Evaluation of Nuclear and Chloroplast gene sequences for feasibility of DNA Barcoding of Musaceae family

Thesis submitted in partial fulfillment
of the requirements for the
Degree of Master of Philosophy in Biotechnology

by
H.Lallawmawma

M.Phil. Registration No: MZU/M.Phil/102 of 10.05.2012

Under the Supervision
of

Dr. Thangjam Robert Singh
Assistant Professor
Department of Biotechnology
Mizoram University, Aizawl – 796004, INDIA

Department of Biotechnology, Mizoram University,
Aizawl, Mizoram 796004
INDIA

2012

DEPARTMENT OF BIOTECHNOLOGY
SCHOOL OF LIFE SCIENCES
MIZORAM UNIVERSITY
(A Central University)
P. B. No. 190
TANHRIL, AIZAWL – 796004: MIZORAM, INDIA

Dr. Thangjam Robert Singh, Ph.D
Assistant Professor

Phone : +91 389-2330859/2330861(O)
Fax : +91 389-2330642 (O)
E-mail : robert_thangjam@yahoo.com
robertthangjam@gmail.com

MZU/BTCH/TRS/2007-M.Phil-01 Date:

Certificate

This is to certify that the present thesis entitled "**Evaluation of Nuclear and Chloroplast gene sequences for feasibility of DNA Barcoding of Musaceae Family"** submitted by **Mr. H. Lallawmawma**, M.Phil. Registration No: MZU/M.Phil/102 of 10.05.2012 in partial fulfillment for the award of the Degree of Master of Philosophy in Biotechnology of Mizoram University has been carried out during 2011-2012 under my supervision. The thesis embodies original research and has not been submitted for any degree elsewhere.

The candidate has fulfilled all the requirements under M. Phil ordinance of the Mizoram University.

(Dr. Thangjam Robert Singh)
Supervisor

Declaration of the Candidate

I, H. Lallawmawma, M.Phil student in Biotechnology Department, Mizoram University, Aizawl, do hereby solemnly declare that the subject matter of this thesis is the record of the work done by me. I have duly worked on my M.Phil. thesis under the supervision of Dr. Thangjam Robert Singh, Assistant Professor, Department of Biotechnology, Mizoram University. This is being submitted to Mizoram University for the degree of Master of Philosophy in Biotechnology and that I have not submitted this work to any other University or Institute for any other degree.

I also declare that the present investigations relate to bonafide M.Phil student works undertaken and the title of the thesis is ***"Evaluation of Nuclear and Chloroplast gene sequences foe feasibility of DNA Barcoding of Musaceae family".***

(Dr. Thangjam Robert Singh)
Supervisor
Assistant Professor
Department of Biotechnology
Mizoram University
Aizawl, Mizoram.

(H. LALLAWMAWMA)
Candidate

ACKNOWLEDGEMENT

First, I would like to wholeheartedly thank **God** for His abundant blessing and care He has given me throughout the period of my work. Without Him, I would have been given up and this thesis would not come into an end. He has been a strong pillar behind the success of my thesis, which endlessly giving me an opportunity to experiences His wonderful creatures through different stages of hardship and Joy.

My words would not be able to express my indebtedness to my supervisor, **Dr. Thangjam Robert Singh**, Assistant Professor, Department of Biotechnology, Mizoram University for his invaluable knowledge, scholarly guidance and patience. His scientific instinct, good hearted and exhorting words build the foundation of my scientific view and carrier in research. He has provided me all the things I need during the time of my work. His nature of understanding the student, broad minded and zeal to sow the working environment mean a lot for my inspiration.

I am immensely indebted to my parents and family who have waited so patiently to enable me complete this work. Their moral support, spending a daily prayer for me, their constants love and affection remind me the value of endurance at the time of trouble while doing my research. This thesis would not become possible without the contribution of their willing souls. My heartfelt gratitude to **Prof. N. Senthil Kumar** and the faculty members of Department of Biotechnology namely **Dr. J Bhattacharya, Dr. Bhim Pratab Singh, Dr. H. Vanlalhruaia,** Mizoram University, Aizawl for their noble knowledge, assistance and encouragement during the course of this work. My sincere appreciate to all my colleagues **P.C Lalfinfela**, **Lalremsiami Hrahsel**, **Atom Annupama Devi, Moirangthem Lakshmi Pyari Devi, Zothansanga, Rita Zomuanpuii**, **Catherene Vanlalruati, Rebecca Vanlalsangi, Esther Lalnunmawii, Liansangmawii Chhakchhuak, Thangjam Premabati** (Research scholar, Biotechnology Deptt., Mizoram University) **Ruth Lalfelpuii, Lalhlimpuia Pachuau** (M.phil Biotechnology Deptt., Mizoram University) for all the laugh, cheerful company, motivation and helping me in finishing my work. I am also extremely thankful to all the non-teaching staffs of the Biotechnology Department, Mizoram University Aizawl , namely – **Baby , Lalnunzira, Vanlalruata, Lalmuanpuia** for their valuable cooperation during the course of this work.

My special thanks to DBT, Ministry of Science and Technology, India for awarding fellowship. I record my appreciation to DBT-BIF, Biotechnology Deptt. for continuous provision of computer and internet facility all the time.

Heartfelt thanks are also due to **David K. Zorinsanga and Lalrinhlua** (Laboratory cum field Attendant in DBT project) for their invaluable assistance during the course my work.

. And I would like to extend my special thanks to all my friends who encourage me and remember me in their prayers.

Dated:

Department of Biotechnology

Mizoram University, Aizawl **(H. LALLAWMAWMA)**

Abbreviations

DNA	Deoxribonucleic Acid
rbcL	Ribulose-biphosphate carboxylase
matK	Maturase K
ITS1	Internal Transcribed Spacer 1
ITS2	Internal Transcribed Spacer 2
rDNA	Ribosomal Deoxyribonucleic Acid
*nr*DNA	Nuclear Ribosomal Deoxribonucleic Acid
*nr*ITS	Nuclear Internal Transcribed Spacer
PCR	Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulphate
CTAB	Cetyltrimethyl Ammonium Bromide
EDTA	Ethylene Diaminetetra Acetic Acid
Nacl	Sodium Chloride
PVP	Polyvinyl Pyrolidone
$MgCl_2$	Magnesium Chloride
dNTP	De-oxyribonucleotide Triphosphate
RNA	Ribulose Nucleic Acid
UV	Ultra Violet
ML	Maximum Likelihood
cv	Cultivar
T	Thymine
C	Cytosine
A	Adenine
G	Guanine

CONTENTS

Chapter No.	Topic	Page No.
1	INTRODUCTION	1-11
2	MATERIALS AND METHODS	12-19
3	RESULTS	20-47
4	DISCUSSION	48-50
5	CONCLUSION	51
6	SUMMARY	52-53
	REFERENCE	54-59

List of Tables

Table Nos.	Title	Page.
1	List of banana plants used for the study	19
2	Details of the edited and unedited ITS2 sequences from the banana samples studied	27
3	Results of the Blast search for similarity between the ITS2 sequences obtained from the banana samples studied and GeneBank data carried out using the MegaBlast program of the Blastn algorithm.	28
4	Details of ITS2 reference sequences of *Ensete spp*.obtained from Genebank(NCBI) used as outgroup for the analysis	28
5	Details of the nucleotide composition of ITS2 sequences obtained from the banana samples studied	29
6	Variable nucleotide sites occurring in the ITS2 non-coding regions among banana samples studied	30
7	Results of the Tajima's Neutrality Test of the ITS2 sequences obtained from the banana samples studied	31
8	Pairwise genetic distance (Tamura and Nei,1993) generated by ITS2 sequences	32
9	Details of the edited and unedited rbcL sequences from the banana samples studied	33
10	Results of the blast search for similarity between the rbcL sequences obtained from the banana samples studied and GeneBank data using MegaBlast of the Blastn algorithm.	34
11	Reference rbcL sequences used as outgroup for the analysis	35
12	Details of the nucleotide composition rbcL sequences obtained from the banana samples studied.	36
13	Variable nucleotide sites occurring in the rbcL coding regions among banana samples.	37
14	Results of the Tajima's Neutrality Test of the rbcL sequences obtained from the banana samples studied.	38
15	Pairwise genetic distance(Tamura and Nei,1993) generated by rbcL sequences in the samples studied	39

16	Analysis of intra and inter sections of Musaceae using ITS2 and rbcL gene sequences	40
17	Wilcoxon signed-rank test for inter-species divergence among loci.	40

List of Figures

Figure Nos.	Title	Page.
1.	Generalized ribosomal DNA unit structure.	9
2.	Electrophoregram of the extracted genomic DNA from the banana samples electrophoresed on a 0.8% agarose gel.	41
3.	PCR amplification profile of the samples generated by ITS2 primers	41
4.	PCR amplification profile of the samples using rbcL primers.	42
5.	Chromatogram of the raw ITS2 sequences of the Musa ABB group cv. Banpawl.	42
6.	Multiple sequence alignment of the ITS2 sequences.	43
7.	Relative distribution between intra- and interspecies divergence of the ITS2 sequences using 0.020 barcode distance units	44
8.	Phylogenetic tree generated by analysis of the ITS2 sequences using Maximum Likelihood method.	44
9.	Multiple alignment of the rbcL sequences analyzed. (generated using the software DNAman ver. 6.0.)	45
10.	Relative distribution between intra- and interspecies divergence of the rbcL sequences using 0.020 barcode distance units.	46
11.	Phylogenetic relationship revealed by rbcL sequences of the banana cultivars.	47
12.	Genetic distance within and between groups of Musaceae as revealed by ITS2 and rbcL sequences in the samples studied.	47

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1. General background

Bananas and plantains belongs to the family Musaceae Jussieu (1789), distributed highly in tropical and sub-tropical regions in the world found growing in more than 120 countries. *Musa* L. (Linnaeus 1753) is one of the three recognized genera of the family Musaceae along with *Ensete* Horaninow (1862) and *Musella* (Fr.) . The genus *Musa* originated from Southeast Asia and consists of around 65 species while *Ensete* is a small genus of 8-9 species distributed in Asia, Sub-Saharan Africa and Madagascar and *Musella* is a monotypic genus native to Southwest China (Cheesman1947; Baker and Simmonds 1953). According to Cheesman (1947), *Musa* is divided into four sections based on the phenotypic traits and basic chromosome numbers, such as Eumusa (x = 11), Rhodochlamys (x = 11), Australimusa (x = 10) and Callimusa (x = 10). Recently Wong *et al*., (2002) divided *Musa* into five sections with the addition of Ingentimusa (x = 7). Among the five *Musa* sections, Eumusa contributed the largest number, most diversified and ancient including the two wild diploid progenitor species of *M.acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome) (Bekele and Shigeta 2011). According to Simmonds and Shepherd (1955) the inter and intra specific crosses between these two species leads to the existence of different varieties with various genome composition resulting several ploidy level ranging from diploids to tetraploids (AA, AB, AAB, ABB, ABBB, AABB etc.) and posing a huge challenge towards its proper identification.

According to Li *et al*. (2010) Musaceae contains only two genera *Musa* L. (Linnaeus, 1753) and *Ensete* Horan (Horaninow, 1862). The genus *Ensete* differs from *Musa* by being monocarpic,

non-suckering (except *E. ventricosum*) with a distinctively swollen base and having large-sized seeds while *Musa* produces suckers and has small seeds (Samson, 1992). However, the clarity on the taxonomy of *Ensete*remains a subject of debate and so far 6 species have been identified (Bekele and Shigeta, 2011). They are distributed in Asia, sub-Saharan Africa and Madagascar (Cheesman1947; Baker and Simmonds, 1953). A few species are also reported from northeast India to the Philippines and New Guinea (Purseglove, 1972). The genus *Musa* contains 30-40 species and the taxonomy is based on plant phenotype, flower orientation and arrangement and also the basic chromosome number (Ude et al., 2002).

1.2. Origin and distribution of banana

The *Musa* domestication process started some 7,000 years ago in Southeast Asia (D' Hont *et al*., 2012). It involves the hybridization between diverse species and subspecies, cultivated by human migrations (Perrier *et al*., 2011) and selection of diploid and triploid seedless, parthenocarpic hybrids thereafter widely dispersed by vegetative propagation. Simmonds (1962) described Southeast Asia as the centre of origin of banana and plantains and the majority of cultivated bananas arose from the *Eumusa* section. This section is the largest, most geographically diversified with ~21 species and mainly distributed in the Indo-Malaysian and Southeast Asian countries to North Australia (Li *et al*., 2010). It also consists of the two wild diploid progenitor species namely - *M. acuminate* Colla(A genome) and *M. balbisiana* Colla(B genome) which leads to the origin of different genome combinations (AB, AAB, ABB and ABBB etc.) through intra and inter-specific crosses (Uma *et al*., 2006). Azhar (2006) stated that either Malaysia (Simmonds, 1962) or Indonesia (Horry *et al*., 1997) is the centre of diversity of *M. acuminata*

whereas *M. balbisiana* is originated from India, Myanmar, Thailand and Philippines. Few of the edible hybrids originated from crosses between *M. acuminata* and *M. schizocarpa* (S genome) as well as from *M.acuminata* and *M. textilis* (T genome) have also been reported (Hřibová *et al*., 2011). *M.acuminata* is the most widespread of the *Eumusa* and the occurrence of spontaneous mutation, or recombination events have resulted the development of natural reproductive barriers within the species, causing subspecies divergence and genetic diversity in the species as a whole. In addition, edibility of mature fruits of diploid *M. acuminata* (AA) resulted due to female sterility and parthenocarpy (Daniells *et al*., 2001). Initially these spontaneously occurring mutant edible cultivars with parthenocarpic fruits were domesticated by farmers and brought into cultivation. From the domesticated stocks of the cultivated bananas further distribution and propagation took place by vegetative means through suckers (Heslop-Harrison and Schwarzacher, 2007). However, the genetic basis of the mutation (or mutations) in the 'A genome' that gives rise to parthenocarpy has not been characterized so far and no parthenocarpy has been identified in 'B genome' diploids, although hybrids of A and B show the character.

The cultivated bananas and plantains differ from their wild relatives by being seedless. Fruits of these cultivated species are typically sterile or have extremely low fertility. Majority of the cultivated bananas are triploid ($2n = 3x = 33$) with a minor contribution of diploid and tetraploid ($2n = 4x = 44$) (Heslop-Harrison and Schwarzacher, 2007). The hybridization events and mutations giving rise to the seedless and parthenocarpic characters have been occurring many hundreds of times in the wild and wherever the fertile plants occur together, hybridization continues to produce new recombinations and parental lines (Pollefeys *et al*., 2004). The hybridization of *M. acuminata* from southeast Asia with *M. balbisiana* from Indian subcontinent occcured mainly due to human intervention leading to the evolution of AB, AAB and ABB

cultivars which are greatly diversified in the region (Daniells *et al*., 2001). The early dispersal of such banana cultivars resulted in the development of distinct sub-groups of varieties in different geographic locations. Secondary diversification within the major sub-groups of cultivated bananas is thought to have been the result of somatic mutations rather than sexual reproduction. Mutations affecting traits of economic or horticultural interest have been selected by farmers over the years and multiplied by vegetative propagation to produce morphotypes.

1.3. Status of research on banana

Banana and plantains (*Musa* spp.) are one of the top world trade commodities and as food security for millions of people in the tropical and sub-tropical regions. It is the fourth most important food crops (after rice, wheat, and maize), highly distributed in more than 120 countries with an annual production of 106 million tonnes (FAO, 2005). It originated from Southeast Asia and then diversified widely across the world (Simmonds and Shepherd, 1955). India is the largest producer of banana in the world with annual production of 13.5 mt from an area of 4.0 lakh (Daniells *et al*., 2001). Besides the cultivated species and their cultivars, majority of the species of *Musa* were found in wild condition and they are widely distributed in the northeast states. The northeastern region of India have been considered as the richest sources of natural banana diversity where *M. balbisiana* from Indian subcontinent meet *M. acuminata* from South East Asia (Molina and Kudagamage, 2002). Uma *et al*. (2001) had made the survey and recorded 13 different bananas and plantains of Mizoram, Northeast India.

For the improvement of banana production and the planting materials at the global level, International Network for the Improvement of Banana and Plantain (INIBAP) and other international research centers were established to emphasize on banana research with the main

objective of sustainable production of banana and plantain for domestic consumption and for local and export markets. Sathiamoorthy *et al*. (2001) reported that the first improvement programmed on banana in India was started in 1949 at Central Banana Research Station in Aduthurai in Tamil Nadu. Its main objectives were collection, conservation and evaluation of *Musa* germplasm and identifying accessions resistant/tolerant to stress by screening their reaction to pests and diseases for sustainable production. Of the 8 banana genebanks in India, National Research Centre on Banana (NRCB), Trichy, Tamil Nadu which was established by the Indian Council of Agricultural Research (ICAR) in 1993 has now a collections of more than 970 indigenous banana from all banana growing areas including southern state, Gujarat, Meghalaya, West Bengal, Bihar. Among these banana 109 were collected from Northeast states (Assam, Meghalaya, Arunachal Pradesh, Mizoram, Tripura and Sikkim). Assam Agricultural University (AAU) in Kahikuchi and Jorhat are the two banana gene banks available in Northeast India where there are 87 and 96 numbers of accessions respectively. Independent collections are also being maintained in the institutes of northeast region such as Mizoram, Arunachal Pradesh, Manipur, Assam, etc.

1.4. DNA Barcoding

DNA Barcoding is a relatively new concept that has been developed for providing a rapid, accurate and automatable species identification using standardized DNA sequences as tags. In fact, it started with the seminal work of Hebert *et al.*,(2003). DNA Barcoding is the process of identification of species based on nucleotide diversity of short DNA segments. However, in plants, due to the difficulty in finding a universally acceptable barcode.it is yet to be well established. Based on the relative efficacy testing, the Consortium for the Barcode of Life-Plant

Working Group has recently identified a few loci as potential barcode candidates and from them a two-locus standard barcode (*rbcl+matK*) has been recommended for initiating the barcoding process of plant species. In DNA barcoding, the unique nucleotide sequence patterns of small DNA fragments (400-800 bp) are used as specific reference collections to identify specimens and to discover overlooked species. Thus, the initial goal of the DNA Barcoding process is to construct on-line libraries of barcode sequences for all known species that can serve as a standard to which DNA barcodes of any identified or unidentified specimens can be matched. This can alleviate several inherent problems associated with traditional taxonomic identification,based on morphological characters, such as wrong identification of species due to phenotypic plasticity and genotypic variability of the characters, overlooking cryptic taxa, difficulty in finding reliable characters due to long maturity periods. DNA barcoding thus provide the taxonomists, conservationists and others who need the identification of species, a cost-effective tool, much as a barcode that identifies supermarket products. It is especially of much use in areas where species identification with morphological characters is not practicable due to extensive damage or delayed expression.. With the support of CBOL, the effort of DNA barcoding has been slowly progressing with controversies and intense debates.

1.5. Basic features of barcoding sequences:

The most important characteristic features of a DNA barcode are its universality, specificity on variation and easiness on employment. This means that the gene segment used as a barcode should be suitable for a wide range of taxa, should have high variation between species but should be conserved within the species, so that the intra-specific variation will be significant.

Additionally, it is often required that degraded DNA is to be used for DNA amplification. Therefore, the gene sequences used for barcoding should be short enough to be PCR-amplified easily. In general, DNA barcoding is based on the use of a short, standard region that enables cost-effective species identification. In plants, the mtDNA has a low substitution rates and a rapidly changing gene content and structure, which makes co1 unsuitable for barcoding. Therefore, based on the information available from phylogenetic studies, a number of gene sequences, both coding and non-coding sequences from the chloroplast DNA along with a gene from nuclear DNA have been examined for their suitability as barcodes. Although barcoding based on the biparentally inherited nuclear DNA segment is expected to provide more information on species identify, including hybridization events, till date internal transcribed spacer (ITS) regions of the ribosomal DNA(rDNA) are the only nuclear DNA that have been tested for suitability as barcodes in plants. The difficulty in obtaining high universality of the PCR amplification of single or low-copy genes, especially from degraded and low-quality DNA and the low species discriminatory power due to conservation of functional genes across large lineages could be the major reasons why such limited numbers of genes are being tested. Although numerous reports on the use of molecular markers including the nuclear and chloroplast sequences in banana have been reported the information on the barcodes are scanty (Carreel *et al*., 2002; Li *et al*., 2010).

Thus the present research proposal was envisaged to evaluate the feasibility of using nuclear and chloroplast gene sequences for DNA barcoding in Musaceae family.

1.6. Molecular markers:

Details of the used molecular markers for the evaluation of nuclear and chloroplast gene sequences for feasibility of DNA barcoding of banana are explained as follows:

Internal Transcribed Spacers regions of nuclear ribosomal cistron: The rDNA cistron is a multigene family encoding the nucleic acid core of the ribosome (Fig. 1). Within the cell, the rDNA is arranged as tandemly repeated units containing *18S,* 5.8*S,* 26*S* coding regions and two internal transcribed spacers (ITS1 & ITS2) present on either side of 5.8*S* region. Generally, the rDNA units are reiterated thousands of times and are organized into large blocks in the chromosome called the nuclear organizer regions. One of the most remarkable features of the rDNA is that the individual unit of this multiple gene family does not evolve independently, instead all the units evolve in a concerted manner such that higher level of overall sequence homogeneity exists among copies of the rDNA within a species, but differs among different species. This high sequence homogeneity is achieved through a process initially termed as horizontal evolution or

co-incidental evolution, but later renamed as concerted evolution which involves unequal crossing over and gene conversions. Currently, *nr*ITS is considered as one of the most useful phylogenetic markers for both plants and animals, because of its ubiquitous nature, biparental inheritance, and comparatively higher evolutionary changes due to less functional constrains. Likewise species-level discrimination and technical ease have also contributed to its wider acceptability as a powerful phylogenetic marker. Another advantage is that the ITS1 and ITS2 regions can be PCR-amplified separately by anchoring primers in the conserved coding regions. This facilitates easy amplification of ITS even from poor quality or degraded DNA. Universal primers are also available for amplification of ITS1 and ITS2 regions.

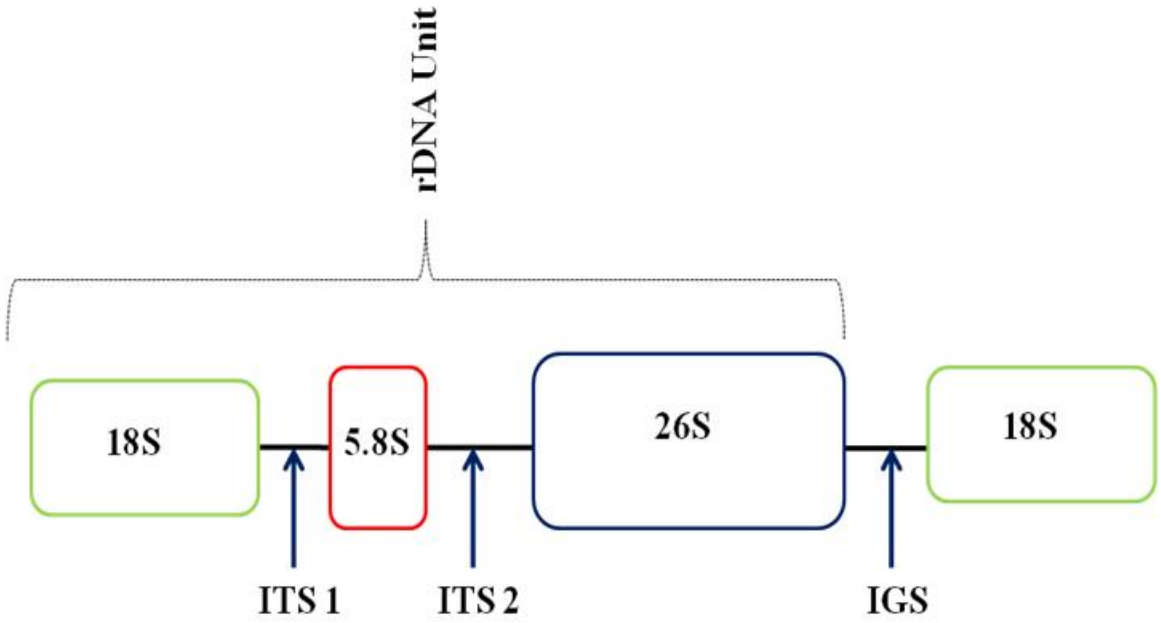


Figure 1: Generalized ribosomal DNA unit structure. The internal transcribed spacers are denoted as ITS1 and ITS2. The intergenic spacer is marked IGS. The 18S, 5.8S and 25S ribosomal RNA genes are indicated**.**

rbcL gene sequence: Among the plastid genes, rbcL is the best characterized gene sequence. Therfore, most of the investigating groups tested its suitability in barcoding. It encodes the large subunit of rubilose-1.5-biphosphate carboxylase/oxygenase (RUBISCO) . As RUBISCO is a critical photosynthetic enzyme, rbcL was the first gene that was sequences from the pants. rbcL has been used so extensively in plant phylogenetic studies that more than 10,000 rbcL sequences are already available in GeneBank. Because of this wide utility, various aspects of the molecular evolution of rbcL have also been studied in detail. Most of the phylogenetic studies suggest that rbcL is best suited to reconstruct the relationships down to the generic levels, but is not useful for species level. Furthermore, inorder to obtain enough species discrimination,, the

entire ~1430bp needs to be sequenced, which acts as a limiting factor for its use as a barcoding sequence because an ideal DNA barcoding region should be sort enough to amplify from degraded DNA and analysed via single-pass sequencing. Primers for PCR amplification and sequencing for such short sequence within the rbcL gene have been developed accordingly for most of the taxa. Owing to the ease in PCR amplification across a wide range of plant groups and the availability of sequence information in many plant groups, the CBOL-Plant Working Group has recently recognized rbcL as one of the most potential gene sequences for DNA barcoding in plants. However, because of the low species discrimination. Most of the investigating groups are of the opinion that rbcL should be used in conjunction with other markers. Therfore, CBOL-Plant Working Group recommended a combination of genes appears to be pragmatic solution to a complex trade-off among universality, sequence quality discrimination and cost. Although numerous studies on the use of molecular markers including the nuclear and chloroplast sequences in banana have been reported (Lannaud et al., 1992; Carreel et al., 2002; Ge et al., 2005; Li, 2010), the information on the availability of an universal barcode(s) has been missing.

Thus the present research proposal was envisaged to evaluate the feasibility of using nuclear and chloroplast gene sequences for DNA barcoding in Musaceae family.

1.7. OBJECTIVES:

- To test the universality of the nuclear and chloroplast gene sequence loci in Musaceae family.
- Evaluation of feasibility of DNA barcoding of Musaceae family.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2

MATERIALS AND METHODS

2.1. Plant materials: For the present study six (6) different banana accessions maintained in the field gene bank of Biotechnology Department, Mizoram University were selected. The details of the samples are given in Table 1.

2.2. DNA extraction:

Theories underlying extraction procedure: DNA is located within the chromosomes inside the nucleus and the complexity of the plant cell composition make the isolation of pure form DNA difficult. The cellular constituents accounting from cell wall to the cell membrane has to be broken in order to release the intact nucleic acids (DNA) into the extraction buffer. This is achieved by using detergents like SDS (Sodium Dodecyl Sulphate) or CTAB (Cetyltrimethyl Ammonium Bromide). During the extraction process the DNA should be protected from degradation by exogenous nuclease and this is accomplished using disodium salt of EDTA (Na_2EDTA) which is included in the extraction buffer to chelate magnesium ions, a necessary co-factor for most nucleases. After extraction, the concentration of extracted DNA can be checked by agarose gel electrophoresis. The purity of DNA could also determined by calculating the ratio of absorbance at 260 nm to that of 280 nm using spectrophotometer. The tissue buffer mixtures emulsified with chloroform-phenol denature and separate the proteins from DNA enabling DNA to be precipitated with either ethanol or Isopropanol. Shearing of DNA should be minimized and proper choice of the leaf tissue and developmental stage of plant is very important for DNA extraction. The use of very young leaf tissues has resulted in poor yields.

2.3. Materials for DNA extraction

Fresh young cigar leaves were collected from each plant maintained in the field gene-bank of the Department of Biotechnology, Mizoram University, Aizawl which represent the respective voucher number. The plant materials were subsequently used for the DNA extraction in the present study.

2.4. Chemicals and reagents used

1.0 M Tris-HCl pH 8.0 (Himedia Pvt. Ltd.)

Glacial acetic acid

5.0M NaCl

0.5M Na_2EDTA pH 8.0

10% SDS

Proteinase K (200 µg)(Bangalore Genei, India)

10% CTAB

10% PVP

Chloroform: Isoamyl alcohol (24:1) containing 5 % phenol

Absolute alcohol

Isopropanol

RNase (Sigma Aldrich Chemicals Pvt. Ltd., Bangalore, India)

10× Buffer (Bangalore Genei, India)

$MgCl_2$ (Bangalore Genei, India)

dNTP Mix (Bangalore Genei, India)

Taq DNA polymerase (Bangalore Genei, India)

6× gel loading buffer (Bangalore Genei, India)

Ethidium bromide: 10 mg/ml (Bangalore Genei, India)

2.5. Extraction Buffer: 200 mM Tris-HCl (pH 8.0), 0.8M NaCl, 25 mM Na_2EDTA (pH 8.0), 0.5 % SDS, 14 µg proteinase K

2.6. 2× CTAB Solution: 2 % CTAB, 100 mM Tris-HCl (pH 8.0), 20 mM Na_2EDTA pH 8.0, 1.4 M NaCl, 5 % PVP.

All the chemicals were purchased from Himedia Pvt. Ltd. India

2.7. DNA extraction protocol

Extraction of genomic DNA was carried out with the modified protocols of Thangjam *et al*. (2003) with minor modifications. Different concentrations of NaCl, Proteinase K and Polyvinyl pyrolidone (PVP) were used to optimize the extraction buffer.

2.8. Procedure for DNA extraction

1. Collect about 0.2 g of leaves from each cultivar, rinsed with tap water and blotted dry.

2. Placed the cultivar in the micro-centrifuged tube (2 ml) containing 400 µl of extraction buffer.

3. Grind the leaves in the buffer inside the tube with sterile glass rod.

4. Incubate at 37°C for at least 90 min in water bath.

5. Add 400 µl of 2 % CTAB solution and incubate at 65°C at least for 120 min.

6. Cool to room temperature then extract by gently adding equal amount of chloroform: isoamyl alcohol (24:1) containing 5% phenol.

7. Centrifuge at 12,000 rpm in a micro centrifuge at 4°C for 5 min.

8. Carefully transfer the upper aqueous layer to a new tube and repeat the extraction process 3 times to remove the cloudiness of the upper layer.

9. Add 2/3rd volume of ice cold isopropanol, mix gently by repeated inversions and incubate at room temperature for 30 min to precipitate DNA.

10. Centrifuge at 12,000 rpm for 5 min at 4°C and remove the supernatant.

11. Wash the pellet with ice cold 70 % ethanol.

12. Decant the supernatant and air-dry the pellet.

13. Resuspend the pellet in 100 µl of TE buffer and remove RNA by adding 1 µl of RNase(50 µg/ml).

14. Incubate at 37°C for approximately 60 min (An overnight RNAse treatment helped achieving in proper genomic DNA)

2.9. Estimation of quantity and quality of the extracted of DNA

The amount of isolated DNA per milligram of leaf tissue was estimated by measuring absorbance at 260 nm and 280 nm using Bio-photometer plus (Eppendorf, Germany) according to the manufacturer's instruction. The purity of DNA was determined by calculating the ratio of

absorbance at 260 nm to that of 280 nm. A 100 ng/µl DNA stock was prepared from the isolated DNA to be used for further experiments.

2.1.0. Electrophoretic separation of the total genomic DNA

The isolated DNA from each of the cultivars was separated on 0.8% agarose gel in 1× TAE buffer. For this, aliquot of 10 µl (1 µg) from 100 ng/µl DNA stock was loaded into the gel and then photographed using a gel documentation system (UVP Ltd, Cambridge, UK).

For PCR amplification a DNA stock of 10 ng/µl for each of the cultivars were prepared and stored at 4 °C until use.

2.1.1. PCR-amplification of the ITS2 region

For the PCR-amplification of the ITS2 region, genomic DNA of each of the 6 banana cultivars were amplified with ITS 2 primer (Chen *et al.,2010*) having the forward sequence of 5′-ATGTCACCACAAACAGAAAC-3′ and having the reverse sequence of 3′-GACGCTTCTCCAGACTACAAT-5′ purchased from Eurofins Genomics Pvt. Ltd., Bangalore, India. The PCR amplification was carried out on a Gene AMPPCR System 9700 (Applied Biosystems, USA) with 25 µl reaction mixture containing 1× PCR buffer, 100 ng of genomic DNA as template, 0.5 µM each of the primers, 1.5 mM of $MgCl_2$, 0.2 mM of dNTP, 1 U of Taq DNA polymerase(Bangalore Genei, India). The PCR reaction was set up with the condition of 94°C for 4 min, 35 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 1 min followed by a final extension at 72°C for 7 min. The amplified fragments were resolved on a 1.5% agarose gel and the gels were stained with ethidium bromide and visualized under UV light with a 100 bp DNA

ladder (Bangalore Genei, India) as a marker and then photograph using a gel documentation system.

2.1.2. PCR-amplification of the rbcL region

For the PCR-amplification of the rbcL region, genomic DNA of each of the 6 banana cultivars were amplified with rbcL primer (Chen *et al.,2010*) having the forward sequence of 5′-ATGTCACCACAAACAGAGACTAAAGC-3′ and having the reverse sequence of 3′-CTTCTGCTACAAATAAGAATCGATCTC-5′.The PCR amplification was carried out on a GeneAMPPCR System 9700 (Applied Biosystems, USA) with 25 µl reaction mixture containing 1× PCR buffer, 100 ng of genomic DNA as template, 0.5 µM each of the primers, 1.5 mM of $MgCl_2$, 0.2 mM of dNTP, 1 U of Taq DNA polymerase(Bangalore Genei, India). The PCR reaction was set up with the condition of 95°C for 4 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min followed by a final extension at 72°C for 7 min. The amplified fragments were resolved on a 1.5% agarose gel and the gels were stained with ethidium bromide and visualized under UV light with a 100 bp DNA ladder (Bangalore Genei, India) as a marker and then photograph using a gel documentation system.

2.13. Sequencing of the PCR fragments

The desired ITS2 and rbcL PCR fragments of 560 bp and 680 bp respectively generated by the six samples were sent directly for sequencing to GCC Biotech Pvt. Ltd. Kolkata, India.

2.1.4. Sequence analysis

The raw sequences obtained were first checked for its accuracy and edited by comparing with the associated chromatogram using Bioedit software ver. 7.5.1.0 (Hall, 1999). Homology search of the sequences was performed using the Blastn (Altschul et al., 1990) against the nucleotide GenBank database (http://www.ncbi.nlm.nih.gov) using the megablast program to search for the homologous sequences. The identified and annotated sequences were aligned using the multiple sequence alignment tool of the DNAMAN software package ver. 6.0.3 (Lynnon, Quebec, Canada). Analysis of nucleotide composition, Tajima's neutrality test (Tajima, 1989), pairwise distance (Tamura and Nei, 1993) and phylogenetic analysis using maximum likelihood method using MEGA software ver 5.0.(Tamura et al., 2011). Evaluation of the ability of the ITS2 and the rbcL gene sequences to analyse the inter-species divergence were performed using Wilcoxon signed-rank test tool of the SPSS software ver. 16.0. For all these analyses 2 species of the genus *Ensete* viz., *E. glaucum* and *E. ventricosum* were taken as an outgroup with their sequences (both ITS2 and rbcL) retrieved from the genbank database (www.ncbi.nlm.nih.gov/genbank).

Table 1: List of banana plants used for the study

Sl. No.	Voucher no.	Name of the species	Section
1.	MZUTRS-13	*Musa* ABB group cv. Banpawl	Eumusa
2.	MZUTRS-14	*Musa* AAB group cv. Balhlasen	Eumusa
3.	MZUTRS-01	*Musa balbisiana* BB group cv. Changthir	Eumusa
4.	MZUTRS-08	*Musa ornata* cv. Changvandawt	Rhodoclamys
5.	MUTRS-28	*Musa laterita* cv. Red	Rhodoclamys
6.	MUTRS-29	*Musa laterita* cv. Green	Rhodoclamys

CHAPTER 3

RESULTS

CHAPTER 3

RESULTS

3.1. Quality of the extracted genomic DNA:

The quality of the extracted genomic DNA from the six samples were evaluated by loading 1 µg of the DNA and electrophorised on 0.8% agarose gel. The electrophoregram showed a distinct sharp band of high molecular weight DNA (Fig. 2)

3.2. PCR amplification of the ITS2 region:

Using the optimized PCR condition the amplification of the genomic DNA of all the 6 banana cultivars resulted in the generation of a single sharp fragment of around 560 bp (Fig. 3).

3.3. PCR amplification of the rbcL region:

Using the optimized PCR condition the amplification of the genomic DNA of all the 6 banana cultivars resulted in the generation of a single sharp fragment of around 680bp (Fig. 4).

3.4. Sequence Analysis for ITS2 sequences:

The raw sequences obtained from the 6 samples on PCR amplification with the IT2 primers were obtained and analyzed with the corresponding chromatogram (Fig. 5).

Table 2 showed the sequence information of the unedited and edited ITS2 sequences obtained from the 6 banana samples**.** The unedited sequences of the samples showed variable length in the order of Changthir (486 bp) > Lailaphu2 (461 bp) > Changvandawt (429 bp) > Lailaphu1 (394 bp) > Balhlasen (375 bp) > Banpawl (368 bp).

The edited sequences obtained after analysis through Blastn and Blast2seq tools were in order of Lailaphu2 (312 bp) > Changthir (282 bp) > Lailaphu1 (241 bp) > Banpawl (215 bp) > Changvandawt (214 bp) > Balhlasen (203 bp).

The homology search result of the edited sequences using MegaBlast program of the Blastn algorithm showed high value of identity (97-99%) with the Genbank database of ITS2 sequences from different banana species (Table 3). For the sequence analysis, the reference ITS2 sequences of *E. glaucum*(GI: 239584269) and *E. ventricosum*(GI: 332714902) retrieved from the genebank database were used as an outgroup (Table 4).

The details of the nucleotide composition of the ITS2 sequences obtained from the 6 banana samples and the reference sequences of *Ensete* are given in Table 5. The average length of the sequences is 210.1 and the nucleotide composition were in the order of G (39.9) > C (30.4) > T (16.3) > A (13.4). *Musa laterita*1 and *Musa laterita*2 have the highest content of T nucleotide, *Musa ornata* cv Changvandawt have the highest content of C nucleotide, *Musa* AAB group cv. Balhlasen have the highest content of A nucleotide and *Musa laterita*1 have the highest content of G nucleotide.

Multiple sequence alignment of all the ITS2 sequences including the samples studied and the outgroup revealed variability in sequence sites including nucleotide substitutions and indels occurred (Fig. 6).

The ITS2 sequences were then again aligned using a software MEGA 5.0 in order to study the variable nucleotide sites occurring in the ITS2 sequences among the banana samples studied (Table 6)**.** In this study, *Musa laterita*2 (Rhodochlamys) was used as a model. With the analysis of sequence variation in the commonly existing sequence, a total of 50 variable nucleotide sites

were found where a sum total of 89 nucleotide changes was found to be occurring among the banana samples studied. In each site it was found that some of the nucleotides have been changed which could be either deletion or insertion as when compared to the model sequences i.e *Musa laterita*2. Among the banana samples studied, *Musa ornata* cv. Changvandawt (Eumusa) have the most variable nucleotide changes of 36, while *Musa* ABB group cv.Banpawl (Eumusa) have the least variable nucleotide changes of 10.

The result of the Tajima's neutrality test of the ITS2 sequences obtained from the banana samples studied along with the reference sequences is shown in Table 7. The result obtained shows that m=8 (number of sequences) , *S*=41 (number of segregating sites) , ps= 0.251534 (*S*/m) , Θ=0.097010 (ps/a1) , π=0.091805 (nucleotide diversity) , *D*=-0.287237 (D is the Tajima test statistic.)

The pairwise genetic distance calculated using Tamura and Nei's (1993) model revealed the maximum distance between *M. ornata* (Rhodochlamys) and *E. glaucum* with a value 0.202 and the minimum value of 0.032 was observed in 3 cases between *Musa* ABB group cv.Banpawl and *Musa* AAB group cv. Balhlasen, *Musa* ABB group cv. Banpawl and *M.laterita2, M. laterita1* and *M.laterita2* (Table 8)**.**

The relative distribution between intra- and inter-species divergence of the ITS2 sequences in the banana samples was represented in a bar diagram (Fig. 7)**.** The genetic distance within the range of 0.020 – 0.040 was observed to have the highest distribution with 2 instances at the intra-species level. At the interspecies level, the range of genetic distance between 0.160 - 0.180 was observed to have the highest occurrence with 6 cases.

The phylogenetic relationships generated by the analysis of the ITS2 sequences using Maximum Likelihood (ML) method revealed the grouping of the samples into 2 clusters (Fig. 8). One cluster shows a bootstrap value of 99% existing between *E.glaucum* and *E.ventricosum* while the other cluster shows 4 sister clusters with a bootstrap value of 79% between Changtir (Eumusa) and Changvandawt (Rhodochlamy), 48% between *Musa Laterita1* and *Musa Laterita2* (Rhodochlamys) ,and 41% between Banpawl (Eumusa) and Balhlasen (Eumusa).

The homology search result of the edited sequences using MegaBlast program of the Blastn algorithm showed high value of identity (95-99%) with the Genbank database of rbcL sequences from different banana species (Table 12**).** For the sequence analysis, the reference rbcL sequences of *E. glaucum* (GI: 227808823) and *E. ventricosum* (GI: 37722337) retrieved from the genbank database were used as an outgroup (Table 13).

3.5. Sequence Analysis for rbcL sequences:

Table 9 showed the sequence information of the unedited and edited sequences obtained from the 6 banana samples. The unedited sequences of the samples showed variable length in the order of Lailaphu2 (665bp) > Banpawl (660bp) > Changthir (614) > Balhlasen (597 bp) > Lailaphu1(575bp) > Changvandawt (413bp). The edited sequences after analysis through Blastn and Blast 2seq tools were in the order of Lailaphu2 (521bp) > Banpawl (520bp) > Balhlasen (513bp) > Changthir (510bp) > Lailaphu1(491bp) > Changvandawt (337bp).

The homology search result of the edited sequences using MegaBlast program of the Blastn algorithm showed high value of identity (95-99%) with the Genbank database of rbcL sequences

from different banana species (Table 10)**.** For the sequence analysis, the reference rbcL sequences of *E. glaucum*(GI: 227808823) and *E. ventricosum*(GI: 37722337) retrieved from the genbank database were used as an outgroup (Table 11).

The details of the nucleotide composition of the rbcL sequences obtained from the 6 banana samples and the reference sequences of *Ensete* are given in Table 12. The average length of the sequences are 487.4 and the nucleotide composition were in the order of T (29.3) > A (26.8) > G (22.9) > C (21.0). *M.ornata* cv. Changvandawt has the highest content of T nucleotide (29.9) and the lowest was found in the 2 cultivars of *M. laterita* with 28.9. The highest number of C was observed in Banpawl (21.4) and the lowest in Changthir (20.3). For the A nucleotide the highest content was observed in *M. laterita* cv. Red (27.6) and the lowest in *M. ornata* (25.1). The highest content of G was found in *M. ornata* cv. Changvandawt (23.9) and the lowest in *M. laterita* cv. Red (22.3).

Multiple sequence alignment of all the rbcL sequences including the samples studied and the outgroup revealed variability in sequence sites including nucleotide substitutions and indels occurred (Fig. 9).

To study the variable nucleotide sites occurring in the rbcL sequences among the banana samples studied *Musa* ABB group cv. Banpawl (Eumusa) was used as a model for reference. With the analysis of sequence variation in the commonly existing sequence, a total of 39 variable nucleotide sites were found where a sum total of 96 nucleotide changes was found to be occurring among the banana samples studied. In each site it was found that some of the nucleotides have been changed which could either deleted or inserted as when compared to the

model sequence. Among the banana samples studied, *M. ornata* cv.Changvandawt (Rhodochlamys) have the most variable nucleotide changes of 28, while *E. ventricosum* have the least variable nucleotide changes of 4. The most number of nucleotide changes occur at the 44^{th} and 246^{th} nucleotide sites (Table 13).

The Tajima's neutrality test of the rbcL sequences obtained from the banana samples studied was also carried out which is shown in the Table 14. In this test, the reference sequences of *Ensete spp.* obtained from the Genebank was used. The results obtained shows that m=8 (m=number of sequences), *S*=22 (number of segregating sites), ps=0.026188 (*S*/m ,Θ=ps/a1), π=0.020723 (nucleotide diversity) , *D*=-1.091682*(D* is the Tajima test statistic).

The pairwise genetic distance calculated using Tamura and Nei's (1993) model revealed the minimum distance between *M. laterita*1(Rhodochlamys) and *M. laterita*2(Rhodochlamys) with a value of 0.003. The maximum distance value of 0.046 was observed in 3 cases between *Musa* AAB group cv.Balhlasen and *Musa* BB group cv.Changthir,*Musa* AAB cv.Balhlasen and *M.ornata* cv.Changvandawt,*Musa* BB group cv.Changthir and *E.ventricosum* (Table 15).

The relative distribution between intra- and inter-species divergence of the rbcL sequences in the banana samples was represented in a bar diagram (Fig. 10)**.** The genetic distance within the range of 0.000 – 0.020 was observed to have the highest distribution with 2 instances at the intra-species level and 15 at the interspecies level.

The phylogenetic relationships generated by the analysis of the rbcL sequences using Maximum Likelihood (ML) method revealed the same grouping of the sequences into 2 clusters with a low bootstrap value (Fig. 11) as compared to that generated by rbcL sequences. One cluster shows a bootstrap value of 75% existing between *E.glaucum* and *E.ventricosum* while the other cluster

shows 4 sister clusters with a bootstrap value of 42% between Banpawl (Eumusa) and Changthir (Eumusa) , 9% between Balhlasen (Eumusa) and*M. laterita2* (Rhodochlamys) and Balhlasen (Eumusa) , 44% between *M.laterita1*(Rhodochlamys) and sister clad(Balhlasen, laterita 2) and 19% between Changvandawt (Rhodochlamys) and another sister clad (Banpawl, Changthir).

3.6. Overall analysis of the ITS2 and rbcL sequences:

The overall analysis of the genetic distance observed at the intra- and inter-sectional levels of the Musaceae family using the ITS2 and rbcL genetic sequences obtained in the study are represented in (Table 16 and Figure 12). The overall genetic distances across all the sections were observed to be 0.0918 with ITS2 and 0.021 with rbcL respectively. Comparison of the ITS2 and rbcL sequences in terms of its ability to generate inter-species divergence was calculated with Wilcoxon signed-rank test revealing the superiority of ITS2 over the rbcL (Table 17)**.**

Table 2: Details of the edited and unedited ITS2 sequences from the banana samples studied

Sample	Raw sequence (length in bp)	Edited sequence (length in bp)
Musa ABB group cv.Banpawl	GCTTCGAAGGTTTTGAAGGCAAGTGGGCTCWAAGSCATCCGGTAATGKACGCCTGCCTGGCGTCAGGTAA AGAAATTCGKGTTGCYCCTCGGGGGGSGGGGSTTTTKCGGAGGTGCCCCCCGTGCGRGGTTGGTTGGCMK GAGGSGCTCGTGTTYGAAGAGGGGGGWTGCCTGGRAACCTAYCSTGCTTGGGACCGGCGGGCCTGGGACC CAATCGGGGGGWTCATGCACGAACGCMCCCCGGTSGGGGGGCCCCGTGGATTAAGCTATAATAGGMGGGA GAAAACTTMAGGTTCCTTTAGASGGSGAGGAACGGGACSCCCYTTTGAAATCGGGGTGGCTCTTCTGATT TAATCGGGAAAACAACAA **(368)**	CTTTCGACGCTTCTTCGTTGCCCCCTCGGGGGGGCGGGGGCGAATGC GGAGGATGGCCCCCCGTGCCGGAAGGTGCGGTTGGCCGAAGACCGGG CCGTCGATGGTTGTCGAACACGACGCGTGGTGGATGCCTTGTGCGAG CCGTACGTCGTGCCTTCGGGACCCGGGCGAGGCCTCGAGGACCCAAG TCGTGGTGCGAGTCGATGCCACGGACC **(215)**
*Musa*AAB group cv.Balhlasen	GCGTTCATAGTTTTGGAYTAAGAAGYGACGAAGCATRCCGKMCATGGCCGCTGCTGGMTACAGAAACGAG GCTTCRTCGTTGCCCCTCGGGGGTGGGMTCTAGYGAGGAGGWCCCCGTGCSGAAGGMGCCSTCCGCCAAG ACGGCTTSTGGTTKCCGACAGATCCTGGGGATWCTTTGKSRASCSGMSKYKKGCTYSGGRACSGGGGGGC CTCRGGCCCAAKSKKGKSKRASYKATGCMCGRACGCACCCSGGYRGGGGGSYYCCCSYKGGTTASMWWTA AWARSGGAGGGAAGAWTTRSRGKTTCCYTTATAMSGSGGSRMCGGGATCGCCACTTGAKATSGSSAYGCY YYTTGATTSARYCAGRAAASTMCAA **(375)**	GGCACGCCTGCCTGGGCGTCACGCTTTCGACGCTTCGTCGTTGCCCC CTCGGGGGGGGTGGGGGCGAACGCGGAGGATGGCCCCCCCGTGCCGG AAGGTGCGGTTGGCCGAAGAGCGGGCCGTCGATGGTTGTCGAACACG ACGCGTGGTGAATGCCTTGTGCGAGCCGTACGTCGTGCCTTCGGAAC CCGGGCGAGGCCTCG **(203)**
Musa BB group cv.Changthir	TGGTCTTAAARGTTMTTTGWAYKCAAGTTGCGCCCGAGGCCMATCCSGGCTAAGGGCACGCCTGCGTGGG CGTCAGCAATSGAGGGGTCGTCGTTGCCCCCTCGGGGGGGGGGGGGGCAAASAAGGAGGKTGGCCCCCCG KGCCGGAAGGTGGGGTTGSCCAAAAARCGGCCCSTCGTTGTTTTTCAAAACCATWCCGGGGGGAATGCTT GTTSSAARCCTTAYTCTGGGCCTTCGGGACCGGGGGGAAGGGCCTTCGGAACCAAATYCGGGGGGGGGGG TCARGGCCCSGGAACCCRACCCCCCGGGGTGGGGGGGGGGGCTCCCCCGYGGGTTTTAAMWWATAAAAAA AARGGGGGGGGGAAAAAAAAATTTMGGGGGGTTTCCCTTTTARTAAGGGGGGGGGGACCGGGGAWTCMC CCCMYTTTGGAAAAKGGGGGGGGKGGGSTTCTTTTTGATTTTGTAMTTGGAAAAAAAMCACAAAAA **(486)**	TTTGAACGCAAGTTGCGCCCGAGGCCATCCGGCTAAGGGCACGCCTG CCTGGGCGTCACGCTTTCGACGCTTCGCCGTTGCCCCCTCGAGGGGT GGGGGCGAACGCGGAGGATGGCCCCCCGGGCCGGAAGGTGCGGTCGG TCGAAGAGCGTGCCGTCGGTGGTTGTCGAACACGACGCGTGGTGGAT GCCTTGTGCGAGCCGTACGTCGTGCCTTCGGGACCCGGGCGAGGCCT CGAGGACCCAAGTCGTGGTGCGAGTCGATGCCACGGACCGCGACCCC **(282)**
*Musa ornata*cv.Changvandawt	CATGGTWRTKTGRKKGKTGTAATMGWGSTTTGGMYAGWWGYCCCGAGGCATCCGGCTTTGCCGCCTGGCT GGGGTCAGAAACGGAGTTCACGTTGCCCCCCTSGGGGTGGTACTAGCGGGGGAGGCCCCCCTGCCGGAGG TCGTCGCCGAAGAGCGGGCGTCGGTGGTTCGAACACGTTYCGGGGGGATGCTTTGTGCGACCKTACSTKG TGCCTTCGGGCCCCGGGCGAGGCCTCGGGGACCAAGTCGTGGTGSGAGTCGATGCCACGRCCCGCSACCC CGGGYMGGGGGGSYCCCCCCGCTGATTTAARCATATAAATASGCGGAGGAAAAAACTTWCGGGTTTTCTC TTAGYSGGGGGGGRGCCGGGWATSSCCCRYTTTGAAAYCGGGCGTTTGCSCCGTTAGATTTAAATCTAGA AAACCCAAA **(429)**	TTCACCGTTGCCCCCCTCGGGGTGTGGGGATGAACGCGGAGGATGGC CCCCCGTGCCGGAAGGTGCGGTCGGCCGAAGAGCGGGCCGTGGGTGT TTGTCGAACACGACGCGTGGGTGGATGCCTTGTGCGAGCCGTACGTC GTGCCTTCGGGCCCCGGGCGAGGCCTCGGGGACCCAAGTCGTGGTGC GAGTCGATGCCACGGCCCGCGACCCC **(214)**
Musa laterita cv. Red (Lailaphu 1)	CAATCCTTGGGAGGTATGCGAATACGCACCYGAGTCTMTCGAGCTATGGGAAGCTTGCTTGGGGTTAGAA AAGAAGCGTGTCGTTGCCCCGTCGGGGGKGGGTTTTAACGAGGGGGAGGGCCCCCCGTGKCGAAGGTGCG GTTGGCCRAGGAGCGGSCCGCGKTGGTTGTCGAAATTTTTTCTGGTGWAGCTTTGYCRAYCTTACGTCTT SCCTTCGGGCCGGGCGAGCCTCGAGACCMTGTCGTGGTGGAGCGAGCCACGACGCSACCCGGTCAGGGGG GCTWCCCTGGTTTAAAAATAAAAGGGGAGGAAGAACCTTCGGTTTTTCTTGAACGCGAGASCCGGTTAGC CCGGTTARAATGGGCGCTCSCCCTTAATTGAGTCGGAGAACACC **(394)**	CTTTCGACGCTTCGTCGTTGCCCCGTCGGGGGGTGGGGGCGAACGCG GAGGATGGCCCCCCGTGCCGGAAGGTGCGGTTGGCCGAAGATCGGGC CGTCGGTGGTTGTCGAACACGACGCGTGGTGGATGCTTTGTGCGAGC CGTACGTCGTGCCTTCGGAAGCCGGGCGAGGCCTCGAGGACCCATGT CGTGGTGCGAGTCGATGCCACGGACCGCGACCCCAGGTCAGGTGGGG CTACCC **(241)**
Musa laterita cv. Green (Lailaphu2)	TGGGTCAGGAGTCTTTGACGCAAGTTGCGCCGAGGCCATCCGGCTAAGGGCACGCCTGCCTGGGCGTCAC GCTTTCGACGCTTCGTCGTTGCCCCCTCGGGGGGTGGGGSCGAAMGCGGAGGATGGCCCCCCGTGTCGGA AGGTGCGGTTGGCCGAAGAGCGGGCCGTCGGTGGTTGTCGAACACGACGCGTGGKGRATGCCTTGTGCGA GCCGTACGTCGTGCCTTCGGGACCCGGGCGAGGCCTCGAGGACCCAAGTSGTGGTGCGAGTCGATGCCAC GGACCGCGACCCCRGGTCRGGKGGGSCTMCCCGCTGAGTTTAAGCATATAAATAAGCGGAGGARRAAGAA ACTTACGAGGATTCCCTTAGTAACGGSGAGCGAACCGGGATCRGCCCAGCTTGARRAATSGGGCGGCYTG CGTCGYCKGAATTTGWAGTCTGGWRRAAASCRWACAAAAAG **(461)**	GAGTCTTTGAACGCAAGTTGCGCCCGAGGCCATCCGGCTAAGGGCAC GCCTGCCTGGGCGTCACGCTTTCGACGCTTCGTCGTTGCCCCCTCGG GGGGTGGGGGCGAACGCGGAGGATGGCCCCCCGTGCCGGAAGGTGCG GTTGGCCGAAGATCGGGCCGTCGGTGGTTGTCGAACACGACGCGTGG TGGATGCCTTGTGCGAGCCGTACGTCGTGCCTTCGGGACCCGGGCGA GGCCTCGAGGACCCAAGTCGTGGTGCGAGTCGATGCCACGGACCGCG ACCCCAGGTCAGGTGGGGCTACCCGCTGAG **(312)**

Table 3: Result of the Blast search for similarity between the ITS2 sequences obtained from the banana samples studied and GenBank data carried out using the MegaBlast program of the Blastn algorithm.

Sl. No.	Sample	Gene Bank ID & Sequence name	Sequence identity	E value
1	*Musa* ABB group cv.Banpawl	GI:332715029 Internal transcribed Spacer2	99%	3e-139
2	*Musa*AAB group cv.Balhlasen	GI:332715004 Internal transcribed Spacer 2	99%	1e-97
3	*Musa* BB group cv.Changthir	GI:343405463 Internal transcribed Spacer 2	99%	2e-141
4	*Musa ornata* cv. Changvandawt	GI:213136971 Internal transcribed Spacer 2	97%	7e-95
5	*Musa laterita* cv. Red (Lailaphu 1)	GI:213136957Intenal transcribed Spacer 2	98%	7e-122
6	*Musa laterita* cv. Green (Lailaphu2)	GI:213136957Intenal transcribed Spacer 2	99%	3e-166

Table 4: Details of ITS2 reference sequences of *Ensete spp.* obtained from Genbank (NCBI) used as outgroup for the analysis.

GI No.	Species name	Sequence length (bp)
239584269	*Ensete glaucum*	TTTCGACGCTTCGTCGGCACCCCCCCGGGGGTCCGCAGGCGAACGCGGAGGATG GCCCCCCGTGCCCGAAGGTGCGGTTGGCCGAAGAGCCGGCCGTCGGTGCTCGTC GGACACGACGCGTGGTGGATGCCCCGTGCGAACCGTACGTCGTTCCTTCGGGAC CCGAACGAGGCCTCGAGGACCCAAGCCGTGGTGCGATTCGATGCGACGCCACGG **(216)**
332714902	*Enseteventricosum*	CTTTCGACGCTTCGTCGGTGCCCCCCGGGGTGTAGGCGAATGCGGAGGATGGCC CCCCGTGCCCAAAGGTGCGGTAGGCCGAAGAGCTGGCCGTCTGTGCTCGTCGGA CACGACGCGTGGTGGATGCCTTGTGCGAACCGTACGTCGTTCCTTCGGGACCCG AACGAGGCCTCGAGGACCCAAGCCGTGGTGCGAGTCGATCGATGCCACGGACC **(215)**

Table 5: Details of the nucleotide composition of ITS2 sequences obtained from the banana samples studied.

Sample	**Nucleotide composition**				
	T	**C**	**A**	**G**	**Total**
Musa laterita cv. Green (Lailaphu 2)	16.5	29.5	13.1	40.9	237.0
Musa ABB group cv.Banpawl	17.1	29.3	13.2	40.5	205.0
*Musa*AAB group cv.Balhlasen	17.1	28.8	12.4	41.8	170.0
Musa BB group cv.Changthir	15.6	30.7	13.2	40.6	212.0
Musa ornata cv. Changvandawt	16.4	31.3	12.1	40.2	214.0
Musa laterita cv. Red (Lailaphu 1)	16.9	28.6	13.0	41.6	231.0
Enseteventricosum	17.1	30.7	15.6	36.6	205.0
Ensete glaucum	14.0	34.3	14.5	37.2	207.0
Average	16.3	30.4	13.4	39.9	210.1

Table 6: Variable nucleotide sites occurring in the ITS2 non-coding regions among banana samples studied.

Nucleotide site	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**	**76**	**78**
Sample																			
Laterita 2	C	G	T	C	A	C	G	C	T	T	T	C	G	A	C	G	C	T	C
Laterita 1	G	.	C	G	T	T	C	.	.	.	A	A	.	.	A	.	.	G	-
Banpawl	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Balhlasen	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Changthir	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Changvandawt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.
E. ventricosum	.	.	.	.	.	G	.	.	.	-	-	.	.	T	.	T	-	-	-
E. glaucum	.	.	.	.	.	G	C	A	.	-	-	C	.	.	.	T	C	C	.

Nucleotide site	**79**	**80**	**90**	**91**	**92**	**93**	**94**	**99**	**100**	**101**	**106**	**107**	**111**	**123**	**131**	**147**	**150**
Laterita 2	G	T	C	T	C	G	G	T	-	-	G	C	C	-	T	T	C
Laterita 1	.	.	G	.	.	.	.	.	-	-	.	.	.	-	.	.	.
Banpawl	T	.	.	.	.	-	.	G	G	C	.	.	T	-	.	.	.
Balhlasen	.	.	.	.	.	.	.	G	G	T	.	.	.	C	.	.	.
Changthir	.	C	.	.	.	.	A	.	-	-	.	.	.	-	G	C	T
Changvandawt	A	C	.	C	T	-	C	.	G	T	A	T	.	-	.	C	.
E.ventricosum	G	T	.	.	T	.	.	.	.	.	.	.	.	-	.	.	A
E. glaucum	G	C	.	.	.	.	.	.	.	.	.	.	.	-	C	C	A

Nucleotide site	**157**	**160**	**166**	**168**	**171**	**190**	**195**	**200**	**229**	**230**	**231**	**247**	**255**	**281**
Laterita 2	T	G	C	G	G	-	G	C	G	A	C	A	A	A
Laterita 1	.	.	.	.	.	-	.	T	A	.	G	.	T	.
Banpawl	C	.	.	A	.	-	.	.	.	.	.	.	.	.
Balhlasen	G	.	.	A	.	-	A	.	A	.	.	-	-	-
Changthir	G	T	.	.	.	-	.	.	.	.	.	.	.	.
Changvandawt	G	.	G	.	T	G	.	.	.	C	.	G	.	C
E.ventricosum	A	A	.	.	.	C	.	G	T	C	C	A	.	G
E. glaucum	A	A	.	.	.	C	T	G	.	C	C	A	.	G

"." means identity to the nucleotide on the top row; **"-"** means nucleotide indel

Table 7: Result of the Tajima's Neutrality Test of the ITS2 sequences obtained from the banana samples studied.

M	*S*	p_s	*Θ*	*Π*	*D*
8	41	0.251534	0.097010	0.087423	-0.529071

***m* = number of sequences; *S* = Number of segregating sites; $p_s = S/m$; $\Theta = p_s/a_1$; π = nucleotide diversity; *D* is the Tajima test statistic**

Table 8: Pairwise genetic distance (Tamura and Nei, 1993) generated by ITS2 sequences.

Species 1 (section)	Species 2 (section)	Distance	Std. Err (±)
Banpawl {eumusa}	Balhlasen {eumusa}	0.032	0.015
Banpawl {eumusa}	Changthir {eumusa}	0.076	0.026
Balhlasen {eumusa}	Changthir {eumusa}	0.069	0.024
Banpawl {eumusa}	Changvandawt {rhodochlamys}	0.109	0.034
Balhlasen {eumusa}	Changvandawt {rhodochlamys}	0.110	0.034
Changthir {eumusa}	Changvandawt {rhodochlamys}	0.083	0.028
Banpawl {eumusa}	Laterita1 {rhodochlamys}	0.068	0.023
Balhlasen {eumusa}	Laterita1 {rhodochlamys}	0.053	0.019
Changthir {eumusa}	Laterita1 {rhodochlamys}	0.083	0.027
Changvandawt {rhodochlamys}	Laterita1 {rhodochlamys}	0.126	0.037
Banpawl {eumusa}	Laterita2 {rhodochlamys	0.032	0.01
Balhlasen {eumusa}	Laterita2 {rhodochlamys}	0.033	0.016
Changthir {eumusa}	Laterita2 {rhodochlamys}	0.046	0.019
Changvandawt {rhodochlamys}	Laterita 2 {rhodochlamys}	0.084	0.028
Balhlasen {eumusa}	Changthir {eumusa}	0.069	0.024

Species 1 (section)	Species 2 (section)	Distance	Std. Err (±)
Laterita1 {rhodochlamys}	Laterita2 {rhodochlamys}	0.032	0.015
Banpawl {eumusa}	Enseteventricosum {ensete}	0.164	0.047
Balhlasen {eumusa}	Enseteventricosum {ensete}	0.179	0.052
Changthir {eumusa}	Enseteventricosum {ensete}	0.182	0.047
Changvandawt {rhodochlamys}	Enseteventricosum {ensete}	0.194	0.050
Laterita 1 {rhodochlamys}	Enseteventricosum {ensete}	0.191	0.050
Laterita 2 {rhodochlamys}	Enseteventricosum {ensete}	0.143	0.040
Banpawl {eumusa}	Enseteglaucum {ensete}	0.165	0.047
Balhlasen {eumusa}	Enseteglaucum {ensete}	0.162	0.049
Changthir {eumusa}	Enseteglaucum {ensete}	0.178	0.052
Changvandawt {rhodochlamys}	Enseteglaucum {ensete}	0.202	0.058
Laterita 1 {rhodochlamys}	Enseteglaucum {ensete}	0.174	0.049
Laterita 2 {rhodochlamys}	Enseteglaucum {ensete}	0.128	0.039
Enseteventricosum {ensete}	Enseteglaucum {ensete}	0.077	0.025
Laterita1 {rhodochlamys}	Laterita2 {rhodochlamys}	0.032	0.015

Table 9: Details of the edited and unedited rbcL sequences from the banana samples studied.

Sample	Raw sequence (length in bp)	Edited sequence (length in bp)
Musa ABB group cv.Banpawl	ACSRTTYTSTTGKGGAGTAAAGATAACAKATAARAMKTATTRTACTCCTGACTACGAGYCAAAGATACTGATATTTTGGCA GCATTCCGAGTAACTCCTCAACCTGGAGTTCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTAC ATGGACAACTGTGTGGACTGATGGACTTACCAGYCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTG GGGAGGAAAATCAATAKATTGCTTATGTAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTTC CATTGTGGGTAATGTATTTGGTTTCAAAGCCTTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAA ACTTTCCAAGGCCCGCCTCACGGCATTCAKRTTGAAAGAGATWWKTTGAACAAGTATGGTCGTCCCCTATTGGGATGCAC TATTAAACCAAAATTTGGGATTATCTGCAAAAACTACGGTAGAGCTGTTTTTATGATGTCTACGTGATGACTTTSMYYYYM CCAAAGATGATGAAACGTAACTCACATCATTWTGSCTGGAGAGATCGATYCTTATTGTARGCARRAGAGACTTCGTCCTG KTTGGGGGGSKGAC **(660)**	GAACTTATTATACTCCTGACTACGAAGTCAAAGATACGTGACATCTTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGT TCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGACTT ACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTATG TAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTATTTGGTTTCA AAGCCTTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAAACTTTCCAAGGCCCGCCTCACGGC ATTCAGGTTGAAAGAGATAAGTTGAACAAGTATGGTCGTCCCCTATTGGGATGCACTATTAAACCAAAATTGGGATTAT CTGCAAAAAACTACGGTAGAGCTGTTTTTATGAATGTCTACG **(520)**
Musa AAB group cv.Balhlasen	GGGRTTTCCRGCTGGTGGGTTGAGGATAACATATTGAMKTATTATWCTCCTGACTACGAGCCAAAGATAYTGATAATTTG GCAGCATTCCGAGTACTCCTCACCTGGAGTTCCGCCCGAAGAAGCAGGGGCTGCGKTAGCTGCYGWATCTTCTACTGGTA CATGRACAACTGTGTGGACTGATGKACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTG GGGAGGAAAATCAATATATTGCTTATGTAGCTTATCYTTTAGACTATTTTGAGAAGGTTCTGTWACTAACATGTTTACTTC CATTGKGGGTAATGWTTTGGTTTCAARGCCTTACGAGCTCWACGTCTGGGGATCTGCRAATYCCSMCTTCTTATTCCAAA YTTTCARGSYCYGTCTCAGGSWMTWAGATTGAAGARWWAKTTGAMAGTWKGYCKYCCCTTTGGRTGCCWATAAACCAA AWTGGRWTAYCTGCAAAAACTACGWAGAGCGTTWAGRAKGYTACTGTGACTTGATTTACCAAGTGRTGRAASGAAMTC CCACCATTATGSKTGGAGRRYGMTCTATTGKRGMRAGRT **(597)**	GGGATTTCCAGCTGGTGGGTTGAGGATAACATATTGAATTATTATTCTCCTGACTACGAGCCAAAGATATTGATAATTTG GCAGCATTCCGAGTACTCCTCACCTGGAGTTCCGCCCGAAGAAGCAGGGGCTGCGTTAGCTGCCGAATCTTCTACTGGTA CATGGACAACTGTGTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTT GGGGAGGAAAATCAATATATTGCTTATGTAGCTTATCCTTTAGACTATTTTGAGAAGGTTCTGTTACTAACATGTTTACTT CCATTGTGGGTAATGTTTTGGTTTCAAAGCCTTACGAGCTCTACGTCTGGGGATCTGCGAATTCCCACTTCTTATTCCAAA CTTTCAGGGCCTGTCTCAGGGAATAAGATTGAAGAGAAATTTGACAGTAGGTCGTCCCTTTGGGTGCCAATAAACCAAA TTGGATTATCTGCAAAAACTACGAAGAGCGTT **(513)**
Musa BB group cv.Changthir	TCTCAGCATSTCAKACCWGCACGTAKTRYACKYMWGWMTACGAAGYCAAAKAYAYTRMYATAAGGCAGCATTCAKRGT ACTCCTSACCTGGAGTTCCGCCCGAAGAAGCAGGGGCTGCGKWAGCTGCMGWATCTTCTACTGGTACATGGAGAACTGT GTGGACTGATGGMCTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGRGGAGGAAAATC AATATATTGCTTATGTAGCTTATCCTTTAGACYTTTTTGAAGAAGGTTCTGTWACTAACATGTTTACTTCCATTGTGGGTAA TGTATTTGGTTTCAAAGCCTTACGAGCTCTACGTCTGGAGGATCTGCRAATTCCSAYTTCTTATYCMAAAAYKTKCTWAGS YCYGYSTSWSGGSAYWYAKRTTGAAAGAGATAWKTTGAACAAGTATGGTCGTCCCCTATTGGGATGCACTATTAAACCA AATTGGGATTATCTGCAAARACTACGGTAGAGCGGTTTRTGATGTCTACGTGTGACTTGATTTYACCAAGATGATGAAACG TAACTCACACCATTTATGCGTTGAGAGATCGATTTCTTWATTGWRGCAGAAGGA **(614)**	GACTTATTATACTCCAGAATACGAAGCCAAAGATACTGCCATAAGGCAGCATTCAGGGTACTCCTCACCTGGAGTTCCG CCCGAAGAAGCAGGGGCTGCGTTAGCTGCMGAATCTTCTACTGGTACATGGAGAACTGTGTGGACTGATGGACTTACCA GTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTATGTAGCT TATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTATTTGGTTTCAAAGCC TTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCAAAAACTTTCCAAGGCCTGTCTCAGGGGATTCA GATTGAAAGAGATAAGTTGAACAAGTATGGTCGTCCCCTATTGGGATGCACTATTAAACCAAATTGGGATTATCTGCAA AAACTACGGGTAAGCGGTTTAATATGTCTACG **(510)**
Musa ornata cv. Changvandawt	ARGATTCTTATTTGTAGGAGGAGAARARAYSACKTTYRYACKYSAGAATACGAARCAAWGGTATTGCTAAAAGGAAACA GTCGGRGTASTCCWCACCTGRTTTCCGCCCGCMGAAGCAGGKGCTGCKTWAGCYGCYATTCTTTTCTGGWACATtgCAACT GTGTGGACTGATGKCTTATTTTCTTGATCGTTACKGGGACGACATACRACCTCSAGCCCGTTGTGGRGCAGTAAAACCAAT ATATGCTTTAGGCGCCTTTTCTTTTAGACATATTARAARAGGTTCWGKWACAATCGGTATGCTTCCATTGRGGGGATGTTT TGTTCCAAGCCGTTGKGAGCTCATTCYATTGGGAATCTTCAACACCCGCTTTCATCCCAAAATTGTGTTTWCTCTGTTTTTG GGAAATTAA **(413)**	AAATTGACTTATTATACTCCTGACTACGAAGTCAAAGATACTGATATCTTGGCAGCATTCGGAGTAACTCCTCAACCTGG AGTTCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGA CTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTA TGTAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTAACTAACATGTTTACTTCCATTGTGGGTGATGTATTTGGTTT CCAAGCCTTACGAGCTC **(337)**
Musa laterita cv. Red	ASARAWCYYYGMTGTGTGWGYAARAGGAAAATWRWMTTATTATACTCCTGACTACGAAGTCAAAGATACTGATATTTG GCAGCATTCCGAGTAACTCCTCAACCTGKAKTTCCGCCCGAAGAAGCMGGGGCTGCGGWAGCTGCCGAATCTTCTACTGG TACATGKACAACTGTGTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCSAGGCCGTTGT TGGGGAGGAAATCAATATATTGCTTATGTAGCTTATCCTTKAGACYTTTTGAAGAAGGTTCTGTTACTAACATGTTRYTTC MATGTGGGTAWTGTATTKGTTTCAAGCYTTACSAGCTCWASTCKGAGRWCTGSRAWTYCCACTTYTTATTCAAAACTTTC ARGSCCGCYYMASGCMTYMRGTTGAAGAYAAGTTAAAAKKTGGCSTCCCWATGGGTGCMYATTAAMCAAWTGGATATC KSAMAAAMWMCGKAARCGTTTWKRTGYRYGTGTGGACTGWTTTACAAMWKKKGAACYKAACTCCACYTTTTGCSTYGG RYSWTCTTTGTGCASAARA **(575)**	AAATTGAATTATTATACTCCTGACTACGAAGTCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCTGG AGTTCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGA CTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTA TGTAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTAACTAACATGTTTACTTCCATTGTGGGTGATGTATTTGGTTT CCAAGCCTTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAAACTTTCCAAGGCCCGCCTCACGG CATTCAGGTTGAAAGAGATAAGTTGAAAAAGTATGGTCGTCCCCTATTGGGATGCACTATTAAACCAAAATTGGGAATA TCTGCAAAAAA **(491)**
Musa laterita cv. Green	GSGKTTTTYTTMTGGTKTAKGGAGGAAAAAAATTGATTATTATACTCCTGTACTACGAAGTCAAAGATACTGATATATGGC AGCATTCCGAGTAACTCCTCAACCTGGAGTTCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTA CATGGACAACTGTGTGGACTGAKGGACTTACCAGYCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTT GGGGAGGAAAATCAATATATTGCTTATGTAGCTTATCCTTTARACYTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTT CCATTGTGGGTAATGTATTTGGTTTCAAAGCCTTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTTCTTATTCCA AAAYTTTCMAGGCCCGCCTCACGGCWTTCAGGTTGAAAGAGATAWGTTTGAACAAKTATGGTCSKYCCCCTATTGGGAT GCMCTATTAAAACCAAAATTKGGGAWTATCTGCAAAAAACYACGGGTAGAGCGGKTTTATGRATGTCTACGGTGGTGGA CYTTGATTTYTMCCCAAGATTRATGAGAACGKTAAACYTCACATCCTATTTTATGCGTTGGAGAGAYCSRWTTTCTTAWTT TGGAGRGSSAAGAARGAATTT **(665)**	AAATTGAATTATTATACTCCTGACTACGAAGTCAAAGATACTGATATCATGGCAGCATTCCGAGTAACTCCTCAACCTGG AGTTCCGCCCGAAGAAGCAGGGGCTGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGA CTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTA TGTAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTATTTGGTTTC AAAGCCTTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAAACTTTCCAAGGCCCGCCTCACGG CATTCAGGTTGAAAGAGATAAGTTGAACAAGTATGGTCGTCCCCTATTGGGATGCACTATTAAACCAAAATTGGGATTA TCTGCAAAAAACTACGGTAGAGCGGTTTATGAATGTCTACG **(521)**

Table 10: Result of the blast search for similarity between the rbcL sequences obtained from the Banana samples studied and GenBank data carried out using the MegaBlast program of the Blastn algorithm

Sl. No	**Species name**	**Gene Bank ID & Sequence name**	**Sequence identity**	**E value**
1	Musa ABB group cv.Banpawl	GI:227808689Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	99%	0.0
2	Musa AAB group cv.Balhlasen	GI:156598412Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	99%	0.0
3	Musa BB group cv.Changthir	GI:227808689Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	95%	0.0
4	*Musa ornata* cv. Changvandawt	GI:227808811Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	99%	2e-174
5	*Musa laterita* cv. Red	GI:227808791Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	99%	0.0
6	*Musa laterita* cv. Green	GI:227808791Ribulose-1'5-bisphosphate carboxylase/oxygenase large subunit(rbcL)	99%	0.0

Table 11: Reference rbcL sequences used as outgroup for the analysis.

GI. No	Species name	Species length(bp)
227808823	*Enseteglaucum*	TACAAATTGACTTATTATACTCCTGACTACGAAGTCAAAGATACTGATATCTTGG CAGCATTCCGAGTAACTCCTCAACCTGGAGTTCCGCCTGAAGAAGCAGGGGCTGC GGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGACTT ACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGGGG AGGAAAATCAATATATTGCTTATGTCGCTTATCCTTTAGACCTTTTTGAAGAAGG TTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTATTTGGGTTCAAAGCC TTACGAGCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAAACTT TCCAAGGCCCGCCTCACGGCATTCAGGTTGAAAGAGATAAGTTGAACAAGTATGG TCGTCCCCTATTGGGATGTACTATTAAACCAAAATTGGGATTATCTGCAAAAAAC TACGGTAGAGCGGTTTATGAATGTCTACGAGGTGG **(530)**
37722337	*Enseteventricosum*	ATGTCACCACAAACAGAGACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAAG ATTACAAATTGACTTATTATACTCCTGACTACGAAGTCAAAGATACTGATATCTT GGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTCCGCCTGAAGAAGCAGGGGCT GCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGATGGAC TTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCACATCGAGGCCGTTGTTGG GGAGGAAAATCAATATATTGCTTATGTAGCTTATCCTTTAGACCTTTTTGAAGAA GGNTCTGTTNCTAACATGTTTACTTCCATTGTGGGTAATGTATTTGGTTTCAAGG CCTTACGANCTCTACGTCTGGAGGATCTGCGAATTCCCACTTCTTATTCCAAAAC TTTCCAAGGCCCGCCTCACGGCATTCAGGTTGAAAGAGATAAGTTGAACAAGTAT GGTCGTCCCCTATTGGGATGTACTATTAAACCAAAATTGGGATTATCTGCAAAAA ACTACGGTAGAGCGGTTTATGAATGTCTACGTGGTGGACTTGATTTTACCAAAGA TGATGAAAACGTGAACTCACAACCATTTATGCGTTGGAGAGACCGTTTCTTATTT TGTGCCGAAGCAATTTATAAAGCGCAGGCCGAAACAGGTGAAATCAAAGGACATT ACTTGAATGCTACTGCGGGTACATGTGAAGAAATGATGAAAAGGGCCGTATGTGC CAGAGAATTAGGAGTTCCTATCGTAATGCACGACTACTTAACTGGTGGATTCACT GCAAATACTAGCTTGGCTCATTATTGCCGTGACAACGGCCTACTTCTTCACATCC ATCGCGCAATGCATGCAGTTATTGATAGACAGAAAAATCATGGTATGCATTTCCG TGTACTAGCTAAAGCATTACGTATGTCTGGTGGAGATCATATTCACGCCGGTACA GTAGTAGGTAAACTGGAAGGGGAACGTGAGATGACTTTAGGTTTCGTTGATTTAT TACGTGATGATTATATTGAAAAAGACCGAAGTCGCGGTATTTTCTTCACTCAAGA TTGGGTCTCTATGCCAGGTGTTCTGCCCGTGGCTTCAGGGGGTATTCATGTTTGG CATATGCCTGCTCTGACCGAAATCTTTGGGGATGATTCCGTACTACAGTTT **(1206)**

Table 12: Details of the nucleotide composition rbcL sequences obtained from the banana samples studied.

Sample	**Nucleotide composition**				
	T	**C**	**A**	**G**	**Total**
Musa ABB group cv.Banpawl	29.3	21.4	26.6	22.6	518.0
Musa AAB group cv.Balhlasen	29.7	20.8	26.6	22.9	515.0
Musa BB group cv.Changthir	29.1	20.3	27.4	23.2	508.0
Musa ornata cv. Changvandawt	29.9	21.1	25.1	23.9	331.0
Musa laterita cv. Red	28.9	21.2	27.6	22.3	485.0
Musa laterita cv. Green	28.9	21.2	27.2	22.7	515.0
Ensete glaucum	29.3	21.2	26.6	22.9	515.0
Ensete ventricosum	29.5	21.1	26.6	22.9	512.0
Average	29.3	21.0	26.8	22.9	487.4

Table 13: Variable nucleotide sites occurring in the rbcL coding regions among samples.

Sample																					
Nucleotide site	**2**	**10**	**15**	**18**	**26**	**35**	**36**	**39**	**40**	**43**	**44**	**56**	**58**	**104**	**112**	**135**	**246**	**3742 5 5**	**302**	**317**	**370**
Banpawl	C	A	T	C	T	C	G	A	C	C	T	C		G	C	C	C	T	A	A	C
Balhlasen	A	T	.	.	.	T	.	.	T	.	.	T	.	T	.	.	T	.	.	.	.
Changthir	.	.	A	A	C	.	-	C	.	A	A	A	G	T	M	G	.	A	.	.	A
Changvandawt	.	.	.	.	.	.	-	.	T	.	.	G	-	-	.	-	-	-	T	C	-
Laterita 1	A	.	.	.	.	.	-	.	T	.	.	.	.	.	.	.	.	.	.	.	.
Laterita 2	A	.	.	.	.	.	-	.	T	.	A	.	.	.	.	.	.	.	.	.	.
E. glaucum	.	.	.	.	.	.	.	.	-	.	T	.	.	.	.	.	T	.	.	.	.
E. ventricosum	.	.	.	.	.	.	.	.	-	.	T	.	.	.	.	.	T	.	.	.	.

Nucleotide site	**382**	**387**	**389**	**394**	**397**	**404**	**424**	**454**	**473**	**493**	**494**	**495**	**499**	**506**	**507**	**508**	**509**	**510**
Banpawl	A	C	C	C	C	G	C	C	T	T	A	G	T	A	T	G	A	A
Balhlasen	G	.	.	.	.	.	.	A	.	.	.	.	G	.	.	.	.	G
Changthir	.	T	T	G	G	A	.	.	.	G	T	A	G	-	A	A	T	.
Changvandawt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Laterita 1	.	.	.	.	.	.	A	.	A	-	-	-	-	-	-	-	-	-
Laterita 2	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
E. glaucum	.	C	.	G	.	.	.	.	.	.	.	.	.	.	G	.	.	.
E.ventricosum	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.

"." means identity to the nucleotide on the top row; "-" means nucleotide indel

Table 14: Result of the Tajima's Neutrality Test of the rbcL sequences obtained from the banana samples studied.

m	***S***	p_s	Θ	π	***D***
8	22	0.067901	0.026188	0.020723	-1.091682

m = number of sequences; S = Number of segregating sites; $p_s = S/m$; $\Theta = p_s/a_1$;

π = nucleotide diversity; D is the Tajima test statistic

Table 15: Pairwise genetic distance (Tamura and Nei,1993) generated by rbcL sequences in the samples studied.

Species 1(section)	Species 2(section)	Distance	Std.Err
Banpawl {eumusa}	Balhlasen {eumusa}	0.019	0.008
Banpawl {eumusa}	Changthir {eumusa}	0.032	0.010
Balhlasen {eumusa}	Changthir {eumusa}	0.046	0.012
Banpawl {eumusa}	Changvandawt {rhodochlamys}	0.016	0.007
Balhlasen {eumusa}	Changvandawt {rhodochlamys}	0.029	0.009
Changthir {eumusa}	Changvandawt {rhodochlamys}	0.046	0.012
Banpawl {eumusa}	Laterita1 {rhodochlamys}	0.006	0.004
Balhlasen {eumusa}	Laterita1 {rhodochlamys}	0.013	0.006
Changthir {eumusa}	Laterita1 {rhodochlamys}	0.039	0.011
Changvandawt {rhodochlamys}	Laterita1 {rhodochlamys}	0.016	0.007
Banpawl {eumusa}	Laterita2 {rhodochlamys	0.009	0.005
Balhlasen {eumusa}	Laterita2 {rhodochlamys}	0.016	0.007
Changthir {eumusa}	Laterita2 {rhodochlamys}	0.036	0.010
Changvandawt {rhodochlamys}	Laterita 2 {rhodochlamys}	0.019	0.007

Species 1(section)	Species 2(section)	Distance	Std.Err
Laterita1 {rhodochlamys}	Laterita2 {rhodochlamys}	0.003	0.003
Banpawl {eumusa}	Enseteventricosum {ensete}	0.013	0.006
Balhlasen {eumusa}	Enseteventricosum {ensete}	0.026	0.009
Changthir {eumusa}	Enseteventricosum {ensete}	0.046	0.012
Changvandawt {rhodochlamys}	Enseteventricosum {ensete}	0.022	0.008
Laterita 1 {rhodochlamys}	Enseteventricosum {ensete}	0.013	0.006
Laterita 2 {rhodochlamys}	Enseteventricosum {ensete}	0.016	0.007
Banpawl {eumusa}	Enseteglaucum {ensete}	0.009	0.005
Balhlasen {eumusa}	Enseteglaucum {ensete}	0.022	0.008
Changthir {eumusa}	Enseteglaucum {ensete}	0.042	0.011
Changvandawt {rhodochlamys}	Enseteglaucum {ensete}	0.019	0.007
Laterita 1 {rhodochlamys}	Enseteglaucum {ensete}	0.009	0.005
Laterita 2 {rhodochlamys}	Enseteglaucum {ensete}	0.013	0.006
Enseteventricosum {ensete}	Enseteglaucum {ensete}	0.009	0.005

Table 16: Analysis of intra and inter sections of Musaceae using ITS2 and rbcL gene sequences

Genetic Distance (Tamura and Nei, 1993)	ITS2 (Mean ± S.E.)	rbcL (Mean ± S.E.)
Overall genetic distance	0.0918 ± 0.0129	0.21 0.004
Distance within groups		
i)Eumusa	0.0725 ± 0.0235	0.032 ± 0.008
ii)Rhodoclamys	0.0788 ± 0.0231	0.012 ± 0.005
iii)Ensete	0.0807 ± 0.0308	0.009 ± 0.005
Distance between groups		
i)Eumusa + Rhodoclamys	0.069 ± 0.016	0.023 ± 0.006
ii)Rhodoclamys + Ensete	0.172 ± 0.041	0.015 ± 0.005
iii)Ensete + Eumusa	0.172 ± 0.042	0.026 ± 0.006
Net between groups		
i)Eumusa + Rhodoclamys	-0.001 ± 0.004	0.001 ± 0.002
ii)Rhodoclamys + Ensete	0.064 ± 0.018	0.004 ± 0.003
iii)Ensete + Eumusa	0.072 ± 0.019	0.005 ± 0.004

Table 17: Wilcoxon signed-rank test for inter-species divergence among loci.

W+	W-	Relative ranks, n, P value	Result
ITS2	rbcL	W+= 3, W- = 0, n = 3, p=0.109	ITS2 > rbcL

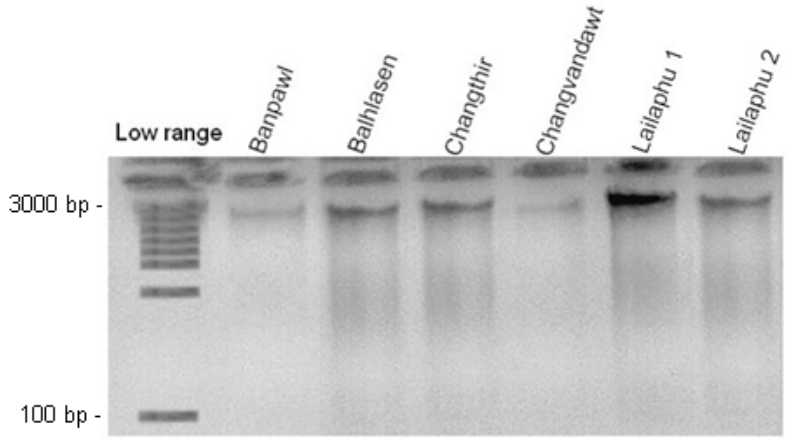


Figure 2: Electrophoregram of the extracted genomic DNA from the banana samples electrophorised on a 0.8% agarose gel

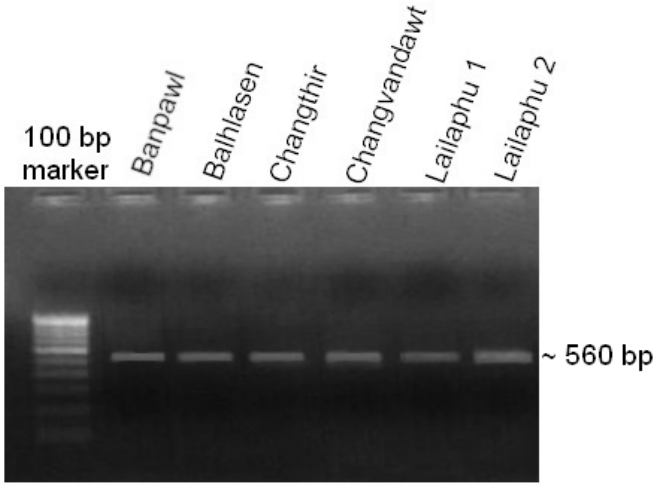


Figure 3: PCR amplification profile of the samples generated by ITS2 primers

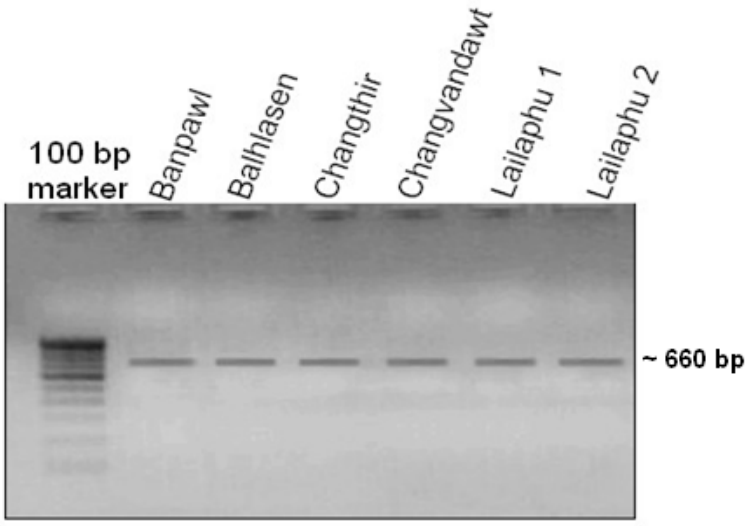


Figure 4: PCR amplification profile of the samples generated by rbcL primers.

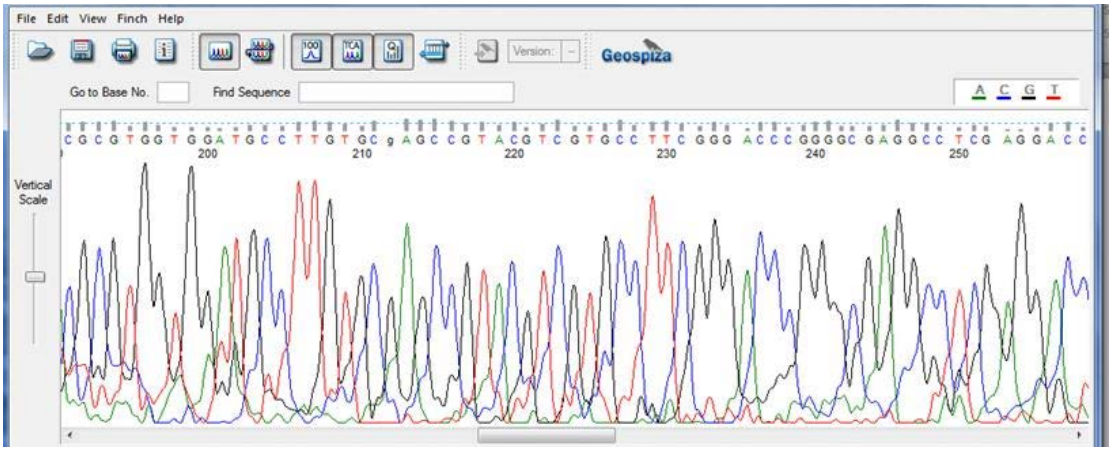


Figure 5: Chromatogram of the raw ITS2 sequences of the Musa ABB group cv.Banpawl.

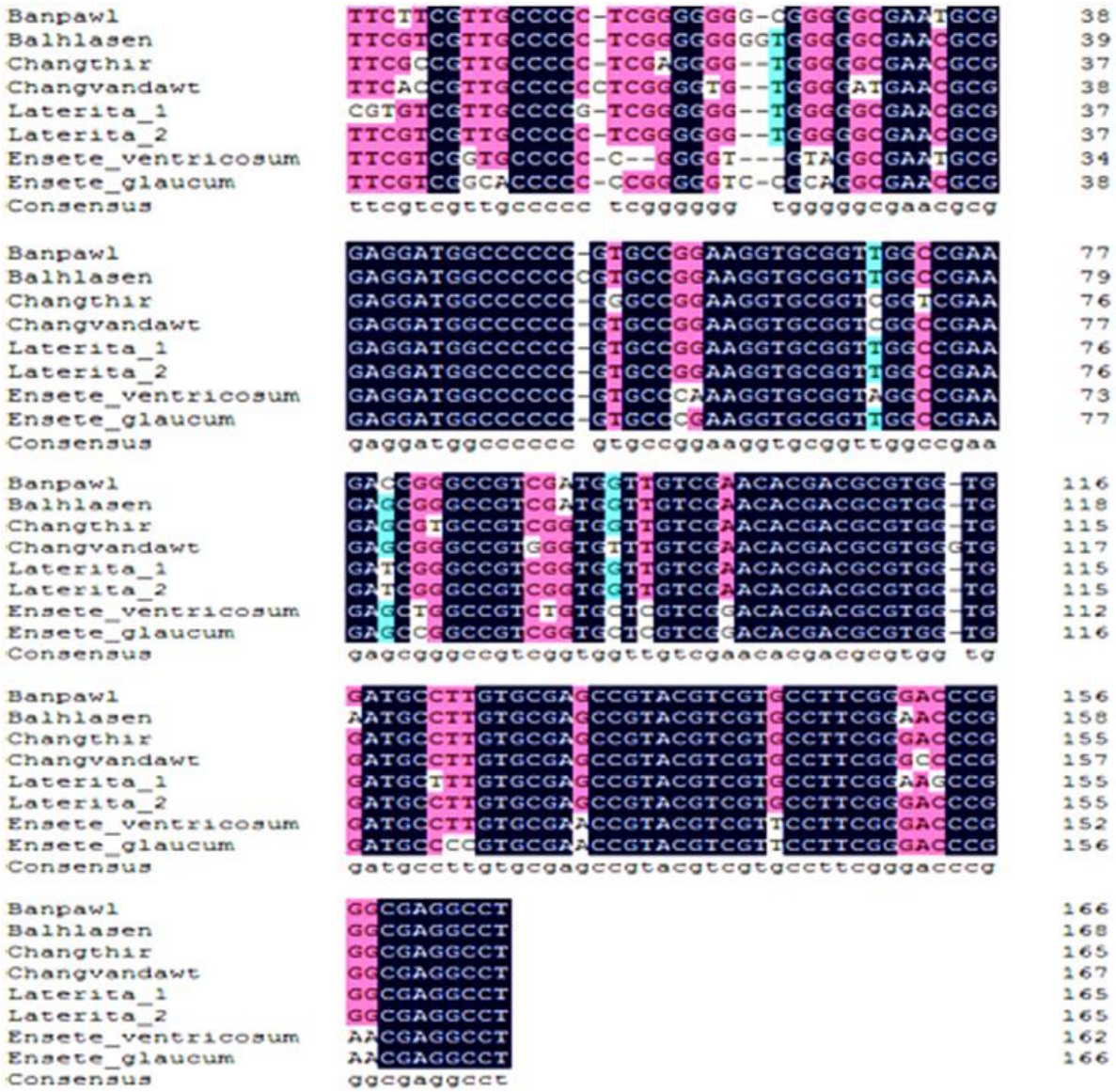


Figure 6: Multiple sequence alignment of the ITS2 sequences.

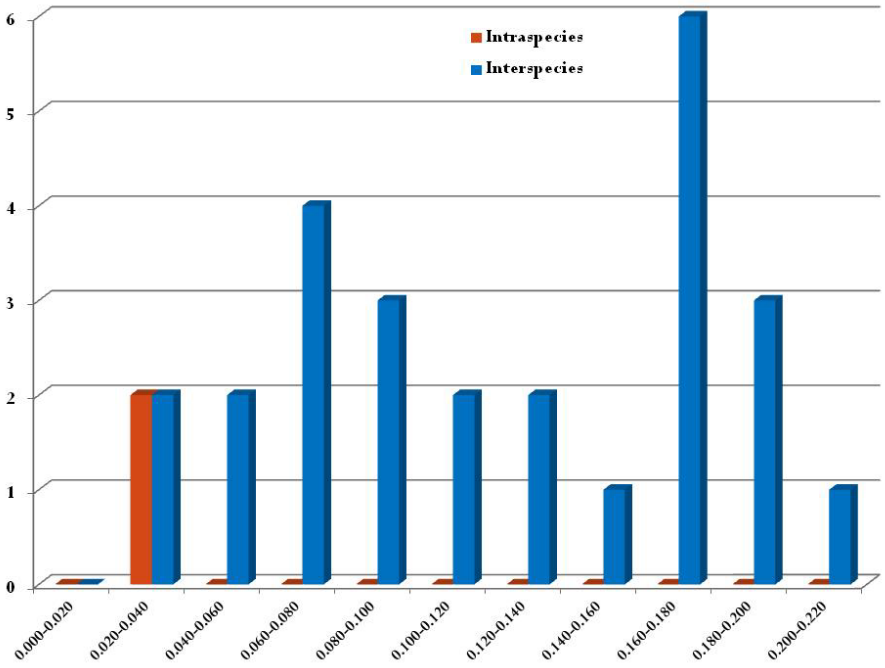


Figure 7: Relative distribution between intra- and interspecies divergence of the ITS2 sequences using 0.020 barcode distance units.

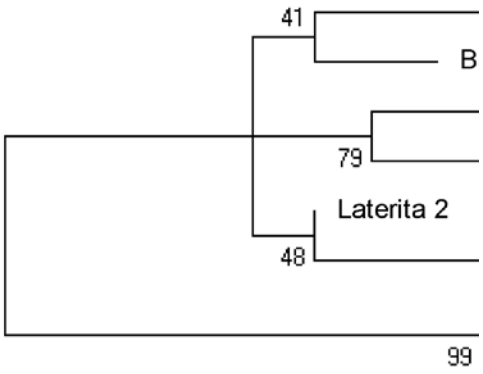


Figure 8: Phylogenetic tree generated by analysis of the ITS2 sequences using Maximum Likelihood method.

Figure 9: Multiple alignment of the rbcL sequences analyzed.

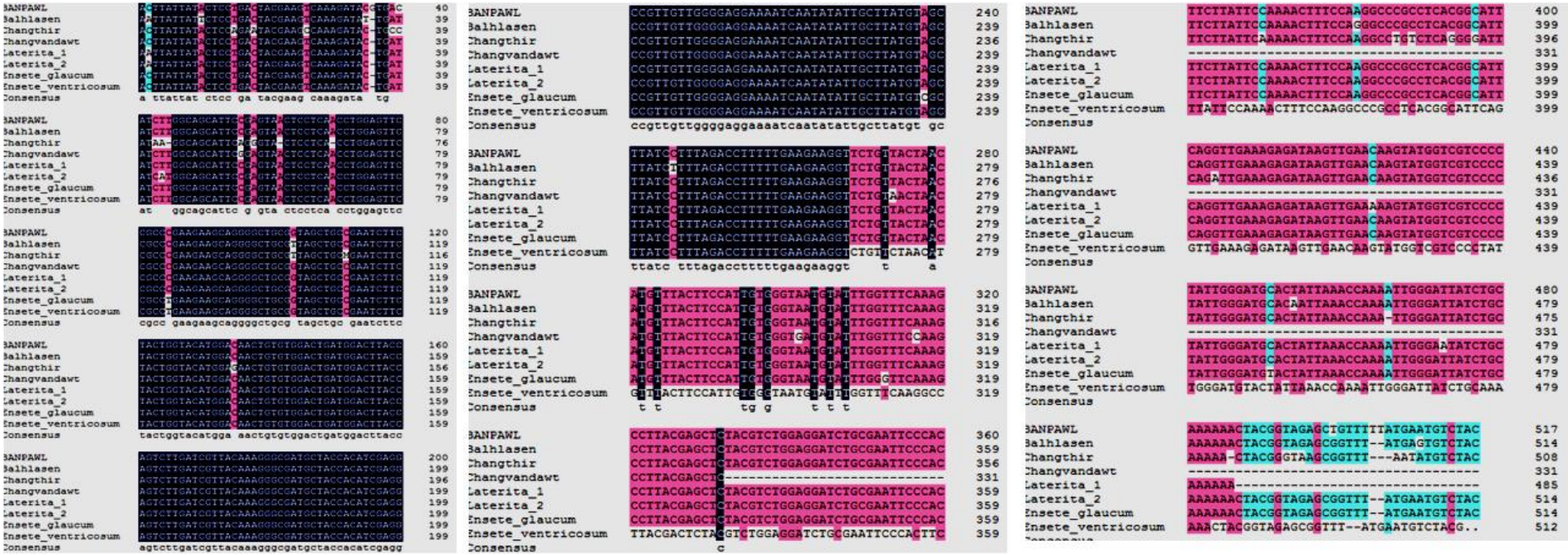
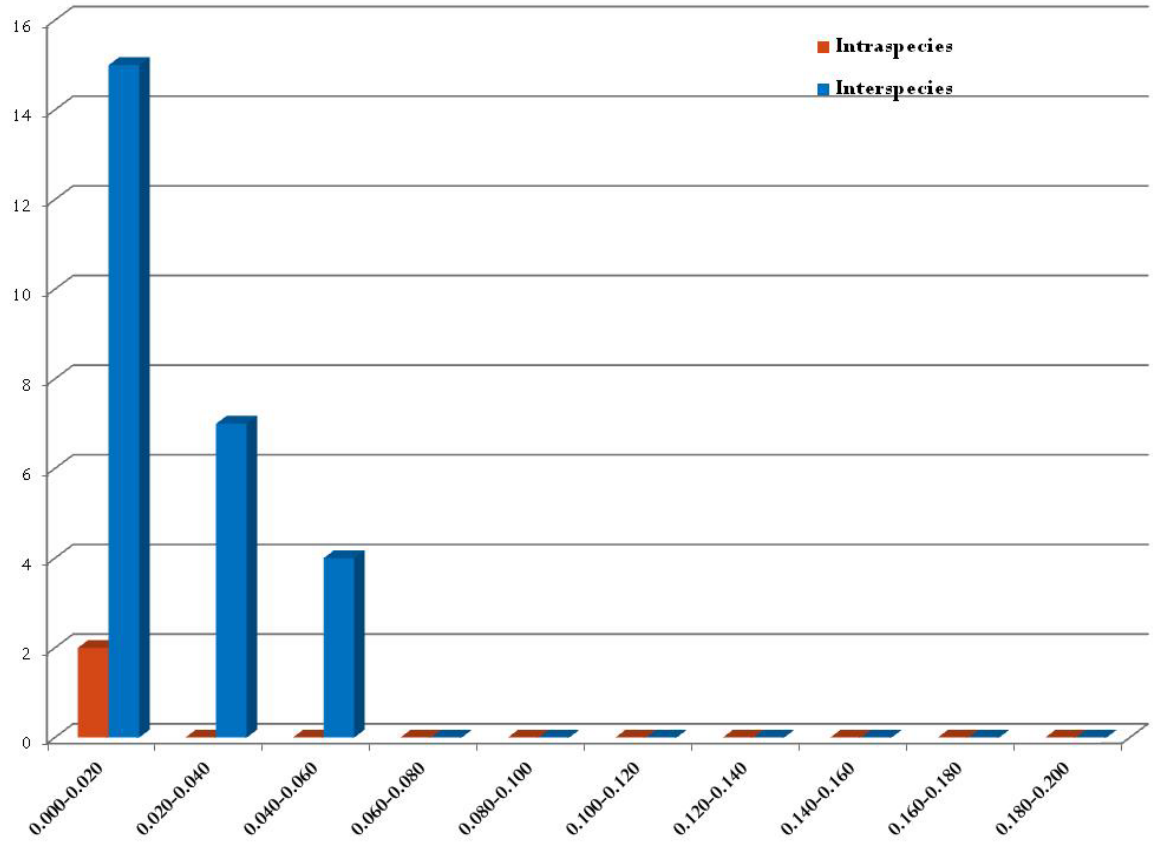


Figure 10: Relative distribution between intra- and inter-species divergence of the rbcL sequences using 0.020 barcode distance units.



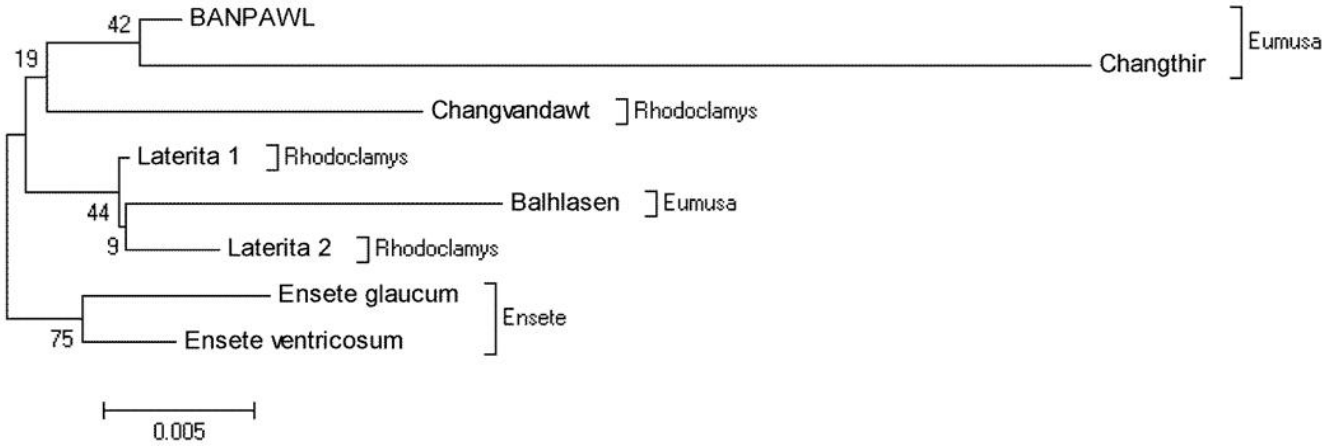


Figure 11: Phylogenetic relationship revealed by rbcL sequences of the banana samples.

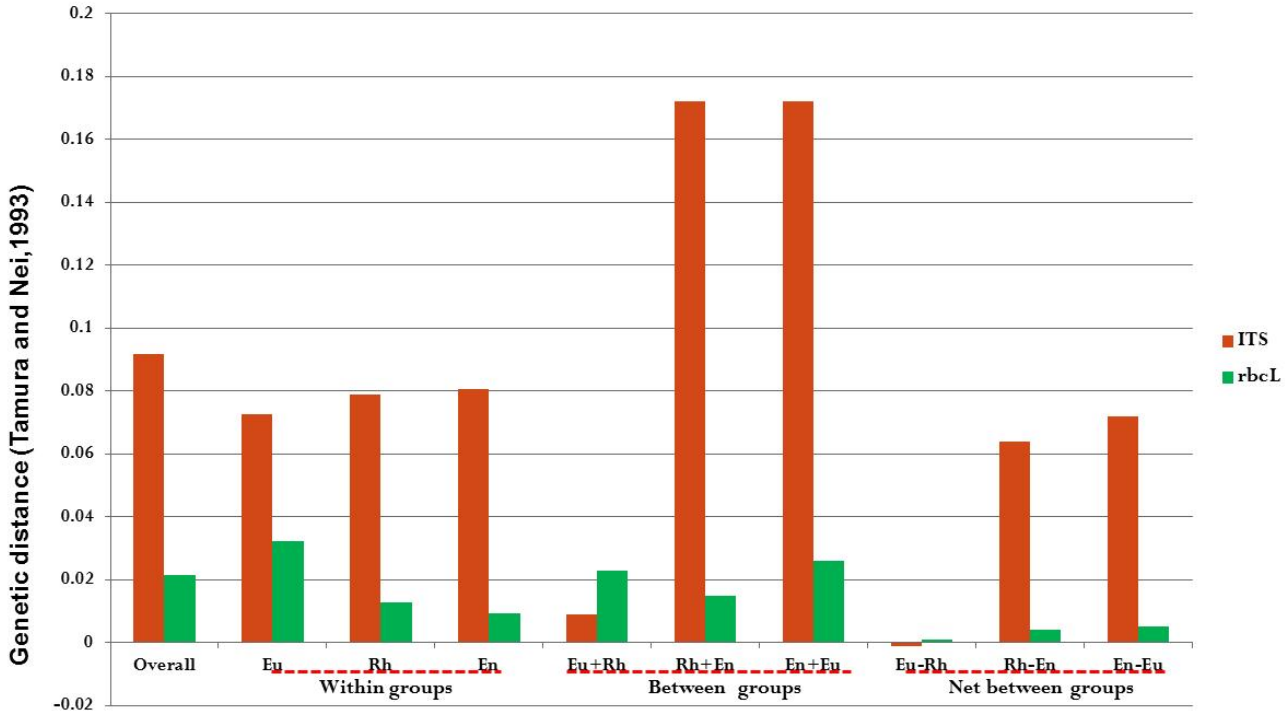


Figure 12: Genetic distance within and between groups of Musaceae as revealed by ITS2 and rbcL sequences in the samples studied.

CHAPTER 4

DISCUSSION

CHAPTER 4

DISCUSSION

In order to carry out a successful PCR-amplification of isolated DNAs, it is essential to optimize the PCR conditions specific for the particular species (Padmalatha and Prasad, 2006; Pandey *et al*., 2012). The presence of unwanted bands or smearing of the PCR products may be observed due to the use of inappropriate PCR conditions such as concentration of template DNA, primer, $MgCl_2$, Taq DNA polymerase and dNTPs. In the present study, several parameters such as concentration of template DNA, $MgCl_2$, dNTPs, primer, Taq DNA polymerase and also various annealing temperatures have been tested to standardize the efficiency of amplification. The optimization PCR runs were conducted using the ITS2 and the rbcL primers before using for the experiment. The ITS2 PCR conditions were optimized for a 25 µl reaction mixture containing 1× PCR buffer, 100 ng of genomic DNA as template, 0.5 µM each of the primers, 1.5 mM of $MgCl_2$, 0.2 mM of dNTP, 1 U of Taq DNA polymerase (Bangalore Genei, India). The PCR reaction was set up with the condition of 94°C for 4 min, 35 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 1 min followed by a final extension at 72°C for 7 min. Similarly using the rbcL primers the optimized PCR conditions were standardised with 25 µl reaction mixture containing 1× PCR buffer, 100 ng of genomic DNA as template, 0.5 µM each of the primers, 1.5 mM of $MgCl_2$, 0.2 mM of dNTP, 1 U of Taq DNA polymerase. The PCR reaction was set up with the condition of 95°C for 4 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min followed by a final extension at 72°C for 7 min.

The need for standardization of PCR conditions for specific analysis and their reasons have been widely reported. Excessive concentrations of genomic DNA and chemicals ($MgCl_2$, dNTP, Taq DNA polymerase and primers) in the PCR-amplification reduced the amplification efficiency thereby leading to smearing and non-specific yield of the products in the gel (Padmalatha and Prasad, 2006; McPherson and Møller, 2006). Williams *et al*., (1990) demonstrated that reduction of Taq DNA polymerase and genomic DNA concentration enhance the fidelity of PCR amplification efficiency when a non- discrete and smear of the products were obtained. If the Mg concentration is too low in the PCR-amplification, the yield is likely to be poor while the excess can reduce the fidelity of Taq DNA polymerase and lead to amplification of nonspecific products (McPherson and Møller, 2006).

A favourable DNA barcoding must contain short sequence length to facilitate amplification and sequence, enough genetic divergence for identification and conserved regions for developing universal primers (Kress *et al*., 2005; Sass *et al*., 2007).

Due to the high primer universality and discriminatory power, the rRNA gene sequences and rbcL are found easy to access due to highly conserved flanking regions allowing for the use of universal primers (Chenuil, 2006). They are routinely used for phylogenetic studies (Schneider *et al*., 2004). In this study, the ITS2 and the partial *rbcL* region have been evaluated for their possible barcoding in the 6 members of Musaceae family with reference to 2 sequences of the *Ensete* genera as a possible outgroup. ITS2 was proposed by Kress *et al*. (2005) as a potential barcode for plants. The rbcL gene sequences have been recommended by CBOL as a candidate region for the barcoding of plants (CBOL, 2009). The rbcL marker proved to be easily in

amplification and sequence with universal primers (Newmaster *et al*., 2006). A number of *rbc*L data are deposited in GenBank.

The analysis of the ITS2 and the rbcL sequences obtained from the PCR amplified fragments generated by the primer pair from the 6 banana samples using blast search revealed a very high identity percentage (97-99%) with the genbank database. For the proper evaluation of the relationship within the Musaceae family reference sequences of 2 species of the genus *Ensete* were retrieved from the genbank database and used as outgroup. The *Ensete spp.* are far distinct from the species of *Eumusa* and *Rhodochlamys* section (Bekele and Shigeta, 2011). The overall analysis of the ITS2 and rbcL sequences revealed the insufficient variable, low interspecific variation and entirely no gap. Therefore the uses of these candidate genes (ITS2 and rbcL) were not successful in generating a substantial barcode for the Musaceae family. It is therefore suggested from this finding that the use of more markers and species for possible barcoding of the family.

CHAPTER 5

CONCLUSION

CHAPTER 5

CONCLUSION

The occurrence of various cultivars belonging to different genome groups in their wild and in cultivated sites further strengthened the observation that Mizoram as a part of the northeast Indian region constitutes the region where the natural hybridization of the banana and their wild relatives occurs.

This work attempted 2 DNA barcodes (ITS2 of the nrDNA ITS region and rbcL of the chloroplast coding region) to evaluate phylogenetic relationship of the Musaceae family. Although rbcL have been recommended by CBOL as a candidate region in their last study (CBOL, 2009) it had a poor performance with insufficient variable, low interspecific variation and no gap. Thus, we suggested that this locus is not a god candidate for the Musaceae family. The ITS2 revealed relatively more variable nucleotide information than the rbcL DNA barcodes, although it cannot be considered as a DNA barcode for Musaceae family yet we came to conclusion that ITS2 is a much better locus than rbcL. The evolution of the members of the Musaceae family studied has been more clearly in combination with multi-locus DNA barcodes than that the single-locus. Thus, this work not only provided more sequence sources of the Musaceae species, but helped further understand the phylogenetic relationship of this taxonomically complex family.

CHAPTER 6

SUMMARY

CHAPTER 6

SUMMARY

Bananas and plantains belongs to the family Musaceae and consists of three genera namely, Musa, Ensete and Musella . The genus Musa is divided into four sections based on the phenotypic traits and basic chromosome numbers, such as Eumusa (x = 11), Rhodochlamys (x = 11), Australimusa (x = 10) and Callimusa (x = 10). Among the five *Musa* sections, Eumusa contributed the largest number, most diversified and ancient including the two wild diploid progenitor species of *M. acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome). Thus the inter and intra specific crosses between these two species leads to the existence of different varieties with various genome composition resulting several ploidy level ranging from diploids to tetraploids (AA, AB, AAB, ABB, ABBB, AABB etc.) and posing a huge challenge towards its proper identification. Numerous reports on the use of molecular markers including the nuclear and chloroplast sequences in banana have been reported the information on the barcodes are scanty.

DNA barcodes are unique sequence patterns of small DNA fragments that are used as specific reference collections to identify specimens and to discover overlooked species. DNA barcoding provides rapid, accurate and automatable species identification using standardized DNA sequences. The most important characteristic features of a DNA barcode are its universality, specificity on variation and easiness on employment. Thus the present investigation was conducted to evaluate the ITS2 sequences of the nuclear ribosomal DNA and the chloroplast rbcL gene sequences for possible DNA barcoding of the members of the Musaceae family.

The findings of the present study are presented briefly as follows:

- Successful PCR amplification of the ITS2 and rbcL regions in the 6 banana samples studied was achieved using the universal primers.
- The genetic sequences derived from the PCR products revealed the variable length and topology.
- Homology search using Blast resulted in the high similarity values of 97-99% in ITS2 and 95-99% in rbcL between the sequences generated and the Genbank databases.
- Using 2 reference sequences of the genus Ensete, the sequences obtained from the banana samples were analysed.
- The overall analysis of the sequences revealed the phylogenetic relationship based on the sections and the classification of Musaceae.
- The ITS2 sequences revealed the superiority in in terms of detecting the inter- and intra-species variations over the rbcL sequences.

However, the DNA barcoding of the members of the Musaceae family could not be achieved and further works on the use of additional markers and members of the family are required. The finding therefore provides glimpses of the possibility of ITS2 as a possible DNA barcode at the interspecies level.

BIBLIOGRAPHY

References

Azhar M (2006). Genome organisation, evolution and biodiversity in Musa: Application to stress-related gene discovery and plant breeding. Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990).Basic local alignment search tool. J MolBiol, 215 403–410.

Baker RED and Simmonds NW (1953).The genus *Ensete* in Africa. Kew Bull, 3: 405–416.

Bekele E and Shigeta M (2011). Phylogenetic relationship between *Ensete* and *Musa* species as revealed by the *trnT trnF* region of cpDNA. Genet Resour Crop Evol, 58: 259-269.

Carreel F (1994). Etude de la diversité des bananiers (genre *Musa*) à l'aide des marqueurs RFLP. Thèse de l'Institut National Agronomique Paris- Grignon, Pp 90.

CBOL Plant Working Group (2009). A DNA barcode for land plants. PNAS, USA 106:12794-12797.

Cheesman EE (1947). Classification of the bananas. I. The genus *Ensete* Horan. Kew Bull. 2: 97–106.

Chen S *et al.*, Validation of the ITS2 region as a model DNA barcode for identifyting medicinal plants species. PLoS ONE., 2010. 5 .e8613

Chenuil A (2006). Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects. Genetica 127:101-201.

D'Hont A, Denoeud F, Aury JM, Baurens FC and Carreel F (2012).The banana (*Musa acuminata*) genome and evolution of monocotyledonous polants.

Daniells J, Jenny C, Karamura D and Tomekpe K (2001).Musalogue: a catalogue of *Musa* germplasm. Diversity in the genus Musa (E. Arnaud and S. Sharrock, compil.).International Network for the Improvement of Banana and Plantain, Montpellier, France.

FAO (2005) http://www.faostat.fao.org/N

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl.Acids.Symp. Ser. 41:95-98.

Hebert, P.D.N., Cywinska, A., Ball,S. L. and deWaard,J.R.,Biological identifications through DNA barcodes. Proc.R.Soc.London, Ser,B,2003,270,313-321.

Hebert,P.D.N and Gregory,T.R., The promise of DNA barcoding for taxonomy. Syst.Biol.,2005,54,852-859

Heslop-Harrison JS and Schwarzacher T (2007).Domestication, Genomics and the Future for Banana. Ann Bot, 100: 1073–1084.

Horaninow PF (1862). Prodromus Monographiae Scitaminarum. Academiae Caesareae Scientiarum, Petropoli (in Latin).

Horry JP and Arnaud E (1997). Musalogue: a catalogue of *Musa*germplasm. Papua New Guinea collecting missions, 1988-1989.International Network for the Improvement of Banana and Plantain, Montpellier, France.

Hřibová E, Čížková J, Christelová P, Taudien S, de Langhe E and Doležel J (2011). The ITS1-5.8S-ITS2 Sequence Region in the Musaceae: Structure, Diversity and Use in Molecular Phylogeny. PLoS ONE 6(3): e17863. doi:10.1371/journal.pone.0017863.

Jussieu AL (1789) Ordo I. Musae. Genera plantarum: secundum ordines naturals disposita, Juxta methodum in Horto regio parisiensi exaratam, anno M. DCC. LXXIV, Paris, pp 61–62

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005). Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U.S.A., 102: 8369-8374.

Li LF, Häkkinen M, Yuan YM, Hao G and Ge XJ (2010). Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. Mol Phylogenet Evol, 57: 1–10.

Linnaeus C (1753). *Musa* L. Species Plantarum, vol. 2.ImpensisLaurentii Salvii, Stockholm, Pp 1043.

McPherson MJ and Møller SG (2006). PCR second edition. Taylor and Francis group, Pp 67.

Molina AB and Kudagamage C (2002). The international network for the improvement of banana and plantain (INIBAP): PGR activities in south asia. *In:* South Asia Network on Plant Genetic Resources (SANPGR) meeting held on December 9-11 at Plant Genetic Resources Center (PGRC), Peradeniya, Sri Lanka, Pp 1-7.

Newmaster SG, Fazekas AJ, Ragupathy S (2006). DNA barcoding in land plant: evalution of rbcL in a multigene tiered approach. Can. J. Bot., 84: 335-341.

Padmalatha K and Prasad MNV (2005). Optimization of DNA isolation and PCR protocol for RAPD analysis of selected medicinal and aromatic plants of conservation concern from Peninsular India. African J Biotech, 5(3): 230-234.

Padmalatha K and Prasad MNV (2005). Optimization of DNA isolation and PCR protocol for RAPD analysis of selected medicinal and aromatic plants of conservation concern from Peninsular India. Afr J Biotechnol, 5 (3): 230-234.

Pandey RJ, Singh DP, Sudhakar G, Thangaraj K and Rao VR (2012). Standardization of PCR conditions for an Ancient DNA Amplification. International Journal of Human Sciences [Online]. 9:1.

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C, Lebot V, Risterucci AM, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P and Denham T (2011). Multidisciplinary perspectives on banana (*Musa* spp.) domestication. Proc Natl Acad Sci USA, 108:11311-11318.

Pollefeys P, Sharrock S and Arnaud DE (2004).Preliminary analysis of the literature on the distribution of wild *Musa* species using MGIS and DIVA-GIS. Montpellier, France:

Purseglove JW (1972).Tropical Crops.Monocotyledons.Vol. 2 Longmans, London.

Samson JA (1992). Tropical Fruits.2nd Edition.Tropical Agriculture series. Longmans, London.

Sass C, Little DP, Stevenson DW, Specht CD (2007). DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads. PLOS ONE, 11: e1154.

Sathiamoorthy S, Uma S and Selvarajan R (2001). Banana research and development programme in India and highlights of NRCB-INIBAP collaborative projects. Advancing banana and plantain R & D in Asia and the Pacific, 10: 67-76.

Schneider H, Russell SJ, Cox CJ, Bakker F, Henderson S, Rumsey F, Barrett J, Gibby M, Vogel JC (2004). Chloroplast phylogeny of asplenioid ferns based on rbcL and trnL-F spacer sequences (Polypodiidae, Aspleniaceae) and its implications for biogeography. Syst. Bot. 29:260-274.

Simmonds NW (1962). Evolution of bananas: Longman, London UK.

Simmonds NW and Shepherd K (1955). The taxonomy and origin of the cultivated banana. J Linn Soc Bot, 55: 302-312.

Tajima F (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism.*Genetics*123:585-595.

Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.*Molecular Biology and Evolution*10:512-526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. MolBiolEvol, 28: 2731-2739.

Thangjam R, Damayanti M and Jitendra GS (2003). A simple and rapid method for isolation of DNA from imbibed embryos of Parkia timoriana (DC.) Merr for PCR analysis. Food Agric Environ, 1(3&4): 36-38.

Ude G, Pillay M, Nwakanma D and Tenkouano A (2002).Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers.Theo Appl Genet, 104: 1239-1245.

Uma S, Siva SA, Saraswathi MS, Manickavasagam M, Durai P, Selvarajan R and Sathiamoorthy S (2006). Variation and intraspecific relationships in Indian wild *Musa balbisiana*(BB) population as evidenced by random amplified polymorphic DNA. Genet Resour Crop Evol, 53:349–355.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res, 18 (22): 6531-6535.

Wong C, Kiew R, Argent G, Set O, Lee SK and GanYY (2002). Assessment of validity in the sections in *Musa* (Musaceae) using AFLP. Ann Bot, 90:231-238.