SYNOPSIS OF

Tapping Zingiberaceae - Wilderness to Mining Plastome

A THESIS

to be submitted by

TUSHAR

Roll No. 08610613

Under the Supervision of

Dr. Latha Rangan

for the award of the degree

of

DOCTOR OF PHILOSOPHY

Department of Biotechnology

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

JANUARY 2014

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

1. INTRODUCTION

Zingiberaceae family is an important natural resource that provides many useful products for food, spices, medicines, dyes, perfume and aesthetics (Jantan et al., 2003) and is distributed widely throughout the tropics, particularly in Southeast Asia. India is one of the richest and diverse regions for Zingiberaceae, having 22 genera and about 170 species (Wu and Larson, 2000). The NE region of India is a zone of greatest concentration where 19 genera and about 88 species are reported (Prakash and Mehrotra, 1995). A large number of species from the family Zingiberaceae is believed to have medicinal values as pointed out by local literature and their documentation would open door to possibility of development of new drugs molecule (Tushar et al., 2010). Despite having tremendous values, spread across diverse economically important sector, Zingiberaceae is not looked upon as a family which would offer large economic benefits. The reasons being poor documentation, difficulty in identification and even more difficult to differentiate some of the species owing to complex and overlapping morphology with narrow window of flowering. As the traditional identification fails to deliver, the need for an alternative mode of identification is felt.

Among the various alternative tools of identification (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; Welsh and McClelland 1990; Bowcock, 1994; vos et al., 1995) and molecular sequences (Neyman, 1974; Bowe et al., 1999; APG II, 2003; Birgitta, et al., 2009; Wu et al., 2009; Perelman, 2011; Rinke et al., 2013) are the most prominent ones. Classification based on morphological data is clouded with problems such as convergent evolution along with high amount of expertise and labour involved in the process whereas the polymorphism widespread in genome is either not reproducible or labour intensive. Therefore the classification and identification based on molecular sequences stands distinctly out. For plants the plastid genome is the most preferable one for the classification and identification purpose because of plastid's conservative mode of evolution (Palmer et al., 1988). Phylogeny among various taxa and super taxa of plants has been resolved using sequences from plastids (Sang et al., 1997; Shaw at al., 2004; Daubes and Paule, 2010). Further with the developments in the field of sequencing with emergence of new generation technologies, the cost of sequencing has gone down (Mardis, 2011) and the prospective of sequencing a large number of plastid genome has opened up. The use of whole plastome for phylogeny construction is especially

useful for several complex cases particularly involving lower taxonomic levels and recent divergence (Parks et al., 2009).

The early works has pointed out towards the complex phylogeny and classification of the Zingiberaceae (Williams et al., 2000; Kress et al., 2002). In this thesis the investigation was carried out towards documenting the ethno-medicinal values of the family followed by the improved identification of Zingiberaceae using plastid genome.

2. REVIEW OF LITERATURE

Zingiberaceae is the largest plant family of the order Zingiberales with 53 genera and over 1200 species (Kress 1990) and is distributed pan-tropically with one genus (*Renealmia*) found in the Neotropics, four genera (*Aframomum, Aulotandra, Siphonochilus, and Renealmia*) found in Africa, and the rest of the genera distributed in east Asia and the Pacific Islands. The family is still poorly known taxonomically with many species and genera newly described in the last several years (Larsen et al., 2001; Williams et al., 2002).

2.1 Ethno-medicinal value of plants and its documentation

Documentation of Traditional Knowledge (TK) have become a major and important issue in the planning, management and conservation of bio resources (Signorini et al., 2009). In the primitive times, there were no means to document any of such knowledge, accrued by apprenticeship or self-directed learning and educational methods, (probably by sharing incidences) and therefore large knowledge may have been lost and might have been re-acquired. Such knowledge have a direct effect on the daily living standards of rural populations, as well as in taking decisions regarding the sustainable use of plant resources (Benz et al., 2000; Shackleton et al., 2002). In developed and developing countries the traditional knowledge is widely threatened by current trends of economic globalization which aims at promoting intensive agriculture and industrialization (Signorini et al., 2009). The increased effort of researcher to counter-act the fast depletion of medicinal usage of plants by ethnic community is evident by the increase in number of reports about ethno-medicinal usage of plants different parts of the world (Lev, 2006).

Members of Zingiberaceae family has been known to have medicinal values and some of the studies have documented ethno-medicinal uses (Ignacimuthu et al., 2005; Igoli et al., 2005; Sajem and Gosai, 2006; Teklehaymanot and Giday, 2007; Lans, 2007). Apart from these ethno-

medicinal usage, a large number of reports on bioactivity of molecules and extracts derived from Zingiberaceae members has been published (Yuliana et al., 2011; Ghosh et al., 2013; Lu et al., 2013; Reddy et al., 2013).

2.2 Plastid 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).

2.2.1 Structure of plastid genome

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). 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 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 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).

2.2.2 Plastome 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) 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 and extensive levels

of intramolecular recombination. With development of easy 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 automation of sequencing further strengthens the study of phylogeny based on variation in sequence information (Olmstead and Palmer, 1994). Kress et al., (2002) conducted a detailed study on the construction of a new classification of Zingiberaceae based on the molecular data (ITS and matK gene) and included a large no of species collected mostly from tropical region in America continent and few from southeast Asia and Africa. Other reported study investigated, either focused on a section of the family (Searle and Hedderson, 2000; Ngamriabsakul et al., 2003; Pedersen, 2003) or only a genera was targeted (Williams et al., 2000; Ngamriabsakul et al., 2000; Yong-Mei et al., 2004). Apart from the two ways of utilizing the plastome for phylogeny as discussed above, recently the use of microsatellite (Powell et al., 1995a, b) and the whole plastome sequence for phylogeny study (Matsuoka, et al., 2002; Leebens-Mack et al., 2005) has increased because of drop in sequencing cost and availability of reference plastome. Among Zingiberaceae members the plastome has been sequenced and assembled only for *Zingiber spectabile* (Barrett et al., 2012).

2.2.3 Plastid genome sequencing and assembly

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#).

3. OBJECTIVES

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

4. BRIEF DESCRIPTION OF RESEARCH WORK

4.1 Collection and documentation of plants and related information

Zingiberaceae is a plant family with some very important members contributing volumes to pharmacognosy (Gilli et al., 2011; Jiang et al 2013) and many yet to be discovered (either the member of the family or the drug molecule). Since the first solid step towards this goal is collection and documentation of the bio-resources along with its ethno-medicinal values, as these property indicates presence of a bio molecule which may be a potent drug. In this chapter we describe about the collection of Zingiberaceae members, their important features and medicinal value. Moreover we developed a web portal containing these information for dissemination of the traditional knowledge.

Excursions were conducted in six states of NE, namely: Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram and Nagaland 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 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 (Table 1). The survey involved collection of fifty one plants belonging to nine genera of the

family Zingiberaceae from NE region of India. Thirty four plants out of fifty one were identified to possess medicinal value. 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. The mode of preparation and administration along with the part of the plant used were also recorded (Figure 1). A medicinal use was accepted as valid only if it was confirmed by at least two separate healers.

Table 1. The detail of various indigenous community interviewed with rounded value of healers along with the number of ethno-medicinally important plants collected.

State	Tribes	Number of healers (approx.)	Number of ethno-medicinally important plants/total no of plants
Arunachal Pradesh	Apatani, Nyishi, Hill Miri, Chakma, Khampits, Adi	130	24/35
Assam	Khampits, Bodo, Tai-Ahom	65	30/32
Manipur	Hmar	25	16/20
Meghalaya	Garo, Jaintia	45	8/9
Mizoram	Chakma, Lushai, Lakher, Pawi	85	12/15
Nagaland	Lushai	20	6/8

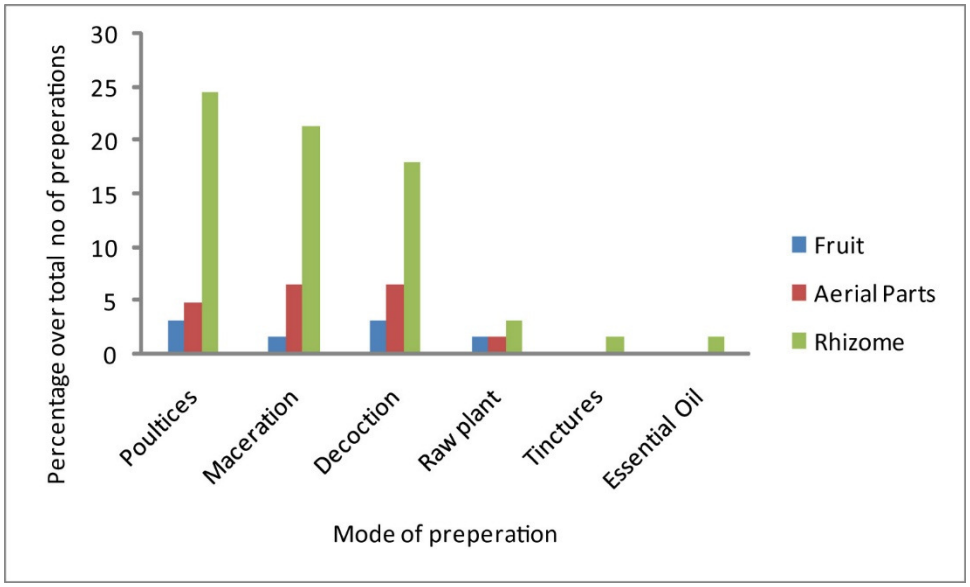


Figure 1. Frequency of plant parts used according to the mode of preparation of herbal medicine. Decoction: aqueous extract of plant material; tinctures: alcoholic extract of plant material; maceration: grinded and taken orally; poultices: grinded and applied locally; essential oil: concentrated, hydrophobic liquid containing volatile aroma compounds from plants; raw plant: specified part of the plant either taken orally or applied locally.

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.

Among the states of NE region, AP was found to be floral rich as far as members of the family Zingiberaceae is concerned (88%). 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 was calculated over total number of preparation). 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 2)

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 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.

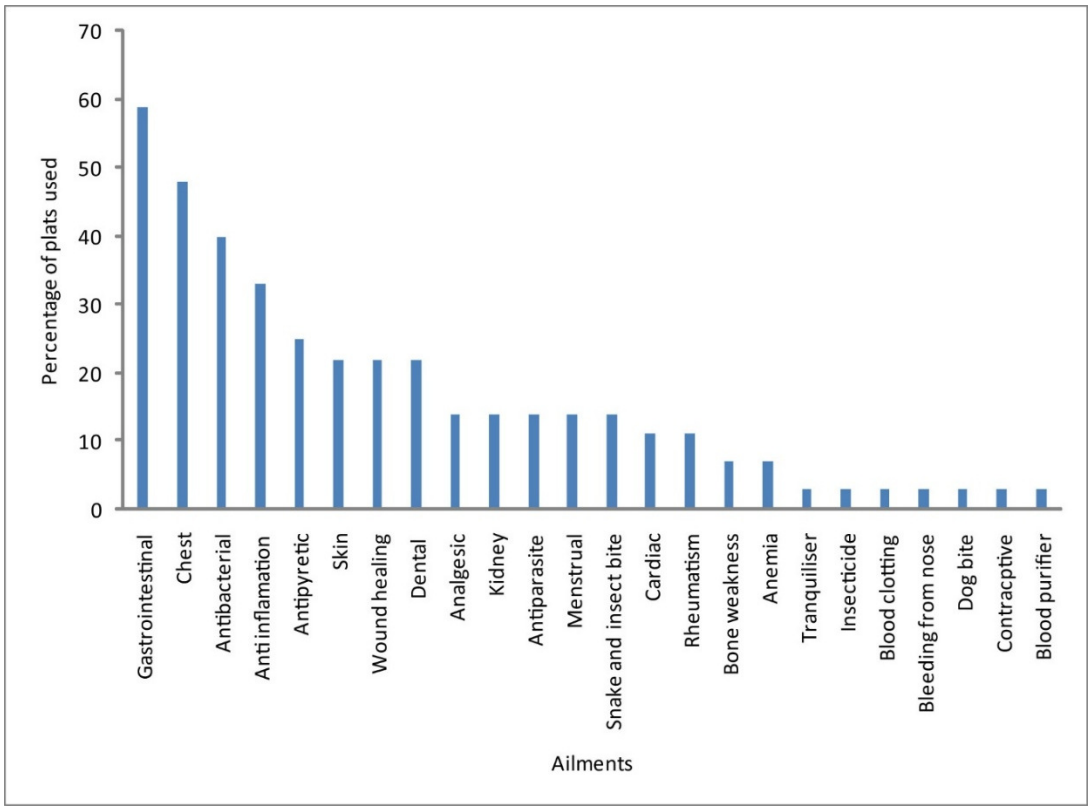


Figure 2. Percentage use of Zingiberaceous plants for treating various human ailments. Highest percentage of plants used is observed for treatment of gastrointestinal ailments that is also supported by large number of pharmacological reports. Percentage was calculated over number of plants documented in the current study.

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. 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 4 and 5)

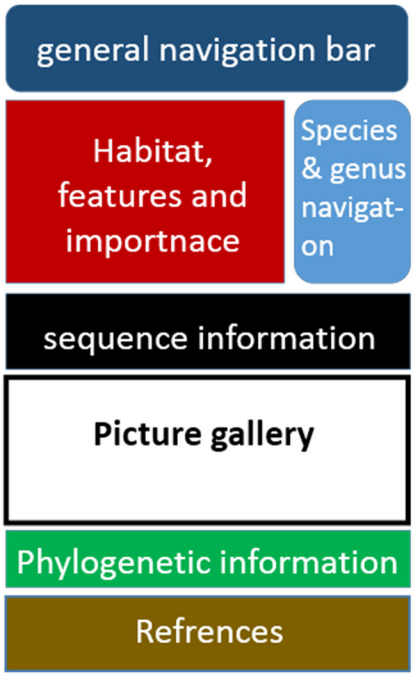
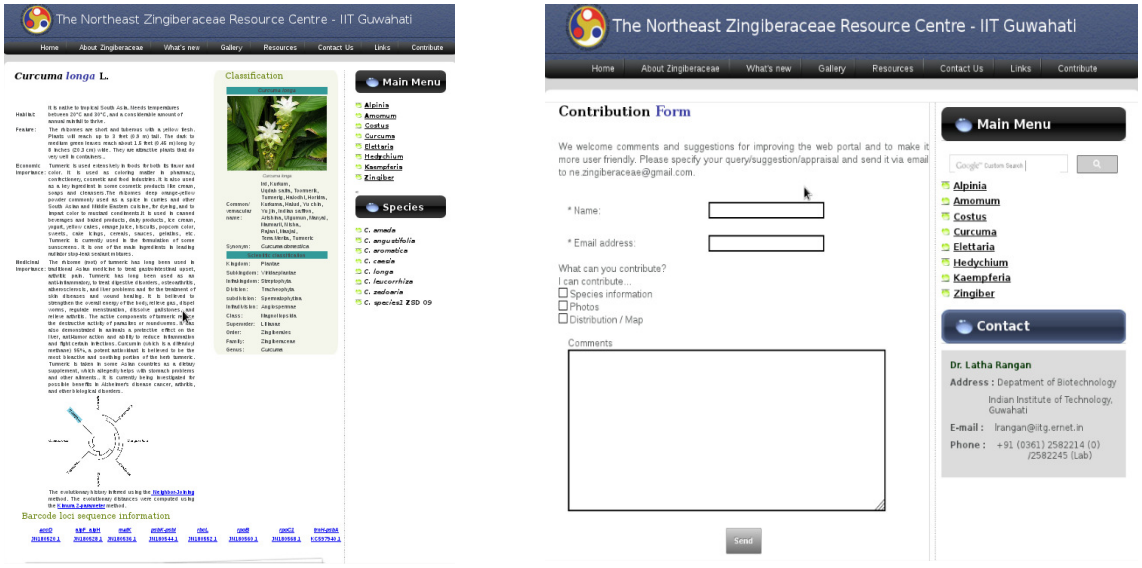


Figure 3 Standard layout of the web portal designed for dissemination of the knowledge accrued. The coloured blocks are different section of an individual page and is discussed in details. The navigation part is depicted in blue (light and dark) while all the other part is reserved for representing the information collected. The dark blue box corresponds to the general navigation bar. 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. 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

Figure 4. A snapshot of the database displaying the genera page. Bottom of the page shows links and downloadable files for the DNA sequences used for the phylogenetic analysis



(a) **(b)**

Figure 5. 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.

4.2 Mining of plastid genome for phylogenetic relationship assessment

Classifying genera into tribes and species into genus has been a daunting task in this family because of defining characters are either not unique to any one taxa or are not universal for all sub taxa within any taxa. Some studies have been undertaken to understand the phylogeny (Kress et al., 2002; Kress et al., 2004; Yong-Mei, 2004) but either they dealt with only one genus or they did not considered samples from northeast India which hosts a large number of Zingiberaceae members (Prakash and Mehrotra, 1995). 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 relevant alterations to the established phylogenetic classification of the family.

Out of the total fifty nine species recorded (as pointed in chapter 2) fourty six species were selected for mining plastid genome for 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-medicinal study and live specimens were not collected due to various unresolvable reason(s). 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). 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). 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. The contigs were assembled using Geneious Pro (5.6.7) (Drummond et al 2006) and were filtered for high value reads.

For the construction of phylogeny we devised two strategy based on the type of information to process from the sequence similarity. The first strategy was aimed at constructing tree using evolutionary distance derived from sequences aligned while the second strategy was using character based information for phylogenetic relationship construction. For the first strategy 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 pairwise distances were calculated using Kimura two parameter (K2P) nucleotide substitution model (Kimura, 1980) using MEGA5 (Tamura, et al., 2011) platform. Subsequently the phylogenetic trees were constructed using neighbor joining method (NJ) using MEGA5 with the option of pairwise deletion. For the second strategy sequences were aligned using MUSCLE (Edgar, 2004) 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. Subsequently dataset was used for construction of tree using maximum likelihood (ML) and Maximum parsimony (MP) analyses on MEGA 5.0 platform. For ML the general time reversible model with the rate considered as uniform across all the sites with using the gaps as missing data for all the sites was used. The Nearest-Neighbor-Interchange (NNI) heuristic approached was used for inferring tree, so as the branch-swapping performed on each tree resulted in a rearrangement resulting alternative trees to differ in only one branching pattern. Subsequently, to test the phylogeny 1000 non parametric bootstrap was performed. Analysis for MP 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.

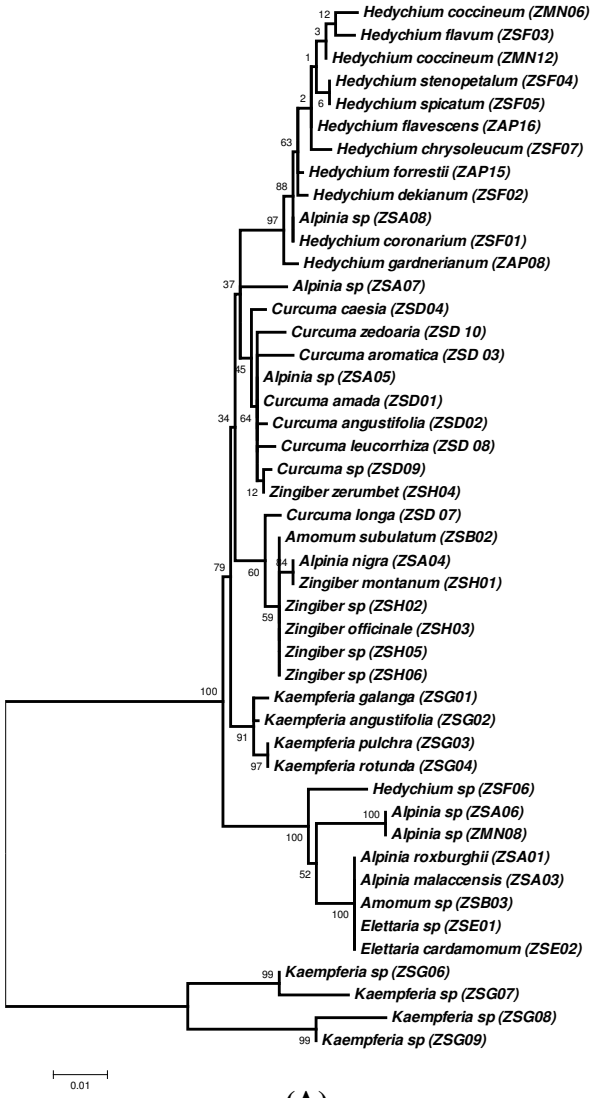
In this study, the Alpinieae tribe consisted of three genera: *Alpinia, Amomum* and *Elettaria*. The best combination for the phylogenetic relationship assessment was observed to be *matK* and *atpF-atpH*. Among the two strategy used for construction of the phylogeny, there was general concordance about the relationship but the resolution was marginally better in case of character based phylogeny. *Alpinia* has been previously found to be polyphyletic (Kress et al., 2005) as was the case with *Amomum* (Xia at al., 2005). The previous reports has not been conclusively proved about the nature of *Elettaria* phylogeny and in our study they appear to be monophyletic, though in this study only two species were included and hence it is too low to prove about the true nature.

Zingibereae consisted of only one genera, *Zingiber*, but since it is an important genera with regard to economical usage, the individual effort for phylogeny construction was focused. Among the various combination of loci tested the combination of *rbcL* and *atpF-atpH* was observed to be

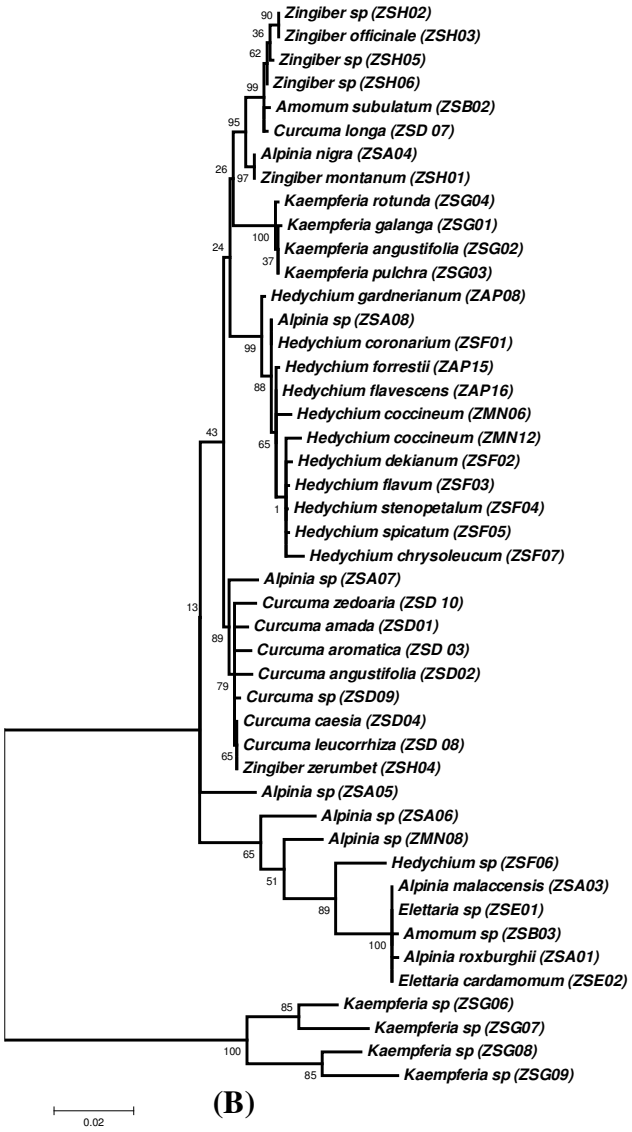
the most effective in resolving phylogeny. *Zingiber zerumbet* is highly diverged from other species of *Zingiber* and its closest relative was observed to be *Z. montanum*.

Hedychieae was the biggest group in this study and included three genera *Curcuma*, *Hedychium*, and *Kaempferia*. The best resolution was observed to be *rpoB* and *matK*. *Hedychium* and *Curcuma* mostly clustered together with the exception of one *Hedychium* species (ZSF06). The *Hedychium* was found to be more close to *Curcuma*, and shared a recent common ancestor, than *Kaempferia*. Further four unidentified species of *Kaempferia* clustered out of the main group and indicates that these are miss identification.

When all the tribe was clustered to provide the phylogeny of the Zingiberaceae family as a whole, two major kind of phylogeny was observed. The first one, based on combination of *rpoB* and *matK* established *Hedychium* and *Curcuma* shared a more recent common ancestor than *Zingiber* which is in accordance with the classical classification (Figure 6). The second being, based on combination of *matK* and *atpF-atpH* established *Hedychium* and *Zingiber* to share a more recent common ancestor than *Curcuma* which is in accordance with the recent molecular phylogeny (Figure 6) report (Kress et al., 2002). Four of *Kaempferia* species were found to cluster out of the main group and this strongly suggested that these could be miss identification, but they serve as a reference to put root to the tree.


Hedychium coccineum (ZMN06)
Hedychium flavum (ZSF03)
Hedychium coccineum (ZMN12)
Hedychium stenopetalum (ZSF04)
Hedychium spicatum (ZSF05)
Hedychium flavescens (ZAP16)
Hedychium chrysoleucum (ZSF07)
Hedychium forrestii (ZAP15)
Hedychium dekianum (ZSF02)
Alpinia sp (ZSA08)
Hedychium coronarium (ZSF01)
Hedychium gardnerianum (ZAP08)
Alpinia sp (ZSA07)
Curcuma caesia (ZSD04)
Curcuma zedoaria (ZSD 10)
Curcuma aromatica (ZSD 03)
Alpinia sp (ZSA05)
Curcuma amada (ZSD01)
Curcuma angustifolia (ZSD02)
Curcuma leucorrhiza (ZSD 08)
Curcuma sp (ZSD09)
Zingiber zerumbet (ZSH04)
Curcuma longa (ZSD 07)
Amomum subulatum (ZSB02)
Alpinia nigra (ZSA04)
Zingiber montanum (ZSH01)
Zingiber sp (ZSH02)
Zingiber officinale (ZSH03)
Zingiber sp (ZSH05)
Zingiber sp (ZSH06)
Kaempferia galanga (ZSG01)
Kaempferia angustifolia (ZSG02)
Kaempferia pulchra (ZSG03)
Kaempferia rotunda (ZSG04)
Hedychium sp (ZSF06)
Alpinia sp (ZSA06)
Alpinia sp (ZMN08)
Alpinia roxburghii (ZSA01)
Alpinia malaccensis (ZSA03)
Amomum sp (ZSB03)
Elettaria sp (ZSE01)
Elettaria cardamomum (ZSE02)
Kaempferia sp (ZSG06)
Kaempferia sp (ZSG07)
Kaempferia sp (ZSG08)
Kaempferia sp (ZSG09)
0.01

(A)


Zingiber sp (ZSH02)
Zingiber officinale (ZSH03)
Zingiber sp (ZSH05)
Zingiber sp (ZSH06)
Amomum subulatum (ZSB02)
Curcuma longa (ZSD 07)
Alpinia nigra (ZSA04)
Zingiber montanum (ZSH01)
Kaempferia rotunda (ZSG04)
Kaempferia galanga (ZSG01)
Kaempferia angustifolia (ZSG02)
Kaempferia pulchra (ZSG03)
Hedychium gardnerianum (ZAP08)
Alpinia sp (ZSA08)
Hedychium coronarium (ZSF01)
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 chrysoleucum (ZSF07)
Alpinia sp (ZSA07)
Curcuma zedoaria (ZSD 10)
Curcuma amada (ZSD01)
Curcuma aromatica (ZSD 03)
Curcuma angustifolia (ZSD02)
Curcuma sp (ZSD09)
Curcuma caesia (ZSD04)
Curcuma leucorrhiza (ZSD 08)
Zingiber zerumbet (ZSH04)
Alpinia sp (ZSA05)
Alpinia sp (ZSA06)
Alpinia sp (ZMN08)
Hedychium sp (ZSF06)
Alpinia malaccensis (ZSA03)
Elettaria sp (ZSE01)
Amomum sp (ZSB03)
Alpinia roxburghii (ZSA01)
Elettaria cardamomum (ZSE02)
Kaempferia sp (ZSG06)
Kaempferia sp (ZSG07)
Kaempferia sp (ZSG08)
Kaempferia sp (ZSG09)
0.02

(B)

Figure 6. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches (A) *rpoB* + *matK* (B) *matK* + *atpf-atpH*

4.3 Whole plastome sequencing and new marker development

Organelles, such as the chloroplast, have prokaryotic ancestral origins as well as a functional and phylogenetic relationship to the evolution of their nuclear genomes. Due to their high levels of conservation and maternal inheritance, chloroplasts can be used to develop molecular markers for phylogenetic and phylogeographic studies. Ever since the first chloroplast was sequenced, it has been observed that the cpDNA of most flowering plants are circular molecules that are highly conserved in terms of gene content, size and structural organization. The size of the chloroplast genome is fairly constrained while the coding regions are critical to fundamental plant processes. Thus, highly conserved genes are found proximal to one another. These serve as optimal primer targets for amplifying short stretches of more divergent, intergenic DNA.

We *de novo* assembled the *Zingiber officinale* chloroplast genome using a shotgun library of total 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 7). Excepting a gap at the IRa/SSC border, the *de novo* assembly shows perfect colinearity with the recently published *Musa* plastome. Sequencing reads evenly cover the entire plastome as the IR regions are effectively identical, reads were randomly placed in either region. Annotation was performed using cpCASAVERA. Generally, there are no major discrepancies in the gene content of *Zingiber officinale* plastome and know plastomes, excepting an expansion of the IRa region as described below. The expansion creates a duplication 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.

The tree generated from alignment of these entire plastomes (Figure 8) corroborates with trees based on combined data from the matK and atpF-atpH gene combination. Both trees indicate

that Hedychium and Zingiber share a more recent common ancestor than the Zingiber and Curcuma, although the divergence of all three lineages appears to have been nearly simultaneous.

Short sequence repeats (SSRs) or "microsatellites" are hypermorphic due to errors during DNA replication. Such sequencers often serve as useful loci for differentiating closely related individuals. Alternatively, DNA barcoding exploits point mutations and short indels that occurring in rapidly evolving sequences flanked by more slowly evolving sequences. The reduced rate of mutations relative to a SSR 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 of a single barcode.

Using *Musa 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 Zingiber officinale sequence and the additional Zingiberaceae data produced in this study and elsewhere, we attempted to generate reliable chloroplastic SSR markers using a comparative approach. Briefly, alignment information from a whole plastome comparison between *Zingiber officinale* and *Zingiber spectabile* was used to find di- and tri-nucleotide repeats that were polymorphic within *Zingiber* genus and, thus, more likely to be polymorphic within the single species. We avoided homopolymer repeats because, generally, they are technically difficult to use

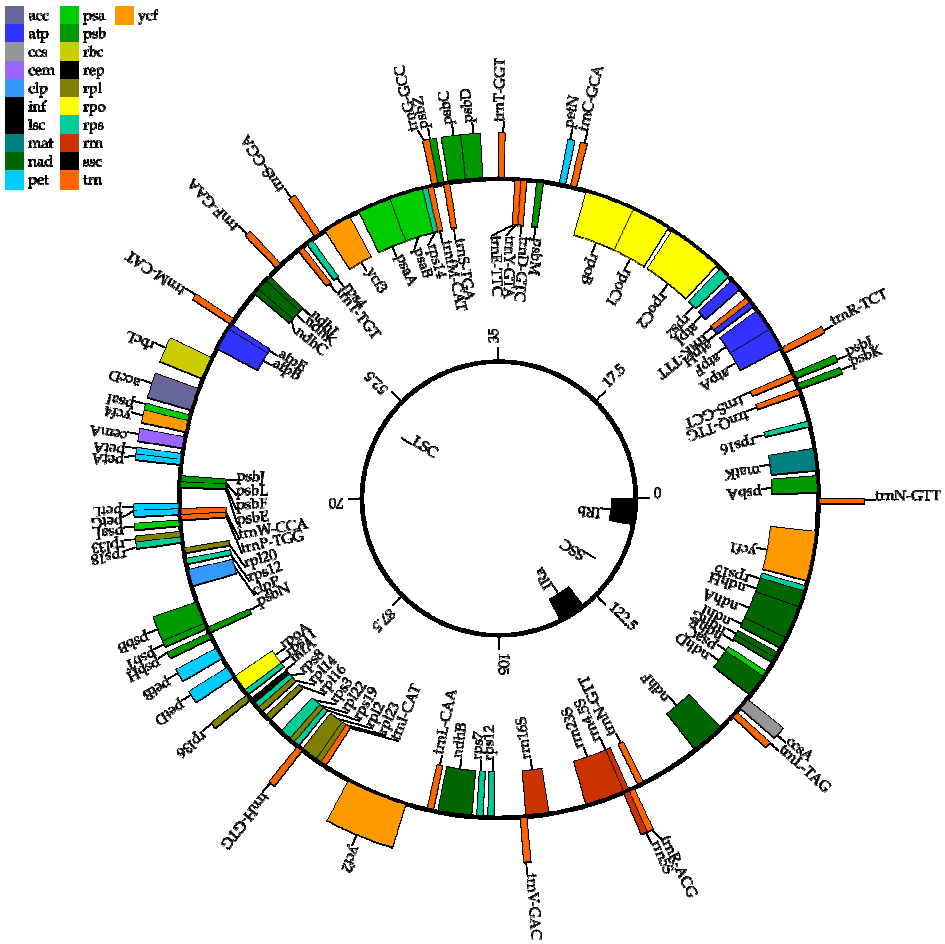


Figure 7. Gene map of Z. officinale chloroplast genome. The inner circle shows the four major regions of the genome: the two copies of the inverted repeat (IRA and IRB) and the large and small single-copy regions (LSC and SSC). The outer circle represents the tobacco genome with the transcribed regions shown as boxes proportional to gene size. Genes inside the circle are transcribed in a clockwise direction, and genes outside of the circle are transcribed counterclockwise

in applied settings. We also exploited alignment information 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, we also identified SSR markers that are likely to be most effective across the Zingiberaceae

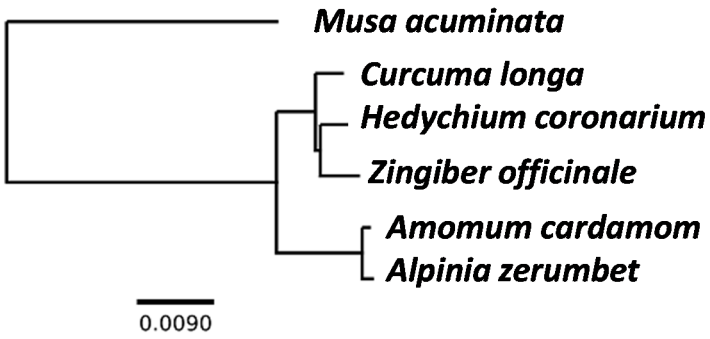


Figure 8. A phylogenetic tree based on the plastome alignment. *Musa acuminata* was used as outgroup. Each branch point has 100% bootstrap support. Scale bar indicates the number of substitutions per site estimated to have occurred for such a length along the tree.

To find suitable barcodes, we searched the alignments 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. Similarly to SSR discovery, we identified markers using the intra-genus comparison for *Zingiber* as well as the comparison across all plastomes from this study. For the intra-genus and intra-family comparison, we required at least 5 and 8 of such SNPs, 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 2.

Table 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

4.4 Comparative study of phylogeny constructed

Phylogeny re-construction is a science based on evidence (fossil records) as well as assumptions (related to morphological or sequence change). Further there are a myriad number of strategies for phylogenetic reconstruction and each have their advantages and disadvantages based on assumptions and the source of variation (Soltis and Kuzoff, 1995). These differences lead to phylogeny which have some variation. While one may be suited more for a particular case, on the other hand one may not. Further the phylogeny is effected by the sampling of taxon (Zwickl and Hillis, 2002).

The main disagreement between phylogeny constructed in the above sections was the relationship of *Zingiber, Curcuma* and *Hedychium*. The phylogeny based on some loci combinations from plastid genome established *Zingiber* and *Hedychium* to be more closely related than either of the two with *Curcuma*. While some of the loci established *Curcuma* and *Hedychium* to be more closely related. When the complete plastid genome was used for phylogeny reconstruction, *Zingiber* and *Hedychium* was observed to be more closely related than *Curcuma*. Though the divergence point was found to be critically close. Further the work by Kress et al., (2002) based on ITS and *matK* region, supports closer relationship of *Hedychium* and *Zingiber.* Further to test the effect of taxon sampling from different region, we downloaded plastid sequences from public domain for *Zingiber* and constructed phylogeny and surprisingly found that *Zingiber* species from China was found to cluster together and their cluster was distinct from *Zingiber* species from northeast India (Figure 9)

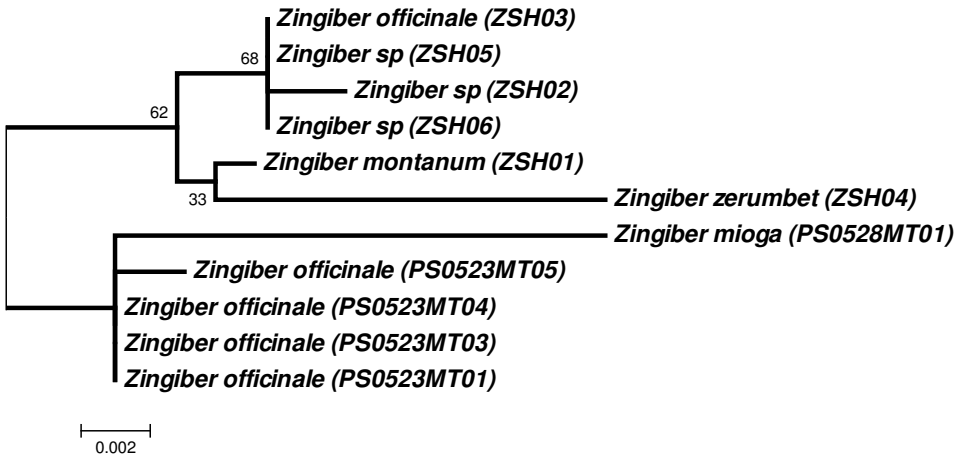


Figure 9. The phylogeny for *Zingiber*, including species from China, using combination of loci *rpoB* + *psbK-psbI* . All the *Zingiber* species from same region clustered closely.

5. CONCLUSION

The present thesis demonstrates the ethno-medicinal importance and phylogenetic relationship of Zingiberaceae. Study on ethno-medicinal usage of Zingiberaceae pointed out some very unique application of these species and will act as the starting point for new drug discovery. The phylogeny constructed using plastid sequences has clearly clarified some unresolved relationships. The plastid genome of *Zingiber officinale* and four other Zingiberaceae members were sequenced for the first time. These genomes will help in better understanding of evolutionary relationship. New markers were also designed using plastid genomes. These markers system could be used for quick classification of members of Zingiberaceae and can also be used for differentiation of *Zingiber* species.

6. PROPOSED THESIS CONTENTS

The outline of thesis is as below

Chapter 1: Introduction

Chapter 2: Review of literature

Chapter 3: Collection and documentation of plants and related information

Chapter 4: Mining of plastid genome and phylogenetic relationship assessment

Chapter 5: Plastome sequencing for developing new marker system

Chapter 6: Comparative study of phylogeny constructed

Conclusion

References

7. 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

8. REFERENCES

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. doi:10.1186/1746-4811-6-22

B. Chevreux, T. Wetter, & S. Suhai. (1999.). Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In German Conference on Bioinformatics, pp. 45-56

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.

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.

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

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.

Daubes, 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.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792–7.

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.

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.

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.

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.

Huang, Y.-Y., Matzke, A. J. M., & Matzke, M. (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*) (*H. Candela*, Ed.)PloS one, 8(8), e74736.

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

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., 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

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.

Lev, E. (2006). Ethno-diversity within current ethno-pharmacology as part of Israeli traditional medicine-A review. Journal of Ethnobiology and Ethnomedicine 2: 4.

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.

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.

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. (J. G. Umen, Ed.)PloS one, 8(6), e67350.

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.

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. doi:10.1017/S0960428603000362

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, Jeffrey D. (1983). Chloroplast DNA exists in two orientations. Nature, 301(5895), 92–93.

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

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.

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.

Ratan, A. (2009). Assembly algorithms for next-generation sequence data. Doctoral Thesis.

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.

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, 6(5), 283–90.

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.

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.

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.

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.

Soltis, D. E., & Kuzoff R. K. (1995). Discordance between Nuclear and Chloroplast Phylogenies in the Heuchera Group (*Saxifragaceae*). Evolution. 49(4), pp. 727-742

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

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.

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.). (J. H. Badger, Ed.)PloS one, 5(9), e12762.

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

Wu TL, Larson K (2000) Zingiberaceae. In: Wu Raven (ed) Flora of China, vol 24. Missouri Botanical Garden Publishing, St. Louis, pp 360–363.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome research, 18(5), 821–9.