Tapping Zingiberaceae - Wilderness to Mining Plastome

A thesis *submitted*

in partial fulfillment of the requirement for the degree

DOCTOR OF PHILOSOPHY

By

TUSHAR

Department of Biotechnology

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

MAY 2014

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biotechnology
Guwahati – 781039

MAY, 2014

CERTIFICATE

It is certified that the work described in this thesis, entitled "**Tapping Zingiberaceae - Wilderness to Mining Plastome**", done by Mr. Tushar for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under our supervision in the Department of Biotechnology, Indian Institute of Technology Guwahati, India, and this work has not been submitted elsewhere for a degree.

Dr. Latha Rangan
Professor
(Thesis Supervisor)

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Department of Biotechnology
Guwahati – 781039

STATEMENT

I do hereby declare that the matter embodied in this thesis is the result of investigations carried out by me in the Department of Biotechnology, Indian Institute of Technology Guwahati, India, under the guidance of Dr. Latha Rangan.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators.

MAY, 2014 TUSHAR

DEDICATED TO MUMMA & PAPA

ACKNOWLEDGEMENT

On the completion of this thesis, it occurred to me that though it is my name which appears on the cover but without the guidance, support, love, advice, encouragement and help of many people there would have been no thesis.

My deepest gratitude to my PhD mentor, Dr. Latha Ragan who provided scientific guidance and always encouraged me to improve and at the same time provided me adequate freedom. Her patience and support always caught and carried me forward when I stumbled.

I express my sincere gratitude to the chairman of doctoral committee, Dr. Biplab Bose for providing me valuable suggestions which definitely helped in concurring my thesis objectives. I am further thankful to his support and care throughout the whole tenure of my doctoral studies but especially towards the end. I express my sincere appreciation to all other DC member, Dr. Rajaram Swamiatahan and Dr. Pinakeshwar Mahanta for their constructive criticism and advices.

I extend my gratitude to Dr. Jeffrey Bennetzen for providing me his expert guidance as well as space and resources. I thank Dr. Srinivas Chaluvadi for helping me out in the lab and offering to help for any problem be it academic or personal. I also than Dr. Justin Vaughn from whom I learnt a lot about doing biology on a computer.

I am thankful to all HODs who helped me in a direct way by approving my request for official leaves do as I may carry out my field activity or indirectly by providing a smooth running of the department. I would not like to miss the opportunity for thanking Mr. Nurul Islam, Mr. Niranjan Barah, Mrs. Prarthana Swargari and Mr. Chandan Kumar Nath for providing me help and support even at odd hours.

I am grateful to IIT Guwahati, whole as an institution including all the deans and staff for making my stay at the campus pleasurable and memorable. I am also thankful to University of Georgia for accepting me as a visiting researcher and making my stay easy and hassle free. I extend my thanks to office of International Education at UGA for hosting me at several events.

I am thankful to DIT and Fulbright-Nehru program for providing me adequate fellowship and USIEF for taking care of every minute details of travel and accommodation.

I am really thankful to my lab members and other colleagues with whom I shared lab space with. I am thankful for providing an atmosphere of cooperation and being friendly all the time. My special thanks to Dr. Vigya Kesari, Dr. Archana Das, Dr. Aadimoolam Ramesh, Dr. Supriyo Basak, Dr. Sudipta ghosh, Sharath Ramachandran, Rimjhim Chaudhary, Akhil Jalutharia, Dr. Ganesh Thapa, Debatosh Das and many more to name.

My long stay at IITG would have seemed longer if I didn't had company of friends. I express thanks to all of them with the deepest core of my heart for their love and support

specially during the period of sickness and economic turmoil. There are many name worthy to be called here but I would specially like to mention Ziyaudin Khan 'Zia bhai', Amit Jaiswal 'Babu', Sahil Batra 'Babajee', Shashank Garg'Bulla', Ritesh Bankar'Paglu'. Some friends left the campus early while some stayed longer, and I would like to specially mention which destiny kept close to me for the later part of my campus life, Dr. Sudipta Ghosh'Sud' and Rajat Pandey'Chulbul'.

I am indebted for my life for some persons who has been with me in this journey for a very long-long time. Anil Kumar'Dealer', Durga Prasad'DP', Rajeev Mishra'Baba', Sachindra'Bihari', Varun Gupta'Dada', Surrendra Chaudhary'Chacha' and many more to name. I am thankful to them and their family for all the wishes and care.

At last I express my deepest thanks to my mother and father for being a source of inspiration, for being a bottom less tank of love, affection and care, for trusting my decisions, for giving me freedom by protecting me from all the worries.

May 2014 Tushar

C O N T E N T S

ABSTRACT

Zingiberaceae is an under-utilized plant family because of difficulty in identification and due to overlapping morphology. Further lack in documentation of medicinal values makes the members of its family as obscured target. The present work describes the ethno-medicinal values of Zingiberaceae and further utilizes the plastid genome for resolving phylogeny.

Identification and documentation of ethno-medicinal values plays a pivotal role for screening and development of new bioactive molecules. For documenting the ethno-medicinal the information was collected from ethnic community during the field trip made for collection of Zingiberaceae species. A total of 51 plants belonging to nine genera was collected out of which 34 plants were found to possess ethno-medicinal usage. Rhizome was found to be most frequently used plant part and poultice was the most preferred mode of preparation. The family was found to treat a variety of disease with highest treatment associated for gastro-intestinal problems. The information gathered was assembled on a web portal, which was built using HTML 5.0 and the navigation was indexed using java script and CSS style was incorporated in these pages

Neighbor joining, maximum likelihood and maximum parsimony analyses of plastid sequence (*accD, matK, rpoB, rpoC1, rbcL, atpF-atpH, psbK-psbI* and *psbA-trnH*) data were used to reconstruct the phylogeny of Zingiberaceae. The phylogeny study was conducted by dividing the family into three tribes Alpinieae, Hydechieae and Zingibereae. *Amomum, Alpinia, Curcuma* and *Zingiber* was found to be polyphyletic. The phylogeny based on character analyses was found to be marginally more distinguishing. The best combination of loci was observed to be *rpoB+matK* and *matK+atpF-atpH*. *Hedychium* was found to share the recent common ancestor with *Zingiber*.

The whole genome shotgun sequencing of *Zingiber officinale*, *Amomum cardamomum, Alpinia zerumbet, Hedychium coronarium,* and *Curcuma longa* was carried out on Illumina platform. The assembly was performed on *edena* and *velvet* platform and the scaffold was constructed by aligning with *Typha latifolia* plastome. The final plastome with 162,598 bp was produced and showed perfect colinearity with *Musa* plastome. Other Zingiberaceae was assembled by mapping total shotgun reads on to *Z. officinale*. An expansion of the IR region at its border with SSC region was observed. Nine pair of new markers were designed by using the five sequenced plastome, which included on SSR marker, six markers for pan Zingiberaceae and two markers for identifying variation among *Zingiber* species.

The comparison of the phylogeny drawn in different section of the thesis and with earlier published works was performed. *Hedychium* and *Zingiber* were established to be close relative. It was seen that species from a geographic region clustered together.

The present thesis demonstrates the ethno-medicinal importance and phylogenetic relationship of Zingiberaceae. The medicinal important plants would be a source of novel bioactive molecules and the phylogeny study will aid in better classification of the family.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The chapter introduces the family Zingiberaceae and discusses the problem associated with family and thus marking the objectives for the thesis work.

Introduction

The animal kingdom has been long dependent on plant kingdom for several of its needs. *Homo sapiens* has been no exception since the time of gatherer to sedentism. Plants has catered to the demand of food, medicines and fuel for thousands of year. One such group of plant species, which has provided almost everything a man can demand from a plant, has been Zingiberaceae. This plant family provides many useful products for food, spices, medicines, dyes, perfume and aesthetics (Jantan et al., 2003) and is distributed widely throughout the tropics (with one genus *Renealmia* found in the neotropics), particularly in Southeast Asia.

Zingiberaceae is the largest family from the order Zingiberales (Kress et al., 2002) and is placed on eumonocots branch of Magnoliophyta. Zingiberaceae shares the order Zingiberales with other prominent family such as Cannaceae, Musaceae and Heliconiaceae. Zingiberaceae has 52 described genera and around 1,300 diverse species which are mainly concentrated in South and Southeast Asia (Wu and Larson, 2000). Though the centre of diversity for the family is Southeast Asia, the basal clades of the family is an African genus *Siphonochilus* and a Bornean genus *Tamijia* (Kress et al., 2002) which is a strong indication that the family have its origin from Gondwana tectonic plate. The classical taxonomy based on both vegetative and floral characteristics divides Zingiberaceae into four tribes Hedychieae, Alpinieae, Zingibereae and Globbeae (Petersen, 1889; Schumann, 1904; Holttum, 1950; Burtt and Smith, 1972; Larsen et al., 1998). A number of morphological features have been used in classifying, defining and distinguishing the four tribes, but the features are either not unique to any one tribe or are not common for all the taxa within any tribe (Kress et al., 2002). The narrow window of flowering further makes the identification more and more complex task. Many plants of the family have a very low survival rate in growing in non-native environment, even in controlled environment of greenhouse. Moreover a large number of genus and species has been described recently and some of the taxa has been shifted and rearranged. Hence either a new set of morphological traits for characterization or an alternative method for characterization of the family is felt. While assigning the new sets of morphological traits to the rearranged taxa would require some serious breakthrough in the field of classical taxonomy, the alternative methods strengthen its claim.

In India, about 22 genera and 178 species have been reported from the North Eastern and peninsular region (Jain and Prakash, 1995), whereas North East India (NEI) alone harbours 19 genera and close to about 88 diverse species (Prakash and Mehrotra, 1995). The number of species and genus recorded and reported from the NEI is on rise with numerous reports of species identified recently (Singh and Kumar, 2011; Thomas et al., 2012; Kumar et al., 2013; Thongam et al., 2013; Thomas et al., 2014). As previously stated, Zingiberaceae has its origin from Godwana tectonic plate, the phylogeny of species from India can provide deeper insights about evolution of the family and the order Zingiberales. The NEI region belongs to the Indo-Burmese hotspot, which is the largest of the four hotspots grouped under south-east Asia, and ironically is the least studied hotspots (Morgan et al., 2011). The region is unique because it has been a part of Gondwana plate but the humid and damp climate favours growth and evolution of herbs. This region is also unique with regards to the presence of more than 130 major native tribes (Tanti et al, 2009) which during the course of time has an intensively developed ethno-medicinal practice. These ethno-medicinal usage serves as target for the new bioactive molecule discovery (Oubre et al., 1997; Katiyar et al., 2012; Omonike et al., 2013; Sucher, 2013). As most of the species belonging to the Zingiberaceae becomes dormant during the winter season, owing to low rainfall, the rhizomes serves as a reservoir for its resurrection on the onset of favourable season. The plant family hence has evolved to manufacture and accumulate a myriad number of molecules which has bioactivity to prevent infestation by microorganisms and insects. Hence it has been observed that a large number of plant species belonging to the family Zingiberaceae has been recorded to be used by the native population belonging to this region (Tushar et al., 2010). In recent times it has been observed that urbanization and migration of forest dweller towards cities has led to severe degradation of indigenous knowledge and hence a dire need for documentation of ethno-medicinal usage of flora and fauna is felt before the valuable information is lost in time. Further as stated earlier about the bottlenecks in identification of the plant, assigning the ethno-medicinal property to the plant species becomes difficult.

Understanding the phylogeny of plants has major practical implications which includes providing information which are at most useful in breeding programmes and germplasm conservation. The correct classification leads to accurate identification of species and also helps in clubbing the variant intraspecific individuals into one. In spite of tremendous utility of the family Zingiberaceae and failure of development a robust classification approach, it is seldom subjected to genetic and molecular biology research (Gao et al., 2006; Xia et al., 2009). Among the various method of phylogeny reconstruction for plants, the method involving DNA sequences from the chloroplast genome or plastome has emerged as the most favoured choice (Wurdack et al., 2005; Chase et al., 2007; Chang and Graham, 2011). The inheritance of plastid is generally uniparental (Birky, 2001) and is homozygous with just one kind of copy and hence provides a strong signal of population and phylogenetic history (Petit and Vendramin, 2007). Further, the plastid genome does not plague with the problem of lateral gene transfer and neither from the high rate heterogeneity as observed in mitochondrial genome (Palmer et al., 2000; Bergthorsson et al., 2004). In some of the studies the assessment of phylogeny relationship, by coding regions, between the plants at lower taxa (species and sub species level) was observed to be confounding and it led to incorporating the noncoding loci in such studies (Kim et al., 1992; Soltis et al., 1993; Chiang and Schaal, 2000; Hu et al., 2000; Hall et al., 2002; Dane and Lang 2004). The previous work on the phylogeny of the family Zingiberaceae either involved use of one or maximum two loci (Kress at al., 2002) or their dataset lacked plants from NEI for phylogenetic evaluation (Vinitha et al., 2014).

With the development of computational resources which was better in handling a larger amount of data and required lesser time for analyses. This development led to use of several loci from plastid genome for carrying out the assessment of phylogenetic relationships (Savolainen et al., 2000; Dong et al., 2012). The use of number of combination of loci witnessed an exponential rise with the advent of a new generation of sequencing techniques. The use of complete plastid genome for estimating the evolutionary distance and reconstruction of the phylogeny has emerged as a main stream method for phylogeny assessment (Shaw et al., 2007; Njuguna et al., 2013; Walker at al., 2014).

Zingiberaceae family, as important as it is economically, this thesis aims at documenting the ethno-medicinal usage of plants belonging to this family, construction of a database for dissemination of the knowledge, reconstruction of phylogeny using eight loci of plastid genome. Further to resolve some of the confounding relationships the plastid genome of five species was sequenced representing five different genera of Zingiberaceae which are distributed across the family. The plastid genome information was used for phylogenetic study and were further used for development of new marker system. Since a number of different methods and dataset has been used in this thesis a comparative study was undertaken as an effort for an improved comprehension of the family.

1.1 Objectives

The specific objectives were identified based on the back ground study to mark out the studies whose outcome would help in efficient utilization of the members of Zingiberaceae, a rich bio resource. The identified specific objectives are as follows

i) Collection and documentation of plants and related information

ii) Mining of plastid genome for phylogenetic relationship assessment

iii) Whole plastome sequencing and new marker development

iv) Comparative study of phylogeny constructed

CHAPTER 2

LITERATURE REVIEW

The chapter describes the importance of the plant family and explains the identifying characters which are routinely employed in classification and the associated perils. The alternative methods are discussed and assessed based on the available literature.

Literature review

Plants are considered to be a mystic bag of a magician which holds all the objects that is needed or desired. One such plant family, which is an aromatic perennial herbs and creeps horizontal tuberous rhizomes, is Zingiberaceae. The importance of the plants from this family has been felt since ancient time with mentions in *Charak Samhita* and *Sushrutha Samhita* ancient *ayurvedic* text, in *Koran* (76: 15–17) and in other numerous text (Ravindran et al., 2007). It was the members of this family which encouraged the traders from Egypt to cross the Red Sea and Indian Ocean to reach the coast of Western Ghats of India. The trader apart from carrying spices from India, also carried numerous fascinating talks of the land and which ultimately lead to discovery of India to the western world by Vasco da Gama in 1498. The members belonging to this family has been source of primary healthcare for India and China with even calling ginger as "*Maha-audhisi*" in Sanskrit, meaning the great cure. In modern times this plant family has been an important source of a number of bioactive molecule targeting a myriad diverse medical conditions (Sharma et al., 2001; Singh et al., 2010; Suthisut et al., 2011; Chuan-Li et al., 2013). Zingiberaceae is the biggest family in its order and with a number of new species added each year, the ability of the family to act as a true mystic bag of a magician is not an exaggerated statement. The proper utilization of the family and developing it as an asset to mankind would require a better understanding of the family and documentation of several of its features. The present study has been undertaken towards an improved understanding of the members of family from the northeastern part of India which is a biodiversity hot spot.

2.1 Zingiberaceae

Zingiberaceae is the largest family belonging to the order Zingiberales with 52 described genera and around 1300 species with its centre of diversity in tropical Asia (Wu and Larson 2000). Zingiberaceae is a perennial aromatic herb with short stems which are often poorly developed. The identifying characters of Zingiberaceae has been very accurately described by Larsen et al. (1998). The morphology descriptions consisted of detailed mention of all the plant parts. It was pointed out that the stems are supported by pseudostems formed by the sheath of leafs. The height of the plants in Zingiberaceae varies considerably and some of the species reaches 8 meters high. The shoot is always unbranched. The leaves are distichous and appear tufted sometimes in lower part of shoot without blade. The shape of the leave is elliptic but are known to occur as sometimes linear or broadly elliptic and may or may not have trichomes. The young leaves are rolled up from one side with a prominent midrib and lateral veins in pinnate-parallel arrangement with petioles present or absent. The position of inflorescence are either terminal on the leafy shoot or terminal on a short separate leafless shoot arising directly from the rhizome. Ridley (1899) observed that plants growing in the dense part of the jungle have their flowers close to the ground while those growing in more open space have flower on the terminal of leafy stem. Inflorescence could be congested or lax are cylindrical or fusiform and are rarely globose. The bracteolate could be cincinni axillary to the bracts of the inflorescence and then representing a thyrse. Rarely the inflorescence could be raceme or a spike. Flowers are epigynous, bisexual and are zygomorphic to facilitate pollination by insect (Giurfa et al., 1999). Calyx are tubular and are 3-lobed or 3-dentate which are sometimes split down first lobe side. Corolla are tubular at base with three lobes and varies in size and shape with the median posterior corolla is larger than other and cucullate. The floral characteristic that binds all the Zingiberaceae is the fact that median stamen of the inner ring is fertile while other two are sterile with the median of the outer ring missing (Weberling, 1992). Functional stamen could have either short or long filament. The other two stamens connate to form the petaloid labellum. The two lateral staminodes of the outer whorl petaloid or inconspicuous, flanking the stamen or adenate to the labellum, which is also an autapomorphical character of the family (Kress, 1990). The morphology of labellum was debated a lot (Holttum 1950) but the study of Rao et al. (1954) conclusively demonstrated that the labellum is formed from the two united staminodes. The median anterior member of outer whorl is invariably reduced. Anthers are dithecal, tetrasporangiate with thecae opening intorsely by longitudinal slits. The gynoecium is 3-carpellate and ovary is invariably trilocular in its first stages but fully developed ovary could be trilocular with axile placentation or unilocular with parietal or basal placentation. The style which are terminal and always placed in a furrow of the filament and held between the thecae is a character which is very unique to Zingiberaceae and one of its sister species, Costaceae. Stigma are papilate funnel shaped and are often lined with cilia. The nectaries found are two in number and are epigynous. Ovules are anatropous, bitegmic and crassinucellar. Fruits born to zingiberaceae are dry with sometimes fleshy capsules with syncarpic fruit is some of the genera. Seeds are found in few to many numbers and varies in shape from Ovoid to ellipsoid to subglobose and are sometimes angular with an operculum next to the radicle. Embryo are straight, linear and central in the endosperm. The perisperm is abundant and starchy with large compound starch grains. The germination of the seeds are hypogeal. The representative floral diagram and formula for the Zingiberaceae is depicted in Figure 2.1.

The underground stem of the plants which serves as storage unit and helps in resurrection on the onset of favourable conditions is sympodial in growth with a variable degree of branching across the family. The dispersion of the plant seeds hs been suggested by ants, mice, squirrels, rain streams, primates and elephants (Ridley, 1899; Maas, 1977; Nishida and Uehra, 1983). Many *Alpinia*, *Hedychium* and *Amomum* plants grows in bunch with several stems arising from the underground rhizome as it grows.

The phytochemistry of rhizomes has revealed that these are a large reservoir of many bioactive molecules. As the primary function of rhizome is as a storage unit and it seems logical, in order to protect itself from bacterial and fungal contamination during the dormant stage these bioactive molecules are synthesized and stored in rhizome. This has led to incorporation of the Zingiberaceae in traditional medicine by the ancient civilization.

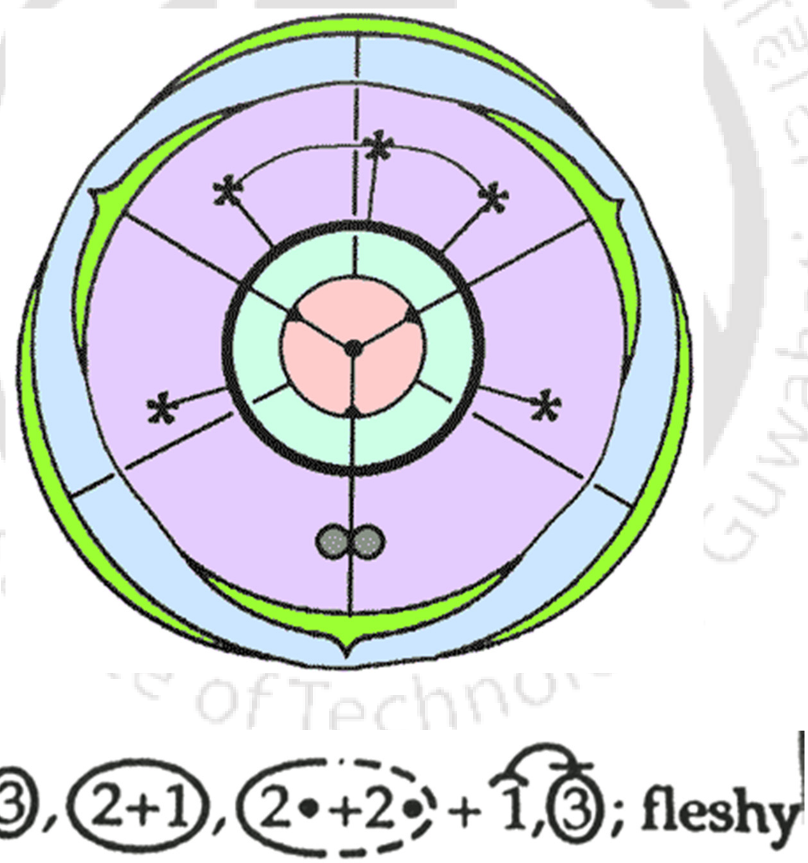


Figure 2.1. Floral diagram and formula of Zingiberaceae (Source: http://www.thewildclass-room.com).

2.2 Traditional knowledge

For several years the bio-resources have been efficiently utilized and maintained by the aboriginal peoples of their region. These hunters and gatherers utilized the most of the bio-resources for an era with just an exception of deep sea and open ocean diversity (Berkes et al., 1993). The practices adopted by these societies just not only derived benefits from the nature but is believed to effectively manage the resources. The native people satiated all of their requirements from these bio-resource and hence with the time developed the healthcare and the whole of the bio-resources acted as their pharmacy. With the arrival of scientific period the importance of plants in providing cure were forgotten by the majority of people until Reverend Edward Stone in 1763 submitted a communication to Royal Society of London and described the use of extracts of willow bark for treatment of fever. Though he demonstrated the use of a plant extract for curing an ailment for the first time, the race for discovery and development of cure from plants did not jettisoned till 1829. French pharmacist Henri Leroux first isolated salicin in 1829 from a willow plant and further demonstrated its antipyretic property. Since then a large number of plant derived molecule has been derived and tested for his efficacy. By an estimate approximately quarter of bestselling drugs worldwide were natural products or derivative in 2001 and 2002 (Butler, 2004). Several of these new drugs has been discovered on ethno-medicinal leads (Heinrich and Teoh, 2004; Pirttila et al., 2004). Ethno-medicinal methods has been established as the first and foremost step in development of new drug molecules (Balunas and Kinghorn, 2005).

Documenting of the Traditional Knowledge have emerged as a central issues for planning and management of bio-resources (Rao, 2006; Signorini et al., 2009). Due to a number of anthropogenic activities, rapid urbanization and industrialization there is a potent threat over rich biodiversity of the world (Tushar et al., 2010). The conservation of the rich biodiversity and ethnobotanical information requires strong steps. Various ethnobotanical surveys and studies are a prerequisite in attaining these goals.

2.3 Traditional knowledge database

Documentation of traditional knowledge is important for conservation of the knowledge as well as bio resource. The documentation puts on check on erosion of knowledge which has been built by thousands of years of practices as well as prevents bio-piracy attempts. The documentation also helps in safeguarding the interest of the aboriginal people (Udgaonkar, 2002). Since the last two decades has seen a number of attempt to gain patents for ethno-medicinal usage of many plants, a large number of such databases has now appeared on the screen.

TKDL (2001) was set up by Indian government by a collaboration between the Council of Scientific and Industrial Research (CSIR) and the Department of Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy (Dept. of AYUSH), Ministry of Health & Family Welfare, Government of India with the objective to protect the traditional knowledge. The database at present have more than quarter of a million medicinal formulations which has been collected by transcribing several books into five languages (English, German, French, Spanish and Japanese) which were earlier available only in regional language. Thomas et al., (2001) developed a database model for integrating and facilitating collaborative ethno-medicinal research which was based on object oriented database technology. Subsequently many database which focuses on various types of traditional knowledge has been developed (Iwu et al., 2002; De Natale et al., 2009; Skoczen and Bussmann, 2006; Hong-lin et al., 2010; Kumar et al., 2010; Mary et al., 2012).

Zingiberaceae has three dedicated databases, ‘Zingiberaceae resource centre’ (http://elmer .rbge.org.uk/ZRC/Acknowledgements.php), ‘The gingers of world’ (http://zingiberaceae.e-monoc ot.org/) and ‘Gingers of India’ (http://www.gingersofindia.com/main.htm). The database Zingiberaceae resource centre has a comprehensible list of almost all the genera and species with the link or mention of the original publication used to describe the species. Some of the other information such as the classification and the list of synonyms are also contained. The database entry lacks any digital image and since the database does not mention the morphology, rather links to a publication which may or may not be an open access document. The gingers of world also have an exhaustive list of species and do considers incorporating the digital image, description and list of literature for that particular species. The database is under construction and hence many of the data are still missing from resource. Gingers of India hosts a collection of Indian Zingiberaceae species. The morphological description of each genus is provided but lacks the same for the species. The collection of digital images are good. In all these three databases the ethno-botanical information for zingiberaceae is missing and none of them have any information about the phylogeny of the family. As the identification of members of this complex family has long been considered as a problem, accumulating the phylogeny would aid in drafting a procedure for easy identification.

2.4 Zingiberaceae phylogeny

As well defined and characterized are the Zingiberales (Tomlinson, 1962; Cronquist, 1981) the same amount of complexity is observed in the characterization of the sub taxa of Zingiberaceae. The Zingiberaceae was first described by Lindley in 1835. The short lived flowers of members of this family poses a major obstacle in classification (Kress, 1990). There has been proposed changes in the phylogeny of the plant, the Costoideae representing the genus *Costus* was excluded from the Zingiberaceae and established as a family (Nakai 1941; Tomlinson, 1962). The basis of this proposition was non-aromatic nature of its vegetative body and spiral phyllotaxy and presence of anther appendages in Costaceae. The classical taxonomy based on floral and vegetative morphology was successful in establishing the Zingiberaceae as a monophyletic group and more.

2.4.1 Classical taxonomy of Zingiberaceae: Basis and shortcomings

The family was classified based on floral and vegetative morphology (Petersen, 1889; Schumann, 1904; Burtt and Smith, 1972; Larsen et al, 1998) and was divided into four tribes Hedychieae, Alpinieae, Zingibereae and Globbeae. Before Burt and Smith there were a number of recognized tribe but they established only four tribes as stated above. In their classification the Largest tribe was Alpinieae with 25 genera followed by Hedychieae (22 genera), Globbeae (4 genera) and Zingibereae with only one genus (representative classification shown in Figure 2.2). This study distinguished the four tribes based on a number of morphological traits but the distinguishing or defining traits lacked specificity to a tribe or were not common to all the genera within a tribe. The characters chosen were plane of distichy of leaves, lateral staminodes, labellum, stamen, ovary, style, placentation and stigma. The examples of discordance between the defining character and genera placed under the tribes are: Distichy of leaves is parallel in the genera *Rhynchanthus* and *Pommereschea* while it is perpendicular for the Alpinieae to which these genera belongs to (Kress et al., 2002); Ovary is partially unilocular in *Caulokaempferia* belonging to the tribe Hedychieae which is a trilocular tribe.

Some of the studies has also raised question about restructuring some of the genera of the family based on morphology and anatomy (Burtt and Smith, 1986; Larsen, 1998; William et al., 2004; Skornickova & Sabu 2005; Sabu 2005). A quarter of the genera in the family are monospecific (Kress et al., 2002). This has led to a failure to come to a consensus of the total number of genera. As classical taxonomy has been inefficient in coming up of a consensual solution a need of alternative method of classification is felt.

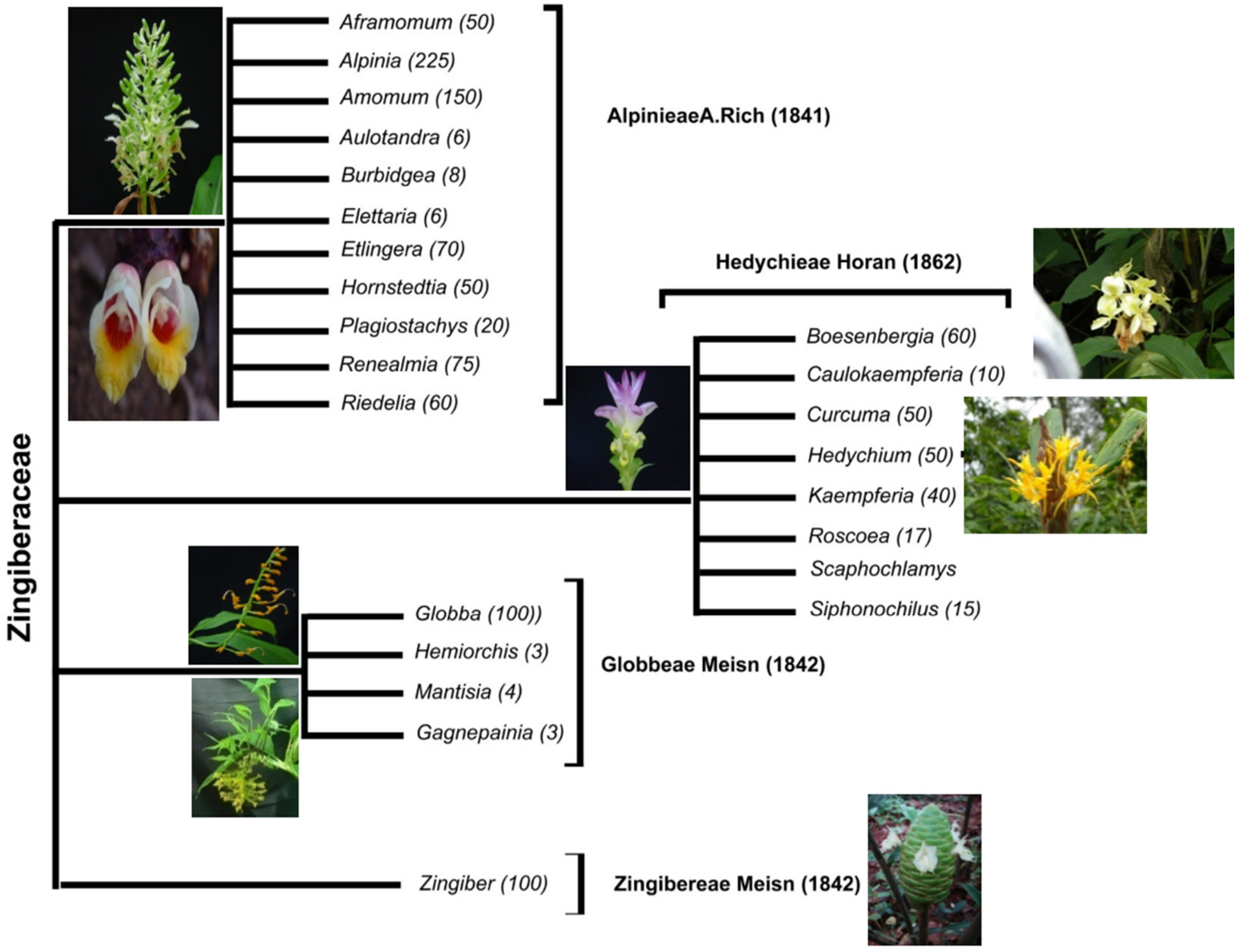


Figure 2.2. Classification of members of Zingiberaceae into four previously recognized tribes of the Zingiberaceae (adopted from Kress et al., 2002). The number next to the name of the genus corresponds to approximate number of species described under the taxa.

2.4.2 Alternative to classical taxonomy

Technological advancement and new discovery paved way into new methods for classification of the organisms. The prominent ones has been classification based on chemical profiles (Goodner et al., 2001; Correa and Goodacre, 2011), morphological data matrices (Tillyard, 1921; Gauthier et al., 1989; Lewis 2001; Lambkin et al., 2013), data matrices for polymorphism widespread in genome (Botstein, 1980; Bowcock, 1994; Welsh and McClelland 1990; Vos et al., 1995), and molecular sequences (Neyman, 1974; APG II, 2003; Bowe et al., 1999; Birgitta, et al., 2009; Wu et al., 2009; Perelman, 2011; Rinke et al., 2013). Among all the above mentioned method the one which has caught the maximum attention and fascination of researchers is classification based on the molecular sequences. The reason for its acceptance was that the evolution of molecular data did not suffer from homoplasy which is a huge problem in plants at higher taxonomic levels (Bremer, 1988). Further the sequence information is relatively easily retrieved and the phylogeny constructed is quite robust. As a plant cell have three compartments of genome hence it gives options to a researcher for selection. The mitochondrial locus has been the most preferred locus for the sequence based phylogenetic study in the animal kingdoms but in plants because of the slow nucleotide substitution rates and extensive levels of intramolecular recombination it was not suited to be utilized for such studies (Provan et al., 2001). Use of nuclear DNA has been used in several studies for elucidating the phylogeny of plants at various taxonomical rank. At higher taxonomic ranks the rRNA genes are used which are slow evolving (Hamby and Zimmer 1988, 1992; Steele et al., 1991; Soltis et al., 1997; Kuzoff et al., 1998; Soltis and Soltis 1998), whereas at lower taxonomic levels intergenic spacers are more commonly employed (Baldwin et al., 1995; Alvarez and Wendel 2003; Bailey et al., 2003). But the major problems with rRNA genes have been the presence of multiple copies per array and multiple arrays per genome, and the indefinite strength of concerted evolution (Small et al., 2004). This leaves chloroplast genome which has several advantages over other genomes to be used for phylogeny study in plants.

2.5 Chloroplast genome

Plastid is one of the most complex organelle in a plant cell and perform several important functions such as high energy molecule generation (ATP and NADPH) by trapping sunlight energy, liberation of oxygen from water molecule and synthesis of carbohydrate by fixing carbon dioxide.

The plastid exists in many form in plants such as chloroplast (the green plastid), chromopalst (the coloured plastid) and leucoplast (the colourless plastid). Plastid is an endosymbiont with its own genome which is inherited maternally. The first time the DNA presence in plastid was reported by Chiba 1951 in a moss Selaginella and from two flowering plants. By the end of 1963 the presence of chloroplast DNA was widely acknowledged (Sugiura, 2002). The plastome size varies from 120-220 kb (Schmitz-Linneweber et al., 2001; Ravi et al., 2008) and the coding genes are found in a close cluster (Sato et al., 1999). The plastid genome is a quadripartite molecule (Huang, 2013) with a pair of inverted repeats (IRs) flanked by a large single copy region (LSC) and a small single copy region (SSC). The IRs are exact inverted repeats, varies in size from 20-30 kb and divides chloroplast into two unequal parts, the LSC and SSC (Kolodner and Tewari 1979) (Figure 2.3). Some genomes have been documented with one copy of the IR missing (Palmer and Thompson, 1982; Raubeson and Jansen, 1992). Two populations of chloroplast has been observed differing in the orientation of the SSC region, resulting from the intramolecular recombination with in the IR (Tsumura et al., 2000). The junctions between IR and single-copy regions is highly unstable and hence the position of these junctions may vary among different and even closely-related species (Logacheva et al., 2009). In the order Zingiberales a major expansion of IR region-a (IRa) has been noted in the Musa lineage (Martin et al., 2013). The inheritance of chloroplast is maternal but some exception has been noticed in the nature (Mogensen, 1996; McCauley et al., 2007).

2.6 Chloroplast genome loci for phylogenetic analysis

The first account of use of plastid genome for phylogenetic analysis involved study on restriction site variation (Palmer and Zamir, 1982; Erickson et al., 1983). With development of user friendly

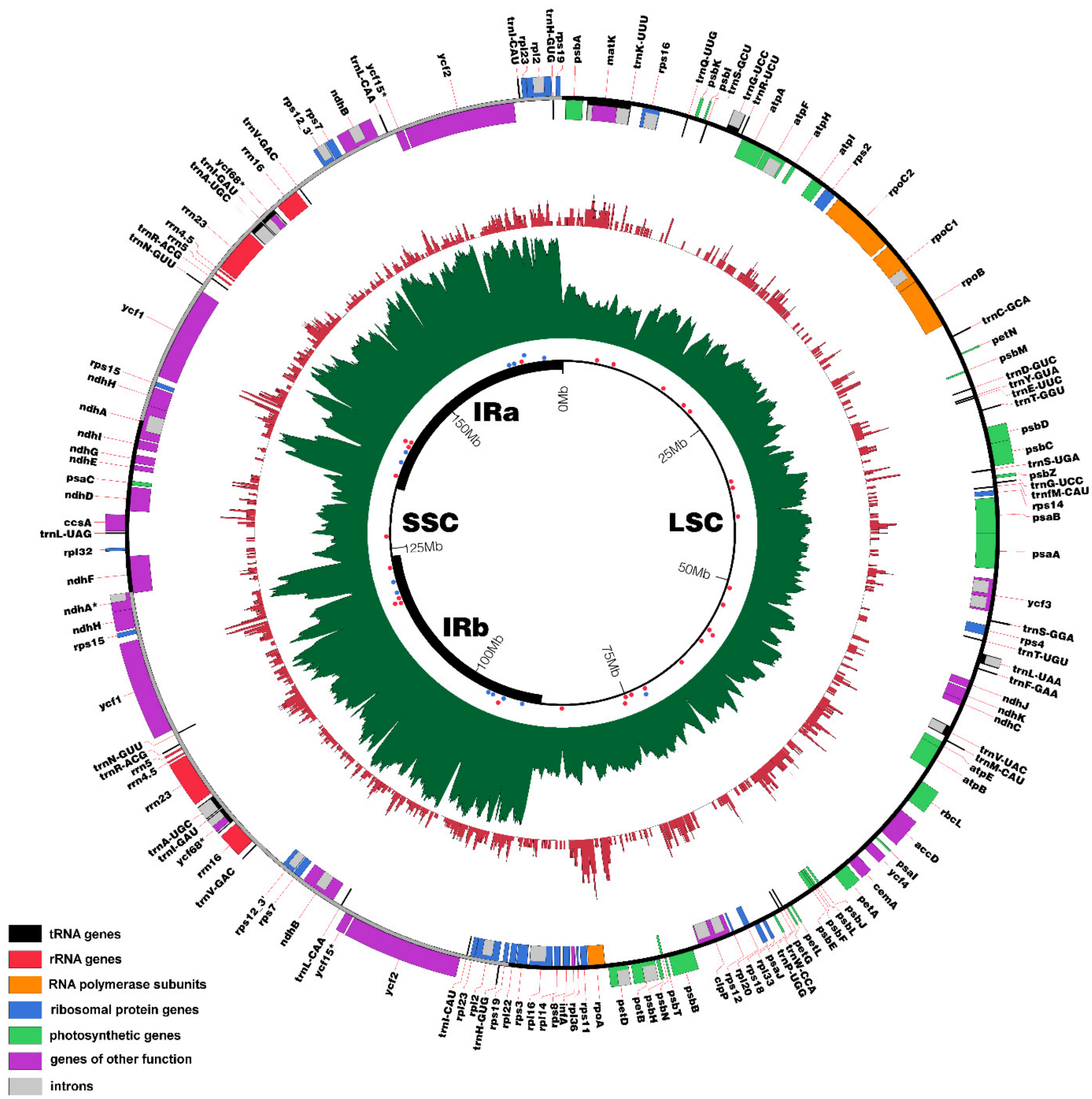


Figure 2.3. The chloroplast map of *Musa acuminata* with genes represented with boxes inside or outside the circle to indicate clockwise or counterclockwise transcription direction respectively. The inner most circle depicts the structure of chloroplast by the region (reproduced from Martin et al., 2013)

and automated sequencing techniques the use of plastid sequence for phylogeny study became prominent (Clegg and Zurawski, 1991). For the use of plastome for studying phylogeny, the use of sequence information has fair advantages over the restriction site variations as the amount of DNA required is less. The initial studies concentrated on use of the coding loci for the phylogeny and ribulose-bisphosphate carboxylase gene (*rbcL*) was the most preferred choice (Palmer et al., 1988; Clegg and Zurawski, 1991) and were focused on higher taxonomic levels. For sub family level the *rbcL* was found to fail at deducing the relationships (Kim et al., 1993; Soltis et al., 1991). This lead to exploration of other loci and to venture into the sea of plastid genome other than coding loci. A number of other loci were tried and tested with the aim to discover a loci which would be able to improve the phylogenetic resolution at lower taxonomic level. Some of the other loci which were utilized included *trnL*, *psbA-trnH, atpB-rbcL, matK, trnL-trnF, ndhF, ycf6-psbM* (Gielly and Taberlet, 1994; Kim et al., 1999; Chiang and Schaal, 2000; Hu et al., 2000; Hall et al., 2002; Dane and Lang 2004). Shaw et al., in 2005 conducted a study to detect any predictable rate heterogeneity among twenty one non-coding Chloroplast loci which has been identified as useful for phylogenetic studies at low levels. The study established that a rate heterogeneity exits among the non-coding regions and established three tier of non-coding DNA loci depending on polymorphic information content. As chloroplast has a very slow rate of evolution hence, even the single non-coding loci which were mostly spacer regions was found to be inefficient at resolving lower taxonomic level relation (Baker et al., 1999; Kellermann and Udovicic, 2008). For constructing a better resolved tree with higher support of clades, various combinations of sequences were employed for analyses (Davis et al., 2001; Li et al., 2008). The combinations of loci increased the phylogenetic tree results and this prompted researcher to explore the realms of using whole chloroplast genome for phylogeny studies.

2.7 Whole chloroplast genome for phylogenetic studies.

With the emergence of next-generation sequencing technology the ways for whole genome sequencing has been completely revised owing to the high-throughput nature of the technique and low per base cost of sequencing (Shendure, 2008). The first chloroplast, though sequenced using

Sanger sequencing in 1986, paved the path to sequence a large number of plastid genome with around 400 plastid sequenced by March 2014 and are available in the GenBank Organelle Genome Resources with majority of them sequenced after 2005 (Shinozaki et al., 1986). With the availability of the genome information in the public domain and development of new approaches for comparison and analyses of such large rich data set has led to addressing the difficult phylogenetic questions aimed at deep nodes in the angiosperm tree (Goremykin et al., 2003, 2004; Leebens-Mack et al., 2005). The use of whole plastid genome established the monocots, as a sister taxa to all other extant lineages in the angiosperm phylogeny (Goremykin et al., 2003, 2004) but by increasing the number of genomes in the study by Leebens-Mack (2005) established that this could be because of long branch attraction and monocots are probably derived from angiosperm. The study clearly demonstrated the importance of taxa sampling when undertaking such and similar study. The phylogenetic studies using the whole plastid needs a careful approach as large indels can arise multiple times and this may confound the phylogeny tree drawn (Bortiri et al., 2008). Inversions in plastid genome is not a very common event but occurs as multiple event in some of the taxa, leading to difficulties in alignment of genomes (Walker et al., 2014). Apart from answering critical questions pertaining to the phylogeny, the comparative chloroplast genome has been used for marking the hotspots of divergence and then from these hotspots new marker systems has been developed (Li et al., 2013; Yang et al., 2013).

2.8 Chloroplast genome sequencing and assembly

The first chloroplast genome was sequenced using shotgun approach, starting from isolation of the organelle followed by restriction digestion of the genome and then preparing a large clone bank (libraries) (Shinozaki et al., 1986). The process was tedious and costly as reflected by only fourty five chloroplast genome sequenced by 2005 (Jansen et al., 2005). In next nine years around 350 plastid genome were sequenced which is attributed to the development of techniques such as large fragment of DNA cloning as Fosmid or bacterial artificial chromosome and next generation of sequencing method. These advances lead to three approaches for chloroplast genome sequencing: polymerase chain reaction (PCR) to amplify chloroplast DNA fragments from genomic DNA (gDNA) extracts (Goremykin et al., 2009), isolation of chloroplast from nuclear extract (Atherton et al., 2010) and whole genome shot gun (Yang et al., 2010). Out of these three strategies the whole shot gun approach has become popular, due to recent price drop of next generation sequencing along with advances in computational management and analyses of these millions of short reads. For initial assembly the De Bruijn assembly method such as Velvet (Zerbino and Birney, 2008) and EULER-SR (Chaisson and Pevezner, 2008) or several related assembly method such as ALLPATHS (Butler et al., 2008), MIRA (Chevreux at al., 1999), YASRA (Ratan, 2009), SOAP2 (Li at al., 2009) are in use. These assembler follows different strategy for reducing the computational resources for assembling millions of short reads. The large contigs generated from such assembler could be de novo assembled if the reads are from more than one sequencing platform or if one of the sequencing platform has mate pair library based sequencing result. Alternatively the contigs can be mapped on the plastid genome of closely related species. The Illumina GA (http://www.illumina.com) short sequence reads of up to 100 base pairs which can be easily converted to contiguous DNA sequence data by reference-guided assembly using an existing genome sequence as a scaffold (Nock et al., 2011).

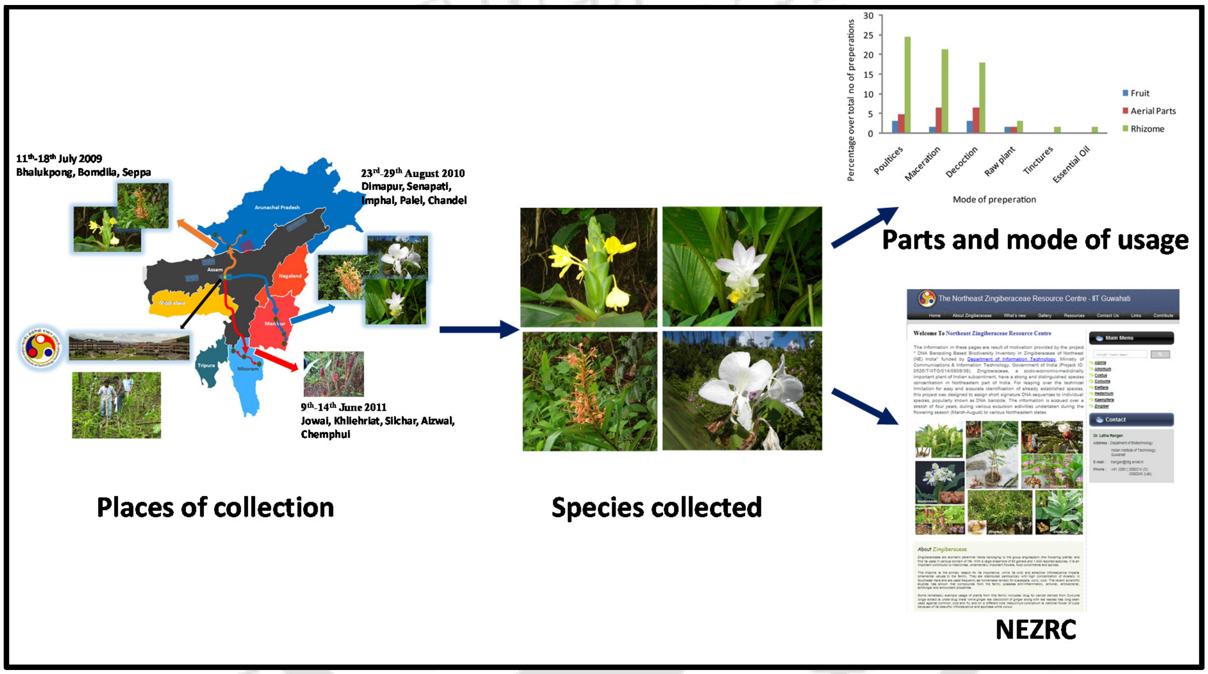
Based on the literature reviewed the objectives were outlined as stated in the section 1.1. The research work was carried out towards achieving the objectives with the aim for increasing the utilization of the plant family by scientific community.

CHAPTER 3

Collection of Zingiberaceae plants from northeast India and documentation of associated ethno medicinal information.

The chapter describes the importance of the plant family and explains the identifying characters which are routinely employed in classification and the associated perils. The alternative methods are discussed and assessed based on the available literature.

GRAPHICAL ABSTRACT



Collection of Zingiberaceae plants from northeast India and documentation of associated ethno medicinal information.

3.1 Introduction

Mankind has been accumulating knowledge since the Homo erectus struck stones to create fire 400,000 years back, and this particular knowledge was reinvented countless time by different isolated herds across the globe (may be at different time). As there was no means to document any of the knowledge accrued by apprenticeship or self-directed learning and educational methods, (probably by sharing incidences) a large knowledge were lost and re-acquired. While the society of lower and middle Paleolithic era has other many different issues to take care of, they bothered very little to retain the information which apparently was not very necessary. But in this contemporary age, the documentation and safeguarding of the traditional knowledge related to ecology is an important issue as it leads to better management and conservation of ecological products.

The family Zingiberaceae well known for its immense medicinal values is distributed widely throughout the tropics, particularly in Southeast Asia. Zingiberaceae family is an important natural resource that provides many useful products for food, spices, medicines, dyes, perfume and aesthetics (Jantan et al., 2003). India is one of the richest and diverse regions for Zingiberaceae, having 22 genera and about 170 species. The NE region of India is a zone of greatest concentration where 19 genera and about 88 species are reported (Prakash and Mehrotra, 1995). Most of the members of Zingiberaceae are found here at wild states, which are yet to be explored. Local literature survey have pointed out that there are numerous medicinal plants described for treatment of many diseases and herbal medicine. However, there was no report on ethnobotanical study of Zingiberaceae of NE to be effective in a very wide range of diseases (Tushar et al., 2010). There is an utmost need for such studies as the environmental scenario of the region is changing drastically due to myriad number of anthropogenic activities and rapid urbanization. Since these information collected is of interest to a diverse group of people (from research/academics to industry to common people) and majority of these demo-graphs do not or cannot access the research papers. Therefore such information if published on web as free material would have broader impact.

Keeping the above points in consideration this study was taken up with the objective to collect the plant specimens from northeast India along with their ethnobotanical information. Further we developed a web portal containing these information for dissemination of the traditional knowledge.

3.2 Methodology

3.2.1 Ethnobotanical study

The current ethnopharmacological survey was conducted in seven states of NE, namely: Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura by making field trips to parts of NE during the period March 2008 to August 2009. Trips were made to villages and areas inhabited by indigenous people during different seasons, so as to include wide number of species of the family. A total of 17 communities (Apatani, Nyishi, Hill Miri, Chakma, Khampits, Adi Khampits, Bodo, Tai-Ahom Hmar Garo, Jaintia Chakma, Lushai, Lakher, Pawi Lushai) were visited during the study. As the members of the family Zingiberaceae are known to bloom during monsoon season, a special account was taken to organize the visits during this season. Most of the Zingiberaceous plants are rare and some are endangered too. Moreover many plants grew on slopes of hills, so at times; it was not only difficult but also inaccessible for collection and subsequent documentation. Since the choice of individual informant to be interviewed was of great importance to the reliability of the information, all contacts were established with elderly and local practitioners. Before interviewing any respondent, the study team members explained the objectives of the study, methods and the plans for use of the data that were to be generated from the interviews. Verbal consent to conduct the interviews was sought from every respondent before the interview and was granted in every case. We selected healers who utilized medicinal plants as therapeutic substance and treated patients outside their immediate circle of family and friends. Particularly in some cases they were the only known healer in the region. A total number of 370 healers were interviewed. The healers were asked to identify the plant species from his/her collection as well as from the natural habitat. A structured interview form (Annexure A2.1) was used to collect information in the local language and respondents were queried for the herbal remedy known to him or her for various diseases. Traditional healers identified some of the not immediately recognized diseases based on the symptoms and physical examinations. Certain diseases are known to healers not by name but by the specific symptoms and condition, as exhibited by patient. These include gastric ulcer, pulmonary tuberculosis, gravel, pneumonia, ascites and diabetes. The name of such disease has been assigned based on the symptoms described by the healers. Since the traditional healers diagnose ailments only by physical examination so at times the disease is identified at a late and acute phase as in pulmonary tuberculosis and ascites. The mode of preparation and administration were also recorded. The acquired data were also cross-checked for its pharmacological significance with the available literature. All doubtful and misleading cases where the informants showed little knowledge concerning the identification of plant species either in the field or from specimens were excluded from this survey and the information provided was discarded. A medicinal use was accepted as valid only if it was confirmed by at least two separate healers. The information obtained was documented in a tabular form. The data included the botanical name, local name, location, plant parts used, mode of preparation, administration and their utility as remedy for treating human diseases. The list of major ailments was compiled and the number of plants utilized in treatment was determined. The collected plants along with the rhizome and flowers were properly tagged and maintained in the departmental green house, IITG and botanical garden of Gauhati University (GU). Hooker (1875) and Petersen (1889) were used as reference for identification of the plants. The features were also preserved as herbaria for future reference at Botanical Garden, GU and a copy of the same is being maintained at IITG with the authors. Taxonomists at Botany Department GU, Assam, later identified the specimens. The botanical name was written as in IPNI database.

3.2.2 Development of database

The database was built using HTML 5.0 and the navigation was indexed using java script and CSS style was incorporated in these pages. The website primarily consists of pages for genus and species covering the respective information for habitat, feature, economic and medicinal importance, scientific classification, synonyms, vernacular name, image of plants various parts (most of them are the image collected by us, but some have been reused with request made to the owner of the image), and the barcode sequence information generated. The portal has been christened as Northeast Zingiberaceae Resource (NEZRC), and is updated regularly and can be accessed at http://www.iitg.ac.in/lrangan/. The basic design conceptualized for the database is explained in Figure 3.1. The page can be broadly divided into two parts, the navigation part and the information part. The navigation part is depicted in blue (light and dark) while all the other part is reserved for representing the information collected. The Dark blue box in Figure 3.1 corresponds to the general navigation bar which would have links to home, about the project page, gallery, contact information for receiving any further information and finally link to the page to provide an easy way to people or organization willing to contribute to the project. While the lighter blue box would bear the links to various genus and species. The red region would have the information such as habitat, features, medicinal importance, economic importance, classification and synonyms. The black parts would have link to sequence information of the barcode loci and would also have downloadable sequence fasta files (to assist standalone version of the database). The white part would act as gallery and the associated pictures with the concerned genus or species would be on run. The green part will discuss about the phylogeny, in case of the page bears information about genus then a brief description would be available about the genus and a link to page discussing the in depth phylogeny of the members of the species. In case of species, the phylogeny drawn using best loci (or combination) was depicted and discussed.

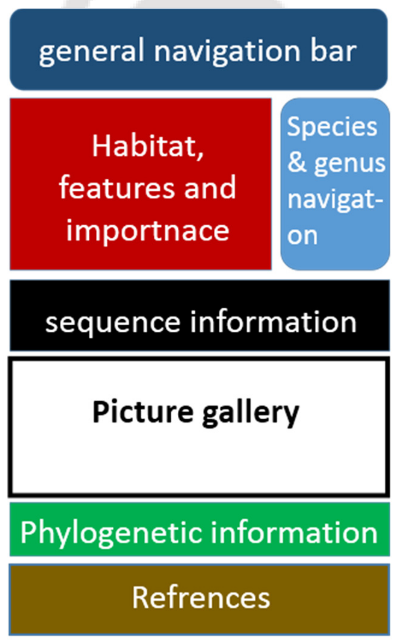


Figure 3.1 Standard layout of the web portal designed for dissemination of the knowledge accrued. The coloured blocks are different section of an individual page and is discussed in details.

3.3 Results and discussion

3.3.1 Ethnobotanical study

The survey involved collection of fifty one plants belonging to nine genera of the family Zingiberaceae from NE region of India. Most of the plant species grow naturally in the different regions and their properties are important in traditional herbal medicine. Therefore some plants are commonly cultivated for its use as spice and vegetables. Thirty four plants out of fifty two were identified to possess medicinal value (Table 3.1 and 3.2). The number of healers from a particular community varied from 3 to 79. During the process of this study, an average of 31 healers per community was interviewed. The average male informer to female informer ratio was found to be 6:5. Interestingly in Garo community from Meghalaya the number of female informer was found to be higher than the male informer. This can be attributed to matrilineal nature of Garo community. Among the states of NE region, AP was found to be floral rich as far as members of the family Zingiberaceae is concerned (88%). Plants like *Zingiber officinale* Rosc. and *Curcuma longa* L. were found to occur in all the 6 sister states and has a wide geographical distribution. All the documented plants were found to be prevalent in use by local practioners for their therapeutic use against human disease and as herbal care as enlisted in Table 3.1. In general the medicinal plants were collected by the males and prepared by the females. Various parts of the plants are being used as herbal medicine, although, rhizome was the most frequently used plant part. This is well supported by the scientific literature as various molecules and extract from the rhizome have shown bioactivity (Ficker et al., 2003; Lantz et al., 2005; Chien et al., 2008; Kumar and Singh, 2008; Lee et al., 2009). Some medicinal preparations were essential oils extracted from various plants parts such as *Curcuma angustifolia* Roxb., *Curcuma caesia* Roxb. showing antifungal activity (Banerjee and Nigam, 1976; 1977), *Alpinia malaccensis* Rosc., *Costus spiralis* Rosc. and *Zingiber cassumunar* Roxb. showing antimicrobial and antioxidant activities (Habsahet al., 2000). The essential oils are a good source of fragrance apart from having the desired therapeutic use. Another thing that emerged out from the survey was that the practitioner used the plants singly or in association with different plant species for treating human ailments. The use of single plant as a therapy for an ailment was observed to be 84% while 16% of herbal preparation involved combination of plants. The remedies, which involved merely the use of single plant, could be of great interest for the development of novel drugs as the exploration of therapeutic activity bearing ingredients from a single plant may be easier (Saikia et al., 2006). The majority of herbal medicines were prepared in aqueous medium and the healers administer the remedies in various forms like poultices (32.3%), maceration (28.41%), decoction (27.4%), raw plant parts (6.5%), essential oil (3.2%) and tincture (1.6%) (The percentage shown is over total number of preparations). The individual percentage breakdown of plant parts used in different type of preparation further shows that tincture and essential oils are prepared only from rhizome (Figure 3.2). The remedies were administered orally or locally applied to the affected portion of the body according to the treated disease and preparation method.

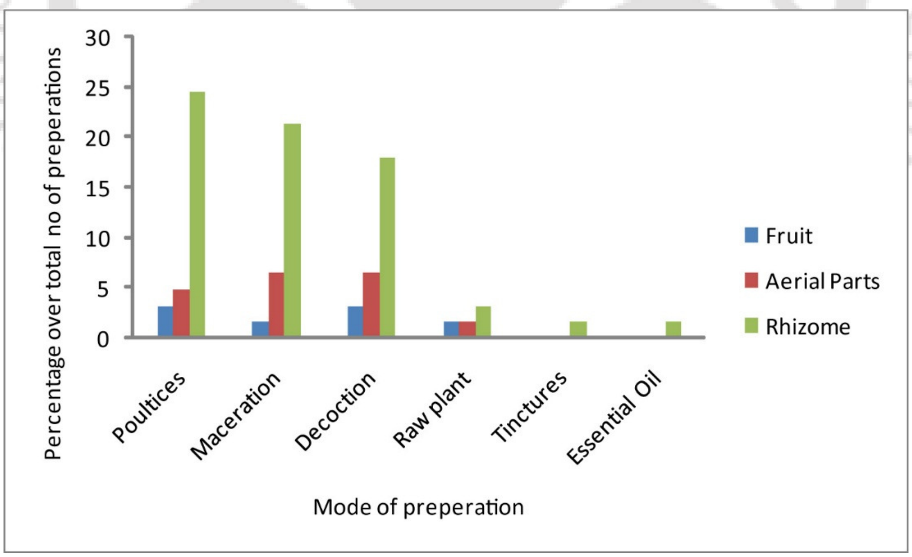


Figure 3.2. Frequency of plant parts used according to the mode of preparation of herbal medicine. Decoction: aqueous extract of plant material; tinctures: alcoholic extract of plant material; maceration: grinded and taken orally; poultices: grinded and applied locally; essential oil: concentrated, hydrophobic liquid containing volatile aroma compounds from plants; raw plant: specified part of the plant either taken orally or applied locally.

In addition to pure herbal preparations, in some cases the drug was administered along with milk, ghee, honey, curd, etc. to enhance the effect of the herbal preparations or to make the preparations palatable. In current survey, the use of only underground plant parts for medicinal purposes was found to be higher (56%) than the case where both above and underground parts were used together (29%). The usages of aboveground plant parts were comparatively less (only 15%). Rhizomes were used in most of the preparations (60%) followed by fruits (14%), leaves (12%), shoot (12%) and flower (2%) (The percentage shown is over total number of preparation). In two cases viz. *Hedychium coronarium* Koenig and *Costus specious* Sm. all plant parts were found to have a medicinal value. Thirty-four medicinal plants were used in curing about 25 different types of ailments, of which the highest number of plant species (58%) were used for the treatment of gastrointestinal disorders and 41% of them for curing chest and lungs related diseases (Figure 3.3) (percentage calculated over total number of applications). The plants also showed effective activity as antipyretic, analgesic and anti-inflammatory. We observed that some plants were used for ailment like cardiac disorder, kidney and urinary disorder, skin related diseases, irregular menstrual cycle, diabetes and as abortifacient. The survey of scientific literature has yielded maximum report of bioactivities against gastrointestinal ailments (Yamahara et al., 1990; Matsuda et al., 2003b; Gilani et al., 2005; Nanjundaiah et al., 2011; Thong-Ngam, 2012) and as an antioxidant (Adlercreutz, 1984; Habsah et al., 2000; Ramsewak et al., 2000; Vankar et al., 2006; Chen et al., 2011; Yeh et al., 2014). A large amount of literature is also available for anti-cancerous activity but this is not a generalized feature of this family rather attributed to promising effect of *Curcuma* genus. A close analysis revealed that about 67.6% plants are known for their use to cure multiple disorders. Among these, *Costus specious* and *Elettaria cardamomum* Maton were primarily effective against wide range of diseases. The rhizome powder from some plants such as

Amomum subulatum Roxb., *Curcuma aromatica* Salisb., *Elettaria cardamomum* and *Hedychium spicatum* Sm. are used as anti-venom for snake and insect bite.

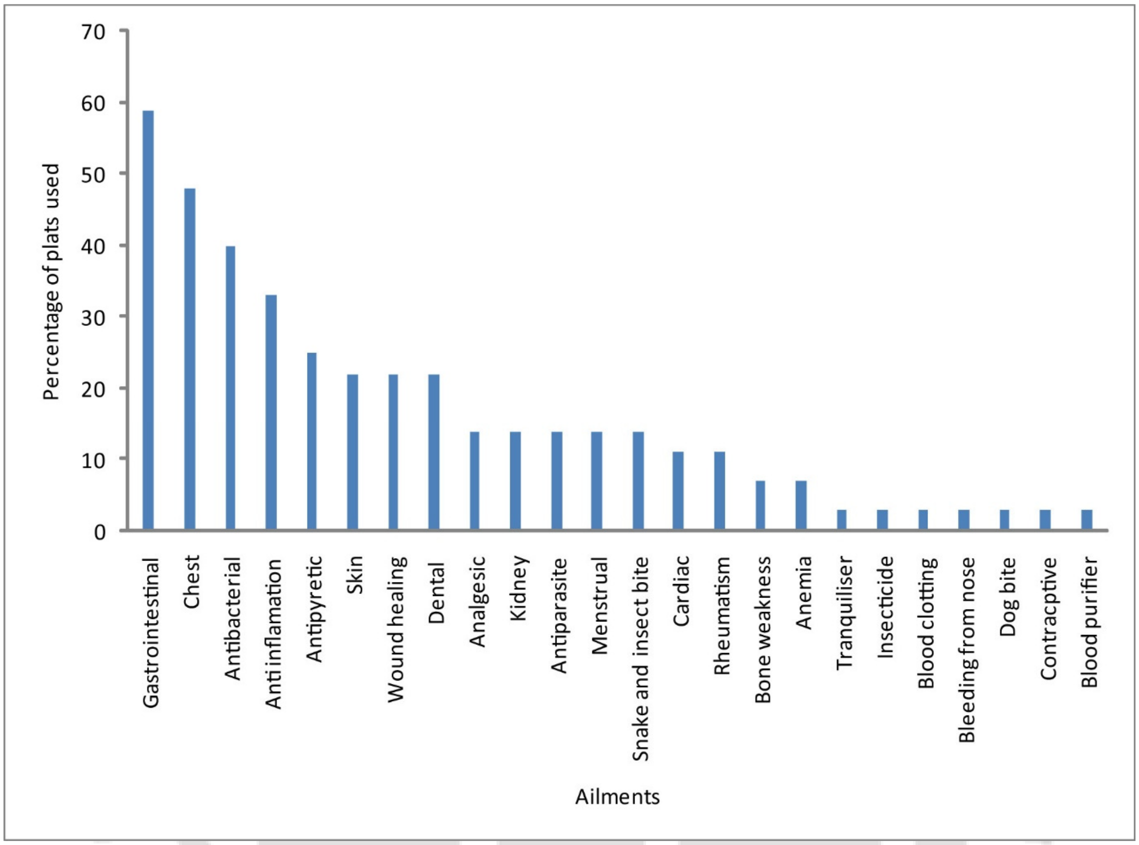


Figure 3.3. Percentage use of Zingiberaceous plants for treating various human ailments. Highest percentage of plants used is observed for treatment of gastrointestinal ailments that is also supported by large number of pharmacological reports. Percentage was calculated over number of plants documented in the current study.

Mostly the rhizome as a whole or as paste is used for medication. In certain cases, preparation of medicine involved use of a mixture of herbs and sometime molasses. The use of mixture of other herbs with Zingiberaceae members were also documented such as *Alpinia galanga* L. along with *Euphorbia neriifolia* L. for treating skin allergy, *Curcuma caesia* along with *Andrographis paniculata* (Burm. f.) Wallich ex Nees. As anti-asthmatic agent and anti-inflammatory in case of insect and snake bite, *Globba clarkei* Baker along with *Adhatoda vasica* Nees. for treating cough, *Zingiber officinale* along with *Nigella sativa* L. as anti-inflammatory and *Zingiber montanum* (Koen.) Link ex A. Dietr. along *with Kaempferia galanga* L. and *Zingiber officinale* for treating diarrhoea. It was interesting to note that most of the plants that were found in all the 7 states of NE region had more or less same utility and bear a similar resemblance from the other parts of India. There were also instances where a particular plant species was used by a particular ethnic community from one region for treating a particular disease and the same plant species with a different name was used by different community from other location for treating an altogether different ailment. For example, *Alpinia galanga* is used as an anti-inflammatory agent by Khampit community in state of AP while the same species is also used as an abortifacient by ethnic community of Barak valley in Assam. Similarly, *Amomum dealbatum* Roxb. is used as an analgesic by indigenous communities in Assam such as Lushai and Jaintia while it is used as an antiseptic in west Mizoram. Since the indigenous communities, which are residing in close proximity uses a plant in a similar way against same ailment, the proximity of distance from one to another community might be an important factor in widening or narrowing the scale of herbal practices. The use of *Amomum dealbatum* species as an analgesic and as an antiseptic has been reported earlier but its use against rheumatic joints is a new report. The communities of Borok apply the crushed, fried rhizomes to relieve the joints pains. The use of *Amomum linguiforme* Benth. and Hook.f. against diabetes has been reported earlier but its use against high blood pressure by the Nyishi community of AP is a new report. Use of *Globba multiflora* Wall. and *Hedychium coccineum* Buch.-Ham. ex Sm. species as an antipyretics and anti-inflammatory (in case of bruise andwounds) respectively, has never been reported till date. Similarly the property of *Costus variagata* L. is known for generation's altogether but its use for curing piles is a new report. From this study certain plants have emerged as critical species, which deserves extra attention by researchers, and are possible source of potent drug molecule. For example use of *Alpinia bracteata* Rosc. crushed rhizome as a cure for abnormal menstruation. Another plant belonging to the same genus as previous, *Alpinia galanga* L. is used as an abortifacient. These two examples are inkling towards presence of a potent molecule, which can moderate hormone levels. Further *in vitro* studies can help in elucidation of the mechanism involved. *Costus specious* seed was found to be used as contraceptive and its importance is augmented by the fact that it is not used in combination with any other herbs. *Zingiber purpureum* Rosc. use against paralysis, points its activity as neural or muscular stimulant. A detailed study of the bioactive of the above species may yield novel drug molecule. Apart from this study, there have been some reports of ethnomedical survey, in which Zingiberaceous species from NE region have been documented. In particular, some species that has been cited in other studies, but not reported in this study include *Alpinia officinarum* Hance. (Sharma et al., 2001) and *Hedychium dekianum* A.S. Rao & Verma (Kala, 2005). As 34 plant species have been reported in this particular study and including the two other members that has been cited in other studies, the count reaches to 36. A total of 88 Zingiberaceae members have been reported from the NE region alone out of which 41% have been found to possess medicinal property. Thus it can be said now that the discovery of different plant species belonging to the family Zingiberaceae used by different ethnic communities of NE India lays the grounds for establishing Zingiberaceae as a medicinal family.

3.3.2 Database

The database was developed and for easy accessibility has been put in a web portal format. The database was designed as originally incepted (Figure 3.1), with very few and minor alterations (Figure 3.4). The top most part consists of navigation tools and also displays the name of the database and the logo of the institution. The home page deals with general information about the Zingiberaceae family and displays the diverse nature of the plant morphology by collections of image. The home page describes features, characteristics and other important information such as social, medicinal, economical importance of the plant family (Figure 3.5). One can access the various genera pages from the home page as well as from any other page, as the links are provided on the left side. Once a genera page is visited, the left side navigation tool extends to display the list of the species contained in the particular genera (Figure 3.6 and 3.7)

Figure 3.4. The detailed layout of the database developed.

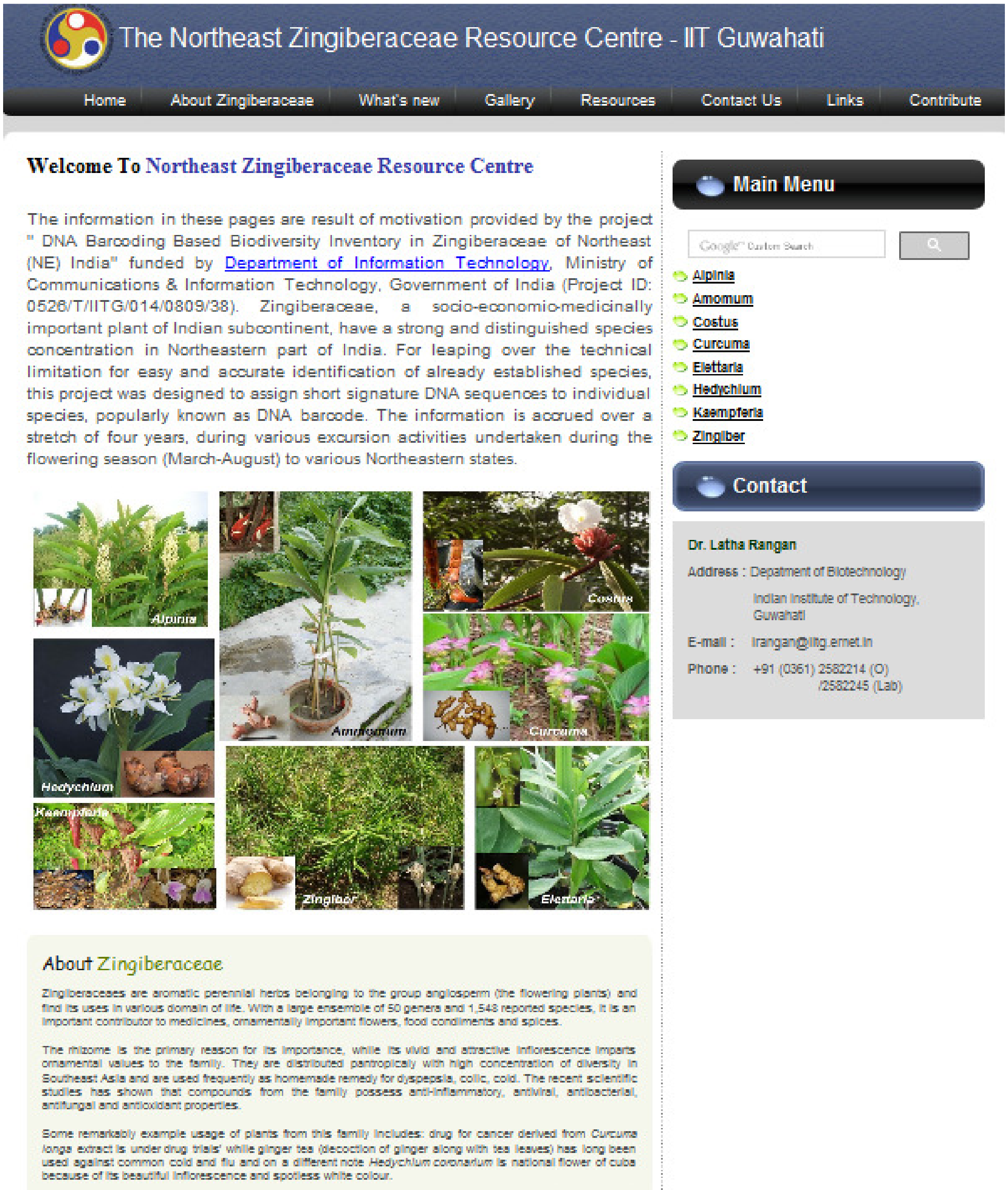


Figure 3.5. Home page showing information about the plant family and various links which can be accessed

Figure 3.6. A snapshot of the database displaying the genera page. The bottom of the page shows links and downloadable files for the DNA sequences used for the phylogenetic analysis.

(a) **(b)**

Figure 3.7. Snapshots from the NEZRC depicting (a) Species page with the phylogenetic information shown in the bottom of the page along. (b) The contribution page which could be used by the users to contribute to the information.

Apart from the botanical features and economic importance the details of the phylogenetic relationship among the various taxa of the family could be accessed from the genera or species page. The phylogenetic relation included in the database was reconstructed on the basis of the DNA sequences obtained from plastid genome (discussed in detail in chapter 4) and is result of various combination of the DNA loci. All the sequence information is also assembled in the database where it can be downloaded as FASTA file, while the link to the public domain where these sequences has been deposited is available along with the accession number. Further for the ease of users who wants to contribute to the database, a contribution page (Figure 3.7) has been designed which is in an e-form format, where one can input information and after it has been checked for authenticity and reliability can be incorporated in the database.

3.4 Conclusion

Northeast India is an important reservoir of rich flora with diverse ethnic population and their traditionally conserved knowledge. In this study the information regarding the medicinal uses of Zingiberaceae family by aboriginal people of NE India was systematically collected and critically analysed and for better dissemination has been put in as a web portal. The web portal are more user friendly and appeals more to general public because of the presence of visuals and well categorised data. A total of 34 plant species belonging to the family were identified to possess medicinal importance and primary activity seems to be against gastrointestinal ailments. This is also reflected by the scientific literature (Yamazaki et al., 2000; Gilani et al., 2005). In at least two ethnic communities, the highest number of species used for the gastrointestinal ailments have been found and it needs to be exploited. The unhygienic livelihood compounded by the excessive alcohol consumption of the surveyed ethnic communities was found to be major concern of the severe gastrointestinal problems. This is followed by the chest related problems, which was due to excessive smoking and chewing of tobacco among the males. The use of single plant as a therapy for an ailment was observed to be 84% which is a good indication for future research leading to drug discovery. Poultices (32.3%) and maceration (28.41%) were the most frequent way of preparation and hence no complex procedure involved in application of these species. *Alpinia bracteata, Alpinia galenga* L., *Costus specious* and *Zingiber purpureum* are some of the important species for future work. The future research focused on coupled phytochemicals-pharmacological studies can help in elucidation of novel phytotherapeutic molecules. Most of the persons from whom we had gathered the information are illiterate and at times there is no script for the language they speak, hence the herbal practices are running in mouth to mouth without any written documents. Therefore there is an utmost need to document the traditional and cultural practices including from remotely located region where the different communities has been living in harmony for century's altogether. The development of the database and making it public aims at exercising the above said objective and would also help in proper utilization of these bio resources and will also help industries who are willing to tap these resources. The database which have been made accessible through a web portal aims at targeting nonscientific community specially the one who are directly involved with the end application of the data acquired, which may include farmers, entrepreneurs, pharmacy industries. While the information which is critically required by the scientific community such as medicinal importance, phylogenetic relationships and the sequence information has also been included. The work has brought to light some hidden but popular prescriptions of ethnic group of the NE region. These new prescription will help mankind in short term, by providing improved phytotherapeutic preparations while an extensive pharmacological study will elucidate new drug molecule. Though some scientific literature is available for most of the species, the specific properties of all the species has not been studied yet. A further pharmacological study of these particular and scientifically unexplored properties of Zingiberaceae appears promising. Further, a concrete list of plants and their utilization will not only provide basic data for all studies aimed at conservation, cultivation, healthcare and economic welfare of rural and indigenous people, but all future pharmacological and clinical studies. Since NE region is rich in species of Zingiberaceae with majority of them possessing medicinal values (41%), the extensive ethnobotancial study should be carried out. An urgent need to take step for conservation and inventory of Zingiberaceae species from NE region including varieties from all the possible location is also felt, as the medicinal value of plants are known to vary with the geographic location (Datte et al., 1998). The pharmacological study to identify the usefulness of plant against a particular ailments and phytochemical study for identifying the exact molecule(s) from such inventory is recommended. This report is the good example, which compiles the ethnobotanical information that may lead to the discovery of novel pharmaceuticals. The database constructed which is compilation of the diverse data, ranging from the morphological information to the molecular phylogeny of the plant. The database would be a very effective tool for researcher who would like to pursue towards utilizing the economic benefits of Zingiberaceae.

Table 3.1: List of plants collected and their application in context to ethnomedicine as prevalent in use by various ethnic communities of NE region, India.

Botanical name	Local name	Location	Parts used	Mode of preparation	Usefulness	Experimental data available
Alpinia allughas (RETX) ROS.	Deotora (Chakma, Arunachal Pradesh)	Arunachal Pradesh	Fruits and rhizome	Crushed fruits are applied. Rhizome extract taken orally.	To treat ring worm. Antipyretic	Antioxidant (Vankar et al., 2006)
Alpinia bracteata ROSC. (Syn *Alpinia calcarata*).	Latara (Khasi)	Arunachal Pradesh	Rhizome	Crushed rhizome	Tooth decay, abnormal menstruation	Antinociceptive (Arambewela et al., 2004), Antioxidant (Thabrew et al., 2001)
Alpinia galanga L.	Karphul, Kulajan (Assamese); Kanghoo (Manipuri); Aaichal (Mizo); King Pang (Khampti, Arunachal Pradesh)	Arunachal Pradesh, Assam, Manipuri, Mizoram, Nagaland	Rhizome and Shoot	The rhizome is used to create a tincture that is applied topically to treat fungal skin infections. Rhizome powder mixed with young leaf paste of *Euphorbia neriifolia* and applied locally. Plant extract	Stimulant and carminative in flatulence, laxative dyspepsia, vomiting and sickness at stomach. Inflammation and skin allergy caused by insect bites or microbes. For abortion	Antifungal (Janssen and Scheffer, 1985), Antiallergic (Matsuda et al., 2003a), Gastroprotective (Matsuda et al., 2003b), Antimycobacterial (Phongpaichit et al., 2006)

Alpinia malaccensis ROSC.	Puprere (Adi, Arunachal Pradesh)	Arunachal Pradesh	Rhizome	Rhizomes boiled with water.	Abdominal pain	Antioxidant (Habsah et al., 2000)
Alpinia nigra BURTT.	Pullei (Manipuri); Tora (Assamese)	Assam, Manipur	Tender shoot and rhizome	Shoot extract Rhizome extract is taken 2-3 times daily.	Bone weakness, irregular menstruation Jaundice, gastric ulcers	Flukicidal (Roy and Tandon, 1999), Antimycobacterial (Phongpaichit et al., 2006)
Amomum aromaticum ROXB.	Borelachi (Chakma, Arunachal Pradesh); Bodaelachi (Assamese)	Arunachal Pradesh, Assam	Seeds	Seed paste	Antidote to scorpion and snakebite.	Antimycobacterial (Phongpaichit et al., 2006)
Amomum dealbatum ROXB.	Aidu (Mizo); Alach (Kokborok); Alachi (Tripura)	Arunachal Pradesh, Assam, Tripura	Rhizome leaves, seeds	Rhizomes are crushed and fried lightly with mustard oil and applied to joints. Leave juice Rhizome extract applied locally. Seeds powder mixed with equal amount of honey.	Joints pain Antiseptic Abscesses Muscular rheumatism	None

Amomum linguiforme BENTH. & HOOK. F.	Karpur (Tai-ahom)	Arunachal Pradesh, Assam	Rhizome.	Epidermis is removed from rhizome under water.	Diabetes, High blood pressure	None
Amomum subulatum ROXB.	Bara elachi (Bengali)	Arunachal Pradesh, Assam	Pods	By boiling 2-3 pods for 30 minutes in water. Strain and drink the extract twice daily for week.	Eye inflammation, kidney and urinary disorder, infection of teeth, prevents and treats throat trouble, congestion of lung and pulmonary tuberculosis, asthma, heart disease, digestive disorder, cold, bladder disease, snake bite, scorpion bite, masticatory.	Antimicrobial (Islam et al., 1990), Anti-inflammatory, Analgesic and Antispasmodic (Al-Zuhair et al., 1996)
Amomum zingiber L.	Ardraka (Sanskrit)	Assam, Meghalaya	Rhizome	Drinking of water extract of dried rhizome for a month.	Filaria	Antibacterial and Antifungal (Islam et al., 1990)
Costus speciosus SM.	Jamlakhuti (Assamese); Khongbam takhelei (Manipuri); Sumbul (Mizo); Jomalkhuti (Chakma, Arunachal Pradesh); Myonpobap	Arunachal Pradesh, Assam, Manipur, Meghalaya, Nagaland, Tripura	Whole plant including rhizome	Rhizome extract	Burning sensation, flatulence, constipation, helminthiasis, leprosy, skin diseases, fever, hiccough, asthma, bronchitis, bleeding from nose and mouth inflammation and anaemia Tonsillitis	Sexual hormones and contraceptives (Warrier et al., 1993)

	(Nyishi, Arunachal Pradesh)			Raw plant is taken orally. Fresh rhizome extract orally taken twice daily for 3 days. Raw seeds are chewed several times. Rhizome essential oil. Stem extract mixed with sugarcane juice is taken orally. Rhizome and stem are eaten raw. Warm stem juice is applied. Dried powdered leaf and young stems mixed with lukewarm water.	Dog bite Contraceptive Jaundice Urinary problems and blood in urine Snake or insect bite Burning wounds For curing piles	

Costus variagata L.	Kewpachla (Assamese)	Assam	Leaf and stem	Leaf and stem extract in water.	For curing piles	None
Curcuma amada ROXB.	Amada (Assamese); Aiengpui (Mizo)	Arunachal Pradesh, Assam, Manipur, Mizoram	Rhizome	Rhizome extract and paste.	Carminative, bronchiolytic and vulnerary	Anti-inflammatory (Mujumdar et al., 2000), Cholesterol lowering activity (Srinivasan et al., 2008) Antioxidant and Antibacterial (Policegoudra et al., 2007)
Curcuma angustifolia ROXB.	Yaipan (Manipuri); Gorusat haladhi (Assamese)	Arunachal Pradesh, Assam, Nagaland	Rhizome	Pudding using rhizome powder with milk and sugar is used as general tonic for children. The powder of rhizomes with honey is applied on the mucous membrane of the oral cavity. Rhizome paste is applied to cattle injured by leech.	Demulcent, antipyretic, effective against gravel Stomatitis Aid in blood coagulation	Antifungal (Banerjee and Nigam, 1977)
Curcuma aromatica SALISB.	Bon haladhi (Assamese); Lam-yaingang (Manipuri)	Arunachal Pradesh, Assam, Manipur, Meghalaya	Rhizome	Paste of rhizome with milk is used for blood dysentery and stomach-ache.	Carminative, antidote to snake bite, astringent and used for bruises, corns and sprains	Inhibits proliferation of hepatoma (Wu et al., 2000), Anticancer activity performed on Eight-week

				Paste of rhizome taken with water.	Kills intestinal worms	old Sprague-Dawley rats (Li et al., 2008), Larvicidal (Madhu et al., 2010)
Curcuma caesia ROXB.	Kalahalud (Assamese); Amuba yaingang (Manipuri); Aihang (Mizo); Chongkah (Khampti, Arunachal Pradesh); Homen (Lohit, Arunachal Pradesh)	Arunachal Pradesh, Assam, Mizoram	Rhizome and seeds	The dried rhizome powder is mixed with powdered seeds of *Andrographis paniculata* and applied during insect, scorpion and snake bite. Fresh rhizome juice along with mustard oil is given daily. Rhizome paste	Against inflammation caused by insect and snake bite (Anti inflammatory) and anti-asthmatic Dysentery For healing of wound	Antimicrobial (Garg and Jain, 1998), Antifungal (Banerjee and Nigam, 1976)
Curcuma domestica (MEDIK) VALH.	Haldi (Assamese); Yaingang (Manipuri)	Arunachal Pradesh, Assam, Manipur, Nagaland	Rhizome	The dried powder of rhizome.	Against burn by fire (anti-inflammatory), flatulence, jaundice, Scabies	Anticancer activity performed on national human prostate epithelial cell line PrEC and normal human bronchial epithelial cell line Beas2B (Shankar and Srivastava, 2007), Antioxidant (Ramsewak et al., 2000), Anti-inflammatory (Ramsewak et al., 2000; Chainani-Wu, 2003), Antifungal (Apisariyakul et al., 1995), Antibacterial

						(Kim et al., 2005), Antiviral (Sindelarova et al. 1996)
Curcuma longa L.	Haladhi (Assamese); Aieng (Mizoram); Khumein Nak (Khampti, Arunachal, Pradesh)	Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura	Rhizome	Crushed rhizomes are taken orally before food. Rhizomes are chewed for relief from asthma.	Dyspepsia Vasodilator	Anti-inflammatory, Antispasmodic activity (Ammon and Wahl, 1991), Anti-HIV, Antioxidant, Anti-tumour, Anti venom (Araújo and Leon, 2001)
Curcuma zedoaria (CHRIST.) ROSCOE.	Keturi (Assamese); Aidizung (Mizo).	Arunachal Pradesh, Assam, Mizoram	Rhizome	Rhizome extract	Blood purifier, cough antiseptic, indigestion, wound healing, toothache, leucoderma, tuberculosis, enlargement of spleen and for promoting menstruation.	Anticancerous activity on Human ovarian cancer OVCAR-3 cells (Syu et al., 1998)
Elettaria cardamomum MATON	Elassi (Assamese)	Arunachal Pradesh, Assam	Seeds and pods	Infusion, powder, milk decoction	Eye inflammation, kidney and urinary disorder, to treat infection of teeth, prevent and treat throat trouble, congestion of lung and pulmonary tuberculosis, asthma, heart disease, digestive disorder, cold, snake bite, scorpion bite, masticatory.	Antimicrobial (Islam et al., 1990), Anti-inflammatory, Analgesic and Antispasmodic (Al-Zuhair et al., 1996)

Globba clarkei BAKER.	Silaadha (Chakma, Arunachal Pradesh)	Arunachal Pradesh	Rhizome	Rhizome extract mixed with leaf extract of *Adhatoda vasica* is taken orally.	To cure cough	None
Globba multiflora BAKER.	Belah (Nyishi Arunachal Pradesh)	Arunachal Pradesh	Rhizome	Crushed rhizome applied locally on wound.	Analgesic and Antipyretics	None
Hedychium coccineum BUCH.-HAM. EX SM.	Aichhia (Mizoram); Mansila (Lohit Arunacahal Pradesh)	Arunachal Pradesh, Mizoram	Rhizome	Rhizome paste applied over swollen part.	Against swelling caused by bruises and wounds (Anti-inflammatory)	None
Hedychium coronarium KOENIG.	Pakhila phul (Assamese); Tora (Chakma Arunachal Pradesh)	Arunachal Pradesh, Assam	Flowers, rhizome and stem	Flower paste Rhizome extract prepared by boiling the rhizome in water. Essential oil from rhizome.	Foetid nostrils Febrifuge, tonic and anti rheumatic swellings Antihelmintic, tonic and mild tranquiliser	Antibacterial (Aziz et al., 2009)
Hedychium spicatum SM.	Takhellei-hanggam-mapan (Manipuri); Aithur (Mizo) Karpurakachari (Bengali)	Arunachal Pradesh, Manipur, Meghalaya, Mizoram	Rhizome	Root decoction	Nausea, Bronchial Asthma, Halitosis, Vomiting and indigestion, expectorant; stimulant; stomachic Treatment of liver complaint, treating	Antibacterial against methicillin and vancomycin resistant *Staphylococcus aureus* and fungal cultures (Bisht et al., 2006)

				Powder of root	fevers, vomiting, diarrhoea, inflammation, pains and snake bite, heating potency to the female, stimulant, expectorant, tonic, carmative	
Kaempferia galanga L.	Chandramula (Assamese); Sying khmoh, Sying shmoh (Khasi)	Arunachal Pradesh, Assam	Leaves and rhizomes	The rhizome is externally used. Rhizome is taken orally. Used together with the rhizomes of *Z. montanum,* and *Z. officinale.*	Treating indigestion, cold, pectoral and abdominal pains, headache, carminative and toothache, menstrual pain, insecticidal. Effective for dandruff or scabs on the head Against poisoning when there is blood vomiting Antidiarrheal	Larvicidal activity (Othman et al., 2006), Inhibits activity of Epstein-Barr virus (Kanjanapothi et al., 2004), kills larvae of the mosquito (Ahn et al., 2008)

Kaempferia pulchra RIDL.	Khanjanburah (Assamese)	Assam, Mizoram	Rhizome leaves and stem	Rhizome paste is used in the treatment of pneumonia and bronchial complaints. Steamed rhizomes, stems and leaves for curing wound.	Pneumonia, bronchial complaints Wound healing	Topical anti-inflammatory activity (Pongprayoon et al., 1996)
Kaempferia rotunda L.	Bhuyichampa (Assamese); Yai-thamna-manbi (Manipuri); Tuktinpar, Tuktin-par (Mizo)	Assam, Manipur, Mizoram	Rhizome and leaves	The rhizomes are used in local medicine by grinding (fresh or dried) and making a paste with water. This paste is mixed with other herbs and applied to sprains and covered with a bandage. The leave paste is used as body lotion.	Sprain ache, stomachic	Insecticidal against neonate larvae of the *Spodoptera littoralis* (Nugroho et al., 1996)
Zingiber cassumunar ROXB.	Bura-ud (Assamese); Naga-shing (Manipuri); Manthing (Mizoram); Banada (Bengali)	Assam, Manipur, Mizoram,Meghalay, Arunachal Pradesh	Rhizome	Rhizome paste with salt is given orally once for two days.	Indigestion/ gas formation	Anti-inflammatory, Anti-allergic, Antioxidant activity and are used against asthma and muscle and joint pain (Jeenapongsa et al., 2003)
Zingiber chrysanthum ROSC.	Sobleksin (Assamese)	Arunachal Pradesh, Assam	Fruit	Rhizome powder is mixed in a glass of water and drank in instances of nausea.	Nausea	Antioxidant (Adlercreutz, 1984)

Zingiber montanum (KOEN.) LINK EX A. DIETR.	Banada (Assamese)	Arunachal Pradesh, Assam	Rhizome	Used together with the rhizomes of *Z. officinale,* and *Kaempferia galanga*. As a rhizome paste.	Antidiarrheal Antidote to snake venom	Anti-inflammatory activity (Khayungamnawee et al., 2009)
Zingiber officinale ROSC.	Ada (Assamese); Shing (Manipuri)	Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura	Rhizome	The juice extracted from the rhizome is mixed with molasses and taken orally. Rhizomes along *Nigella sativa* and spider are made into paste and applied locally.	Promotes digestive power, cleanses the throat and tongue, dispels cardiac disorders and cures vomiting, ascites, cough, dyspnoea, anorexia, fever, anaemia, flatulence, colic, constipation, swelling, elephantiasis and dysuria. Inflammation from irritation caused by caterpillar. Antidiarrheal	Rheumatism and inflammation of liver (Aiyer and Kolammal, 1966; Kurup et al., 1979). Ingredient of *Indukantam kashaya, Suranadi lehya, Talisapatravataka, Visvamrta* etc. (Sivarajan and Balachandran, 1994)

				Used together with the rhizomes of *Z. montanum,* and *Kaempferia galanga*.		
Zingiber purpureum ROSC.	Bon ada (Assamese) Pale; ramthing (Mizoram); Borahu (Assamese)	Arunachal Pradesh, Assam, Mizoram	Rhizome	Juice extracted from rhizome.	Paralysis	Antifungal (Ficker et al., 2003)
Zingiber zerumbet ROSC. EX SM.	Gathian (Assamese); Yaiimu (Manipuri)	Arunachal Pradesh, Assam, Manipur	Rhizome	Mixture of Rhizome powder and ripe Noni fruit (*Morinda citrifolia*) is applied on sprains. Cooked and softened rhizome was pressed into the hollow and left for as long as was needed. Ground and strained rhizome material is mixed with water and drunk.	Treats severe pain Tooth ache and cavity cough, asthma, worms, leprosy and other skin diseases Stomach ache	Anti-inflammatory (Elliott and Brimacombe, 1987) and Anti cancer activity on HL60 (human promyelocytes); LS174T, LS180, COLO205, COLO320DM (each human colon adenocarci-noma cells); CCD-18 Co (human colon normal fibroblast); AS52 cells (Chinese hamster ovary cells); RAW264.7 cells (mouse macrophages); 2F0-C25 cells (human normal fibroblasts) (Murakami et al., 2002)

Table 3.2: List of plants collected from NEI with no ethno-medicinal values observed.

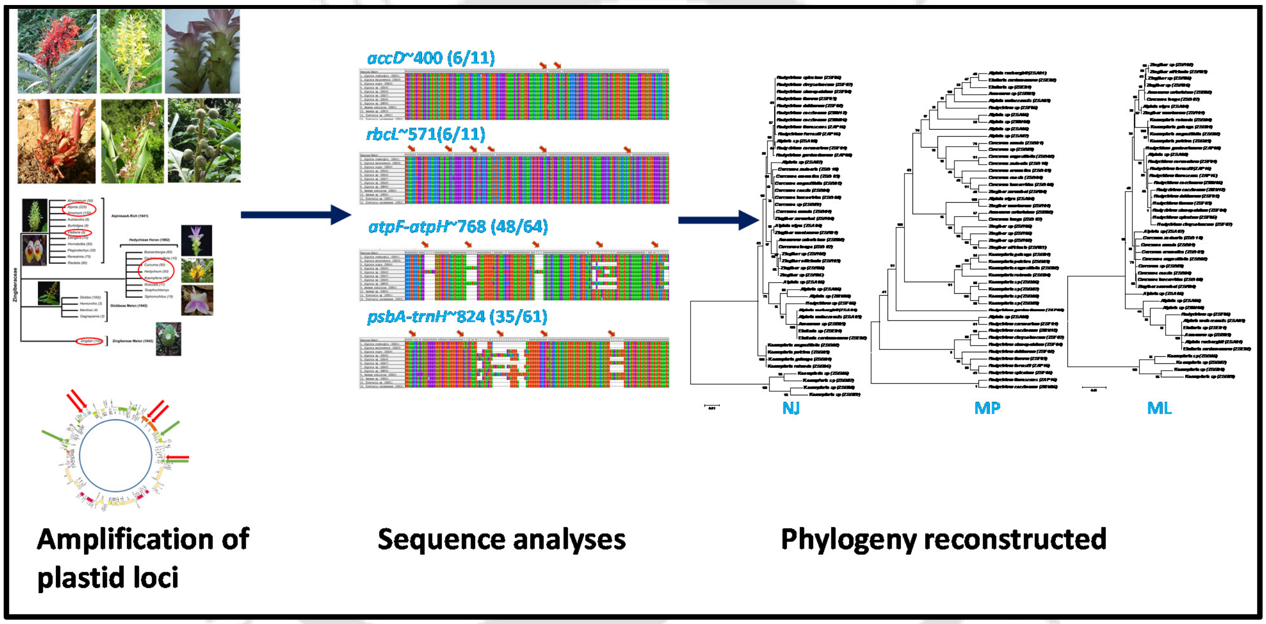
Serial no	Species name	Unique tag no	Collected from
1 ***Alpinia***			
1.1	*A. species*	ZSA 05	Amingaon (Assam)
1.2	*Alpinia sp*	ZSA 08	Seijusha (A.P.)
2. *Amomum*			
2.1	*Amomum species*	ZSB 03	Sonapur (Assam)
3 *Eletteria*			
3.1	*E. big*	ZSE 01	NEDFi
4 *Curcuma*			
4.1	*C. wild*	ZSD 09	Kokrajhar (Assam)
5 *Kampferia*			
5.1	*K. angustifolia*	ZSG 02	Jalukbari (Assam)
5.2	*K. ornamentalIV*	ZSG 05	Goalpara (Assam)
5.3	*K. ornamental I*	ZSG 06	Sonapur (Assam)
5.4	*K. ornamental II*	ZSG 07	Sonapur (Assam)
5.5	*K. ornamental III*	ZSG 08	Sonapur (Assam)
6	***Zingiber***		
6.1	*Z. moran*	ZSH 02	Sivasagar (Assam)
6.2	*Z. species*	ZSH 05	Darrang (Assam)
6.3	*Z. species*	ZSH 06	Zowai (Mizoram)

CHAPTER 4

Mining of plastid genome for phylogenetic relationship assessment

The chapter deals with reconstruction of phylogeny of Zingiberaceae by using plastid loci in singular and combinations

GRAPHICAL ABSTRACT



Mining of plastid genome for phylogenetic relationship assessment

4.1 Introduction

With the cognizant thinking came the desire to discover and with discovery came the urge to classify events, objects and phenomenon into categories. The earliest classification would have been for animals as they were source of food and danger. In retrospect one of the first basis of such classification must have been how many hunters required to hunt a 'game'. Some thousands later and almost 250 years ago from present time came a Swedish biologist Carolus Linnaeus who agreeing with an ancient Chinese proverb "The beginning of wisdom is to call things by their right names" classified approximately 12,000 living beings (http://www.smithsonianmag.com/science-nature/organization-man-151908042/) by grouping them with the most similar visible characteristics in groups. Since then a number of ways to classify living organism has been developed and adopted such as chemical profiles (Goodner et al., 2001; Correa and Goodacre, 2011), morphological data matrices (Tillyard, 1921; Gauthier et al., 1989; Lewis 2001; Lambkin et al., 2013), data matrices for polymorphism widespread in genome (Botstein, 1980; Bowcock, 1994; Welsh and McClelland 1990; Vos et al., 1995), and molecular sequences (Neyman, 1974; APG II, 2003; Bowe et al., 1999; Birgitta, et al., 2009; Wu et al., 2009; Perelman, 2011; Rinke et al., 2013). These all attempts were focused towards more accurate taxonomy as accurate taxonomy and correct sample identification are crucial to many branches of science which have direct impact on the human life and society.

Taxonomy based on morphological character(s) has long been the most accepted way of classifying organisms but the extreme challenge of identifying and classifying all species in this way is laborious and due to numerous evolutionary and ecological phenomenon such as convergent evolution or even mimicry. Hence the use of molecular sequence for reconstructing plant phylogeny has emerged as a populist approach for classification after the pioneer work of Woese and Fox (1977).

In the plant kingdom, Zingiberaceae is a large family with pan tropical distribution. The classical taxonomy of the family recognizes four tribes (Globbeae, Hedychieae, Alpinieae, and Zingibereae) based on morphological features such as number of locules, number of placentation in the ovary, staminodes development, fertile anther modifications and orientation of the rhizome, shoot and leaf with respect to each other (Petersen, 1889; Schumann, 1904; Holttum, 1950; Burtt and Smith, 1972; Larsen et al., 1998). But the classification based on morphological characters has been a daunting task in this family because defining characters are either not unique to any one particular taxa or are not universal for all the sub taxa clustered into one group. Some studies have been undertaken to understand the phylogeny of Zingiberaceae (Kress et al., 2002; Kress et al., 2004; Yong-Mei, 2004) but either they dealt with only one genus or there was absence of samples from northeast India which hosts a large number of Zingiberaceae members (Prakash and Mehrotra, 1995). The study involving phylogeny of Zingiberaceae from northeast India is specifically important because of the presence of geographical barrier between other south-east Asian species, especially China.

The objective in this study is to use molecular sequence data from plastid genome to test the past hypotheses on the phylogenetic relationships among the genera of the Zingiberaceae and make detailed observations to the established phylogenetic classification of the family.

4.2 Methodology

4.2.1 Specimen sampling

Out of the total fifty nine species recorded (as pointed in chapter 2) fourty six species were selected for carrying out the phylogeny study. The selection of these fourty six species was based on a single factor i.e. availability of green leaves for DNA extraction, since some of the species were only recorded for the ethno-botanical study and live specimens were not collected due to various reasons. While some of the species failed to grow in the glass house because the plant specimen, at the time of collection, had small rhizomes and once uprooted from their natural habitat they failed to grow, even in controlled environment. The list of the species used in the study is mentioned in Table 4.1. For the detailed analysis, fourty six species belonging to seven genera were categorised into three tribes viz. Alpinieae (Rich, 1841), Hedychieae (Horan, 1862) and Zingibereae (Meisn, 1842). The details of the species and there categorization as used in the study is as shown in the Table 4.1

4.2.2 DNA extraction, amplification and sequencing

DNA was extracted using DNeasy Plant Mini Kit (Qiagen) as per the instruction provided by the manufacturer. The DNA was amplified for eight plastid loci, five of them being from coding regions (*rbc*L, *rpo*C1, *rpo*B, *acc*D and *mat*K) and the other three from inter-genic region (*atp*F-*atp*H, *psb*K-*psb*I and *trn*H-*psb*A). Amplification of eight loci was done with the primers which have been developed by various research group and has been documented by Consortium for the Barcode of Life (CBOL: http://www.barcoding.si.edu/pdf/reactionconditionsusedincbolpnas-paper.pdf), the primer details is mentioned in Annexure A 4.1. The amplicons were ran on agarose gel and eluted from the gel using QIAquick Gel Extraction Kit (Qiagen) as per the instructions provided by the manufacturer. The sequencing job was outsourced to Macrogen inc. (South Korea) and sequenced on ABI 3730xl sequencer (Applied Biosystems) using the same primers which were used for amplifications. The sequencing was performed using both the forward and reverse primer in duplicates.

4.2.3 Sequence cleanup and contig generation

Contig assembly and the generation of consensus sequences were performed using Geneious Pro (5.6.7) (Drummond et al., 2006). The bases in sequence with the quality value (QV= Q = -10 log10 (Pe) where Pe is error probability of a particular base call) of less than 30 was only included if the duplicate sequence or the complementary read had QV more than 30 at that particular position. The sequences with the average QV value less than 50 were discarded.

4.2.4 Sequence analyses

The sequences were aligned using ClustalW (Larkin, 2007) with manual adjustments made by eye to improve the alignment. Numerous observations were made to deduce the informative content offered by each locus by looking at insertion-deletions and stop codon. The datasets were analysed using neighbor joining (NJ), maximum likelihood (ML) and maximum parsimony (MP) methods. For the NJ analyses, the pairwise distances were calculated using Kimura two parameter (K2P) nucleotide substitution model (Kimura, 1980) using MEGA5 (Tamura et al., 2011) platform. The K2P model evaluates distances for calculating the divergence of the species.

Divergence of i^{th} species = $\Sigma A_{ij} / \Sigma \Sigma A_{mn}$

ΣA_{ij} = Sum of species i from all other species.

ΣA_{mn} = Total divergence of all species from each other.

Subsequently the phylogenetic trees were constructed using neighbor joining method (NJ) using MEGA5 with the option of pairwise deletion.

Further the sequences were aligned using muscle (Edgar, 2004) which is a fast progressive alignment allowing re-optimizations of columns during the whole process and hence fairly corrects any errors made during preceding steps. To select the best fitting substitution model, jModelTest (http://darwin.uvigo.es/software/jmodeltest.html) was used and the best model was chosen according to the corrected Akaike information criterion (ACCc). Maximum likelihood analyses were done using MEGA 5.0. The Nearest-Neighbor-Interchange (NNI) heuristic approach was used for inferring tree, so as the branch-swapping performed on each tree resulting in a rearrangement into alternative trees so as to differ in only one branching pattern. Subsequently, to test the branch support 1000 non parametric bootstrap was performed.

Maximum parsimony was performed using MEGA5.0 with gaps treated as missing information. For tree inference "tree bisection-reconnection branch-swapping (TBR)" was performed with 1000 random addition and keeping the search level as one, with the option to retain trees as 10. To test the phylogeny, bootstrap was performed using 1000 reiterations.

For using the combination of locus, the two sequences were concatenated for combination so as to include one coding and one noncoding region (Table of combination tested in Annexure A 4.2), the concatenation was done using Geneious Pro (5.6.7) (Drummond et al., 2006). The phylogenetic inference was done as stated above for each individual genera, tribe and finally for the whole family.

Table 4.1. List of species used for phylogenetic study using the plastid loci. The genera has been classified into tribes based on Rich (1841), Horan (1862) and Zingibereae (Meisn, 1842).

Serial no	Species name	Unique Tag no	Collected from
		Tribe: Alpinieae	
1	***Alpinia***		
1.1	*A. bracteata*	ZSA 01	Darrang (Assam)
1.2	*A. galanga*	ZSA 02	Jalukbari (Assam)
1.3	*A. malaccensis*	ZSA 03	NEDFi
1.4	*A. nigra*	ZSA 04	Sivasagar (Assam)
1.5	*A. species*	ZSA 05	Amingaon (Assam)
1.6	*Alpinia sp*	ZSA 08	Seijusha (A.P.)
2	***Ammomum***		
2.1	*A. subulatum*	ZSB 02	Tizu (Nagaland)
2.2	*A. cardamom*	ZSB 03	Sonapur (Assam)
2.3	*A. aromaticum*	ZSB 04	Chandel (A.P.)
2.4	*A. linguiforme*	ZSB 05	Chandel (A.P.)
2.5	*A. zingiber*	ZSB 06	Chandel (A.P.)
3	***Elleteria***		
3.1	*E. big*	ZSE 01	NEDFi
3.2	*E. cardamomum*	ZSE 02	Sivasagar
		Tribe: Hedychieae	
1	***Curcuma***		
1.1	*C. amada*	ZSD 01	Kahikuchi (Assam)
1.2.	*C. angustifolia*	ZSD 02	Shillong (Meghalaya)
1.3	*C. aromatica*	ZSD 03	Kamrup (Assam)
1.4	*C. caesia*	ZSD 04	Goalpara (Assam)
1.5	*C. longa*	ZSD 07	Kamrup (Assam)
1.6	*C. leucorrhiza*	ZSD 08	Amingaon (Assam)

1.7	*C. wild*	ZSD 09	Kokrajhar (Assam)
1.8	*C. zedoaria*	ZSD 10	Darrang (Assam)
2	***Hedychium***		
2.1	*H. coronarium*	ZMN 09	Pokpi (Manipur)
2.2	*H. dekianium*	ZSF 02	Barpeta (Assam)
2.3	*H. flavum*	ZSF 03	Barpeta (Assam)
2.4	*H. stenopetalum*	ZSF 04	Guwahati (Assam)
2.5	*H. spicatum*	ZSF 05	Panbazar (Assam)
2.6	*H. (AP)*	ZSF 06	Pakke (A.P)
2.7	*H. chrysoleucum*	ZSF 07	Margherita (Assam)
2.8	*H. bracteata*	ZSF 08	NEDFi
2.9	*H. gardnerianum*	ZAP 08	Bomdila (Arunachal Pradesh)
2.10	*H. forrestii*	ZAP 15	Seppa
2.11	*H. flavescens*	ZAP 16	Ziro (Arunachal Pradesh)
2.12	*H. coccinium*	ZMN 06	Chandel
2.13	*H. aurantiacum*	ZMN12	Lamkangkhunou (Manipur)
3	***Kaempferia***		
3.1	*K. galanga*	ZSG 01	Khetri (Assam)
3.2	*K. angustifolia*	ZSG 02	Jalukbari (Assam)
3.3	*K. pulchera*	ZSG 03	Jalukbari (Assam)
3.4	*K. rotunda*	ZSG 04	NEDFi
3.5	K. ornamental IV	ZSG 05	Goalpara (Assam)
3.6	K. ornamental I	ZSG 06	Sonapur (Assam)
3.7	K. ornamental II	ZSG 07	Sonapur (Assam)
3.8	K. ornamental III	ZSG 08	Sonapur (Assam)
Tribe: Zingibereae			
1	***Zingiber***		
1.1	*Z. casuamonar*	ZSH 01	Rongia (Assam)

1.2	*Z. moran*	ZSH 02	Sivasagar (Assam)
1.3	*Z. officinale*	ZSH 03	Khliehriat (Mizoram)
1.4	*Z. zerumbet*	ZSH 04	Chemphui (Mizoram)
1.5	*Z. species*	ZSH 05	Darrang (Assam)
1.6	*Z. species*	ZSH 06	Zowai (Mizoram)

4.3 Results and Discussion

4.3.1 Alpinieae

4.3.1.1 Sequence analysis

The species from the genus *Alpinia, Amomum* and *Elleteria* were combined into one group since the classical taxonomy groups them into one tribe: Alpinieae (Rich, 1841). The DNA extraction and PCR amplification did not pose any problems except for the amplification involving the primer "*psbA-trnH*" which required many rounds of optimization. The sequencing was mostly good in quality and did not require re-sequencing. Alignment of the coding loci was fairly easy and was performed using the auto features for ClustalW alignment, as implemented by MEGA 5.0, while the alignment of the non-coding regions required manual adjustments to improve the alignment. The genic regions were observed to be fairly conserved with low level of variation (1.9-4.0 %) and even lower parsimony informative content (PIC) (1.00-3.25%). Inter-genic regions were fairly diverse (7.4-11.3%), but the 5 prime and 3 prime were relatively conserved as the primer were designed from the genic region while the intervening DNA showed high variation and had large insertion-deletions in this region. The PIC value was observed to be high (4.2-7.9 %) for the non-coding regions (intervening DNA). Out of all the loci, *matK* was found to stand out as the PIC was fairly high (4.3%) and the alignment was easy to perform. To estimate the discriminatory power of the various loci, the overall mean distance was calculated using K2 parameter (Figure 4.1). *PsbK-psbI* was found to possess highest discriminatory power but the alignment also required manual attention. The loci *atpF-atpH* and *matK* were found to be next in line for discriminatory power and appeared as much better choice since they were observed to be highly informative and the alignment was relatively easy requiring very little manual attention.

4.3.1.2 Neighbor Joining

The reconstruction of the tree based on NJ using single loci failed to resolve the branches and a high level of polytomy was observed. The phylogeny constructed using coding and non-coding regions were similar but the resolution was very poor in case of phylogeny constructed based on information from coding region. The tribe Alpinieae was broadly divided into two major cluster with *Alpinia* and *Amomum* distributed in both the cluster. *Alpinia malaccensis* and *A. roxburghii* clustered with *Eletteria* species (*E. cardamomum* and *Eletteria* unidentified species ZSE01). The evolutionary history clearly establishes the polyphyletic nature of the *Alpinia* and *Amomum*. Among all the eight single loci, the phylogeny drawn based on *matK* and *atpF-atpH* was found to provide high resolution and the two phylogeny drawn broadly agreed with each other (Figure 4.2). *Elettaria* clustered closely with *Amomum* this is in agreement with Fischer (1956) who in his classification, of the family Zingiberaceae, has indicated similarity of *Elettaria* to *Amomum* based on the absence of lateral staminodes (which are small and narrow or absent) and to *Alpinia* based on the inflorescence type.

In order to check if the polytomy observed is a true polytomy, i.e. hard polytomy, or were result of lower information content, i.e. soft poytomy, combinations of loci were used for phylogenetic relationship assessment. The combinations were tried for a maximum of two sequences, with one coding and another inter-genic region. The basis of selection, of this combination strategy, was to reduce manual effort for alignment. Since *matK* is type II intron and

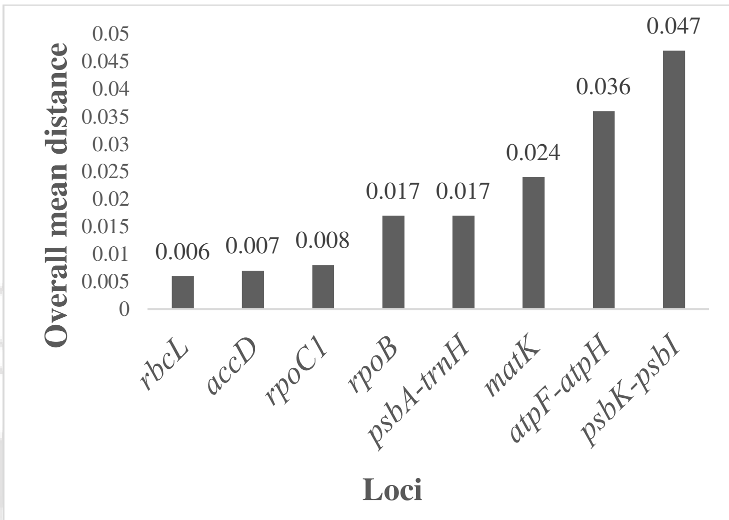


Figure 4.1. Mean distance calculated using K2 parameter between species belonging to the tribe Alpinieae as a measure of the discriminatory power.

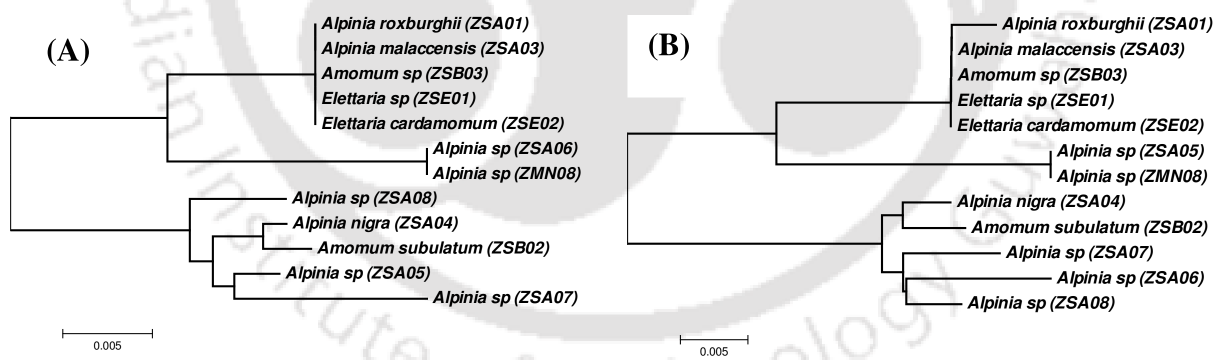


Figure 4.2. The evolutionary history for the tribe Alpinieae, reconstructed using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (A) *matK* (B) *atpF-atpH*.

involved in splicing, hence by definition is a gene therefore its combination were also tested with other inter-genic regions. A total of nineteen combinations were also tested (Annexure A 4.2) and phylogenetic tree was constructed and tested for their ability to resolve the phylogeny. Among the nineteen combinations, the pairing of *rbcL* with *atpF-atpH* and *matK* showed a higher level of resolution. Apart from the above mentioned combinations, *matK* and *atpF-atpH* produced highly resolved phyogeny (Figure 4.3). The tree constructed using all the combinations were observed invariably to have two clusters, and the members in each cluster was fixed except the case of *Alpinia* species (ZSA05 and ZSA06) which were found to interchange their position in the tree. The net divergence between the two cluster was found to be 0.016 (number of base substitutions per site from estimation of net average between the two cluster).

The first cluster consisted of *Alpinia roxburghii, Alpinia malaccensis, Elettaria* species*, E. cardamomum, Amomum* species and *Alpinia* species (ZSA05 and ZMN08). All of the seven member clustered closely and it was most of the time difficult to discriminate with the exception of two *Alpinia* species (ZSA05 and ZMN08). The loci combination of *rbcL* and *atpf-atpH* was most successful in resolving the relationship. Species belonging to the genera *Elettaria* was observed to be present in one cluster, but since the number of species were low from the genera therefore it does not conclusively proves the monophyletic nature of the genera. Previous studies on phylogeny of *Elettaria* has not been able to emphatically prove about the nature of this taxa, due to either lack of species from other taxa or lack of enough no of species (Babu et al., 2012). Although a master's thesis result mentioned on University of Oslo website (http://www.mn.uio.no-/ibv/studier/masteroppgaver/okologi-og-evolusjon/poulsen1.html) indicates that *Elettaria* may not be monophyletic. The second cluster consisted of *Alpinia* species (ZSA06, ZSA07 and ZSA08), *Alpinia nigra* and *Amomum subulatum*. *Amomum* is the second largest genera belonging to Zingiberaceae with around 170 species and is believed to be polyphyletic (Kress et al., 2004; Lamxay, 2011,). *Alpinia* registered a fair presence in both the cluster, reiterating its polyphyly (Kress et al., 2004).

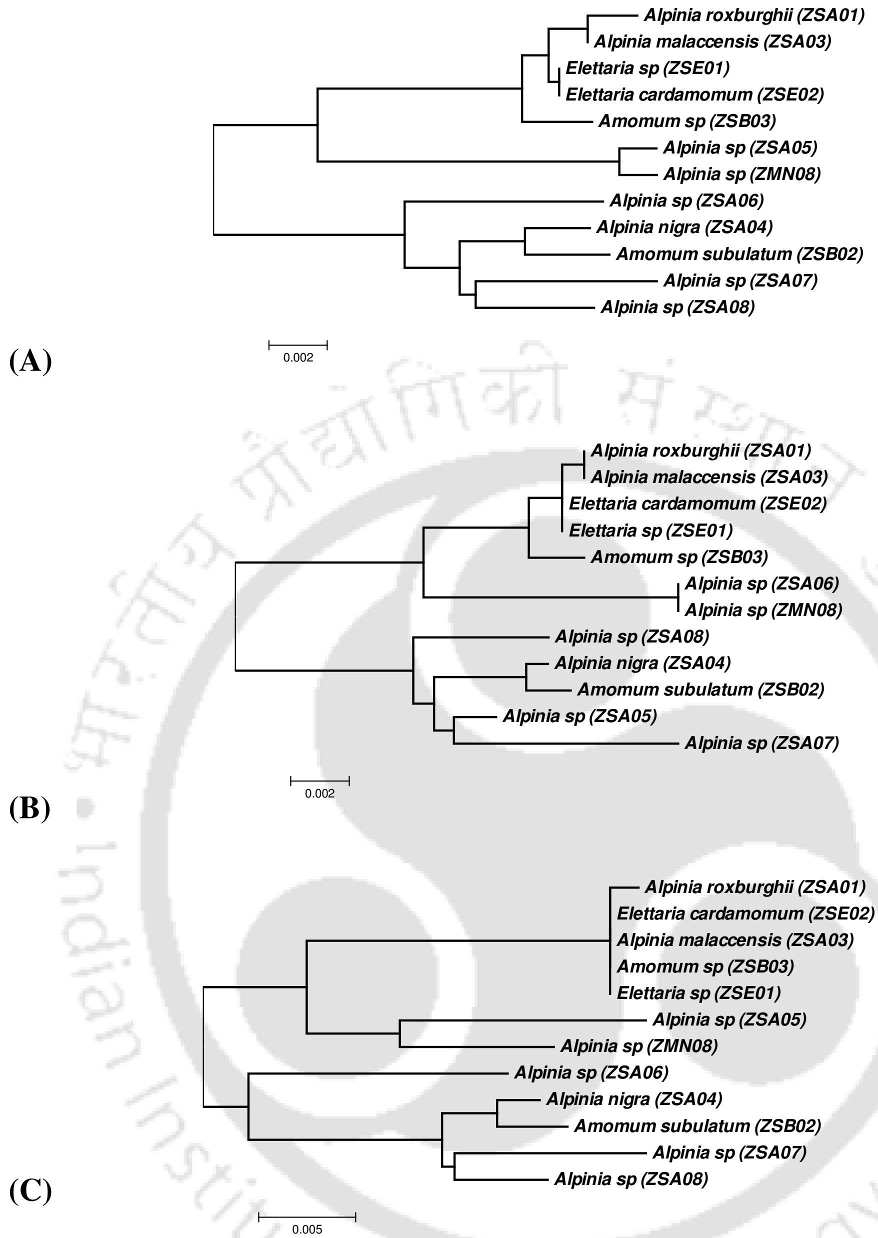


Figure 4.3. The evolutionary history reconstructed for the tribe Alpinieae, using combination of loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (A) *rbcL* + *atpf-atpH* (B) *rbcL* + *matK* (C) *matK* + *atpf-atpH*.

4.3.1.3 Maximum likelihood

The construction of phylogenetic tree using maximum likelihood involves finding a tree which would have the likelihood of observing the DNA sequence under a particular model of sequence evolution. To test for the best suited model for DNA substitution applicable to the data set, jModelTest (Darriba et al., 2012) was used and the suitable model were chosen according to the corrected Akaike information criterion (Posada and Buckley, 2004). Maximum-likelihood were reconstructed using MEGA5.0 software and by using the most suitable model (Table 4.2) and the γ parameter as estimated in the analyses.

The resulting ML trees were found to be in congruence with the topology of tree reconstructed using NJ method, with two major clades (Figure 4.4). As was the case with phylogeny drawn using NJ method, the phylogeny reconstructed based on information from single locus using ML method failed to resolve phylogeny while the combination of loci improved the phylogenetic resolution. The best phylogeny was observed by the same combination of loci as was with NJ method i.e. *rbcL+atpf-atpH*, *rbcL+matK* and *matK+atpf-atpH*. The percentage of trees in which the associated taxa clustered together is shown next to the node. All the nodes were supported by large percentage of tree agreeing with the node as constructed in final tree(s) (Figure 4.4).

4.3.1.4 Maximum parsimony

Maximum parsimony tree reflects the ancestral relationship and uses all known evolutionary information and are computationally known to be faster to calculate than ML trees. The calculation of MP tree involves computing the minimum number of evolutionary changes for each topology that explains the entire sequence evolution i.e. tree length (TL). Among the number of topology examined, the topology with the smallest TL value is chosen as the preferred tree (Takahashi and Nei, 2000). Among the various locus and their combinations used the best locus for tree construction was not chosen based on the smallest tree length as the optimality scores of the MP

Table 4.2. The best substitution model for the tribe Alpinieae, as predicted by jModelTest along with associated parameter for individual locus and combination of loci.

Particulars / Loci	Substitution model (AICc)	Proportion invariable sites	Gamma shape parameter
accD	HKY+I	0.8730	-
atpF-atpH	GTR+G	-	0.3120
matK	HKY	-	-
psbA-trnH	HKY+G	-	0.1280
psbK-psbI	GTR	-	-
rbcL	HKY	-	-
rpoB	HKY	-	-
rpoC1	HKY	-	-
accD+atpF-atpH	GTR+I+G	0.7780	0.7040
accd+matK	HKY+G	-	0.4020
accd+psbA-trnH	HKY+I	0.7650	-
accD+psbK-psbI	GTR+I	0.6870	-
matK+atpF-atpH	GTR+I+G	0.4180	0.0180
matK+psbA-trnH	GTR+G	-	0.0280
matK+psbK-psbI	HKY+G	-	0.1590
rbcL+atpF-atpH	GTR+I+G	0.7660	0.7750
rbcL+matk	HKY+G	-	0.1500
rbcL+psbA-trnH	GTR+I	0.8640	-
rbcL+psbK-psbI	HKY+I	0.7160	-
rpoB+atpF-atpH	GTR+I	0.8430	-
rpoB+matK	HKY+G	-	0.6190
rpoB+psbA-trnH	HKY+I	0.7840	-
rpoB+psbK-psbI	GTR+I	0.5030	-
rpoC1+atpF-atpH	GTR+G	-	0.0210
rpoC1+matK	HKY	-	-
rpoC1+psbA-trnH	HKY+I+G	0.7100	0.3620
rpoC1+psbK-psbI	HKY+G	-	0.0240

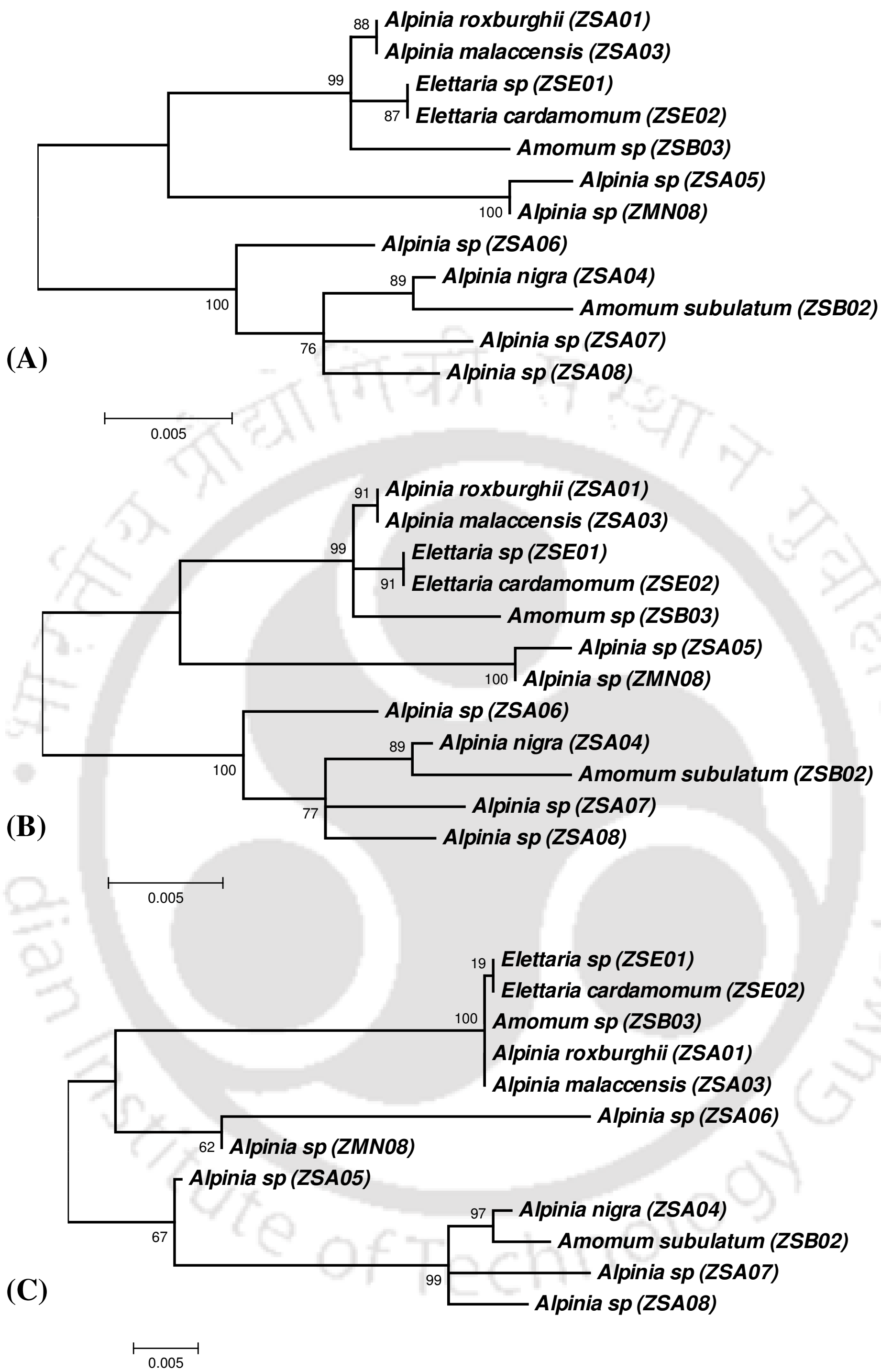


Figure 4.4. Phylogenetic tree constructed for the tribe Alpinieae, using the Maximum Likelihood method based on the different substitution model as predicted using jModeTest (Table 4.2). The tree with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (A) *rbcL+atpF-atpH* (B) *rbcL+matK* (C) *matK+atpf-atpH*.

trees are always smaller than or equal to those of the true tree in case when the number of nucleotides examined is small in a dataset but the number of sequences are large (Nei et al., 1998) (Table 4.3). The MP tree based on the locus combination which showed promising results in the previous analyses are shown in Figure 4.5. The shortest tree produced form the dataset *rbcL+atpF-atpH* with Consistency Index (CI) = 0.79 and Retention Index (RI) = 0.89, *rbcL+matK* with CI = 0.90 and RI = 0.96 and *matK+atpf-atpH* with CI = 0.71 and RI = 0.84. The combination *rbcL+matK* was found to have the shortest TL among the three combinations with length equal to 75. The ML tree was found to be in complete congruence with the topology of NJ and ML tree with the exception of exchange of *Alpinia* species ZSA06 and ZSA05 position.

Table 4.3. The characteristics of tree resulting from Maximum Parsimony analyses for single and combined locus of the tribe Alpinieae. CI and RI calculated from informative characters.

Gene	CI	RI	Number of trees	Tree length
accD	0.857143	0.909091	12	13
atpF-atpH	0.913043	0.956522	8	64
matK	0.948718	0.980769	9	57
psbA-trnH	0.931034	0.962963	9	60
psbK-psbI	0.939394	0.973333	5	61
rbcL	1.00000	1.000000	11	11
rpoB	0.928571	0.971429	11	17
rpoC1	1.000000	1.000000	12	11
accD+atpF-atpH	0.787879	0.867925	8	90
accd+matK	0.937500	0.974576	5	73
accd+psbA-trnH	0.880952	0.929577	6	65
accD+psbK-psbI	0.883721	0.941860	2	76
matK+atpF-atpH	0.714286	0.838095	5	153
matK+psbA-trnH	0.804878	0.900000	7	127
matK+psbK-psbI	0.824176	0.913514	6	134
rbcL+atpF-atpH	0.794118	0.888889	3	91
rbcL+matk	0.903846	0.960317	6	75
rbcL+psbA-trnH	0.869565	0.926829	8	68
rbcL+psbK-psbI	0.875000	0.938144	2	78
rpoB+atpF-atpH	0.769231	0.879195	5	99
rpoB+matK	0.943396	0.978417	9	75
rpoB+psbA-trnH	0.843137	0.914894	5	75
rpoB+psbK-psbI	0.888889	0.947368	3	84
rpoC1+atpF-atpH	0.833333	0.920000	9	82
rpoC1+matK	0.956522	0.984496	8	69
rpoC1+psbA-trnH	0.800000	0.891566	4	82
rpoC1+psbK-psbI	0.812500	0.910000	7	79

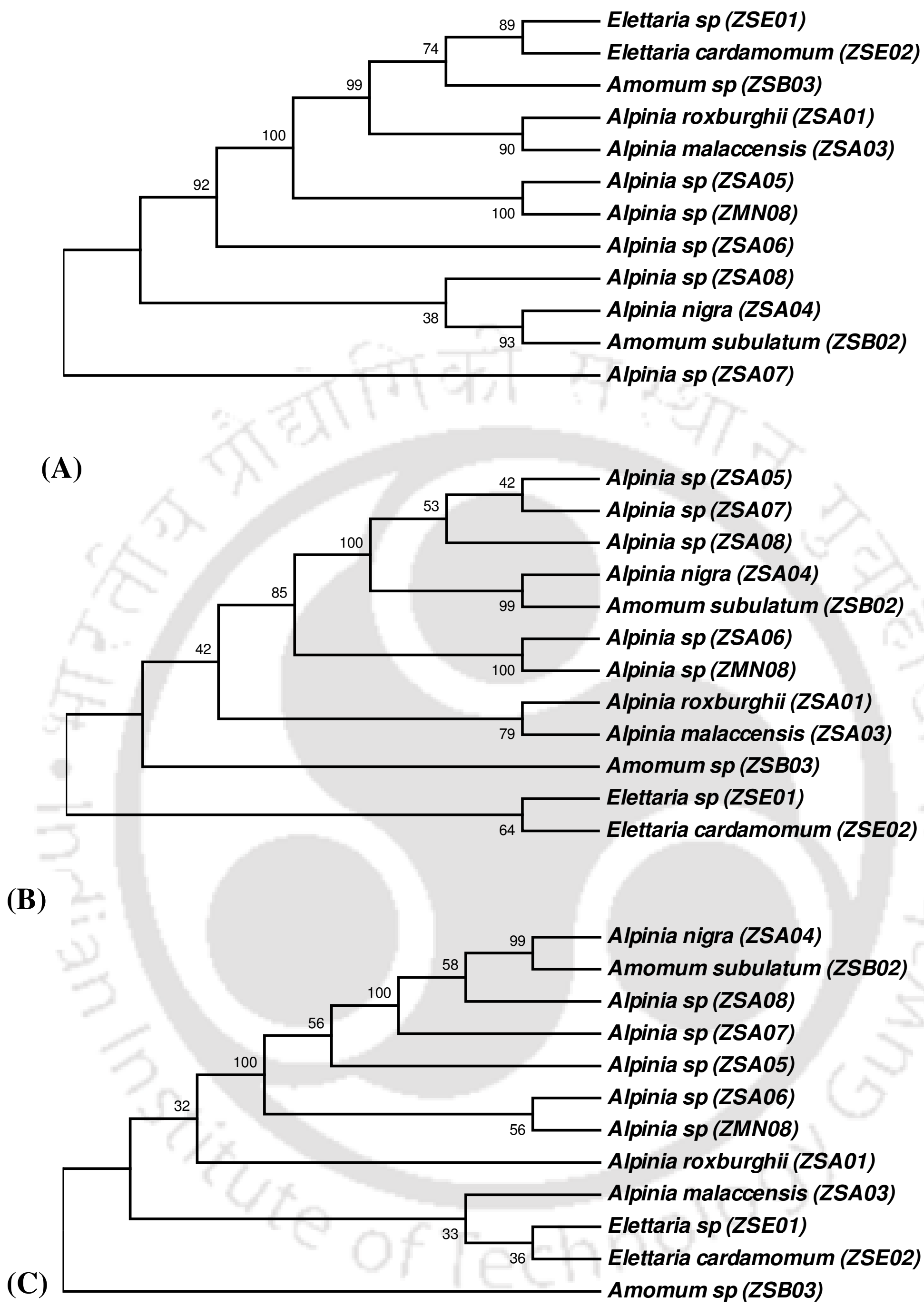


Figure 4.5. The evolutionary history for the tribe Alpinieae, inferred using the Maximum Parsimony method. The most parsimonious trees is shown with their consistency index and retention index mentioned in Table 4.3. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10000 replicates). (A) *rbcL+atpF-atpH* (B) *rbcL+matK* (C) *matK+atpf-atpH*.

4.3.2 Hedychieaea

4.3.2.1 Sequence Analysis

The species from the genus *Curcuma*, *Hedychium* and *Kaempferia* were combined into one group since the classical taxonomy groups them into one tribe: Hedychieae (Horan, 1862). The DNA extraction and PCR amplification did not pose any problems except for the amplification involving the primer *"psbA-trnH"* which required optimization for each genera. The sequencing was mostly good in quality and did not require re-sequencing. Alignment of the coding loci was fairly easy and was performed using the auto features for ClustalW and muscle alignment, as implemented by MEGA 5.0, whereas the alignment of the non-coding regions required manual adjustments by eye for improving the alignment. The genic regions were observed to show a large range of variation (3.9-17.1%) and PIC (1.8-13.57%). The locus *rbcL* was found to be least variable while *matK* was found to be the most variable loci among the coding loci. Same pattern was observed for PIC value. Inter-genic regions were found to be fairly diverse (9.2-20.0%) with their 5' and 3' relatively conserved as the primer were designed from the genic region while the intervening DNA showed high variation with large insertion-deletions. The PIC value was observed to be high (3.3-11.32%) for the non-coding regions. Out of all the loci, *matK* was found to stand out as the PIC was fairly high (17.1%) and the alignment was easy to perform. To estimate the discriminatory power of the various loci, overall mean distance was calculated using K2 parameter (Figure 4.6). *PsbK-psbI* was found to possess highest discriminatory power but the alignment also required manual attention. The loci *atpF-atpH* and *matK* were found to be next in line for discriminatory power and appeared as much better choice since they were observed to be highly informative and the alignment was relatively easy requiring very little manual attention.

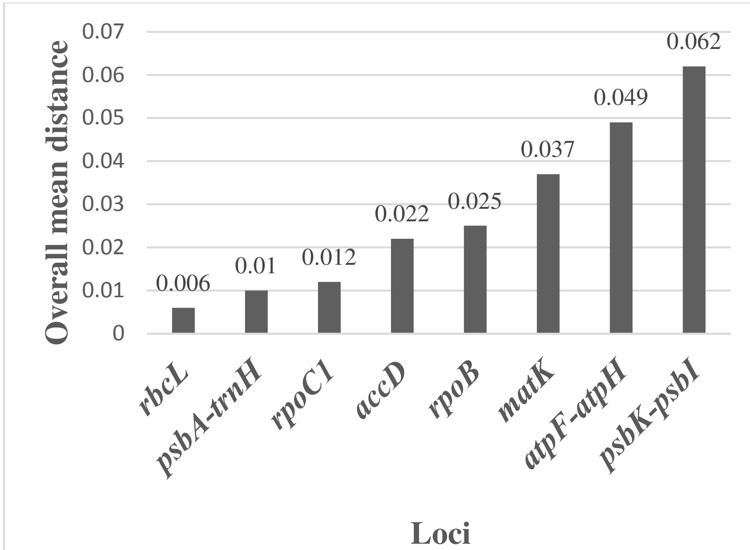


Figure 4.6. Mean distance calculated using K2 parameter between species belonging to the tribe Hedychieae as a measure of the discriminatory power.

4.3.2.2 Neighbor joining

The reconstruction of tree based on NJ using single loci established two major clusters, the smaller clustered mostly consisted of four unidentified *Kaempferia* species while the bigger cluster invariably consisted of three sub cluster which mostly distributed the rest of Hedychieae according to their genera. Hence in most of the trees based on single locus the member belonging to a genus were found to be clustered together. The phylogeny constructed using coding and non-coding regions were similar. The *Hedychium* species were found to cluster closely with the exception of *Hedychium* species ZSF06 which was found to be scattered in the phylogeny constructed using different loci. The resolution for *Hedychium* species was found low with most of the species clustering on a single branch (polytomy) both with phylogeny constructed using information from coding and non-coding loci. *Curcuma* was found to cluster mostly closely, though *Curcuma caesia* was found to be scattered in the phylogeny constructed using different loci. *Kaempferia* species found in the larger cluster were found to occur compactly on the tree and the resolution among the

Kaempferia species was fairly high as comapred with other subcluster belonging to this large cluster. Among the all eight single loci, the phylogeny drawn based on *matK* and *atpF-apH* was found to provide high resolution and the two phylogeny drawn broadly agreed with each other (Figure 4.7).

Since the single loci failed to resolve most of the phylogeny, combination of loci were used for phylogenetic relationship assessment. The choice of combination of loci was made as described in the section 4.3.1.1 and the list of combination is described in Annexure A 4.2. Among the combination of loci tested, the *rpoB* with *matK* and *accD* with *psbK-psbI* combinations showed a higher level of resolution (Figure 4.8). The tree constructed using all the combinations were predominately observed to have two clusters, with the species falling consistently in the same cluster. The net divergence between the two cluster was found to be 0.08 (number of the base substitutions per site from estimation of net average between the two cluster). The loci combination of *rpoB* + *matk* was most successful in resolving the relationship.

The phylogeny, as mentioned earlier, consisted of two cluster with one of them being the big cluster with all the *Curcuma* and *Hedychium* species and four *Kaempferia* species *(K. angustifolia, K. pulchra, K. galanga and K. rotunda*), while the smaller cluster just consisted of 4 unidentified species of *Kaempferia*. In the larger cluster the species belonging to the same genus clustered consistently as a distinct cluster indicating the monophyletic nature of the genera. *Hedychium* has been earlier reported to be monophyletic by Wood et al., (2000). *Curcuma* appears as monophyletic in the current data set assembled as per classification of the tribe Hedychieae but has been reported as not being a polyphyletic group (Kress et al., 2002; Ngamriabsakul et al., 2003; Zaveska et al., 2012). The reason for this discrepancy could be best explained by the influence of sampling of species on phylogenetic relationship assessment (Cori and Ellegren, 2013). The phylogenetic tree construction by including species from other Zingiberaceae tribe would help in better understanding about the phylogenetic nature of the genus *Curcuma*. *Hedychium* is a known monophyletic genus as was the case with the phylogeny constructed in the current study with an exception of *Hedychium* species ZSF06. Further in all the previous phylogenetic studies involving *Kaempferia*, it was observed that this is a true clade (i.e. monophyletic taxa) though a study published in 2010 by Techaprasan et al., showed that this taxa is polyphyletic. The study involved twenty five species (identified, new and unidentified) and used *psbA-trnH* and *petA-psbJ* as the data source for phylogenetic and found that one species *K. candida* clustered outside the *Kaempferia* cluster. Interestingly a report published in 2011 (Techaprasan and Skornickova) has proposed that the *K. candida* should be included in the *Curcuma* clad. Conclusively based on these studies, the monophyletic nature of the genus *Kaempferia* is established. If one consider only the large cluster in this current study, the monophyletic nature of the genus is reiterated but the presence of four unidentified *Kaempferia* species which are present on the smaller cluster raises lots of questions. The true reason(s) of this deviation from the established fact could only be established in the section where the phylogenetic relationship of the entire Zingiberaceae is undertaken.

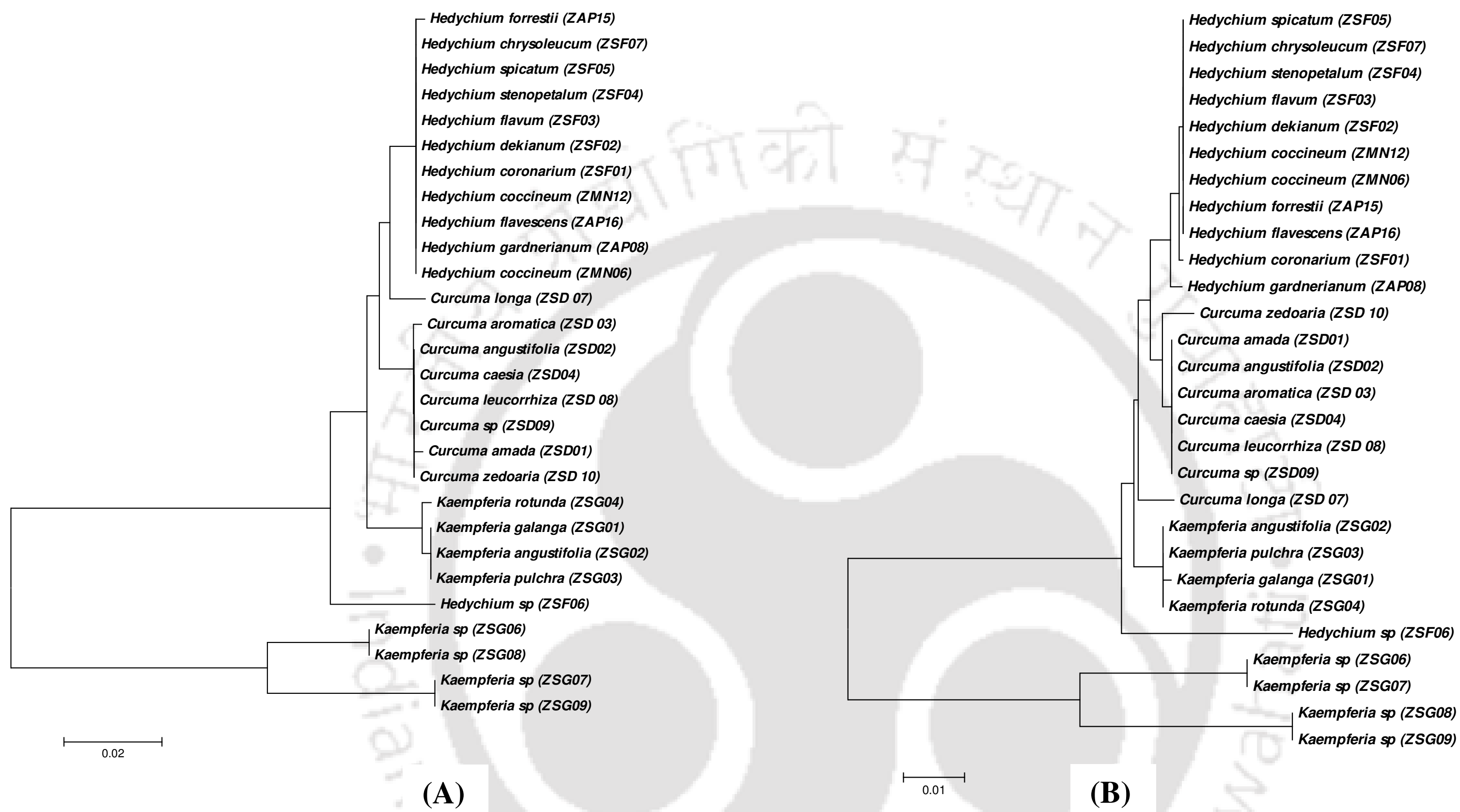


Figure 4.7. The evolutionary history reconstructed for the tribe Hedychieae, using single loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (A) *atpF-atpH* (B) *matK*.

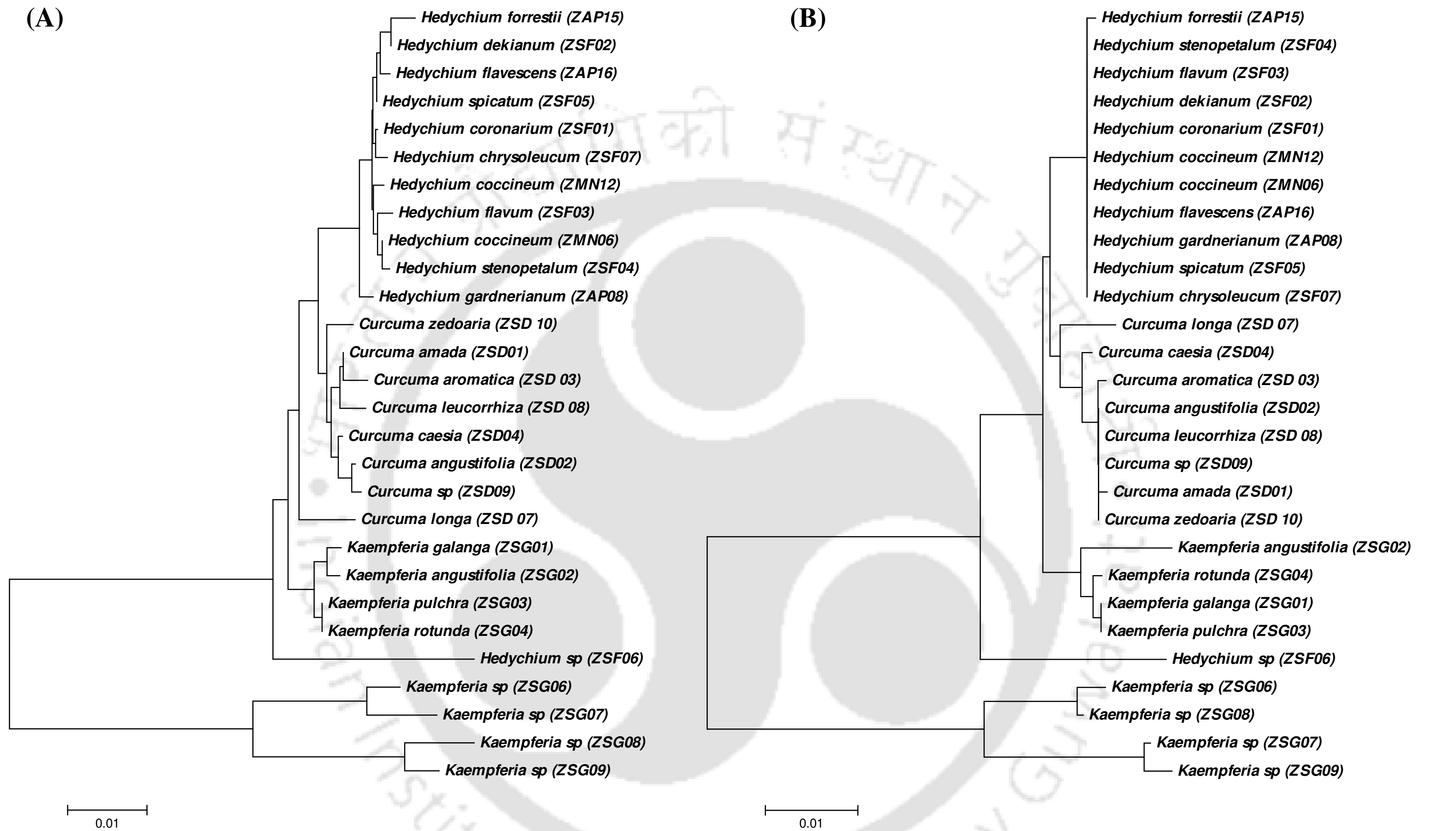


Figure 4.8. The evolutionary history reconstructed for the tribe Hedychieae, using combination of loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree NJ (A) *rpoB* + *matK* (B) *rbcL* + *atpF*.

4.3.2.3 Maximum likelihood

The construction of phylogenetic tree using maximum likelihood was performed as mentioned earlier in section 4.3.1.3. Maximum-likelihood phylogenies were reconstructed by using MEGA5.0 software and by using the most suitable model (Table 4.4) and the γ parameter as estimated in the analyses.

The ML trees constructed based on the model and parameter mentioned in Table 4.4 were found to be in congruence with the topology of tree reconstructed using NJ method. As was the case with the phylogeny drawn using NJ method, the phylogeny reconstructed based on information from single locus using ML method failed to produce trees with fair resolution, whereas the combination of loci was observed to definitely improve the resolution. A good resolved phylogeny was obtained with the combination of *rpoB* + *matK* and *rbcL*+*atpF-atpH* (Figure 4.9), the same combination of loci as was with NJ method. The *rpoB* + *matK* ML tree was observed to have higher resolution and stronger support for branches. The two cluster was well supported with 100% of trees with the smaller cluster consisting of all the unidentified *Kaempferia* species. The cluster of *Hedychium-Curcuma* and *Kaempferia* was well supported by 55% and 67% of trees. Interestingly the one *Hedychium* species (ZSF06) which was represented outside the *Hedychium* cluster was found to have 100% of tree support. *Kaempferia* clad in the large cluster was divided into two cluster with 80% of tree support and established *K. galanga* and *K. angustifolia* to be more closely related whereas *K. pulchra* and *K. rotunda* were found to be more closely related.

Table 4.4. The best substitution model for the tribe Hedychieae as predicted by jModelTest along with associated parameter for individual locus and combinations of loci.

Particulars / Loci	Substitution model (AICc)	Proportion invariable sites	Gamma shape parameter
accD	HKY+I	0.558	-
atpF-atpH	GTR+I	0.400	-
matK	GTR+G	-	0.593
psbA-trnH	HKY+I	0.633	-
psbK-psbI	GTR+I	0.431	-
rbcL	GTR+I	0.808	-
rpoB	HKY	-	-
rpoC1	HKY	-	-
accD+atpF-atpH	GTR+I	0.618	-
accd+matK	HKY+I	0.673	-
accd+psbA-trnH	HKY+G	-	0.268
accD+psbK-psbI	HKY+G	-	0.234
matK+atpF-atpH	GTR+I+G	0.402	0.705
matK+psbA-trnH	GTR+G	-	0.336
matK+psbK-psbI	GTR+G	-	0.470
rbcL+atpF-atpH	GTR+I+G	0.572	0.981
rbcL+matk	GTR+I+G	0.542	0.818
rbcL+psbA-trnH	HKY+I	0.842	-
rbcL+psbK-psbI	GTR+G	-	0.110
rpoB+atpF-atpH	GTR+G	-	0.257
rpoB+matK	GTR+I+G	0.454	0.809
rpoB+psbA-trnH	HKY+G	-	0.264
rpoB+psbK-psbI	HKY+I	0.614	-
rpoC1+atpF-atpH	GTR+I	0.699	-
rpoC1+matK	GTR+G	-	0.239
rpoC1+psbA-trnH	HKY+I	0.607	-
rpoC1+psbK-psbI	HKY+G	-	0.129

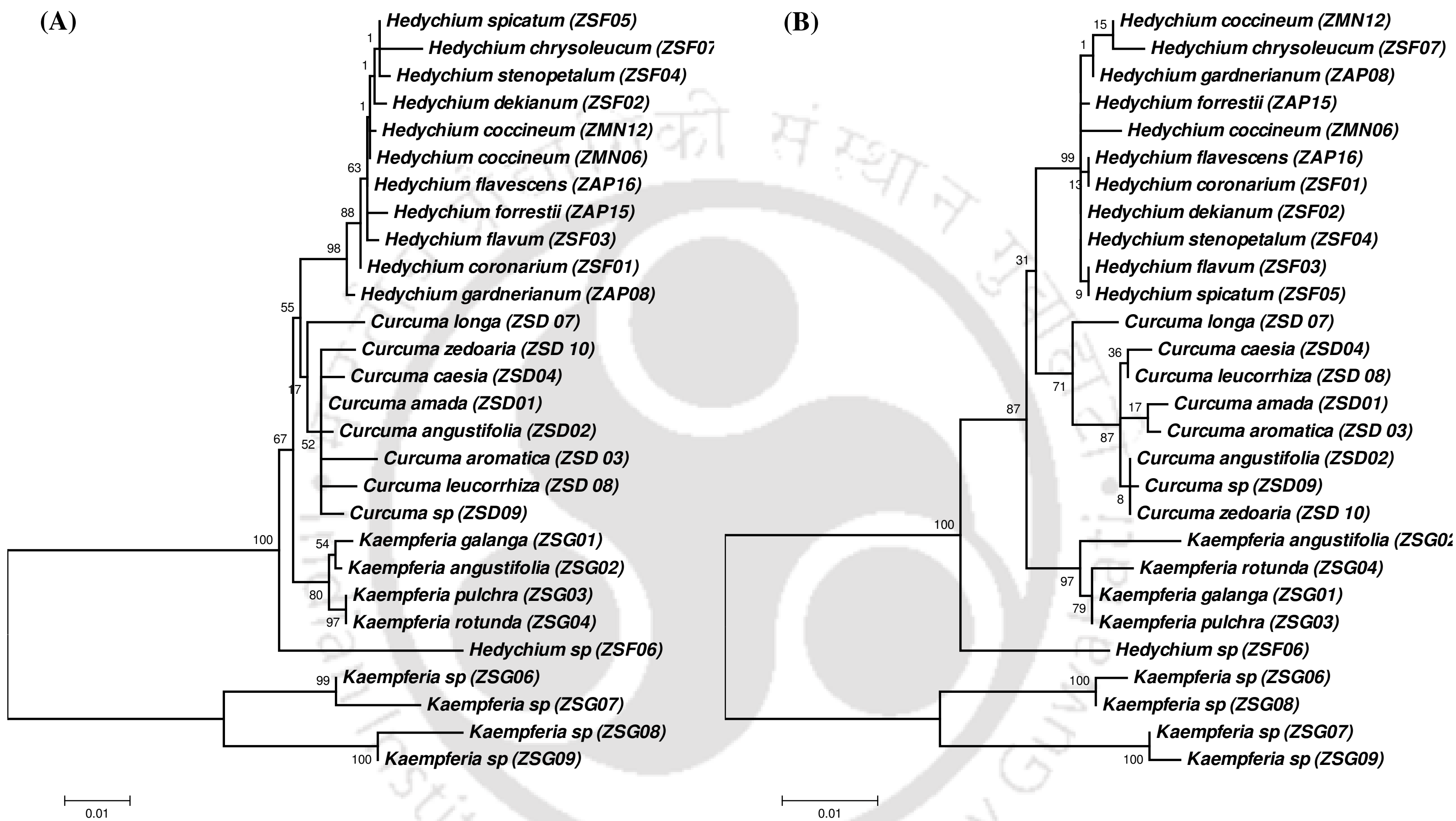


Figure 4.9. Phylogenetic tree constructed using the Maximum Likelihood method based on the different substitution model as predicted using jModeTest (Table 4.4). The tree with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (A) *rpoB* + *matK* (B) *rbcL* + *atpf-atpH*.

4.3.2.4 Maximum Parsimony

The MP trees based on the loci combinations which showed promising result in the precious analyses (NJ and ML) are shown in Figure 4.10. The shortest tree produced from the dataset *rpoB* + *matK* with CI = 0.79 and RI = 0.90 and *rbcL* + *atpF-atpH* with CI = 0.85 and RI = 0.93 (Table 4.5). The MP tree reconstructed using *rbcL* + *atpF-atpH* and *rpoB* + *matK* clubbed the unidentified *Kaempferia* species along with other identified *Kaempferia* species, this could be attributed to the long branch attraction which is very high and prominent in MP analyses and tree reconstruction.

4.3.3 Zingibereae

4.3.3.1 Sequence Analysis

In the current study the taxa Zingibereae consisted of species only from *Zingiber*. The DNA extraction and PCR amplification did not pose any problems except for the amplification involving the primer "psbA-trnH" which required optimization for each genera. The sequencing was mostly good in quality and did not required re-sequencing. Alignment of the coding loci was fairly easy and was performed using the auto features for ClustalW and muscle alignment, as implemented by MEGA 5.0, whereas the alignment of the non-coding regions required manual adjustments by eye for improving the alignment. The genic regions were observed to show a low range of variation (0.19 - 2.22%) and much lower PIC (0.00 - 0.23%). The locus *rpoC1* was found to be least variable while *rbcL* was found to be the most variable loci among the coding loci. Inter-genic regions were found to be more diverse (1.56 - 3.25%) with their 5 prime and 3 prime relatively conserved as the primer were designed from the genic region while the intervening DNA showed high variation with large insertion-deletions. The PIC value was observed to be in the range of 0.13 - 1.41% for the non-coding regions. The discriminatory power of the various loci were estimated by calculating the overall mean distance using K2 parameter (Figure 4.11). Loci *psbA-trnH* was found to possess highest discriminatory power followed by *atpF-atpH* and *psbK-psbI*.

Table 4.5. The characteristics of tree for the tribe Hedychieae resulting from maximum parsimony analyses single and combined locus. CI and RI calculated from informative characters.

Gene	CI	RI	Number of trees	Tree length
accD	0.947368	0.974684	26	57
atpF-atpH	0.948905	0.976898	14	157
matK	0.905109	0.959375	24	167
psbA-trnH	0.857143	0.958763	14	73
psbK-psbI	0.912088	0.964912	20	145
rbcL	0.833333	0.954545	8	24
rpoB	0.900000	0.960396	20	65
rpoC1	1.000000	1.000000	25	37
accD+atpF-atpH	0.871795	0.935401	14	243
accd+matK	0.822115	0.912322	11	262
accd+psbA-trnH	0.837500	0.933673	14	153
accD+psbK-psbI	0.782609	0.884488	2	253
matK+atpF-atpH	0.792049	0.896499	21	387
matK+psbA-trnH	0.858696	0.941176	10	260
matK+psbK-psbI	0.85654	0.936449	14	343
rbcL+atpF-atpH	0.852071	0.928977	9	215
rbcL+matk	0.832298	0.925824	10	209
rbcL+psbA-trnH	0.692308	0.892617	4	110
rbcL+psbK-psbI	0.786408	0.911647	4	184
rpoB+atpF-atpH	0.811321	0.904077	9	284
rpoB+matK	0.794393	0.901124	14	281
rpoB+psbA-trnH	0.812500	0.932432	17	154
rpoB+psbK-psbI	0.789809	0.899083	15	258
rpoC1+atpF-atpH	0.911765	0.958678	24	209
rpoC1+matK	0.878613	0.944882	22	222
rpoC1+psbA-trnH	0.921569	0.972414	7	110
rpoC1+psbK-psbI	0.842520	0.930314	9	205

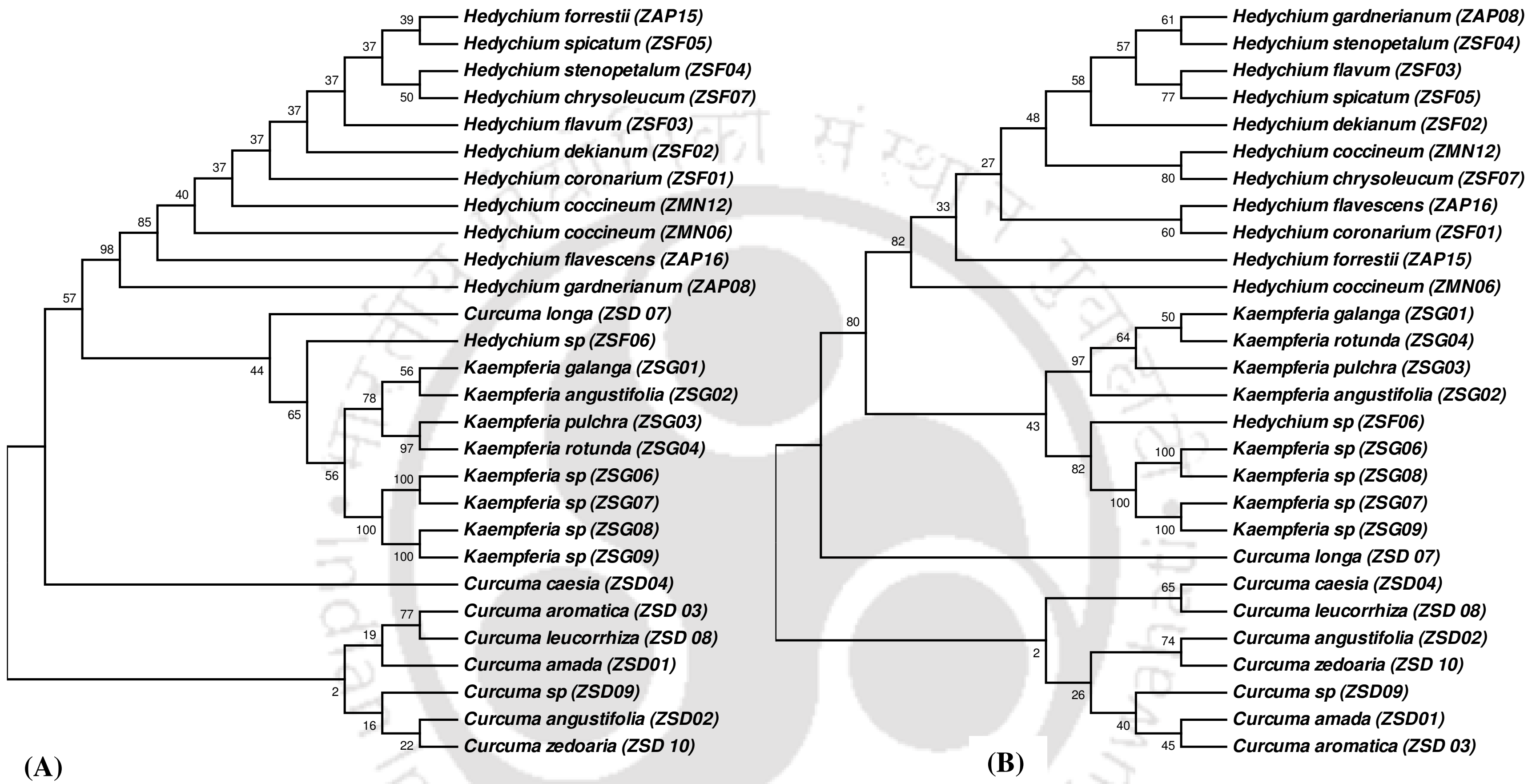


Figure 4.10. The evolutionary history was inferred for the tribe Hedychieae using Maximum Parsimony method. The most parsimonious trees is shown with their consistency index and retention index mentioned in Table 4.5. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10000 replicates).

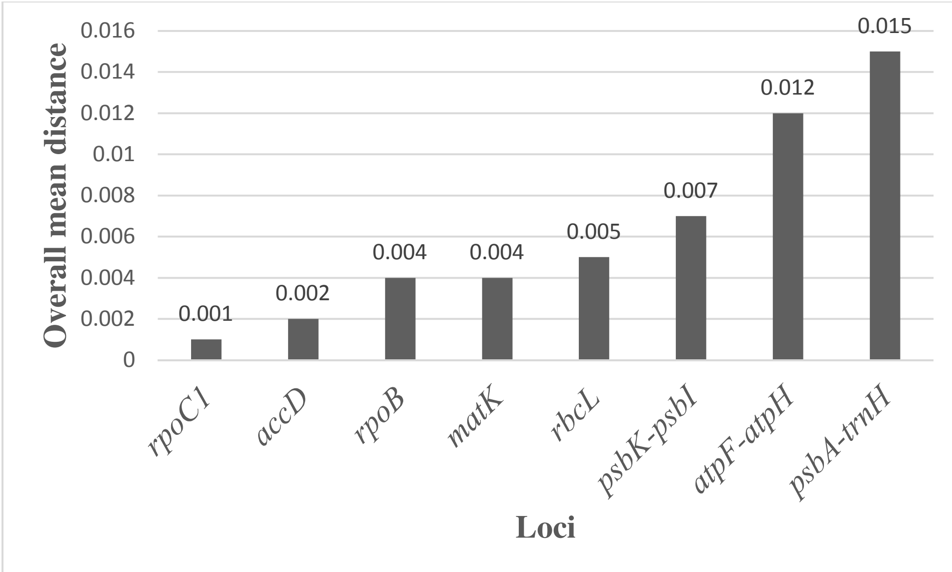


Figure 4.11. Mean distance calculated using K2 parameter between species belonging to the tribe Zingibereae as a measure of the discriminatory power.

4.3.3.2 Neighbor Joining

The reconstruction of the tree based on NJ using single loci failed to resolve the branches and a high level of polytomy was observed. The phylogeny constructed using coding and non-coding regions were similar but the resolution was very poor in case of phylogeny constructed based on information from coding region. The tribe Zingibereae was broadly divided into two major cluster. While the smaller cluster consisted of only one species, *Z. zerumbet* and was found to be the most distant *Zingiber* species. The larger cluster consisted of the rest of the *Zingiber* species, with all the unidentified species grouped with the *Z. officinale*. Since all the unidentified species was found to be morphologically diverse from the *Z. officinale*, hence two assumption can be made for these unidentified species. Either they are some ecotype variant of the *Z. officinale* or are a different species which is closely related to the *Z. officinale*. Among all the eight single loci, the phylogeny drawn based on *matK, rbcL* and *psbK-psbI* was found to provide high resolution and the phylogenies drawn broadly agreed with each other (Figure 4.12).

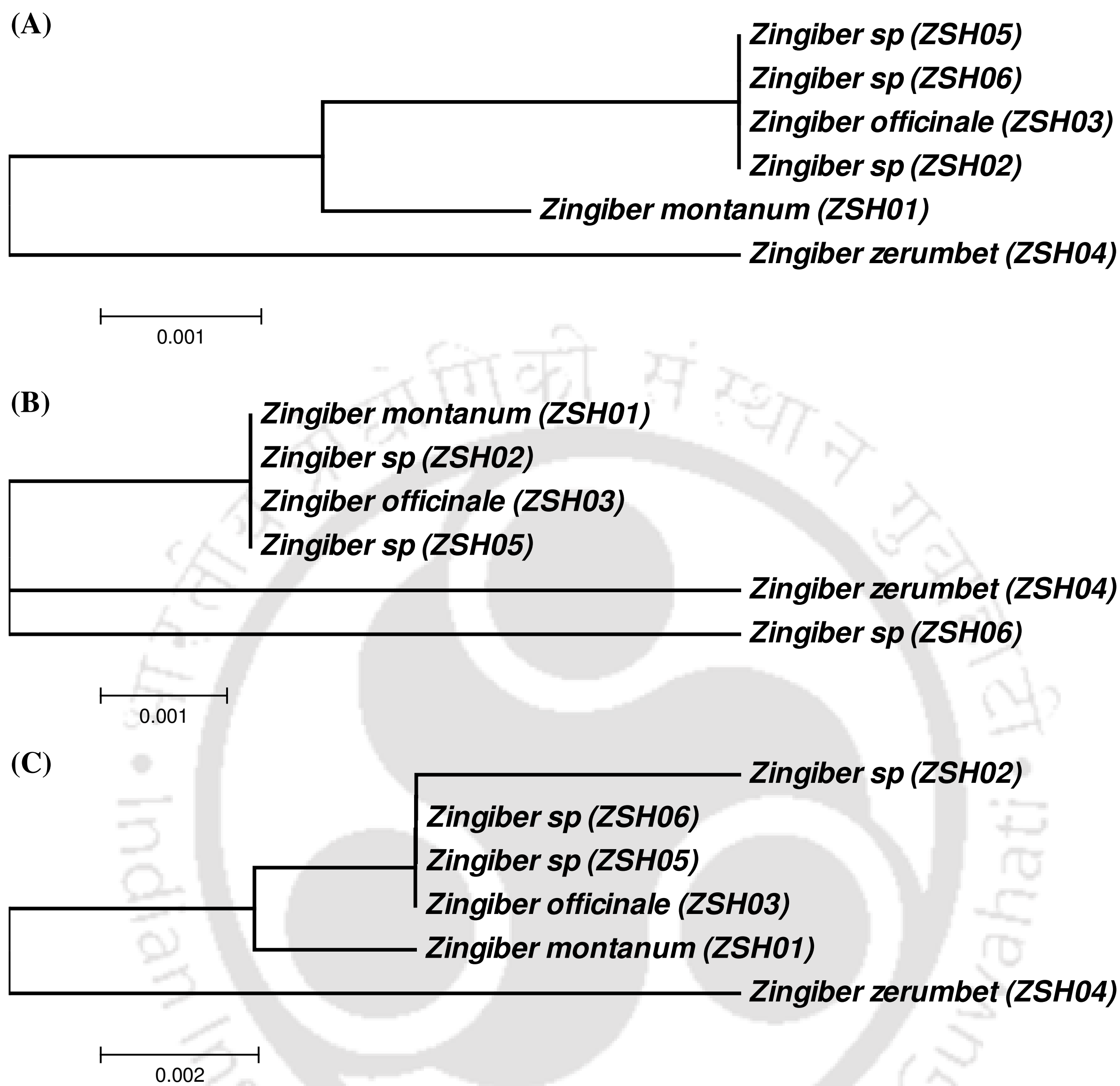


Figure 4.12. The evolutionary history reconstructed for the tribe Zingibereae using single loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (A) *matK* (B) *rbcL* (C) *psbK-psbI*.

Since the single loci failed to resolve most of the phylogeny, combination of loci were used for phylogenetic relationship assessment. The choice of combination of loci was made as described in the section 4.3.1.1 and the list of combination is described in Annexure A 4.2. Among the combination of loci tested, the *matK* with *psbK-psbI* and *rbcL* with *atpF-atpH* combinations showed a higher level of resolution (Figure 4.13). The tree constructed using all the combinations were predominately observed to have two clusters, with the species falling consistently in the same cluster. The divergence between the two cluster was found to be 0.015 (number of the base substitutions per site from estimation of net average between the two cluster). The loci combination of *rbcL* + *atpF-atpH* was most successful in resolving the relationship.

The phylogeny based on *matK* + *psbK-psbI*, consisted of two cluster with one of them being the big cluster with all *Zingiber* species except *Z. zerumbet*, while the smaller cluster consisted of only *Z zerumbet*. In the larger cluster the unidentified species clustered compactly with *Z. officinale*. The phylogeny based on *rbcL* + *atpF-atpH* further resolved the relationship among the unidentified species and *Z. officinale*. Fair amount of diversity was observed with all of the unidentified species, with an exception of *Zingiber* species ZSH02, ending up on distinct branch.

4.3.3.3 Maximum Likelihood

The construction of phylogenetic tree using maximum likelihood was performed as stated earlier (Section 4.3.1.3). Maximum-likelihood phylogenies were reconstructed by using MEGA5.0 software and by using the most suitable model (Table 4.6) and the γ parameter as estimated in the analyses. The ML trees constructed based on the model and parameter mentioned in Table 4.6 were found to be in congruence with the topology of tree reconstructed using NJ method. As was the case with the phylogeny drawn using NJ method, the phylogeny reconstructed based on information from single locus using ML method failed to produce trees with fair resolution, whereas the combination of loci was observed to definitely improve the resolution. A good resolved phylogeny was obtained with the combination of *matK* +*psbK-psbI* and *rbcL*+ *atpF-atpH*

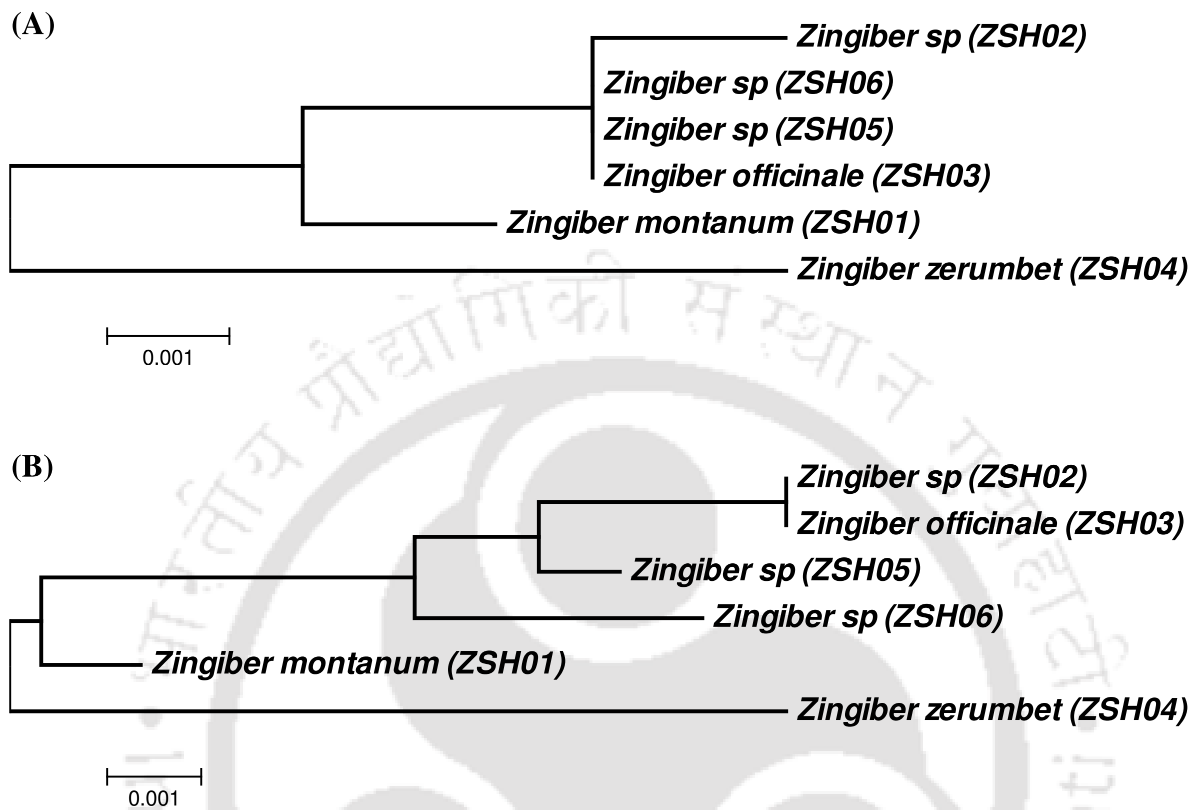


Figure 4.13. The evolutionary history reconstructed for the tribe Zingibereae using combination of loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (A) *matK* + *psbK-psbI* (B) *rbcL* + *atpF-atpH.*

(Figure 4.14), the same combination of loci as was with NJ method. The *rbcL*+ *atpF-atpH* tree was observed to have higher resolution and stronger support for branches. Among the unidentified species in the larger cluster, the *Zingiber* species ZSH06 clustered separately with 97% of tree support While the compact clustering, and hence the closer relationship, of *Zingiber* species ZSH02 with *Z. officinale* was supported by 96% of tree populations.

Table 4.6. The best substitution model as predicted by jModelTest along with associated parameter for individual locus and combinations of loci.

Particulars / Loci	Substitution model (AICc)	Proportion invariable sites	Gamma shape parameter
accD	HKY	-	-
atpF-atpH	HKY	-	-
matK	HKY	-	-
psbA-trnH	HKY	-	-
psbK-psbI	F81	-	-
rbcL	HKY	-	-
rpoB	HKY	-	-
rpoC1	HKY	-	-
accD+atpF-atpH	HKY	-	-
accd+matK	HKY	-	-
accd+psbA-trnH	HKY	-	-
accD+psbK-psbI	HKY	-	-
matK+atpF-atpH	HKY	-	-
matK+psbA-trnH	GTR	-	-
matK+psbK-psbI	HKY	-	-
rbcL+atpF-atpH	HKY	-	-
rbcL+matk	HKY + I	0.938	-
rbcL+psbA-trnH	GTR+I	0.845	-
rbcL+psbK-psbI	HKY+I	0.886	-
rpoB+atpF-atpH	HKY	-	-
rpoB+matK	HKY	-	-
rpoB+psbA-trnH	HKY	-	-
rpoB+psbK-psbI	HKY	-	-
rpoC1+atpF-atpH	HKY	-	-
rpoC1+matK	HKY	-	-
rpoC1+psbA-trnH	GTR	-	-
rpoC1+psbK-psbI	F81	-	-

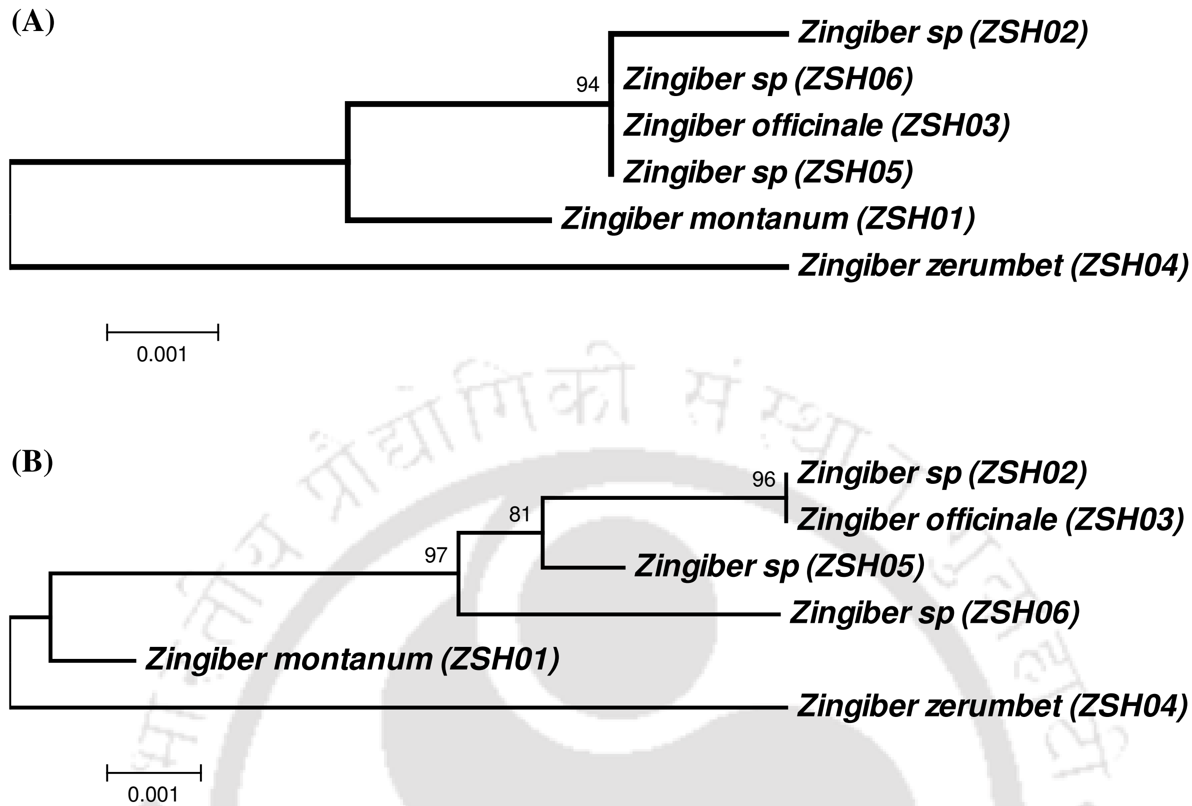


Figure 4.14. Phylogenetic tree constructed for the tribe Zingibereae using the Maximum Likelihood method based on the different substitution model as predicted using jModeTest (Table 4.6). The tree with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (A) *matK* + *psbK-psbI* (B) *rbcL* + *atpF-atpH*.

4.3.3.4 Maximum Parsimony

The MP trees were drawn as described earlier in section 4.3.1.4. The MP trees based on the loci combinations which showed promising result in the precious analyses (NJ and ML) are shown in Figure 4.15. The shortest tree produced from the dataset *matK* + *psbK-psbI* with CI = 1.00 and RI = 1.00 and *rbcL* + *atpF-atpH* with CI = 0.91 and RI = 0.91 (Table 4.7). The topology of the MP tree was similar and in congruence with the NJ and ML tree.

Table 4.7. The characteristics of tree for the tribe Zingibereae resulting from maximum parsimony analyses single and combined locus. CI and RI calculated from informative characters.

Gene	**CI**	**RI**	**Number of trees**	**Tree length**
accD	No PIC			
atpF-atpH	1.0000	1.000000	1	17
matK	1.000000	1.000000	6	8
psbA-trnH	1.000000	1.000000	6	24
psbK-psbI	1.000000	1.000000	6	12
rbcL	1.000000	1.000000	6	7
rpoB	No PIC			
rpoC1	No PIC			
accD+atpF-atpH	1.000000	1.000000	1	19
accd+matK	1.000000	1.000000	6	10
accd+psbA-trnH	1.000000	1.000000	6	26
accD+psbK-psbI	1.000000	1.000000	6	12
matK+atpF-atpH	1.000000	1.000000	1	25
matK+psbA-trnH	1.000000	1.000000	6	32
matK+psbK-psbI	1.000000	1.000000	6	20
rbcL+atpF-atpH	0.909091	0.909091	1	25
rbcL+matk	0.750000	0.666667	5	16
rbcL+psbA-trnH	0.900000	0.888889	5	32
rbcL+psbK-psbI	0.666667	0.500000	6	18
rpoB+atpF-atpH	1.000000	1.000000	1	21
rpoB+matK	1.000000	1.000000	6	12
rpoB+psbA-trnH	1.000000	1.000000	6	28
rpoB+psbK-psbI	1.000000	1.000000	6	14
rpoC1+atpF-atpH	1.000000	1.000000	1	18
rpoC1+matK	1.000000	1.000000	6	9
rpoC1+psbA-trnH	1.000000	1.000000	6	25
rpoC1+psbK-psbI	1.000000	1.000000	6	13

(A)

(B)

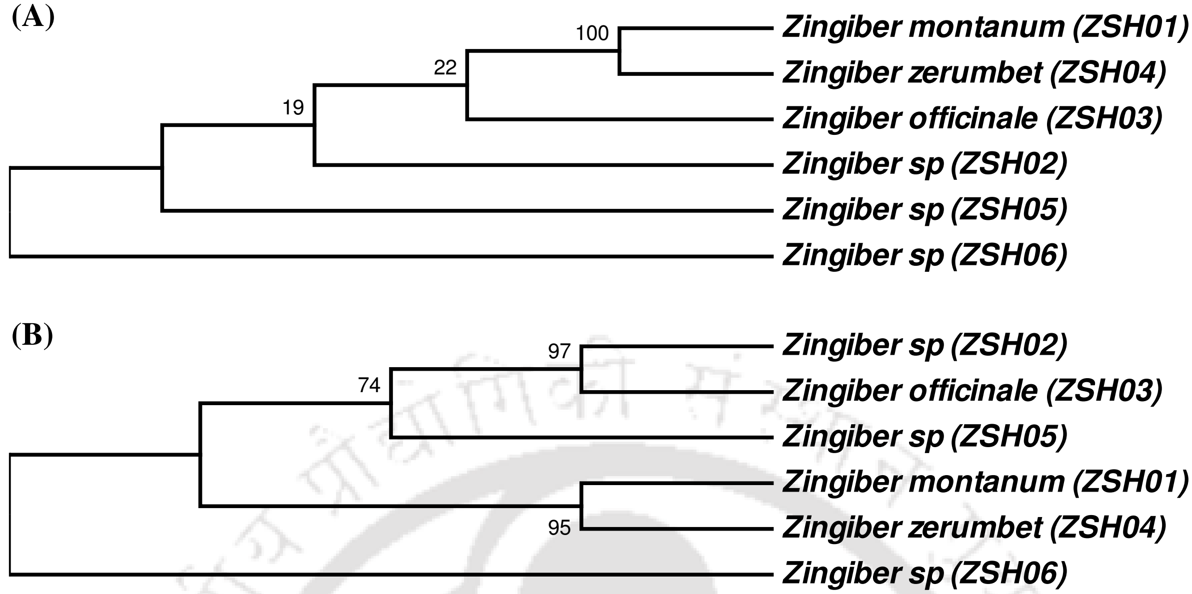


Figure 4.15. The evolutionary history for the tribe Zingibereae was inferred using the Maximum Parsimony method. The most parsimonious trees is shown with their consistency index and retention index mentioned in Table 4.7. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10000 replicates)(A) *matK* + *psbK-psbI* (B) *rbcL* + *atpF-atpH.*

4.3.4 Zingiberaceae

4.3.4.1 Sequence Analyses

The species from the tribe Alpinieae, Hedychieae and Zingibereae were combined into one group for assessing the phylogenetic relationship and also for testing whether these tribes are true clades. Alignment of the coding loci was fairly easy and was performed using the auto features for ClustalW alignment, as implemented by MEGA 5.0, while the alignment of the non-coding regions required manual adjustments by eye to improve the alignment. The genic regions were observed to show a high range of variation (5.26-19.76%) and parsimony informative content (PIC) (2.63-16.91%). Inter-genic regions were fairly diverse (17.05-24.13%), but the 5' and 3' were relatively conserved as the primer were designed from the genic region while the intervening DNA showed high variation and had large insertion-deletions in this region. The PIC value was observed to be high (12.50-22.13%) for the non-coding regions (intervening DNA). Out of all the loci, *matK* was found to stand out as the PIC was fairly high (16.91%) and the alignment was easy to perform. To estimate the discriminatory power of the various loci the overall mean distance was calculated using K2 parameter (Figure 4.16). *PsbK-psbI* was found to possess highest discriminatory power but the alignment also required manual attention. The loci *atpF-atpH* and *matK* were found to be next in line for discriminatory power and appeared as much better choice since they were observed to be highly informative and the alignment was relatively easy requiring very little manual attention.

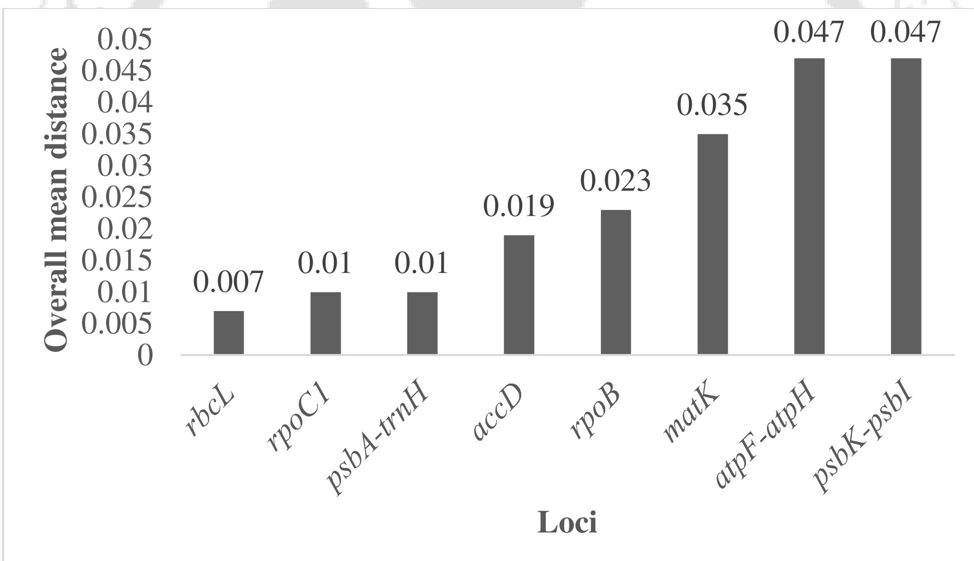


Figure 4.16. Mean distance calculated using K2 parameter between species belonging to the family Zingiberaceae as a measure of the discriminatory power.

4.3.4.2 Neighbor Joining

The resolution of phylogeny reconstructed using single chloroplast loci was observed to increase from use of single genic loci to single noncoding loci. This was also apparent from the percentage of variation observed. The reconstruction of the tree based on NJ using single loci failed to resolve several branches and a high level of polytomy was observed. The phylogeny constructed using coding and non-coding regions were similar but the resolution was very poor in case of phylogeny constructed based on information from coding region. The family Zingiberaceae was invariably divided into two primary clusters, the smaller cluster consisted of all the unidentified species belonging to the genus *Kaempferia.* Among all the eight single loci, the phylogeny drawn based on *rpoB, matK* and *atpF-atpH* was found to provide high resolution and the two phylogeny drawn broadly agreed with each other (Figure 4.17 – 4.19).

In order to improve the resolution of the phylogeny, combinations of loci were tried. The choice of combination of loci was made as described in the section 3.3.1.1 and the list of combination is described in Annexure A 4.2. A total of nineteen combinations were tried and phylogenetic tree was constructed and tested for their ability to resolve the phylogeny. Among the nineteen combinations the pairing of *rbcL* with *matK* with *rpoB* and *atpF-atpH* showed a higher level of resolution (Figure 4.20 & 4.21). The tree constructed using all the combinations were observed majorly to have two clusters, and the members in each cluster was fixed. The average of net divergence between the two cluster, based on the two above mentioned loci combinations, was found to be 0.895 (number of base substitutions per site from estimation of net average between the two cluster).

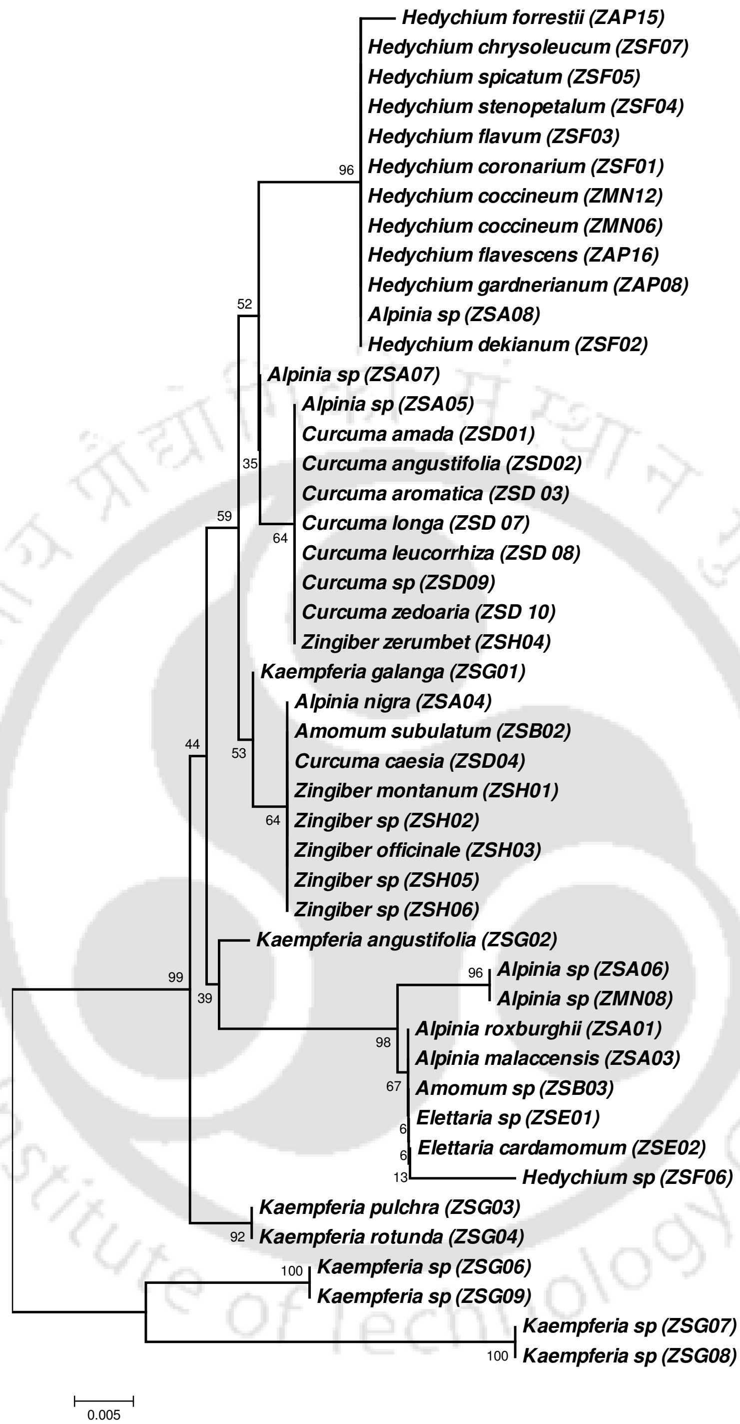


Figure 4.17. The evolutionary history reconstructed for the family Zingiberaceae using *rpoB* loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

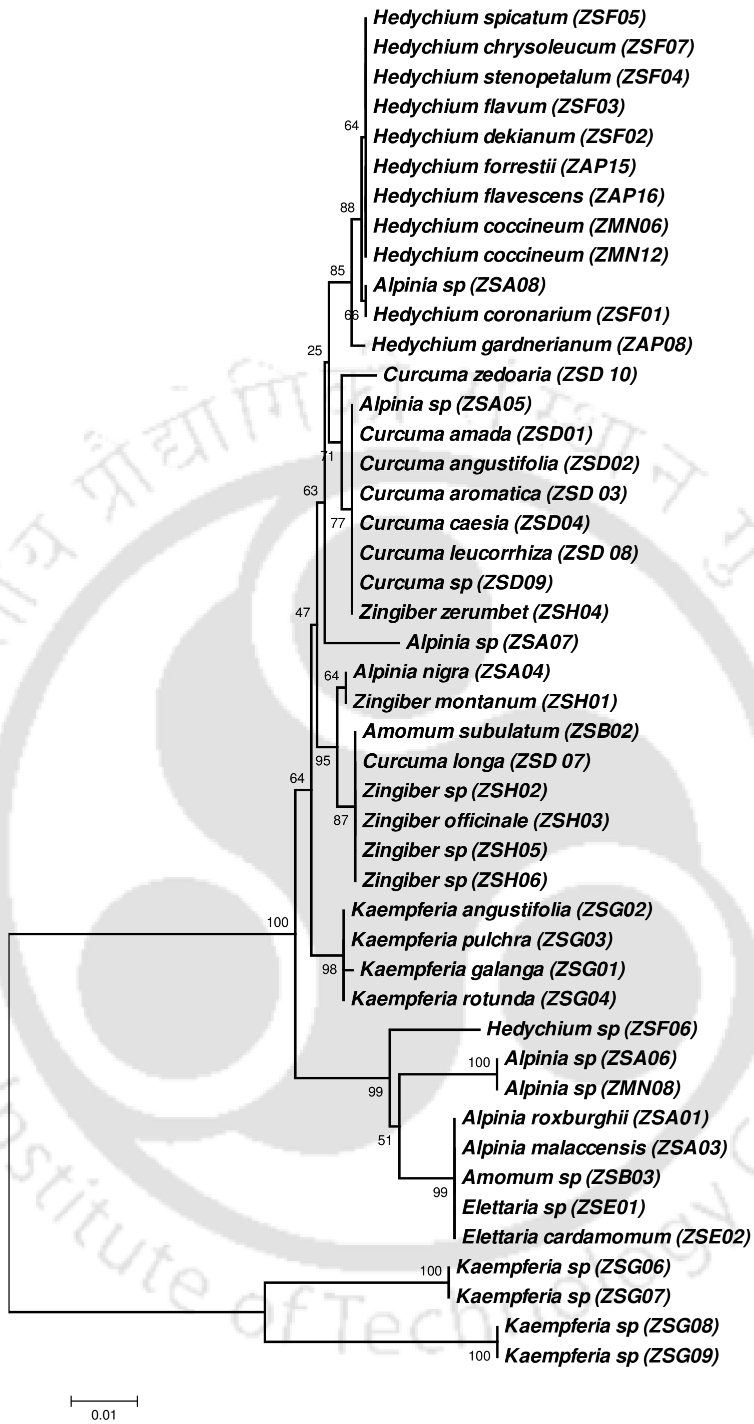


Figure 4.18. The evolutionary history reconstructed for the family Zingiberaceae using *matK* loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

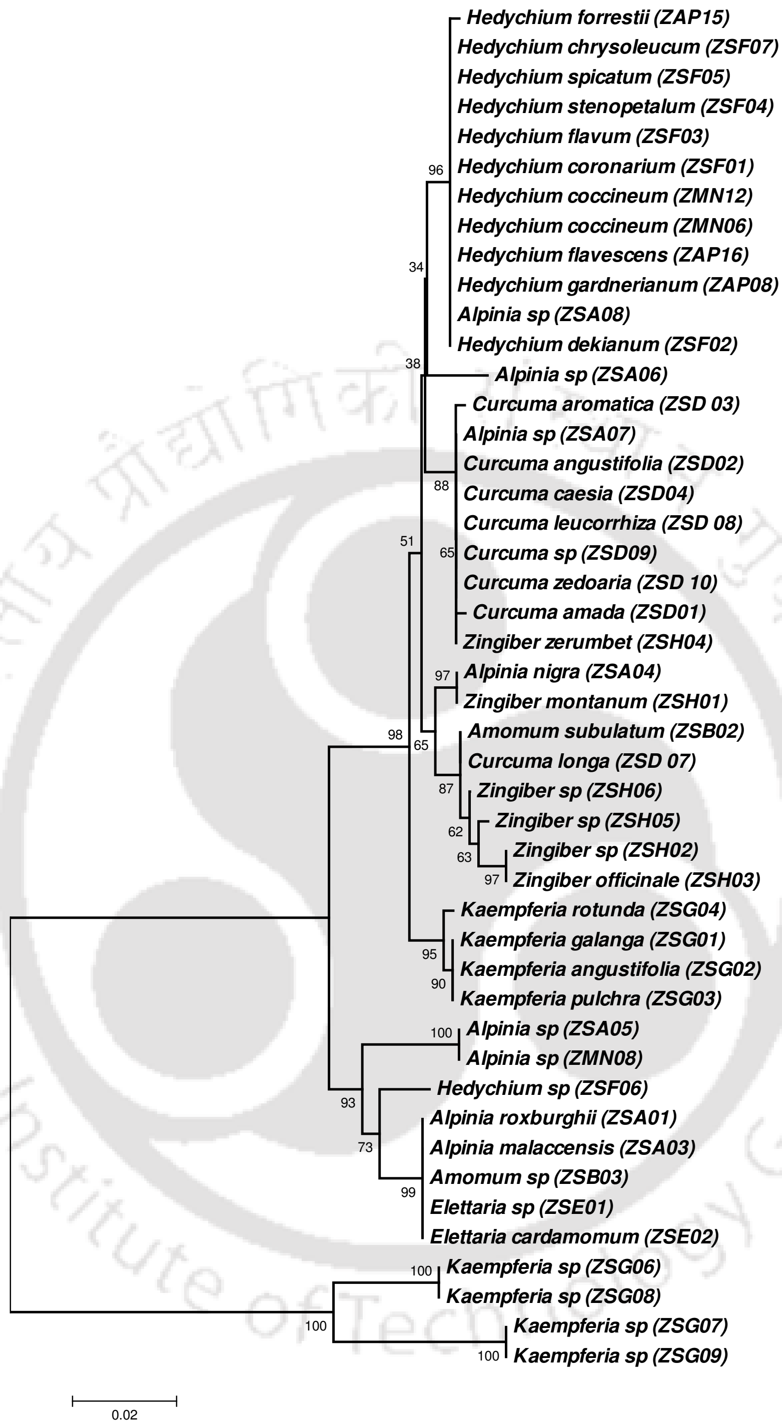


Figure 4.19. The evolutionary history reconstructed for the family Zingiberaceae using *atpF-atpH* loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

The phylogenetic tree, constructed using either single loci or combinations of loci, consisted of two clusters. The smaller cluster mostly consisted of *Kaempferia* unidentified species and since these species clustered separately from other *Kaempferia* species their case was observed critically. The *Kaempferia* is believed to be monophyletic (Techaprasan et al. 2010, Techaprasan and Skornickova 2011 as mentioned in section 4.3.2.2) and hence the distant position of *Kaempferia* species was reexamined. The reassessment of morphological characters of these species revealed that they probably belongs to the family Aeraceae and were retained as an outgroup. The phylogeny based on *rpoB* + *matK* (Figure 4.20) larger cluster of the tree was subdivided into two cluster *viz*., IA and IB. The cluster IA consisted of *Hedychium, Curcuma, Zingiber* and *Kaempferia. Kaempferia* and *Hedychium* was observed to be monophyletic with the exception of an unidentified *Hedychium* species ZSF06 falling into the cluster IB, which primarily consisted of species from the tribe Alpinieae. Since the *Hedychium* shares most of its vegetative morphological characters with *Alpinia* hence this plausibly is a case of miss identification. *Curcuma* was observed to be polyphyletic with the presence of *C. longa* in the *Zingiber* cluster. *Zingiber* was observed as a polyphyletic group with *Z. zerumbet* clustering in the *Curcuma* cluster. *Alpinia* appeared as a polyphyletic group with species distributed in all the clusters except the *Kaempferia. Hedychium* shared the most common ancestor with *Curcuma* while the *Hedychium-Curcuma* shared the most common ancestor with *Zingiber. Eletteria* was found to be monophyletic with the species clustered in the *Alpinia* group (IB cluster). Amomum was found to be distributed in both *Zingiber* and *Alpinia* cluster (IB cluster). The phylogeny based on *matK* + *atpF-atpH* (Figure 4.21) was in congruence with the phylogeny based on *rpoB* + *matK* with some changes in the topology, with most prominent change being the exchange of position for cluster IB and *Kaempferia.* Both the phylogenies had 100% bootstrap support for the two major cluster.

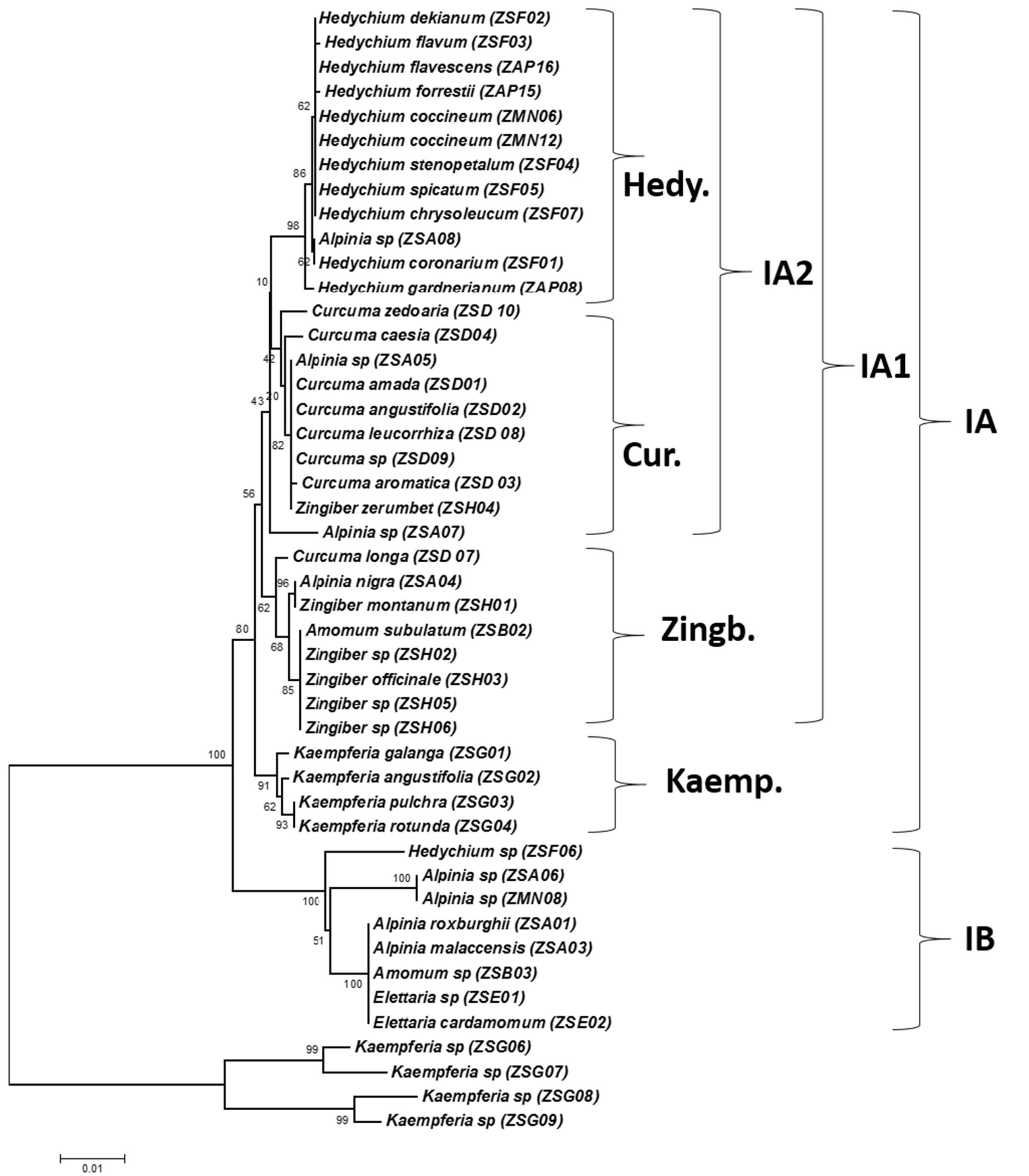


Figure 4.20. The evolutionary history reconstructed for the family Zingiberaceae using *rpoB* + *matK* loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches

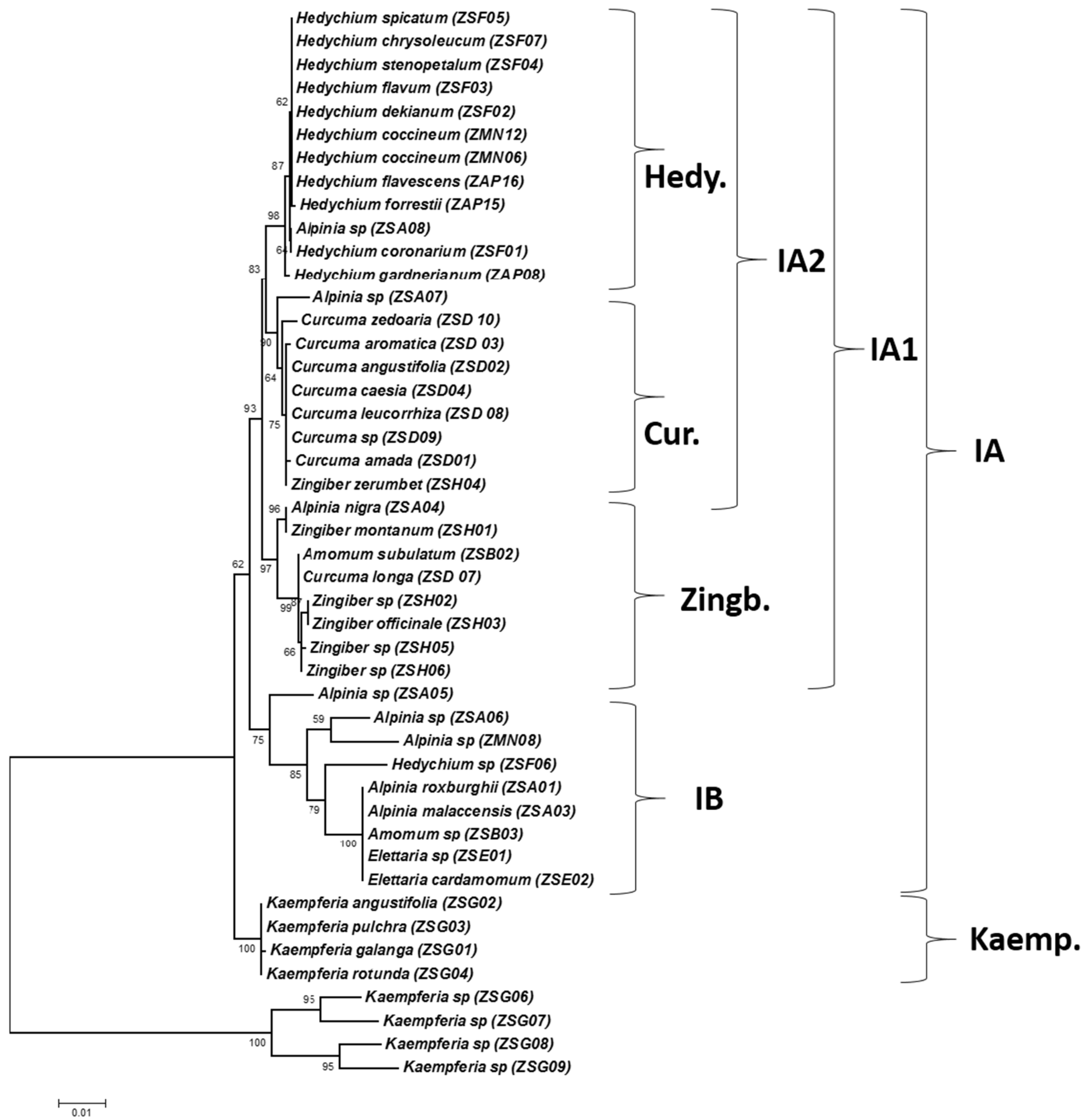


Figure 4.21. The evolutionary history reconstructed for the family Zingiberaceae using *matK* + *atpF-atpH* loci by utilizing Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches

4.3.4.3 Maximum Likelihood

The construction of phylogenetic tree using maximum likelihood was performed as maentioned earlier in section 4.3.1.3. Maximum-likelihood phylogenies were reconstructed by using MEGA5.0 software and by using the most suitable model (Table 4.8) and the γ parameter as estimated in the analyses.

The ML trees constructed based on the model and parameter mentioned in Table 4.8 were found to be mostly in congruence with the topology of tree reconstructed using NJ method. The most important difference in topology was regarding the relationship of *Zingiber* with *Hedychium* and *Curcuma*. As was the case with the phylogeny drawn using NJ method, the phylogeny reconstructed based on information from single locus using ML method failed to produce trees with fair resolution, whereas the combination of loci was observed to definitely improve the resolution. A good resolved phylogeny was obtained with the combination of *rpoB* + *matK* and *matK* + *atpF-atpH* (Figure 4.22 and 4.23), the same combination of loci as was with NJ method. The phylogeny based on *rpoB* + *matK* was in complete congruence with NJ phylogeny and was observed that *Hedychium* was closely related and shared recent common ancestor with *Curcuma*. While with *matK* + *atpF-atpH* the *Hedychium* was observed to be closely related and shared recent common ancestor with *Zingiber*. Based on classical morphology, the genus *Curcuma* falls under the tribe Hedychieae along with *Hedychium,* while according to a new classification proposed by Kress et al (2002), based on molecular taxonomy, the tribe Hedychieae and Zingibereae are combined into one tribe with *Zingiber* sharing the recent common ancestor with *Hedychium*. The node where the partition of the three genus occur, is marginally better in case of *Zingiber* sharing recent common ancestor with *Hedychium* (43%) than in case of *Curcuma* sharing recent common ancestor with Hedychium (34%), but is still poor in both the cases.

Table 4.8. The best substitution model for the family Zingiberaceae as predicted by jModelTest along with associated parameter for individual locus and combinations of loci.

Particulars / Loci	Substitution model (AICc)	Proportion invariable sites	Gamma shape parameter
accD	HKY+G	-	0.516
atpF-atpH	GTR+G	-	0.496
matK	GTR+G	-	0.779
psbA-trnH	HKY+G	-	0.219
psbK-psbI	GTR+G	-	1.36
rbcL	HKY	-	-
rpoB	HKY+G	-	0.546
rpoC1	HKY	-	-
accD+atpF-atpH	GTR+G	-	0.227
accd+matK	GTR+G	-	0.361
accd+psbA-trnH	HKY+I+G	0.577	0.719
accD+psbK-psbI	HKY+I+G	0.432	0.989
matK+atpF-atpH	GTR+I+G	0.310	0.619
matK+psbA-trnH	GTR+I+G	0.397	0.773
matK+psbK-psbI	GTR+G	-	0.418
rbcL+atpF-atpH	GTR+I+G	0.564	0.726
rbcL+matk	GTR+I+G	0.615	0.79
rbcL+psbA-trnH	HKY+G	-	0.219
rbcL+psbK-psbI	HKY+I+G	0.638	0.851
rpoB+atpF-atpH	GTR+G	-	0.402
rpoB+matK	GTR+G	-	0.293
rpoB+psbA-trnH	HKY+I+G	0.564	0.632
rpoB+psbK-psbI	HKY+I	0.629	-
rpoC1+atpF-atpH	GTR+G	-	0.213
rpoC1+matK	HKY+G	-	0.336
rpoC1+psbA-trnH	HKY+I+G	0.562	0.675
rpoC1+psbK-psbI	GTR+G	-	0.252

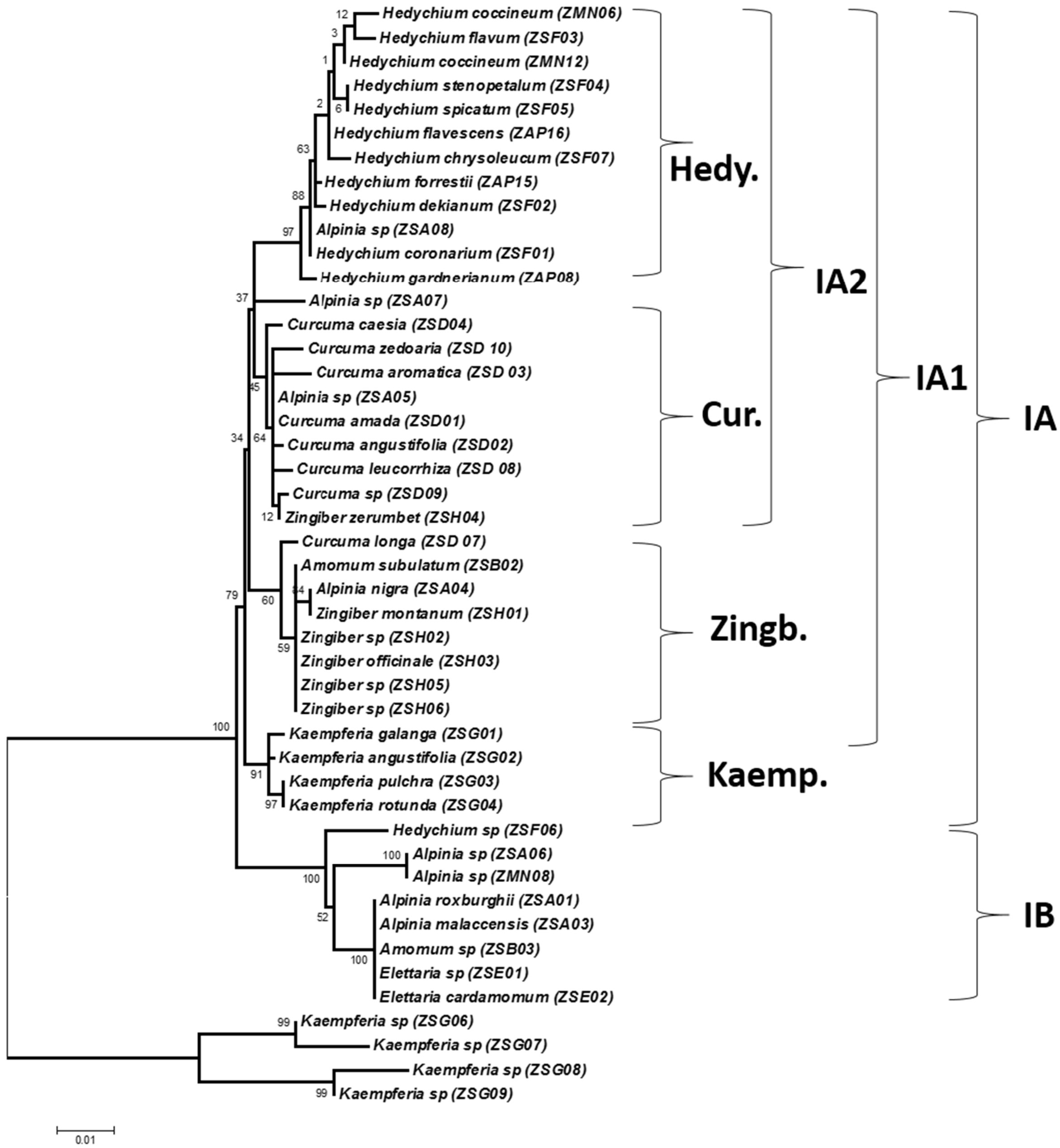


Figure 4.22. Phylogenetic tree constructed for the family Zingiberaceae using *rpoB* + *matK* using the Maximum Likelihood method based on the different substitution model as predicted using jModeTest (Table 4.8). The tree with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

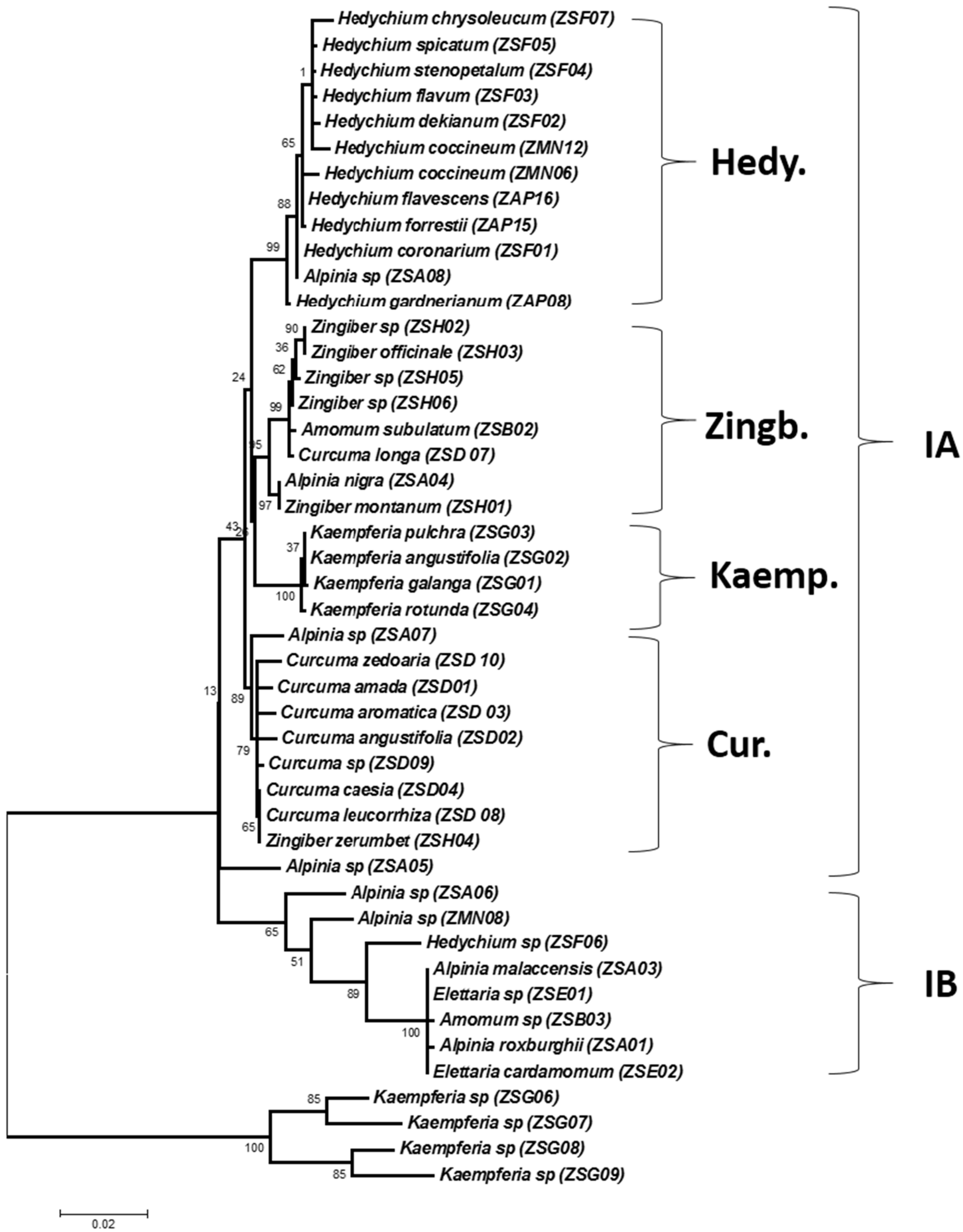


Figure 4.23. Phylogenetic tree constructed for the family Zingiberaceae using *matK* + *atpF-atpH* using the Maximum Likelihood method based on the different substitution model as predicted using jModeTest (Table 4.8). The tree with the highest log likelihood are shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

4.3.4.4 Maximum Parsimony

Among the various locus and their combinations used the best locus for tree construction were not chosen based on the smallest tree length as the shortest tree is best for a given dataset, but given that many dataset are considered the same criteria (shortest tree based on several dataset) cannot be applied for selection of best tree. The MP trees based on the loci combinations which showed promising result in the previous analyses (NJ and ML) are shown in Figure 4.24 and 4.25. The shortest tree produced from the dataset *rpoB* + *matK* with CI = 0.79 and RI = 0.92 and *matK* + *atpF-atpH* with CI = 0.75 and RI = 0.90 (Table 4.9). The topology of the MP tree was similar and in congruence with the ML tree.

Table 4.9. The characteristics of tree for the family Zingiberaceae resulting from maximum parsimony analyses single and combined locus. CI and RI calculated from informative characters.

Gene	CI	RI	Number of trees	Tree length
accD	0.893617	0.954955	42	67
atpF-atpH	0.888889	0.959720	24	221
matK	0.887574	0.963035	11	199
psbA-trnH	0.884058	0.969582	33	139
psbK-psbI	0.859375	0.953125	36	193
rbcL	0.882353	0.975309	8	32
rpoB	0.862745	0.961111		
rpoC1	1.000000	1.000000	43	44
accD+atpF-atpH	0.800000	0.915594	17	338
accd+matK	0.811024	0.925697	29	316
accd+psbA-trnH	0.716763	0.886047	2	262

accD+psbK-psbI	0.748792	0.895582	4	300
matK+atpF-atpH	0.754202	0.897727	11	537
matK+psbA-trnH	0.786713	0.922687	5	412
matK+psbK-psbI	0.802360	0.927095	21	448
rbcL+atpF-atpH	0.767790	0.904762	17	309
rbcL+matk	0.718876	0.888712	2	300
rbcL+psbA-trnH	0.730000	0.912621	3	187
rbcL+psbK-psbI	0.720812	0.887064	1	277
rpoB+atpF-atpH	0.819549	0.933610	9	289
rpoB+matK	0.785185	0.920000	34	323
rpoB+psbA-trnH	0.716049	0.901288	10	265
rpoB+psbK-psbI	0.750000	0.907563	6	314
rpoC1+atpF-atpH	0.822878	0.930836	3	316
rpoC1+matK	0.876777	0.957655	14	265
rpoC1+psbA-trnH	0.789474	0.929619	11	200
rpoC1+psbK-psbI	0.820809	0.936992	8	244

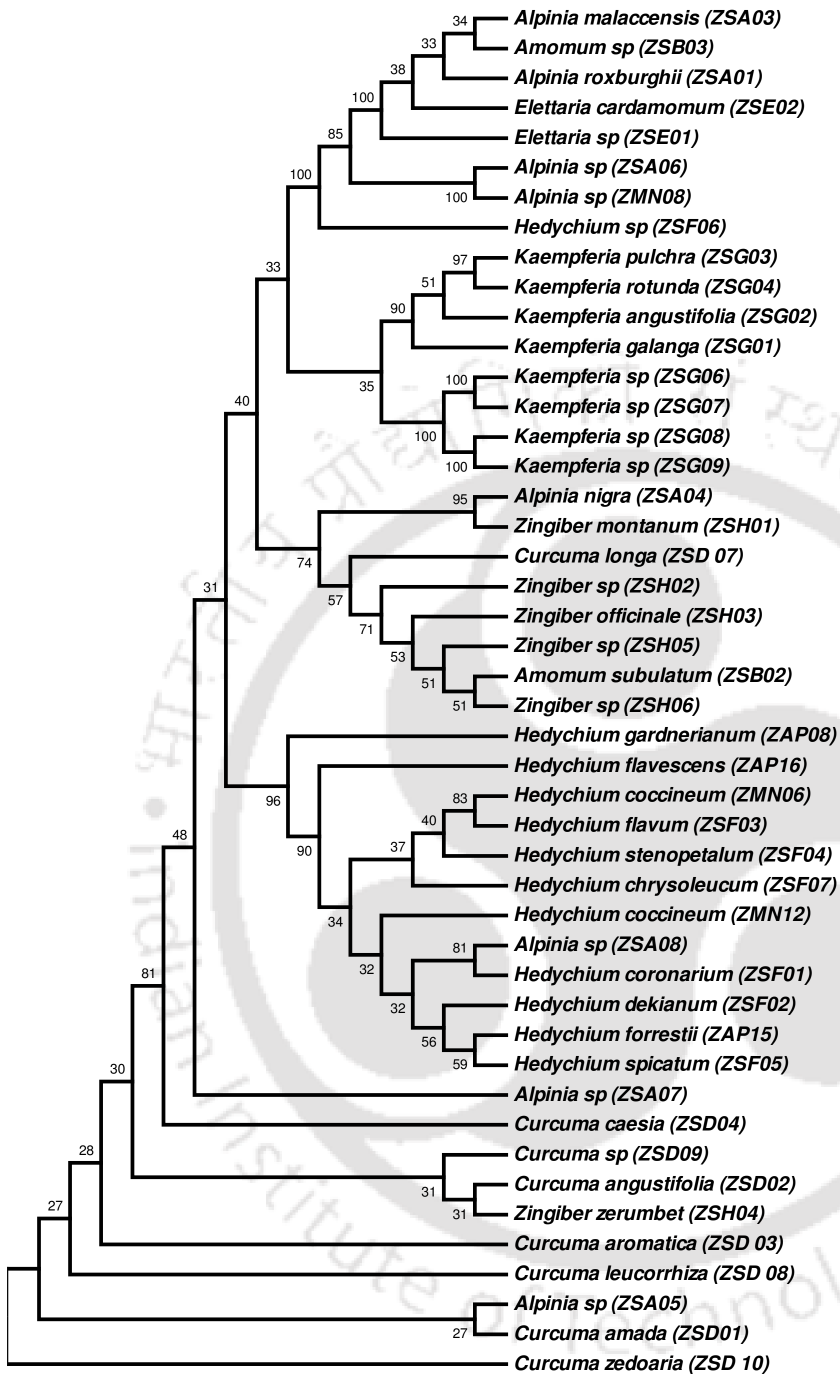


Figure 4.24. The evolutionary history was inferred using the Maximum Parsimony method based on *rpoB* + *matK* dataset. The most parsimonious trees is shown with their consistency index and retention index mentioned in Table 4.9. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10000 replicates).

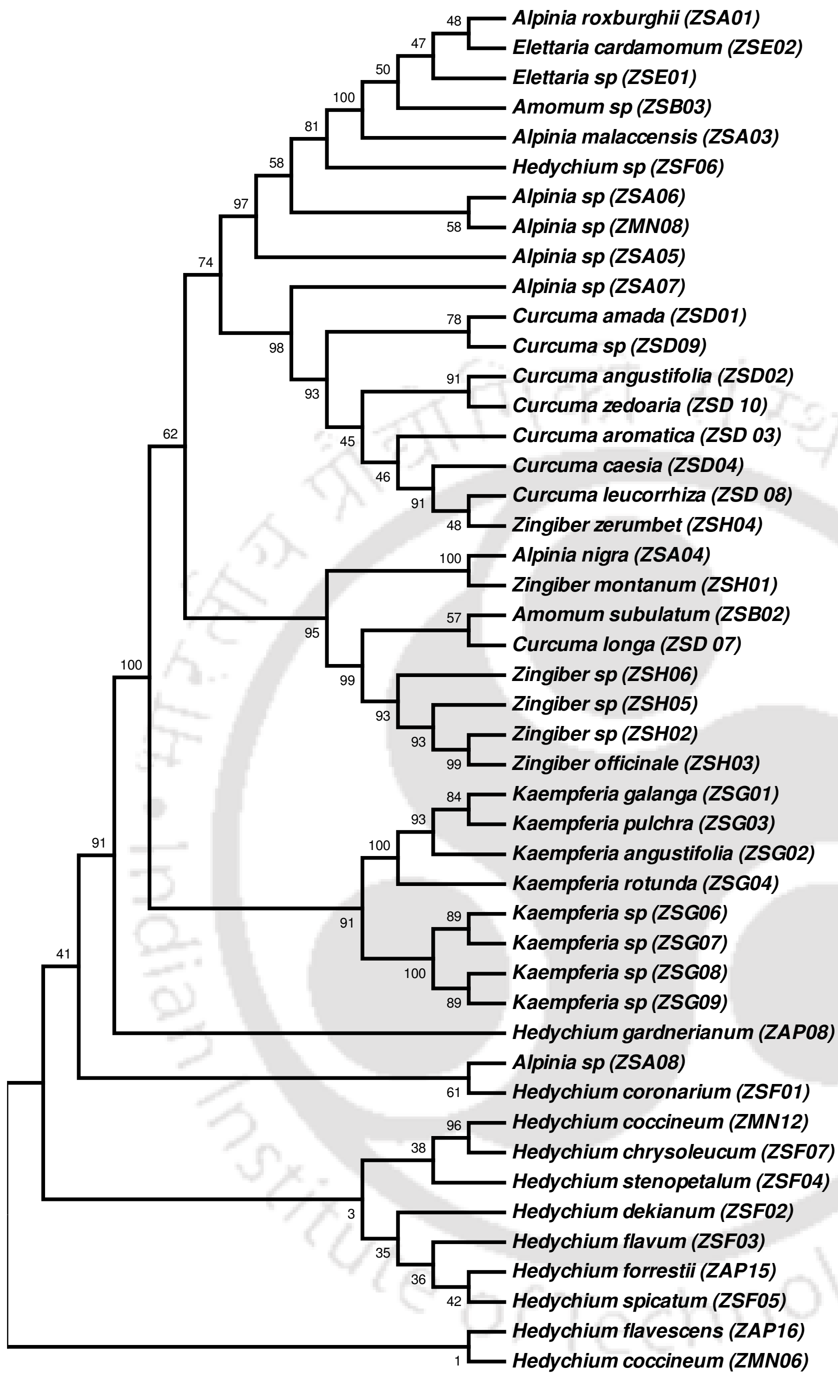


Figure 4.25. The evolutionary history was inferred using the Maximum Parsimony method based on *matK* + *atpF-atpH* dataset. The most parsimonious trees is shown with their consistency index and retention index mentioned in Table 4.9. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10000 replicates).

4.4 Conclusion

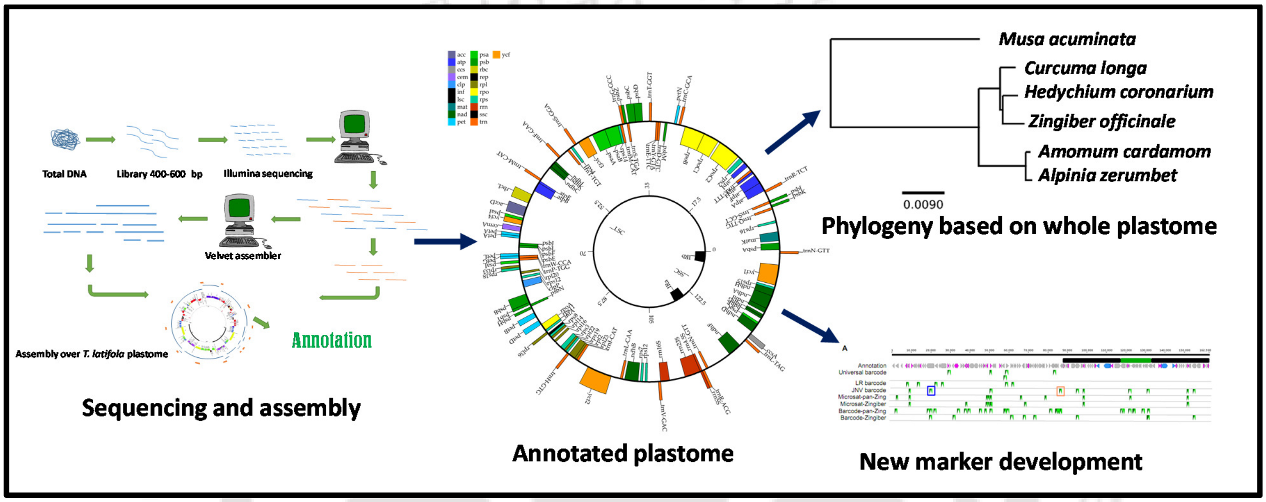
Northeast India is an important reservoir of Zingiberaceae and is considered unique as breaking of Indian tectonic plate from Gondwana plate took place around 140 million years ago. It is estimated that, Indian plate collided Eurasian plate around 35-50 million years ago (Gaina et al 200; Klootwijk 1992) while the estimated date of first appearance of Zingiberaceae is around 70 MYA (D'Hont et al). This suggests that Zingiberaceae on Indian plate evolved independently and hence may differ from Zingiberaceae species from rest of world. In this study the phylogenetic relationship among the Zingiberaceae was reconstructed using NJ, ML and MP analyses. The analyses were done individually for each tribe and then for the whole family. The *genus Eletteria, Hedychium* and *Kaempferi*a were found to be monophyletic while *Alpinia, Amomum, Curcuma* and *Zingiber* were found to be polyphyletic. The relationship between *Curcuma-Hedychium-Zingiber* was found to be conspicuous and future studies should be concentrated on using several Nuclear loci and complete plastid genome sequence for true assessment of relations.

CHAPTER 5

Whole plastome sequencing and new marker development

The chapter deals with sequencing and assembly of plastid genome of five Zingiberaceae species for phylogeny and development of new marker systems

GRAPHICAL ABSTRACT



Whole plastome sequencing and new marker development.

5.1 INTRODUCTION

The endosymbiosis of a cyanobacteria lead to acquiring a great power, to split the water molecule and produce the oxygen, to the eukaryotes. Many millions year after the birth of plastid, the presence of DNA in plastid was reported by Chiba 1951 in a moss *Selaginella* and from two flowering plants. By the end of 1963 the presence of chloroplast DNA was widely acknowledged (Sugiura, 2002). The sequencing of first plastid genome, *Marchantia polymorpha* (Ohyama, 1986) opened several new avenues for phylogenetic study. The plastid genome is a quadripartite molecule (Huang, 2013) with a pair of inverted repeats (IRs) flanked by a large single copy region (LSC) and a small single copy region (SSC). The IRs are exact inverted repeats, varies in size from 20-30 kilo base pair and divides chloroplast into two unequal parts, the LSC and SSC (Kolodner and Tewari 1979). Two populations of chloroplast has been observed differing in the orientation of the SSC region, resulting from the intramolecular recombination with in the IR. The plastome size varies from 120-220 kilo base pair (Schmitz-Linneweber et al., 2001; Ravi et al., 2008) and the coding genes are found in a close cluster (Sato et al., 1999). The junctions between IR and single-copy regions is highly unstable and hence the position of these junctions may vary among different, even closely-related, species (Logacheva et al., 2009). In the order Zingiberales a major expansion of IR region-a (IRa) has been noted in the Musa lineage (Martin et al., 2013).

The first account of use of plastid genome for phylogenetic analysis involved study on restriction site variation (Palmer and Zamir, 1982; Erickson et al., 1983) and it was noticed that the use of plastid DNA for phylogenetic analysis has upper hand over mitochondrial DNA (Provan et al., 2001) as mitochondrial DNA has large size, slow nucleotide substitution rates, extensive levels of intramolecular recombination and high level of ploidy. The whole plastome sequence for phylogeny study (Matsuoka, et al., 2002; Leebens-Mack et al., 2005) has recently increased because of drop in sequencing cost and availability of reference plastome. Along with the drop in cost of sequencing the rapid development of genome- scale data collection and analyses has augmented the task of resolving phylogenetic relationships. The emergence of large number of plastid genome information has also lead to exploitation of the microsatellite (Powell et al., 1995a, b) from plastome especially at low taxonomic levels (Yang et al 2013).

For the plastid genome sequencing and assembly there are three major strategies utilized; polymerase chain reaction (PCR) to amplify chloroplast DNA fragments from genomic DNA (gDNA) extracts (Goremykin et al., 2009), isolation of chloroplast from nuclear extract (Atherton et al., 2005) and whole genome shot gun (Yang et al., 2010). Out of all the three strategies the third one has gained prominence, attributed to recent price drop of next generation sequencing along with advances in computational management and analyses of these millions of short reads. For initial assembly the De Bruijn assembly method such as Velvet (Zerbino and Birney, 2008) and EULER-SR (Chaisson and Pevezner, 2008) or several related assembly method such as ALLPATHS (Butler et al., 2008), MIRA (Chevreux at al., 1999), YASRA (Ratan, 2009), SOAP2 (Li at al., 2009) are in use. These assembler follows different strategy for reducing the computational resources for assembling millions of short reads. The large contigs generated from such assembler could be de novo assembled if the reads are from more than one sequencing platform or if one of the sequencing platform has mate pair library based sequencing result. Alternatively the contigs can be mapped on the plastid genome of closely related species. In last five years more than 200 new plastid genomes has been sequenced and assembled (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid#). Among Zingiberaceae members the plastome has been sequenced and assembled only for Zingiber spectabile (Barrett et al., 2012).

The objective of this study is to sequence and assemble plastome for five species from the family Zingiberaceae: *Zingiber officinale*, *Amomum cardamomum, Alpinia zerumbet*, *Hedychium coronarium*, and *Curcuma longa*. The plastid sequenced was further used for assessing the phylogenetic relationship and development of new marker system.

5.2 METHODOLOGY

5.2.1 Sampling of Zingiberaceae accessions

Zingiberaceae accessions used in the current study were collected from US and India (Table 5.1) and are being maintained in green-houses at the University of Georgia, Athens, GA and the Indian Institute of Technology Guwahati, Assam, India.

Table 5.1. General features of the material used for sequencing and the sequencing method.

Species	Site of Collection	Sequencing Method	Insert Size (bp)	Total Reads (10^6)	Chloroplast Reads (10^3)	Coverage
Zingiber officinale (TARS18100)	USDA Tropical Agriculture Research Station, Puerto Rico	MiSeq-PE150	400-600	3.0	52.9	50x
Curcuma longa (ZSD07)	Assam, India	MiSeq-PE150	400-600	1.3	58.1	54x
Alpinia zerumbet (TARS1718)	USDA Tropical Agriculture Research Station, Puerto Rico	HiSeq-PE100	400-600	3.0	35.4	22x
Amomum cardamomum (ZSB03)	Assam, India	HiSeq-PE100	400-600	10.8	112.4	70x
Hedychium coronarium (UGA-Hyd)	UGA Green House, Athens, GA	HiSeq-PE100	400-600	20.2	337.1	211x

5.2.2 DNA extraction and Illumina Sequencing

DNA extractions were carried out using PlantDNeasy mini kits (Qiagen, USA) following the manufacturer’s instructions. The TruSeq sample preparation kit (Illumina, USA) was used to prepare DNA libraries with inserts from 400-600 base pair (bp) for paired-end multiplexed sequencing. The sequencing was peformed on two different Illumina machines, HiSeq2000 and MiSeq. Library preparations and sequencing on an Illumina HiSeq2000 instrument were carried out at the Interdisciplinary Center for Biotechnology Research, University of Florida. Library preparations and sequencing on an Illumina MiSeq instrument were performed at the Georgia Genomics Facility, University of Georgia. The reads for individual samples from the pooled library was separated using library specific indexed adapter

5.2.3 Assembly and annotation

Based on the highest coverage and superior quality of reads, the shotgun reads from the *Z. officinale* were first assembled. Total reads were first assembled using edena with default parameters. Reads where then filtered to remove read pairs with polynucleotide runs of >30bp and consecutive Ns of >75. Contigs resulting from edena assembly were fed to velvet as long pseudo-reads ('-long' parameter) to perform a second round of assembly in paired-end mode with k-mer size set to 51 bp and the minimum k-mer coverage set to 2 ('-cov_cutoff 2'). Additionally only 1 long read was required to scaffold two contigs ('-long_mult_cutoff 1'). Masked reads were then mapped back to the final assemblies using bowtie-2 in local mode (--local), and scaffolded gaps were manually curated. This hybrid strategy for assembly was used as a combination of overlap-consensus-layout and k-mer graph approaches by observing that though velvet was more robust to sequencing errors and could extend contigs further whereas edena was less prone in making errors when resolving the ends of duplicated sequences. The CpGAVAS web service

(Liu et al., 2012), based on the MAKER pipeline (Drescher et al., 2000) was used to annotate the *Z. officinale* plastome.

For the assembly of other species in this study, total shotgun reads from a given species were mapped over *Z. officinale* plastome using Geneious (6.0.6) with 'Medium Sensivity' and 5 iterations. Matches with equal scores were randomly assigned to an appropriate position. A consensus sequence was generated such that 50% of the bases across all mapped reads were required to be identical. An 'N' was called if the quality score was <20. Total shotgun reads were remapped to these consensus sequences and manually curated.

The MAUVE aligner ('mauveAligner' algorithm) (Rissman et al., 2009), was used to create whole plastome alignments between *Musa acuminata* sequence (gi|525312436|emb||HF677508.1) and all Zingiberacae sequences generated in this study. A seed weight of 8 was used, and plastomes were assumed to be colinear. Resultant alignments and PhyML were used to estimate the topology, branch length, and substitution rate of a tree relating these species. A Kimura (K80) substitution model was used with 50 bootstrap trials. *M. acuminata* was chosen as the outgroup.

5.2.4 Identifying optimal markers

The method employed for identification of, both, microsatellite and barcode markers used an alignment to maximize both the variability of the marker and the affinity of the primer to the binding sites. To achieve the above stated property, two consensus sequences were generated from whole plastome alignments: the "best-base" consensus, for which the majority base in every column was used and gaps were ignored, and the "ambiguous" consensus, for which 100% of bases were required to match and gaps were considered. Thus, explicit bases (i.e., A, C, G, and T) in the ambiguous consensus are indicative of a perfectly conserved column.

For microsatellite marker development, we used the best-base consensus in conjunction with MISA software (Thiel et al., 2003) to identify di- and tri-nucleotide repeats of 4 units or more. Using a custom Perl program, the resultant coordinates were then screened against the ambiguous consensus to determine if they were in fact polymorphic across the species analyzed, such that an ambiguous base was required to be present within the microsatellite or within four bp to the left or right of it. Primer3 (Untergasser et al., 2012) was then used to identify primer sites within the ambiguous consensus that amplify polymorphic positions, with the following parameters: PRIMER_PRODUCT_SIZE_RANGE=100-280, PRIMER_MAX_END_STABILITY=250, and PRIMER_LIBERAL_BASE=1. Primers were not picked from ambiguous sequence, so, by using the ambiguous consensus, only highly conserved regions were considered.

A similar approach was used for barcode development, except that instead of MISA a custom Perl program was written to find three hundred bp blocks that 1) had less than two percent divergence, 2) did not contain more than seven consecutive ambiguous bases, 3) did not have more than ten percent or fifteen percent divergence for genera or family comparisons, respectively, and 4) had more than four or more than eight single-nucleotide polymorphisms with five conserved flanking bases for genera or family comparisons, respectively. Primer sequences were sought for within three hundred bp upstream and three hundred bp downstream of a three hundred bp block passing the above criteria. The same parameters were used as for microsatellite primer identification (excepting the different size range).

5.3 RESULTS AND DISCUSSION

5.3.1 Assembly and annotation

The *de novo* assembly on the *Z. officinale* chloroplast genome was peformed using DNA sequences from a shotgun library of total cell DNA. Assembly resulted in two major contigs of 87,626 and 45,356. These were aligned to the *Typha latifolia* (gi|289065068|ref|NC_013823.1) plastome and scaffolded with smaller contigs. Borders between the inverted repeat regions (IRa and IRb) and the long and short single copy regions (LSC and SSC, respectively) were manually assembled and curated by iterative mapping of reads. A final plastome with 162,598 bp was produced (Figure 5.1). Excepting a gap at the IRa/SSC border, the *de novo* assembly shows perfect colinearity with the *Musa acuminata* (banana) plastome (Martin et al., 2013) (Figure 5.2). Sequencing reads evenly cover the entire plastome (Figure 5.3) - as the IR regions are effectively identical, reads were randomly placed in either region.

Annotation was performed using CpGAVAS (Liu et al., 2012). In general, there are no major discrepancies in the gene content of the *Z. officinale* plastome compared to known plastomes, excepting an expansion of the IRa region as described below. The expansion creates a duplication of the first 3,912 bp of the *ycf1* gene. YCF1 and YCF2 are the longest proteins encoded by known plastomes and appear to be indispensable to plant survival (Drescher et al., 2000). Though the *Zingiber* plastome encodes a full-length YCF1, it is unknown if the large fragment associated with the IRa/SSC boundary is also expressed.

5.3.2 Repeat expansion and structural variation

The borders of inverted repeats and single copy regions are known to be hypermorphic for major structural variation. A major expansion of IRa has been noted in the *Musa* lineage (Martin et al., 2013). As seen in Figure 5.2, the *Z. officinale* plastome does not completely share this expansion, although it does have a 3kb expansion at this site relative to the otherwise highly collinear *Amborella trichopoda* (gi|34500893|ref|NC_005086.1) and *T. latifolia* plastomes (Figure 5.4) (because of their deep ancestry, the colinear relationship between *Amborella* and *Typha*, among others, suggests that this is the ancestral plastome structure). Further as the IRa/SSC boundary was supported by high coverage of fragments (Figure 5.3). Thus, it appears that an initial 3kb IRa expansion occurred in the *Musa/Zingiber* ancestor

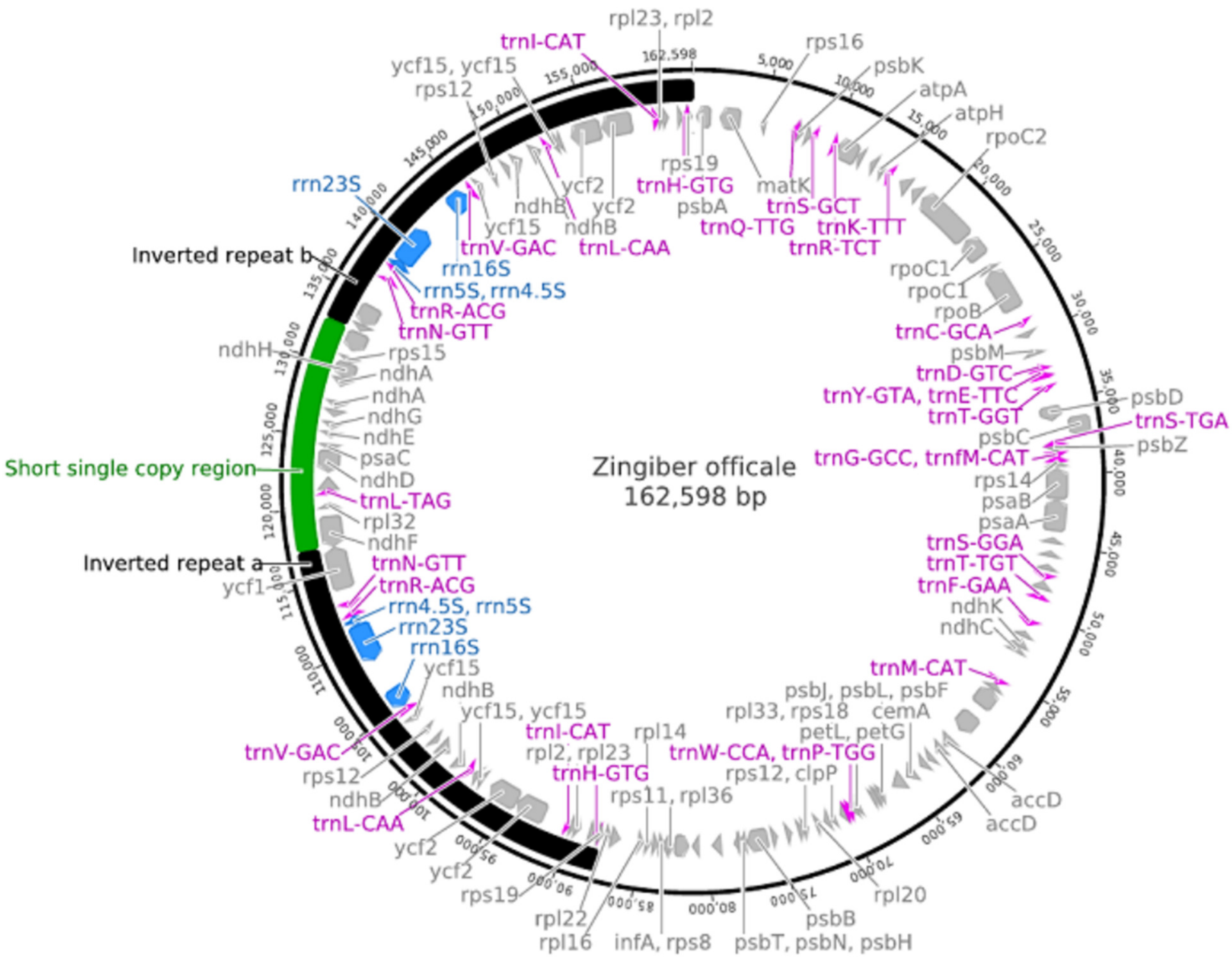


Figure 5.1. The Z. officinale plastome is shown in circular form with IRa and IRb indicated. The short single copy region between the two inverted repeats is shown in green. Protein coding genes are shown in grey, tRNA genes in pink, and rRNA genes in blue.

and was further expanded in the Musa lineage by another 4.7kb. Alternatively, a 7.7 kb expansion could have occurred in the *Musa*/*Zingiber* ancestor followed by a deletion at the IRa/SSC boundary in the *Zingiber* lineage after divergence from the *Musa* lineage. The structure of the recently sequenced *Zingiber spectibale* plastome (Barrett et al., 2013) suggests that these were separate events and that the short expansion has only occurred in *Z. officinale*. By mapping our sequencing reads from the *C. longa* plastome to the *Z. spectibale* scaffold, it is apparent that the scaffold has collapsed the 3kb Zingiberale IRa expansion into the SSC. We see no such discrepancy in coverage when reads from *C. longa* or the

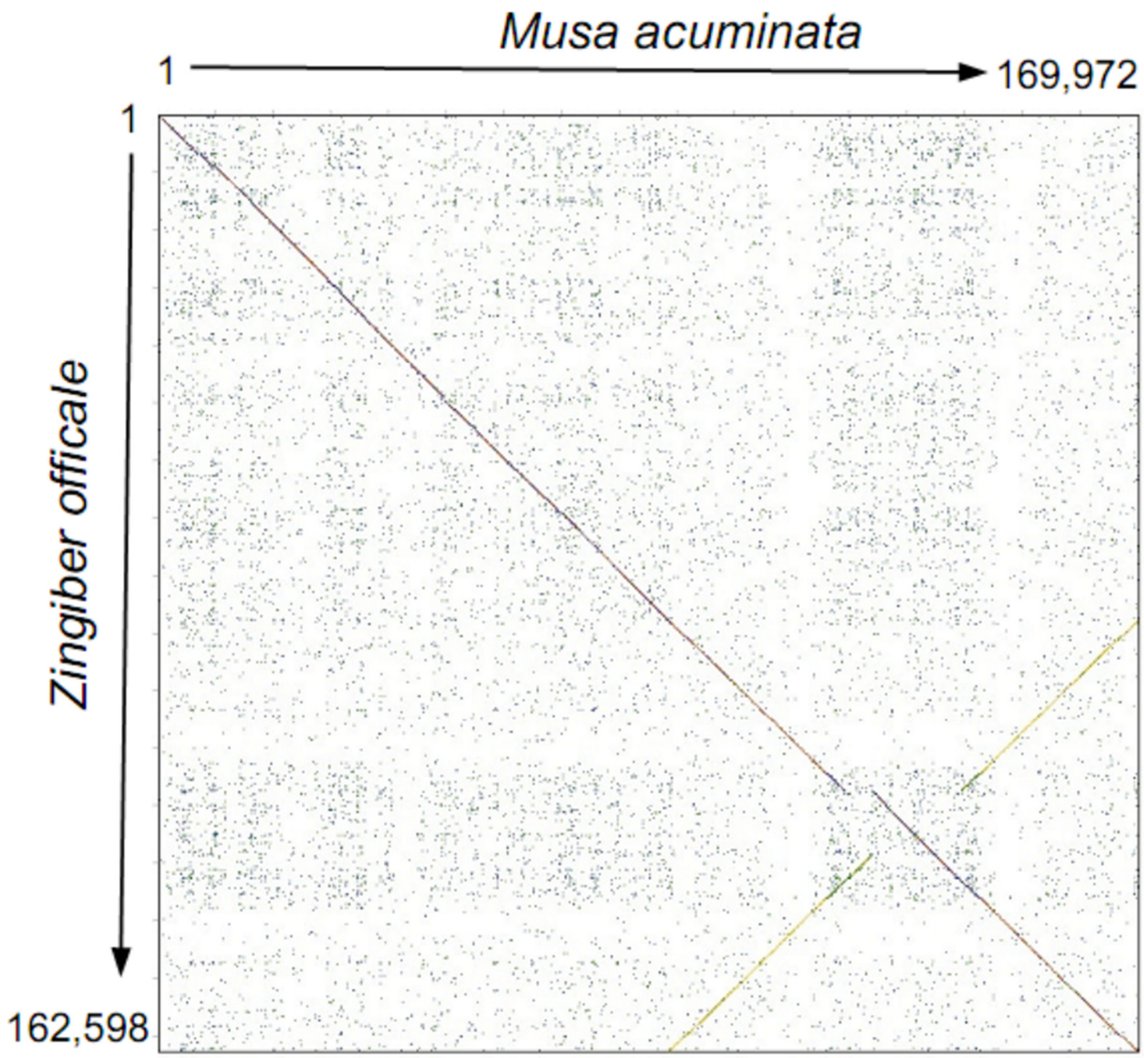


Figure 5.2. Dotplot comparison of the *M. acuminata* and *Z. officinale* plastome sequence. Inverted repeat dotplot is spotted in the extreme down right corner.

more distantly related *A. cardamom* are mapped to the *Z. officinale* plastome. Since this small expansion appears to have occurred at the root of the Zingiberaceae, the large expansion in *Musa* may have occurred in multiple steps. Whether this sequence is predisposed to expansion or such expansions are preferentially retained by natural selection remain open questions.

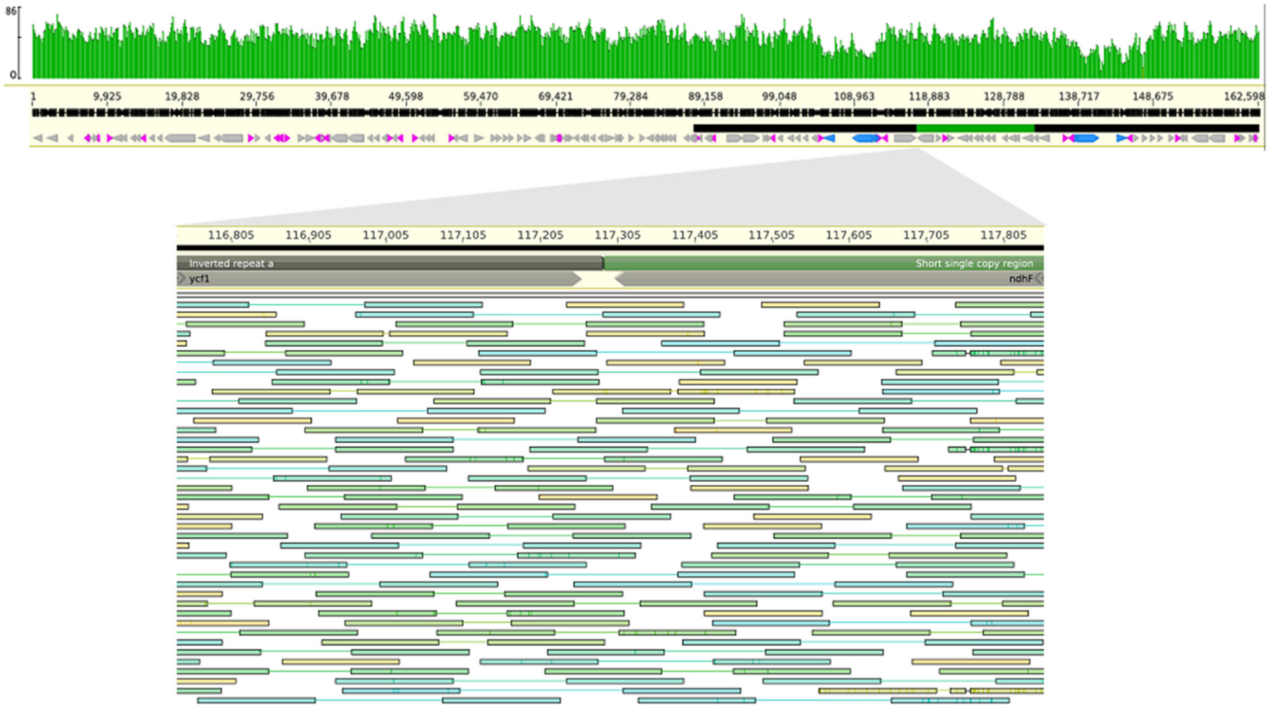
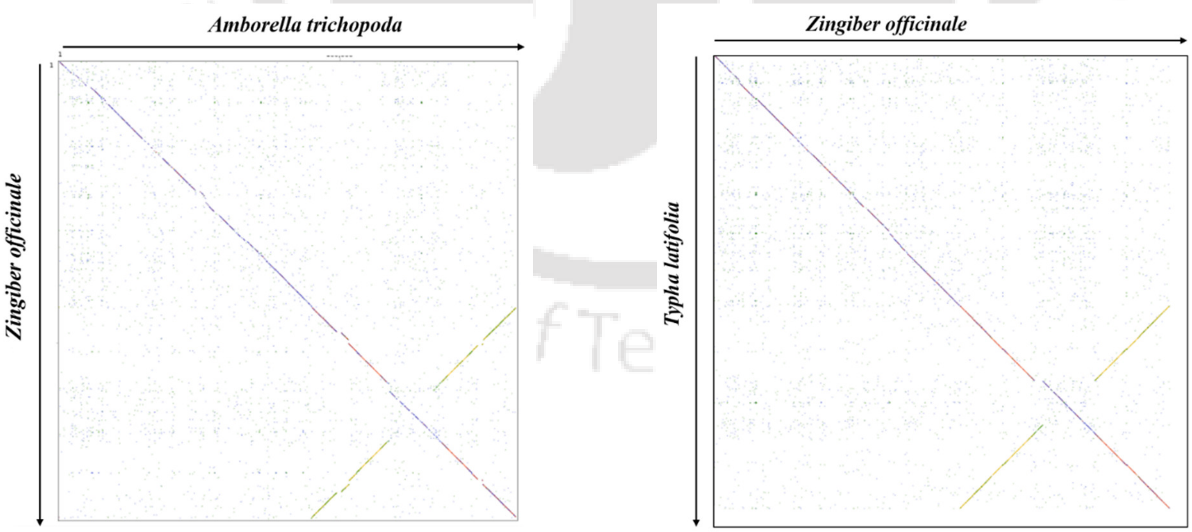


Figure 5.3. Total coverage across the *Z. officinale* plastome. The annotation color scheme is as depicted in A. The lower panel magnifies the border between the IRa and the SSC, where the large indel that differentiates Musa and Zingiber is seen to occur. Reads are shown individually; green is very near the expected insert size, while blue is slightly smaller and gold is slightly longer, respectively, than the expected size. The entire stack of reads is clipped due to space limitations.



(A) (B)

Figure 5.4. Dotplot comparison of the plastome sequence of *Zingiber officinale* with (A) *M. acuminata* and (B) *Typha latifolia* plastome sequence. Inverted repeat dotplot is spotted in the down right corner.

5.3.3 pan-Zingiberaceae variation

For estimating the variation across the Zingiberaceae family, *Z. officinale* plastome was used as a reference on which to assemble data from four other libraries of lower coverage or shorter read length (Table 5.1). The final sequence reported for each plastome is the consensus sequence of these mapped reads.

Generally, the IRs, which encode ribosomal RNAs, are highly conserved relative to both the LSC and SSC regions. Within the LSC and SSC, highly variable regions correlate with regions that either encode tRNAs or short genes and have a low gene density. Interestingly, regions of the IR with similarly low gene density exhibit much higher levels of conservation. This may be related to regulatory constraint outside of the protein coding regions or to a homology-dependent DNA repair mechanism specific to the IR region (Birky and Walsh, 1992).

The tree generated from alignment of these plastomes is shown in Figure 5.5. The tree indicate that *Hedychium* and *Zingiber* share a more recent common ancestor than *Zingiber* and *Curcuma*, although the divergence of all three lineages appears to have been nearly simultaneous.

5.3.4. Microsatellite and barcode marker development

Short sequence repeats (SSRs) or "microsatellites" are hypermorphic due to errors during DNA replication. Such sequences often serve as useful loci for differentiating closely related individuals. In contrast, DNA barcoding exploits point mutations and short indels that occur in rapidly evolving sequences flanked by more slowly evolving sequences. The lower rate of mutations in these barcodes relative to SSRs is somewhat compensated for by the presence of multiple informative sites in a single barcode sequence. Which of these strategies is most appropriate depends on the application, the taxonomic group under study, and the available instrumentation. Indeed, a very short SSR may serve as one among many informative sites within a single barcode.

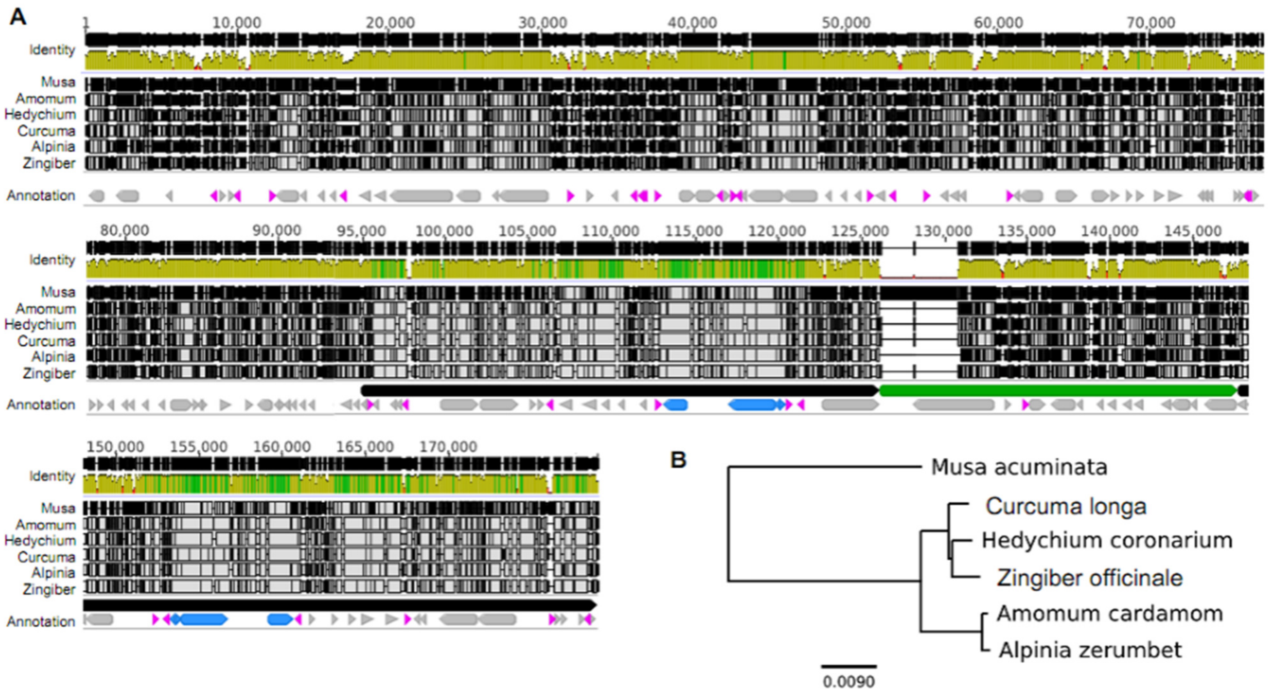


Figure 5.5. Plastome sequence variation across five Zingiberaceae (A) Alignment of whole plastome alignment between *M. acuminata* and Zingiberaceae species sequenced in this study. The annotation color scheme is as depicted in Figure 5.1. In the coverage plot, color indicates percent identity: green, 100%; gold, >30%; red, <30%. Black bars within each species row indicate polymorphisms relative to consensus. (B) A phylogenetic tree based on the plastome alignment. *M. acuminata* was used as an outgroup. Each branch point has 100% bootstrap support. Scale bar indicates the number of substitutions per site estimated to have occurred for such a length along the tree

Using *M. accuminata* plastome sequence, researchers were able to identify 15 chloroplastic SSR markers that were polymorphic within the species (Martin et al., 2013). By using both the *Z. officinale* sequence and the additional Zingiberaceae data produced in this study and elsewhere (Barrett et al., 2013) reliable chloroplastic SSR markers were generated using a comparative approach. Briefly, alignment information from a whole plastome comparison between *Z. officinale* and Z*. spectabile* was used to find di- and tri-nucleotide repeats that were polymorphic within the *Zingiber* genus. The homopolymer repeats were avoided because, generally, they are technically difficult to use as markers. The alignment information was exploited in order to generate reliable primers for the amplification of these polymorphic regions. Using a similar pipeline for the alignment of all plastomes in this study (Figure 5.5, excluding *Musa*), such SSR markers were identified that are likely to be most effective across the Zingiberaceae.

To find suitable barcodes, the alignments were searched for regions that have at least 2% divergence. The regions were also filtered based on the number of single nucleotide polymorphisms (SNPs) that were bordered by 5 conserved bases on both sides of the SNP. Similar to SSR discovery, the markers were identified using the intra-genus comparison for *Zingiber* as well as the comparison across all plastomes from this study. For the intra-genus and inter-genus comparison, at least 5 and 8 of such SNPs were required, respectively, across a 300 bp region. As with the SSR marker discovery, primers were only chosen from perfectly conserved sequence blocks.

A total of 9 pair of primers were designed, of which seven were to be used in pan Zingiberaceae while 2 were developed for *Zingiber* species. The details of the primer is shown in Table 5.2. The Figure 5.6 demonstrates some of the frequently used markers and their position on the genome. The "Universal barcode" primers in Figure 5.6 are from previously proposed sets designed to work across all flowering plants (Kress et al., 2005). Additionally, the "LR barcode" track plots of the marker used in the chapter four (the primers with more than two mismatches are not shown). Figure 5.6 also indicates the four sets of primer pairs resulting from our marker discovery pipeline. Looking across all four sets, we chose a subset of primers that exhibit a high number of polymorphisms at both the intra-genus and intra-family levels ("JNV barcode"). As shown, there is very little overlap between the "JNV barcode" markers and those used in previous studies. While these mismatched primers may still produce a clear PCR product, they are not optimal with regard to heterogeneous starting material. Again, we aimed to make our primers as unbiased as possible by selecting from only highly conserved sites. Thus, in theory, “JNV barcodes” are optimal markers for the Zingiberaceae.

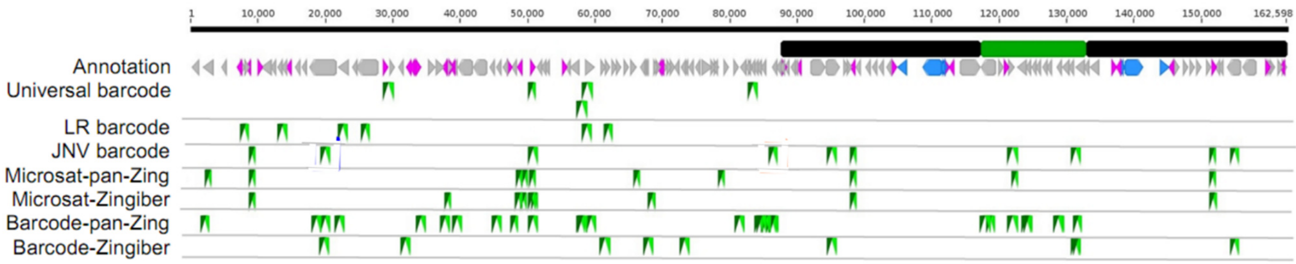


Figure 5.6. Assorted primer combinations from this and prior studies. "Universal barcode" primers are those described in (Kress et al., 2005). "LR barcode" primers are from the chapter four. "JNV barcode" primers are associated with this study and were generated based on the microsatellite and barcode primers identified by the computational pipeline described above. These are indicated in the figure as the last four tracts, where "*Zingiber*" indicates the intraspecific *Z. officinale* and *Z. spectabile* comparison and "*pan-Zing*" indicates a comparison across all sequences described in this study.

Table 5.2. List of primers identified as optimal primers.

Oligo Name	Sequence	Oligo Name	Sequence
MpanZing18F	ACGAAAGATTTATTCCCCCG	MpanZing27R	CCAAATTATGGTGTTGACGC
MpanZing25F	ATGAGGATGGGTCATTCGAG	MpanZing34R	GGTCATGTCATATAGGCCCG
BpanZingib27F	TGAAATGCACCAATCCGTAA	BpanZingib27R	GCCGAACAATGCAAAAGAAT
BpanZingib21F	CGAGTCACACACTAAGCATAGCA	BpanZingib21R	GGAAGCATCGAAGAATTACAGG
BpanZingib10F	TCAAGTCCCTCTATCCCCAA	BpanZingib10R	CCCGGAAAGTCAAAGTAACG
BpanZingib2F	AATTGACCTCTACGGTCCCA	BpanZingib2R	GTGCTGGAACGTCCACTTTT
kress_e	GGTTCAAGTCCCTCTATCCC	kress_f	ATTTGAACTGGTGACACGAG
Bzingiber14F	CATTGCTCTTGCTAATGCGA	Bzingiber14R	CCACCTTAACGACCGAAAAA
Bzingiber11F	TTGCACCTAAAAAGATTCTGTGA	Bzingiber12R	TCGAATCAACTCGTCTAGCTTTT

5.4 CONCLUSION

There is growing consensus that the entire chloroplast genome sequence is an optimal locus in plant identification and phlyogenetics (Parks and Liston, 2009; Nock et al., 2011; Zhang et al., 2011; Yang et al., 2013). This study supports this consensus as markers optimized to focus on the most variable (but still alignable) regions of these plastomes fails to produce enough informative sites for differentiating the most closely related species. Moreover, given the low cost of next-generation sequencing, the price of sequencing an entire plastome versus amplifying and Sanger sequencing a single barcode are approaching equivalence. This is particularly true when a reference plastome is available to facilitate assembly using lower read coverage.

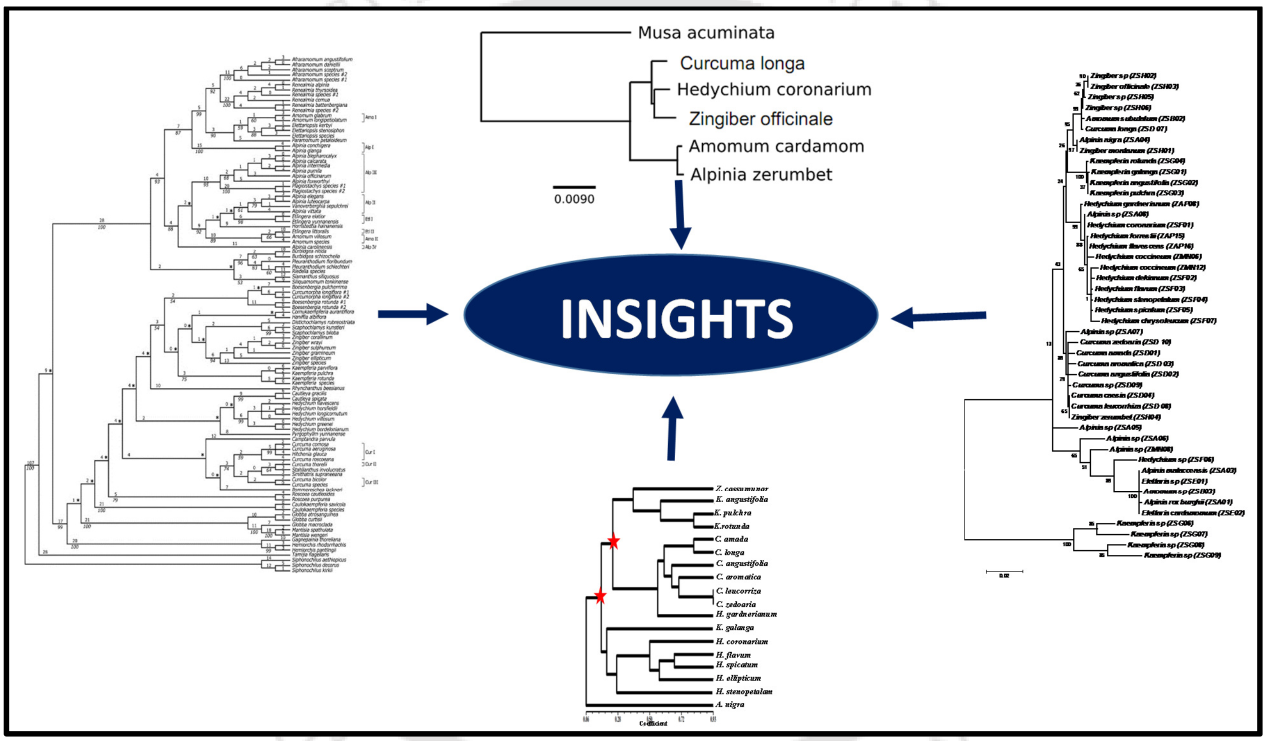
Still, chloroplast sequencing and assembly strategies will be ineffective when faced with a heterogeneous pool of samples from similar species, as might be associated with an ecological study or with herbal/spice contamination detection. Both the former and latter applications are of particular interest within the Zingiberaceae. Therefore, exploiting the range of chloroplast sequences presented in this study, we have also developed a new set of barcode markers optimized for the Zingiberaceae. The test for this markers has to be performed in various sets of conditions, specifically if closely related species could in some cases be resolved. A combination of these markers along with other pre-established marker is still likely to be required for confident identification of the constituent species in a mixed pool.

CHAPTER 6

Comparative study of phylogeny constructed

The chapter deals with comparison of phylogeny drawn using various dataset in this thesis with earlier published work

GRAPHICAL ABSTRACT



Comparative study of phylogeny constructed

6.1 Introduction

Phylogeny re-construction is a science based on evidence (fossil records) as well as assumptions (related to morphological or sequence change). The process of phylogenetic analysis is contingent on estimating evolutionary history by analysing a data set (Swofford 1996). Over the time a number of methods for analyzing the data set has been developed, with each method having its own advantages and perils. Advances in theoretical and mathematical implementation of phylogenetic methodology have been made in recent decades leading to an unprecedented set of choices on methodological issues (Sanderson and Shaffer, 2002). Tree-building methods are broadly classified into distance-based and character-based methods. The distance-based method computes pairwise distances and then, the actual data is discarded and the calculated distance matrix are used in the construction of trees. Trees using character-based method have been optimized according to the distribution of actual data patterns in relation to a specified character. The three major categories of phylogenetic inference methods includes distance based phylogeny, maximum parsimony (MP), and maximum likelihood (ML). The ML methods are especially useful for sequence sets with varying extents of sequence diversity as evident by low problem of long branch attraction (Felsenstein 2001; Mount, 2008).

Apart from a dilemma for choosing the method, the method of taxon sampling is also an important parameter which has deep impact on the re-constructed tree. An insufficient in-group sampling and inappropriate choice of out-group taxa can lead to a long branch attraction artifact which generates controversial findings (Philippe et al., 2009; Pick et al., 2010). Long branches in a phylogeny can strongly influence the topology of early branching in-groups (Philippe and Laurent 1998; Rota-Stabelli and Telford 2008).

As phylogeny reconstruction is affected by such a large parameter and invariably it has been observed that different research group uses different parameter for the reconstruction of tree. In the previous study it was observed that there was a discordance between some relationships (Chapter 4 and 5). The objective of this chapter is to minutely observe the discordant relationship under the family Zingiberaceae. A comparison of phylogeny drawn in earlier work with the present study along with reconstruction of the phylogeny by utilizing the sequence information present in public domain so as to study the effect of taxon sampling.

The literature survey has pointed out two important work which had substantial number of species distributed across the family. The first work was of Kress et al., published in 2002 which involved 104 species from 41 genera and had representation from all the four tribes of the family. These specimens were collected from a diverse region with a large number from Southeast Asia and Polynesia region. The other regions from where the samples were collected from were American continent and some from Africa. The representation of the samples were low from India with only two species. The work involved sequencing of plastid matK and nuclear ITS region. The 2014 publication of Vinitha et al., was a study focused over selection of barcode loci for easy identification of the family. A total of 19 species belonging to seven genera were sampled and had representation of all the four tribes. As the aim of the study was selection of barcode loci, three individual for each species were sampled and a total of ten loci were sequenced. The ten loci sequenced included nine from the plastid genome (accD, matK, ndhJ, rbcL, rpoB, rpoC1, trnH-psbA, trnL-F, and rpl36-rps8 and one from nuclear genome (ITS). Out of the 19 species collected four were collected from Northeast India. The *matK* data from these two mentioned work would supplement the dataset from the current work and hence this study was conducted. Further to observe the dynamics of variation of species from a region to that of a geographically isolated region was carried out by analysing other plastid sequences obtained from public domain (GenBank).

6.2 Methodology

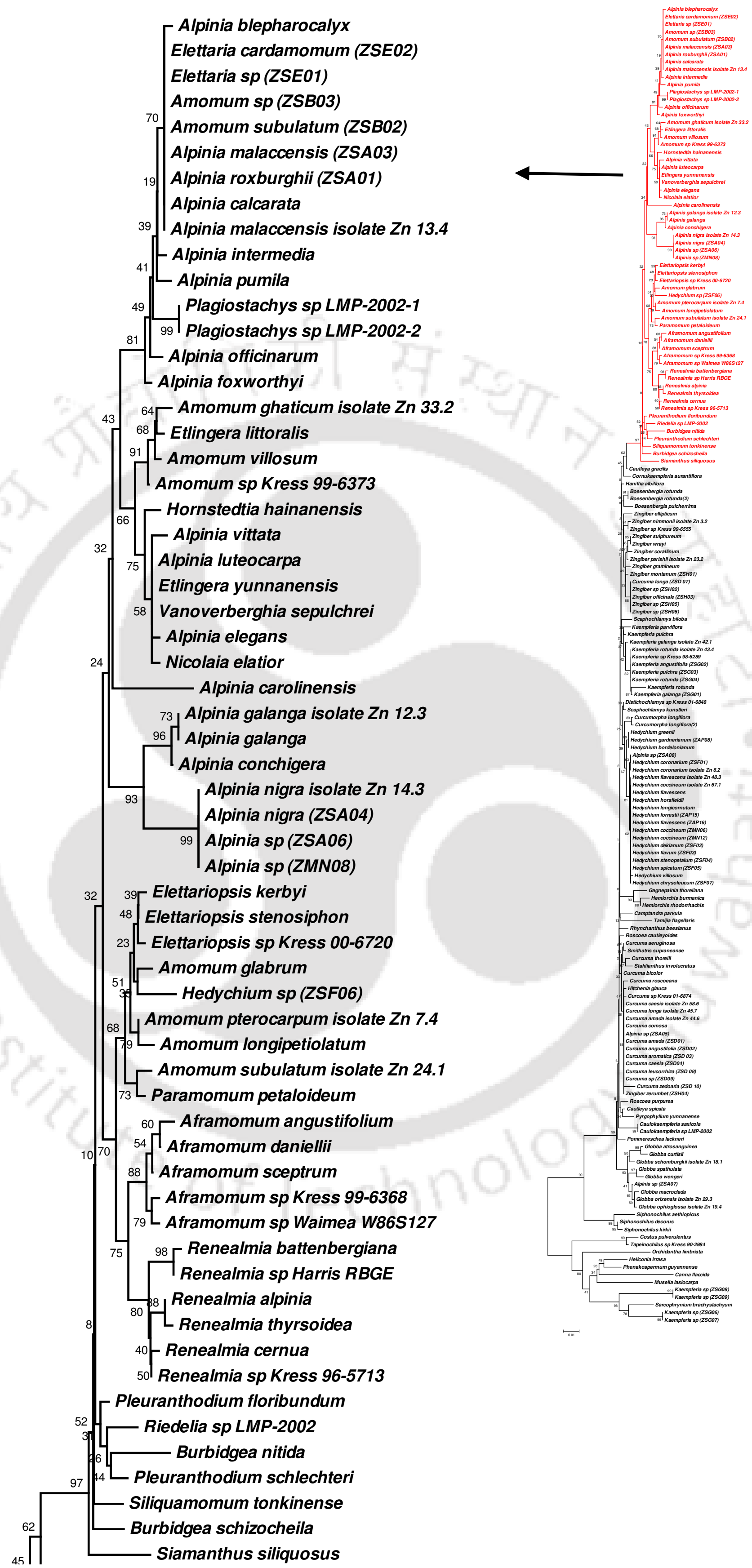
The public domain database for nucleotides were searched for population set of Zingiberaceae species, and sequences were downloaded for the population sets. Two dataset were created, one consisted of *matK* sequences from earlier published work (Kress et al., 2002; Vinitha et al., 2014) combined with the sequences from the current work and the species in the dataset spanned the whole Zingiberaceae family. The second dataset was created, keeping the requirement in sight that a number of plastid loci sequences for all the species to be considered for this part of study should be available. Only species belonging to *Zingiber* was obtained from GenBank, based on the above stated requirement. The sequences from five plastid loci (*matK, psbA-trnH, psbK-psbI, rbcL* and *rpoB*) were retrieved for six species and combined with the same sequences for *Zingiber* from the current study. The sequences were then concatenated as mentioned in Annexure 4.2 (which ever combinations was applicable). Both the two data set was subjected to analyses as mentioned in the section 4.2.4.

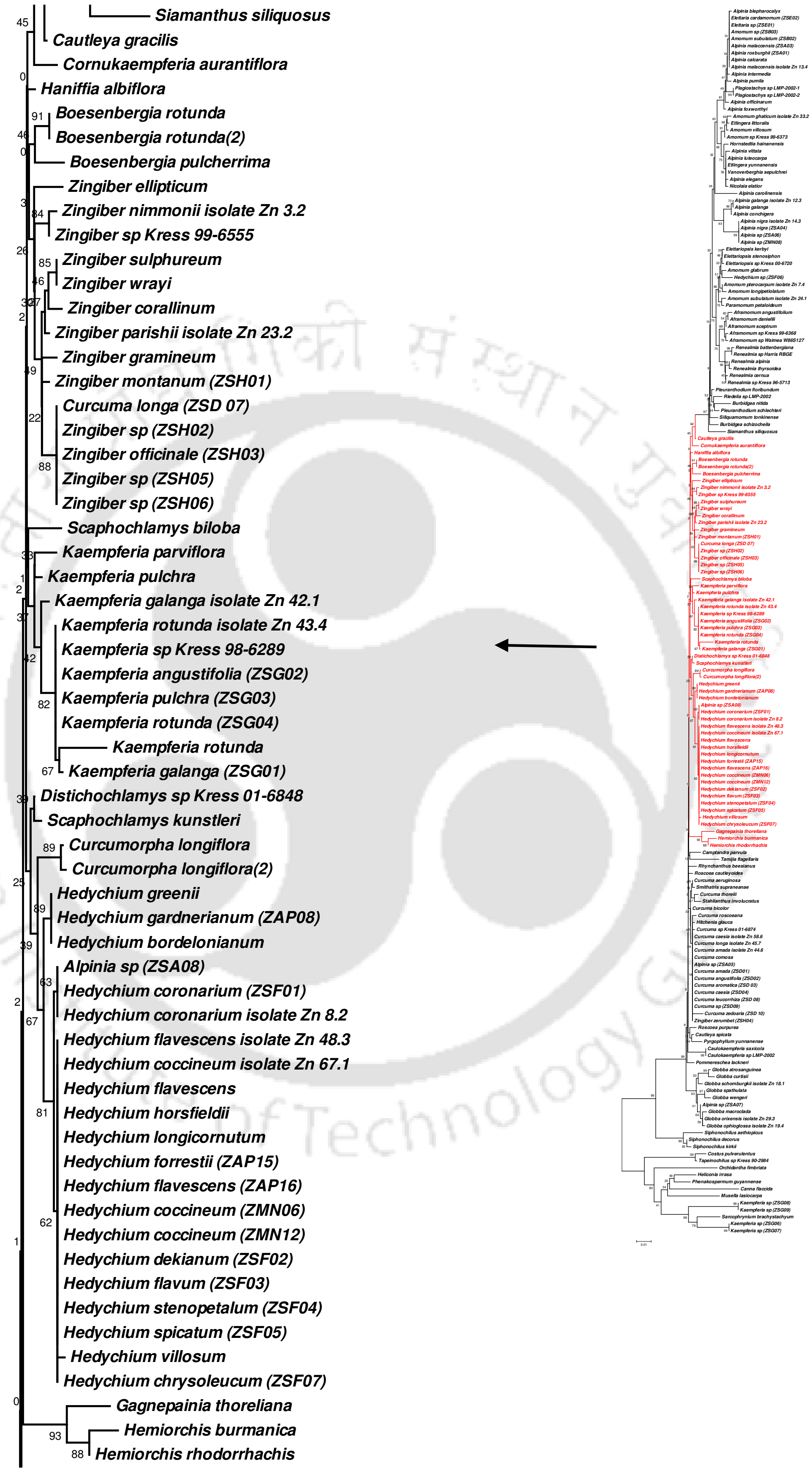
6.3 Results and discussion

6.3.1 Zingiberaceae: Prospect of phylogeny with global sampling

Alignment of the loci was fairly easy and was performed using the auto features for muscle alignment, as implemented by MEGA 5.0. The maximum size of the dataset consisted of 3261 characters but the final dataset was of only 706 sites with large unaligned length, as some of the retrieved sequences were full length gene submission. The PIC value in percentage was observed to be fairly high (21.28%) with average GC content of 29.8%. The reconstruction of the tree based on NJ method failed to fully resolve the tree as was noticed earlier in section 4.3.1.2 but it was able to conclusively clear the relationship between *Hedychium* and *Zingiber* and established them to be closer to each other as compared to *Curcuma* with either of the two (Figure 6.1 and 6.2). Though the bootstrap support was low for the node (37%) but along with the support from the data from the phylogeny drawn using whole plastid genome (Figure 5.5, Page 118) the dilemma ends. Further some of the plants which were identified to the species level but were placed on a branch which seemed counter intuitive (*Alpinia nigra* and *Amomum subulatum*) were found to be placed in the right cluster. The phylogeny drawn using MP method was also congruent with the above stated results (Figure 6.2). The effect of the increase in taxon sampling is hence clearly observed on the outcome of phylogeny. Comparing the phylogeny with the Kress (2000) phylogeny it appears that most of the *Alpinia* in India belongs to the I and III group of *Alpinia*. An interesting thing observed was that most of the species, irrespective of the genera, were placed in *Alpinia* II group was collected from Philippines or from across the ocean *i.e.* Polynesia and China. Such grouping of species into a clad, which belongs to different genera but collected from the same region is interesting and arises many questions regarding the evolution and cross breeding. As the information about the exact place of collection of the samples was not available, the answer to such question can only be answered by a detailed phylogeographic study.

The four unidentified species of *Kaempferia* which we earlier suspected be a miss identification is emphatically established as not belonging to Zingiberaceae. The occurrence of these four species along with the species which were outgroup in the study clearly establishes the above statement as fact.


Alpinia blepharocalyx
Elettaria cardamomum (ZSE02)
Elettaria sp (ZSE01)
Amomum sp (ZSB03)
Amomum subulatum (ZSB02)
Alpinia malaccensis (ZSA03)
Alpinia roxburghii (ZSA01)
Alpinia calcarata
Alpinia malaccensis isolate Zn 13.4
Alpinia intermedia
Alpinia pumila
Plagiostachys sp LMP-2002-1
Plagiostachys sp LMP-2002-2
Alpinia officinarum
Alpinia foxworthyi
Amomum ghaticum isolate Zn 33.2
Etlingera littoralis
Amomum villosum
Amomum sp Kress 99-6373
Hornstedtia hainanensis
Alpinia vittata
Alpinia luteocarpa
Etlingera yunnanensis
Vanoverberghia sepulchrei
Alpinia elegans
Nicolaia elatior
Alpinia carolinensis
Alpinia galanga isolate Zn 12.3
Alpinia galanga
Alpinia conchigera
Alpinia nigra isolate Zn 14.3
Alpinia nigra (ZSA04)
Alpinia sp (ZSA06)
Alpinia sp (ZMN08)
Elettariopsis kerbyi
Elettariopsis stenosiphon
Elettariopsis sp Kress 00-6720
Amomum glabrum
Hedychium sp (ZSF06)
Amomum pterocarpum isolate Zn 7.4
Amomum longipetiolatum
Amomum subulatum isolate Zn 24.1
Paramomum petaloideum
Aframomum angustifolium
Aframomum daniellii
Aframomum sceptrum
Aframomum sp Kress 99-6368
Aframomum sp Waimea W86S127
Renealmia battenbergiana
Renealmia sp Harris RBGE
Renealmia alpinia
Renealmia thyrsoidea
Renealmia cernua
Renealmia sp Kress 96-5713
Pleuranthodium floribundum
Riedelia sp LMP-2002
Burbidgea nitida
Pleuranthodium schlechteri
Siliquamomum tonkinense
Burbidgea schizocheila
Siamanthus siliquosus


Siamanthus siliquosus
Cautleya gracilis
Cornukaempferia aurantiflora
Haniffia albiflora
Boesenbergia rotunda
Boesenbergia rotunda(2)
Boesenbergia pulcherrima
Zingiber ellipticum
Zingiber nimmonii isolate Zn 3.2
Zingiber sp Kress 99-6555
Zingiber sulphureum
Zingiber wrayi
Zingiber corallinum
Zingiber parishii isolate Zn 23.2
Zingiber gramineum
Zingiber montanum (ZSH01)
Curcuma longa (ZSD 07)
Zingiber sp (ZSH02)
Zingiber officinale (ZSH03)
Zingiber sp (ZSH05)
Zingiber sp (ZSH06)
Scaphochlamys biloba
Kaempferia parviflora
Kaempferia pulchra
Kaempferia galanga isolate Zn 42.1
Kaempferia rotunda isolate Zn 43.4
Kaempferia sp Kress 98-6289
Kaempferia angustifolia (ZSG02)
Kaempferia pulchra (ZSG03)
Kaempferia rotunda (ZSG04)
Kaempferia rotunda
Kaempferia galanga (ZSG01)
Distichochlamys sp Kress 01-6848
Scaphochlamys kunstleri
Curcumorpha longiflora
Curcumorpha longiflora(2)
Hedychium greenii
Hedychium gardnerianum (ZAP08)
Hedychium bordelonianum
Alpinia sp (ZSA08)
Hedychium coronarium (ZSF01)
Hedychium coronarium isolate Zn 8.2
Hedychium flavescens isolate Zn 48.3
Hedychium coccineum isolate Zn 67.1
Hedychium flavescens
Hedychium horsfieldii
Hedychium longicornutum
Hedychium forrestii (ZAP15)
Hedychium flavescens (ZAP16)
Hedychium coccineum (ZMN06)
Hedychium coccineum (ZMN12)
Hedychium dekianum (ZSF02)
Hedychium flavum (ZSF03)
Hedychium stenopetalum (ZSF04)
Hedychium spicatum (ZSF05)
Hedychium villosum
Hedychium chrysoleucum (ZSF07)
Gagnepainia thoreliana
Hemiorchis burmanica
Hemiorchis rhodorrhachis

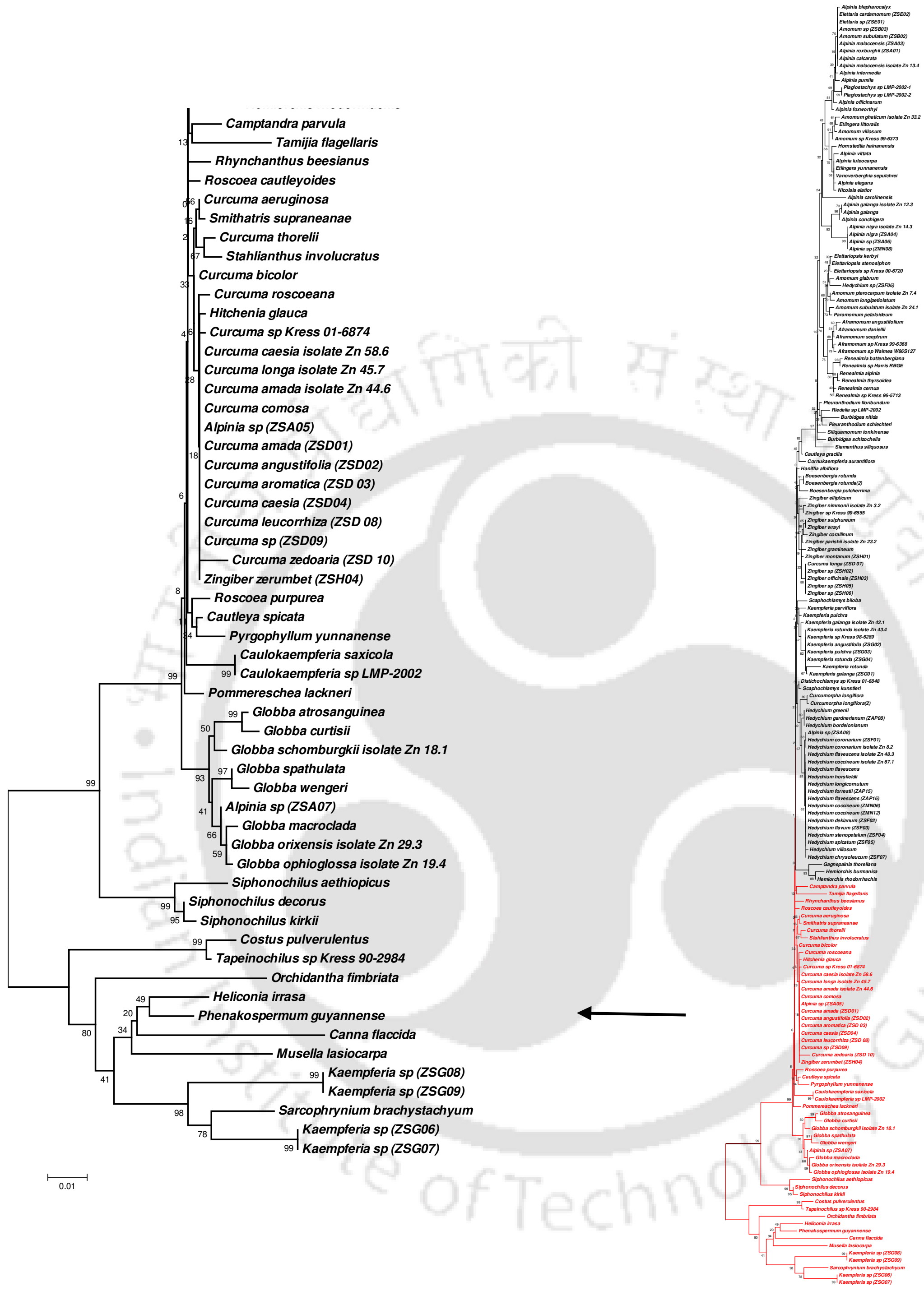
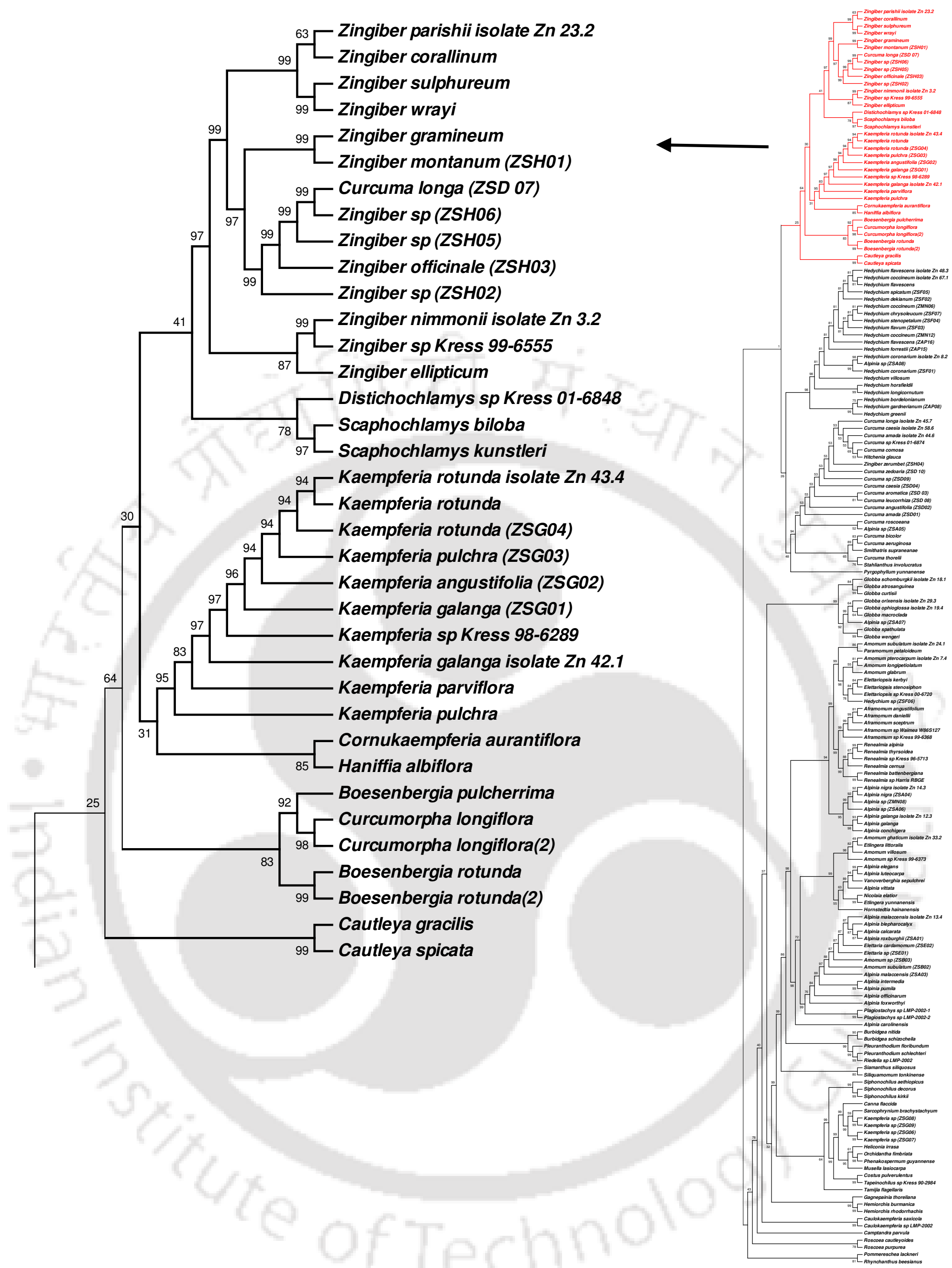
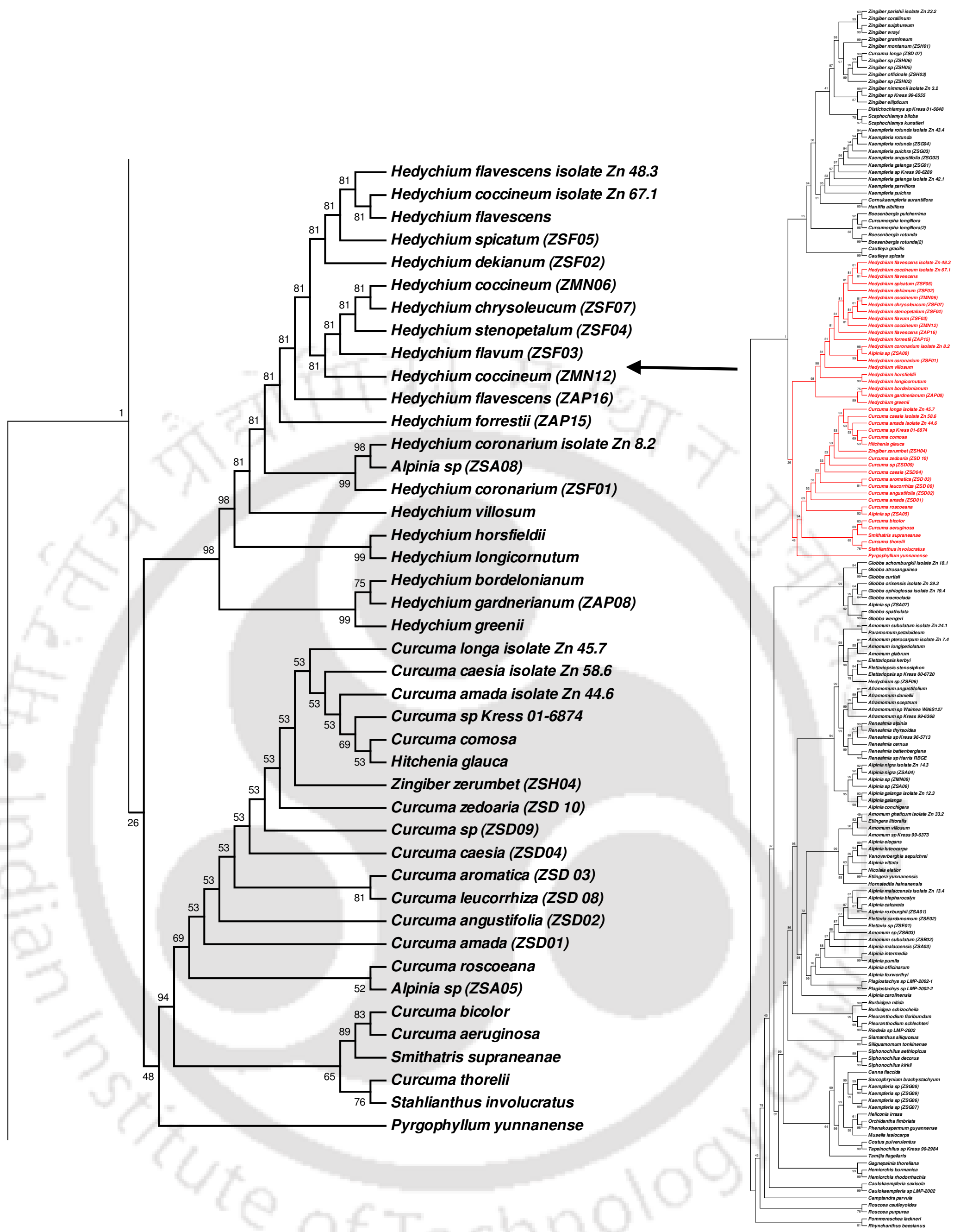
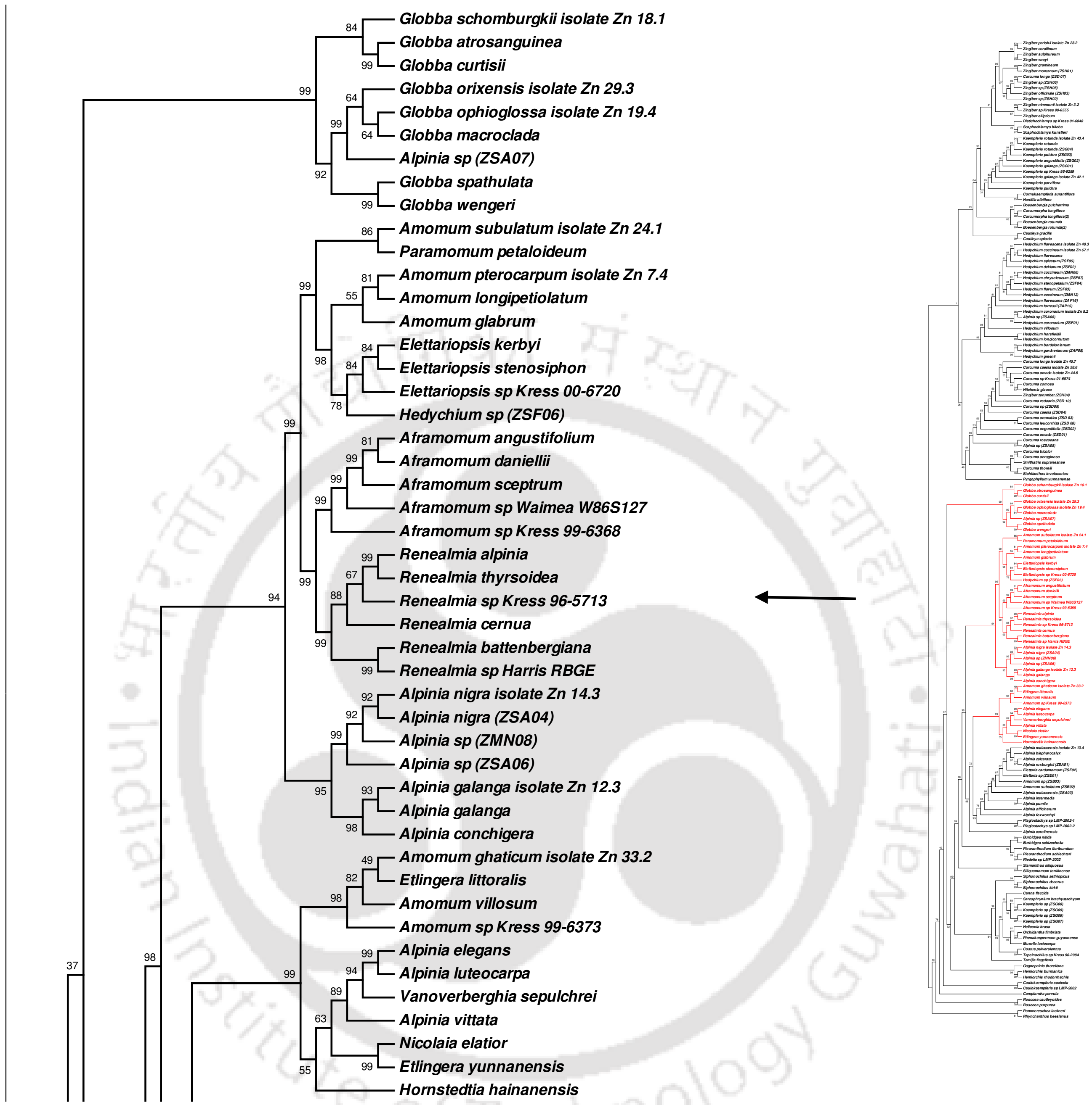


Figure 6.1. The evolutionary history for Zingiberaceae, with sequences from GeBank, inferred using the NJ method. The optimal tree with the sum of branch length = 0.88392967 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.


Zingiber parishii isolate Zn 23.2
Zingiber corallinum
Zingiber sulphureum
Zingiber wrayi
Zingiber gramineum
Zingiber montanum (ZSH01)
Curcuma longa (ZSD 07)
Zingiber sp (ZSH06)
Zingiber sp (ZSH05)
Zingiber officinale (ZSH03)
Zingiber sp (ZSH02)
Zingiber nimmonii isolate Zn 3.2
Zingiber sp Kress 99-6555
Zingiber ellipticum
Distichochlamys sp Kress 01-6848
Scaphochlamys biloba
Scaphochlamys kunstleri
Kaempferia rotunda isolate Zn 43.4
Kaempferia rotunda
Kaempferia rotunda (ZSG04)
Kaempferia pulchra (ZSG03)
Kaempferia angustifolia (ZSG02)
Kaempferia galanga (ZSG01)
Kaempferia sp Kress 98-6289
Kaempferia galanga isolate Zn 42.1
Kaempferia parviflora
Kaempferia pulchra
Cornukaempferia aurantiflora
Haniffia albiflora
Boesenbergia pulcherrima
Curcumorpha longiflora
Curcumorpha longiflora(2)
Boesenbergia rotunda
Boesenbergia rotunda(2)
Cautleya gracilis
Cautleya spicata
63
99
99
99
99
97
99
99
99
99
97
41
99
87
78
97
94
94
94
94
96
97
97
83
95
30
31
85
64
92
98
83
99
25
99


Hedychium flavescens isolate Zn 48.3
Hedychium coccineum isolate Zn 67.1
Hedychium flavescens
Hedychium spicatum (ZSF05)
Hedychium dekianum (ZSF02)
Hedychium coccineum (ZMN06)
Hedychium chrysoleucum (ZSF07)
Hedychium stenopetalum (ZSF04)
Hedychium flavum (ZSF03)
Hedychium coccineum (ZMN12)
Hedychium flavescens (ZAP16)
Hedychium forrestii (ZAP15)
Hedychium coronarium isolate Zn 8.2
Alpinia sp (ZSA08)
Hedychium coronarium (ZSF01)
Hedychium villosum
Hedychium horsfieldii
Hedychium longicornutum
Hedychium bordelonianum
Hedychium gardnerianum (ZAP08)
Hedychium greenii
Curcuma longa isolate Zn 45.7
Curcuma caesia isolate Zn 58.6
Curcuma amada isolate Zn 44.6
Curcuma sp Kress 01-6874
Curcuma comosa
Hitchenia glauca
Zingiber zerumbet (ZSH04)
Curcuma zedoaria (ZSD 10)
Curcuma sp (ZSD09)
Curcuma caesia (ZSD04)
Curcuma aromatica (ZSD 03)
Curcuma leucorrhiza (ZSD 08)
Curcuma angustifolia (ZSD02)
Curcuma amada (ZSD01)
Curcuma roscoeana
Alpinia sp (ZSA05)
Curcuma bicolor
Curcuma aeruginosa
Smithatris supraneanae
Curcuma thorelii
Stahlianthus involucratus
Pyrgophyllum yunnanense


Globba schomburgkii isolate Zn 18.1
Globba atrosanguinea
Globba curtisii
Globba orixensis isolate Zn 29.3
Globba ophioglossa isolate Zn 19.4
Globba macroclada
Alpinia sp (ZSA07)
Globba spathulata
Globba wengeri
Amomum subulatum isolate Zn 24.1
Paramomum petaloideum
Amomum pterocarpum isolate Zn 7.4
Amomum longipetiolatum
Amomum glabrum
Elettariopsis kerbyi
Elettariopsis stenosiphon
Elettariopsis sp Kress 00-6720
Hedychium sp (ZSF06)
Aframomum angustifolium
Aframomum daniellii
Aframomum sceptrum
Aframomum sp Waimea W86S127
Aframomum sp Kress 99-6368
Renealmia alpinia
Renealmia thyrsoidea
Renealmia sp Kress 96-5713
Renealmia cernua
Renealmia battenbergiana
Renealmia sp Harris RBGE
Alpinia nigra isolate Zn 14.3
Alpinia nigra (ZSA04)
Alpinia sp (ZMN08)
Alpinia sp (ZSA06)
Alpinia galanga isolate Zn 12.3
Alpinia galanga
Alpinia conchigera
Amomum ghaticum isolate Zn 33.2
Etlingera littoralis
Amomum villosum
Amomum sp Kress 99-6373
Alpinia elegans
Alpinia luteocarpa
Vanoverberghia sepulchrei
Alpinia vittata
Nicolaia elatior
Etlingera yunnanensis
Hornstedtia hainanensis

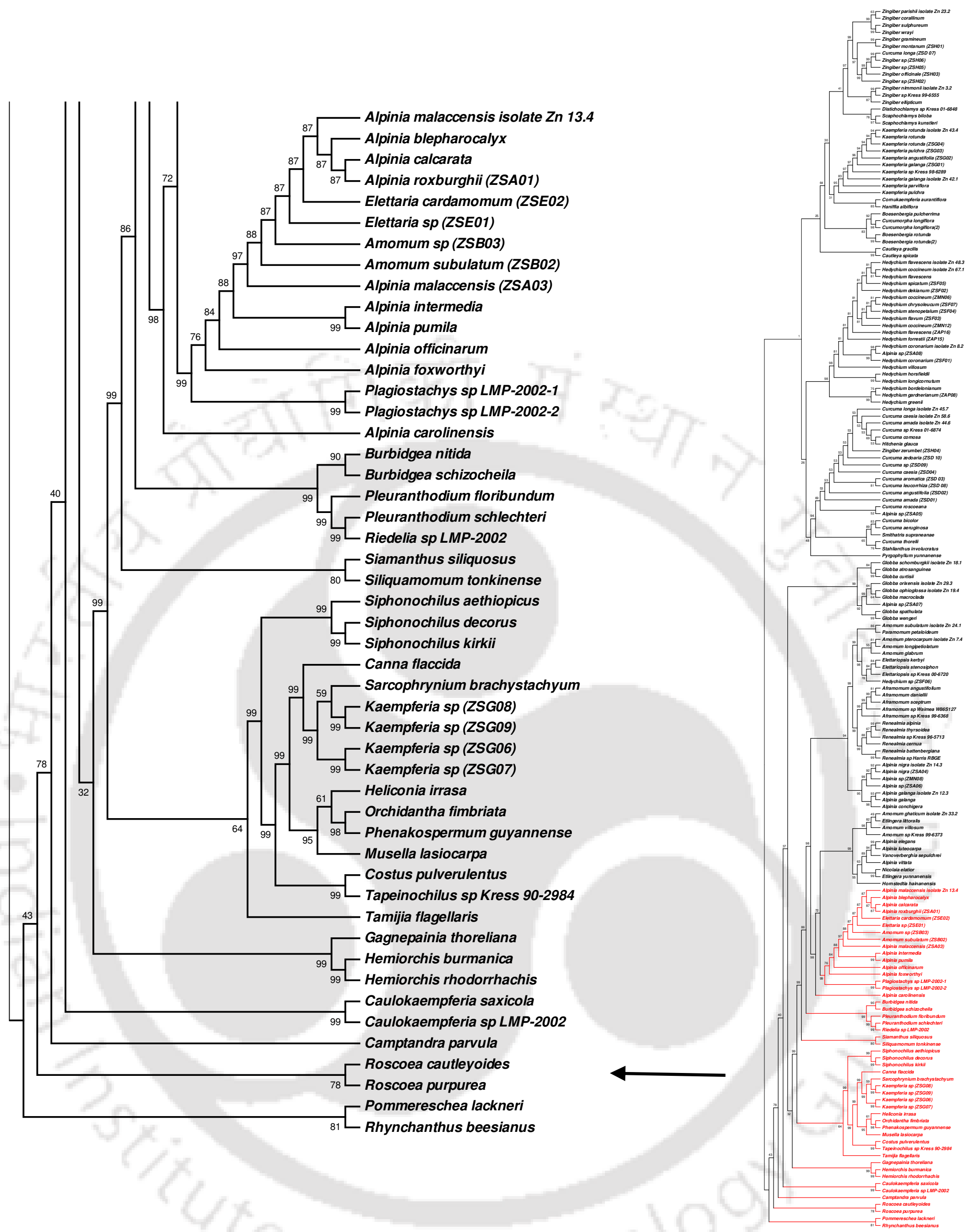


Figure 6.2. The evolutionary history for combined *matK* dataset was inferred using the MP method. The bootstrap consensus tree is shown which was inferred from 1000 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

6.3.2 *Zingiber*: Variation among geographically isolated species

The phylogenetic analyses of the second dataset unearthed an interesting and unique finding. The second dataset consisted of *Zingiber* species from our work, which have been collected from India and the *Zingiber* species sequences collected from the GenBank which on investigation was found to be samples from China. The phylogenetic analysis revealed that Zingiber species clustered together not on the basis of the species but rather the clustering was observed to be based on the location of the collected species (Figure 6.3). This strange and unique finding inkles towards that Zingiberaceae or at least *Zingiber* suffers from inter-species hybridization. For a conclusive proof, study of nuclear genome is suggested specifically for the species with such unique cases identified.

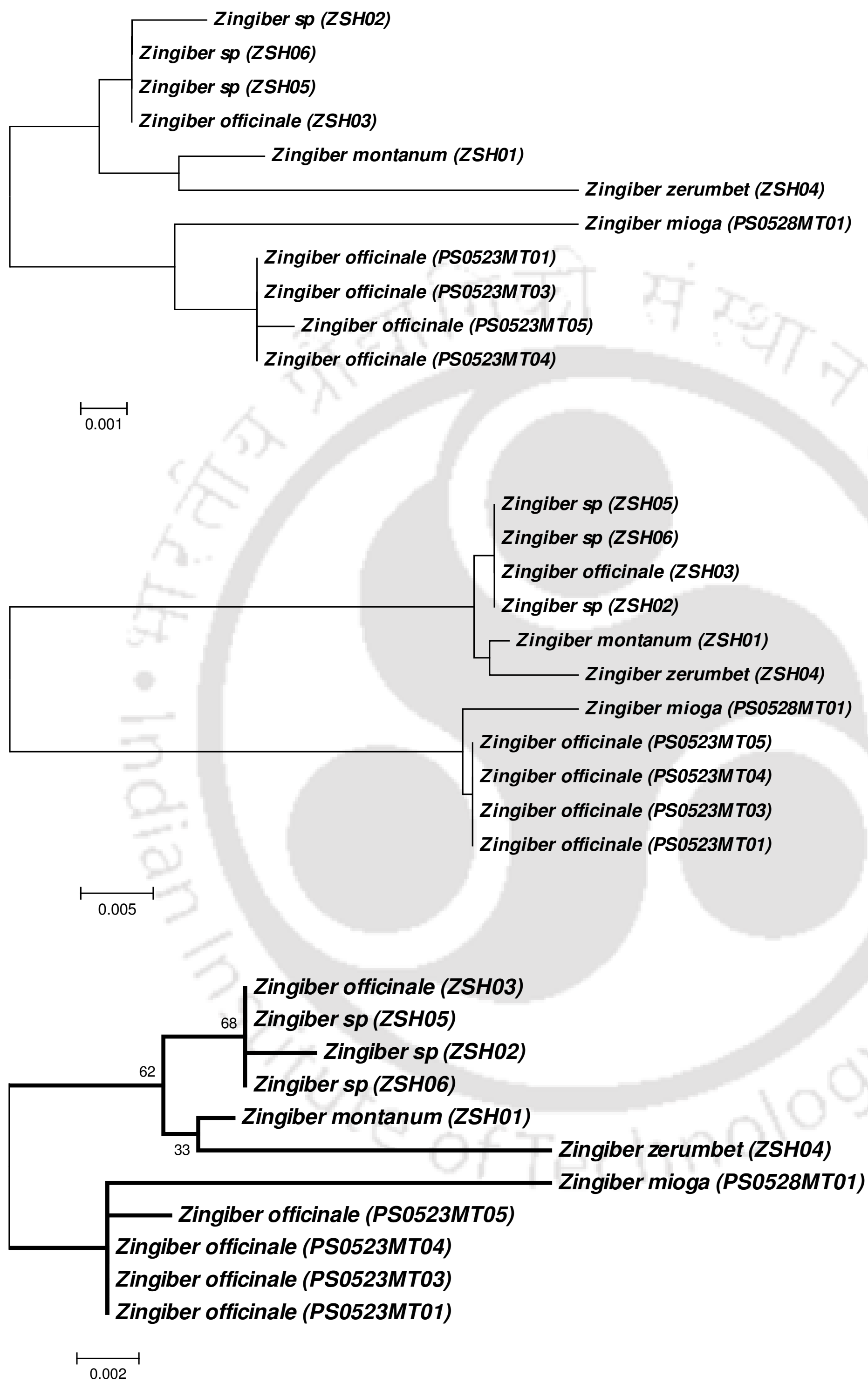


Figure 6.3. Phylogeny of Zingiber based on ML analysis. The Zingiber species with accession no starting with "Z" are from India while accession strating with "P" are from China. (A) *matK+psbK-psbI* (B) *rbcL+ matK* (C) *rpoB+psbK-psbI.*

6.4 Conclusion

Zingiberaceae is taxonomically a very complex family and for a more agreeable classification a large number of questions and phenomenon has to be understood. In this study relationship between *Curcuma-Hedychium-Zingiber* was resolved and the variation or closeness of species collected from a region irrespective of the taxa to which they belong was observed. *Hedychium* and *Zingiber* was found to be closer though by classical taxonomy they belonged to different tribes. A need of detailed phylogeographic study is required for understanding the complexities of the family.

CHAPTER 7

Concluding Remarks

The chapter deals with highlighting the salient features of this report along with discussing the future directions

Concluding remarks

7.1 Summary

In recent year the global losses in biodiversity and ongoing development pressures on the environment have led to responsible authorities and institutions to introduce policies which are aimed at achieving no net loss of biodiversity across areas (Gardner et al., 2013). In NEI the practice of shifting cultivation has also started to take its toll (Zonunsanga et al., 2014). Amidst the growing need of new bioactive molecules, the bio conservation is all the more need of the hour. Zingiberaceae being a plant family of medicinally important plant species, its conservation is of paramount nature. Along with the conservation the importance of documenting the medicinal usage in folklore is felt so that these valuable leads could be investigated. The conservation of the plant species is incumbent upon its identification which is in turn requires an agreeable classification.

In the present study excursion trips were made to the villages and remote places habited by native tribal people. The information regarding usage of Zingiberaceae in medicinal practice were recorded along with collection of the specimens. The information was documented and has been uploaded as an electronic resource. It was found that Zingiberaceae family members are very commonly used by native people for treating gastro-intestinal and pulmonary ailments with rhizome being the most frequently used plant part. Collected species were subjected to phylogeny study by utilizing the information from its plastid genome. The phylogeny study highlighted and resolved some crucial relationship between the Zingiberaceae genera. In order to understand deeper phylogenetic relationships, plastid genome of five species were sequenced and assembled.

The information was used for building phylogeny and development of new marker systems. A comparison across the plastid genome in the order Zingiberales a 3 kb indel was observed at the border of IRa and SSC in *Z. officnale* which possibly has been extended in *M. accuminata* by 4.7 kb. As the most of our sampling were restricted to NEI, hence the GenBank was mined for such sequences which would complement our dataset and would result in an improved phylogeny. Some very interesting observations were made. The species who are more likely to interact with each other cluster closely together irrespective of the geography. This is the first intensive ethno-medicinal survey for usefulness of Zingiberaceae along with its phylogeny. The plastid genome sequenced is first report of whole plastid genome being sequenced from Zingiberaceae for resolving phylogeny within the family. Hence this investigation is a potent effort towards improved utilization and conservation of the family.

The important outcomes of the study are:

1) A large number of members from the family was found to be regularly used as medicine by ethnic people (34 out of collected 52 plants). The rhizomes appeared as the most potent target for the screening of drug molecules as it was the most used plant parts. The use of plant parts as an abortifacient and for regulating the menstrual cycle is of special notice.

2) Use of eight loci and their combination for construction of phylogeny along with incorporation of the sequences from the GenBank helped in resolving some of the confounding relationships. It also highlighted the area which would require work in future with recommendations to how to proceed.

3) The information from plastid genome helped in resolving the *Curucma-Hedychium-Zingiber* relationship which is very significant since Curcuma and Zingiber are medicinally and economically important genera.

4) The development of markers from the plastid genomes (both the barcode and SSR marker) was seen to be highly informative by the *in silico* analysis and would be an indispensable tool for future classification and study of the family.

7.2 Future directions

The areas which are interesting to explore on the basis of the work carried out in this thesis are:

1) Collection of some of the obscure genera which was lacking representation or were under represented. The collection of genus *Amomum* and *Elettaria* from the NEI should be taken up along with the phylogeny study.

2) A phylogeography study of the species along with a detailed account of the geographical barrier would help in addressing many questions.

3) To address the occurrence of inter species hybridization nuclear loci which are single copy or low copy number should be mined and analysed.

4) Important ethno-medicinal leads should be investigated by isolating the molecules and performing bioactivity assays.

REFERENCES

REFERENCES

Alvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution, 29(3), 417–34.

Amit Kumar, M. (2010). HimFlorIS – an information system forflora in Himachal Pradesh, India. Current Science, 99(1), 98 – 101.

APG II. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: (2003). Botanical Journal of the Linnean Society, 141(4), 399–436.

Atherton, R. A., McComish, B. J., Shepherd, L. D., Berry, L. A., Albert, N. W., & Lockhart, P. J. (2010). Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant Methods, 6(1), 22.

Bailey, C. D., Carr, T. G., Harris, S. A., & Hughes, C. E. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics and Evolution, 29(3), 435–55.

Baker, W. J., Asmussen, C. B., Barrow, S. C., Dransfield, J., & Hedderson, T. A. (1999). A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trnL-trnF* region. Plant Systematics and Evolution, 219(1-2), 111–126.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., & Donoghue, M. J. (1995). The *ITS* Region of Nuclear Ribosomal DNA: A Valuable Source of Evidence on Angiosperm Phylogeny. Annals of the Missouri Botanical Garden, 82(2), 247.

Balunas, M. J., & Kinghorn, A. D. (2005). Drug discovery from medicinal plants. Life Sciences, 78(5), 431–41.

Barakat, A., Matassi, G., & Bernardi, G. (1998). Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. Proceedings of the National Academy of Sciences, 95(17), 10044–10049.

Barrett, C. F., Davis, J. I., Leebens-Mack, J., Conran, J. G., & Stevenson, D. W. (2013). Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics, 29(1), 65–87.

Barrett, C. F., Specht, C. D., Leebens-Mack, J., Stevenson, D. W., Zomlefer, W. B., & Davis, J. I. (2014). Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales). Annals of Botany, 113(1), 119–33.

Bendich, A. J. (2004). Circular chloroplast chromosomes: the grand illusion. The Plant Cell, 16(7), 1661–6.

Benditt, John (1989) "Cold water on the fire: a recent survey casts doubt on evidence for early use of fire." Scientific American, May 1989, pp. 21-22.

Benz, B. F., Cevallos E., J., Santana M., F., Rosales A., J., & Graf M., S. (2000). Losing knowledge about plant use in the sierra de manantlan biosphere reserve, Mexico. Economic Botany, 54(2), 183–191.

Bergthorsson, U., Richardson, A. O., Young, G. J., Goertzen, L. R., & Palmer, J. D. (2004). Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proceedings of the National Academy of Sciences of the United States of America, 101(51), 17747–52.

Berkes, Fikret. (1993). Traditional ecological knowledge in perspective. inInglis. (ed.), Traditional ecological knowledge: concepts and cases. Ottawa: International Program on Traditional Ecological Knowledge, and International Development Research Center, 1-10.

Gardner, T.A., Amrei Von Hase, Susie Brownlie, Jonathan M. M. Ekstrom, John D. Pilgrim, Conrad E. Savy, R. T. Theo Stephens7, Jo Treweek, Graham T. Ussher, Gerri Ward & Kerry Ten Kate. Biodiversity Offsets and the Challenge of Achieving No Net Loss Volume 27, Issue 6, pages 1254–1264, December 2013

Borsch, T., & Quandt, D. (2009). Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Systematics and Evolution, 282(3-4), 169–199.

Bortiri, E., Coleman-Derr, D., Lazo, G. R., Anderson, O. D., & Gu, Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. BMC Research Notes, 1(1), 61.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 32(3), 314–31.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R., Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457.

Bowe, L. M., Coat, G., & dePamphilis, C. W. (2000). Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proceedings of the National Academy of Sciences, 97(8), 4092–4097.

Bremer, B., Bremer, K., Chase, M., Fay, M., Reveal, J., Soltis, D., & Stevens, P. (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(2), 105–121.

Burtt B.L., & Smith, R.M. (1972). Tentative keys to the subfamilies, tribes and genera of the Zingiberales. Notes from the Royal Botanic Garden Edinburgh, 31, 171-176.

Burtt, B. L. 1986. General introduction to papers on Zingiberaceae. Notes Roy. Bot. Gard. Edinburgh 31: 155-165.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C., Jaffe, D.B. (2008) ALLPATHS: De novo assembly of whole-genome shotgun microreads. Genome Res. 18:810–820.

Butler, M. S. (2004). The role of natural product chemistry in drug discovery. Journal of Natural Products, 67(12), 2141–53.

Cesalpino, A. (1583). De Plantis libri xvi. Florence: Marescot.

Chaisson, M. J., & Pevzner, P. A. (2008). Short read fragment assembly of bacterial genomes. Genome Research, 18(2), 324–30.

Chan, W.C.E., Lim, Y.Y., & Wong, S.K. (2011). Antioxidant properties of ginger leaves: An overview. Faculty of Applied Sciences, UCSI University, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia School of Science, Monash University Sunway Campus, Bandar Sunway, 46150 Petaling Jaya, Selangor, Malaysia, 1(1): 6–16.

Chang, Y., & Graham, S.W. (2011). Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. American Journal of Botany, 98, 839-849.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., Berg, C. van den, Madriñán, S., Petersen, G., & Wilkinson, M. (2007). A Proposal for a Standardised Protocol to Barcode All Land Plants. Taxon, 56(2), 295-299.

Chevreux, B., Wetter, T., Suhai, S. (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computational Science. Biology.: Proc. German Conference on Bioinformatics GCB'99 GCB, pp. 45–56.

Chiang, T.Y., and Barbara A. Schaal. Molecular evolution and phylogeny of the atpB– rbcL spacer of chloroplast DNA in the true mosses. Genome43: 417–426 (2000)

Chiba, Y. (1951) Cytochemical studies on chloroplasts. I. Cytologic demonstration of nucleic acids in chloroplasts. Cytologia (Tokyo) 16: 259–264

Clegg, M. T., & G. Zurawski. (1991). Chloroplast DNA and the study of plant phylogeny. Pp. 1- 13 in P. S. SOLTIS, D. E. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants. Chapman & Hall, New York.

Corl, A., & Ellegren, H. (2013). Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. Molecular Phylogenetics and Evolution, 67(2), 358–66.

Cronquist. (1981). An Integrated System of Classification of Flowering Plants. Columbia University. Press, New York.

Curtis, S., & Clegg, M. (1984). Molecular evolution of chloroplast DNA sequences. Molecular Biology Evolution, 1(4), 291–301.

D'Hont, A., Denoeud, F., Aury, J.-M., Baurens, F.-C., Carreel, F., Garsmeur, O., & Wincker, P. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature, 488(7410), 213–7.

Dane, F., & Lang, P. (2004). Sequence variation at cpDNA regions of watermelon and related wild species: implications for the evolution of *Citrullus* haplotypes. American Journal of Botany, 91(11), 1922–9.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature Methods, 9(8), 772–772.

Das, A., Kesari, V., & Rangan, L. (2010). Aromatic Joha Rice Of Assam-A Review. Agricultural Reviews, 31(1), 1–10.

Davis, C. C., Anderson, W. R., & Donoghue, M. J. (2001). Phylogeny of Malpighiaceae: evidence from chloroplast *ndhF* and *trnl-F* nucleotide sequences. American Journal of Botany, 88(10), 1830–46.

De Natale, A., Pezzatti, G. B., & Pollio, A. (2009). Extending the temporal context of ethnobotanical databases: the case study of the Campania region (southern Italy). Journal of Ethnobiology and Ethnomedicine, 5(1), 7.

De Queiroz, A., Donoghue, M. J., & Kim, J. (1995). Separate Versus Combined Analysis of Phylogenetic Evidence. Annual Review of Ecology and Systematics, 26(1), 657–681.

Degradation of Land and Forest Resources: The Story of Shifting Cultivation and Loss of Biodiversity in North-East India R. Zonunsanga, Ch. Udaya Bhaskara Rao, P. Rinawma Climate Change and Biodiversity Advances in Geographical and Environmental Sciences 2014, pp 259-267

Dobeš, C., & Paule, J. (2010). A comprehensive chloroplast DNA-based phylogeny of the genus *Potentilla* (Rosaceae): Implications for its geographic origin, phylogeography and generic circumscription. Molecular Phylogenetics and Evolution, 56(1), 156–175.

Dong, W., Liu, J., Yu, J., Wang, L., Zhou, S. (2012) Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. PLoS One 7:e35071.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797.

Erickson, L. R., Straus, N. A., & Beversdorf, W. D. (1983). Restriction patterns reveal origins of chloroplast genomes in Brassica amphiploids. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 65(3), 201–6.

Erickson, L. R., Straus, N. A., & Beversdorf, W. D. (1983). Restriction patterns reveal origins of chloroplast genomes in *Brassica amphiploids*. Theoretical and Applied Genetics, 65(3), 201–206.

Fischer C.E.C. 1956. Flora of the Presidency of Madras. Part VIII. Ulmaceae to Xypridaceae. Government of India

Gaina, C., Müller, R. D., Brown, B., Ishihara, T., & Ivanov, S. (2007). Breakup and early seafloor spreading between India and Antarctica. Geophysical Journal International, 170(1), 151–169. doi:10.1111/j.1365-246X.2007.03450.x

Gao, J.Y., Yang, Z.H., Ren, P.Y., & Li, Q.J. (2006) Reproductive ecology of Rhynchanthus beesianus W.W. Smith (Zingiberaceae) in South Yunnan, China: A ginger with bird pollination syndrome. Journal of Integrative Plant Biology, 48, 1294 – 1299.

Gauthier, J., Annatella, D.C., Ueiroz, K.,De.Q., Luge, A.G.K, & Owe, T.R. (1989). Tetrapod phylogeny. In The hierarchy of life (B. Fernholm, K. Bremer, and H. Jornvall, eds.). Elsevier, Amsterdam, 337 – 353.

Ghosh, S., Padilla-González, G. F., & Rangan, L. (2013). Alpinia nigra seeds: A potential source of free radical scavenger and antibacterial agent. Industrial Crops and Products, 49, 348–356.

Gielly, L., & Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. Mol. Biol. Evol., 11(5), 769–777.

Gilli, C., He, Z., But, P., Schinnerl, J., Valant Vetschera, K., & Greger, H. (2011). Chemodiversity and biological activity of the genus Alpinia (Zingiberaceae). Planta Medica, 77(12), PG71.

Giurfa, M., Dafni, A., & Neal, P. (1999). Floral Symmetry and Its Role in Plant-Pollinator Systems. International Journal of Plant Sciences, 160(S6), S41–S50.

Goodner, K.L., Rouseff, R.L., & Hofsommer, H.J. (2001). Orange, Mandarin, and Hybrid Classification Using Multivariate Statistics Based on Carotenoid Profiles . Journal of Agricultural Food Chemistry, 49(3), 1146–1150.

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. Mol Biol Evol 26: 99–110.

Goremykin, V. V, Hirsch-Ernst, K. I., Wolfl, S., & Hellwig, F. H. (2003). Analysis of the Amborella trichopoda chloroplast genome sequence suggests that amborella is not a basal angiosperm. Molecular Biology and Evolution, 20(9), 1499–505. doi:10.1093/molbev/msg159

Goremykin, V. V, Hirsch-Ernst, K. I., Wölfl, S., & Hellwig, F. H. (2004). The chloroplast genome of Nymphaea alba: whole-genome analyses and the problem of identifying the most basal angiosperm. Molecular Biology and Evolution, 21(7), 1445–54. doi:10.1093/molbev/msh147

Gowda, V., Kress, W. J., & Htun, T. (2012). Two new species of Gingers (Zingiberaceae) from Myanmar. PhytoKeys, 13(13), 5–14. doi:10.3897/phytokeys.13.2670

Hall, J. C., Sytsma, K. J., & Iltis, H. H. (2002). Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. American Journal of Botany, 89(11), 1826–42. doi:10.3732/ajb.89.11.1826

Hamby, R. K., & Zimmer, E. A. (1988). Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Plant Systematics and Evolution, 160(1-2), 29–37. doi:10.1007/BF00936707

Heinrich, M., & Lee Teoh, H. (2004). Galanthamine from snowdrop--the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. Journal of Ethnopharmacology, 92(2-3), 147–62. doi:10.1016/j.jep.2004.02.012

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. PloS One, 6(5), e19254. doi:10.1371/journal.pone.0019254

Holttum R.E. (1950). The Zingiberaceae of the Malay Peninsula. Gardener's Bulletin of Singapore 13, 1-249.

Hong-lin, Z., Zhi-qiong, W., Lin, W., Hong-ying, Z., Kan-zhi, Z., & Yun, D. (2010). The Construction of the Ethnomedicinal Database: A Case Study of the Construction of the Databank of Pattra Medicine (Editorial Department), Journal of Yunnan University of Nationalities, Natural Sciences edition, Kunming 650031, China.

Hu, J.-M., Lavin, M., Wojciechowski, M. F., & Sanderson, M. J. (2000). Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast trnK/matK sequences and its implications for evolutionary patterns in Papilionoideae. Am. J. Botany, 87(3), 418–430.

Huang, Y.-Y., Matzke, A. J. M., & Matzke, M. (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (Cocos nucifera). PloS One, 8(8), e74736. doi:10.1371/journal.pone.0074736

Ignacimuthu, S., Ayyanar, M., & Sivaraman K, S. (2006). Ethnobotanical investigations among tribes in Madurai District of Tamil Nadu (India). Journal of Ethnobiology and Ethnomedicine, 2(1), 25. doi:10.1186/1746-4269-2-25

Igoli, J., Ogaji, O., Tor-Anyiin, T., & Igoli, N. (2005, March 22). Traditional Medicine Practice amongst the Igede People of Nigeria. Part II. African Journal of Traditional, Complementary and Alternative Medicines. African Ethnomedicines Network (Nigeria). doi:10.4314/ajtcam.v2i2.31112

Iwu, M.M., & Wootton, J. (2002). Ethnomedicine and Drug Discovery. Elsevier (p. 344)

Jacobsen, F., & Omland, K. E. (2011). Species tree inference in a recent radiation of orioles (Genus Icterus): multiple markers and methods reveal cytonuclear discordance in the northern oriole group. Molecular Phylogenetics and Evolution, 61(2), 460–9. doi:10.1016/j.ympev.2011.06.017

Jain S.K., & Prakash, V. (1995) 22 genera and 178 species in India concentrated mainly in North eastern and the Peninsular region (Zingiberaceae in India) Phytogeography and endemism. Rheedea, 5(2), 154-169.

Jansen, R. K., Raubeson, L. A., Boore, J. L., dePamphilis, C. W., Chumley, T. W., Haberle, R. C., & Cui, L. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in Enzymology, 395, 348–84. doi:10.1016/S0076-6879(05)95020-9

Jantan, I.B., Yassin, I.B., Chin, C.B., Chen, L.L., Sim, N.L. (2003). Antifungal activity of the essential oils of nine Zingiberaceae species. Pharmaceutical Biology, 41, 392–397.

Katiyar, C., Gupta, A., Kanjilal, S., & Katiyar, S. (2012). Drug discovery from plant sources: An integrated approach. Ayu, 33(1), 10–9. doi:10.4103/0974-8520.100295

Kellermann, J., & Udovicic, F. (2008). Large indels obscure phylogeny in analysis of chloroplast DNA (trn L-F) sequence data: Pomaderreae (Rhamnaceae) revisited. Telopea, Royal Botanic Gardens and Domain Trust, 12(1) 1–22

Kim, K.J., Jansen, R.K., Wallace, R. S., Michaels, H.J., & Palmer, J.D. (1992). Phylogenetic implications of *rbcL* sequence variation in the Asteraceae. Annals of Missouri Botanical Garden, 79, 428-445.

Kim, S.-C., Crawford, D. J., Jansen, R. K., & Santos-Guerra, A. (1999). The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribeSonchinae (Asteraceae:Lactuceae). Plant Systematics and Evolution, 215(1-4), 85–99. doi:10.1007/BF00984649

Klaus Kubitzki, H. Huber Flowering Plants. Monocotyledons: Alismatanae and Commelinanae (except Gramineae). (1998) (p. 511). Springer.

Klootwijk, C. T., Gee, J. S., Peirce, J. W., Smith, G. M., & McFadden, P. L. (1992). An early India-Asia contact: Paleomagnetic constraints from Ninetyeast Ridge, ODP Leg 121. Geology, 20(5), 395–398. doi:10.1130/0091-7613(1992)020

Kolodner, R., & Tewari, K. K. (1979). Inverted repeats in chloroplast DNA from higher plants. Proceedings of the National Academy of Sciences of the United States of America, 76(1), 41–5.

Kress, W. J. (1990). The Phylogeny and Classification of the Zingiberales. Annals of the Missouri Botanical Garden, 77(4), 698. doi:10.2307/2399669

Kress, W. J., & Larsen, K. (2001). Smithatris, a New Genus of Zingiberaceae from Southeast Asia. Systematic Botany, 26(2), 226–230. doi:10.1043/0363-6445-26.2.226

Kress, W. J., D Mood, J., Sabu, M., Prince, L. M., Dey, S., & Sanoj, E. (2010). Larsenianthus, a new Asian genus of Gingers (Zingiberaceae) with four species. PhytoKeys, 1(1), 15–32. doi:10.3897/phytokeys.1.658

Kress, W. J., Liu, A.-Z., Newman, M., & Li, Q.-J. (2005). The molecular phylogeny of *Alpinia* (Zingiberaceae): a complex and polyphyletic genus of gingers. American Journal of Botany, 92(1), 167–78. doi:10.3732/ajb.92.1.167

Kress, W. J., Prince, L. M., & Williams, K. J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. American Journal of Botany, 89(10), 1682–96.

Kumar, R., Mood, J., Singh, S. K., & Sinha, B. K. (2013). A new species of *Zingiber* (Zingiberaceae) from Northeast India. Phytotaxa, 77(4), 61–64. doi:10.11646/phytotaxa.77.4.2

Kuzoff, R. K., Sweere, J. A., Soltis, D. E., Soltis, P. S., & Zimmer, E. A. (1998). The phylogenetic potential of entire 26S rDNA sequences in plants. Molecular Biology and Evolution, 15(3), 251–63.

Lambkin, C. L., Sinclair, B.J., Pape, T., Courteney, G.W., Skevington, J.H., Meier, R., Yeates, D.K., Blagoderov, V., & Wiegmann, B.M. (2013). The phylogenetic relationships among infraorders and superfamilies of Diptera based on morphological evidence. Systematic Entomology, 38(1): 164–79.

Lamxay, V., & Newman, M. (2012). A revision of Amomum (Zingiberaceae) in Cambodia, Laos and Vietnam. Edinburgh Journal of Botany, 69(1).

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. & Higgins D.G. (2007) ClustalW and ClustalX version 2. Bioinformatics, 23(21), 2947-2948.

Larsen, K.J., Lock, M., Maas, H., Maas, P.J.M. (1998). Zingiberaceae. In K. Kubitzki [ed.], The families and genera of vascular plants, Springer-Verlag, Berlin, Germany, IV, 474–495.

Leebens-Mack, J., Raubeson, L. A., Cui, L., Kuehl, J. V, Fourcade, M. H., Chumley, T. W., … depamphilis, C. W. (2005). Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Molecular Biology and Evolution, 22(10), 1948–63. doi:10.1093/molbev/msi191

Leong-Škorničková, J., Lý, N.-S., Poulsen, A. D., Tosh, J., & Forrest, A. (n.d.). Newmania: A new ginger genus from central Vietnam. International Association for Plant Taxonomy.

Leong-Skornicková, J., Sída, O., Jarolímová, V., Sabu, M., Fér, T., Trávnícek, P., & Suda, J. (2007). Chromosome numbers and genome size variation in Indian species of Curcuma (Zingiberaceae). Annals of Botany, 100(3), 505–26. doi:10.1093/aob/mcm144

Leroux H. J Chem Med 1830;6:341.

Lescot, M., Piffanelli, P., Ciampi, A. Y., Ruiz, M., Blanc, G., Leebens-Mack, J., … Town, C. D. (2008). Insights into the Musa genome: syntenic relationships to rice and between Musa species. BMC Genomics, 9, 58. doi:10.1186/1471-2164-9-58

Lewis, P.O. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Systems Biology, 50(6): 913–25.

Li, R., Ma, P.-F., Wen, J., & Yi, T.-S. (2013). Complete sequencing of five araliaceae chloroplast genomes and the phylogenetic implications. PloS One, 8(10), e78568. doi:10.1371/journal.pone.0078568

Logacheva, M. D., Penin, A. A., Valiejo-Roman, C. M., & Antonov, A. S. (2009). Structure and evolution of junctions between inverted repeat and small single copy regions of chloroplast genome in non-core Caryophyllales. Molecular Biology, 43(5), 757–765. doi:10.1134/S0026893309050070

Löhne, C., & Borsch, T. (2005). Molecular evolution and phylogenetic utility of the petD group II intron: a case study in basal angiosperms. Molecular Biology and Evolution, 22(2), 317–32. doi:10.1093/molbev/msi019

Lu, C.-L., Zhao, H.-Y., & Jiang, J.-G. (2013). Evaluation of multi-activities of 14 edible species from Zingiberaceae. International Journal of Food Sciences and Nutrition, 64(1), 28–35. doi:10.3109/09637486.2012.694852

Lu, Chuan-Li; Zhao, Hai-Yan; Jiang, Jian-Guo (2013)Evaluation of multi-activities of 14 edible species from Zingiberaceae Informa HealthcareLondon

Maas, P. J. M. 1977. Flora Neotropica Monograph No. 18 Renealmia (Zingiberaceae - Zingiberoideae), Costoideae (Additions) (Zingiberaceae). New York Botanical Garden, New York.

Martin, G., Baurens, F.-C., Cardi, C., Aury, J.-M., & D'Hont, A. (2013). The complete chloroplast genome of banana (Musa acuminata, Zingiberales): insight into plastid monocotyledon evolution. PloS One, 8(6), e67350. doi:10.1371/journal.pone.0067350

Mary, J. A., Priyadharshini, K. C., Amal, G. P. R., Ramya, G., Nithya, R., Ambika, M. B., & Shenbagarathai, R. (2012). MEDDB: a medicinal plant database developed with the information gathered from tribal people in and around Madurai, Tamil Nadu. Bioinformation, 8(8), 391–3. doi:10.6026/97320630008391

Matassi, G., Montero, L. M., Salinas, J., & Bernardi, G. (1989). The isochore organization and the compositional distribution of homologous coding sequences inthe nuclear genome of plants. Nucleic Acids Research, 17(13), 5273–5290. doi:10.1093/nar/17.13.5273

Matsuoka, Y., Yamazaki, Y., Ogihara, Y., & Tsunewaki, K. (2002). Whole Chloroplast Genome Comparison of Rice, Maize, and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. Molecular Biology and Evolution, 19(12), 2084–2091. doi:10.1093/oxfordjournals.molbev.a004033

McCauley, D. E., Sundby, A. K., Bailey, M. F., & Welch, M. E. (2007). Inheritance of chloroplast DNA is not strictly maternal in Silene vulgaris (Caryophyllaceae): evidence from experimental crosses and natural populations. American Journal of Botany, 94(8), 1333–7. doi:10.3732/ajb.94.8.1333

Megarry, Tim (1995) Society in Prehistory: The Origins of Human Culture. New York: New York University Press

Mogensen, H. L. (2008). The Hows and Whys of Cytoplasmic Inheritance in Seed Plants.

Morgan, K., O'Loughlin, S. M., Chen, B., Linton, Y.-M., Thongwat, D., Somboon, P., … Walton, C. (2011). Comparative phylogeography reveals a shared impact of pleistocene environmental change in shaping genetic diversity within nine Anopheles mosquito species across the Indo-Burma biodiversity hotspot. Molecular Ecology, 20(21), 4533–49. doi:10.1111/j.1365-294X.2011.05268.x

Morrison, D. A. (2006). L. A. S. JOHNSON REVIEW No. 8. Multiple sequence alignment for phylogenetic purposes. Australian Systematic Botany, 19(6), 479. doi:10.1071/SB06020

Nakai T. 1941. Notulae ad plantas Asiae Orientalis (XVI). Japanese Journal of Botany 17 : 189 – 203

Nanjundaiah, S.M., Annaiah, H.N.M., & Dharmesh, S.M. (2011). Gastroprotective Effect of Ginger Rhizome (Zingiber officinale) Extract: Role of Gallic Acid and Cinnamic Acid in H+, K+-ATPase/H. pylori Inhibition and Anti-Oxidative Mechanism. Evidence-Based Complementary and Alternative Medicine, Volume 2011, Article ID 249487, 13 pages.

Nei, M., Kumar, S., & Takahashi, K. (1998). The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proceedings of the National Academy of Sciences, 95(21), 12390–12397. doi:10.1073/pnas.95.21.12390

Neyman, J. (1974). Molecular studies: A source of novel statistical problems. In: Gupta SS, Yackel J. (eds), Statistical Decision Theory and Related Topics, pp. 1–27. Academic Press, New York.

Ngamriabsakul, C., Newman, M. F., & Cronk, Q. C. B. (2000). Phylogeny And Disjunction In Roscoea (Zingiberaceae). Edinburgh Journal Of Botany, 57(01), 39–61.

Ngamriabsakul, C., Newman, M. F., & Cronk, Q. C. B. (2004). The Phylogeny Of Tribe Zingibereae (Zingiberaceae) Based On Its (nrdna) and trnl–F (cpdna) sequences. Edinburgh Journal of Botany, 60(03), 483–507.

Nishida, T. & Uehara, S. (1983). Natural diet of chimpanzees (Pantroglodytes schweinfurthii): Long-term record from the Mahale Mountains, Tanzania. African Study Monographs, 3: 109-30.

Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., & Bassil, N. (2013). Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. Molecular Phylogenetics and Evolution, 66(1), 17–29. doi:10.1016/j.ympev.2012.08.026

Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., & Henry, R. J. (2011). Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnology Journal, 9(3), 328–33. doi:10.1111/j.1467-7652.2010.00558.x

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., … Ozeki, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature, 322(6079), 572–574. doi:10.1038/322572a0

Omonike, O.O., Adekunle, A.J., Edith, A.O., & Festus, A.D. (2013). Anti-poliovirus activity of medicinal plants selected from the Nigerian ethno-medicine. African Journal of Biotechnology, 12(24), 3878-3883.

Oubré, A. Y., Carlson, T. J., King, S. R., & Reaven, G. M. (1997). From plant to patient: an ethnomedical approach to the identification of new drugs for the treatment of NIDDM. Diabetologia, 40(5), 614–7. doi:10.1007/s001250050724

Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. Nature, 301(5895), 92–93. doi:10.1038/301092a0

Palmer, J. D., & Olmstead, R. G. (1994). Chloroplast DNA systematics: A review of methods and data analysis. American Journal of Botany, 81.

Palmer, J. D., & Zamir, D. (1982). Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. Proceedings of the National Academy of Sciences, 79(16), 5006–5010.

Palmer, J. D., & Zamir, D. (1982). Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. Proceedings of the National Academy of Sciences, 79(16), 5006–5010. doi:10.1073/pnas.79.16.5006

Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qiu, Y.-L., & 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(13), 6960–6966. doi:10.1073/pnas.97.13.6960

Palmer, J. D., R. K. Jansen, H. J. Michaels, M. K. Chase, And J. R. Manhart. 1988. Chloroplast DNA variation and plant phylogeny. Ann. MO. Bot. Garden 75: 1180-1206.

Palmer, J.D., & Thompson, W.F. (1982). ChloroplastDNArearrangements are more frequent when a large inverted repeated sequence is lost. Cell, 29: 537-50.

Pedersen, L. B. (2004). Phylogenetic analysis of the subfamily Alpinioideae (Zingiberaceae), particularly Etlingera Giseke, based on nuclear and plastid DNA. Plant Systematics and Evolution, 245(3-4), 239–258. doi:10.1007/s00606-004-0126-2

Perrings, C. A., Mäler, K.-G., Folke, C., Holling, C. S., & Jansson, B.-O. (Eds.). (1995). Biodiversity Conservation (Vol. 4). Dordrecht: Springer Netherlands. doi:10.1007/978-94-011-0277-3

PETERSEN , O. G. 1889. Musaceae, Zingiberaceae, Cannaceae, Marantaceae. In A. Engler and K. Prantl [eds.], Die Natu ̈rlichen Pflanzenfamilien, 1st ed., vol. 2 (6), 1–43. Verlag von Wilhelm Engelmann, Leipzig, Germany.

Petersen, O.G. (1889). Musaceae, Zingiberaceae, Cannaceae, Marantaceae A. Engler, K. Prantl (Eds.), Die Natürlichen Pflanzenfamilien (1st ed.), vol. 2Verlag von Wilhelm Engelmann, Leipzig, Germany, 1.

Petit, R.J., & Vendramin G.G. (2007). Plant phylogeography based on organelle genes: An introduction. In S. Weiss and N. Ferrand [eds.], Phylogeography of southern European refugia, 23–97. Springer, Dordrecht, Netherlands.

Pirttila T, Wilcock G, Truyen L, Damaraju CV. Long-term efficacy and safety of galantamine in patients with mild-to-moderate Alzheimer's disease: multicenter trial. Eur J Neurol 2004;11:734-41.

Pirttilä, T., Wilcock, G., Truyen, L., & Damaraju, C. V. (2004). Long-term efficacy and safety of galantamine in patients with mild-to-moderate Alzheimer's disease: multicenter trial. European Journal of Neurology : The Official Journal of the European Federation of Neurological Societies, 11(11), 734–41. doi:10.1111/j.1468-1331.2004.00885.x

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Systematic Biology, 53(5), 793–808. doi:10.1080/10635150490522304

Powell, W., Morgante, M., Andre, C., McNicol, J. W., Machray, G. C., Doyle, J. J., … Rafalski, J. A. (1995). Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Current Biology, 5(9), 1023–1029.

Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G., & Rafalski, J. A. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proceedings of the National Academy of Sciences, 92(17), 7759–7763. doi:10.1073/pnas.92.17.7759

Prakash, V., & Mehrotra, B.N. (1995) Zingiberaceae of north-east India: Diversity and taxonomic status. In: Proceedings of the 2nd symposium on the family Zingiberaceae, 262-273.

Prasad, S., & Bharat, B (2011). Aggarwal.Turmeric, The Golden Spice of Life. Benzie, I.F.F., Wachtel, G.S. (eds). Herbal Medicine: Biomolecular and Clinical Aspects. 2nd edition. Boca Raton (FL): CRC Press.

R. M. SMITH. 1972. Tentative keys to the subfamilies, tribes and genera of Zingiberaceae. Notes Roy. Bot. Gard. Edinburgh 31: 171-176.

Rao , V . S. , Karnik , H . & Gupte , K . (1954) . The floral anatomy of some Scitamineae . I . J. Indian bot. Soc, 33 , 118

Ratan, A. (2010). Assembly algorithms for next-generation sequence data, DAI/B 71-0.

Raubeson, L. A., & Jansen, R.K. (1992a). A rare chloroplast-DNA mutation is shared by all conifers. Biochemical Systematics and Ecology, 20: 17-24.

Ravi, V., Khurana, J. P., Tyagi, A. K., & Khurana, P. (2007). An update on chloroplast genomes. Plant Systematics and Evolution, 271(1-2), 101–122. doi:10.1007/s00606-007-0608-0

Ravindran, P.N., Babu, K.N., & Sivaraman, K. (eds). (2007). Turmeric, The genus Curcuma. CRC Press, Print ISBN: 978-0-8493-7034-2 eBook ISBN: 978-1-4200-0632-2.

Reddy, P. R., Sudhakar, C., Kumar, J. N., & Das, B. (2013). The First Stereoselective Total Synthesis of Naturally Occurring, Bioactive (3 R ,5 R)-1-(4-Hydroxyphenyl)-7-

phenylheptane-3,5-diol and the Synthesis of Its Enantiomer. Helvetica Chimica Acta, 96(2), 289–295. doi:10.1002/hlca.201200165

Ridley, H. (1899). The Scitamineae of the Malay Peninsula. Singapore.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature, 499(7459), 431–7. doi:10.1038/nature12352

S.S. Rao Indigenous knowledge organization: An Indian scenario International Journal of Information Management, 26 (2006), pp. 224–233
S.S. Rao Indigenous knowledge organization: An Indian scenario International Journal of Information Management, 26 (2006), pp. 224–233
Sabu, M. (2005). The recircumscription of *Curcuma* L. to include the genus Paracautleya RM Sm. Gard. Bull. Singapore.

Sang, T., Crawford, D., & Stuessy, T. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). Am. J. Botany, 84(8), 1120–.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes, 6(5), 283–90.

Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., De Bruijn, A.Y., Sullivan, S., & Qiu, Y.L. (2000). Phylogenetics of Flowering Plants Based on Combined Analysis of Plastid *atpB* and *rbcL* Gene Sequences. Systematic Biology, 49(2), 306-362.

Schmitz-Linneweber, C., Maier, R. M., Alcaraz, J. P., Cottet, A., Herrmann, R. G., & Mache, R. (2001). The plastid chromosome of spinach (Spinacia oleracea): complete nucleotide sequence and gene organization. Plant Molecular Biology, 45(3), 307–15.

Schumann , K. 1904. Zingiberaceae. In A. Engler [ed.], Das Pflanzenreich 4/46, 1–458. Leipzig, Germany.

Shackleton, S. E., Shackleton, C. M., Netshiluvhi, T. R., Geach, B. S., Ballance, A., & Fairbanks, D. H. K. (2002). Use Patterns and Value of Savanna Resources in Three Rural Villages in South Africa1. Economic Botany, 56(2), 130–146.

Sharma, Ricky A.; McLelland, Heather R.; Hill, Kirsti A.; Ireson, Christopher R.; Euden, Stephanie A. et al. (2001)Pharmacodynamic and Pharmacokinetic Study of Oral Curcuma Extract in Patients with Colorectal Cancer. Clin. Cancer Res. 7 (7) p. 1894-1900

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., & Small, R. L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany, 92(1), 142–66.

Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany, 94(3), 275–88.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135–45.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., … Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. The EMBO Journal, 5(9), 2043–2049.

Signorini, M. A., Piredda, M., & Bruschi, P. (2009). Plants and traditional knowledge: an ethnobotanical investigation on Monte Ortobene (Nuoro, Sardinia). Journal of Ethnobiology and Ethnomedicine, 5(1), 6.

Sigrist, M. S., Pinheiro, J. B., Ilho, J. A. A. Z., Colombo, C. A., Bajay, M. M., Lima, P. F., … Zucchi, M. I. (2009). Development and characterization of microsatellite markers for turmeric (Curcuma longa). Plant Breeding, 21(5), 570–573.

Singh, S. K., & Kumar, R. (2011). *Globba wardii* (Zingiberaceae) – a Very Rare Species, New Record for India. Taiwania, 56(3), 261-264.

Singh, Sailendra; Kumar, Jonnala Kotesh; Saikia, Dharmendra; Shanker, Karuna; Thakur, Jay Prakash et al. (2010)A bioactive labdane diterpenoid from Curcuma amada and its semisynthetic analogues as antitubercular agents European Journal of Medicinal Chemistry vol. 45 (9) p. 4379-4382

Skoczen S, Bussmann RW: ebDB – Filling the gap for an international Ethnobotany Database. [http://www.quantumimagery.com/software-ebDB.html] Lyonia 2006, 11(2):71-81

Skornickova, J., & Sabu, M. (2005). Curcuma roscoeana Wall.(Zingiberaceae) in India. Gard. Bull. Singapore.

Small, R.L., Cronn, R.C., & Wendel, J.F. (2004). Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany, 17: 145–70.

Smith, J., Kress, W., & Zimmer, E. Phylogenetic Analysis of the Zingiberales Based on rbcL Sequences, Annals of the Missouri Botanical Garden (1993).

Soltis DEet al.(1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Annals of the Missouri Botanical Garden 84 , 1–49.

Soltis PS, Soltis DE (1998) Molecular evolution of 18S rDNA in angiosperms: implications for character weighting in phylogenetic analysis. In ‘Molecular systematics of plants II: DNA sequencing’. (Eds DE Soltis, PS Soltis, JJ Doyle) (Kluwer Academic Publishers: Boston, MA)

Soltis, D. E., Nickrent, P. S., Johnson, D. L., Hahn, L.A., W. J., Hoot, S. B., & Sytsma, K. J. (1997). Angiosperm Phylogeny Inferred from 18S Ribosomal DNA Sequences. Annals of the Missouri Botanical Garden, 84(1), 1.

Soltis, D. E., D. R. Morgan, A. Grable, P. S. Soltis, And R. Kuzoff. 1993. Molecular systematics of Saxifragaceae sensu stricto. Am. J. Bot. 80: 1056- 1081

Steele KP, Holsinger KE, Jansen RK, Taylor DW (1991) Assessing the reliability of 5S ribosomal-RNA sequence data for phylogenetic analysis in green plants. Molecular Biology and Evolution 8 , 240–248

Stone, E. (1763). An account of the success of the bark of the willow in the cure of agues. Philosophical Transactions Royal Society of London, 53:195-200.

Sucher, N. J. (2013). The application of Chinese medicine to novel drug discovery. Expert Opinion on Drug Discovery, 8(1), 21–34. doi:10.1517/17460441.2013.739602

Sugiura, M. (2002). History of chloroplast genomics. In Discoveries in Photosynthesis. Springer Netherlands 1057-1063.

Suthisut, Duangsamorn; Fields, Paul G.; Chandrapatya, Angsumarn (2011).Contact Toxicity, Feeding Reduction, and Repellency of Essential Oils from Three Plants from the Ginger Family (Zingiberaceae) and their Major Components Against Sitophilus zeamais and Tribolium castaneum Journal of Economic Entomology vol. 104 (4) p. 1445-1454

Takahashi, K., & Nei, M. (2000). Efficiencies of Fast Algorithms of Phylogenetic Inference Under the Criteria of Maximum Parsimony, Minimum Evolution, and Maximum Likelihood When a Large Number of Sequences Are Used. Molecular Biology and Evolution, 17(8), 1251–1258. doi:10.1093/oxfordjournals.molbev.a026408

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Method. Molecular Biology and Evolution, 28, 2731-2739.

Tanti, B., Gurung, L., Sarma, H.K., & Buragohain, A.K. (2010). Ethnobotany of starter cultures used in alcohol fermentation by a few ethnic tribes of Northeast India. Indian Journal of Traditional Knowledge, 9(3), 463-466.

Tawatchai TANEE, P. M. (2008). Two new species of Curcuma (Zingiberaceae) used as cobra-bite antidotes.

Techaprasan, J., & Leong-Škorničková, J. (2011). Transfer of Kaempferia candida to Curcuma (Zingiberaceae) based on morphological and molecular data. Nordic Journal of Botany, 29(6), 773–779. doi:10.1111/j.1756-1051.2011.00970.x

Techaprasan, J., Klinbunga, S., Ngamriabsakul, C., & Jenjittikul, T. (2010). Genetic variation of Kaempferia (Zingiberaceae) in Thailand based on chloroplast DNA (psbA-trnH and petA-psbJ) sequences. Genetics and Molecular Research : GMR, 9(4), 1957–73. doi:10.4238/vol9-4gmr873

Teklehaymanot, T., & Giday, M. (2007). Ethnobotanical study of medicinal plants used by people in Zegie Peninsula, Northwestern Ethiopia. Journal of Ethnobiology and Ethnomedicine, 3:12.

Thai, J. (2012). Protective Effects of Curcumin on Gastric Inflammation and Liver Diseases. Thong-Ngam D (eds.) Gastroenterol Vol.2012:13 (2).

The Construction of the Ethnomedicinal Database:A Case Study of the Construction of the Databank of Pattra Medicine ZHUANG Hong-lin1,WAN Zhi-qiong1,WANG Lin1,ZHOU Hong-ying2,ZHANG Kan-zhi2,DAI Yun1(1.Editorial Department of Journal,Yunnan University of Nationalities,Kunming 650031,China;2.School of Chemistry and Biotechnology,Yunnan University of Nationalities,Kunming 650031,China)

Thomas, M. B., Lin, N., & Beck, H. H. (2001). A Database Model for Integrating and Facilitating Collaborative Ethnomedicinal Research.

Thomas, V. P., Sabu, M., & Chathurvedi, S. K. (2012). Amomum carnosum (Zingiberaceae): a new species from Nagaland, north-east India. Kew Bulletin, 67(3), 549–553. doi:10.1007/s12225-012-9391-3

Thongam, B., Konsam, B., & Sarangthem, N. (2013) *Globba sherwoodiana* (Zingiberaceae) – A new record for India from Manipur. Rheedea, 23(1) 34-36.

Tillyard R. J. 1921. A new classification of the order Perlaria. Canadian Entomologist, 53, 35-43.

Tomlinson, P. B. (1962). Phylogeny of the Scitamineae-morphological and anatomical considerations. Evolution, 16: 192-213.

Tseng, Y.-H., & Wang, C.-C. (2011). Alpinia oui (Zingiberaceae), a New Species from Taiwan. Novon: A Journal for Botanical Nomenclature, 21(2), 270–273. doi:10.3417/2009111

Tsumura, Y., Suyama, Y., & Yoshimura, K. (2000). Chloroplast DNA Inversion Polymorphism in Populations of Abies and Tsuga. Molecular Biology and Evolution, 17(9), 1302–1312. doi:10.1093/oxfordjournals.molbev.a026414

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

Udgaonkar. (2002). The recording of traditional knowledge: Will it prevent 'bio-piracy'? Current Science, 82:4, 413-19.

Upadhya, V., Hegde, H.V., Mesta, D., Belchad, S., Hampannavar, V., Kholkute, S.D. (2010). Digital database on ethno medicinal plants of Western Ghats. Current Science - Bangalore-, 99(12), 1658-1659.

Use of nuclear genes for phylogeny reconstruction in plants Randall L. Small A,D , Richard C. Cronn B and Jonathan F. Wendel C. Australian Systematic Botany 17, 145–170

V. P., T., Nissar, V. A. M., & Gupta, U. (2014). Amomum sabuanum (Zingiberaceae): A new species from Sikkim, India. Phytotaxa, 159(2), 122.

Vinitha, M. R., Kumar, U. S., Aishwarya, K., Sabu, M., & Thomas, G. (2014). Prospects for discriminating Zingiberaceae species in India using DNA barcodes. Journal of Integrative Plant Biology. doi:10.1111/jipb.12189

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., … Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23(21), 4407–4414. doi:10.1093/nar/23.21.4407

Walker, J. F., Zanis, M. J., & Emery, N. C. (2014). Comparative analysis of complete chloroplast genome sequence and inversion variation in Lasthenia burkei (Madieae, Asteraceae). American Journal of Botany, 101(4), 722–9. doi:10.3732/ajb.1400049

Weberling, F. (1992). Morphology of Flowers and Inflorescences (p. 405). CUP Archive.

Welsh, J., & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research, 18(24), 7213–8.

Williams, K. J., Kress, W. J., & Htun, T. (2003). A Striking New Epiphytic Hedychium (Zingiberaceae) From Myanmar With A Discussion On Some Anomalous Related Genera. Edinburgh Journal Of Botany, 60(01), 43–48. Doi:10.1017/S0960428603000040

Williams, K. J., Kress, W. J., & Manos, P. S. (2004). The phylogeny, evolution, and classification of the genus Globba and tribe Globbeae (Zingiberaceae): appendages do matter. American Journal of Botany, 91(1), 100–14. doi:10.3732/ajb.91.1.100

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America, 74(11), 5088–90.

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States of America, 84(24), 9054–8.

Wood, T. H., Whitten, W. M., & Williams, N. H. (2000). Phylogeny of *Hedychium* and related genera (zingiberaceae) based on its sequence data. Edinburgh Journal of Botany, 57(02), 261–270.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., … Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature, 462(7276), 1056–60.

Wu, T.L., & Larson, K. (2000). Zingiberaceae. In: Wu Raven (ed) Flora of China, vol 24. Missouri Botanical Garden Publishing, St. Louis, 360-363.

Wurdack, K. J., Hoffmann, P., & Chase, M. W. (2005). Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid RBCL and TRNL-F DNA sequences. American Journal of Botany, 92(8), 1397–420. doi:10.3732/ajb.92.8.1397

Xia, Y.M., Gao, X.M., & Li, Q.J. (2009). Identification and expression of floral organ homeotic genes from Alpinia oblongifolia (Zingiberaceae). Journal of Integrative Plant Biology, 51, 155 – 166.

Xia, Y.-M., Kress, W. J., & Prince, L. M. (2004b). Phylogenetic Analyses of *Amomum* (Alpinioideae: Zingiberaceae) Using ITS and matK DNA Sequence Data. Systematic Botany, 29(2), 334–344. doi:10.1600/036364404774195520

Yadav, S.K., Sah, A.K., Jha, R.K., Sah, P., & Shah, D.K. (2013). Turmeric (curcumin) remedies gastroprotective action Pharmacognosy Reviews, 7(13): 42–6.

Yang, J.-B., Tang, M., Li, H.-T., Zhang, Z.-R., & Li, D.-Z. (2013). Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evolutionary Biology, 13(1), 84. doi:10.1186/1471-2148-13-84

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., … Yu, J. (2010). The complete chloroplast genome sequence of date palm (Phoenix dactylifera L.). PloS One, 5(9), e12762. doi:10.1371/journal.pone.0012762

Yeh, H., Chuang, C.H., Chen, H.C., Wan, C.J., Chen, T. L., & Lin, L.Y. (2014). Bioactive components analysis of two various gingers (Zingiber officinale Roscoe) and antioxidant effect of ginger extracts - Food Science and Technology Volume 55(1): 329–34.

Yuliana, N. D., Iqbal, M., Jahangir, M., Wijaya, C. H., Korthout, H., Kottenhage, M., … Verpoorte, R. (2011). Screening of selected Asian spices for anti obesity-related bioactivities. Food Chemistry, 126(4), 1724–1729.

Záveská, E., Fér, T., Šída, O., Krak, K., Marhold, K., & Leong-Škorničková, J. (n.d.). Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata. International Association for Plant Taxonomy.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research, 18(5), 821–29.

Zwickl, D. J., & Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. Systematic Biology, 51(4), 588–98.

VISIBLE RESEARCH OUTCOME

VISIBLE RESEARCH OUTCOME

Journals

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

Vaughn J.N., Chaluvadi, S.R., ***Tushar***, Rangan, L. and Bennetzen, J.L. Whole plastome sequences of five major spices clarify recombination history in the Zingiberales and facilitate marker development (Manuscript under preparation)

Tushar, Chaluvadi, S.R., Bennetzen, J.L. and Rangan, L. Phylogeny of Zingiberaceae using plastid sequences. (Manuscript under preparation)

Conferences

Tushar, Aggarwal, S., Satyanarayana, V.M., Parida, A. and Rangan, L., 2010. Mining of Curcuma species from Assam using plastid specific DNA barcodes. National Conference on Biotechnology, Bioinformatics and Bioengineering. Dharmapuri

Tushar, Chaluvadi, S.R., Bennettzen, J., Rangan, L.,2011 Mining of Zingiberoideae plastid genome for assessment of phylogenetic relationships. Plant Genome Evolution. Amsterdam.

Database

NEZRC accessible at http://www.iitg.ac.in/lrangan/

Genebank submission

Genera	Accession number (Genebank)
Alpinia	KC597943-50, KC597844-51, KC597951-58, KC598059-66, KC597959-66, KC597967-74, KC597904-11, KC597896-903.
Amomum	KC597894-95, KC597852-53, KC597975-76 KC597977-78, JN180547-54, KC597979-80 KC597981-82, KC598057-58
Curcuma	JN180515-22, JN180523-30, JN180531-38 JN180539-46, JN180547-54, JN180555-562, JN180563-70
Elettaria	KC598073-74, JN180523-30, KC597983-84, KC597985-86, JN180547-54, KC597987-88, KC597989-90, KC597892-93
Hedychium	JN180571-82, JN180583-91, JN180503-14 JN180592-603, JN180604-15, JN180491-502 JN180616-26, KC597923-34
Kaempferia	KC597991-98, KC598047-54, KC597912-19 KC597874-81, KC597999-06, KC598007-14 KC598015-22, KC597882-89
Zingiber	KC598067-72, KC597856-61, KC598023-28 KC5988023-28, KC598029-34, KC598035-40, KC598041-46, KC597868-73

ANNEXURES

ANNEXURES

A3.1.

Interview form for collection of information from the traditional healers

i) Birth place
ii) Gender
iii) Ethnicity
iv) Village and District
v) Where has he/she spent most of the time growing up?
vi) Number of times visited modern city
vii) Occupation
viii) Learned healing from
- a) If from family member then relation to that family member?
- b) If from some other person, is the community same?
- c) How many generations have the healer have been practicing, from whom the art/science was learned

ix) How many patient does he/she treats from his community and other community
x) What disease he/she treats along with
- a) Symptoms
- b) Plants and preparation used
- c) Who makes the preparations?
- d) Days taken for the patient to recuperate?

xi) Does he/she grows the plant or take it from wild?
xii) Ask him/her to identify plants in field (if grown) and wild
xiii) Common name of the plant
xiv) From where does he collected plants
xv) Agricultural practices involved with plants
xvi) Any other use of the plants (as food, as other economically important usage)
xvii) Does he/she share the information with other healers
xviii) If not why?

A 4.1.

List of primers for the plastid loci used for the phylogenetic study.

Gene	Sequence (5′ 3′)	PCR Product (bp)
accD	AGTATGGGATCCGTAGTAGG TTTAAAGGATTACGTGGTAC	~400
matK	CGTACAGTACTTTTGTGTTTACGAG ACCCAGTCCATCTGGAAATCTTGGTTC	~850
rpoB	AAGTGCATTGTTGGAACTGG CCGTATGTGAAAAGAAGTATA	~425
rpoC1	GTGGATACACTTCTTGATAATGG TGAGAAAACATAAGTAAACGGGC	~525
rbcL	GTAAAATCAAGTCCACCRCG ATGTCACCACAAACAGAGACTAAAGC	~600
atpF- *atpH*	ACTCGCACACACTCCCTTTCC GCTTTTATGGAAGCTTTAACAAT	~750
psbK- *psbI*	TTAGCCTTTGTTTGGCAAG AGAGTTTGAGAGTAAGCAT	~750
trnH- *psbA*	GTTATGCATGAACGTAATGCTC CGCGCATGGTGGATTCACAATCC	~775

A 4.2.

The combination of the loci used for phylogenetic analyses. The order of concatenation is as shown in the respective cell.

	accD(+)	*rbcL*(+)	*rpoB*(+)	*rpoC1*(+)	*matK*(+)
atpF-atpH(-)	+-	+-	+-	+-	+-
psbA-trnH(-)	+-	+-	+-	+-	+-
psbK-psbI(-)	+-	+-	+-	+-	+-
matK(-)	+-	+-	+-	+-	