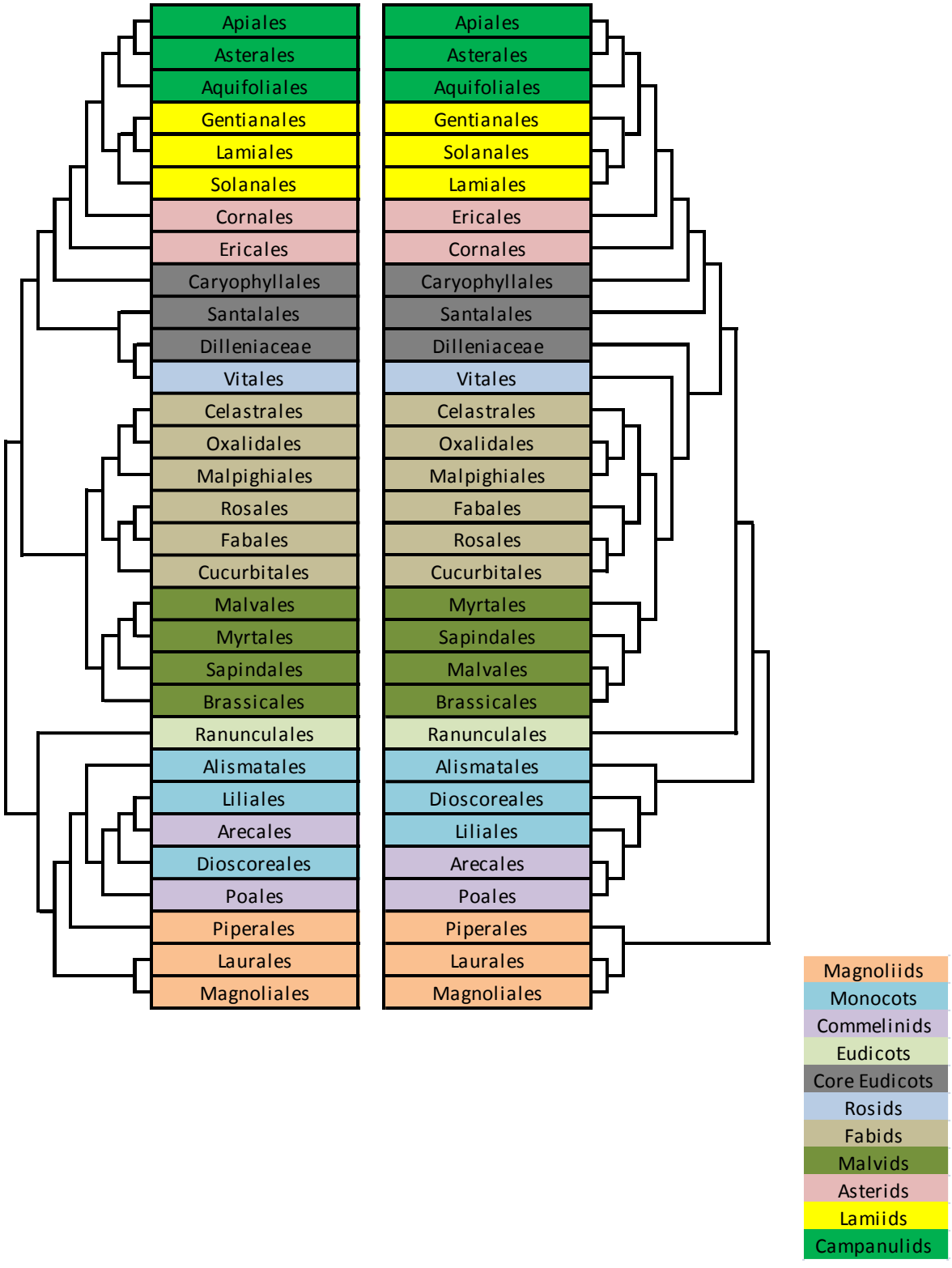


Figure 4.8 Comparison of ordinal topologies of the APG III phylogenetic tree (right) with ML phylogenetic tree generated based on matK+rbcL marker (left)

5 DISCUSSION

5.1 The matK and rbcL sequences recoverability and quality

The rbcL universality as DNA barcode was observed in many studies (e.g., CBOL Plant Working Group 2009, Hollingsworth et al 2009a, 2009b) and was confirmed by the result of this study. There was no observable problem in performing PCR using a universal primer of rbcL and to obtain bi-directional consensus sequences. This study is consistent with other similar studies (e.g., Lahaye et al 2008, Gonzales et al 2009, Parmentier et al 2013) that DNA sequences were easily obtained with rbcL primers from a wide range of tropical plant species.

In this study, the amplification and sequencing success rate with matK was lower compared to rbcL, as many similar studies have shown (Kress and Erickson 2007, Chen et al 2010, CBOL Plant Working Group 2009, Hollingsworth et al 2009a, 2009b). The tropical flora seems more difficult to be amplified using matK as shown by this study and a study by Gonzales et al 2009, compared to temperate flora (e.g., de Vere et al 2012, Bruni et al 2012). This might be due to a higher rate of evolution in the tropical flora compared to the temperate flora (Gillman et al 2010).

The PCR of matK were performed in this study using two pairs of primer which were found to be effective to generate DNA barcodes from specific taxa, such as Tetrastigma spp. (Fu et al 2011), Hedyotis spp. (Guo et al 2011), Asteraceae (Gao et al 2010). But these primers became less effective when they were used for a wide range of species, as shown in this study and other studies (Kress et al 2010, Gonzales et al 2009). A certain primer pair did not always yield a PCR product in all members of a group of seemingly closely related taxa, indicating that the primer themselves are not conserved.

Despite the rigorous efforts in performing PCR and sequencing, the matK recoverability rate was hardly improved (52%). Several studies showed higher recoverability rate for matK (e.g. 69% by Kress et al 2010, 80% by Burgess et al 2011, 91% by Dong et al 2015). However, the term of sequence recoverability was not clearly defined in those studies; the reported success

rates might actually represent the sequencing success rate only. In this study, the sequence recoverability rate was calculated as the proportion of the successful sequencing to the total amplification.

The use of matK as a barcode has been criticized mainly because no universal primers were available (Chase et al 2007). A study by Fazekas et al (2008) showed a relatively high rate of sequencing success for this marker after using up to 10 primer pairs. The usefulness of matK primers is proven when they are used in specific taxa, such as spices (De Mattia et al 2011), tea plants (Stoeckle et al 2011), palms (Jeanson et al 2011). This suggests that the performance of matK as DNA barcode may be improved by using a certain combination of primers and specific-taxa primers. In a recent review of the most optimal barcode for plants, Hollingsworth et al (2011) indicated that matK still needed optimization of primer combinations and needs to be adapted to specific taxonomic groups.

Furthermore, sequence quality was acceptable for both markers. Levels of bidirectional reads were high, averaging from 81% to 98% for rbcL and matK. This is similar to a study by de Vere et al (2012). There were considerably high percentages of sequences with low-quality bases less than 1% for both markers, but substitutions and internal gaps more than 1% were found in 50% of matK sequences. The sequences of rbcL have less proportion of substitutions and internal gaps, but in overall the mean of sequence quality value of this marker was at the same level with matK. This might be due to the mean length of the matK sequences which was approximately 100 bp longer than rbcL sequences.

Using the CBOL Plant Working Group criteria for high-quality sequences (see sub-chapter 3.5.1), the percentage of high-quality sequences of matK and rbcL were only 25% and 41%, respectively. The estimation for matK was not far from the estimation by de Vere et al (2012), but for rbcL was much lower compared to the same study. The reason for this lower estimation is unclear. Burgess et al (2011) state that DNA barcoding across a broad range of taxa with only a small number of samples per species may fail to produce high-quality contigs. Therefore, a higher number of representations per species is recommended for further studies.

5.2 Plant species identification success using matK and rbcL

5.2.1 Identification success of DNA barcoding compared to morphological identification

As one way to evaluate the species identification success, comparisons were made between morphological identification results with molecular identification results. Some authors suggested a superiority of molecular identification in comparison with morphological identification (Newmaster et al 2009, Stace 2005). Identification keys based upon morphology could be difficult to use when some features are not visible, as happens when specimens are not well developed or outside specific life stages (e.g. flowering period). Thus, the absence of some morphological features could make the identification impossible. In this study, DNA barcoding was proven to be a useful method to support the morphological identification.

However, this study showed that DNA barcoding alone is not sufficient to assign all of the DNA sequences to a correct species name. Only 22-30% of the samples were able to be correctly identified to species level. The majority of correct identifications were limited to genus level (46-51%). This indicates that DNA barcoding is useful for specimen identification at least up to genus level. Therefore, to assign specimen to a certain species name, morphological identification is still needed.

Several cases of mismatch between morphological identification results with DNA identification results were found in this study. An identification result was considered to be mismatched when matK and rbcL sequences both assigned to species names which belong to a family different from the family of the species name according to morphological identification.

Several factors could be the reason of misidentification. A sample could be misidentified when it was found to have the highest similarity to a reference sequence that was falsely identified. In this study, every sequence were compared to reference sequences from two databases (GenBank and BOLD), thus the identification results could be cross-checked one to another.

The mismatch between morphological and molecular identification could also happen when the taxonomist mistakenly identified the voucher. Morphological identification is difficult without

the presence of certain features, such as flowers or fruits, especially when dealing with species-rich groups. In the case of incorrect morphological identification, the herbarium vouchers of corresponding samples should be verified morphologically once again.

Another reason of misidentification is specimen mislabeling or contamination, either during the specimen collection or when the specimen was processed in the laboratory. For example, one sample collected during this study was morphologically identified as *Cyrtococcum oxyphyllum* (Poaceae). Meanwhile, the DNA barcodes of matK and rbcL showed this sample was actually *Dioscorea sp.* (Taccaceae). The other sample that was positioned adjacent to the previous sample in the DNA extraction plate was identified morphologically and molecularly as *Dioscorea sp.* as well. This indicates that during the DNA extraction, the latter must have been falsely labeled as the former. In another case, a sample was morphologically identified as *Vitex pinnata* (Lamiaceae). According to matK, this sample was identified as *Dianella montana* (Xanthorrhoeaceae), but rbcL confirmed the morphological identification. The adjacent sample was also identified as *Dianella sp*., indicating that in this case the latter sample was falsely located into the well labeled as the former during the amplification using matK primers. Similar cases were happened at least 60 times when one of the barcodes confirmed the morphological identification but the other barcode was resulting in misidentification. Moreover, one DNA plate of ninety-six samples was contaminated by specimen belongs to *Hanguana malayana* (Hanguanaceae). All of the samples in this plate was misidentified as *Hanguana malayana* when using rbcL but correctly identified with matK. It seems that the contamination was happening in any stage between amplification and sequencing using rbcL. When samples were suspected to be mislabeled or contaminated during lab work, the DNA of these samples should be analyzed once again.

Mislabeling/contamination during field work is untraceable but the frequency must not as high as mislabeling/contamination during lab work. Nevertheless, the number of all misidentified samples were considerably low compared to the total number of investigated samples (3%). A better sample collection and lab work management should be able to minimize the incidence of mislabeling and contamination in the further studies.

The success of species identification using DNA barcoding depends very much on the taxa in question, as much as the utilized marker. For example, in this study, the family Piperaceae resulted in high species-matched identification when using matK (60%) but no success at all when using rbcL. Meanwhile, in family Asteraceae, the species-matched identification was higher with rbcL (50%) than with matK (30%). This is consistent with a study by de Vere et al (2012), showing that DNA barcoding using different markers for different taxa will result in a different level of successful identification.

Another factor which affects the success of species identification using DNA barcoding is the availability of nucleotide data of the corresponding taxa in the DNA sequences database such as GenBank and BOLD. Of 780 species included in the analysis, 41% have no nucleotide data at all available in Genbank and BOLD databases at the time of the study. It means there was non-negligible proportion of samples which belong to species absent from the reference databases, which will increase the rates of unassigned samples and of wrong identifications. The cause of incorrect specimen assignment is more because of the incompleteness of molecular datasets rather than in the data analysis (Bruni et al 2010, Burgess et al 2011, Cowan and Fay 2012).

Little and Stevenson (2007) suggested that using a reference database in which all species are represented will provide the most reliable identification. Nowadays, GenBank is still the largest repository of sequences for all markers used in plant DNA barcoding. However, the reference sequences stored in this database show a high level of incorrect species assignment (Bidartondo 2008). BOLD, on the other hand, is largely incomplete but has already reached a good level of accuracy (Ratnasingham and Hebert 2007). Thus, an accurate and completed molecular database, especially for plant species, is still far from being achieved in the present state. Such database will hopefully be developed in the future as many studies and projects of plant DNA barcoding are going on. Nevertheless, currently, available databases will always be useful for barcoding morphologically unidentified specimens.

5.2.2 Identification success according to the best-close match analysis

Best-close match analysis with TaxonDNA (Meier et al 2006) was conducted to see if each DNA barcode generated in this study was able to differentiate the species represented by the corresponding barcode from the other species represented by other barcodes included in the dataset. In this analysis, genetic distances were calculated from each barcode against other barcodes. A genetic distance percentage of 0,6 was assigned in this study as a threshold to decide whether a pair of barcodes were matched to each other. This threshold was much lower compared to 3% threshold suggested by Hebert et al (2003b). This higher threshold was recommended especially for DNA barcoding closely-related taxa; meanwhile, this study investigated distantly-related species which likely would have a low mean of intra-specific genetic distance. Will and Rubinoff (2004) suggested that the assignment of threshold value is arbitrary and depends on the investigated species.

The percentage of correct identification, as expected, was higher when using matK than rbcL (78% and 71%, respectively). The ambiguous and incorrect identifications were also lesser with matK, but the number of unidentified sequences was higher compared to rbcL. This proves that matK performs better in species identification, though, more sequences would be likely unidentified due to its high variability. A study of DNA barcoding of Amazonian trees reported by Gonzalez et al (2009) showed a lower percentage of correct identification (34 - 44%) with 0,5% threshold. This indicates that the barcodes generated in this study have a relatively high identification success. Moreover, the matK+rbcL sequences were succeeded to identify more than 80% of the species included in the dataset (161 species). Unfortunately, there was no comparable result of a similar study using the same analysis with the combination of matK+rbcL. However, the identification success of matK+rbcL is consistent with the identification success of the same marker in a study reported by Parmentier et al (2012) which used different analysis for African rainforest plant species.

The identification success was not uniform for all of the families included in the analysis. As has been mentioned in the previous sub-chapter, the success of DNA barcoding in species identification depends on the investigated taxa and the markers used. This is clearly shown in

Figure 4.3 (sub-chapter 4.4) that different families have different identification success and different markers have different performance in each of the family. This is consistent with a study by de Vere et al (2012) in which the identification success of DNA barcoding of massive collection of flowering plants using matK and rbcL were reported. Several DNA barcoding studies of specific plant families such as Asteraceae (Gao et al 2010), Myristicaceae (Newmaster et al 2007), Rutaceae (Luo et al 2010) have been reported, showing the effectiveness of using matK and rbcL as DNA barcode for certain taxa. This means that to achieve the high success of DNA barcoding, one should consider the best marker to be used according to the taxa in interest.

Furthermore, the identification success on species/genus level is also influenced very much by the marker. There were cases when a species/genus could not be correctly identified with a certain marker but the other species/genera from the same family were successfully identified with the same marker. Some examples for this case are shown by Fu et al (2011), Clement and Donoghue (2012), and Ren et al (2011). In this study, several species from different families (shown in Table 4.6 sub-chapter 4.4) failed to be correctly identified with one or all of the markers used, as follow:

- Two species of *Uvaria* from family Annonaceae were identified ambiguously when rbcL was used. The other genera of this family did not have the same problem. It seems that rbcL is not a suitable barcode for *Uvaria*.
- *Hunteria zeylanica* from family Apocynaceae was not able to be identified by matK and the combined marker when it was correctly identified by rbcL. The reason of this is because the sequences of matK and matK+rbcL of this species have intra-specific divergence higher than the threshold, thus the sequences did not match to any others.
- In family Burseraceae, all of the sequences belong to genus *Canarium* and *Santiria* were ambiguously/incorrectly identified with all of the markers, except for *Santiria apiculata* which was correctly identified with matK and *Santiria laevigata* which was correctly identified with rbcL. According to the analysis result, the mean value of inter-specific divergences within these two genera were considerably low (matK 0,0042; rbcL 0,0034). Many studies reported similar results when using matK and rbcL as barcodes for closely-

related taxa, such as Myristicaceae (Newmaster et al 2007), Rutaceae (Luo et al 2010), Vitaceae (Fu et al 2011), Lamiaceae (De Mattia et al 2011. Low level of inter-specific divergence indicates that plastid markers such as matK and rbcL are not sufficiently variable to distinguish sister species.

- The genus *Clerodendrum* from Lamiaceae was unsuccessfully identified with all of the markers. The species of this genus represented in the analysis showed no variation in the plastid DNA region, thus could not be distinguished from each other with the markers used.
- Four out of six of the species of *Ficus* from Moraceae were unable to be identified using all of the markers. A study by Olivar et al (2014) reported that the inter-specific divergence within the rbcL region of this genus was lower than the intra-specific. Therefore, the sequences would match the sequences of other species rather than match the own species, for example, *Ficus aurata* matched *Ficus sagittata*. The identification of sequences from this genus using BLAST always resulted in different species names with the same score. A certain identification result for this genus was impossible to be obtained when using both matK and rbcL.
- *Spermacoce exilis* of Rubiaceae was unidentified with all the markers. The intra-specific divergence between two sequences of this species was very high indicating that the sequences were obtained from two different species. The BLAST results showed that one of the sequences was actually identified as *Hedyotis sp.* of Rubiaceae. A double-check is needed to see whether the morphological or molecular identification is correct.

5.3 Discriminatory power of matK and rbcL

The effectiveness of the DNA barcodes in discriminating a certain taxon from the others is assessed by calculating the genetic distances between the sequences belong to the corresponding species to the sequences belong to the other species. When the intra-specific distance between a pair of conspecific sequences was smaller than the inter-specific distances between these sequences with any allospecific sequences, these sequences were considered to be discriminated.

The calculation of intra-specific and inter-specific divergences in this study involved 161 species with each species was represented by two sequences from each marker used. The two representations for each species were considered sufficient as shown by other studies (Burgess et al 2011 and Bruni et al 2012). These studies have shown that the inclusion of more accessions of each species would have very little effect to the DNA barcoding of distantly-related taxa which tends to has very low intra-specific genetic distances but high inter-specific distances. Indeed, this study showed that the mean value of intra-specific divergences was very low (0,0008-0,0014) and the mean value of the inter-specific divergences was significantly higher (0,0964-0,3054). The highest variation was obtained with matK followed by matK+rbcL and rbcL. Combining variable marker such as matK with a less variable marker such as rbcL 'dilutes' the genetic divergence, thus the two-loci has intermediate variation.

However, none of the markers used in this study successfully obtained a DNA barcoding gap. An ideal barcode can be determined by the presence of a barcoding gap, which occurs when the minimum value of the inter-specific divergence found among the dataset sequences is higher than the maximum level of intra-specific divergence (Meyer and Paulay 2005). All of the minimum values of inter-specific divergence obtained from three different markers were lower than the maximum values of intra-specific divergence. Thus, a barcoding gap was not clearly observed. A similar study by Lahaye et al (2008) reported the presence of barcoding gap though it was not sufficiently large. Unfortunately, it is difficult to predict when an ideal barcoding gap might be found in plant DNA barcodes. As a comparison, a large barcoding gap is typically present in the CO1 barcode for animals (Hebert et al 2003a, 2003b, 2004a, 2004b). Nevertheless, Ross et al (2008) suggested that barcoding gap is not a major concern as the degree of overlap between the intra-specific and inter-specific divergences is a poor predictor of identification success.

Despite the absence of barcoding gaps, the barcodes generated in this study has relatively high discriminatory power (80-89% for sequence discrimination, 82-91% for species discrimination). According to Hollingsworth et al (2011), most of the plant barcodes would have discriminatory

power more than 70%. Studies by Kress et al (2009) and Burgess et al (2011) showed that barcoding of distantly-related taxa typically results in high level of discriminatory power.

As expected, matK has higher discrimination level compared to rbcL (80% and 73%, respectively) but the different was not significant. The combination of matK and rbcL significantly improved the discrimination by 10%. This is because the use of two-loci barcodes maximized the genetic variation, thus minimizing the number of identical barcode between different species.

All of the species that were unable to be discriminated (Table 4.12 in sub-chapter 4.5) have barcode identical to other species belong to the same family. Identical barcodes across different genera of the same family were uncommon with matK but more common with rbcL. However, matK and rbcL mostly failed to discriminate different species belong to one genus. It means that these two plastid markers are not variable enough to be effective barcodes for closely-related species in certain taxa.

Low variation of matK and rbcL to discriminate closely-related species were clearly observed in groups of species-rich genera, such as *Ficus*, *Santiria*, *Litsea*. Four species of *Santiria* (*Santiria tomentosa*, *S. oblongifolia*, *S. laevigata*, *S. rubiginosa*) shared a single haplotype of matK, as well as five species of *Ficus* (*F. grossularioides*, *F. aurata*, *F. ribes*, *F. variegata, F. schwarzii*) shared a single haplotype of rbcL. These two species groups may contain enough morphological variation for discrimination in the field, but appear to be genetically identical at matK and rbcL. Shared chloroplast DNA sequences between sister species might be caused by chloroplast capture events which occur frequently in species with reproductive compatibility (e.g., Acosta and Premoli 2010, Fehrer et al 2007). Chloroplast capture happens when the chloroplast genome of one species is replaced by that of another species through hybridization (Rieseberg and Soltis 1991). As a consequence, these two separated species would have identical chloroplast DNA.

5.4 The phylogeny of flowering plants of Jambi based on matK and rbcL

5.4.1 Taxonomic resolution in the reconstructed phylogenetic trees

The estimation of the taxonomic resolution allows assessing the effectiveness of a certain marker as a DNA barcode. Fazekas et al (2008) remarked that the taxonomic resolution in a local context may provide a useful indication for predicting the relative DNA barcoding success of the examined taxa at a wider geographic context.

In this study, the taxonomic resolution was estimated at three different levels: family-level, genus-level, and species-level. The resolution figures were based on the percentage of the monophyletic clades of each level that were found to have a bootstrap more than 70% in each phylogenetic tree reconstructed using Neighbor Joining (NJ) method, Maximum Parsimony (MP) method, and Maximum Likelihood (ML) method with matK, rbcL, and matK+rbcL as the markers. The clades with a bootstrap value less than 70% were considered unreliable (Hillis and Bull 1993) and thus excluded from the estimation.

Both matK and rbcL showed high family-level resolution (90-98%). This means most of the families included in the dataset were resolved to be monophyletic clades with bootstrap values more than 70%. The combination of matK and rbcL succeeded to resolve all of the families into monophyletic clades with high bootstrap value.

Furthermore, the taxonomic resolution in the genus-level (63-72%) was much lower compared to the family-level which was as expected. Oddly, the genus-level monophyletic percentages were found slightly lower compared to the species-level in all of the trees, except in MP and ML trees using rbcL. A similar study by Gonzalez et al (2009) reported higher number monophyletic genera (71-77%) compared to monophyletic species (57-71%). This abnormality can be explained by the fact that the proportion of distantly-related species included in the dataset in this study was higher than the proportion of closely-related species. Thus, the probability of resolving monophyletic-species clade was higher than to resolve the monophyletic-genus clade.

Finally, the species-level resolution of 56-74% using matK and rbcL in this study is comparable to similar studies (De Vere et al 2012, Gonzalez et al 2009). However, the two-loci barcode did not make a lot of improvement in the species-level resolution (73-75%). Combining two chloroplast markers was not sufficient to provide 100% of species monophyly.

5.4.2 The topology of the reconstructed phylogenetic trees

The phylogenetic analysis was conducted in this study to see if matK and rbcL barcodes resolve the investigated taxa into appropriate taxonomical grouping. The topologies of the nine phylogenetic trees (Appendixes 4.1 – 4.9) reconstructed in this study based on three statistical analyzes were generally congruent but there were some differences in the clade positions and bootstrap values. These differences, however, were insignificant as visualized in the estimation of the monophyletic percentage in the previous sub-chapter.

At the family level, both matK and rbcL succeeded to separate all of the seventy-six families of flowering plants. However, at the genus and species level, these two markers seemed to have problems in providing a clear separation between some genera and species as described in following sections.

5.4.2.1 Unresolved genera

Twenty non-monophyletic genera (Table 4.14 in sub-chapter 4.6) were found in the phylogenetic trees reconstructed in this study. Fifteen genera (*Dacryodes, Canarium, Santiria, Horsfieldia, Myristica, Croton, Artocarpus, Sterculia, Shorea, Nephelium, Litsea, Fordia, Alstonia, Panicum, and Rourea*) were found to be non-monophyletic in all trees. Of seventy-six families included in the phylogenetic tree reconstruction, Burseraceae and Phyllanthaceae were families with the highest number of unresolved genera.

In most cases, these above-mentioned genera were not able to be separated from the other genus/genera of the same family. For example, *Canarium*, *Dacryodes*, and *Santiria* of Burseraceae were resolved together into several paraphyletic clades; as well as *Myristica* and

Horsfieldia of Myristicaceae. Most of the species in these genera were found to have identical sequences, thus could not be separated from each other. Identical sequences between species of different genera could be common to find when the marker was not variable enough, such as matK and rbcL. In this study, it was revealed that matK and rbcL were not sufficiently variable for species-rich groups such as Burseraceae and Myristicaceae.

The non-monophyly of genera found in this study was comparable to previous studies. A study by Harnelly (2013) reported that matK and rbcL could not completely resolve *Shorea* into a monophyletic clade, and a study by Li et al (2004) showed that *Litsea* was confirmed to be paraphyly based on matK. In other studies, the species-rich genus, such as *Panicum* (Zimmerman et al 2012), *Croton* (Berry et al 2005), and *Artocarpus* (Nyree et al 2010), were reported to have some monophyletic sections when the other sections were non-monophyletic. Meanwhile, phylogenetic studies using non-coding chloroplast DNA found that *Nephelium* (Zamzuriada et al 2009) actually is a monophyletic genus. All of these studies reflect the accuracy of phylogenetic analysis depends on factors such as the utilized marker and the taxa group in question.

A closer look at the unresolved genera in this study revealed that some of them (Kamiya et al 2011, Knobloch 1972) were likely to have had a history of hybridization, which may be contributing to the variation among the species within the corresponding genus. Knobloch (1972) remarked that there was an enormous number of hybridization in flowering plants which commonly happened between species of one genus or between species of different genera. Nevertheless, the genetic variation caused by hybridization cannot be simply detected by plastid marker as its lack of variation (Fazekas et al 2008, 2009). Twyford and Ennos (2011) suggested the use of nuclear marker to accurately identify genetic variation caused by hybridization.

A special attention was given to a species-rich genus, *Artocarpus*, which was found to be paraphyletic with *Prainea limpato* in all of the trees reconstructed in this study. According to The Plant List (www.theplantlist.org), *Prainea limpato* is an unresolved taxon and synonym to

Artocarpus limpato. This might be one of the reasons why matK and rbcL were not able to separate *Prainea limpato* from the clade of *Artocarpus*.

5.4.2.2 Unresolved species

There were sixteen species that were found to be non-monophyletic in all of the trees reconstructed in this study. The phylogenetic trees based on the rbcL marker, as expected, resulted in the highest number of unresolved species compared to matK. At least eighteen species were non-monophyletic according to rbcL but monophyletic according to matK.

The non-monophyly of unresolved species found in this study could happen due to of two reasons. First, these species has identical genetic information with other species belong to the same genera/family. Second, these species has intra-specific divergence higher than the inter-specific, thus they were grouped with the allospecies than with the conspecies.

Species of Burseraceae and Moraceae were examples of the first case. Many species of these two rich families has identical DNA sequences, at least according to plastid markers. This is confirmed by the BLAST results in this study, that most of the sequences belong to the *Canarium spp.*, *Santiria spp.*, *Dacryodes spp.* (Burseraceae) and *Ficus spp.* (Moraceae) have 100% similarity with reference sequences belong to allospecies of the same genera/family.

The second case did not occur so often. In this study, *Ficus aurata* was grouped together with *Ficus grassularoides*, as well as *Litsea robusta* with *Litsea oppositifolia.* Based on the genetic distance calculation, *Ficus aurata* and *Litsea oppositifolia* have intra-specific divergence higher than inter-specific divergence. In order to properly distinguish two different species, the intra-specific divergence of conspecific sequences should be lower than the inter-specific divergence of allospecific sequences (Meier et al 2006). Again, this emphasizes the weakness of plastid markers in providing optimal genetic variation in order to distinguish closely-related species.

5.4.2.3 Ordinal topology of flowering plants in the study area

The ordinal topologies of flowering plants (Figure 4.6 – 4.8 in subchapter 4.6) in the study area were found to be congruent with the ordinal topology according to Angiosperm Phylogeny Group III system (The APG III 2009). This system included 59 orders and 4 families which do not belong to any order. The dataset in this analysis included 30 orders and 1 unplaced family.

There were several disagreements found in the ordinal topologies in this study when using matK or rbcL alone. However, the combination of these two markers provided clearer topologies and corrected the position of orders that were misplaced. All of the orders were resolved accordingly to their ordinal clades outlined by APG III though the relationships between these orders were not perfectly match the APG III' topology. The relationship between Monocots clade and Commelinids clade was not fully resolved. According to APG III (2009), Commelinids is actually a group within Monocots, thus it's difficult to separate these two clades clearly.

6 CONCLUSIONS AND OUTLOOKS

A series of analysis were performed to evaluate the effectiveness of DNA barcodes generated in this study. Criteria of DNA barcodes effectiveness lies on four parameters: (1) the barcode recoverability and quality, (2) the identification success, (3) the discriminatory power, and (4) the taxonomic resolution. Comparisons were made between matK, rbcL, and the combination of these two markers.

In term of barcode recoverability and quality, rbcL was far higher than matK. Approximately 30% of amplification using matK was failed. It took much more effort to generate a good-quality barcode using matK. A robust primer combination might help to improve the amplification and sequencing success of matK in such study where highly variable taxa were sampled.

The identification success was calculated by comparing morphological with molecular identification results and by performing best-close match analysis. Morphological identification was done to each sample and the results were cross-checked with the molecular identification based on nucleotide databases, GenBank and BOLD. As a result, matK has more success in assigning the samples into the correct species name compared to rbcL. The best-close match analysis showed the similar result, confirming the advantage of using matK in plant identification.

The discriminatory power tells about the percentage of barcodes that were successful to differentiate a particular species from other species. Both markers have relatively high discriminatory power as successfully discriminating more than 70% of species included in the analysis. Moreover, the combination of matK and rbcL improved the discriminatory power up to 90%.

Finally, the taxonomic resolution was estimated by analyzing nine phylogenetic trees that were reconstructed based on matK, rbcL, and the two-loci barcodes using three different methods (Neighbor Joining, Maximum Parsimony, and Maximum Likelihood). This analysis showed that either matK and rbcL alone nor the combined marker were able to give a clear taxonomic

resolution over a considerable amount of species. On the other hand, the analysis of matK and rbcL showed that these two markers were not sufficiently variable in certain taxa.

As an overview of the effectiveness of matK and rbcL as plant barcodes, this study showed that these two plastid markers were working well in identifying flowering plant species in the study site, at least up to the genus level. However, there were groups of taxa that were difficult to be distinguished using matK and rbcL. These taxa mostly belong to species-rich groups which tend to have low intra-specific divergences. As reported by many other studies, DNA barcoding of closely-related species results in low success, especially when using coding plastid markers, such as matK and rbcL. Up to date, as indicated by Hollingsworth et al (2011), none of the plant barcodes are perfect in every respect.

A number of constraints were considered to be the limiting factors in DNA barcoding of plant species, such as slow evolution rates and high incidence of hybridization. Accurate methods of analysis are essential, but some degree of uncertainty will always be present due to the difficulties linked to plant species definition. Indeed, one of the future challenges for plant DNA barcoding is to find the most suitable marker to tackle these problems. As the DNA sequencing technology and bioinformatic tools are progressively advancing, the development of new primers soon will be much easier and at the end will increase the success of DNA barcoding.

In this study, the number of misidentified samples was found to be low, yet limiting the total number of sequences included in the analysis. The reasons of misidentification could be due to incorrect reference sequences, incorrect morphological identification, and mislabeling and contamination of the specimen during field work or lab work. Certain measures should be taken to minimize the misidentification, for example, to include only reference sequences that were taxonomically confirmed so correct molecular identification is achievable; using well-presented vouchers for each sample in morphological identification is highly recommended though it is not always possible; implementing a good management of specimen collection to eliminate the possibility of mislabeled specimen; maintaining a clean and well-ordered procedure of DNA analysis should be able to minimize the mislabeling/contamination incidences during lab work.

Ideally, the future study would include all congeneric species from a geographic region and maximize the geographic diversity of samples for each species. Such experimental design should be applied to properly test the performance of DNA barcodes. Moreover, utilization of supplement markers, such as psbA-trnH or ITS/ITS2 is highly recommended in combination with matK and rbcL.

References

Aceto S, Gaudio L (2011) The MADS and the beauty: Genes involved in the development of orchid flowers. Current Genomics 12: 342–356.

Acosta MC, Premoli AC (2010) Evidence of chloroplast capture in South American Nothofagus (subgenus Nothofagus, Nothofagaceae). Molecular Phylogenetics and Evolution 54(1): 235-242.

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular Phylogenetic Evolution 29: 380–395.

Aliabadian M, Nijman V, Mahmoudi A, Naderi M, Vonk R, Vences M (2014) ExcaliBAR: a simple and fast software utility to calculate intra- and interspecific distances from DNA barcodes. Contributions to Zoology 83(1): 79-83.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

Appanah S and Turnbull JM (Ed.) (1998) A review of dipterocarps: taxonomy, ecology, and silviculture. Bogor, Indonesia: Center for International Forestry Research.

APG III (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105–121.

Armenise L, Simeone MC, Piredda R, Schirone B (2012) Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. European Journal of Forest Research 131: 1337–1353.

Asis AMJM, Lacsamana JKM, and Santos MD (2014) Illegal trade of regulated and protected aquatic species in the Philippines detected by DNA barcoding. Mitochondrial DNA, Early Online: 1–8.

Ashton PS (1982) Dipterocarpaceae. Flora Malesiana 9: 237–552.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631.

Baker CS, Steel D, Choi Y, Lee H, Kim KS, Choi SK, Ma Y-U, Hambleton C, Psihoyos L, Brownell RL, Funahashi N (2010) Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. Biology Letter 6: 647-650.

BAPPENAS (2003). Indonesia Biodiversity Strategy and Action Plan 2003 – 2020. Jakarta: BAPPENAS.

Barber P and Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proceedings of the Royal Society B: Biological Sciences 273 (1597): 2053–2061.

Barthlott W, Mutke J, Rafiqpoor D, Kier G, and Kreft H (2005) Global centers of vascular plant diversity. Nova Acta Leopoldina NF 92(342): 61–83.

Berry PE, Hipp AL, Wurdack KJ, Van Ee B, Riina R (2005) Molecular phylogenetics of the giant genus Croton and tribe Crotoneae (Euphorbiaceae sensu stricto) using ITS and TRNL-TRNF DNA sequence data. American Journal of Botany 92(9): 1520-34.

Beukema, H., et al. (2007). "Plant and bird diversity in rubber agroforests in the lowlands of Sumatra, Indonesia." Agroforestry System 70: 217–242.

Bidartondo M (2008) Preserving accuracy in Genbank. Science 319: 1616.

Biemont C (2010) From genotype to phenotype: what do epigenomics and epigenetics tell us? Heredity 105: 1–3.

Bina Perl RG, Nagy ZT, Sonet G, Glaw F, Wollenberg KC, Vences M (2014) DNA barcoding Madagascar's amphibian fauna. Amphibia-Reptilia 35(2).

Bitanyi S, Bjørnstad G, Nesje M, Ernest EM, Mdegela RH, Røed KH (2012) Molecular identification versus local people's information for accurate estimates of bushmeat utilization from the Serengeti ecosystem, Tanzania. African Journal of Biotechnology 11(1): 243-252.

Bleeker W, Klausmeyer S, Peintinger M, Dienst M (2008) DNA sequences identify invasive alien Cardamine at Lake Constance. Biological Conservation 141: 692–698.

BOLD Systems (2014) BOLD Systems. Retrieved from: http://www.boldsystems.org.

Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, Labra M (2010) Identification of poisonous plants by DNA barcoding approach. International Journal of Legal Medicine 124: 595 -603.

Bruni I, De Mattia F, Martellos S, Galimberti A, Savadori P, Casiraghi M, Nimis PL, Labra M (2012) DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). PLoS ONE 7(9): e43256.

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, et al. (2011) Discriminating plant species in a local temperate flora using the rbcL plus matK DNA barcode. Methods in Ecology and Evolution 2: 333–340.

Cawthorn D-M, Steinman HA, Witthuhn RC (2012) DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. Food Research International 46: 30–40.

CBOL (2010-2015) Barcode of life: identifying species with DNA barcoding. Retrieved from http://www.barcodeoflife.org/content/about/what-cbol.

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106: 12794–12797.

Chase M.W., Salamin N., Wilkinson M., Dunwell J.M., Kesanakurthi R.P., Haidar N., Savolainen V. (2005) Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society B-Biological Sciences 360:1889-1895.

Chase MW, Cowan RS, Hollingsworth PM et al. (2007) A proposal for a standardized protocol to barcode all land plants. Taxon 56: 295–299.

Chen S, Yao H, Han J et al. (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5: e8613.

Cho Y, Mower JP, Qiu Y-L, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proceedings of the National Academy of Sciences 101: 17741–17746.

Clement WL and Donoghue MJ (2012) Barcoding success as a function of phylogenetic relatedness in Viburnum, a clade of woody angiosperms. BMC Evolutionary Biology (12):73.

Couvreur TLP, Chatrou LW, Sosef MSM, Richardson JE (2008) Molecular phylogenetics reveal multiple tertiary vicariance of the African rain forest trees. BMC Biology 6: 54.

Cowan RS and Fay MF (2012) Challenges in the DNA barcoding of plant material. Molecular Biology 862: 23-33.

Crawford AJ, Cruz C, Griffith E, Ross H, Ibáñez R, Lips KR, Driskell AC, Bermingham E, Crump P (2013) DNA barcoding applied to ex situ tropical amphibian conservation program reveals cryptic diversity in captive populations. Molecular Ecology Resources 13(6): 1005-18.

Cuenoud P., Savolainen V., Chatrou L.W., Powell M., Grayer R.J., Chase M.W. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. American Journal of Botany 89: 132-144.

Danielsen F, Beukema H, Burgess ND, Parish F, Brühl CA, Donald PF, Murdiyarso D, Phalan B, Reijnders L, Struebig M, Fitzherbert EB (2009) Biofuel plantations on forested lands: double jeopardy for biodiversity and climate. Conservation Biology 23(2): 348-358.

Davis SD, Heywood VH, and Hamilton AC (1995) Centres of plant diversity: a guide and strategy for their conservation (Vol. 2: Asia, Australia, and the Pacific). IUCN Publications Unit, Cambridge.

Dayanandan S., Ashton P.S., Williams S.M., Primack R.B. (1999) Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast rbcL gene. American Journal of Botany 86: 1182-1190.

de Groot GA, During HJ, Maas JW, Schneider H, Vogel JC, et al. (2011) Use of rbcL and trnL-F as a Two-Locus DNA Barcode for Identification of NWEuropean Ferns: An Ecological Perspective. PLoS one 6: e16371.

De Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, and Labra M (2011) A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. Food Research International 44 (3): 693–702.

Dentinger BM, Margaritescu S, Moncalvo J-M (2010) Rapid and reliable high-throughput methods of DNA extraction for use in barcoding and molecular systematic of mushrooms. Molecular Ecology Resources 10: 628–633.

De Salle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conservation Biology 20(5): 1545–1547.

de Vere N, Rich TCG, Ford CR, Trinder SA, Long C, et al. (2012) DNA Barcoding the Native Flowering Plants and Conifers of Wales. PLoS ONE 7(6): e37945.

de Wilde, WJJO (1989) Sumatra. In D.G. Campbell and H.D. Hammond, eds., Floristic Inventory of Tropical Countries, pp. 105-107. New York: The New York Botanical Garden.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S (2015) ycf1, the most promising plastid DNA barcode of land plants. Scientific Reports 5: 8348.

Eaton MJ, Meyers GL, Kolokotronis S-O, Leslie MS, Martin AP, Amato G (2009) Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conservation Genetics 11(4): 1389-1404.

Echi, P. C., Suresh, K. U, George S., et al (2013) Contribution towards the development of a DNA barcode reference library for West African mammals. African Journal of Biotechnology 12(48): 6704-6708.

Elsasser SC, Floyd R, Herbert PDN and Schulte-Hostedde AI (2009) Species identification of North American guinea worms (Nematoda: Dracunculus) with DNA barcoding. Molecular Ecology Resources 9: 707–712.

Fazekas AJ, Burgess KS, Kesanakurti PR et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3: e2802.

Fazekas AJ, Kesanakurti PR, Burgess KS et al (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources 9: 130–139.

Fehrer J, Gemeinholzer B, Chrtek Jr. J, Bräutigam S (2007) Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in Pilosella hawkweeds (Hieracium, Cichorieae, Asteraceae). Molecular Phylogenetics and Evolution 42(2): 347-361.

Finkeldey R, Gailing O, Leinemann L (2009) Molecular Genetic Tools to Infer the Origin of Forest Plants and Wood. Applied Microbiology Biotechnology 85(5): 1251-1258.

Fu Y-M, Jiang W-M, Fu C-X (2011) Identification of species within Tetrastigma (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. Journal of Systematics and Evolution 49 (3): 237–245.

FWI/GFW (2002) The State of the Forest: Indonesia. Jakarta: Forest Watch Indonesia, and Washington DC: Global Forest Watch.

Gao T, Yao H, Song J, Liu C, Zhu Y, et al. (2010) Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. Journal of Ethnopharmacology 130: 116–121.

Gao T, Yao H, Song J, Zhu Y, Liu C, Chen S (2010) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evolutionary Biology 10: 324.

Gautam M, Lele U, Kartodihardjo H, Khan A, Erwinsyah, Rana S (2000) Indonesia: the challenges of World Bank involvement in forests (evaluation country case study series). Washington DC: The World Bank.

Gaveau DLA, Wich S, Epting J, Juhn D, Kanninen M and Leader-Williams N (2009) The future of forests and orangutans (Pongo abelii) in Sumatra: predicting impacts of oil palm plantations,

road construction, and mechanisms for reducing carbon emissions from deforestation. Environmental Research Letters 4 (11pp).

Gillman LN, Keeling DJ, Gardner RC, Wright SD (2010) Faster evolution of highly conserved DNA in tropical plants. Journal of Evolutionary Biology 23: 1327-1330.

Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. BioEssays 33: 135–147.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Petronelli P, et al. (2009) Identification of Amazonian Trees with DNA Barcodes. PLoS one 4: e7483.

Guo X, Simmons MP, But PP-H, Shaw P-C, Wang RJ (2011) Application of DNA barcodes in Hedyotis L. (Spermacoceae, Rubiaceae). Journal of Systematics and Evolution 49: 203–212.

Hajibabaei M. Smith A, Janzen DH, Rodriguez JJ, Whitfield JB and Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes (2006) 6: 959–964.

Harnelly E (2013) DNA sequence-based identification and molecular phylogeny within subfamily Dipterocarpoideae (Dipterocarpaceae). Ph.D. Thesis. Georg-August University: Goettingen.

Hausmann A, Haszprunar G, Segerer AH, Speidel W, Behounek G, Hebert PDN (2011) Now DNA-barcoded: the butterflies and larger moths of Germany (Lepidoptera: Rhopalocera, Macroheterocera). Spixiana 34 (1): 47-58.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. (2003a) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270: 313-321.

Hebert PDN, Ratnasingham S and De Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Philosophical Transactions of the Royal Society B 270: S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences 101: 14812–14817.

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Systematic Biology 54(5): 852–859.

Heimeier, D, Lavery S, Sewell MA (2010) Using DNA barcoding and phylogenetics to identify Antarctic invertebrate larvae: Lessons from a large-scale study. Marine Genomics 3(3–4): 165-177.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology 55:729–739.

Hidayat T, Pancoro A, and Kusumawaty D (2011) Utility of matK gene to assess evolutionary relationship of genus Mangifera (Anacardiaceae) in Indonesia and Thailand. BIOTROPIA 18(2): 74-80.

Hillis DM and Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182–192.

Hilu KW, Liang H (1997) The matK gene: sequence variation and application in plant systematics. American Journal of Botany 84: 830–839.

Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on matK sequence information. American Journal of Botany 90: 1758-1776.

Hollingsworth PM, Clark A, Forrest LL, Richardson J, Pennington RT, et al (2009a) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Research 9: 439–457.

Hollingsworth PM, Forrest LL, Spouge JL, et al (2009b) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America (106): 12794–12797.

Hollingsworth PM (2011a) Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America (108): 19451–19452.

Hollingsworth PM, Graham SW, Little DP (2011b) Choosing and Using a Plant DNA Barcode. PLoS ONE 6 (5): e19254.

Holmes DA (2002) Indonesia: Where Have All the Forests Gone? Washington DC: The World Bank.

Janzen DH (2005) in Plant conservation: A natural history approach. Krupnick G, Kress WJ (eds) University of Chicago Press, Chicago.

Jeanson ML, Labat J-N and Little DP (2011) DNA barcoding: a new tool for palm taxonomists?. Annals of Botany 108(8): 1445.

Kamiya K, Gan YY, & Lum SKY, Khoo MS, Chua SC, Faizu NNH (2011) Morphological and molecular evidence of natural hybridization in Shorea (Dipterocarpaceae). Tree Genetics & Genomes 7: 297–306.

Kesanakurti PR, Fazekas AJ, Burgess KS, Percy DM, Newmaster SG, et al. (2011) Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. Molecular Ecology 20: 1289–1302.

Kiew R (2002) State of biodiversity studies: Sumatra and Peninsular Malesia compared. Flora Malesiana Bulletin (13): 191–196.

Knobloch, IW (1972) Intergeneric hybridization in flowering plants. Taxon 21 (1): 97-103.

Ko HL, Wang YT, Chiu TS, Lee MA, Leu MY, et al. (2013) Evaluating the accuracy of morphological Identification of larval fishes by applying DNA barcoding. PLoS ONE 8(1): e53451.

Koh LP, Miettinen J, Liew SC and Ghazoula J (2011) Remotely sensed evidence of tropical peatland conversion to oil palm. PNAS 108(12): 5127–5132.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences of the USA 102: 8369–8374.

Kress WJ and Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2:e508.

Kress WJ, Erickson DL, Andrew Jones F, Swenson NG, Perez R, et al. (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences of the United States of America 106: 18621–18626.

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, et al. (2010) Advances in the Use of DNA Barcodes to Build a Community Phylogeny for Tropical Trees in a Puerto Rican Forest Dynamics Plot. PLoS ONE 5 (11): e15409.

Kress WJ and Erickson DL [eds.] (2012) DNA barcodes: methods and protocols. Methods in Molecular Biology 858.

Lahaye R, Van der Bank M, Bogarin D et al. (2008) DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the United States of America 105: 2923–2928.

Lambert FR, Collar NJ (2002) The future for Sundaic lowland forest birds: long-term effects of commercial logging and fragmentation. Forktail 18: 127–146.

Laumonier Y (1990) Search for phytogeographical provinces in Sumatra. In: P. Baas, C. Kalkman & R. Geesink (eds.). The plant diversity of Malesia. Proceedings of. The Flora Malesiana Symposium 1989: 193-211.

Laumonier, Y (1997). The Vegetation and physiography of Sumatra. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences of the United States of America 108: 19641–19646.

Li J, Christophel DC, Conran JG, and Li HW (2004) Phylogenetic relationships within the 'core' Laureae (Litsea complex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. Plant Systematic Evolution 246: 19–34.

Lim BK (2012) Preliminary assessment of Neotropical mammal DNA barcodes: an underestimation of biodiversity. The Open Zoology Journal (Suppl 1-M3): 10-17.

Linder HP (2008) Plant species radiations: where, when, why? Philosophical Transactions of Royal Society B 363: 3097–3105.

Little DP and Stevenson DW (2007) A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 23: 1-21.

Luo K, Chen S, Chen K, Song JY, Yao H, Ma XY, Zhu YJ, Pang XH, Yu H, Li HW, Liu Z (2010) Assessment of candidate plant DNA barcodes using the Rutaceae family. Science China 53(6): 701-708.

Maddox T, Priatna D, Gemita E and Salampessy A (2007) The conservation of tigers and other wildlife in oil palm plantations, Jambi Province, Sumatra, Indonesia. ZSL Conservation Report No. 7. London: The Zoological Society of London.

Mankga LT, Yessoufou K, Moteetee AM, Daru BH, van der Bank M (2013) Efficacy of the core DNA barcodes in identifying processed and poorly conserved plant materials commonly used in South African traditional medicine. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (Eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research. ZooKeys 365: 215–233.

Margono BA, Turubanova S, Zhuravleva I, Potapov P, Tyukavina A, Baccini A, Goetz S and Hansen MC (2012) Mapping and monitoring deforestation and forest degradation in Sumatra

(Indonesia) using Landsat time series data set from 1990 to 2010. Environmental Research Letters 7 (16 pp.).

Madden T (2002) The BLAST sequence analysis tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US).

Meier R, Kwong S, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology (55): 715-728.

Meijer W (1981). Sumatra as seen by a botanist. Indonesia Circle 25: 17-27.

Meyer CP and Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3(12): e422.

Miettinen J, Shi C and Liew SC (2011) Deforestation rates in insular Southeast Asia between 2000 and 2010. Global Change Biology 17 (7): 2261–2270.

Miller JA, Beentjes KK, van Helsdingen P, IJland S (2013). Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. ZooKeys (365): 245–261.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol 2(10): 1529–1531.

Nagy ZT, Sonet G, Glaw F, Vences M (2012) First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS ONE 7(3): e34506.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2007) Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Research 8: 480–490.

Newmaster SG, Ragupathy S, Janovec J (2009) A botanical renaissance: state-of-the-art DNA barcoding facilitates an Automated Identification Technology system for plants. International Journal of Computer Applications in Technology 35: 50–60.

Nicolè S, Negrisolo E, Eccher G, Mantovani R, Patarnello T, Erickson DL, Kress WJ and Barcaccia G (2012) DNA Barcoding as a reliable method for the authentication of commercial seafood products. Food Technology and Biotechnology 50 (4): 387–398.

Nyree JC, Zerega MN, Nur Supardi, and Motley TJ (2010) Phylogeny and recircumscription of Artocarpeae (Moraceae) with a focus on Artocarpus. Systematic Botany 35(4): 766-782.

Olivar JEC, Brillantes RY, Rubite RR, Alejandro GJD (2014) Evaluation of three candidate DNA barcoding loci in selected Ficus L. (Moraceae). International Journal of Scientific & Technology Research 3(9).

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, and Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proceedings of the National Academy of Sciences 97: 6960–6966.

Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, Chuyong GB, Cruaud C, Hardy OJ (2013) How effective are DNA barcodes in the identification of African rainforest trees? PLoS One 8: e54921.

Pauls SU, Blahnik RJ, Zhou X, Wardwell CT, and Holzenthal RW (2010) DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (Trichoptera: Hydropsychidae). Journal of the North American Benthological Society 29 (3): 1058-1074.

Prance GT (1977) Floristic inventory of the Tropics: where do we stand? Annals of the Missouri Botanical Garden 64: 659-684.

Prendini L (2005) Comment on ‘‘Identifying spiders through DNA barcodes’’. Canadian Journal of Zoology 83: 498–504.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species, and populations in Odonata. Proceedings of the Royal Society B 275: 237–247.

Ratnasingham S and Hebert PD (2007) BOLD: The barcode of life data system (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355-364.

Ren H, Lu L, Wang H, Li DZ (2011) DNA barcoding of Gaultheria L. in China (Ericaceae: Vaccinioideae). Journal of Systemic Evolution 49: 411–424.

Renaud AK, Savage J, Adamowicz SJ (2012) DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. BMC Ecology 12: 24.

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of Neotropical rainforest trees. Science 293: 2242–2245.

Rieseberg LH and Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5(1): 66-84.

Roos MC, Keßler PJA, Gradstein SR and Baas P (2004) Species diversity and endemism of five major Malesian islands: diversity–area relationships. Journal of Biogeography (31): 1893–1908.

Roy S, Tyagi A, Shukla V et al. (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian Berberis species. PLoS One 5: e13674.

Sanches A, Tokumoto PM, Peres WAM, Nunes FL, Gotardi MST, Carvalho CS, Pelizzon C, Godoi TG and Galetti M (2012) Illegal hunting cases detected with molecular forensics in Brazil. Investigative Genetics 3: 17.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society B: Biological 360: 1850–1811.

Soltis DE, Kuzoff RK, Mort ME, Zanis M, Fishbein, et al. (2001) Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. Annals of the Missouri Botanical Garden 88: 669–693.

Stace CA (2005) Plant taxonomy and biosystematics - does DNA provide all the answers? Taxon 54: 999–1007.

Stoeckle, MY, Gamble CC, Kirpekar R, Young G, Ahmed S, and Little DP (2011). Commercial teas highlight plant DNA barcode identification successes and obstacles. Scientific Reports 1: 42.

Taberlet P, Coissac E, Pompanon F et al (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research 35: e14.

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends in Ecology & Evolution 18: 70-74.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL-W – Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Research 22: 4673–4680.

Twyford AD and Ennos (2011) Next-generation hybridization and introgression. Heredity 108: 179-189.

Vaidya G, Lohman DJ, and Meier R (2011), SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27: 171–180.

Valentini A, Pompanon F and Taberlet P (2008) DNA barcoding for ecologists. Trends in Ecology and Evolution 24(2): 110-117.

Wang W, Wu Y, Yan Y, Ermakova M, Kerstetter R, et al. (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biology 10.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological 360: 1847–1857.

Ward RD, Costa FO, Holmes BH, Steinke D (2008) DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but Zeus faber and Lepidopus caudatus each probably constitute two species. Aquatic Biology 3: 71–78.

Waugh WJ (2011) DNA barcoding the birds of New Zealand. Ph.D. Thesis. Massey University: NZ.

Weeks A, Dalyb DC, Simpson BB (2004) The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. Molecular Phylogenetics and Evolution 35: 85–101.

Whitmore, TC and Tantra IGM (1986) Tree flora of Indonesia: checklist for Sumatra. Forest Research and Development Centre, Bogor.

Whitten T, Damanik SJ, Anwar J, Hisyam N (2000) The ecology of Sumatra. Oxford University Press.

Will K and Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20: 47–55.

Wong LL (2011) DNA Barcoding and related molecular markers for fish species authentication, phylogenetic assessment and population studies. Ph.D. Thesis. Auburn University: USA.

World Bank (2001) Indonesia. Environment and natural resource management in a time of transition. World Bank, Washington DC, USA.

Xiang XG, Hu H, Wang W, Jin XH (2011) DNA barcoding of the recently evolved genus Holcoglossum (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Molecular Ecology Research 11: 1012–1021.

Zamzuriada AS, Mahani MC and Choong CY (2009) Phylogenetic study of Nephelium species based on the chloroplast DNA sequences. Proceedings of the 8th Malaysia Congress on Genetics, 4-6 August 2009, Genting Highlands, Malaysia.

Zimmermann T, Bocksberger G, Brueggemann W, Berberich T (2012) Phylogenetic relationship and molecular taxonomy of African grasses of the genus Panicum inferred from four chloroplast DNA barcodes and nuclear gene sequences. Journal of Plant Research 126(3).

Appendixes

Appendix 4.1 List of vascular plant species collected in Bukit Duabelas National Park and Harapan Rainforest, Jambi, Indonesia

MAGNOLIOPHYTA
Alismatales
Araceae
Alocasia cf. wongii
Alocasia longiloba
Amydrium medium
Anadendrum affine
Anadendrum marginatum
Homalomena cordata
Pothos wallichii
Rhaphidophora sylvestris
Rhaphidophora versteegii
Schismatoglottis calyptrata
Scindapsus hederaceus
Apiales
Araliaceae
Polyscias diversifolia
Polyscias elliptica
Schefflera acutissima
Trevesia burckii
Aquifoliales
Aquifoliaceae
Ilex cymosa
Cardiopteridaceae
Gonocaryum gracile
Gonocaryum macrophyllum
Stemonuraceae
Gomphandra javanica
Gomphandra quadrifida
Stemonurus secundiflorus
Arecales
Arecaceae
Caryota mitis
Daemonorops hirsuta
Daemonorops palembanica
Elaeis guineensis
Licuala spinosa
Asparagales
Amaryllidaceae
Bromheadia finlaysoniana
Corymborkis veratrifolia
Plocoglottis javanica
Asparagaceae
Dracaena elliptica
Hypoxidaceae
Molineria latifolia
Orchidaceae
Eulophia spectabilis
Xanthorrhoeaceae
Dianella ensifolia
Asterales
Asteraceae
Acmella paniculata
Ageratum conyzoides
Chromolaena odorata
Clibadium surinamense
Crassocephalum crepidioides
Cyanthillium cinereum
Mikania micrantha
Rolandra fructicosa
Sphagneticola trilobata
Synedrella nodiflora
Tarlmounia elliptica
Vernonia arborea
Austrobaileyales
Schisandraceae
Kadsura scandens
Schisandra elongata
Brassicales
Capparaceae
Capparis micracantha
Capparis micracantha subsp. korthalsiana
Cleomaceae
Cleome rutidosperma
Caryophyllales
Ancistrocladaceae
Ancistrocladus tectorius
Celastrales
Celastraceae
Euonymus cochinchinensis
Euonymus glaber
Euonymus japonicus
Kokoona ochracea
Lophopetalum beccarianum
Lophopetalum javanum
Salacia korthalsiana
Salacia miqueliana
Chloranthales
Chloranthaceae
Sarcandra glabra
Commelinales
Commelinaceae

Appendix 4.1 (continued)

Amischotolype mollissima	*Diospyros frutescens*
Commelina diffusa	*Diospyros javanica*
Hanguanaceae	*Diospyros korthalsiana*
Hanguana malayana	*Diospyros lanceifolia*
Cornales	*Diospyros sumatrana*
Cornaceae	*Diospyros truncata*
Alangium cf. kurzii	*Diospyros venosa*
Alangium javanicum	**Lecythidaceae**
Alangium ridleyi	*Barringtonia gigantostachya*
Alangium salviifolium	*Barringtonia lanceolata*
Alangium uniloculare	*Barringtonia macrostachya*
Mastixia rostrata	*Barringtonia scortechinii*
Crossosomatales	*Planchonia valida*
Staphyleaceae	**Pentaphylacaceae**
Turpinia cf. brachypetala	*Adinandra cf. integerrima*
Cucurbitales	*Adinandra cf. sarosanthera*
Anisophylleaceae	*Adinandra dumosa*
Anisophyllea disticha	*Eurya acuminata*
Begoniaceae	**Primulaceae**
Begonia cf. aberrans	*Ardisia fuliginosa*
Cucurbitaceae	*Ardisia korthalsiana*
Scopellaria marginata	*Ardisia odontophylla*
Zehneria mucronata	*Ardisia pterocaulis*
Tetramelaceae	*Ardisia purpurea*
Tetrameles nudiflora	*Ardisia sanguinolenta*
Dioscoreales	*Ardisia villosa*
Dioscoreaceae	*Embelia parviflora*
Dioscorea bulbifera	*Embelia ribes*
Dioscorea cf. orbiculata	*Grenacheria amentacea*
Dioscorea hispida	*Maesa ramentacea*
Dioscorea pyrifolia	*Marantodes pumilum*
Dioscorea salicifolia	**Sapotaceae**
Tacca bibracteata	*Gluta wallichii*
Tacca integrifolia	*Madhuca motleyana*
Ericales	*Madhuca penicillata*
Actinidiaceae	*Madhuca sericea*
Saurauia cf. javanica	*Palaquium gutta*
Saurauia cf. tristyla	*Palaquium hexandrum*
Ebenaceae	*Palaquium obovatum*
Artabotrys hexapetalus	*Palaquium ridleyi*
Diospyros borneensis	*Palaquium sumatranum*
Diospyros britannoborneensis	*Payena acuminata*
Diospyros coriacea	*Payena leerii*
Diospyros daemona	*Planchonella maingayi*
Diospyros dictyoneura	*Planchonella nitida*
Diospyros discocalyx	*Planchonella obovata*

Appendix 4.1 (continued)

Pouteria malaccensis
Stryracaceae
Styrax benzoin
Styrax parralleloneuron
Symplocaceae
Symplocos cochinchinensis
Symplocos fasciculata
Symplocos ophirensis
Theaceae
Gordonia borneensis
Gordonia oblongifolia
Escalloniales
Escalloniaceae
Polyosma illicifolia
Fabales
Fabaceae
Acacia pennata
Archidendron bubalinum
Archidendron clypearia
Archidendron fagifolium
Archidendron jiringa
Archidendron microcarpum
Bauhinia acuminata
Bauhinia bidentata
Bauhinia kockiana
Bauhinia semibifida
Bauhinia stipularis
Caesalpinia sumatrana
Callerya artopurpurea
Callerya vasta
Calopogonium mucunoides
Centrosema pubescens
Dalbergia junghuhnii
Dalbergia rostrata
Dalbergia stipulacea
Derris amoena
Dialium indum
Dialium platysepalum
Fordia nivea
Fordia splendidissima
Fordia stipularis
Koompassia malaccensis
Kunstleria ridleyi
Millettia sericea
Mimosa pudica
Mucuna biplicata
Ormosia sumatrana
Paraderris elliptica
Parkia speciosa
Parkia timoriana
Peltophorum pterocarpum
Pericopsis mooniana
Pongamia pinnata
Pterocarpus indicus
Pueraria phaseoloides
Rourea minor
Saraca declinata
Sindora coriacea
Sindora leiocarpa
Sindora wallichii
Spatholobus cf. macropterus
Spatholobus ferrugineus
Spatholobus gyrocarpus
Spatholobus littoralis
Polygalaceae
Polygala paniculata
Salomonia cantoniensis
Xanthophyllum cf. arnottianum
Xanthophyllum eurhynchum
Xanthophyllum flavescens
Xanthophyllum korthalsianum
Xanthophyllum nigricans
Xanthophyllum rufum
Xanthophyllum stipitatum
Xanthophyllum vitellinum
Xanthophyllum wrayi
Fagales
Fagaceae
Castanopsis acuminatissima
Castanopsis argentea
Castanopsis costata
Castanopsis inermis
Castanopsis javanica
Castanopsis schefferiana
Lithocarpus blumeanus
Lithocarpus cf. cantleyanus
Lithocarpus gracilis
Lithocarpus hystrix
Quercus argentata
Quercus lineata
Gentianales
Apocynaceae

Appendix 4.1 (continued)

Allamanda cathartica	*Kohautia cynanchica*
Alstonia angustifolia	*Laisanthus tomentosus*
Alstonia pnematophora	*Lasianthus attenuatus*
Alstonia scholaris	*Lasianthus inaequalis*
Alyxia pilosa	*Lasianthus inodorus*
Aporosa frutescens	*Lasianthus reticulatus*
Chilocarpus cf. suaveolens	*Lasianthus stercorarius*
Chilocarpus costatus	*Lasianthus verticillatus*
Chonemorpha verrucosa	*Morinda villosa*
Dischidia cf. reniformis	*Mussaenda frondosa*
Dyera costulata	*Nauclea orientalis*
Epigynum cf. ridleyi	*Nauclea subdita*
Hunteria zeylanica	*Neonauclea calycina*
Kibatalia maingayi	*Neonauclea excelsa*
Leuconotis eugeniifolia	*Nostolachma densiflora*
Melodinus orientalis	*Oldenlandia corymbosa*
Parsonsia cf. alboflavescens	*Oxyceros longiflorus*
Tabernaemontana macrocarpa	*Pertusadina eurhyncha*
Tabernaemontana pandacaqui	*Porterandia anisophylla*
Tabernaemontana pauciflora	*Psychotria leucocarpa*
Telosma accedens	*Psychotria malayana*
Toxocarpus villosus	*Psychotria penangensis*
Willughbeia angustifolia	*Psychotria robusta*
Willughbeia beccariana	*Psychotria rostrata*
Willughbeia coriacea	*Psychotria viridiflora*
Willughbeia tenuiflora	*Psychotria viridis*
Wrightia laevis	*Rothmannia macrophylla*
Loganiaceae	*Saprosma arboreum*
Strychnos ignatii	*Saprosma cf. scortechinii*
Rubiaceae	*Spermacoce cristata*
Canthium horridum	*Spermacoce exilis*
Chassalia curviflora	*Spermacoce laevis*
Chonemorpha fragrans	*Spermacoce latifolia*
Coptosapelta flavescens	*Spermacoce ocymifolia*
Coptosapelta tomentosa	*Spermacoce ocymoides*
Diodella sarmentosa	*Streblosa polyantha*
Discospermum abnorme	*Streblosa tortilis*
Fagraea racemosa	*Timonius cf. esherianus*
Gardenia tubifera	*Timonius flavescens*
Gynochthodes coriacea	*Timonius wallichianus*
Gynotroches axillaris	*Uncaria acida*
Ixora grandifolia	*Uncaria cordata*
Ixora paludosa	*Uncaria elliptica*
Ixora reticulata	*Uncaria gambir*
Ixora salicifolia	*Uncaria lanosa var. ferrea*
Jasminum elongatum	*Urophyllum arboreum*

Appendix 4.1 (continued)

Urophyllum corymbosum	**Lauraceae**
Urophyllum macrophyllum	*Actinodaphne cf. glomerata*
Urophyllum trifurcum	*Actinodaphne oleifolia*
Lamiales	*Alseodaphne bancana*
Acanthaceae	*Alseodaphne lancilimba*
Andrographis paniculata	*Alseodaphne nigrescens*
Asystasia gangetica	*Beilschmiedia madang*
Staurogyne elongata	*Beilschmiedia maingayi*
Gesneriaceae	*Cinnamomum iners*
Codonoboea platypus	*Cinnamomum porrectum*
Cyrtandra cf. reticosa	*Cryptocarya costata*
Cyrtandra pendula	*Cryptocarya crassinervia*
Lamiaceae	*Cryptocarya densiflora*
Callicarpa cf. candicans	*Cryptocarya ferrea*
Callicarpa pentandra	*Cryptocarya pulchrinervia*
Clerodendrum adenophysum	*Dehaasia cf. membranacea*
Clerodendrum cf. laevifolium	*Dehaasia firma*
Clerodendrum deflexum	*Dehaasia incrassata*
Clerodendrum disparifolium	*Dehaasia microcarpa*
Clerodendrum laevifolium	*Endiandra cf. clavigera*
Clerodendrum phyllomega	*Endiandra immersa*
Clerodendrum laevifolium	*Endiandra rubescens*
Gomphostemma javanicum	*Lindera insignis*
Hyptis capitata	*Lindera lucida*
Peronema canescens	*Litsea aurea*
Petraeovitex sumatrana (cf.)	*Litsea castanea*
Teijsmanniodendron coriaceum	*Litsea cf. forstenii*
Vitex pinnata	*Litsea cf. javanica*
Vitex quinata	*Litsea cf. machilifolia*
Vitex vestita	*Litsea cubeba*
Linderniaceae	*Litsea elliptica*
Legazpia polygonoides	*Litsea firma*
Lindernia crustacea	*Litsea forstenii*
Lindernia diffusa	*Litsea glutinosa*
Torenia violacea	*Litsea grandis*
Oleaceae	*Litsea lanceolata*
Chionanthus curvicarpus	*Litsea machilifolia*
Chionanthus montanus	*Litsea monopetala*
Chionanthus polygamus	*Litsea noronhae*
Scrophulariaceae	*Litsea oppositifolia*
Scoparia dulcis	*Litsea resinosa*
Verbenaceae	*Litsea robusta*
Lantana camara	*Litsea umbellata*
Stachytarpheta indica	*Neolitsea cinnamomea*
Stachytarpheta jamaicensis	*Ocotea beulahiae*
Laurales	*Persea rimosa*

Appendix 4.1 (continued)

Phoebe elliptica	*Orophea cumingiana*
Phoebe grandis	*Orophea hexandra*
Monimiaceae	*Phaeanthus splendens*
Kibara coriacea	*Polyalthia cauliflora*
Liliales	*Polyalthia lateriflora*
Liliaceae	*Polyalthia microtus*
Apostasia wallichii	*Polyalthia obliqua*
Smilacaceae	*Polyalthia rumphii*
Heterosmilax micrantha	*Popowia hirta*
Smilax calophylla	*Popowia pisocarpa*
Smilax cf. lanceifolia	*Popowia tomentosa*
Smilax cf. zeylanica	*Pseuduvaria reticulata*
Smilax leucophylla	*Trivalvaria macrophylla*
Smilax setosa	*Uvaria cuneifolia*
Magnoliales	*Uvaria excelsa*
Annonaceae	*Uvaria grandiflora*
Artabotrys cf. wrayi	*Uvaria hirsuta*
Artabotrys maingayi	*Uvaria littoralis*
Artabotrys suaveolens	*Uvaria lobbiana*
Artabotrys tomentosus	*Xylopia caudata*
Cyathocalyx bancanus	*Xylopia elliptica*
Cyathocalyx sumatranus	*Xylopia ferruginea*
Dasymaschalon dasymaschalum	*Xylopia glauca*
Drepananthus biovulatus	*Xylopia malayana*
Drepananthus carinatus	**Magnoliaceae**
Drepananthus ramuliflorus	*Magnolia elegans*
Fissistigma latifolium	**Myristicaceae**
Fissistigma manubriatum	*Gymnacranthera bancana*
Friesodielsia biglandulosa	*Gymnacranthera farquhariana*
Friesodielsia borneensis	*Gymnacranthera forbesii*
Friesodielsia cuneiformis	*Horsfieldia cf. fulva*
Friesodielsia glauca	*Horsfieldia glabra*
Goniothalamus costulatus	*Horsfieldia grandis*
Goniothalamus macrophyllus	*Horsfieldia macrothyrsa*
Goniothalamus malayanus	*Horsfieldia polyspherula*
Gonystylus forbesii	*Horsfieldia pulcherrima*
Gonystylus maingayi	*Horsfieldia punctatifolia*
Maasia glauca	*Horsfieldia superba*
Maasia hypoleuca	*Knema cinerea*
Maasia sumatrana	*Knema furfuracea*
Meiogyne cf. virgata	*Knema glaucescens*
Mezzettia parviflora	*Knema latifolia*
Miliusa longipes	*Knema laurina*
Mitrella kentii	*Myristica gigantea*
Neo-uvaria acuminatissima	*Myristica iners*
Orophea corymbosa	*Myristica maingayi*

Appendix 4.1 (continued)

Myristica maxima	*Aporosa cf. falcifera*
Myristica villosa	*Balakata baccata*
Malpighiales	*Blumeodendron tokbrai*
Achariaceae	*Botryophora geniculata*
Hydnocarpus cf. curtisii	*Cephalomappa malloticarpa*
Hydnocarpus cf. sumatrana	*Claoxylon longifolium*
Hydnocarpus cf. wrayi	*Clonostylis forbesii*
Hydnocarpus polypetalus	*Croton argenteus*
Ryparosa acuminata	*Croton argyratus*
Ryparosa caesia	*Croton cascarilloides*
Ryparosa cf. hirsuta	*Croton caudatus*
Ryparosa cf. scortechinii	*Croton hirtus*
Ryparosa fasciculata	*Croton leiophyllus*
Ryparosa javanica	*Croton oblongus*
Calophyllaceae	*Endospermum diademum*
Calophyllum pulcherrimum	*Hancea penangensis*
Calophyllum soulattri	*Hevea brasiliensis*
Centroplacaceae	*Macaranga bancana*
Bhesa paniculata	*Macaranga cf. sumatrana*
Bhesa robusta	*Macaranga conifera*
Chrysobalanaceae	*Macaranga gigantea*
Atuna racemosa	*Macaranga heynei*
Parastemon urophyllus	*Macaranga hosei*
Parinari sumatrana	*Macaranga hullettii*
Clusiaceae	*Macaranga hypoleuca*
Garcinia atroviridis	*Macaranga javanica*
Garcinia gaudichaudii	*Macaranga pruinosa*
Garcinia graminea	*Macaranga trichocarpa*
Garcinia griffithii	*Mallotus macrostachyus*
Garcinia macrophylla	*Mallotus mollissimus*
Garcinia parvifolia	*Mallotus paniculatus*
Garcinia rostrata	*Mallotus peltatus*
Dichapetalaceae	*Melanolepis multiglandulosa*
Dichapetalum cf. rugosum	*Neoscortechinia kingii*
Dichapetalum gelonioides	*Pimelodendron griffithianum*
Dichapetalum gelonioides subsp. sumatranum	*Pimelodendron zoanthogyne*
Dichapetalum sordidum	*Ptychopyxis bacciformis*
Dichapetalum timoriense	*Ptychopyxis costata*
Erythroxylaceae	*Suregada cf. glomerulata*
Erythroxylum cuneatum	*Suregada cf. multiflora*
Euphorbiaceae	**Hypericaceae**
Agrostistachys glaudichaudii	*Cratoxylum cochinchinense*
Agrostistaehys hookeri	*Cratoxylum formosum*
Alchornea tilifolia	*Cratoxylum sumatranum*
Antidesma neurocarpum	**Irvingiaceae**
Antidesma tomentosum	*Irvingia malayana*

Appendix 4.1 (continued)

Ixonanthaceae
Ixonanthes petiolaris
Linaceae
Hugonia costata
Reinwardtia patens
Malphigiaceae
Hiptage cf. sericea
Ochnaceae
Gomphia serrata
Pandaceae
Galearia aristifera
Galearia filiformis
Galearia fulva
Galearia maingayi
Passifloraceae
Adenia macrophylla
Paropsia varecifomis
Passiflora foetida
Piriqueta racemosa
Peraceae
Trigonopleura malayana
Phyllanthaceae
Antidesma cf. velutinum
Antidesma coriaceum
Antidesma cuspidatum
Antidesma forbesii
Antidesma leucopodum
Antidesma stipulare
Antidesma velutinosum
Aporosa antennifera
Aporosa arborea
Aporosa benthamiana
Aporosa cf. antennifera
Aporosa cf. octandra
Aporosa lucida
Aporosa lunata
Aporosa nervosa
Aporosa octandra
Aporosa octandra var. malesiana
Aporosa prainiana
Aporosa subcaudata
Aporosa symplocoides
Aporosa whitmorei
Baccaurea cf. pyriformis
Baccaurea cf. velutina
Baccaurea deflexa
Baccaurea dulcis
Baccaurea javanica
Baccaurea lanceolata
Baccaurea macrocarpa
Baccaurea macrophylla
Baccaurea minor
Baccaurea mollis
Baccaurea parviflora
Baccaurea polyneura
Baccaurea pubera
Baccaurea pyriformis
Baccaurea reticulata
Baccaurea sumatrana
Breynia racemosa
Bridelia glauca
Bridelia insulana
Bridelia tomentosa
Dicoelia sumatrana
Glochidion cf. leucocarpum
Glochidion lutescens
Glochidion philippicum
Glochidion rubrum
Glochidion sericeum
Glochidion superbum
Glochidion zeylanicum var. arborescens
Koilodepas brevipes
Leptonychia caudata
Phyllanthus oxophyllus
Phyllanthus urinaria
Putranjivaceae
Drypetes longifolia
Rhizophoraceae
Carallia brachiata
Carallia suffruticosa
Pellacalyx axillaris
Pellacalyx lobbii
Salicaceae
Casearia capitellata
Homalium caryophyllaceum
Homalium cf. foetidum
Scolopia spinosa
Ventilago oblongifolia
Trigoniaceae
Trigoniastrum hypoleucum
Violaceae
Rinorea anguifera

Appendix 4.1 (continued)

Rinorea cf. sclerocarpa	*Sterculia rubiginosa*
Malvales	*Trichospermum kurzii*
Dipterocarpaceae	*Urena lobata*
Anisoptera costata	**Thymelaeaceae**
Cleistanthus megacarpus	*Aquilaria malaccensis*
Hopea beccariana	*Aquilaria sinensis*
Hopea sangal	*Enkleia malaccensis*
Parashorea lucida	*Gonystylus acuminatus*
Shorea acuminata	*Gonystylus affinis*
Shorea bracteolata	**Myrtales**
Shorea gibbosa	**Combretaceae**
Shorea lepidota	*Combretum elmeri*
Shorea ovalis	*Combretum nigrescens*
Shorea parvifolia	*Terminalia bellirica*
Shorea singkawang	*Terminalia foetidissima*
Malvaceae	*Terminalia oblonga*
Alangium villosum	*Terminalia subspathulata*
Byttneria curtisii	**Lythraceae**
Byttneria reinwardtii	*Barringtonia pendula*
Clerodendrum villosum	*Cuphea carthagenensis*
Commersonia bartramia	*Lagerstroemia speciosa*
Durio excelsus	**Melastomataceae**
Durio oxleyanus	*Bellucia pentamera*
Durio zibethinus	*Clidemia hirta*
Grewia laevigata	*Diplectria stipularis*
Heritiera sumatrana	*Dissochaeta affinis*
Leptonychia heteroclita	*Dissochaeta gracilis*
Microcos florida	*Ewyckia cyanea*
Microcos henrici	*Melastoma malabathricum*
Microcos hirsuta	*Memecylon edule*
Microcos paniculata	*Memecylon garcinioides*
Pentace borneensis	*Memecylon myrsinoides*
Pentace triptera	*Memecylon paniculatum*
Pimelodendron cf. amboinicum	*Pternandra azurea*
Pterocymbium tubulatum	*Pternandra caerulescens*
Pterospermum subpeltatum	*Sonerila heterostemon*
Scaphium affine	**Myrtaceae**
Scaphium linearicarpum	*Ctenolophon parvifolius*
Scaphium macropodum	*Decaspermum cf. parviflorum*
Sida rhombifolia	*Psidium guajava*
Sterculia coccinea	*Rhodamnia cinerea*
Sterculia lanceolata	*Syzygium acuminatissimum*
Sterculia macrophylla	*Syzygium aemulum*
Sterculia megistophylla	*Syzygium cf. borneense*
Sterculia oblongata	*Syzygium chloranthum*
Sterculia parvifolia	*Syzygium claviflorum*

Appendix 4.1 (continued)

Syzygium creaghii
Syzygium euneuron
Syzygium everettii
Syzygium fastigiatum
Syzygium glabratum
Syzygium hemsleyanum
Syzygium hirtum
Syzygium laxiflorum
Syzygium leptostemon
Syzygium lineatum
Syzygium palembanicum
Syzygium paludosum
Syzygium polyanthum
Syzygium pseudoformosum
Syzygium racemosum
Syzygium rostratum
Syzygium splendens
Syzygium tetrapterum
Onagraceae
Ludwigia octovalvis
Oxalidales
Connaraceae
Agelaea borneensis
Agelaea macrophylla
Agelaea trinervis
Cnestis palala
Connarus semidecandrus
Connarus villosus
Pycnarrhena cauliflora
Rourea asplenifolia
Rourea mimosoides
Roureopsis emarginata
Taeniochlaena acutipetala
Elaeocarpaceae
Elaeocarpus acronodia
Elaeocarpus cf. oxypyren
Elaeocarpus cf. stipularis var. brevipes
Elaeocarpus floribundus
Elaeocarpus glaber
Elaeocarpus grandiflorus
Elaeocarpus mastersii
Elaeocarpus nitidus
Elaeocarpus parvifolius
Elaeocarpus petiolatus
Elaeocarpus salicifolius
Elaeocarpus serratus
Elaeocarpus stipularis
Elaeocarpus stipularis var. brevipes
Oxalidaceae
Oxalis barrelieri
Sarcotheca diversifolia
Pandanales
Stemonaceae
Stemona javanica
Piperales
Aristolochiaceae
Thottea cf. grandiflora
Thottea piperiformis
Piperaceae
Piper baccatum
Piper caninum
Piper coactile
Poales
Cyperaceae
Cyperus diffusus
Cyperus haspan
Cyperus platystylis
Fimbristylis dichotoma
Hypolytrum nemorum
Mapania cuspidata
Mapania sessilis
Mapania tenuiscapa
Rhynchospora colorata
Rhynchospora corymbosa
Scleria ciliaris
Scleria sumatrensis
Flagellariaceae
Flagellaria indica
Poaceae
Axonopus compressus
Centotheca lappacea
Cyrtococcum oxyphyllum
Cyrtococcum patens
Cyrtococcum patens var. latifolium
Imperata cylindrica
Leptaspis urceolata
Ottochloa nodosa
Panicum laxum
Panicum sarmentosum
Paspalum conjugatum
Paspalum dilatatum
Pennisetum polystachion

Appendix 4.1 (continued)

Proteales	*Ficus deltoidea*
Proteaceae	*Ficus fulva*
Helicia excelsa	*Ficus glandulifera*
Helicia fuscotomentosa	*Ficus globosa*
Helicia robusta	*Ficus grossularioides*
Ixonanthes icosandra	*Ficus heteropleura*
Ranunculales	*Ficus hirta*
Menispermaceae	*Ficus padana*
Albertisia papuana	*Ficus parietalis*
Ampelocissus spicifer	*Ficus racemosa*
Anamirta cocculus	*Ficus ribes*
Arcangelisia flava	*Ficus sagittata*
Coscinium fenestratum	*Ficus schwarzii*
Cyclea laxiflora	*Ficus variegata*
Diploclisia glaucescens	*Ficus virens*
Fibraurea darshanii (cf.)	*Parartocarpus venenosa*
Fibraurea tinctoria	*Prainea limpato*
Hypserpa nitida	*Sloetia elongata*
Hypserpa polyandra	*Streblus asper*
Limacia scandens	*Streblus elongatus*
Menispermum glabrum	**Rhamnaceae**
Pericampylus glaucus	*Ventilago malaccensis*
Pycnarrhena tumefacta	*Ziziphus angustifolia*
Tinomiscium petiolare	*Ziziphus horsfieldii*
Tinospora glabra	*Ziziphus jujuba*
Rosales	**Rosaceae**
Cannabaceae	*Prunus arborea*
Gironniera hirta	*Prunus cf. grisea*
Gironniera nervosa	*Prunus polystachya*
Gironniera subaequalis	**Urticaceae**
Trema orientalis	*Dendrocnide sinuata*
Trema tomentosa	*Dendrocnide stimulans*
Moraceae	*Elatostema parasiticum*
Antiaris toxicaria	*Elatostema raapii*
Artocarpus anisophyllus	*Elatostema sinuatum*
Artocarpus dadah	*Poikilospermum suaveolens*
Artocarpus elasticus	**Santalales**
Artocarpus heterophyllus	**Loranthaceae**
Artocarpus hispidus	*Dendrophthoe cf. pentandra*
Artocarpus integer	**Olacaceae**
Artocarpus kemando	*Ochanostachys amentacea*
Artocarpus nitidus	*Strombosia zeylandica*
Ficus aurata	**Opiliaceae**
Ficus chartacea	*Champereia manillana*
Ficus crassiramea	**Sapindales**
Ficus delosyce	**Anacardiaceae**

Appendix 4.1 (continued)

Bouea macrophylla	*Aglaia sexipetala*
Buchanania sessilifolia	*Aglaia silvestris*
Campnosperma auriculatum	*Aglaia simplicifolia*
Dracontomelon dao	*Aglaia spectabilis*
Drimycarpus luridus	*Aglaia tomentosa*
Mangifera caesia	*Ailanthus integrifolia*
Mangifera foetida	*Aphanamixis polystachya*
Mangifera laurina	*Arytera littoralis*
Mangifera odorata	*Chisocheton ceramicus*
Mangifera torquenda	*Chisocheton macrophyllus*
Melanochyla beccariana	*Chisocheton patens*
Melanochyla caesia	*Chisocheton tomentosus*
Melanochyla tomentosa	*Dysoxylum acutangulum*
Parishia insignis	*Dysoxylum alliaceum*
Semecarpus cf. caesia	*Dysoxylum arborescens*
Semecarpus heterophylla	*Dysoxylum cauliflorum*
Burseraceae	*Dysoxylum densiflorum*
Canarium album	*Dysoxylum excelsum*
Canarium caudatum	*Dysoxylum parasiticum*
Canarium cf. denticulatum	*Lansium parasiticum*
Canarium cf. gracile	*Pseudoclausena chrysogyne*
Canarium cf. patentinervium	*Sandoricum koetjape*
Canarium dichotomum	**Rutaceae**
Canarium hirsutum	*Acronychia pedunculata*
Canarium littorale	*Aglaia cf. silvestris*
Canarium megalanthum	*Aglaia palembanica*
Canarium pilosum	*Citrus nobilis*
Dacryodes costata	*Clausena excavata*
Dacryodes laxa	*Glycosmis pentaphylla*
Dacryodes rostrata	*Lunasia amara*
Dacryodes rugosa	*Luvunga eleutherandra*
Santiria apiculata	*Luvunga motleyi*
Santiria griffithii	*Luvunga sarmentosa*
Santiria oblongifolia	*Melicope cf. triphylla*
Santiria rubiginosa	*Melicope euneura*
Santiria tomentosa	*Melicope glabra*
Triomma malaccensis	*Melicope latifolia*
Meliaceae	*Paramignya cf. scandens*
Aglaia argentea	*Paramignya mindanaensis*
Aglaia crassinervia	*Tetractomia tetrandra*
Aglaia cucullata	*Zanthoxylum myriacanthum*
Aglaia lawii	**Sapindaceae**
Aglaia leucophylla	*Acer laurinum*
Aglaia malaccensis	*Allophylus cobbe*
Aglaia odoratisissima	*Dictyoneura acuminata*
Aglaia rubiginosa	*Dimocarpus longan*

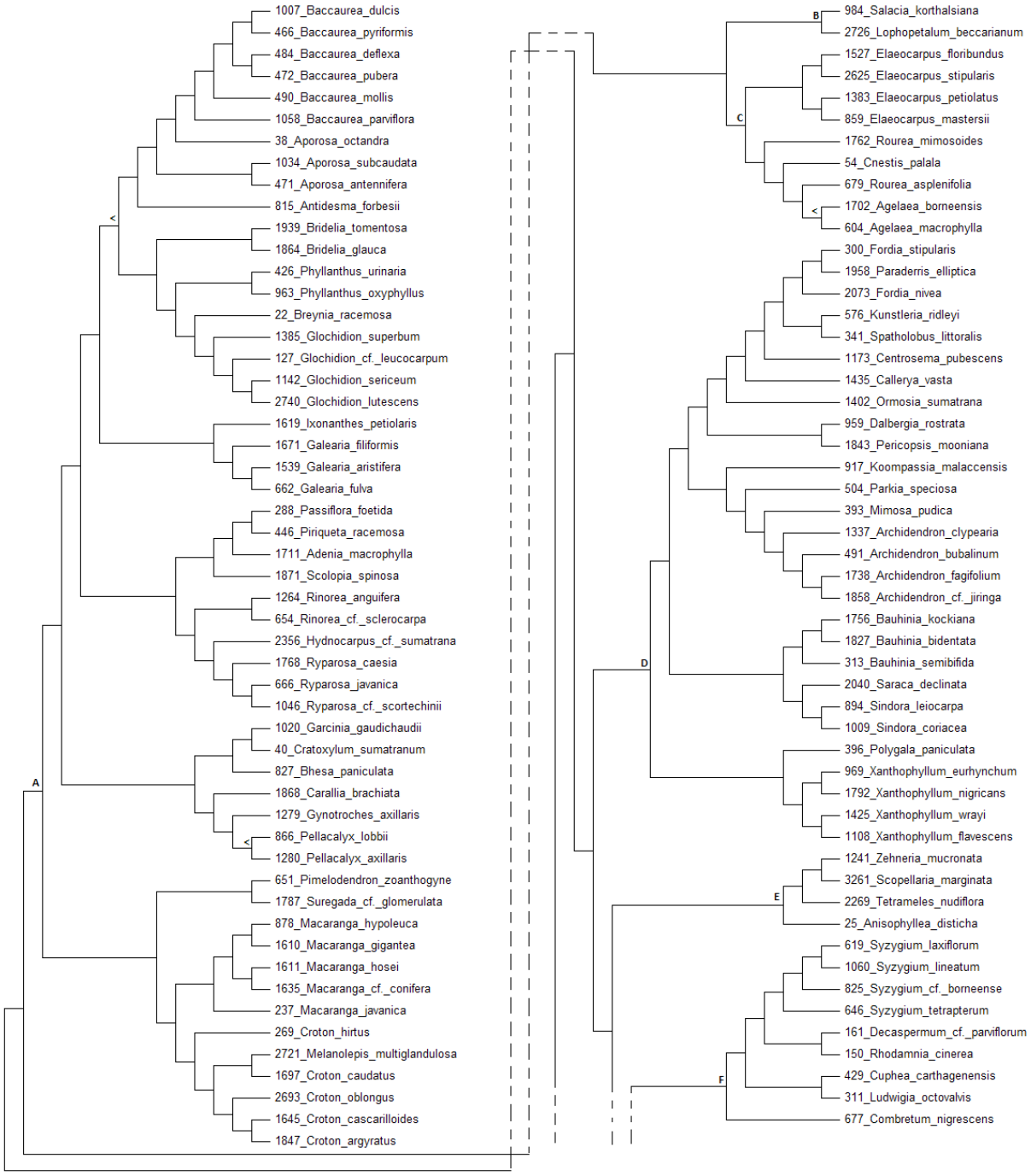
Appendix 4.1 (continued)

Ganophyllum falcatum	*Tetrastigma papillosum*
Guioa multijuga	**Zingiberales**
Lepisanthes tetraphylla	**Costaceae**
Mischocarpus sumatranus	*Cheilocostus globosus*
Mischocarpus sundaicus (cf.)	*Cheilocostus speciosus*
Nephelium cuspidatum	*Costus speciosus*
Nephelium cuspidatum var. cuspidatum subvar. dasyneurum	**Lowiaceae**
Nephelium cuspidatum var. eriopetalum	*Orchidantha cf. siamensis*
Nephelium eriopetalum	**Marantaceae**
Nephelium juglandifolium	*Donax canniformis*
Nephelium lappaceum	*Phrynium pubinerve*
Nephelium laurinum	**Zingiberaceae**
Nephelium maingayi	*Globba pendula*
Nephelium mutabile	**No order**
Nephelium rubescens	**Dilleniaceae**
Nephelium subfalcatum	*Dillenia excelsa*
Otophora amoena	*Dillenia eximia*
Paranephelium xestophyllum (cf.)	*Tetracera akara*
Pometia pinnata	*Tetracera indica*
Xerospermum laevigatum	*Tetracera scandens*
Xerospermum noronhianum	**Icacinaceae**
Simaroubaceae	*Iodes cirrhosa*
Eurycoma longifolia	*Phytocrene macrophylla*
Solanales	**PTERIDOPHYTA**
Convolvulaceae	**Cyatheales**
Erycibe crassipes	**Cyatheaceae**
Erycibe sumatrensis	*Schizocaena molucana*
Erycibe tomentosa	**Gleicheniales**
Ipomoea sumatrana	**Gleicheniaceae**
Merremia umbellata	*Dicranopteris linearis*
Merremia umbellata subsp. orientalis	**Marattiales**
Solanaceae	**Marattiaceae**
Solanum cf. rudepannum	*Angiopteris evecta*
Vitales	**Ophioglossales**
Vitaceae	**Ophioglossaceae**
Ampelocissus ascendiflora	*Helminthostachys zeylandica*
Ampelocissus elegans	*Ophioglossum pendulum*
Ampelocissus thyrsiflora	**Polypodiales**
Cayratia geniculata	**Aspleniaceae**
Cayratia japonica	*Asplenium glaucophyllum*
Cayratia mollissima	*Asplenium longissimum*
Cissus nodosa	*Asplenium nidus*
Leea indica	**Athiryaceae**
Pterisanthes cf. heterantha	*Athyrium bantamense*
Pterisanthes eriopoda	*Diplazium cordifolium*
Pterisanthes polita	**Blechnaceae**

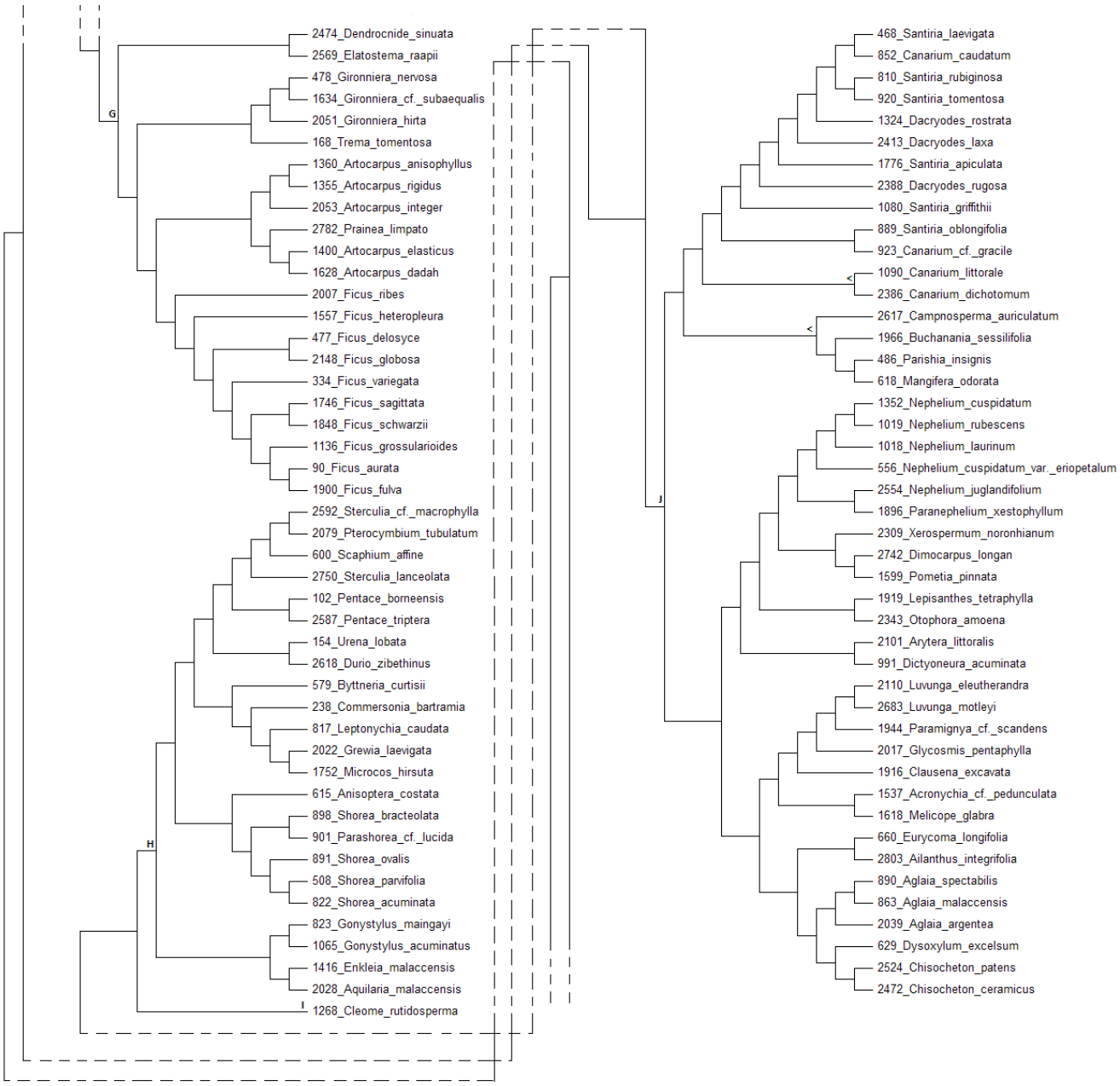
Appendix 4.1 (continued)

Blechnum orientale
Stenochlaena palustris
Davalliaceae
Davallia cf. denticulata
Davallia triphylla
Dennstaedtiaceae
Lindsaea ensifolia
Lindsaea nitida
Microlepia cf. hancei
Dryopteridaceae
Teratophyllum aculeatum
Teratophyllum ludens (cf.)
Nephrolepidaceae
Nephrolepis acutifolia
Nephrolepis biserrata
Polypodiaceae
Goniophlebium verrucosum
Leptochilus cf. decurrens
Phymatosorus cf. membranifolium
Pyrrosia lanceolata
Pyrrosia piloselloides
Tectariaceae
Tectaria barberi
Tectaria cf. semipinnata
Thelypteridaceae
Christella dentata
Cyclosorus cf. heterocarpus
Pronephrium triphyllum
Sphaerostephanos polycarpa
Pteridales
Adiantaceae
Adiantum latifolium
Pteridaceae
Antrophyum callifolium
Taenitis blechnoides
Vittariaceae
Vittaria elongata
Schizaeales
Lygodiaceae
Lygodium microphyllum
Lygodium cf. longifolium
Lygodium circinatum
Lygodium flexuosum
Lygodium microphyllum
Lygodium salicifolium
Schizaeaceae
Schizaea dichotoma
Schizaea digitata
Schizaea malaccana
GNETOPHYTA
Gnetales
Gnetaceae
Gnetum cuspidatum
LYCOPODIAPHYTA
Lycopodiales
Lycopodiaceae
Lycopodiella cernua
Selaginellales
Selaginellaceae
Selaginella intermedia
PINOPHYTA
Pinales
Podocarpaceae
Nageia wallichiana

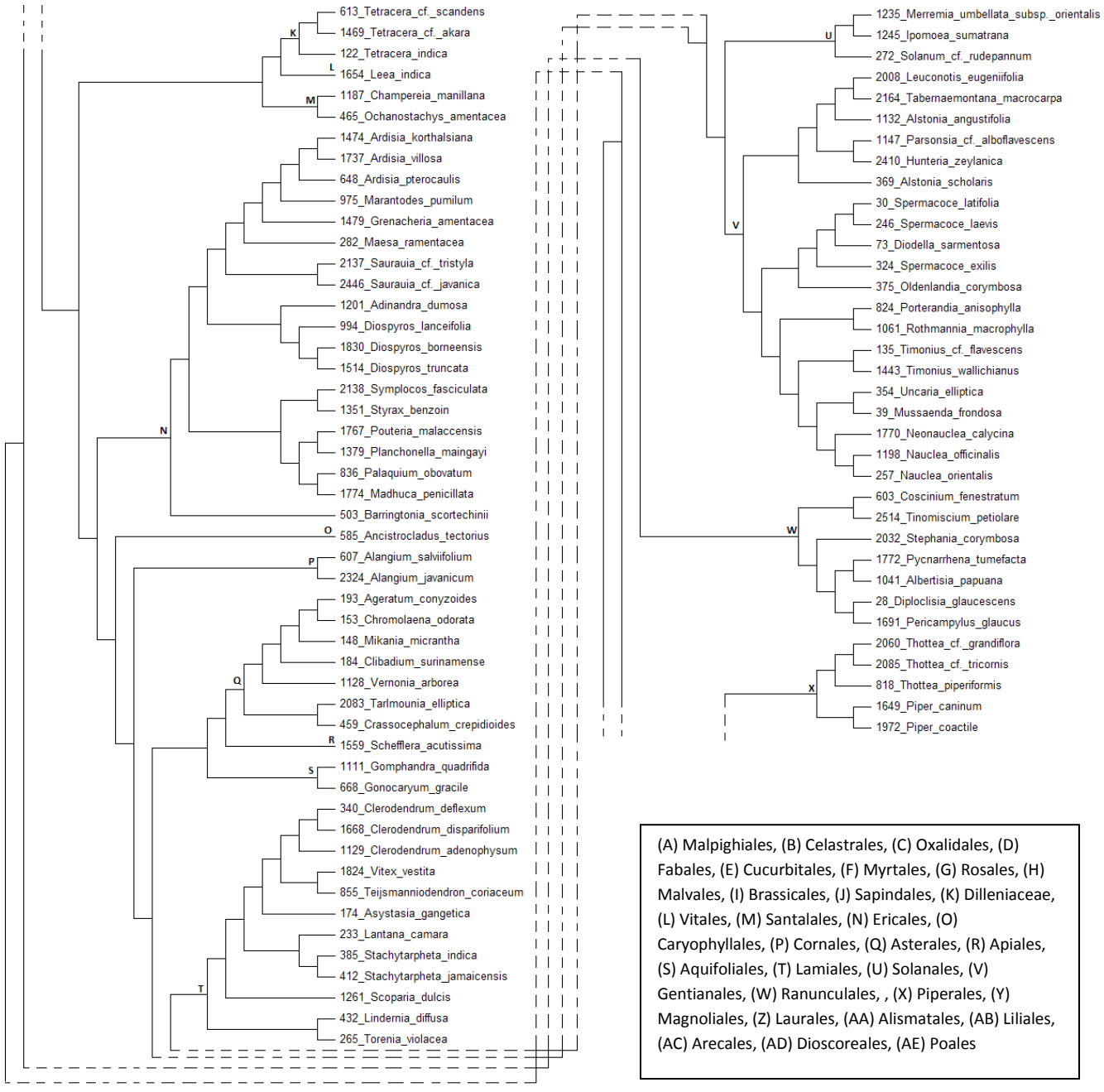
Appendix 4.2 Phylogenetic tree reconstructed using ML method based on matK barcodes



Appendix 4.2 (continued)

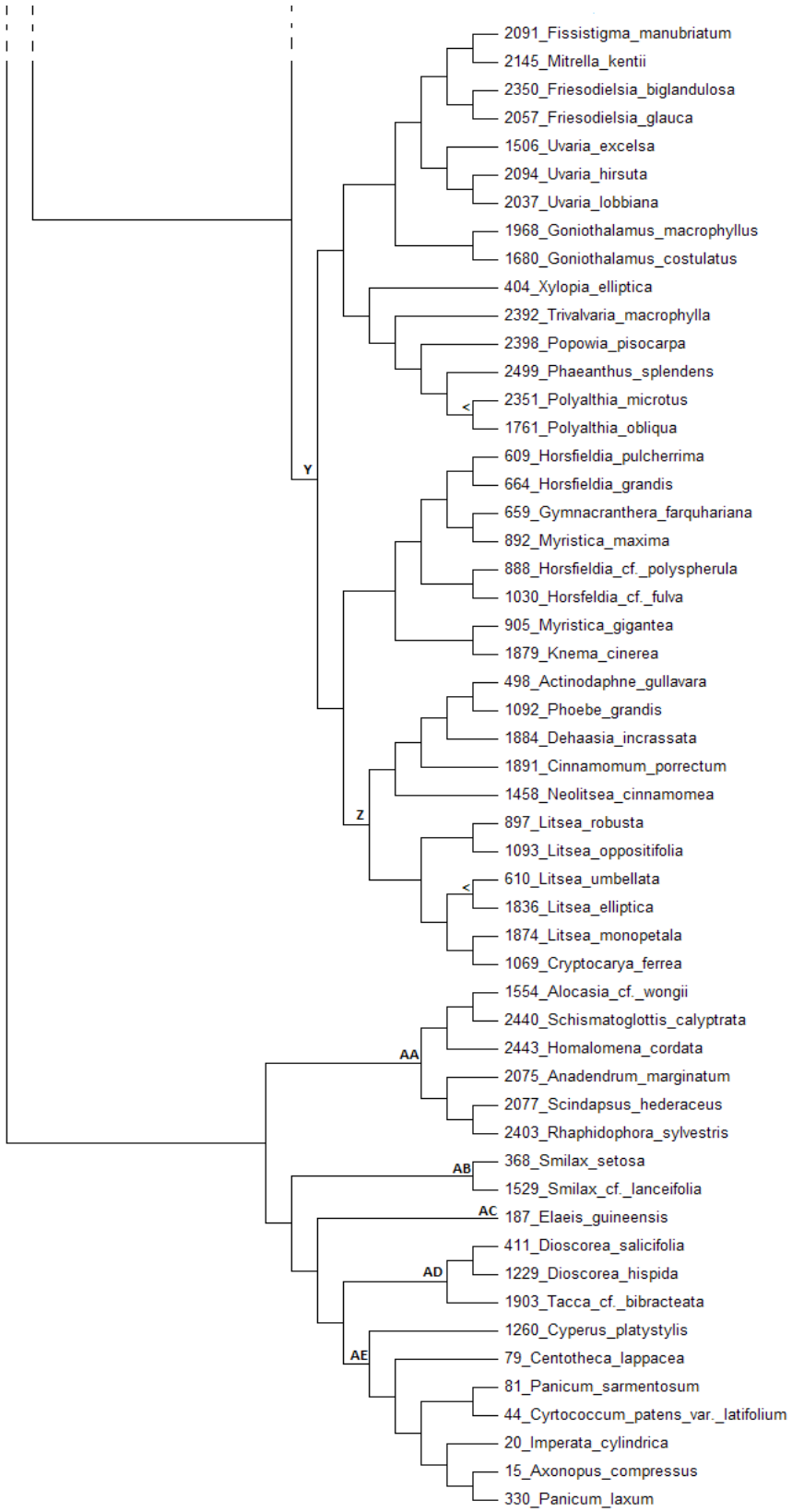


Appendix 4.2 (continued)

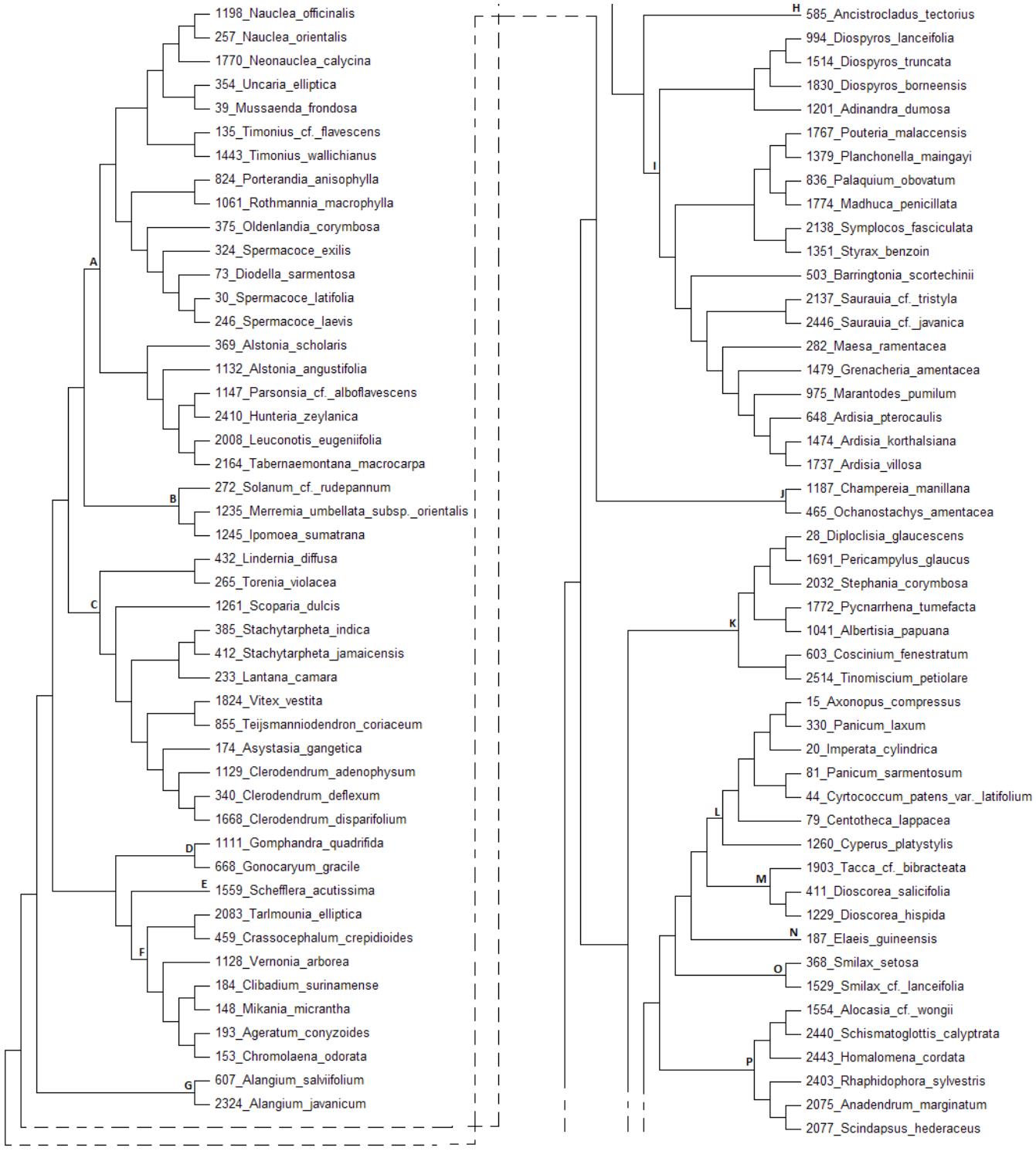


(A) Malpighiales, (B) Celastrales, (C) Oxalidales, (D) Fabales, (E) Cucurbitales, (F) Myrtales, (G) Rosales, (H) Malvales, (I) Brassicales, (J) Sapindales, (K) Dilleniaceae, (L) Vitales, (M) Santalales, (N) Ericales, (O) Caryophyllales, (P) Cornales, (Q) Asterales, (R) Apiales, (S) Aquifoliales, (T) Lamiales, (U) Solanales, (V) Gentianales, (W) Ranunculales, , (X) Piperales, (Y) Magnoliales, (Z) Laurales, (AA) Alismatales, (AB) Liliales, (AC) Arecales, (AD) Dioscoreales, (AE) Poales

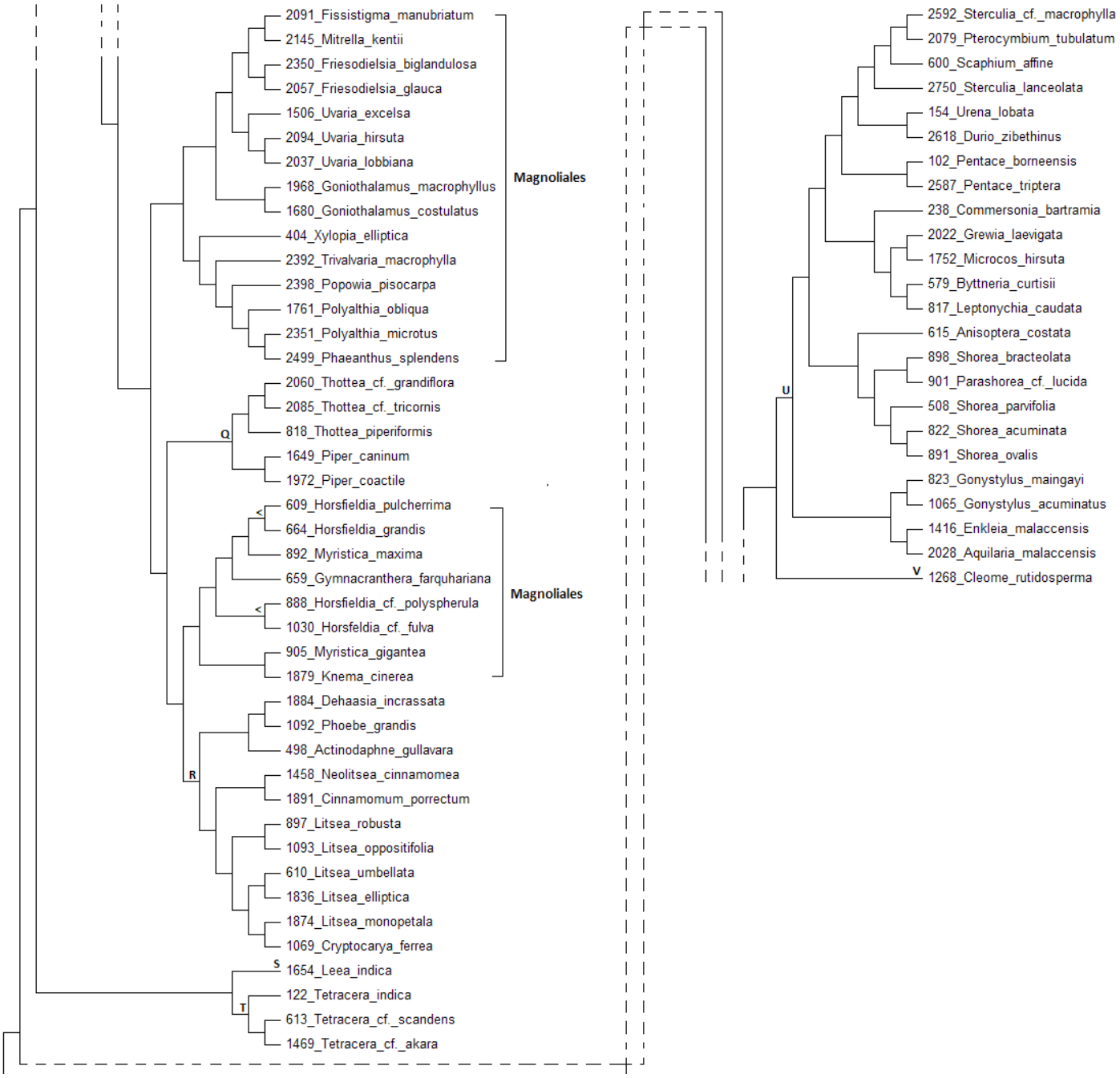
Appendix 4.2 (continued)



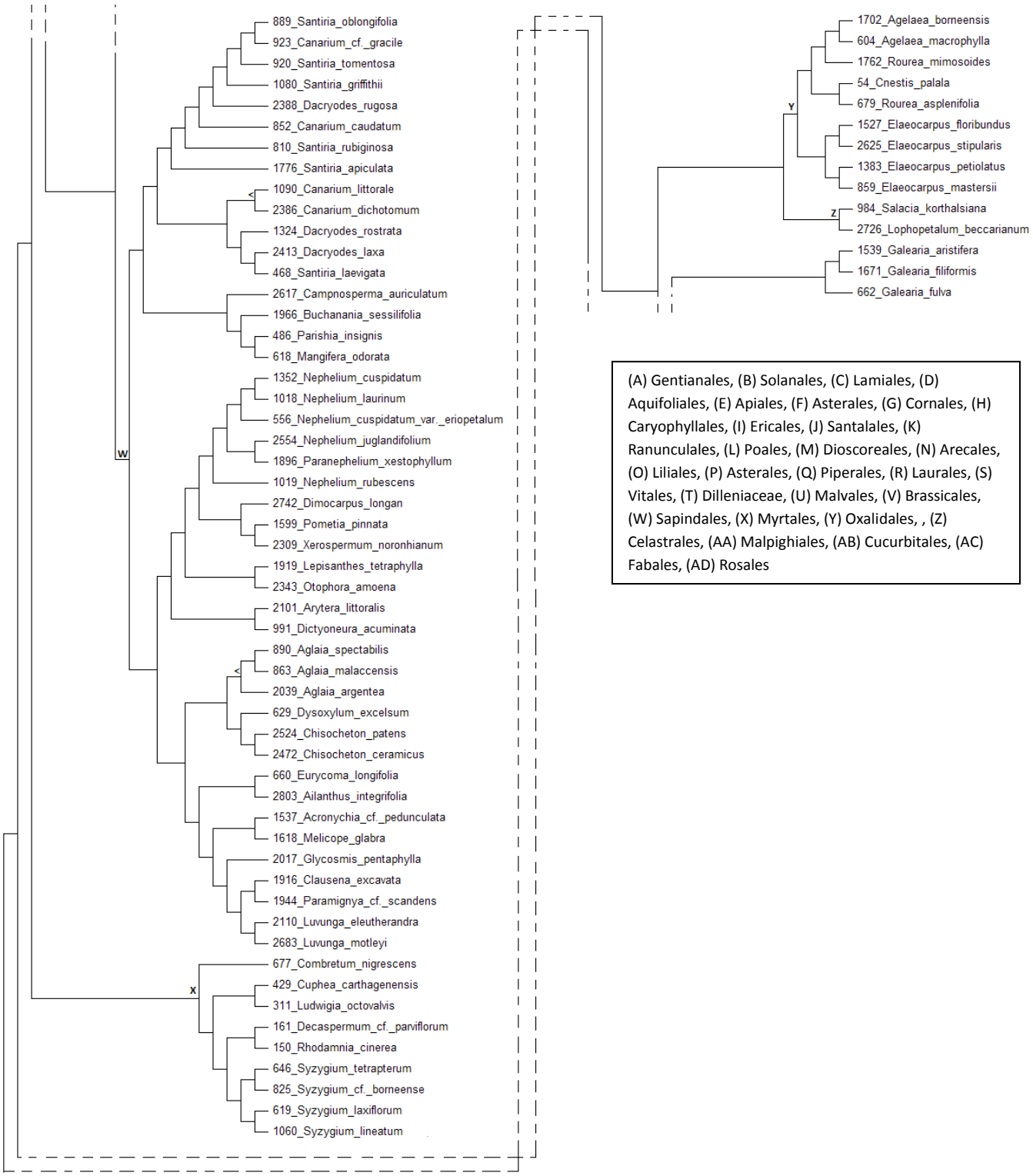
Appendix 4.3 Phylogenetic tree reconstructed using MP method based on matK barcodes



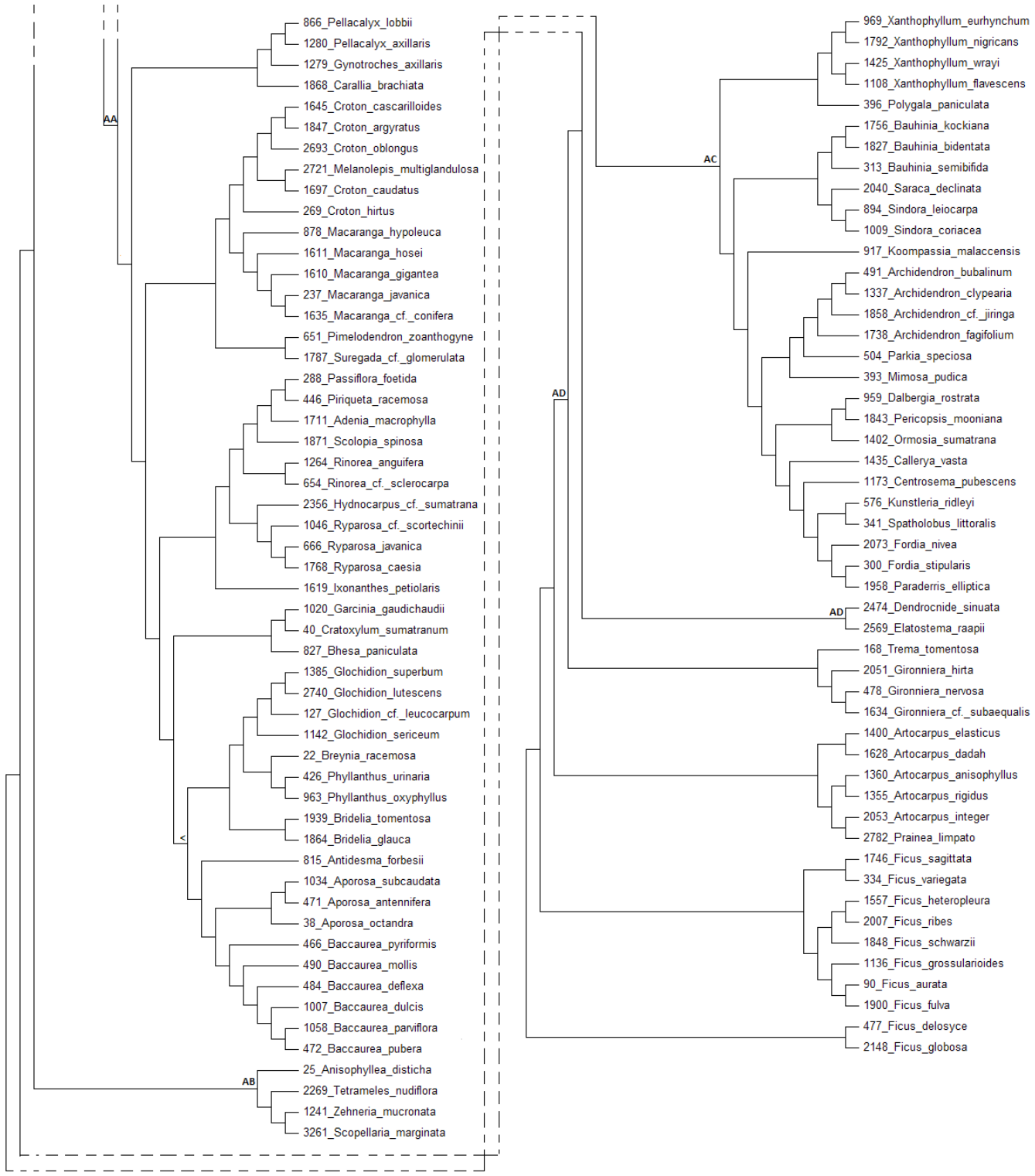
Appendix 4.3 (continued)



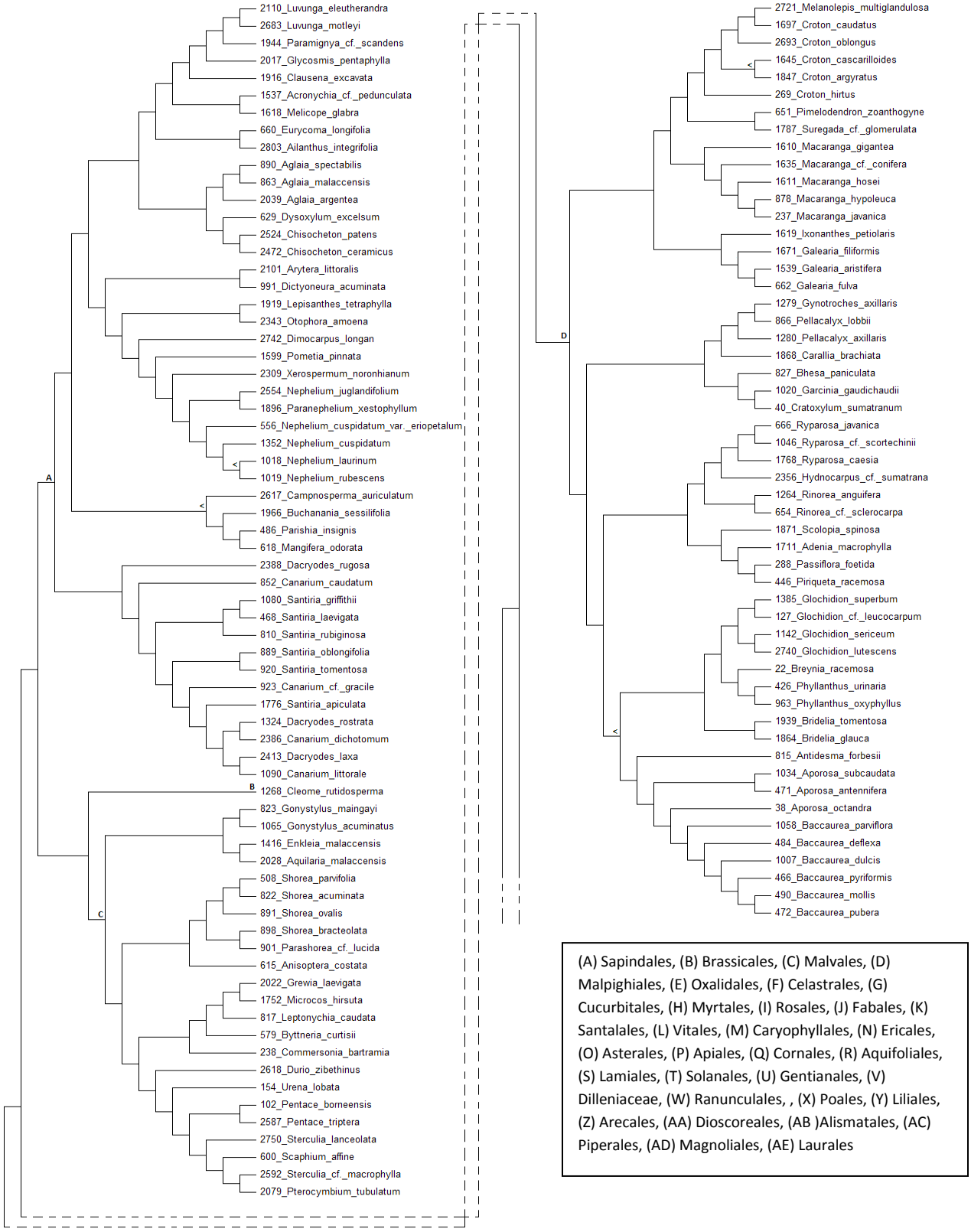
Appendix 4.3 (continued)



Appendix 4.3 (continued)

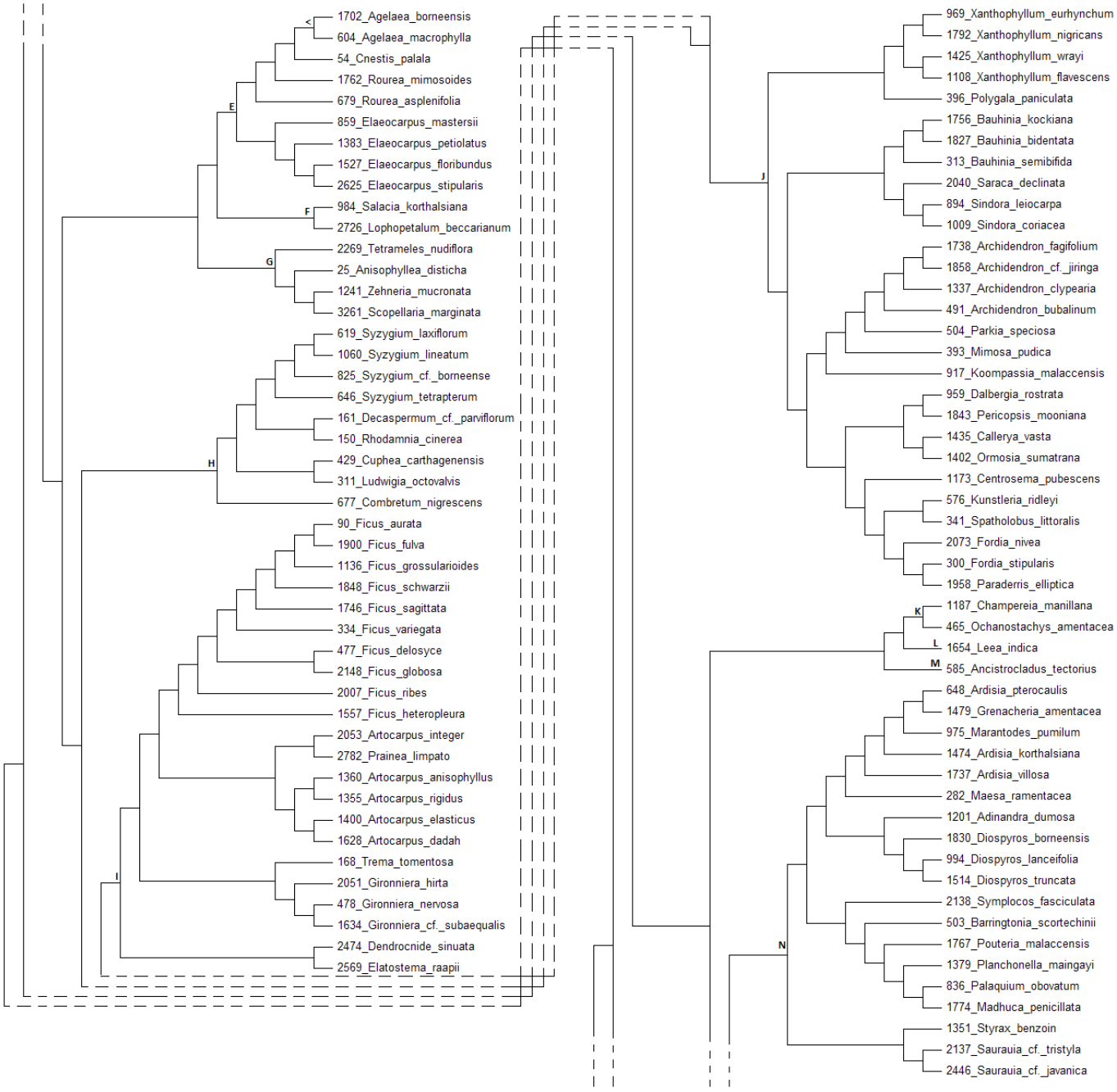


Appendix 4.4 Phylogenetic tree reconstructed using NJ method based on matK barcodes

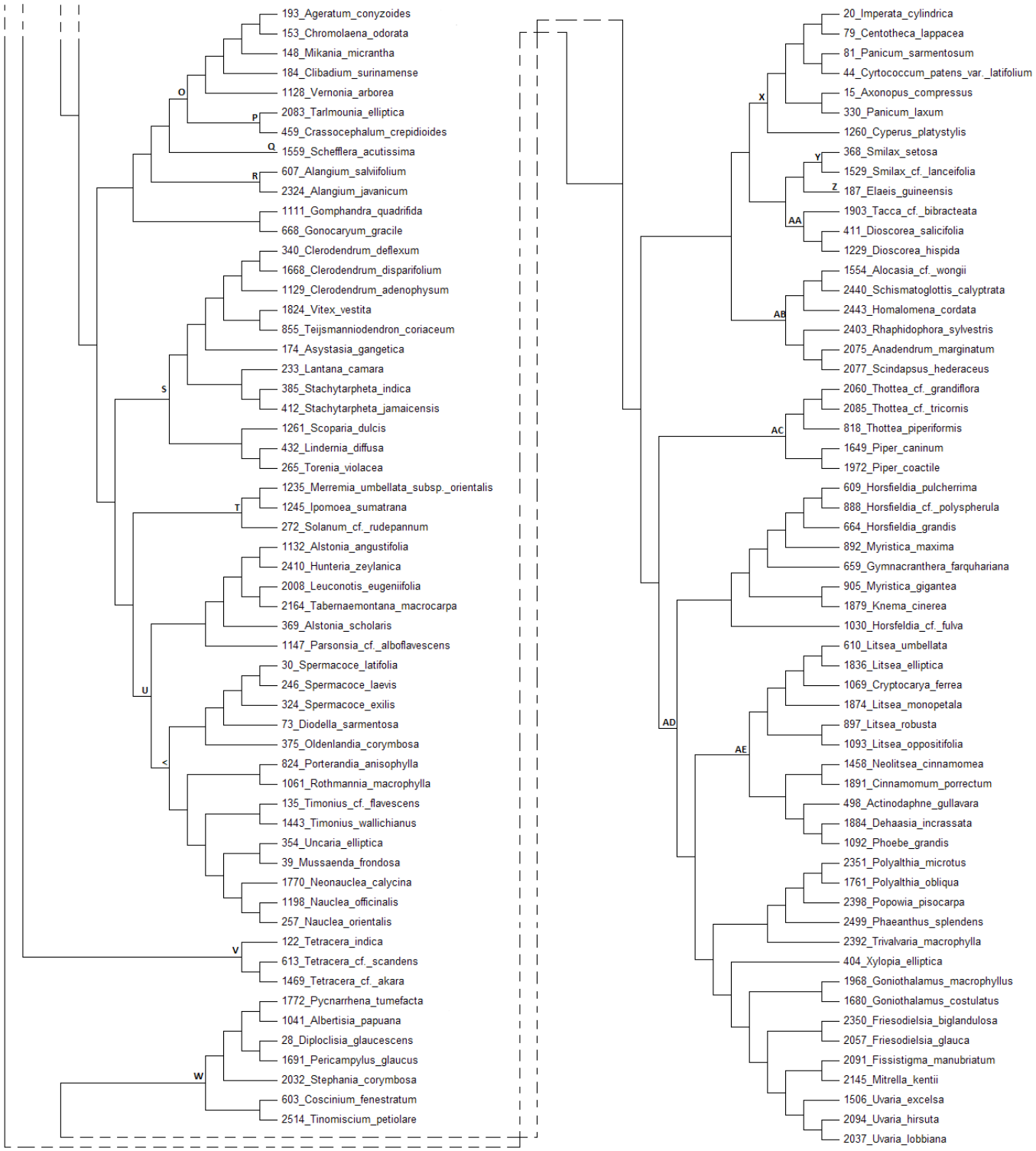


(A) Sapindales, (B) Brassicales, (C) Malvales, (D) Malpighiales, (E) Oxalidales, (F) Celastrales, (G) Cucurbitales, (H) Myrtales, (I) Rosales, (J) Fabales, (K) Santalales, (L) Vitales, (M) Caryophyllales, (N) Ericales, (O) Asterales, (P) Apiales, (Q) Cornales, (R) Aquifoliales, (S) Lamiales, (T) Solanales, (U) Gentianales, (V) Dilleniaceae, (W) Ranunculales, , (X) Poales, (Y) Liliales, (Z) Arecales, (AA) Dioscoreales, (AB)Alismatales, (AC) Piperales, (AD) Magnoliales, (AE) Laurales

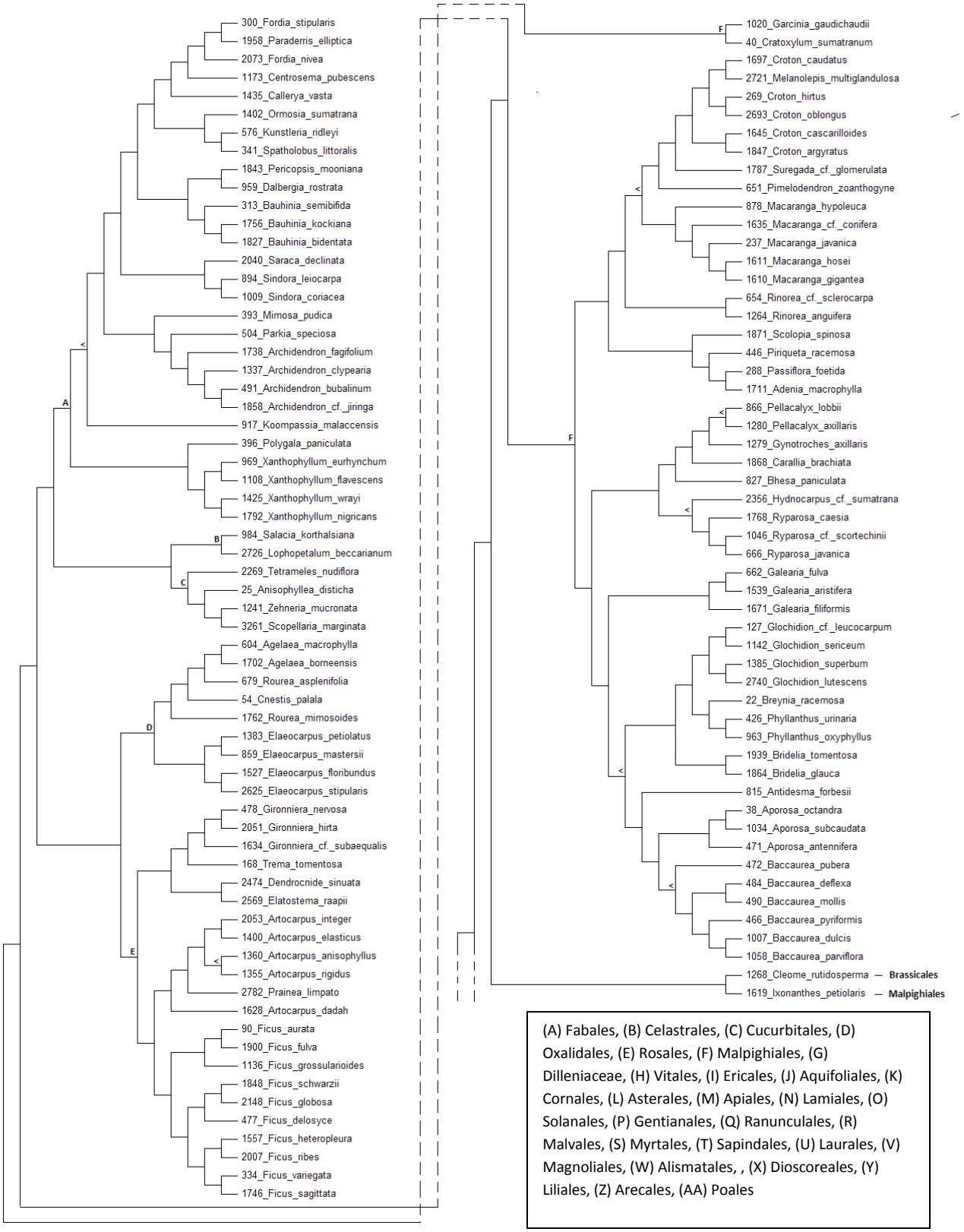
Appendix 4.4 (continued)



Appendix 4.4 (continued)

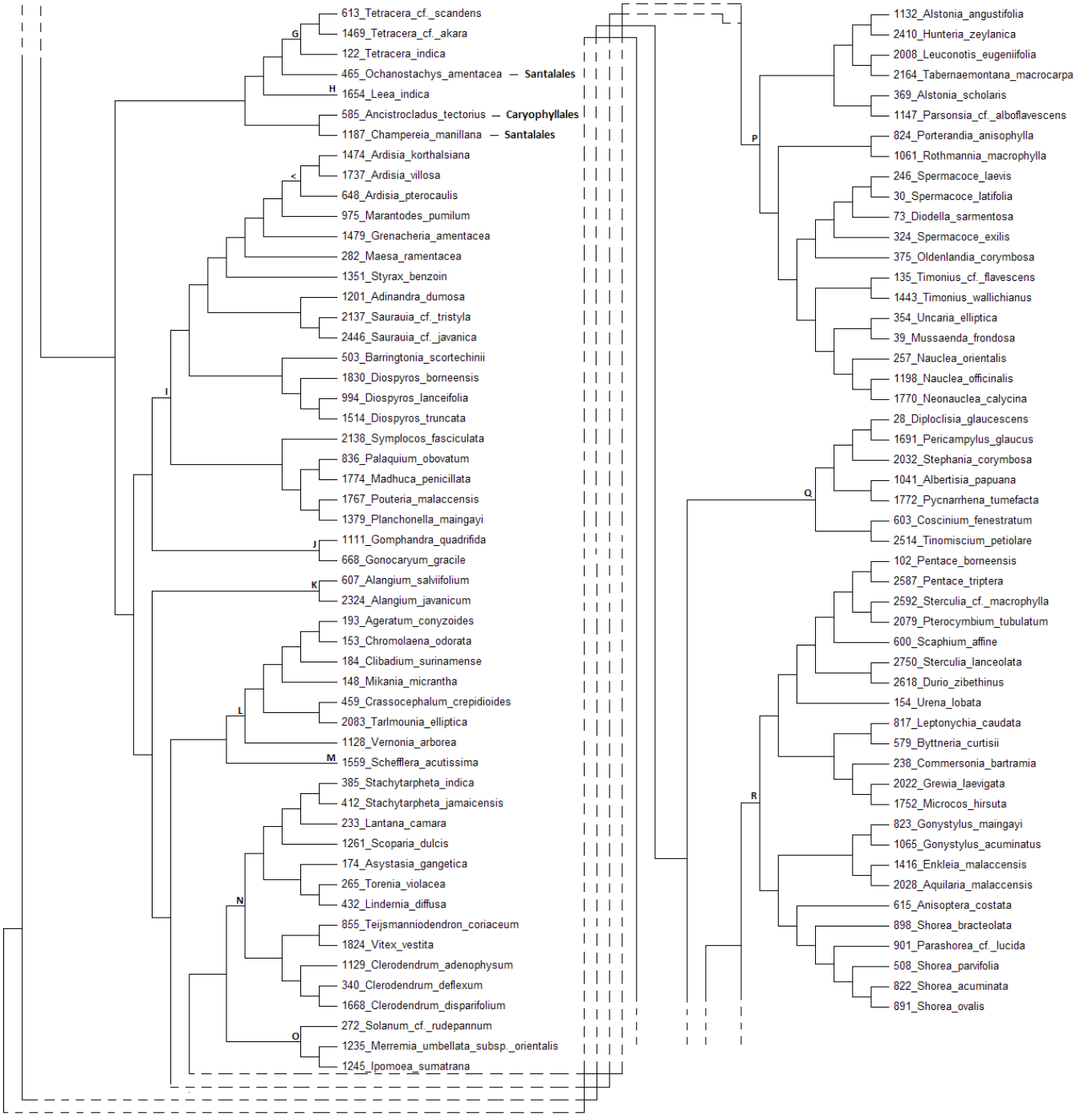


Appendix 4.5 Phylogenetic tree reconstructed using ML method based on rbcL barcodes

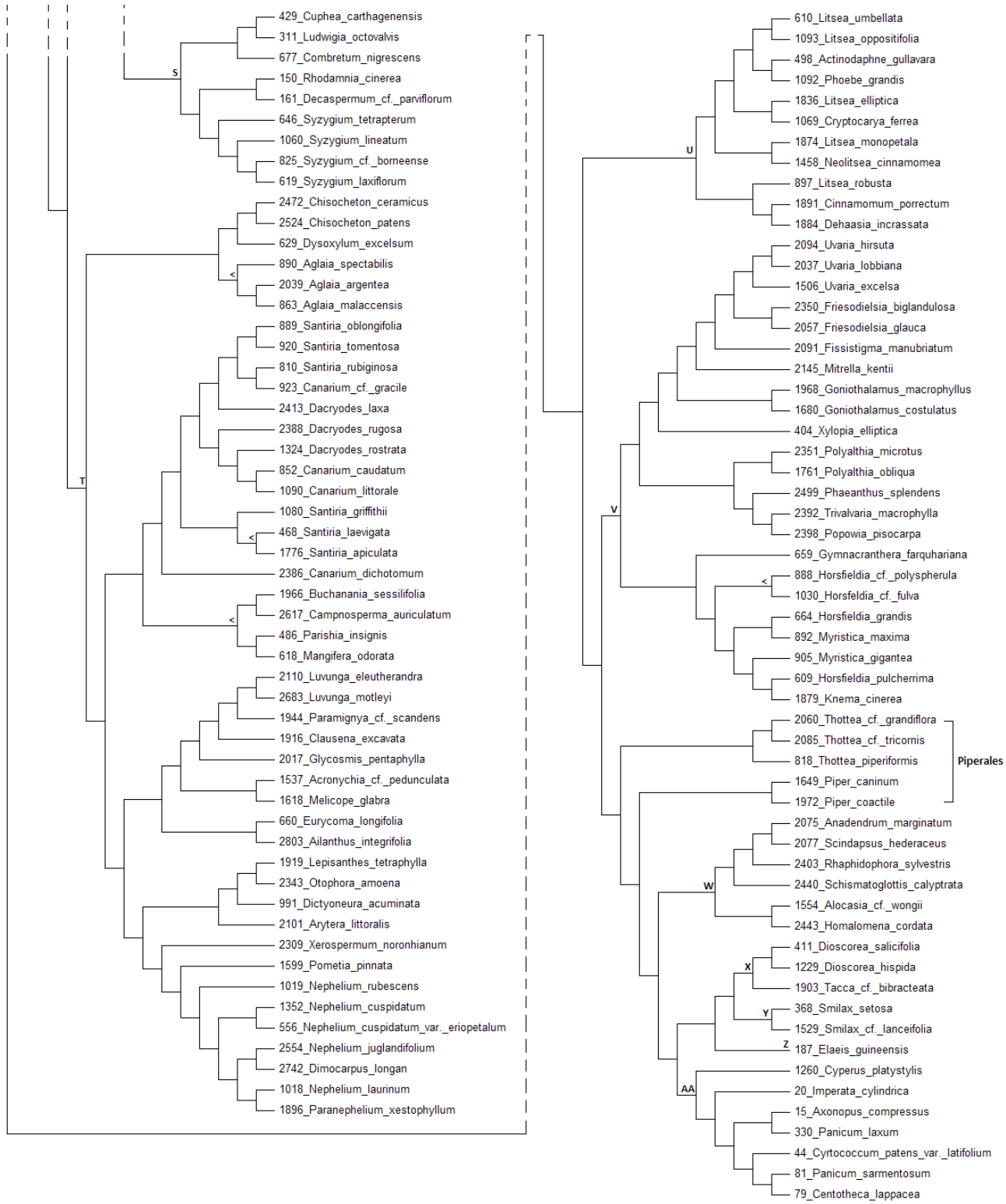


(A) Fabales, (B) Celastrales, (C) Cucurbitales, (D) Oxalidales, (E) Rosales, (F) Malpighiales, (G) Dilleniaceae, (H) Vitales, (I) Ericales, (J) Aquifoliales, (K) Cornales, (L) Asterales, (M) Apiales, (N) Lamiales, (O) Solanales, (P) Gentianales, (Q) Ranunculales, (R) Malvales, (S) Myrtales, (T) Sapindales, (U) Laurales, (V) Magnoliales, (W) Alismatales, , (X) Dioscoreales, (Y) Liliales, (Z) Arecales, (AA) Poales

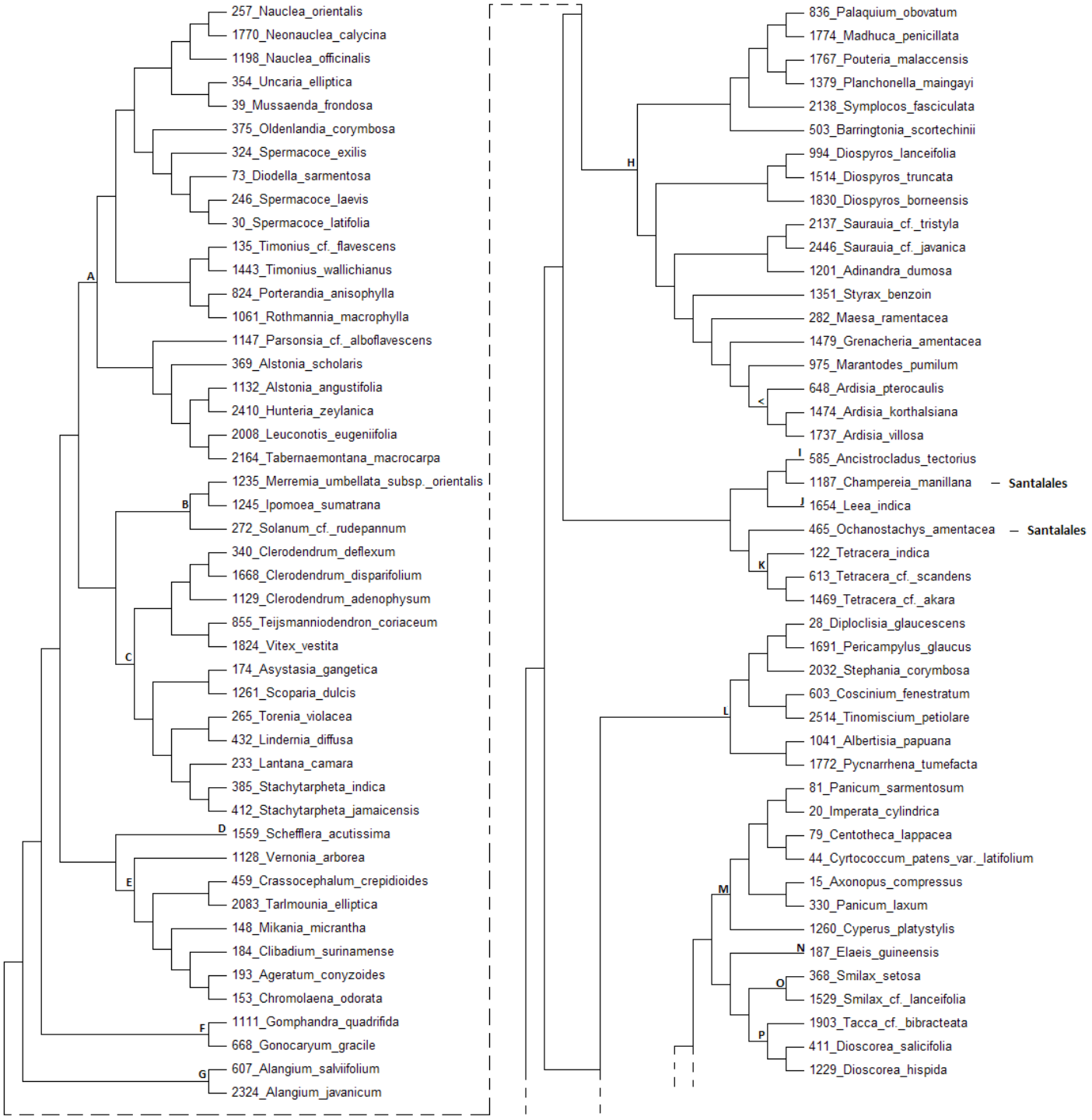
Appendix 4.5 (continued)



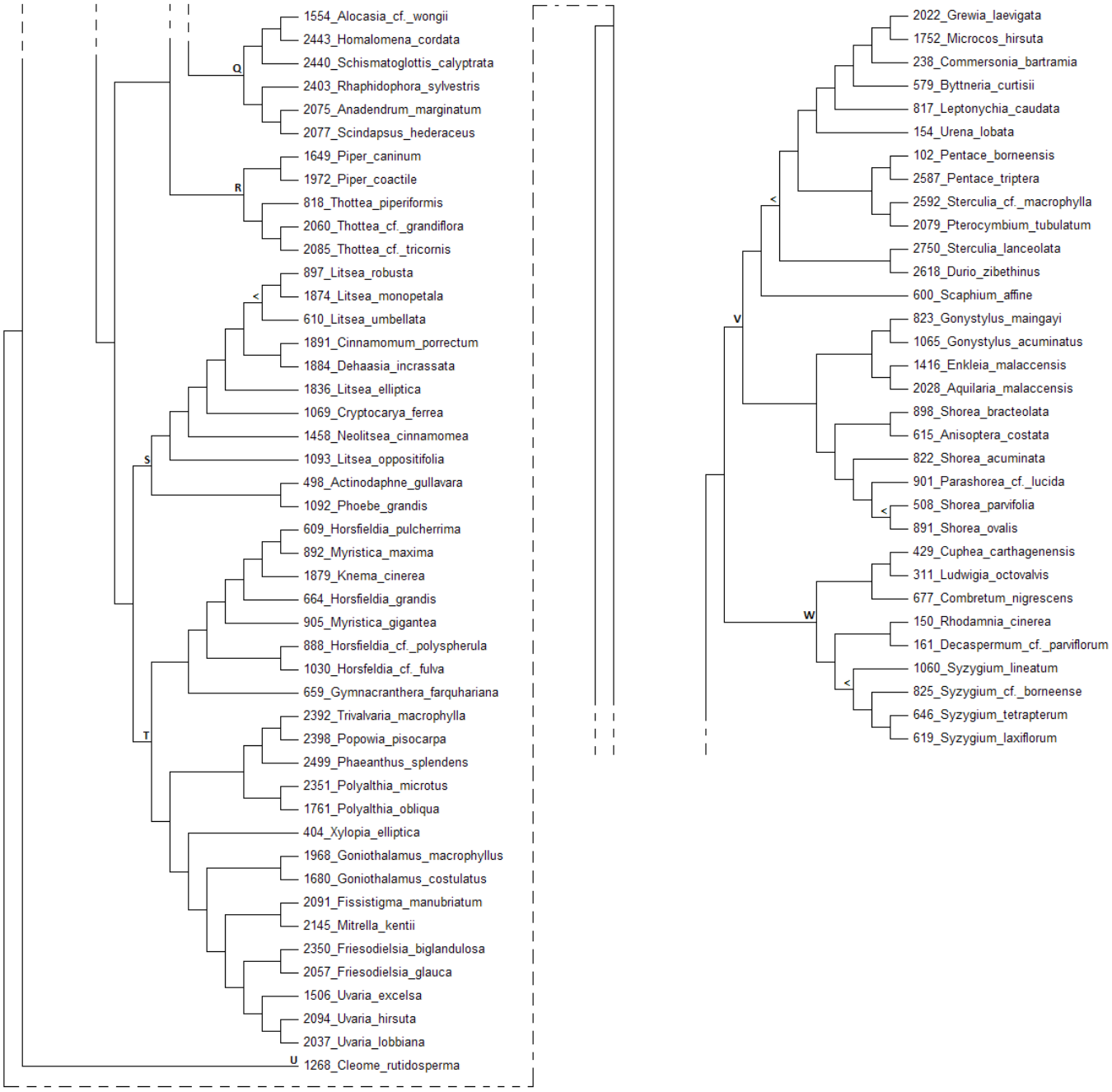
Appendix 4.5 (continued)



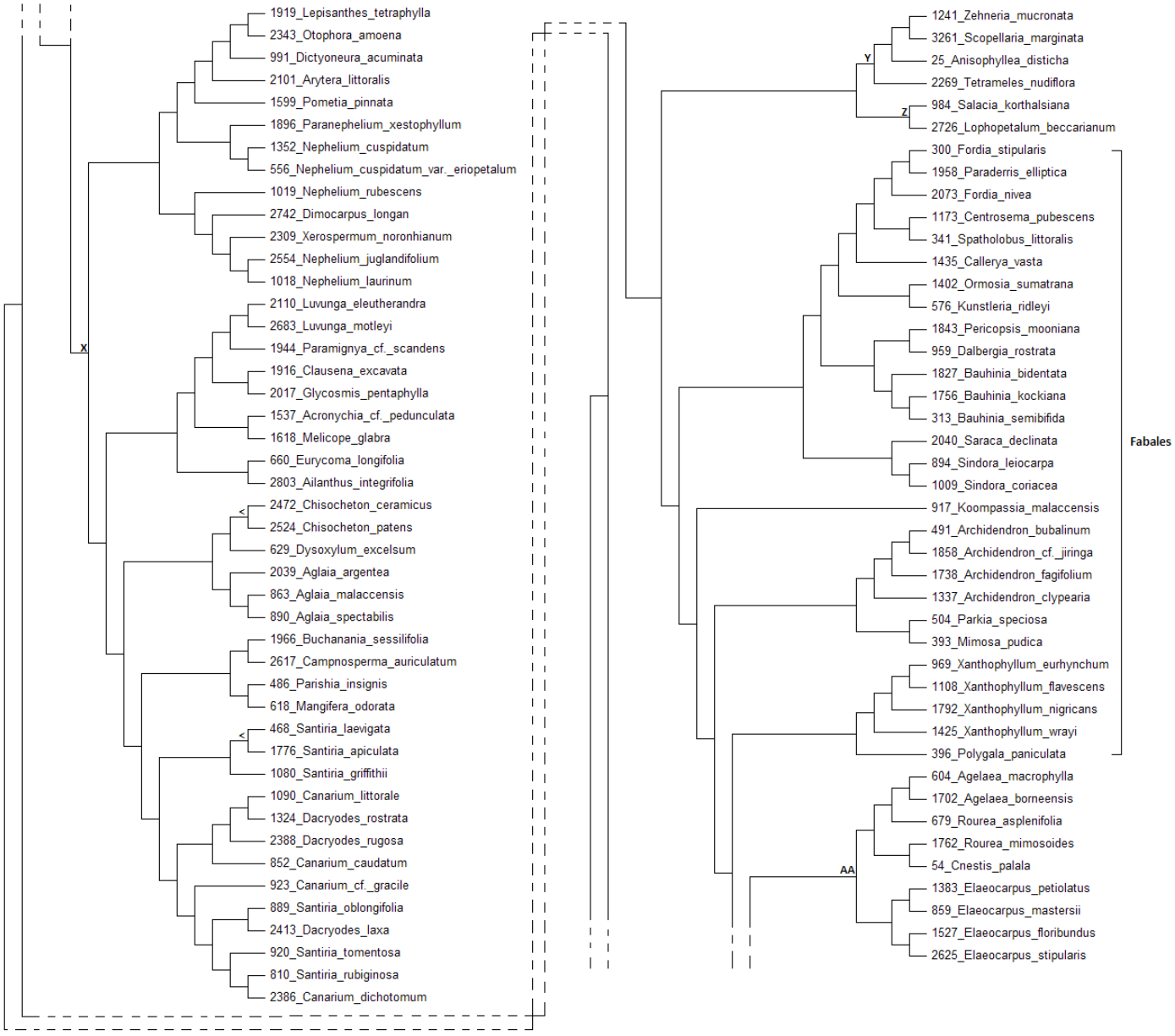
Appendix 4.6 Phylogenetic tree reconstructed using MP method based on rbcL barcodes



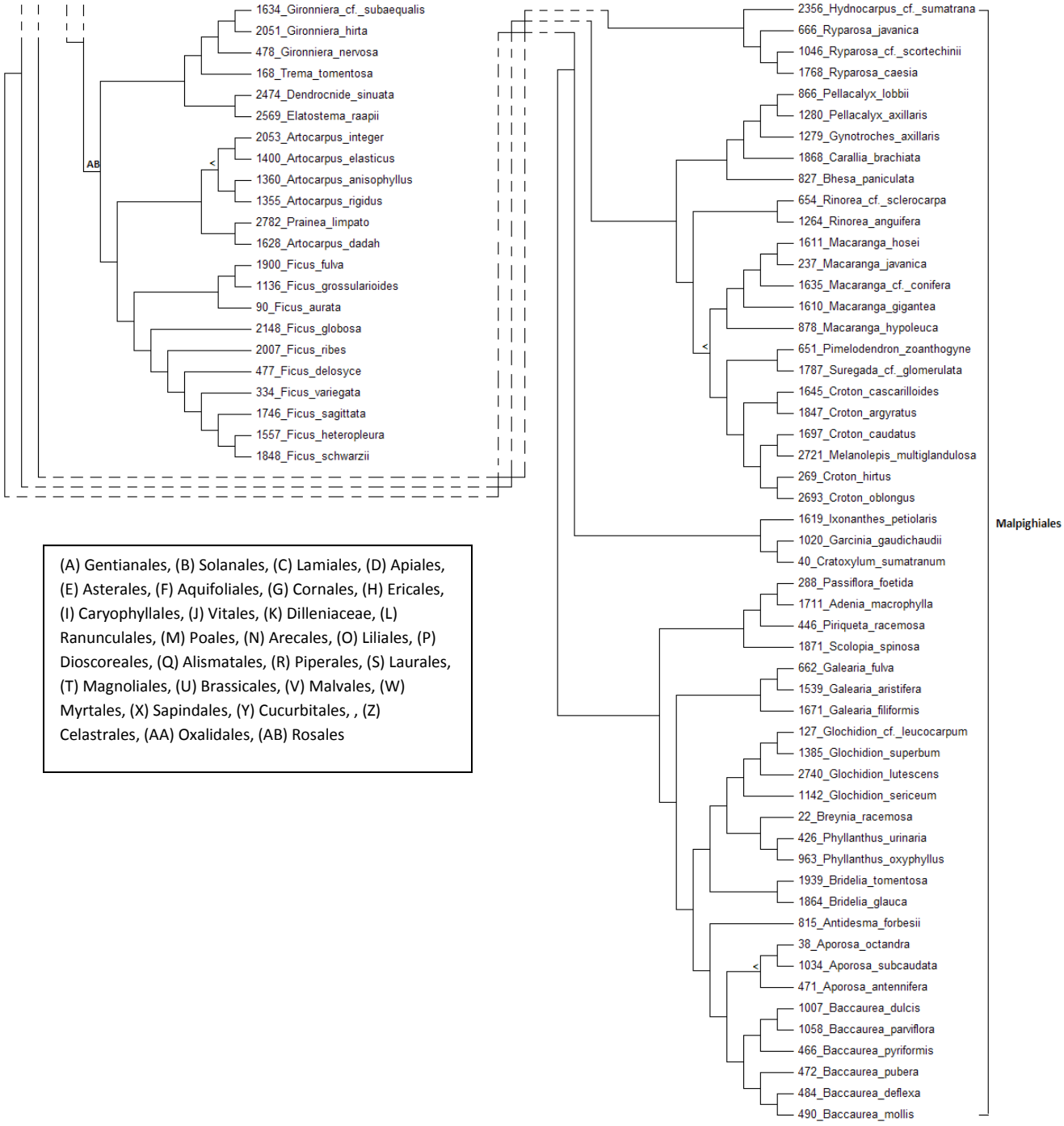
Appendix 4.6 (continued)



Appendix 4.6 (continued)

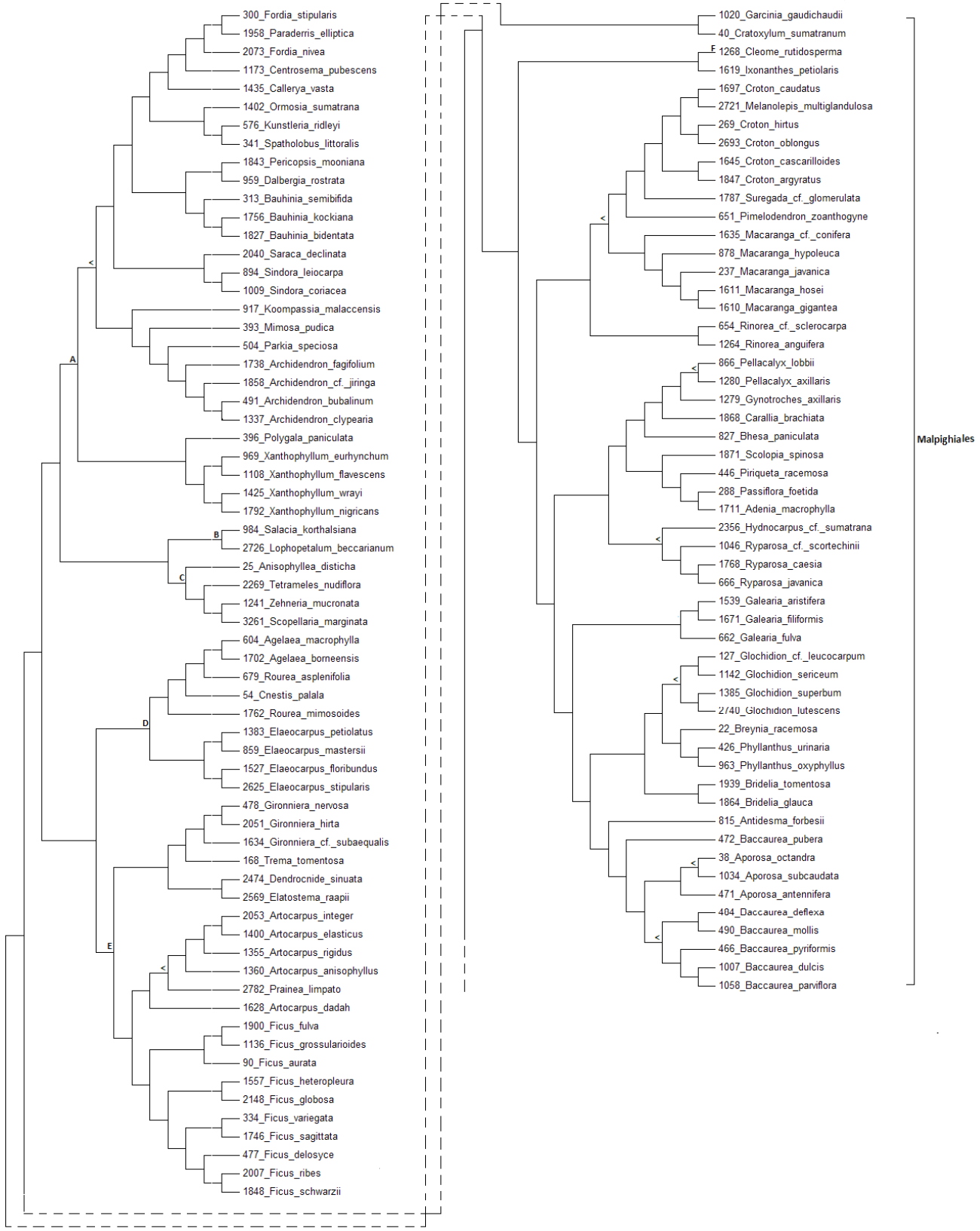


Appendix 4.6 (continued)

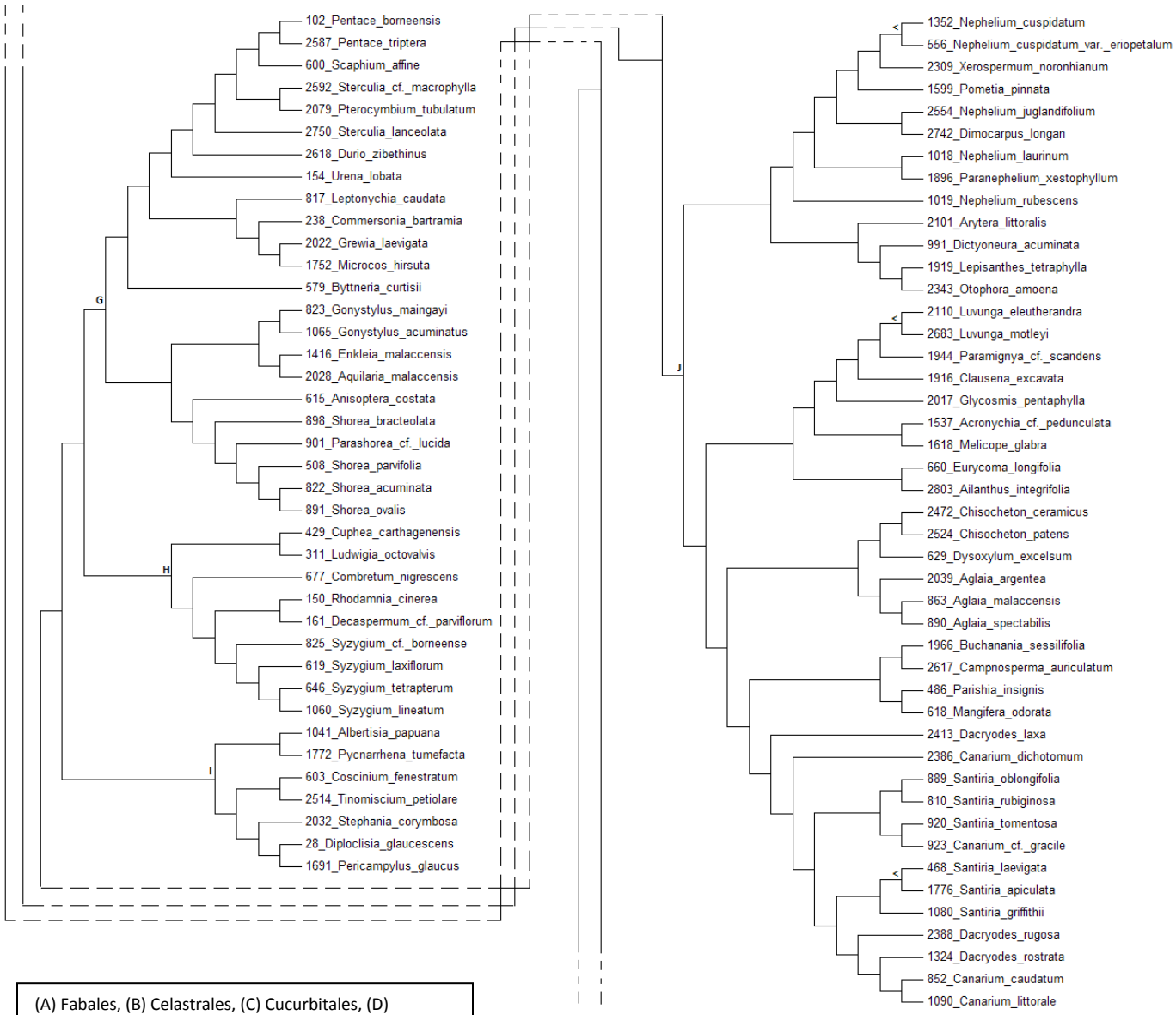


(A) Gentianales, (B) Solanales, (C) Lamiales, (D) Apiales, (E) Asterales, (F) Aquifoliales, (G) Cornales, (H) Ericales, (I) Caryophyllales, (J) Vitales, (K) Dilleniaceae, (L) Ranunculales, (M) Poales, (N) Arecales, (O) Liliales, (P) Dioscoreales, (Q) Alismatales, (R) Piperales, (S) Laurales, (T) Magnoliales, (U) Brassicales, (V) Malvales, (W) Myrtales, (X) Sapindales, (Y) Cucurbitales, , (Z) Celastrales, (AA) Oxalidales, (AB) Rosales

Appendix 4.7 Phylogenetic tree reconstructed using NJ method based on rbcL barcodes

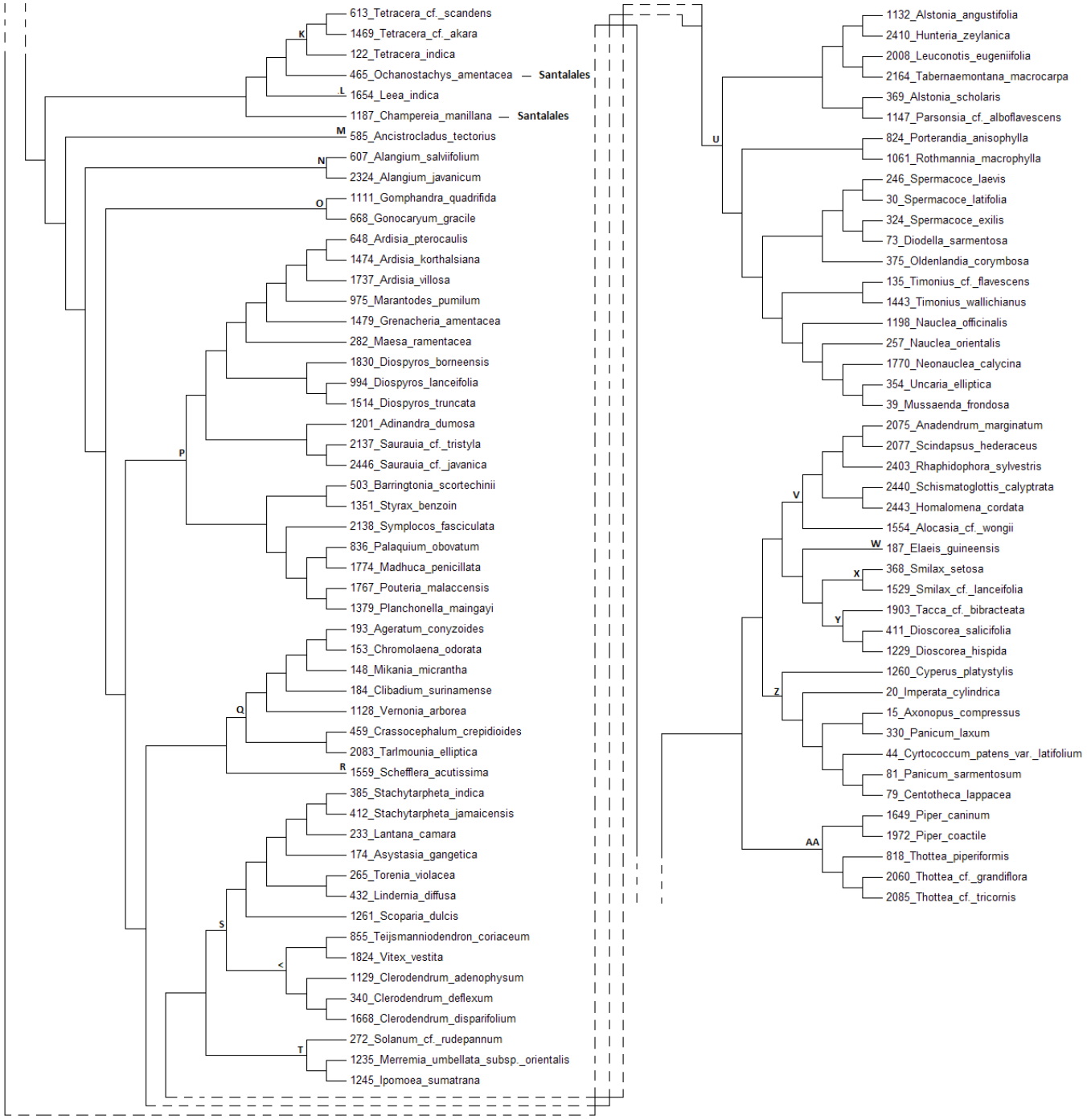


Appendix 4.7 (continued)

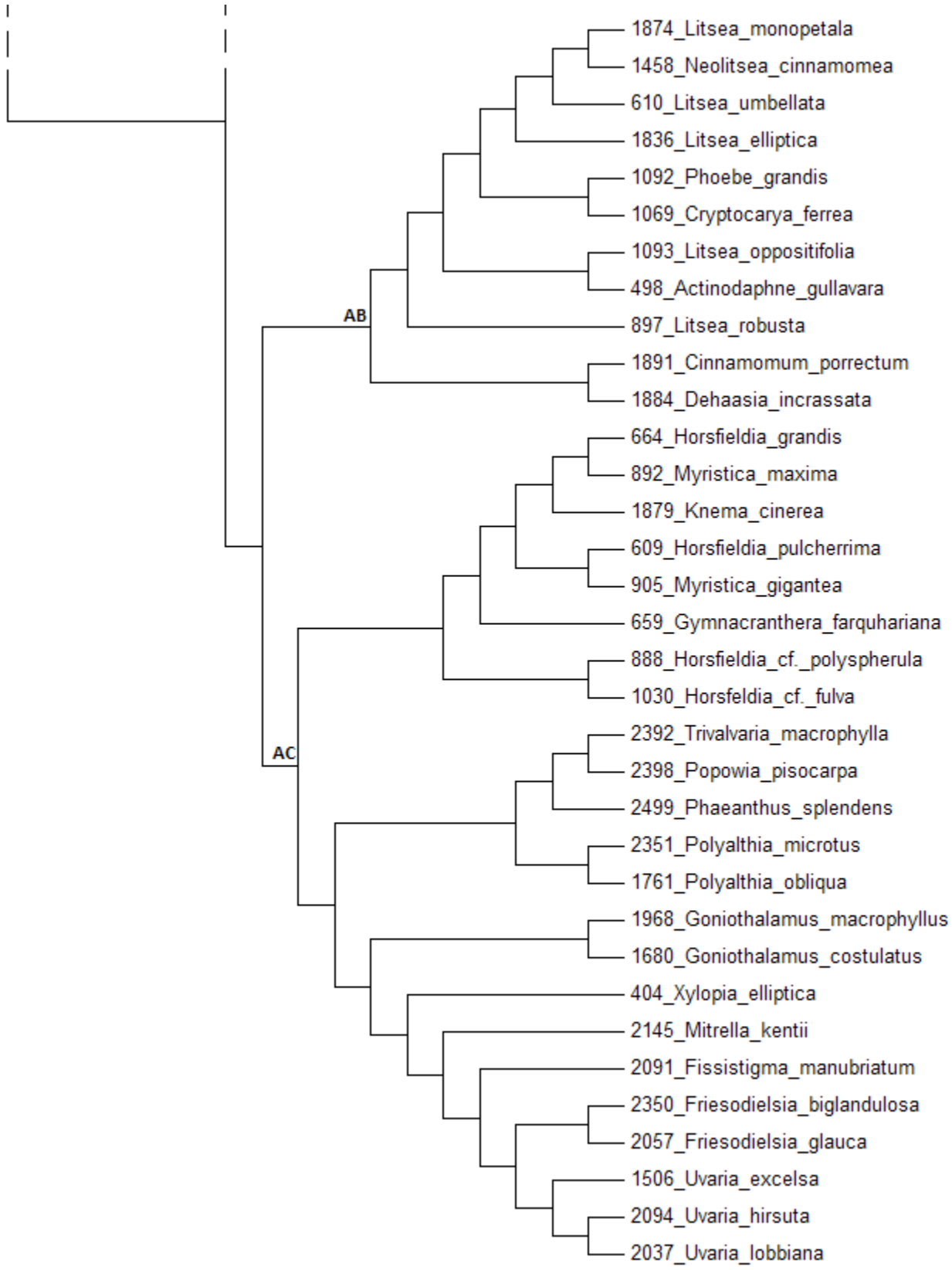


(A) Fabales, (B) Celastrales, (C) Cucurbitales, (D) Oxalidales, (E) Rosales, (F) Brassicales, (G) Malvales, (H) Myrtales, (I) Ranunculales, (J) Sapindales, (K) Dilleniaceae, (L) Vitales, (M) Caryophyllales, (N) Cornales, (O) Aquifoliales, (P) Ericales, (Q) Asterales, (R) Apiales, (S) Lamiales, (T) Solanales, (U) Gentianales, (V) Alismatales, (W) Asterales, (X) Liliales, (Y) Dioscoreales, (Z) Poales, (AA) Piperales, (AB) Laurales, (AC) Magnoliales

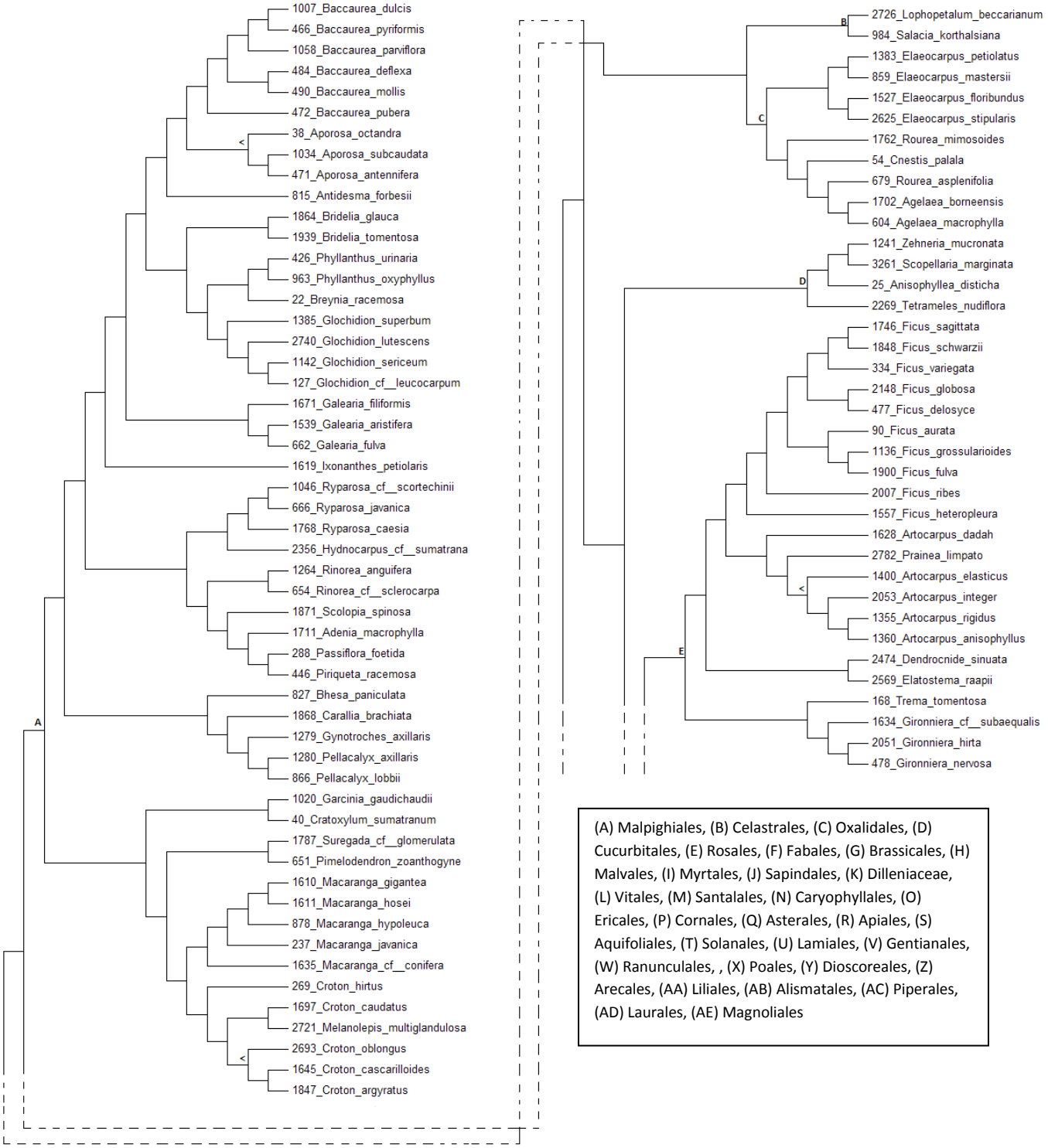
Appendix 4.7 (continued)



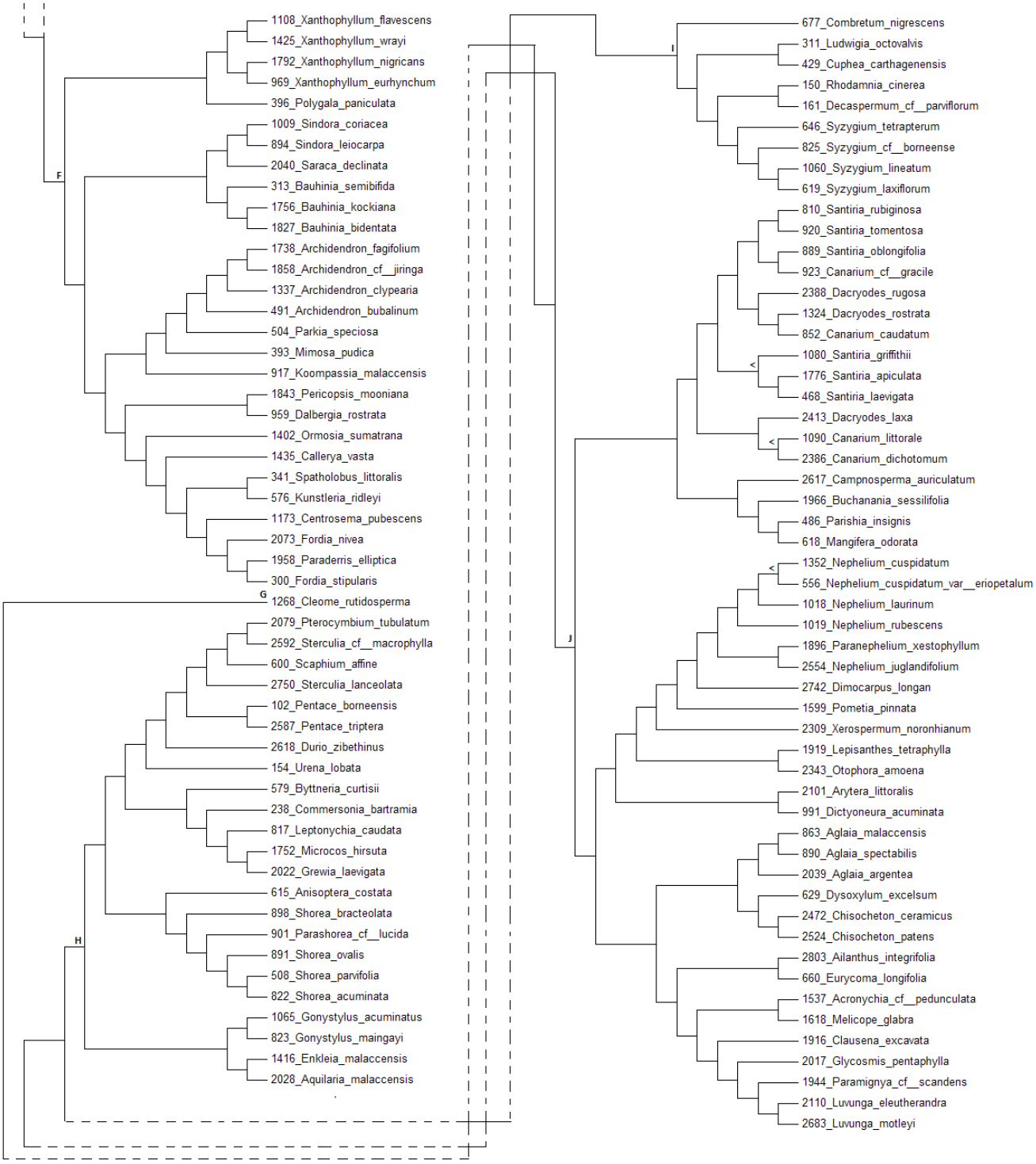
Appendix 4.7 (continued)



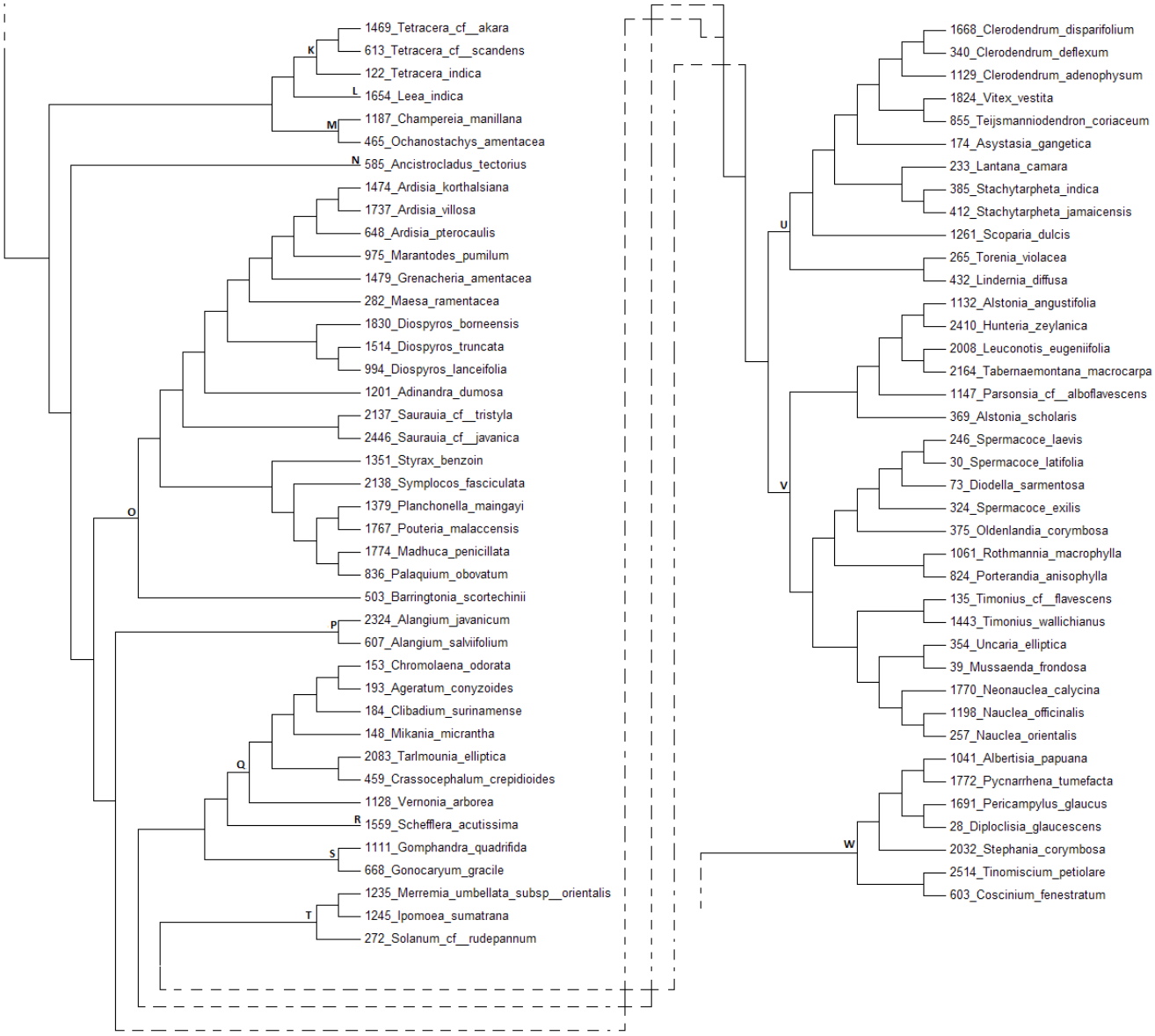
Appendix 4.8 Phylogenetic tree reconstructed using ML method based on matK+rbcL barcodes



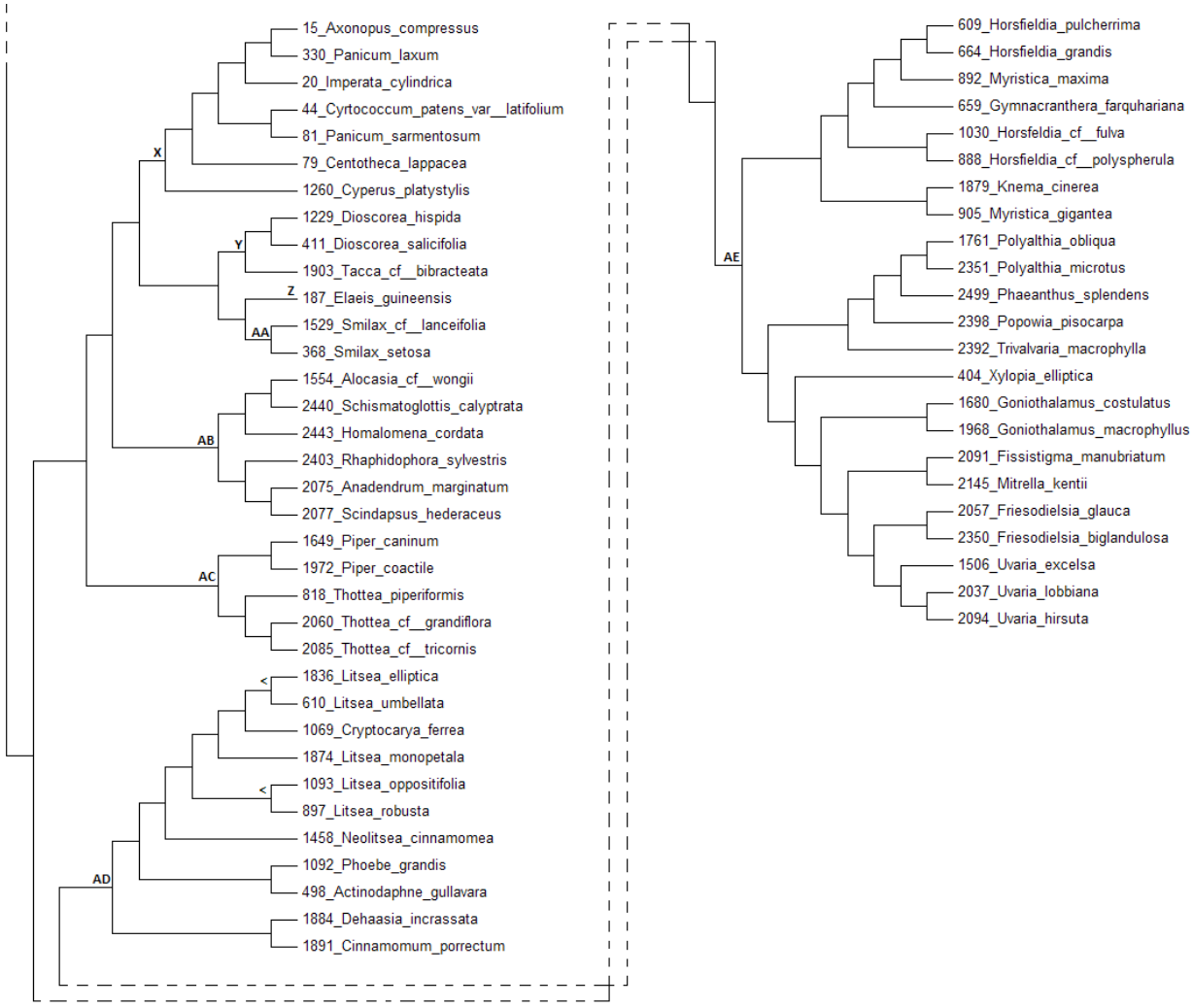
Appendix 4.8 (continued)



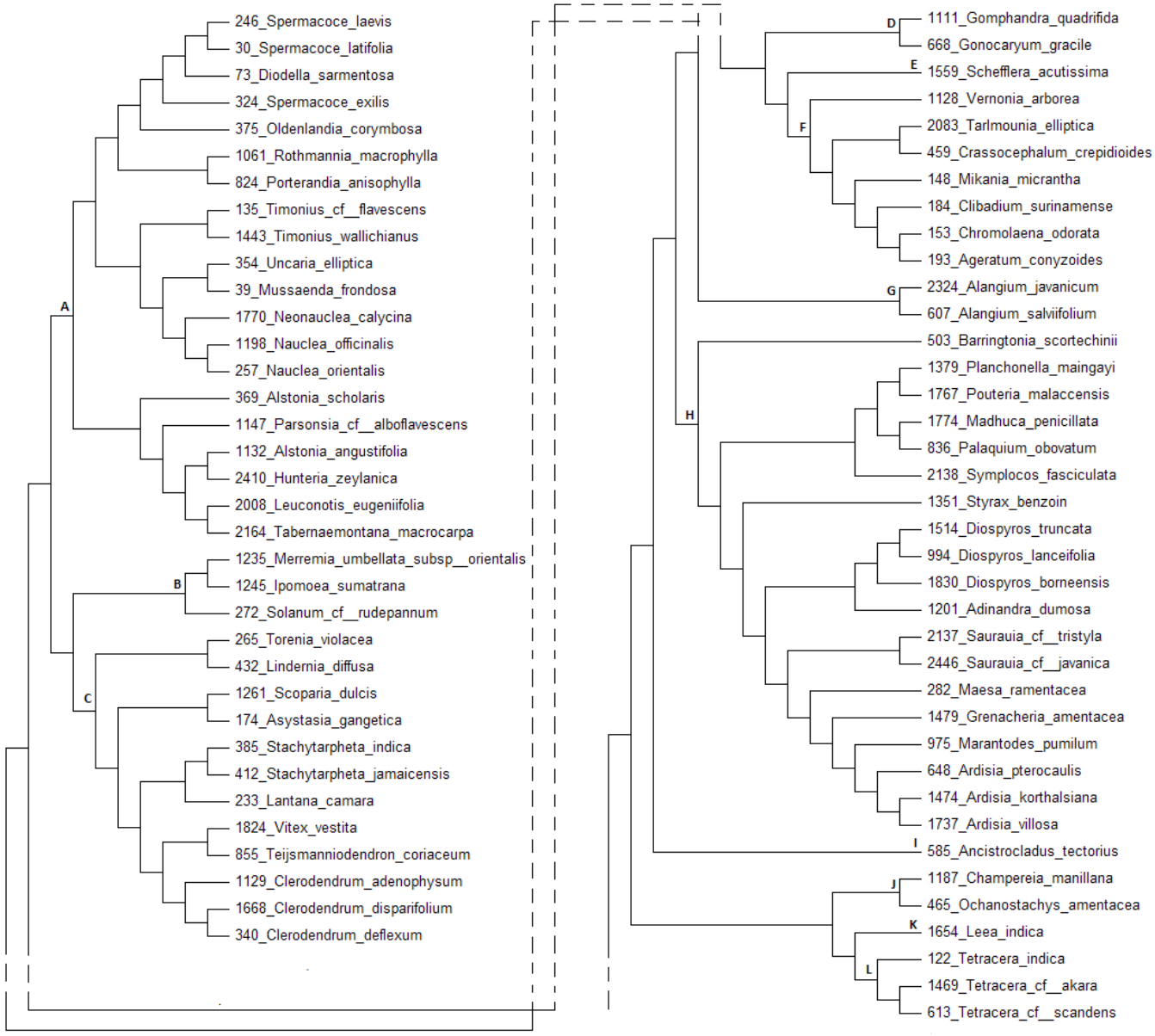
Appendix 4.8 (continued)



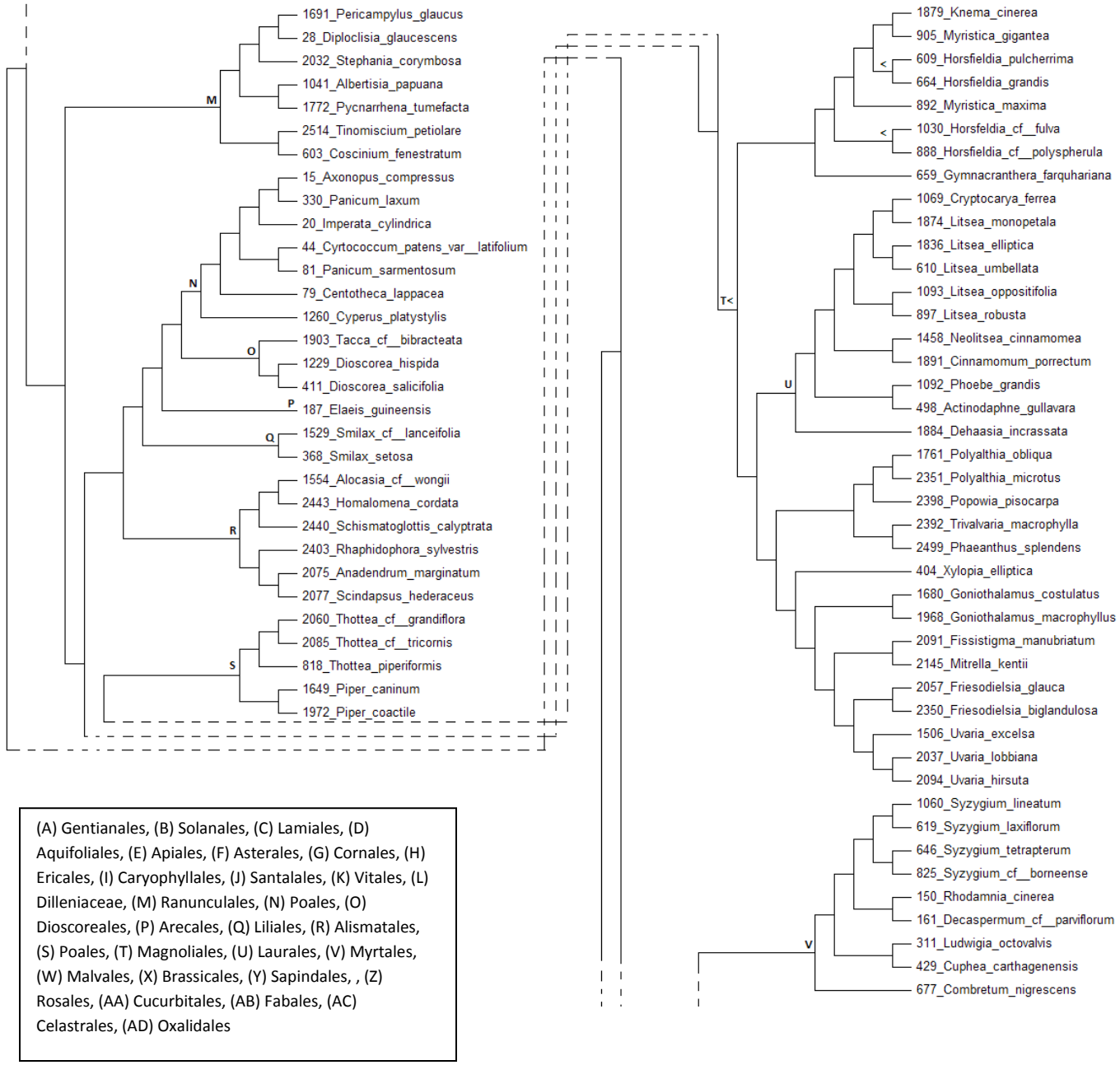
Appendix 4.8 (continued)



Appendix 4.9 Phylogenetic tree reconstructed using MP method based on matK+rbcL barcodes

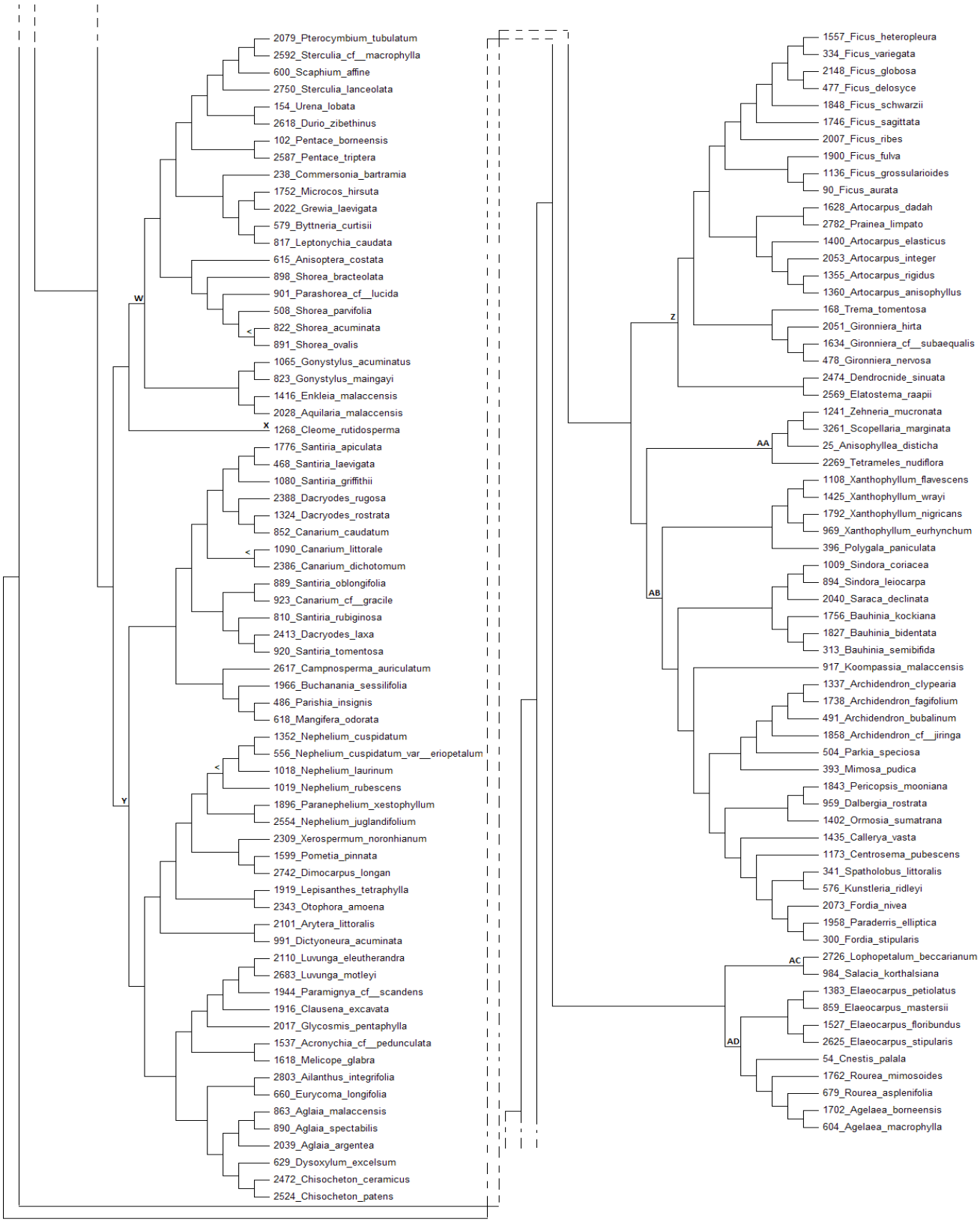


Appendix 4.9 (continued)

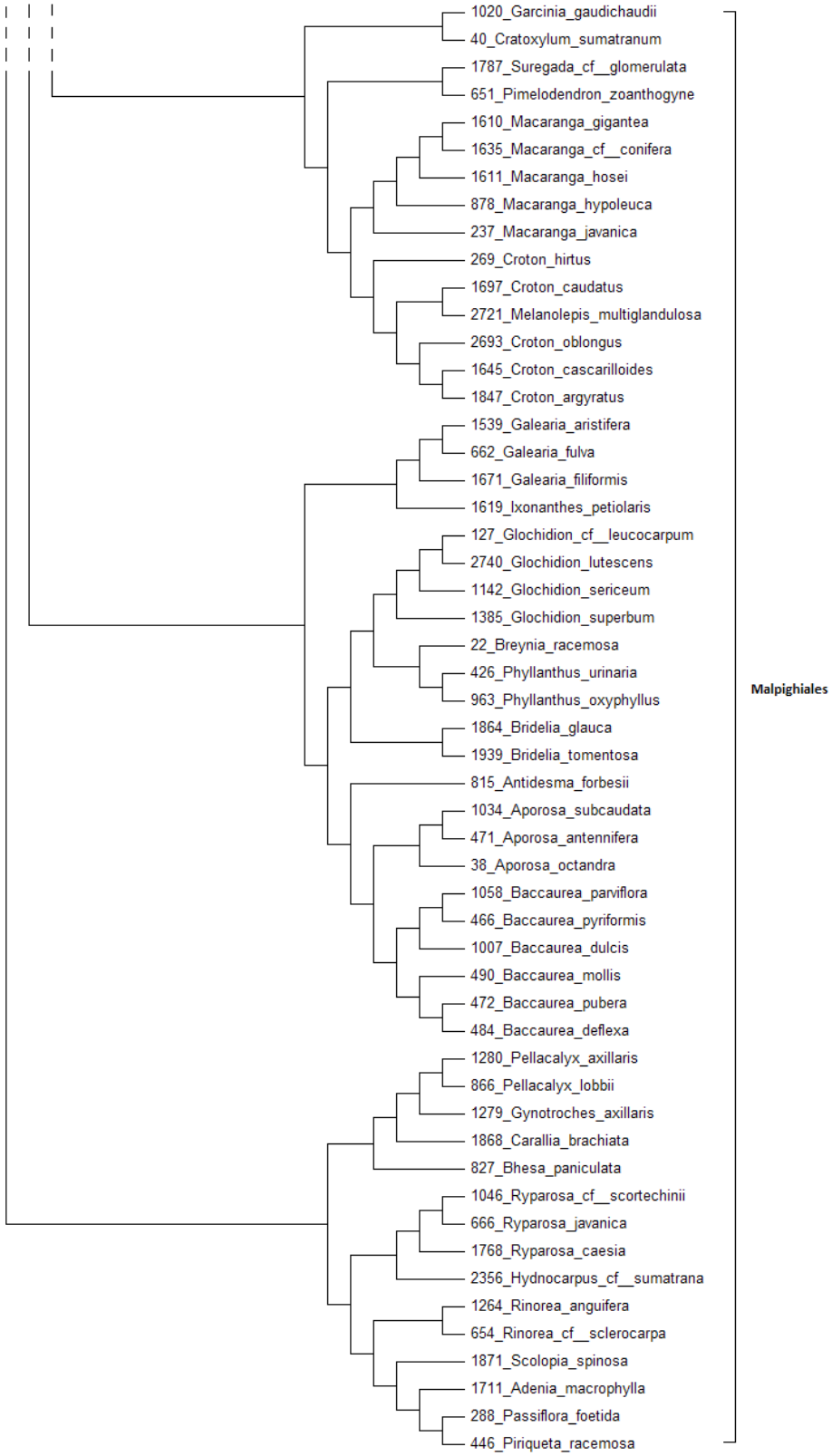


(A) Gentianales, (B) Solanales, (C) Lamiales, (D) Aquifoliales, (E) Apiales, (F) Asterales, (G) Cornales, (H) Ericales, (I) Caryophyllales, (J) Santalales, (K) Vitales, (L) Dilleniaceae, (M) Ranunculales, (N) Poales, (O) Dioscoreales, (P) Arecales, (Q) Liliales, (R) Alismatales, (S) Poales, (T) Magnoliales, (U) Laurales, (V) Myrtales, (W) Malvales, (X) Brassicales, (Y) Sapindales, , (Z) Rosales, (AA) Cucurbitales, (AB) Fabales, (AC) Celastrales, (AD) Oxalidales

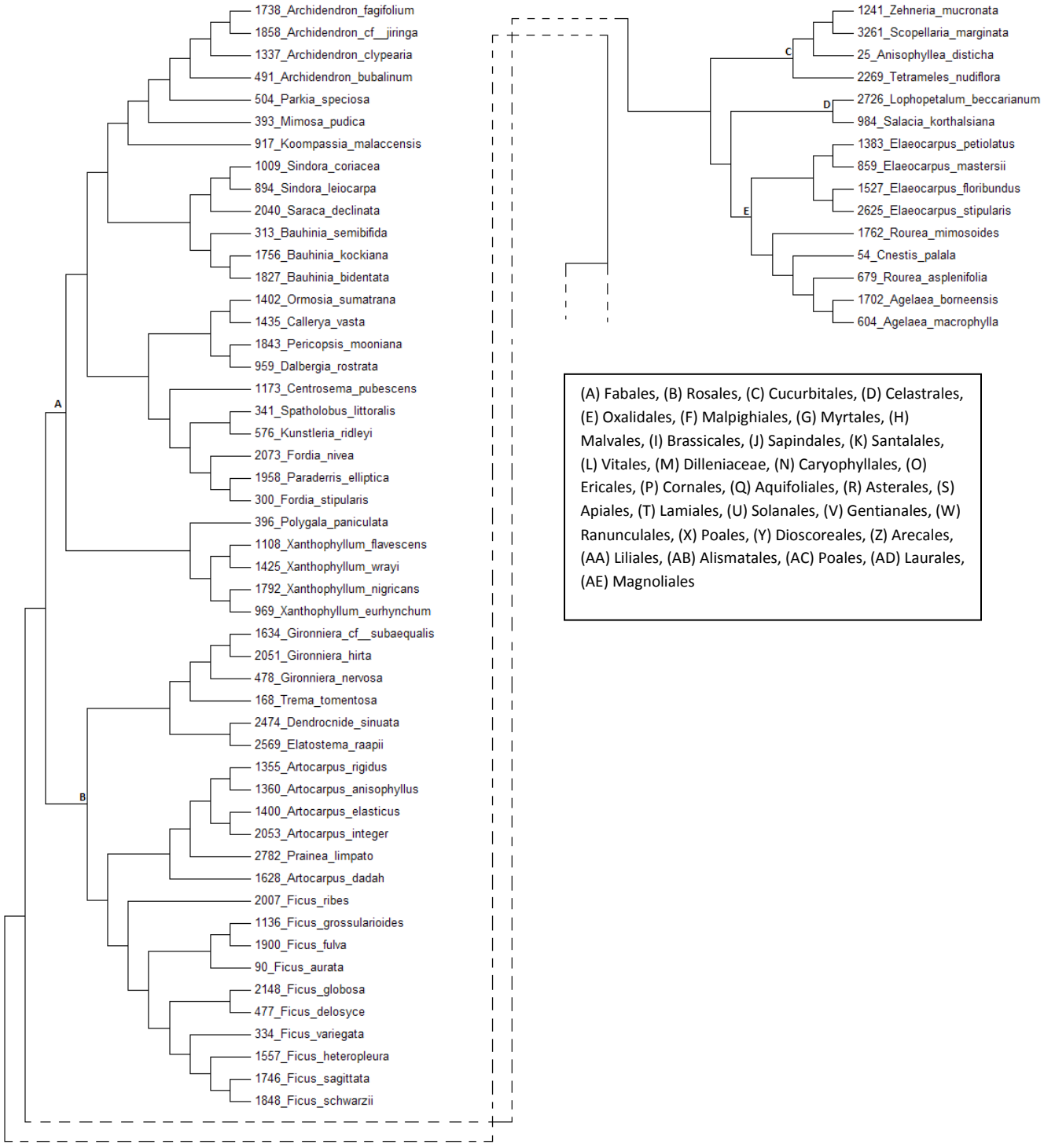
Appendix 4.9 (continued)



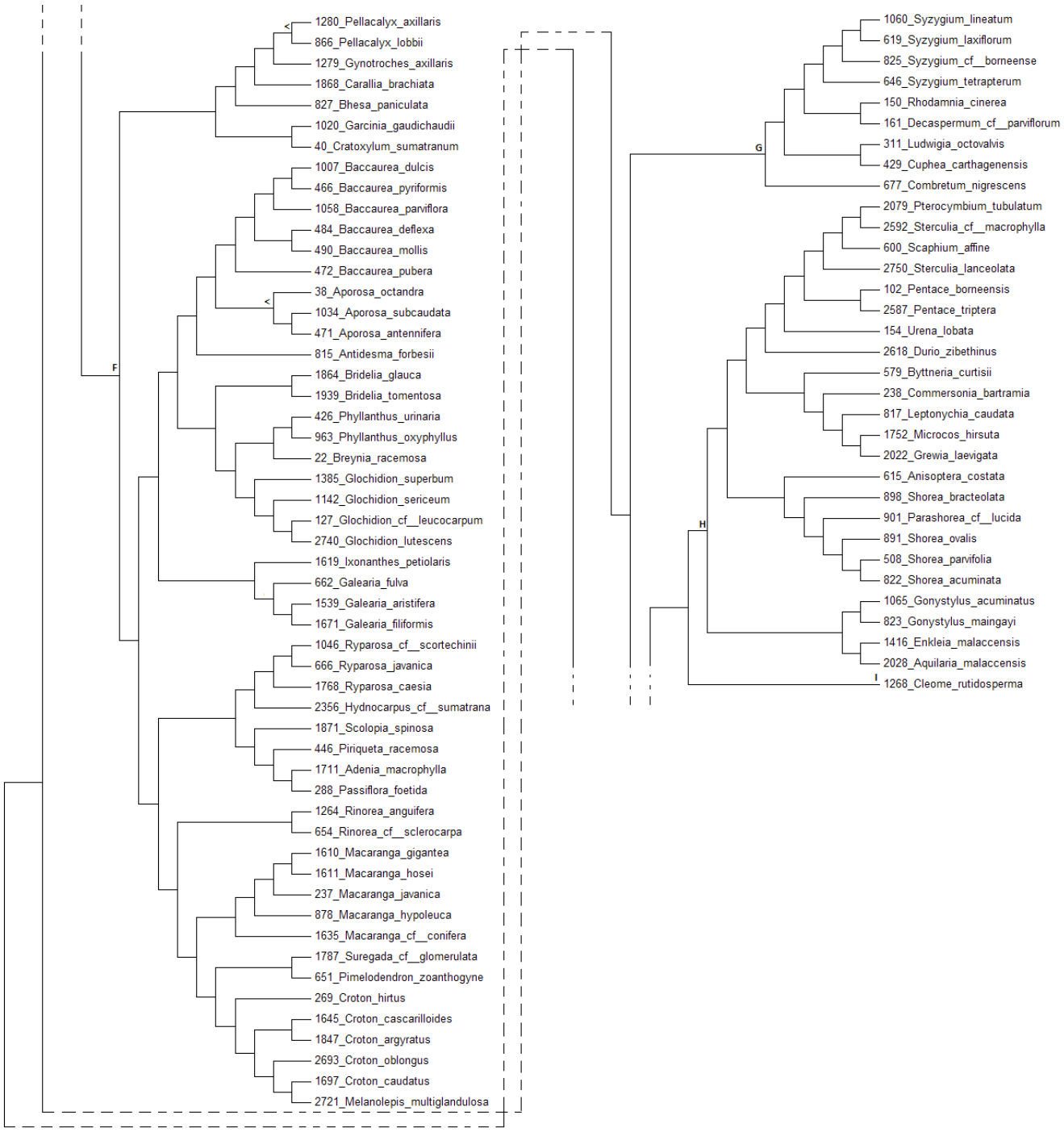
Appendix 4.9 (continued)



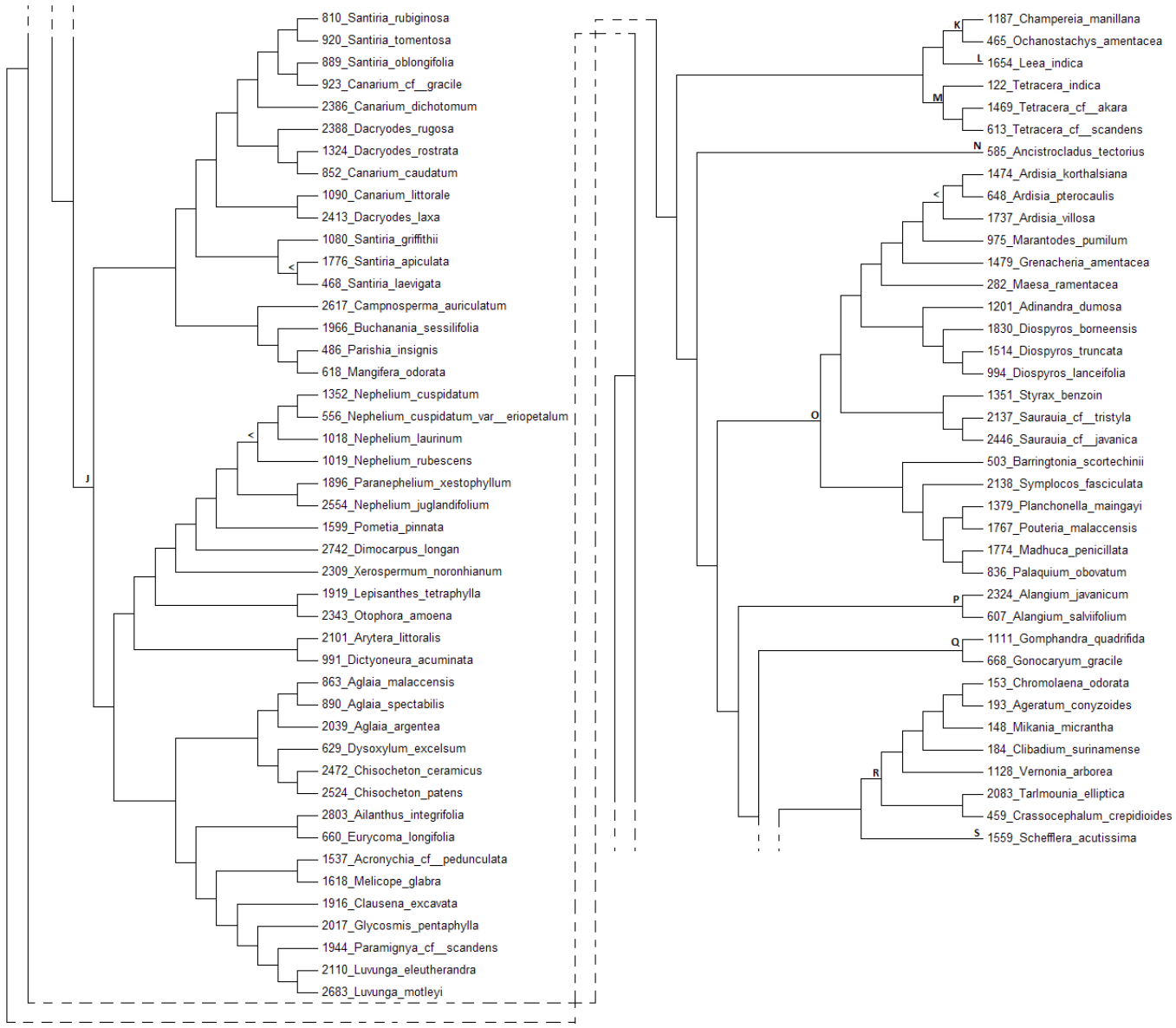
Appendix 4.10 Phylogenetic tree reconstructed using NJ method based on matK+rbcL barcodes



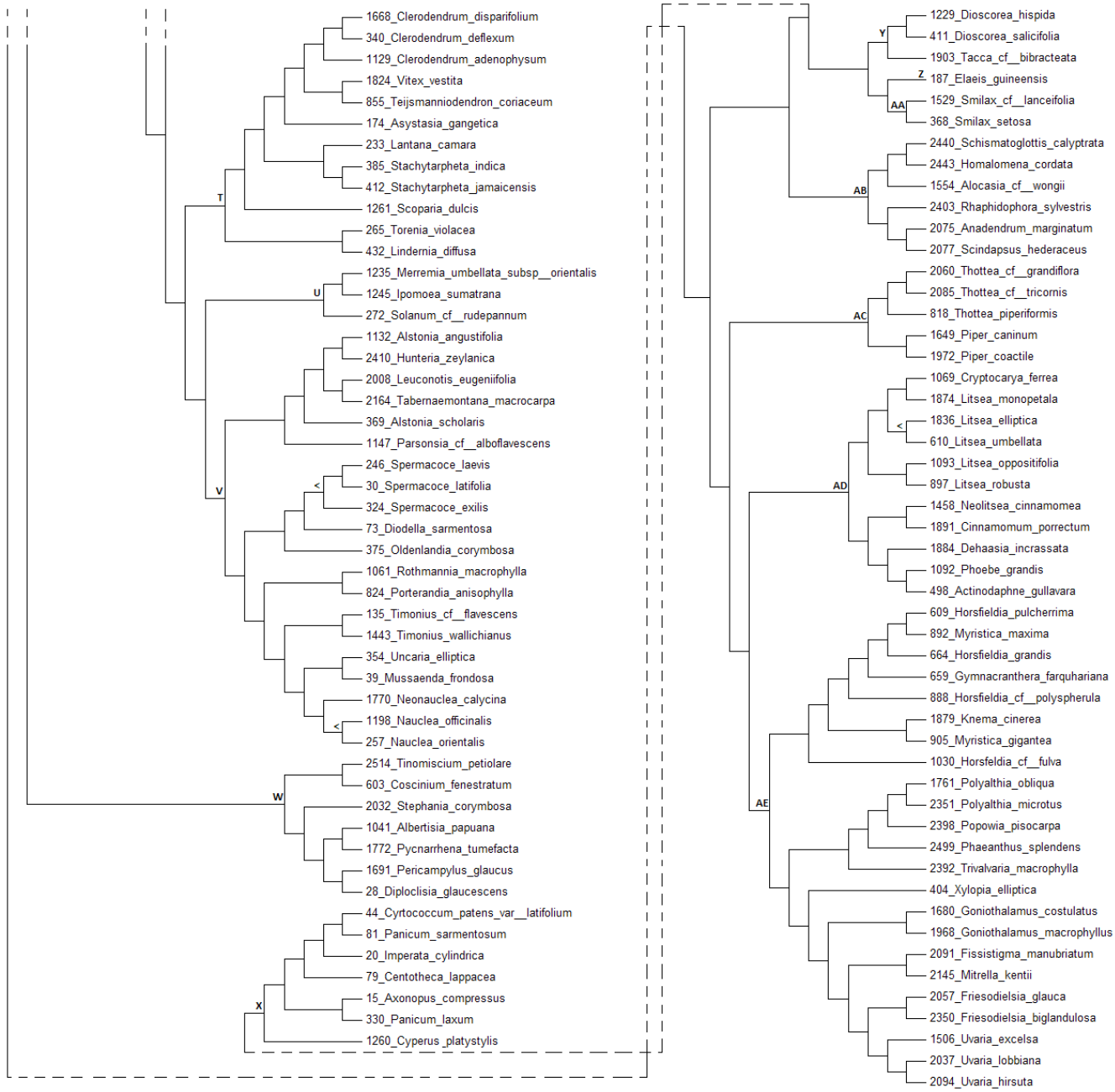
Appendix 4.10 (continued)



Appendix 4.10 (continued)



Appendix 4.10 (continued)



Declaration of originality and certificate of authorship

I, Fitri Yola Amandita, hereby declare that I am the sole author of this dissertation entitled "DNA barcoding of flowering plants in Jambi, Indonesia". All references and data sources that were used in the dissertation have been appropriately acknowledged. I furthermore declare that this work has not been submitted elsewhere in any form as part of another dissertation procedure.

Göttingen, December 2015

Curriculum vitae

Personal information

Surname, first name : Amandita, Fitri Yola
Place and date of birth : Jakarta, 13rd July 1983
Citizenship : Indonesia

Schooling

1990 – 1995 Elementary school in Jakarta
1995 – 1998 Junior high school in Jakarta
1998 – 2001 Senior high school in Jakarta

University education

2001 – 2005 Bachelor study in Faculty of Forestry at Bogor Agricultural University, Indonesia

Thesis: *The financial analysis of small-scale logging enterprises (a case study in South Borneo, Indonesia)*

2010 – 2012 Master study in Faculty of Forest Sciences and Forest Ecology at Georg-August University of Göttingen, Germany

Thesis: *Genetic variation pattern of Shorea johorensis in Borneo, Indonesia using microsatellite markers*

2010 – 2015 Doctoral study in Faculty of Forest Sciences and Forest Ecology at Georg-August University of Göttingen, Germany

Thesis: *DNA barcoding of flowering plants in Jambi, Indonesia*

Internships and work experiences

2003	Work experiences at the Meru Betiri National Park, Indonesia (1 month)
2004	Work experiences at the Sumpol Timber, Co., Indonesia (2 months)
2005	Internship at the Rawa Aopa Watumohai National Park, Indonesia (2 months)
2007 onwards	Work at the Ministry of Forestry of the Republic of Indonesia
2011	Field project “Forest Management Planning in Wanagama Educational Forest, Indonesia (1 month)