DNA Barcoding of Goan Mangroves and Functional Characterization of Sodium/Proton Antiporter in *Rhizophora apiculata*

THESIS

Submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

ANKUSH ASHOK SADDHE

Under the Supervision of

Dr. Kundan Kumar

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

(RAJASTHAN) INDIA

2018

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This thesis is submitted under Regulation 8.20 (a) of the Academic Regulations for Doctoral Programmes which allows a faculty member of the Institute/Professional to do Ph.D. research without the benefit of a supervisor.

This is to certify that the thesis entitled **DNA barcoding of Goan mangroves and functional characterization of sodium/proton antiporter in *Rhizophora apiculata*** and submitted by **Mr. Ankush Ashok Saddhe** ID No **2013PHXF0403G** for award of Ph.D. Degree of the Institute embodies my original work.

Signature: 

Name in Block Letters: ANKUSH ASHOK SADDHE

ID No.: 2013PHXF0403G

Designation: Ph.D. Scholar

Date: August 2018

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled **DNA Barcoding of Goan Mangroves and Functional Characterization of Sodium/Proton Antiporter in *Rhizophora apiculata*** and submitted by **Mr. Ankush Ashok Saddhe** ID No **2013PHXF0403G** for award of Ph.D. of the Institute embodies original work done by him/her under my supervision.

Signature of the Supervisor 

Name in capital letters: Dr. KUNDAN KUMAR

Designation: Associate Professor

Date: August 2018

ACKNOWLEDGMENT

"An educational journey is an endless process", where you can meet thousands of brain, who teach you and carve your life. This is a great platform to appreciate their contribution to our life directly or indirectly. "The reason behind whomever, I am present and "whatever I will be in future mostly influenced by nature, books and countless teachers". I am obliged to everyone who shaped my life and career. I do believe the greatest work always accomplished through highly motivated, directional, and inspired minds who supervised by a teacher. During my research journey, I learned how to execute an idea successfully in a laboratory until scientific application. I would like to express my gratitude to Dr. Kundan Kumar, BITS Pilani, K. K. Birla Goa Campus India, who inspired, motivated and carved me as an experienced researcher.

I am extremely grateful to Prof. Souvik Bhattacharyya (Vice Chancellor, BITS, Pilani), Prof. Raghurama G (Director, BITS Pilani K. K. Birla Goa Campus), Prof. S. K. Verma (Dean, Academic Research, BITS, Pilani), Prof. Sunil Bhand (Dean, Sponsored Research and Consulting, BITS Pilani), Prof. Bharat M. Deshpande (Associate Dean, Academic Research, BITS Pilani K. K. Birla Goa Campus), and Prof. Meenal Kowshik (Associate Dean, Sponsored Research and Consulting, BITS Pilani K. K. Birla Goa Campus), for providing me with the facilities to conduct my research work at BITS, Pilani K. K. Birla Goa Campus.

I would like to express my gratitude to the members of the Doctoral Advisory Committee, Prof. Utpal Roy, and Prof. Srikanth Mutnuri, Department of Biological Sciences for their guidance and co-operation. I express my sincere thanks Prof. Srikanth Mutnuri Head, Department of Biological Sciences, and Prof. Dibakar Chakrabarty, Convener, Departmental Research Committee, for their support, all faculty members of the Biological sciences departments Dr. Judith Braganca, Prof. Vijayashree Nayak, Dr. Anusuya Ganguly, Dr. Sumit Biswas, Dr. Angshuman Sarkar, Dr. Veeky Baths, Dr. Malabika Biswas, Dr. Sukanta Mondol, Dr. Indrani Talukdar, Dr. Raviprasad Aduri, Dr. Rajesh Mehrotra, Dr. Sandhya Mehrotra, and Dr. Arnab Banerjee for their timely help throughout this work.

I would like to express my gratitude towards Forest Department, Goa, India to permit me for mangrove sample collection. Also, I would like acknowledge to Dr. Edward Blumwald, University of California, Davis, USA and Dr. Kees Venema, Spanish National

Research Council, Madrid Spain for their generous gift of Yeast mutant strains. Moreover, I would like to thanks Dr. Alok K. Sinha for his support in performing confocal imaging.

I would emphatically acknowledge to the University Grant Commission, Government of India and BITS Pilani for providing financial assistance in the form research fellowship. Also, I would like to acknowledge DST-SERB for International travel grant to attend iBOL, 2017 at Kruger National Park, South Africa.

It is my pleasure to thanks Dr. Rahul Jamdade for his assistance and generous help during DNA barcoding data analysis. Moreover, I would like to thanks KK lab members, Rakesh Manuka, Shruti Apte, Nilesh Dahibhate, Divya Gupta, Manali Malvankar, and Suhas Karle for their help and great time we shared. This work would not have been possible without the support of all the research scholars of the department. I am forever thankful to all research scholars of Biological Sciences for their timely help. Also, I would like to thanks Ms. Kamna, Ms. Pallavi, Mr. Mahadev and Mr. Mahaling for their departmental help.

Finally, I would not have been able to accomplish this thesis without the understanding, trust, and unconditional support of my parents Mr. Ashok Bapurao Sadhye and Ms. Mangal Ashok Sadhye, giving me the opportunity to pursue my dreams. I would like to thank my younger brother Mr. Yogesh Sadhye, sister-in-law Ms. Pratibha Sadhye, brother-in-law Mr. Santosh Kadoo, nephew Sarthak and niece Swara. I am forever indebted to my life partner, Dr. Shruti Gupta Sadhye. I would like to sincerely thanks to my in-laws Mr. Vivekanand Gupta and Ms. Anant Gupta, Mr. Lokesh Kumar Gupta, Ms. Roopali Gupta, Mr. Ashwani Kumar Gupta, and Ms. Vandana Gupta.

I thank to the Almighty god for all the wonderful things and surrounding life he is giving to me every moment and hope his blessing will be continue on my journey.

Ankush Ashok Saddhe

Dedicated to my beloved sister

Abstract

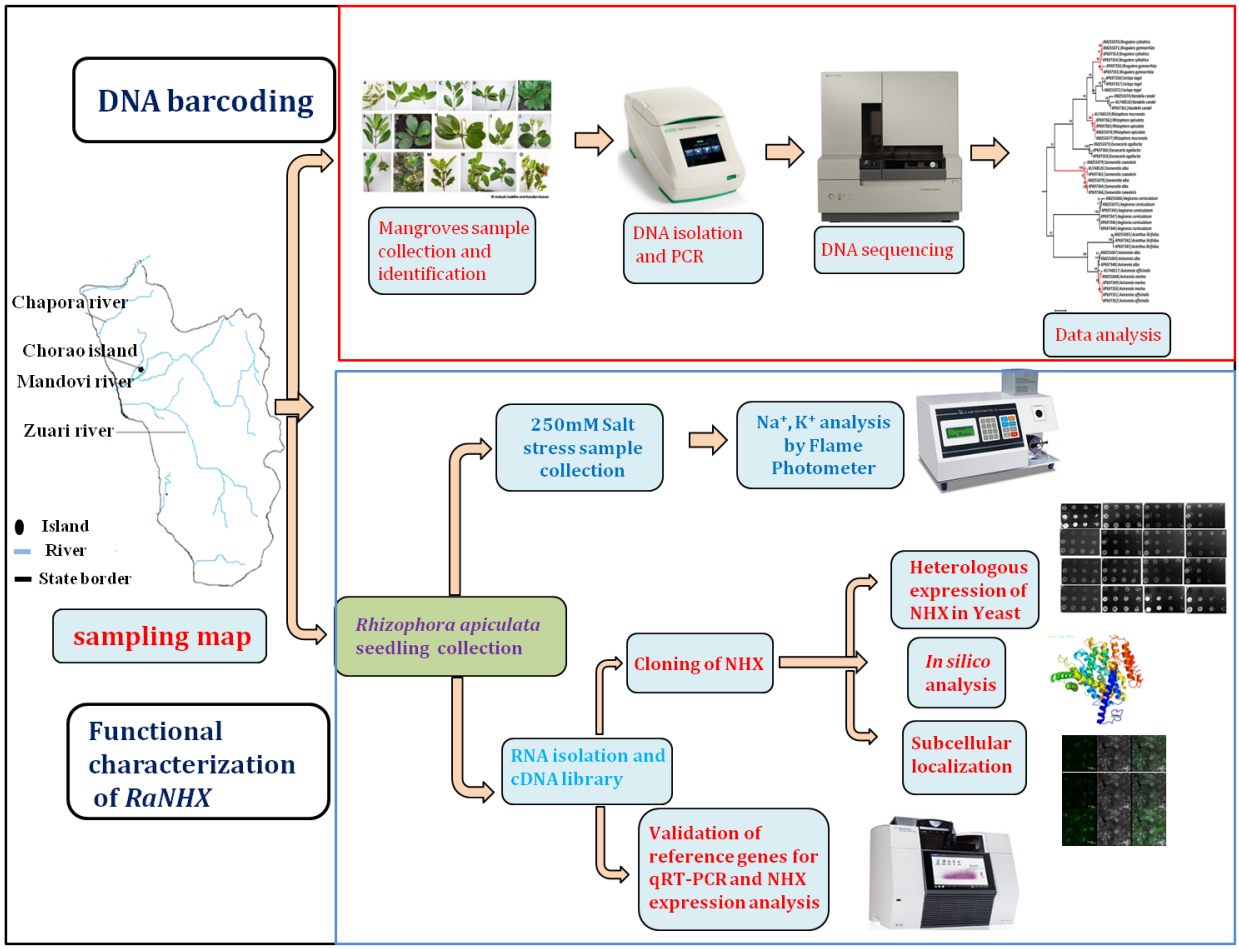
The aim of this work was DNA barcoding of mangroves plant species from West Coast of Goa and the functional characterization of sodium/proton antiporter (NHX) in *Rhizophora apiculata* for salt stress tolerance. Mangroves are diverse estuarine ecosystem prevalent in the tropical and subtropical zone, but anthropogenic activity turned them into the vulnerable ecosystem. There is a need to build a molecular reference library of mangrove plant species based on molecular barcode marker along with morphological characteristics. In this study, we tested the core plant barcode (*rbcL* and *matK*) and four promising complementary barcodes (ITS2, *psbK-psbI*, *rpoC1* and *atpF-atpH*) in 14 mangroves species belonging to 5 families from West Coast India. With a single locus analysis, ITS2 exhibited the higher discriminatory power (87.82%) and combinations of *matK* + ITS2 provided the highest discrimination success (89.74%) rate except for *Avicennia* genera. The *matK*+ITS2 marker based on GMYC method resolved 57.14% of mangroves species and TaxonDNA, ABGD, and PTP discriminated 42.85% of mangrove species. Further, we explored 3 additional markers (*psbK-psbI*, *rpoC1*, and *atpF atpH*) for *Avicennia* genera (*A. alba*, *A. officinalis* and *A. marina*) and *atpF-atpH* locus was able to discriminate three species of *Avicennia* genera. Our analysis underscored the efficacy of *matK* + ITS2 markers along with *atpF-atpH* as the best combination for mangrove identification.

Rhizophora apiculata is a halophytic, small mangroves tree distributed along the coastal region of the tropical and subtropical area of the world. However, there are no reports available on the selection of candidate reference genes (RG) for quantitative real-time polymerase chain reaction (qRT-PCR) in *R. apiculata* different tissues and in stress conditions. We demonstrated that, *EF1α* followed by *ACT* and *β-TUB* was found to be the most stable reference genes in *R.*

apiculata tissues under normal condition. In salt stress, *EF1α* was comprehensively recommended top-ranked reference gene followed by *ACT* and 18S.

Moreover, we cloned the full-length coding sequence of a putative vacuolar Na^+, K^+/H^+ antiporter from *R. apiculata* (*RaNHX1*). *In-silico* analysis predicted highly conserved signature domain, motif such as cation/proton exchanger 1 (CPA1), Na^+/H^+ antiporter domain, amiloride-binding "ND" motif, and 12 transmembrane helices, which confirmed they are the member of plant NHX family. RaNHX1 translated protein shared 86% sequence identity with *Arabidopsis* AtNHX1 and AtNHX2. Moreover, a phylogenetic relationship revealed RaNHX1 clustered with the *Arabidopsis* vacuolar Na^+/H^+ antiporter group. Transcript analysis of *RaNHX1* by qRT-PCR revealed their differential transcript regulation pattern in various tissues and under salt stress libraries. The relative expression of *RaNHX1* showed higher transcript abundance in shoot tissue followed by primary root, anchor root and young leaf tissues. Elemental analysis (Na^+ and K^+), confirmed that leaves were major Na^+ storage organ in *R. apiculata* under salt stress. The heterologous expression of pYES2.0-RaNHX1 in AXT3 strain showed partial complementation of sensitive phenotype under NaCl and KCl stress indicating their involvement in sequestering Na^+ and K^+ in vacuole. The sensitivity of yeast AXT3 mutant and AXT3-RaNHX1 were tested against hygromycin-B antibiotic. Moreover, elemental analysis (Na and K) of AXT3 RaNHX1 cells performed well compared to AXT3 mutant under stress conditions. The subcellular localization study showed the RaNHX1 localized in the stomatal guard and subsidiary cells.

Thesis graphical abstract



The **s**chematic presentation of graphical abstract was covered DNA barcoding of Goan mangroves using the plastid and nuclear DNA barcodes. Selection and validation of reference genes for qRT-PCR was performed. Molecular cloning, gene expression analysis and heterologous characterization of sodium/ proton antiporters from *Rhizophora apiculata* were performed.

TABLE OF CONTENTS

Page No

List of Figures

List of Tables

List of Symbols and Abbreviations

Symbols	Description
mg	Mili gram
mL	Mili liter
mM	Mili molar
$mg.ml^{-1}$	Milligram per milliliter
µg	Micro gram
µl	Micro liter
µmol	Micro moles
$\mu mol.mg^{-1}$	Micromoles per mili gram
ng	Nano gram
nm	Nano meter
pH	Acidic/Basic measurement unit
PPM	Part per million
s	Second
m	Minute
h	Hour
M	Molarity
3D	Three dimensional
°C	Degree Celsius
%	Percentage
>	Greater than
<	Less than

Abbreviations

RbcL	Ribulose 1, 5-bisphosphate carboxylase/oxygenase large subunit
MatK	Maturase K
rpoB and *rpoC1*	Chloroplast RNA polymerase subunit
trnH-psbA	Intergenic spacer
atpF- atpH	ATP synthase subunits CFO I and CFO III respectively

psbK -psbI	Polypeptides K and I
ycf1	Tic214 complex
ITS	Internal transcribed spacer
LSC	Large single copy
SSC	Small single-copy
IR	Inverted repeat (IRA, IRB)
ACT	Actin
B-TUB	β-tubulin
UBQ	Ubiquitin
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
EF1α	Elongation factor 1α
18S	18S ribosomal RNA
qRT-PCR	Quantitative real-time PCR
RT-PCR	Reverse transcription polymerase chain reaction
Ra	*Rhizophora apiculata*
Ct	Cycle thresholds
HKT	High affinity potassium transporter
SOS	Salt Overly Sensitive
NHX	Sodium/hydrogen exchanger
CPA	Cation/proton exchanger
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
CBOL	The Consortium for the Barcode of Life
K2P	Kimura2-parameter
MCMC	Monte Carlo Markov Chains
CTAB	Cetyl-trimethyl ammonium bromide
SDS	Sodium dodecyl sulphate
EDTA	Ethylene diamine tetra-acetic Acid
TE buffer	Tris EDTA buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
PVP	Polyvinyl pyrrolidone

dNTP	Deoxyribonucleotides tri phosphate
$(dT)_{18}$	Deoxythymine 18
DEPC	Diethyl pyrocarbonate
MEGA	Molecular Evolutionary Genetics Analysis
DNA	Deoxyribonucleic acid
cDNA	Complementary deoxyribonucleic acid
gDNA	Genomic Deoxyribonucleic acid
RNA	Ribonucleic acid
OD_{600nm}	Optical density at 600 nm
BOLD	Barcode of Life Database system
NJ	Neighbor-joining
ABGD	Automated Barcode Gap Discovery
JC	Jukes-Cantor
GMYC	General Mixed Yule-Coalescent
HKY	Hasegawa-Kishino-Yano
PTP	Poisson Tree Process model
ANOVA	Analysis of Variance
NTC	Negative control
SD	Standard deviation
SE	Standard error
MIQE	Minimum Information for publication of qRT-PCR experiments
SD media	Synthetic define media
GFP	Green fluorescent protein
CaMV	Cauliflower mosaic virus
rpm	Revolution per minutes
OTU	Operational taxonomic unit

CHAPTER 1

Introduction and Review of Literature

CHAPTER 1

Introduction and Review of Literature

1.1 Mangroves biodiversity

Indian biodiversity is one of the richest in the world and categorized into four major hotspots such as the Western Ghats, the Eastern Himalaya, the Indo-Burma, and the Sundaland (Myers et al., 2000; Chitale et al., 2014). Indian plant diversity is ranked 10^{th} position in the world and 4^{th} in the Asian continent. It sheltered approximately 45,500 plant species, which is almost equal to 11% of the world's floral diversity (Singh and Chaturvedi, 2017). The Indian forest ecosystems are diverse and categorized into six major forest systems such as tropical dry forest, tropical wet forest, montane subtropical forest, montane temperate forest, sub-alpine forest, and alpine scrub (Singh and Chaturvedi, 2017). Mangroves are unique coastal ecosystem distributed in the tropical and subtropical parts of the world (Kathiresan, 2018). Mangroves are highly productive wetland having unique flora and fauna acclimatized to the local environment such as fluctuated water level, high temperature, salinity, and low aeration condition (Tomlinson, 1986; Hutchings and Saenger, 1987; Ragavan et al., 2016). The term 'mangroves' are referred to either individual plant or intertidal ecosystem or both, as 'Mangrove plants' and 'Mangrove ecosystem' (MacNae, 1968). However, in this context we used mangrove term as a mangrove plants. They have unique features such as aerial breathing roots, extensive supporting roots, buttresses, salt-excreting leaves and viviparous propagules (Duke, 1992; Shi et al., 2005) as shown in Figure 1.1.

World atlas of mangroves reported presence of total 73 mangrove species with few identified mangrove hybrids species located in 123 countries covering 1,50,000 km^2 areas (Spalding et al., 2010). Indian mangroves vegetation is third largest in the world followed by

Indonesia and Australia (Kathiresan, 2018). A recent report on Indian mangrove vegetation distribution occupied 4740 km^2 area (Kathiresan, 2018). The Indian mangroves categorized into the three zones such as the East Coast covering 2754 km^2, the West Coast covering 1372 km^2, and Andaman & Nicobar Islands share 617 km^2 as shown in Figure 1.2. The East Coast zone ranges from Sundarban forest of West Bengal to Cauvery estuary of Tamil Nadu and comprises 70% mangrove and characterized by cliffs, promontories, lakes, lagoons and back waters (Jagtap et al., 1993; Sanyal et al., 1998; Untawale and Jagtap, 1992; Kathiresan, 2018).

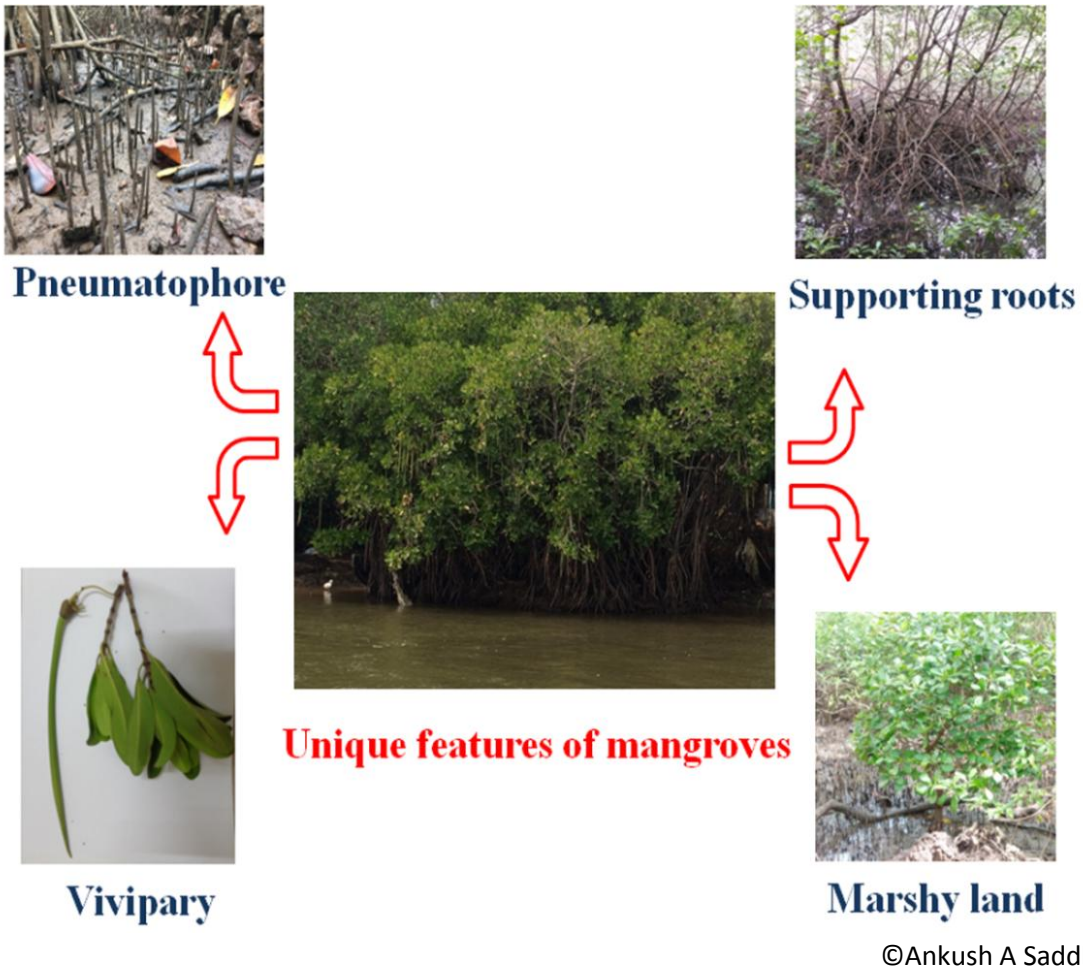


Figure 1.1 Unique features of the mangrove which allows them to grow in the marshy land such as supporting roots, stilt roots, pneumatophores for gaseous exchange, and viviparous mode of reproduction. All the pictures were captured from the Mandovi River, Chorao island Goa, India.

Sundarban forest is one of the largest single blocks of mangrove vegetation in the world located at the East Coast of India. West Coast region stretches from Bhavnagar estuary of Gujarat to Cochin estuary of Kerala and constitute 28% mangrove (Kathiresan, 2018). The West Coast of India is more or less steeply shelved, lack major deltas, river estuaries and dominated by sandy and rocky substratum. The West coast also harbors one of the world's biodiversity hotspot of the Western Ghats in India. It includes the states such as Gujarat, Maharashtra, Goa, Karnataka and Kerala, which harbors 27 mangroves species (16 genera under 11 families) (Ragavan et al., 2016; Kathiresan, 2018).

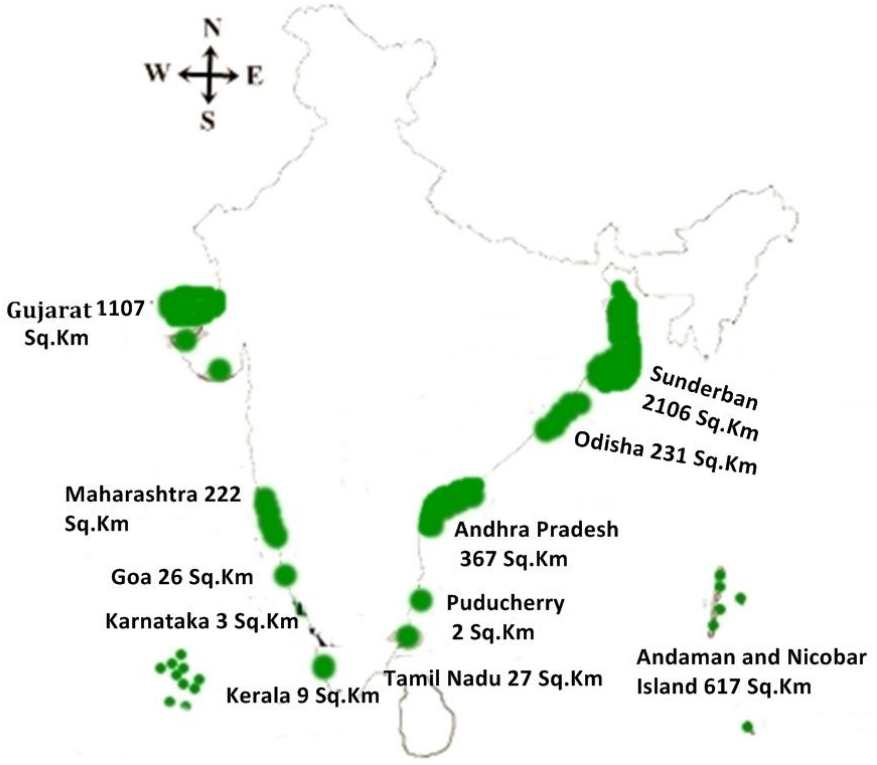


Figure 1.2 Distribution of Indian mangrove vegetation in the different states of India depicted on the map along with covered area. (Area coverage in the square kilometer was taken from Kathiresan, 2018). East Coast mangroves area: Sunderban, Odisha, Andhra Pradesh, Puducherry and Tamil Nadu mangroves. West Coast covered Gujarat, Maharashtra, Goa, Karnataka and Kerala.

The prevalent mangroves were reported along the West Coast of India are *Xylocarpus granatum, X. moluccensis, Avicennia officinalis, A. marina, Excoecaria agallocha, Rhizophora mucronata, R. apiculata, Sonneratia alba, S. caseolaris, Bruguiera gymnorrhiza, B. parviflora, Cariops tagal, Heretiera littoralis,* and *Lumnitzera racemosa* (Kathiresan and Bingham, 2001). Indian mangrove flora constitutes 46 species belonging to 22 genera and 14 families, which include 42 species and 4 natural hybrids (Ragavan et al., 2016). There are about 15, 22, 16, 10, and 19 species of mangroves reported along the coast of Gujarat, Maharashtra, Goa, Karnataka, and Kerala respectively in Western Coast India (Ragavan et al., 2016). Goa state is located in western coast of India and mangrove vegetation in Goa occupies 26 km^2 of area (Ragavan et al., 2016; Kathiresan, 2018). The two river channels of Mandovi and Zuari are linked by Cumbarjua canal, forming an estuarine complex which supports mangrove vegetation (Figure 1.3). D'Souza and Rodrigues (2013) reported the presence of 17 mangrove species in Goa that include 14 true and 3 associated mangrove species (D'Souza and Rodrigues, 2013).

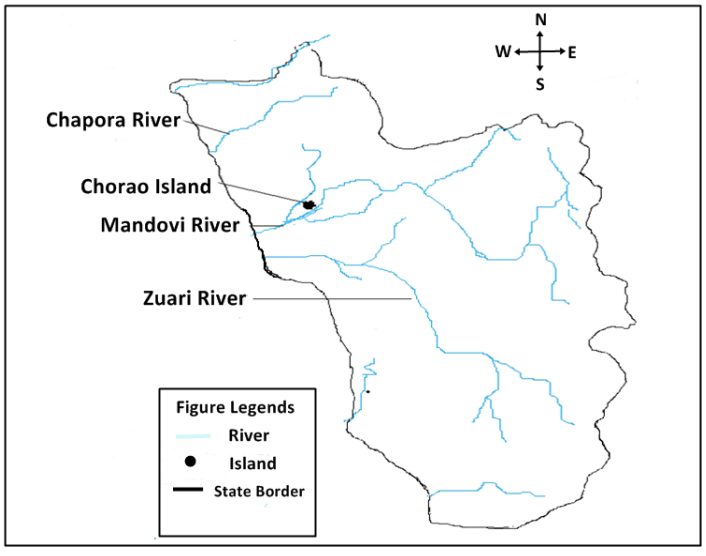


Figure 1.3 Major river systems and mangroves distribution in the Goa state depicted on the map. The Mandovi and the Zuari rivers are the largest among Goa where major mangrove along the coast proliferate.

Plant species identification is a preliminary step to conserve and utilization of biodiversity, which apparently impeded by reduce of plant taxonomic skills. Mangrove diversity was reported based on the morphology, reproductive physiology, vegetation types and ecology. The phenotypic characters altered substantially on the local climatic condition of the habitat. Unlike phenotypic characters, molecular markers are stable and don't prone to environmental influences and they are helpful to demarcate species and hybrids in evaluation of intra and interspecific variation and distribution (Chalmers et al., 1992; Powell, 1992; Waugh and Powell, 1992). Hypervariable regions of DNA are most preferably studied markers such as Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Restriction Fragment Length Polymorphism (RFLP), and Inter-Simple Sequence Repeat **(**ISSR) in evaluation of genetic diversity of mangroves (Parani et al., 1997a, 1997b; Schwarzbach and Ricklefs, 2000; Zhou et al., 2005; Mukherjee et al., 2006; Lo, 2010; Sun and Lo, 2011; Das et al., 2014). Parental issue of *Rhizophora* hybrid was unraveled using molecular marker that exhibited 96.5% genetic similarity with parental species *R. stylosa* and *R. apiculata* (Parani et al., 1997a; Parani et al., 1997b; Lakshmi et al., 1997; Parani et al., 1998; Lakshmi et al., 2000). However, the use of molecular markers is continuously marred by problems such as poor reproducibility and homoplasy (Agarwal et al., 2008; Caballero et al., 2008; Mondini et al., 2009). Alternative method for rapid global biodiversity assessment and species identification has been recommended by The Consortium for the Barcode of Life (CBOL). It uses nuclear as well as organelle DNA markers such as *rbcL*, *matK*, trnH-psbA, an ITS2 (CBOL, 2009; Hollingsworth et al., 2011). The nuclear DNA barcode, internal transcribed spacer (ITS) had been reported as a potential barcoding region for cryptic and complex plant species (Chen et al., 2010; China Plant BOL group, 2011). The plastid markers are emerging as informative tools to

study mangrove molecular systematics and evolution, because of its reasonable size, high substitution rate, evenly distributed codon position variation, low transition and transversion ratio. The *rbcL* and *matK* genes has been contributed to understand the angiosperms genesis, association of dicotyledons, monocotyledons and phylogenetic reconstruction at class, order, family and genus levels (Chase, 1993).

Available information on Indian mangrove is highly scattered and no worthwhile report available to address issues like taxonomy, diversity, systematic and evolution. On the other hand, delineating mangrove species from putative hybrids using morphological characters are always questionable. Putative hybrids were reported within the major genera of *Rhizophora*, *Sonneratia*, *Lumnitzera* and *Bruguiera* (Tomlinson, 1986; Duke and Ge, 2011). In the present study, we have elucidated molecular identification of mangrove species using DNA barcode markers such as plastid coding loci *rbcL*, *matK*, *psbK-I*, *atpF-H* and nuclear ITS2. Goa mangroves are rich in diversity and accounted 14 species belonging to four order and six families. This is our first step towards understanding molecular relationship among mangroves based on plastid barcode genes. Our study might be helpful in identification of mangroves and also for the development of various conservation strategies.

1.2 Plant DNA barcode

DNA barcoding is molecular tools that boost up instant and precisely identification of animal and plant species (Hebert et al., 2003a, 2003b). This technique first time successfully implemented in an animal taxa based on the mitochondrial cytochrome oxidase I (COI) locus as a potential DNA barcode to correctly assign respective animal taxa (Hebert et al., 2003a; Hebert et al., 2003b). In contrast, the plant mitochondrial loci have low substitutions and evolutionary

rates which make them unsuitable for plant barcoding (Chase et al., 2005; Kress et al., 2005; Newmaster et al., 2006). Alternatively, plant DNA barcoding preferred nuclear as well as organellar DNA markers to assign respective plant taxa. Moreover, plastid genome has some unique features such as the maternal inheritance, non-recombination, and genetic stability. Many plastid genes have been assessed for their efficacy as plants barcodes (Chase et al., 2005; Kress et al., 2005; Newmaster et al., 2006). Some inherent problems in plant taxa such as cryptic and closely related taxa, genotypic and phenotypic variability, and natural hybridization which hide the success rate of DNA barcoding in some plant taxa (Vijayan and Tsou, 2010). To overcome this issue, multiple and enormous DNA markers with different combinations were evaluated ranging from plastid coding (*rbcL, matK*) to non-coding regions (*trnH-psbA*), nuclear spacer (ITS) (Vijayan and Tsou, 2010). We have described the commonly used DNA barcodes in the following paragraphs and in the Table 1.1.

1.2.1 The *RbcL* locus

RbcL gene encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) protein which is of 1350 bp long and involved in the CO_2 fixation process. The *rbcL* was the first gene that was sequenced from the plants and choice for inference of phylogenetic relationships at higher taxonomic levels (Chase, 1993). *RbcL* gene strictly follows the DNA barcoding gene selection criteria such as present ubiquitously and eases of amplification which makes them all the time favorite barcode. *RbcL* region was evaluated and recommended by Plant CBOL group (CBOL, 2009).

1.2.2 The *matK* gene locus

The maturase gene *matK* is about 1500 bp long and located within the *trnK* gene encoding the tRNALys (UUU). It encodes the enzyme maturase K involved in the splicing of type-II introns from RNA transcripts (Bhadalkar et al., 2014). It is embedded in the group II introns of the lysine gene *trnK.* Although substitution rates in plastid DNA are generally low compared to nuclear DNA however, the substitution rate of the *matK* gene is among the highest in plastid genes (Johnson and Soltis, 1995; Kress and Erickson, 2007). Sometimes *matK* gene alone was adequate to resolve phylogenetic relationship and identification of plant species in a broad taxonomic range (Johnson and Soltis, 1995; Hilu et al., 2003; Kress and Erickson, 2007; Lahaye et al., 2008). The *matK* gene was rapidly evolved and ubiquitously present in plants because of it used in the plant DNA barcoding and to construct phylogenies. The plastid marker *matK* can differentiate about 90% of species in the Orchidaceae (Orchid family) but less than 49% in the Myristicaceae (nutmeg family) (Kress et al., 2005; Newmaster et al., 2008). CBOL recommended *rbcL* and *matK* gene for plant DNA barcoding (CBOL, 2009).

1.2.3 The *rpoB* and *rpoC1* gene loci

The *rpoB* and *rpoC1* produce plastid RNA polymerase subunits (Serino and Maliga, 1998). The efficacy of *rpoB* and *rpoC1* were evaluated by CBOL group and did not recommend for DNA barcoding (CBOL, 2009). Other studies suggested that *rpoB* and *rpoC1* genes can be used in combination with *matK* which will achieve higher species discrimination (Chase et al., 2007). The *rpoC1* recommended for DNA barcoding the bryophytes group (Liu et al., 2010). Thus, further research on this gene is required for deciding their suitability as a barcode.

1.2.4 The *atpF-atpH intergenic* spacer

The *atpF-atpH* intergenic spacer is a non-coding region between the genes *atpF* and *atpH* which code for ATP synthase subunits CFO I and CFO III respectively (Drager and Hallick, 1993). The efficacy of the *atpF-H* inter-generic spacer between these two genes was evaluated using flora of the Kruger National Park, South Africa (Lahaye et al., 2008). It was demonstrated that, PCR amplification was high but sequence analysis cumbersome due to variation in sequence length (Lahaye et al., 2008). However, it was suggested as a supplementary DNA barcode combination with *matK* (Drager and Hallick, 1993; Fazekas et al., 2008). The CBOL Plant Working Group reported high amplification and sequencing result of *atpF-atpH* intergenic spacer but low species identification rate compared to *rbcL* and *matK* genes (CBOL, 2009).

1.2.5 The *psbK-psbI* intergenic spacer

It is a non-coding region between the *psbK* and *psbI* genes which code two K and I photosystem II polypeptides (Meng et al., 1991). This region is highly conserved from unicellular algae to higher land plants (Knauf and Hachtel, 2002; McNeal et al., 2007; Vijayan and Tsou, 2010). The efficacy of the *psbK–psbI* intergenic region was evaluated for the Kruger National Park flora (Lahaye et al., 2008). Overall, the performance of the *psbK–psbI* was good and proposed in combination with *matK*, *trnH–psbA* and *atpF–atpH* for barcoding of plants (Lahaye et al., 2008; CBOL, 2009). The CBOL group observed that the potential discrimination of *psbK–psbI* was higher than the *matK* and other loci, but lower than *trnH–psbA* (CBOL, 2009).

1.2.6 Nuclear DNA barcode locus

The ribosomal DNA is tandemly arranged multigene family consisting of 18S, 5.8S, 26S coding regions and two internal transcribed spacers (ITS1 and ITS2) (Figure 1.4) (Vijayan and Tsou, 2010). Nuclear internal transcribed spacer (nrITS) region is a useful phylogenetic marker in plants and animals. It is ubiquitously present, exhibit biparental inheritance, and higher evolutionary pattern. However, the nuclear genes and their non-coding regions (introns) are not preferred for DNA barcoding due to lack of universal primers for amplification (Kress et al., 2005; Chen et al., 2010). Kress et al. (2005) suggested the nuclear 5.8S ribosomal gene and ITS as a probable DNA barcode for plants (Figure 1.4). It was thoroughly evaluated for species identification and discrimination. Further, ITS2 region was validated as a potential barcode for authentication of medicinal plant species (Chen et al., 2010). Moreover, efficacy of ITS barcode was assessed using the Euphorbiaceae family members (Pang et al., 2010). The China Plant BOL Group (2011) has strongly advocated the inclusion of ITS in the core barcode for plants along with *matK* and *rbcL*. All the recommendations for the universal barcode region(s) were based on floristic studies involving distantly related species.

Table 1.1 List of commonly used potential plant DNA barcode.

Barcode	**Genomic location**	**Gene coding/non-coding regions**	**Amplicon size (bp)**
rbcL	Chloroplast	Coding for large subunit of RubisCO enzyme	550-734
matK	Chloroplast	Coding for maturase K protein	734-930
trnH-psbA	Chloroplast	Non-coding intergenic spacers	296-1120
ITS1	Nuclear	Internal transcribed spacers of rRNA genes	683-724
ITS2	Nuclear	Internal transcribed spacers of rRNA genes	492-506
rpoB	Chloroplast	Coding gene for chloroplast ribosomal protein B	298-510
rpoC1	Chloroplast	Coding gene for chloroplast ribosomal protein C1	467-564
trnL-F	Chloroplast	Non-coding intergenic spacer	254-767
atpF-H	Chloroplast	Intergenic spacer of ATPase synthase genes	196-573
psbK-I	Chloroplast	Intergenic spacer between genes for low molecular weight proteins for photosystem II	444-492
Ycf5	Chloroplast	Protein coding geneencodesTic214 complex	221-382

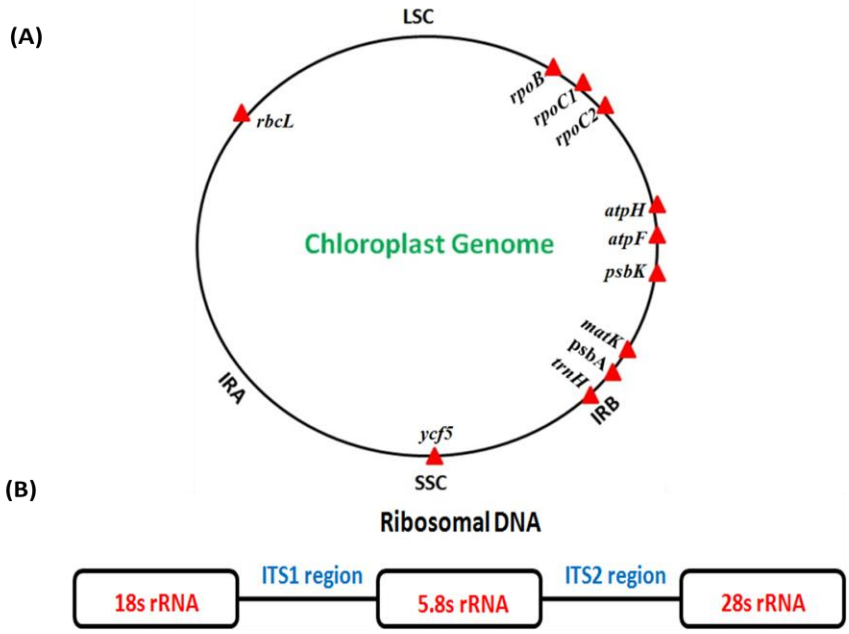


Figure 1.4 Schematic representations of plastid (A) and nuclear (B) markers commonly used in plant DNA barcoding. Abbreviations used: LSC-large single copy region, SSC-small single-copy region, IR-Large Inverted repeat (IRA, IRB), *rbcL*-Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, matK- Maturase K, *rpoB* and *rpoC1*codes for chloroplast RNA polymerase subunit, *trnH-psbA*- intergenic spacer, *atpF* and *atpH* encode ATP synthase subunits CFO I and CFO III respectively, *psbK* and *psbI* genes encode two polypeptides K and I, *ycf1* gene encodesTic214 complex, ITS - Internal Transcribed Spacer.

1.3 *Rhizophora apiculata,* an unexplored mangrove species

Mangroves are natural salt tolerant plant species also called as "halophyte" which is evolutionarily adapted to the intertidal coastal ecosystem (Kathiresan, 2018). However, the mangroves have a unique mechanism to cope with salinity stress such as ultrafiltration, minimum uptake and compartmentalization of Na^+, K^+ and Cl^- into vacuoles, synthesis of an osmoprotectant, and excess ions secreted through salt glands (Menon and Soniya, 2014). The *R. apiculata* is a hardy woody fast growing mangrove tree which belongs to Rhizophoraceae family. They are distributed throughout Indian coastal region but the dominant population prevalent in the southern coast of India. They are non-secretor and store surplus salt into their

leaves (Menon and Soniya, 2014). They are natural salt tolerant plant species but very few reports are available on salt tolerance mechanism and stress associated genes. They can tolerate salinity upto 65 parts per thousand (ppt) and showed optimum growth at 8-15 ppt salinity (Robertson and Alongi, 1992). In order to survive in harsh conditions the plants have developed some specialized trait such as viviparous propagules, aerial extensive supporting roots and high content of secondary metabolites. Several salt induced genes were isolated and characterized from *R. apiculata* using suppression subtractive hybridization technique (Menon and Soniya, 2014). The isolated genes belong to nine functional categories and highly upregulated salt-induced genes were confirmed through qRT-PCR analysis using actin (*ACT*) as a reference gene (Menon and Soniya, 2014). Recently, whole genome sequencing and comparative transcriptome analysis of *R. apiculata* was reported (Xu et al., 2017*).* Moreover, transcriptome analysis of Rhizophoraceae family members such as *B. gymnorrhiza, K. obovata,* and *C. tagal* were studied to understand the adaptive evolutionary mechanism in the harsh intertidal habitats (Guo et al., 2017). This is an ideal plant species to study salt stress tolerance mechanism and isolate related genes. However, there are no systematic reports available on selection and validation of reference gene for qRT-PCR in *R. apiculata* species.

1.4 Selection and validation of candidate reference genes for qRT-PCR

Several techniques are available to investigate gene expression analysis including, semi-quantitative reverse transcription polymerase chain reaction, northern blotting, in situ hybridization, and quantitative real-time PCR (qRT-PCR). The qRT-PCR is a reliable, sensitive, and wide quantification range gene expression analysis technique (Bustin, 2002). Moreover, reference gene for qRT-PCR normalization is not universally standardized and it varies

according to plant tissue material and experimental conditions (Bustin et al., 2009). For precise quantification and reproducible profiling, selection and validation of stable candidate reference genes are crucial steps prior to qRT-PCR for data normalization. Some commonly used reference genes include Actin (*ACT*), β-tubulin (*β-TUB*), Ubiquitin (*UBQ*), Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), Elongation factor 1α (*EF1α*) and 18S ribosomal RNA (18S) are preferred to normalize the expression profiles of candidate reference genes. These reference genes are involved in basic cellular functions, maintaining cell size and shape, and cellular metabolism (Bustin, 2002). However, several reports have shown that the level of reference genes expression varies in different cultivars, tissues, and stress conditions (Sinha et al., 2015; Reddy et al., 2015; Nikalje et al., 2018). Hence, it is very important to select and validate most appropriate reference genes involved in a various experimental conditions before proceeding to gene expression analysis. A various web-based tools and algorithms are available to address validation of candidate reference genes including, comparative ΔCt (cycle thresholds) (Silver et al., 2006), NormFinder (Andersen et al., 2004), BestKeeper (Pfaffl et al., 2004), and geNorm algorithm (Vandesompele et al., 2002). RefFinder is a web-based program which provides a comprehensive ranking of reference genes (Xie et al., 2012).

1.4.1 The minimum information for publication of qRT-PCR experiments

MIQE provides a set of necessary guidelines for detection and measurement of tiny amounts of nucleic acids in a diverse source of samples in a qPCR experiments (Bustin et al., 2009). Bustin et al., (2009) devised qPCR guidelines and suggestions for how to perform, analyse and publish qPCR experiment. Moreover, they revised and introduced unique nomenclature and abbreviations system such as qPCR for quantitative PCR and RT-qPCR for reverse

transcription–qPCR. They suggested that normalization genes should be called as reference genes instead of housekeeping genes and TaqMan probes term replaced with hydrolysis probes. Moreover, the threshold cycle (Ct), crossing point (Cp), and take-off point (TOP) are most commonly used terminology in the literature but they replaced by quantification cycle (Cq). Besides this, Bustin et al., (2009) defined all conceptual qPCR terminology such as analytical and clinical sensitivity, specificity, accuracy, repeatability, and reproducibility (Bustin et al., 2009). An analytical sensitivity of qPCR is defined as the minimum number of nucleic acid copies in a sample that can be measured accurately, whereas clinical sensitivity is referred to the percentage of individuals detected with a given disease as positive for that condition. An analytical specificity is defined as discrimination and detecting ability of qPCR between target sequences from nonspecific sequences present in a sample. The term accuracy defined as the difference between experimental measured versus actual concentrations presented as fold changes or copy number. The initial quantity and quality of nucleic acid (DNA and RNA) is very important for qPCR experiment.

1.5 Salinity stress and SOS mechanism in plants

Abiotic stresses such as drought, salinity, heat, cold, and anaerobic stress are imposing negative effect on plant growth and productivity (Cavanagh et al., 2008; Munns and Tester, 2008; Chinnusamy and Zhu, 2009; Mittler and Blumwald, 2010; Kumar et al., 2013). Compared to other abiotic stresses, soil salinity is one of the brutal climatic factors which impose hyperionic and osmotic stress leads to the limiting crop productivity (Munns and Tester, 2008; Kumar et al., 2013). Soil salinity is a global issue which affected approximately 45 million hectares of irrigated land and about 1.5 million hectares of productive land turned into non-fertile lands

every year (Munns and Tester, 2008). Salt stress is affecting plants in several ways such as ion imbalances due to Na^+ and Cl^- accumulation, nutritional disorders, oxidative stress, alteration of metabolic processes, membrane disorganization, enhanced lipid peroxidation and increased production of reactive oxygen species reduction of cell division and expansion (Kumar et al., 2013). Recent studies have identified various adaptive responses to salinity stress at cellular, molecular, physiological and biochemical levels. The Na^+ enters into the cell through various membrane transporters families such as HKT family and plasma-membrane non-selective cation channels (NSCCs) (Gupta and Huang, 2014). However, plants are evolved with several counter mechanisms to cope with salinity stress including minimum uptake of Na^+, compartmentalization into the vacuoles and effluxing surplus Na^+ out of the cell (Ji et al., 2013; Sandhu et al., 2017). Discovery of a novel salt overly sensitive (SOS) pathway in plants opened the new horizon of salt tolerance mechanism (Ji et al., 2013; Sandhu et al., 2017). Currently, there are three key components of SOS pathways which regulates the SOS activity in salt stress such as SOS1 (NHX), SOS2 (protein kinase) and SOS3 (SCaBP8, calcium sensor) as shown in figure 1.5 (Ji et al., 2013; Sandhu et al., 2017). Interestingly, the important role of SOS1 in the halophyte, *Thellungiella halophila* was demonstrated by knockdown studies of SOS1, which led to loss of halophytic characteristics (Oh et al., 2009).

The prevalent function of SOS1 (NHX) under salt stress is extrusion of Na^+ out of cell. However, the sequestration of excess Na^+ into vacuole is another crucial process to minimize cellular $Na^{+,}$ toxicity and also helped to build osmotic potential inside the cell to facilitate water uptake into the cell (Gupta and Huang, 2014). Plant NHX belongs to cation/proton exchanger 1 (CPA1) superfamily, which consists of three subfamilies (Sandhu et al., 2017).

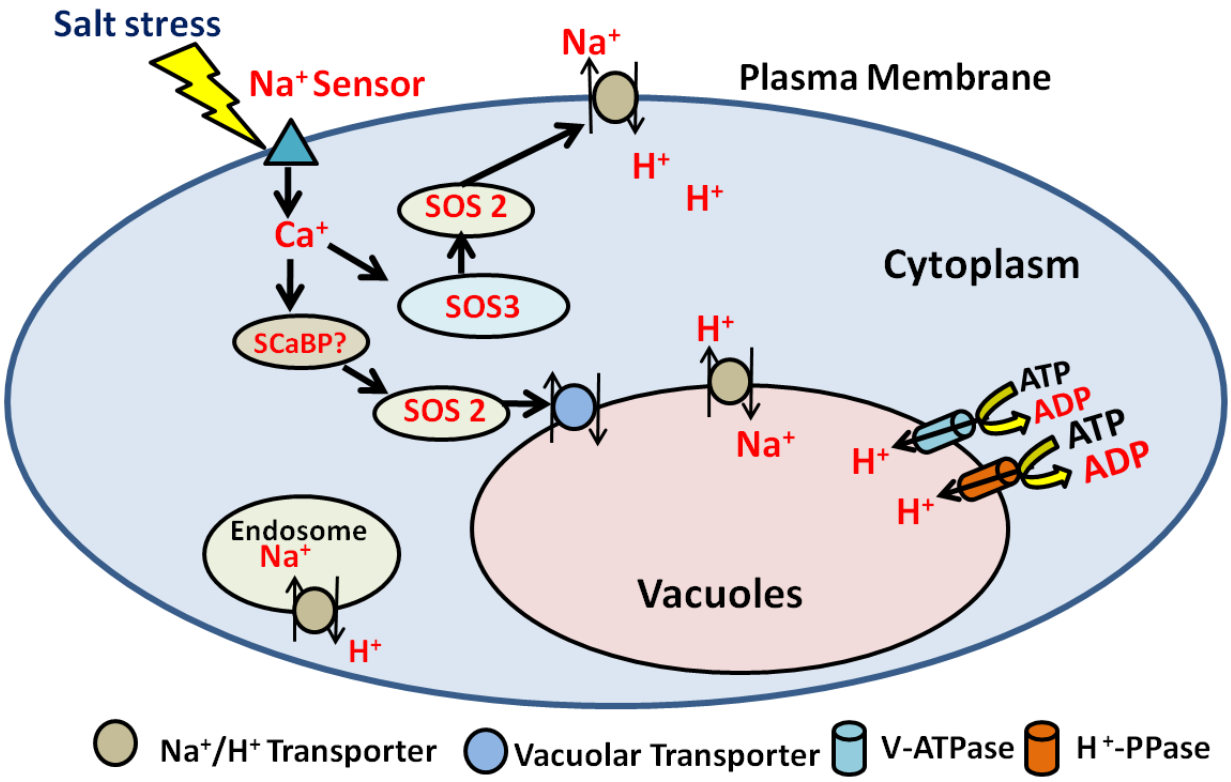


Figure 1.5 Schematic presentation of salt overlay sensitive (SOS) mechanism in plant cells. Under salt stress unknown membrane sensor responds to stress and release Ca^{2+}. Ca^{2+}activates SOS3 and/or ScaBP (calcium sensor) which will activates SOS1 (NHX7/8) transporter through SOS2 protein kinase. SOS-salt overlay sensitive transporter present on plasma membrane, NHX- sodium hydrogen exchanger localized on endomembrane and vacuolar membrane, V-ATPase pumps H^+ inside vacuoles using ATP, and H^+-PPase- pumps H^+ inside vacuoles using ATP, ATP-adenosine triphosphate, ADP-adenosine diphosphate, SCaBP-calcium binding protein.

Based on the phylogenetic analysis, group I is mainly localized to vacuolar membrane including AtNHX1-AtNHX4 members. Similarly group II is localized to endosomal membrane such as vesicles, Golgi, trans-Golgi network, and pre-vacuolar compartment and it includes AtNHX5 and AtNHX6 members (Figure 1.6). Moreover, group III members are present on plasma membrane which includes *AtNHX7* and *AtNHX8* (Pittman et al., 2012; Pires et al., 2013).

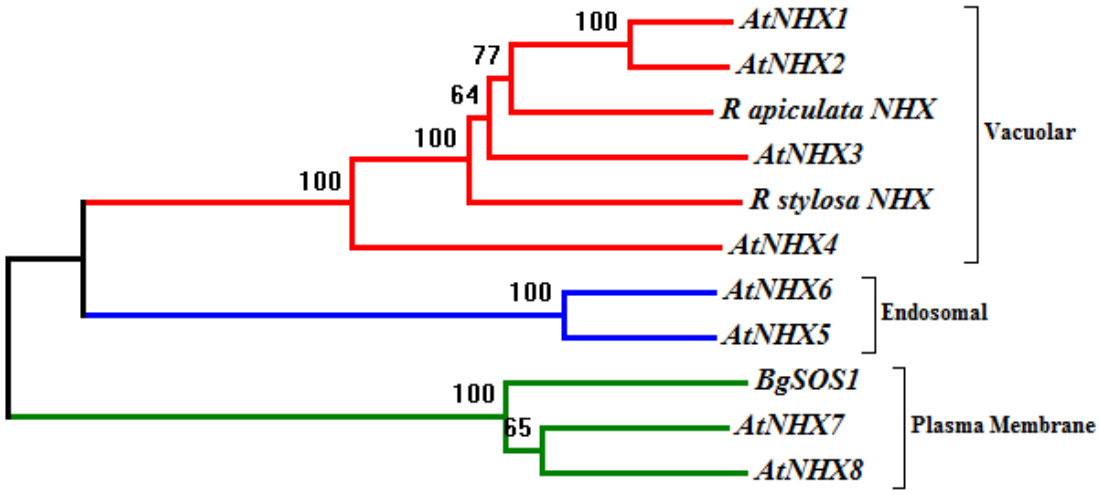


Figure 1.6 Phylogenetic analyses of *Arabidopsis thaliana* NHX gene family members with mangroves NHX members such as *Rhizophora stylosa RsNHX*, *R. apiculata RaNHX*, and *Bruguiera gymnorrhiza SOS1*. NHX family categorized in the three classes based on the subcellular localization such as vacuolar, endosomal and plasma membrane.

Many reports are available on functional characterization of plant NHX under salinity stress and also their involvement in several physiological process such as transport of Na^+ as well as K^+, osmotic adjustment and water uptake, growth and development of cell, vesicular trafficking and protein targeting, calcium signaling, stomatal movements as well as flowering (Andres et al., 2014; Bassil and Blumwald, 2014; Bassil et al., 2011; Pittman, 2012; Reguera et al., 2014; Apse et al., 2003). First NHX gene was successfully isolated from *Arabidopsis* (NHX1) which was localized to the vacuole and showed significant transcript abundance in salt and ABA stress (Apse et al., 1999; Shi and Zhu, 2002). Later on, several NHX members were identified in different plant species such as *Oryza sativa* (Fukuda et al., 1999), *Gossypium hirsutum* (Wu et al., 2004), *Zea mays* (Zorb et al., 2005), *Triticum aestivum* (Brini et al., 2005), *Solanum lycopersicum* (Rodríguez-Rosales et al., 2008), *Vigna radiata* (Mishra et al., 2014), *Vigna unguiculata* (Mishra et al., 2015), *Pennisetum glaucum* (Bhaskaran and Savithramma, 2011), *Medicago truncutula* (Sandhu et al., 2017). In *Arabidopsis thaliana*, total eight members of

NHX family were reported which was localized on the three subcellular compartments. Previous studies showed the different expression patterns and response to various abiotic stresses (Yokoi et al., 2002; Aharon et al., 2003). The transcript abundance of *AtNHX1* and *AtNHX2* were shown to be expressed significantly in all plant tissues, while *AtNHX3* and *AtNHX4* were prevantely expressed in root and flower respectively (Yokoi et al., 2002; Aharon et al., 2003). Moreover, low expression pattern of *AtNHX5* was observed in all plant tissues, while *AtNHX6* expression reported in shoots and roots (Yokoi et al., 2002). Recently, six members of *Medicago truncutula* NHX family was characterized and performed transcript abundance in leaves and roots libraries (Sandhu et al., 2017). It was observed that *MtNHX1*, *MtNHX6,* and *MtNHX7* were significantly expressed in the roots and leaves but the expression of *MtNHX3*, *MtNHX6,* and *MtNHX7* were triggered in the roots under salinity stress (Sandhu et al., 2017). Similar attempts were reported in halophytic plants including *Mesembryanthemum crystallinum* (Chauhan et al., 2000), *Atriplex gmelini* (Hamada et al., 2001), *Beta vulgaris* (Xia et al., 2002), *Populus euphratica* (Ye et al., 2009), *Salicornia brachiata* (Jha et al., 2011), *Suaeda salsa* (Ma et al., 2004), *Halostachys caspica* (Guan et al., 2011), and *Zygophyllum xanthoxylum* (Bao et al., 2015). In *Mesembryanthemum crystallinum* plant *NHX* transcript was up-regulated in leaves and stems, and involved in the leaf tissues to pump surplus sodium into vacuoles under stress condition (Chauhan et al., 2000). Ye et al., (2009) reported six members of the NHX family in the *Populus euphratica.* Moreover, *PeNHX1*, 3, and 6 transcripts levels were significantly higher in the roots, stems, and leaves compared with the *PeNHX2,* 4, *and* 5 (Ye et al., 2009). All *PeNHX* were complemented the yeast (R100) salt sensitive phenotype which showed their importance under salt stress (Ye et al., 2009).

Overexpression studies of NHX from different plant species conferred salt tolerance as well as maintained ion homeostasis in *Arabidopsis* (Apse et al., 1999), *Solanum lycopersicum* (Zhang and Blumwald, 2001), *Brassica napus* (Zhang et al., 2001), and *Triticum aestivum* (Xue et al., 2002). Similarly, ectopic expression of *OsNHX1* in *Oryza sativa* showed improved tolerance under salinity stress (Fukuda et al., 1999, 2004). Moreover, overexpression of *Gossypium hirsutum GhNHX1* in tobacco had been shown to confer salt tolerance (Wu et al., 2004). Overexpression and mutant analysis of *Arabidopsis* NHX members have underscored their important adaptive roles in salinity as well as to maintain physiological functions (Apse et al., 2003). Mutant lines of *AtNHX1* have shown decreased Na^+, K^+/H^+ activity, variation in leaf shape and size compared with wild-type plants, suggesting that *NHX1* plays an important role in the plant development (Apse et al., 2003). Constitutive expression of *Hordeum brevisubulatum NHX1* (*HbNHX1*) in tobacco showed improved phenotypes in salt and drought stress (Lu et al., 2005). Similarly, overexpression of *Pennisetum glaucum NHX1* (*PgNHX1*) in rice and in *Brassica juncea* conferred salinity tolerance in transgenic plants (Verma et al., 2007; Rajagopal et al., 2007). Expression of *Medicago sativa MsNHX1* showed enhanced salt tolerance in *Arabidopsis* (Bao-Yan et al., 2009). Moreover, NHX was co-expressed with many other regulatory genes which conferred tolerance against salinity stress. Co-expression of xerophyte *Zygophyllum xanthoxylum ZxNHX* with *ZxVP1-1* (H^+-PPase) conferred improved salt stress tolerance in sugar beet (*Beta vulgaris L.*) (Wu et al., 2015). Similarly, co-expression of tonoplast cation/H^+-antiporter and *ZxVP1-1* (H^+-pyrophosphatase) from xerophyte *Zygophyllum xanthoxylum* improved alfalfa plant growth under salinity, drought, and field conditions (Bao et al., 2015). Co-overexpression of a plasma membrane sodium/proton antiporter (*AtSOS1*) and a vacuolar membrane sodium/proton antiporter (*AtNHX1*) significantly enhanced salinity

tolerance in *Arabidopsis* under 250 mM NaCl and concluded that the stacked overexpression of two genes could significantly enhanced tolerance against multiple abiotic stresses (Pehlivan et al., 2016). Co-expression of *Pennisetum glaucum* vacuolar Na^+/H^+ antiporter (*PgNHX*) and *Arabidopsis* H^+-pyrophosphatase (H^+-PPase) in the tomato showed improved salt tolerance, vigorous growth in the presence of 200 mM NaCl while WT plants exhibited chlorosis and died (Bhaskaran and Savithramma, 2011). Shen et al., (2015) performed the co-overexpression of *AVP1* (H^+-pyrophosphatase) and *AtNHX1* in cotton improved drought and salt tolerance in transgenic plants (Shen et al., 2015). Expression of wheat Na^+/H^+ antiporter (*TaNHX1*) and H^+-pyrophosphatase (*TaVP1*) genes in tobacco enhanced salt tolerance (Gouiaa et al., 2012).

Based on the review of literature, it is envisaged to clone and functionally characterize Na^+/H^+ antiporter (NHX1) gene from *Rhizophora apiculata* and its transcript profiling under salt stress.

1.6 Gaps in Existing Research

1. Earlier reports are mainly focused on the mangroves diversity and distribution, floral physiology, ecology, and structure of vegetation types.
2. There are no worthwhile reports available to address molecular taxonomy of mangroves and morphology based taxonomic parameters extensively influenced by the environmental condition of the habitat.
3. There are no reports available on the selection and validation of reference genes for qRT-PCR for *R. apiculata*
4. There are no published reports available on cloning and functional characterization of *R. apiculata* NHX.

1.7 To address the aforementioned gaps the following objectives were framed:

1. Identification of mangrove species prevalent in Goa based on DNA barcode markers.
2. Molecular cloning and *in-silico* analysis of sodium/proton antiporter (NHX) genes from *Rhizophora* species.
3. Functional characterization of NHX in heterologous system.

CHAPTER 2

DNA Barcoding of Goan Mangroves

CHAPETR 2

DNA barcoding of Goan mangroves, West Coast India

2.1 Introduction

DNA barcoding is a molecular tool that enables rapid and accurate identification of plant species (Li et al., 2015). The Consortium for the Barcode of Life plant working group (CBOL) evaluated 7 leading candidate DNA regions (*rbcL*, *matK*, *trnH–psbA* spacer, *atpF–atpH* spacer, *rpoB*, *rpoC1*, and *psbK–psbI* spacer) (CBOL, 2009). The CBOL plant working group recommended two-locus combinations of *rbcL* and *matK* as the core plant barcode complemented with *trnH-psbA* intergenic spacer based on the parameters of recoverability, sequence quality, and levels of species discrimination (Kress et al., 2005; CBOL, 2009; Hollingsworth et al., 2011). China Plant Barcode of Life recommended the internal transcribed spacer (ITS) as an additional candidate plant DNA barcode (China Plant BOL, 2011).

In the present study, we have evaluated the efficacy of commonly used DNA barcode such as *rbcL*, *matK*, ITS2, *rpoC1*, *atpF-atpH* spacer, and *psbK-psbI* spacer for Goan mangrove identifications. Goa state is located in West Coast of India and mangrove vegetation occupies 26 km^2 of area (Ragavan et al., 2016; Kathiresan, 2018). The Cumbarjua canal (15 km) links the two river channels of Mandovi and Zuari, forming an estuarine complex which supports a substantial mangrove extent. Recent study revealed that, total 17 mangrove species were distributed in Goa that include 14 true and 3 associated mangrove species (D'Souza and Rodrigues, 2013). Further, acquired barcode (*rbcL*, *matK*, ITS2, *atpF-atpH*, *rpoC1* and *psbK-psbI*) data analyzed using various barcoding methods which can be divided to four classes: (I) similarity match methods like BLAST analysis (II) classical phylogenetic method such as a

Neighbor-Joining and Maximum Likelihood/Bayesian algorithms (III) k-nearest Neighbor based on the K2P distance and statistical approaches based on classification algorithms, and (IV) genealogical methods based on coalescent theory using maximum likelihood/Bayesian algorithms based on Monte Carlo Markov Chains (MCMC) (Austerlitz et al., 2009).

2.2 Materials and Methods

2.2.1Ethical statement

The mangrove species leaf samples were collected from different sampling points of Goa, West Coast of India. All the mangroves samples were collected with the permission of the Principal Chief Conservator of Forest, Goa Forest Department, India. Moreover, mangroves species are not endangered or protected species.

2.2.2 Sample collection

In the present study, leaves samples of 14 mangrove species were collected from Goa, located on the West Coast of India with geographical latitude of 15.5256°N and longitude of 73.8753°E. Mangrove species identification was performed based on morphological characteristics using a mangroves identification guide (Duke and Bunt, 1979; Duke and Jackes, 1987; Duke et al., 1991; Duke and Ge, 2011; Dhargalkar et al., 2014). Herbarium of these specimens was deposited at Botanical Survey of India, Western Regional Centre Pune, India. The sequences obtained using barcode markers: *rbcL*, *matK*, ITS2, *rpoC1*, *psbK-psbI*, and *atpF-atpH* were submitted to the NCBI GenBank and publicly accessible through the dataset of project DNA Barcoding of Indian Mangroves (Project code: IMDB) in Barcode of Life Data systems (BOLD) (doi:10.5883/DS-IMDBNG) (Ratnasingham and Hebert, 2007).

2.2.3 DNA extraction

Mangroves are exposed to harsh environmental conditions and in order to sustain they are rich in mucilage, latex, phenolics, secondary metabolites and polysaccharides content. So these plants are more difficult system for protein and nucleic acid isolation. Cetyl-trimethyl ammonium bromide (CTAB) protocol of plant DNA extraction was modified (Parani et al., 1997a). Leaf tissue was pulverized in liquid nitrogen and 0.2 g tissues were mixed with CTAB buffer (25 mM EDTA; 1.4 M NaCl; 2% PVP-30; 1% β-mercaptoethanol; 10% SDS and 10 $mg.ml^{-1}$ proteinase K). The suspension was incubated at 60°C for 60 min with gentle mixing and centrifuged at 14,000 rpm for 10 m at room temperature with equal volume of chloroform: isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and DNA was precipitated with 0.6 volume of cold isopropanol (-20°C) and chilled 7.5 M ammonium acetate was added followed by incubation at -20°C for 1 h. The precipitated DNA was centrifuged at 14,000 rpm for 10 m at 4°C followed by washing with 70% ethanol. DNA was finally dissolved in TE buffer (10 mM Tris–HCl, 1 mM Na_2EDTA, pH 8.0) and quantity and quality confirmed on the 1% agarose gel and the spectrophotometer (Nanodrop, Thermo Scientific, USA).

2.2.4 PCR and sequencing

Amplification of plastid as well as nuclear genes (*rbcL, matK,* ITS2*, rpoC1, psbK-psbI, and atpF-atpH*) were performed in final 50µl PCR reaction mixture containing 10-20 ng of template DNA, 200 µM of dNTPs (Thermo Scientific, USA), 0.1 µM of each primers and 1 unit of Taq DNA polymerase (Sigma Aldrich, USA). The reaction mixture containing the template DNA was amplified in Bio-Rad (T100 model, Bio-Rad) thermal cycler with temperature profile for *rbcL* (94°C for 4 m; 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 m; repeated for 35

cycles, final extension 72°C for 10 m) and for *matK* (94°C for 1 m; 35 cycles of 94°C for 30 s, 50°C for 40 s, 72°C for 40 s; repeated for 37 cycles, final extension 72°C for 5 m). The temperature profile for ITS2 (94°C for 4 m; 35 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min; final extension 72°C for 10 m), *atpF-atpH* (94°C for 1 m; 35 cycles of 94°C for 30 s, 50°C for 40 s, 72°C for 40 s; final extension 72°C for 5 m), *psbK-psbI* (94°C for 5 m; 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s; final extension 72°C for 10 m), *rpoC1* (94°C for 5 m; 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s; final extension 72°C for 10 m). The amplified products were separated on 1.2% agarose gel and stained with ethidium bromide (Maniatis et al., 1982). Two pairs of universal primers *rbcL* (*rbcL*a_F and *rbcL*a_R) and *matK*_390f and *matK*_1326r were used for the amplification of DNA samples (Kress and Erickson, 2007; Vinitha et al., 2014; Chen et al., 2015) (Table 2.1). To amplify *R. apiculata matK* locus, we designed *matK*_RA reverse primer. PCR Similarly, a universal vascular plants ITS2 pair of primers was used to amplify ITS region (Chen et al., 2010; White et al., 1990). Moreover, primers sequences of *rpoC1*, *psbK-psbI,* and *atpF-atpH* were reported in CBOL and other studies (CBOL, 2009; Lahaye et al., 2008) (Table 2.1). PCR products were purified according to manufacturer's instruction (Chromous Biotech Pvt. Ltd) and further sequencing reactions were carried out using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on ABI 3500xL Genetic Analyzer (Applied Biosystems).

Table 2.1 List of primers pairs used in DNA barcoding of Goan mangroves.

Sr. No.	Primer label	Primer Sequence 5'-3'	References
1	*rbcLa_F*	ATGTCACCACAAACAGAGACTAAAGC	Levin, 2003
2	*rbcLa_R*	GTAAAATCAAGTCCACCRCG	Kress and Erickson, 2007
3	*matK*_390f	CGATCAATTCATTCACTATTTC	Cuenoud et al., 2002
4	*matK*_1326r	AAAGTTCGTTTGTGCCAATGA	Cuenoud et al., 2002
5	*matK*_Rar	AAAGTTCGTTTGTGCCAATGA	Present study
6	ITS2_S2F	ATGCGATACTTGGTGTGAAT	Chen et al., 2010
7	ITS4_R	TCCTCCGCTTATTGATATGC	White et al., 1990
8	*rpoC1*_2F	GGCAAAGAGGGAAGATTTCG	CBOL, 2009
9	*rpoC1*_4R	CCATAAGCATATCTTGAGTTGG	CBOL, 2009
10	*psbK*_F	TTAGCCTTTGTTTGGCAAG	CBOL, 2009
11	*psbI*_R	AGAGTTTGAGAGTAAGCAT	CBOL, 2009
12	*atpF*_F	ACTCGCACACACTCCCTTTCC	Lahaye et al., 2008
13	*atpH*_R	GCTTTTATGGAAGCTTTAACAAT	Lahaye et al., 2008

2.2.5 Data Analysis

Sequence alignment and assembly was achieved in Codon code Aligner v.3.0.1 (Codon Code Corporation) and MEGA 7 (Kumar et al., 2016). The NCBI BLAST was performed to confirm identity of specimens (Altschul et al., 1990). All known mangroves sequences were searched with our sequenced data using BLAST tool against NCBI database and highest-scoring hit from each query is taken as the mangrove identification. Intraspecific, interspecific and barcode gap analysis was performed in BOLD systems. Further, *rbcL* concatenated with *matK* and *matK* concatenated with ITS2 sequences using DNASP v5.10 tool. All the individual barcode as well as concatenated sequences were analyzed in MEGA 7 for their resolution inference (Rozas, 2009). Neighbor-joining (NJ) trees with K2P genetic distance model were constructed using MEGA 7 and node support was assessed based on 1000 bootstrap replicates. Species with multiple individuals forming a monophyletic clade in phylogenetic trees with a bootstrap value above 60% were considered as successful identification.

2.2.6 TaxonDNA

TaxonDNA v1.6.2 analysis for species identification with 'Best Match' and 'Best Closest Match' method was performed (Meier et al., 2006). The threshold (T) was set at 95% and all the results above the threshold (T) value were treated as 'incorrect'. Similarly, all matches of the query sequence were below threshold (T) considered to be the 'correct' identification. The matches of the query sequences were good but corresponded to a mixture of species treated as ambiguous identification (Meier et al., 2006).

2.2.7 Automated Barcode Gap Discovery (ABGD)

The ABGD is a web server based distance method, which can partition the sequences into potential species based on the barcode gap, whenever the divergence within the same species is smaller than organisms from different species. The ABGD analysis was performed with two relative gap width (X = 1.0, 1.5) and three distance metrics (Jukes-Cantor, K2P, and p-distance) with default parameters (Puillandre et al., 2012; Zou et al., 2016; Yang et al., 2016).

2.2.8 General Mixed Yule-Coalescent (GMYC)

The GMYC method requires a fully resolved ultrametric tree for analysis. The Bayesian tree was built using BEAST v1.8 (Drummond, 2006; Drummond and Rambaut, 2007). Input file (XML) for BEAST was compiled in BEAUti v1.83 with an HKY+G molecular evolutionary model for the ITS2 dataset and GTR+G for concatenated dataset of *matK*+ITS2. These models were derived using PartitionFinder V1.1.1. Tree prior was set to Yule process and the length of Markov chain Monte Carlo (MCMC) chain was 40,000,000 generation and sampling was performed at every 4000 step. However, all other settings were kept as default. Convergence of

the BEAST runs to the posterior distribution. The adequacy of sampling was based on the Effective Sample Size (ESS) and diagnostic was assessed with Tracer v1.4. After removing the first 20% of the samples as burn-in, all other runs were combined to generate posterior probabilities of nodes from the sampled trees using TreeAnnotator v1.7.4. Estimation of the number of species included in the tree was analyzed using GMYC with single and multiple thresholds in R by the APE and SPLITS packages (Zou et al., 2016; Drummond and Rambaut, 2007; Drummond, 2006; Gernhard, 2008; Kumar et al., 2009; R Core Team, 2012; Paradis et al., 2004; Ezard et al., 2009).

2.2.9 Poisson Tree Process model (PTP)

The PTP model is a tree-based method that differentiates specimen into populations and species level using coalescence theory (Puillandre et al., 2012; Zou et al., 2016; Yang et al., 2016). The RaxML tree was constructed using CIPRES portal and input data was generated for bPTP analysis. The calculations were conducted on the bPTP webserver (http://species.h-its.org), with the following parameters (500,000 MCMC generations, thinning 100 and burn-in 25%).

2.3 Results

2.3.1 Mangroves identification

DNA barcoding is a molecular tool which helps to correctly assign respective species, but prior to that traditional taxonomic identification must be required. In the current work, we have collected total 14 mangroves species, which belongs to 9 genera and 5 families. List of 14 mangroves species and their morphological identification keys (peculiar characteristics of leaves, flower and fruits) were given in the Table 2.2 and Figure 2.1.

Table 2.2 Morphological key features of 14 Mangroves species. Morphological identification of mangroves species based on leaves, flowers, and fruits.

	Species name	**Morphological Key features**			
		Leaves	**Flowers**	**Fruits**	**References**
1	*Sonneratia alba*	Glabrous simple, elliptic, oblong or sub orbicular, obtuse at apex	Inflorescence 1 to few flowered cyme, stamens white in color	Green, calyx cup shaped	Dhargalkar et al., 2014
2	*Sonneratia caseolaris*	Simple, opposite decussate apex becomes acute at young stage	Inflorescence solitary cyme, stamens white above and reddish below	A berry, green	Dhargalkar et al., 2014
3	*Rhizophora mucronata*	Dark green, broadly ovate, leaf tip mucronate	Inflorescence axillary cyme, petals hairy in the margin	Coriaceous, one celled, one seeded, epigeal viviporous	Duke et al 1999; Dhargalkar et al., 2014
4	*Rhizophora apiculata*	10 to 18 cm long and 2.5 to 7.5 cm broad, opposite decussate, leaf tip acute	Inflorescence cyme, sessile, corolla and petals white	Capsule 3-5 cm long, oval, viviparous	Duke et al., 1999; Dhargalkar et al., 2014
5	*Kandelia candel*	Opposite, oblong elliptic, apex obtuse	Inflorescence axillary cyme , petals 5-6 bifide white	1.5-2.5 cm long, obclavate, viviparous, up to 40 cm maturity	Duke et al., 1999; Dhargalkar et al., 2014
6	*Ceriops tagal*	Simple, coriaceous, obovate-oblong, leathery apex rounded	Inflorescence axillary cyme or glomerules, petals white with clavate appendanges	Slightly conical, 1.5- 2 cm long, germination epigeal, viviparous	Dhargalkar et al., 2014
7	*Bruguiera cylindrica*	Simple, opposite decussate, cauline, yellowish	Inflorescence axillary cyme, flower yellow-green, calyx tube ribbed, petal margin hairy	Radicle sub-cylindrical, germination epigeal, viviparous, hypocotyle 15 cm	Duke and Ge, 2011; Dhargalkar et al., 2014
8	*Bruguiera gymnorrhiza*	Opposite decussate, slightly acuminate, petiole reddish green in colour, acute apex	Inflorescence solitary cyme, flower scarlet, calyx deep orange red/ yellow, campanulate	Capsule pendulous with persistent calyx, Showy with reddish calyx	Dhargalkar et al., 2014
9	*Avicennia alba*	Simple, opposite, lanceolate, long and pointed at apex, dark green on upperside	Flowers orange-yellow colour with around 10-30 flowers on each unit	Conical shape, 1-4cm long, smooth velvety outer skin	Dhargalkar et al., 2014
10	*Avicennia marina*	Opposite, elliptic-oblong beneath, acute at apex	Inflorescence capitate, pale yellow in terminal condensed cyme	Capsule, russet brown, almond shaped	Dhargalkar et al., 2014
11	*Avicennia officinalis*	Entire, elliptic, oblong	Terminal or axillary cyme, corolla orange yellow	Capsule, crypto-viviparous	Duke and Bunt, 1979; Dhargalkar et al., 2014

12	*Acanthus ilicifolius*	Simple, oblong or ovate, sinuate margin with spine tipped lobes	Inflorescence raceme, violet flower	Green, kidney shaped	Duke and Bunt, 1979; Dhargalkar et al., 2014
13	*Aegiceras corniculatum*	Alternate, obtuse	Inflorescence umbellate, terminal or axillary white flower	Sharply pointed, green yellowish brown	Duke and Jackes, 1987; Dhargalkar et al., 2014
14	*Excoecaria agallocha*	Simple and cluster to the end of shoot	Male inflorescence catkin and female mixed cyme	3 lobed, Schizocarp	Duke and Jackes,1987; Dhargalkar et al., 2014

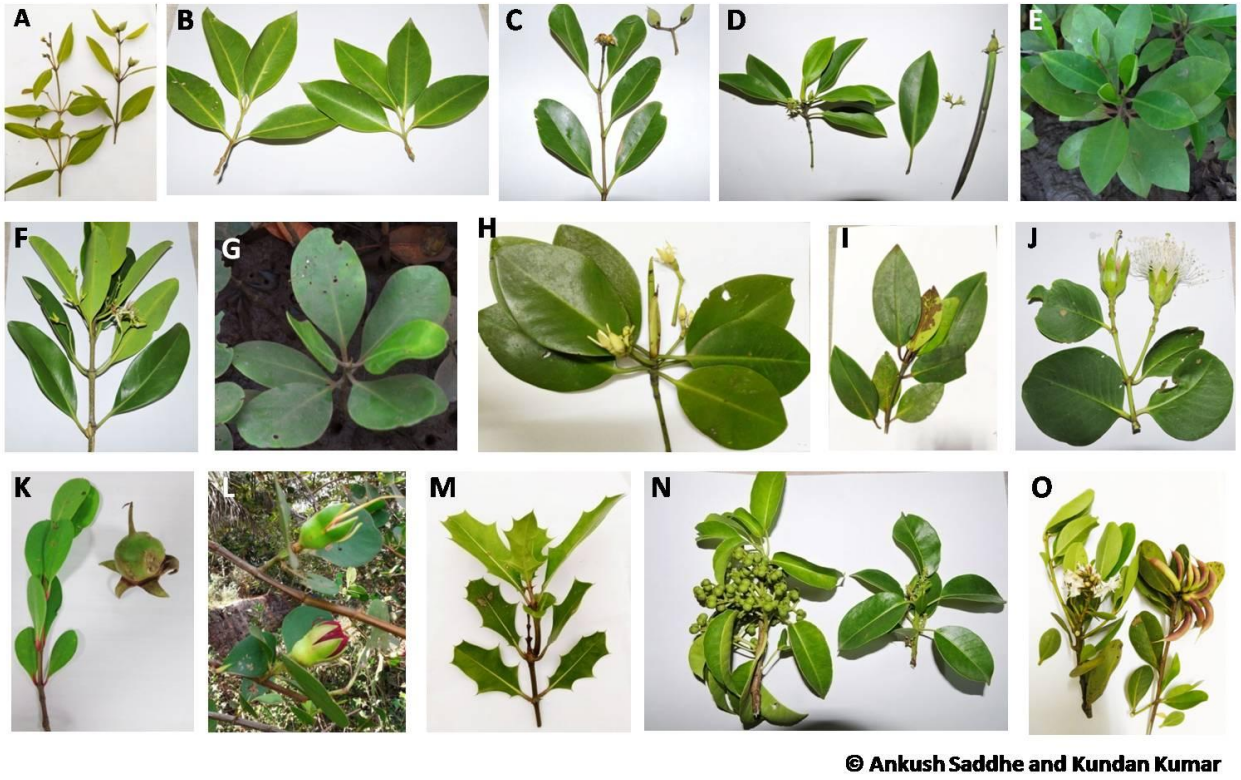


Figure 2.1 Photographs of 14 mangroves species used in our study (A) *Avicennia marina* (B) *A. alba* (C) *A. officinalis* (D) *Bruguiera cylindrica* (E) *B. gymnorrhiza* (F) *Kandelia candel* (G) *Ceriops tagal* (H) *Rhizophora mucronata* (I) *R. apiculata* (J) *Sonneratia alba* (K, L) *S. caseolaris (*M) *Acanthus ilicifolius* (N) *Excoecaria agallocha* and (O) *Aegiceras corniculatum.*

A total of 148 sequences (44 *rbcL*, 43 *matK*, 40 ITS2, 9 *atpF-atpH*, 6 *psbK-psbI* and 6 *rpoC1*) were acquired from 44 specimens of mangrove belonging to 14 species, 9 genera, and 5 families. The sequences (*rbcL*: 510 bp, *matK*: 712 bp, ITS2: 445 bp, *atpF-atpH*: 511 bp, *psbK-psbI*: 360 bp and *rpoC1*: 451 bp) with few insertions and deletions, without stop codon, along with the specimen collection details were submitted to the Barcode of Life Data Systems (BOLD) in form of a project `IMDB' (dx.doi.org/10.5883/DS-IMDBNG). These sequences were submitted to the NCBI GenBank through BOLD systems and their accession numbers were obtained (Table 2.3).

Table 2.3 List of NCBI GenBank accession numbers used in the present work.

Specimen	**Accession No. *rbcL***	**Accession No. *matK***	**Accession No. ITS2**
Avicennia officinalis	KP697351, KP697352, KU748517	KP725238, KP725239	KU876892, KU876893
Avicennia marina	KP697349, KP697350, KM255068	KP725236, KM255083, KP725237	KU876889, KU87689, KU876891
Avicennia alba	KM255067, KM255069, KP697348	KM255082, KM255084, KP725235	KU876886, KU876887, KU876888
Bruguiera cylindrica	KP697354, KM255070, KP697353	KP725241, KM255085, KP725240	KU876894,KU87689, KU876896
Bruguiera gymnorrhiza	KM255071, KP697355, KP697356	KM255086, KP725242, KP725243	KU876897,KU876898, KU876899
Rhizophora mucronata	KM255077, KU748519	KM255092, KU748522, KU748523	KU876910,KU876911, KY250446
Rhizophora apiculata	KP697362, KP697363, KM255076	KP725249, KP725250, KM255091	KU876908, KU876909, KY250445
Aegiceras corniculatum	KM255066, KP697344, KP697345	KM255081, KP725231, KP725232	KU876881, KU876882, KU87688
Excoecaria agallocha	KM255073, KP697360, KP697359	KM255088, KP725247, KP725246	KU876903, KU876904, KU876905
Kandelia candel	KP697361, KM255074	KP725248, KM255089	KU876906, KU87690, KY250444
Ceriops tagal	KM255072, KP697358 KP697357	KM255087, KP725244, KP725245	KU876900,KU876901, KU876902
Sonneratia alba	KM255078, KP697364	KM255093, KP725251	KY250447, KY250448, KY250449
Sonneratia caseolaris	KP697365,KP697366, KM255079	KP725252, KP725253, KM255094	KY250450, KY250451

Acanthus ilicifolius	KM255065, KP697342, KP697343	KM255080, KP725229, KP725230	KY250442,KY250443
Specimen	**Accession No. *atpF-atpH***	**Accession No. *psbK-psbI***	**Accession No. *rpoC1***
Avicennia officinalis	KY754573, KY754574, KY754575	KY754564, KY754565, KY754566	KY754187, KY754188, KY754189
Avicennia marina	KY754570, KY754571, KY754572	KY754561, KY754562, KY754563	KY754184, KY754185, KY754186
Avicennia alba	KY754567, KY754568, KY754569		

2.3.2 Genetic distance analysis

The scatter plot represented the number of individuals in each species against their maximum intraspecific distances, as a test for sampling bias (Figure 2.2). Sequence analysis was performed to estimate the average GC content of the corresponding locus. The average GC content in a DNA sequences were observed in *rbcL* was 43.29% (SE = 0.09), while in *matK*, it was 33.18% (SE = 0.18). Moreover, GC contents were observed in ITS2-63.11%, *atpF-atpH*-35.18%, *psbK-psbI*-31.22% and *rpoC1*-44.6% loci.

The genetic distances were calculated for individual barcode marker by K2P model on the BOLD system. Mangrove exhibited absolute average interspecific differentiation of 0.35% and 0.9% in *rbcL* and *matK* respectively, while for species average intraspecific variability was 0.24% in *rbcL* and 0.20% in *matK* with low species resolution in few taxa (Table 2.4). The mean intraspecific distance for ITS2, *atpF-atpH, psbK-psbI* and *rpoC1* was calculated as 1.85%, 0.11%, 1.63% and 0.37% respectively. While mean intrageneric distance for ITS2, *atpF-atpH, psbK-psbI* and *rpoC1* was calculated as 5.8%, 1.03%, 2.16% and 0.3% respectively (Table 2.4). The intraspecific and interspecific analysis for *rbcL* revealed largest average pairwise distance of 0.68, while in *matK*, it was observed as 2.05 and 2.32 respectively.

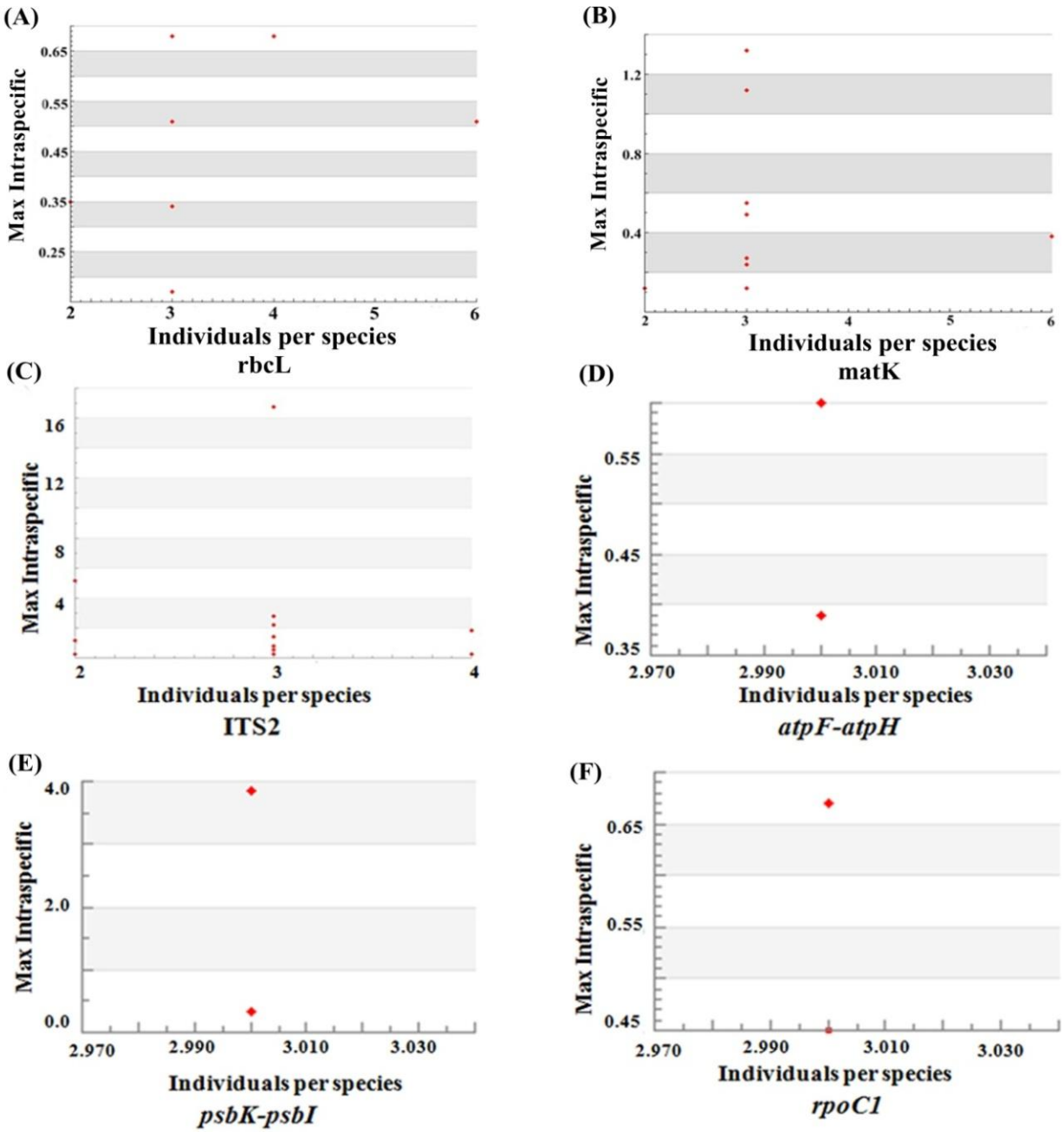


Figure 2.2 The scattered plot showed the number of individuals in each species against their maximum intra-specific distances, as a test for sampling bias. (A) *rbcL* (B) *matK* (C) ITS2 (D) *atpF-atpH* (E) *rpoC1* and (F) *psbK-psbI*.

Table 2.4 (A) Genetic divergence table calculated for all sequenced DNA barcode. (B) Distribution of intra and inter specific Kimura 2 parameter (K2P) mean divergence for *atpF-atpH, psbK-psbI and rpoC1* are represented in table for *Avicennia* genus. N- Number of specimens, Min- Minimum, Max- Maximum, Dist- Distance, SE- Standard error, and NN- Nearest Neighbor.

(A)

Barcode	Level	N	Taxa	Comparisons	Min Dist (%)	Mean Dist (%)	Max Dist (%)	SE Dist (%)
rbcL	Species	44	14	50	0	0.24	0.68	0
	Genus	26	4	53	0	0.35	0.68	0
	Family	29	2	132	1.71	2.63	4.01	0
matK	Species	45	14	50	0	0.2	1.32	0.01
	Genus	25	4	45	0	0.9	2.32	0.02
	Family	29	2	141	2.11	5.82	13.37	0.02
ITS2	Species	40	14	39	0	1.85	16.75	0.1
	Genus	25	4	45	0	5.8	35.14	0.25
	Family	28	2	133	5.72	12.35	40.26	0.08
matK+ ITS2	Species	39	14	37	0	0.51	4.02	0.02
	Genus	24	4	43	0	1.76	7.84	0.05
	Family	28	2	133	3.35	7.39	19.89	0.03
atpF-atpH	Species	9	3	9	0	0.11	0.6	0.02
	Genus	9	1	27	0.39	1.03	1.62	0.02
psbK-psbI	Species	6	2	6	0	1.63	3.85	0.27
	Genus	6	1	9	0.96	2.16	4.94	0.14
rpoC1	*Species*	6	2	6	0.22	0.37	0.67	0.03
	Genus	6	1	9	0	0.3	0.67	0.02

(B)

	atpF-atpH		*psbK-psbI*		*rpoC1*	
	Max. Intraspecific	Min Interspecific NN	Max. Intraspecific	Min Interspecific NN	Max. Intraspecific	Min Interspecific NN
A. officinalis	0.39	0.8	3.85	0.96	0.67	0
A. marina	0	0.39	0.32	0.96	0.45	0
A. alba	0.6	0.39	NA	NA	NA	NA

The highest range of congeneric differentiation in *Bruguiera* and *Avicennia* were observed in *rbcL* from 0 to 0.68, whereas for *matK*, it ranged from 1.29 to 2.31 in *Avicennia*, suggesting significant genetic divergence within *Avicennia* genus. The barcode gap analysis revealed highest intraspecific distance (> 2%) in 9 specimens of *rbcL* and 6 specimens of *matK*, while

low intraspecific distance (< 2%) in 11 specimens of *rbcL* and 9 specimens of *matK*. Here, low intraspecific distance (< 2%) suggests low species resolution, thus leading to species overlap. With *rbcL,* the largest nearest neighboring distance of 8.43 was observed in *Avicennia alba* with mean intraspecific distance of 0.11. The maximum intraspecific distance of 0.68 was observed within individuals of *K. candel, B. gymnorrhiza, A. officinalis* and *S. caseolaris*. With *matK*, maximum intraspecific distance of 2.05 was observed in *E. agallocha* with three individuals per species, while largest distance to the nearest neighbor of 24.65 was observed in *A. officinalis* with mean intraspecific distance of 0.12. Overall average nearest neighboring divergence observed among mangroves using *rbcL* was 1.39% (S.E = 0.17) and *matK* was 4.07% (S.E = 0.5) (Figure 2.3).

Highest intraspecific distance (> 2%) for ITS2 was observed in 19.51% individuals and *S. alba* exhibited highest intraspecific distance of 16.75%. While lower intrageneric distances (<2%) for ITS2 were observed in 50.98% individuals and *A. marina* showed the lowest intrageneric distance of 0%. Higher intraspecific distances for *matK*+ITS2 were observed in 9.30% individuals and *S. alba* exhibited the highest distance of 4.01%. While lower intrageneric distances were observed in almost 90.69% individuals (Table 2.4). In some species intraspecific distance was higher than the intrageneric distance. Six species (*A. alba*, *A. officinalis*, *A. marina*, *B. cylindrica*, *B. gymnorrhiza* and *R. mucronata*) were resolved with ITS2, while in concatenation of *matK*+ITS2, error rates were minimized in two species (*A. officinalis* and *A. marina*). *Avicennia* genus analysis has revealed low resolution among them based on *rbcL*, *matK* and ITS2. To resolve this cryptic genus, we used few supplementary markers such as *atpF-atpH, psbK-psbI* and *rpoC1*. *Avicennia* genus showed intraspecific distance ranging from 0%-1.0% with almost all barcode markers, with highest intraspecific distance (> 2%) was

observed in *psbK-psbI* (3.85%) (Table 2.4). While lower intrageneric distance (< 2%) was observed in nearly all barcode markers, except for *psbK-psbI* (4.94%).

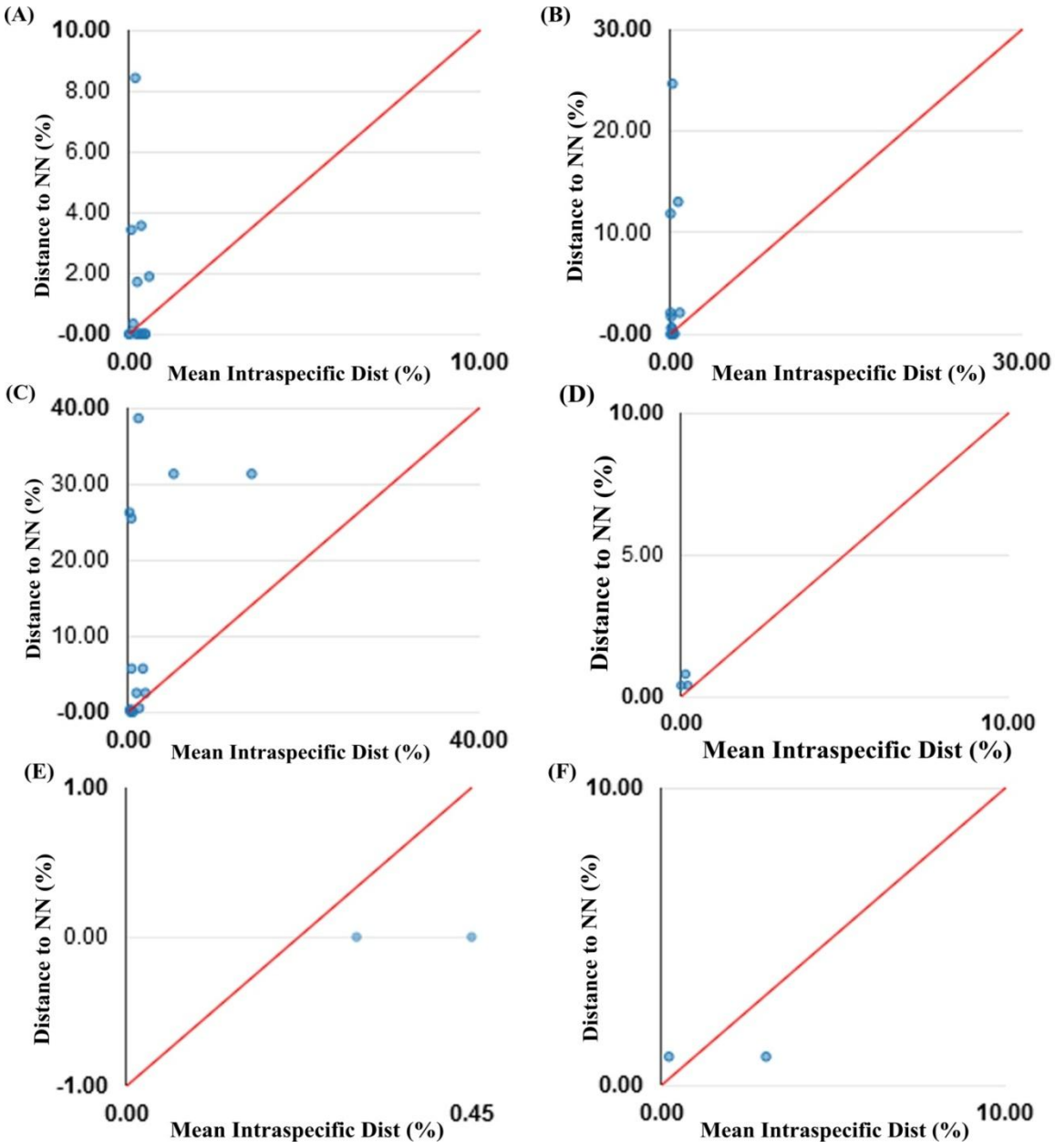
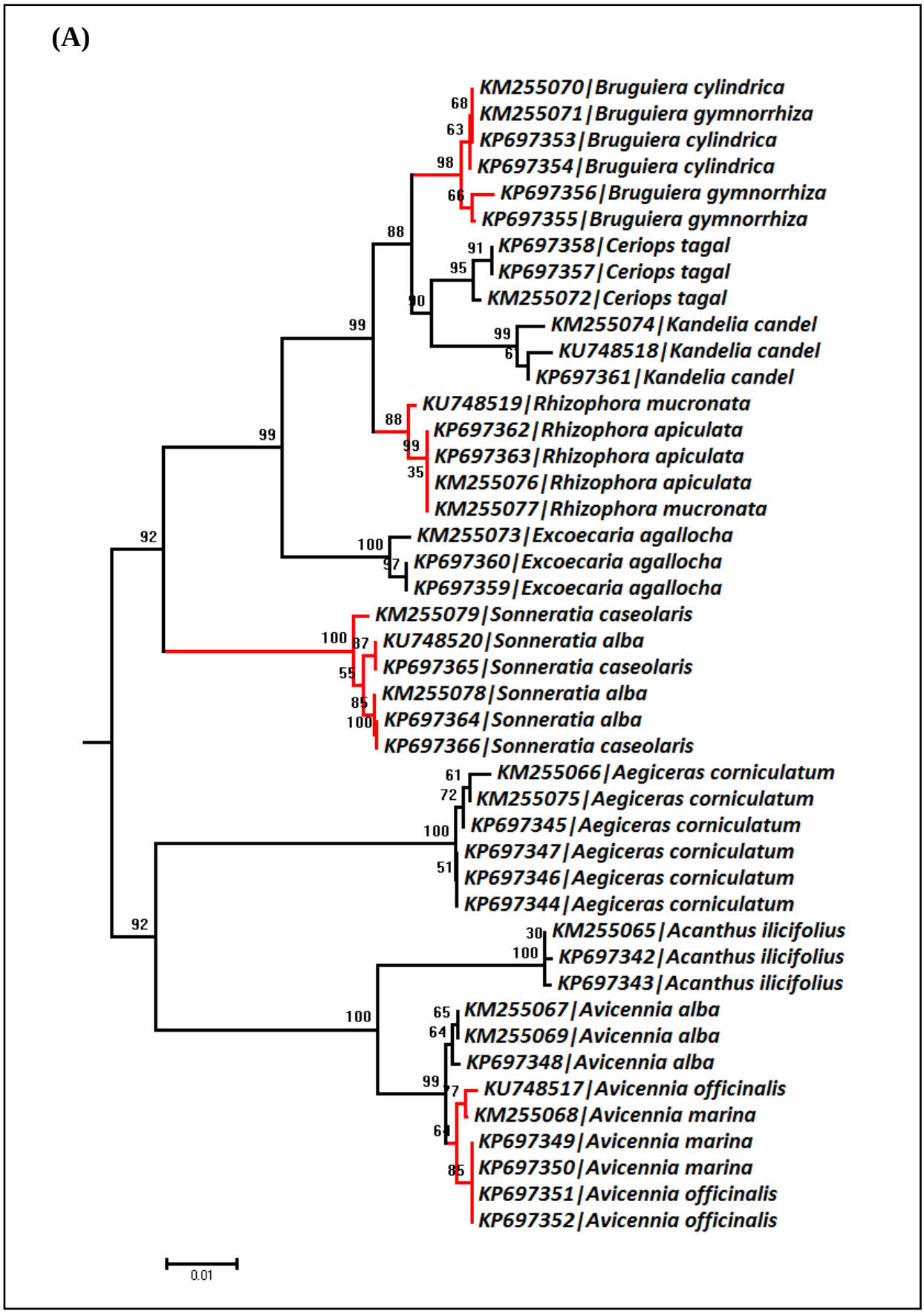
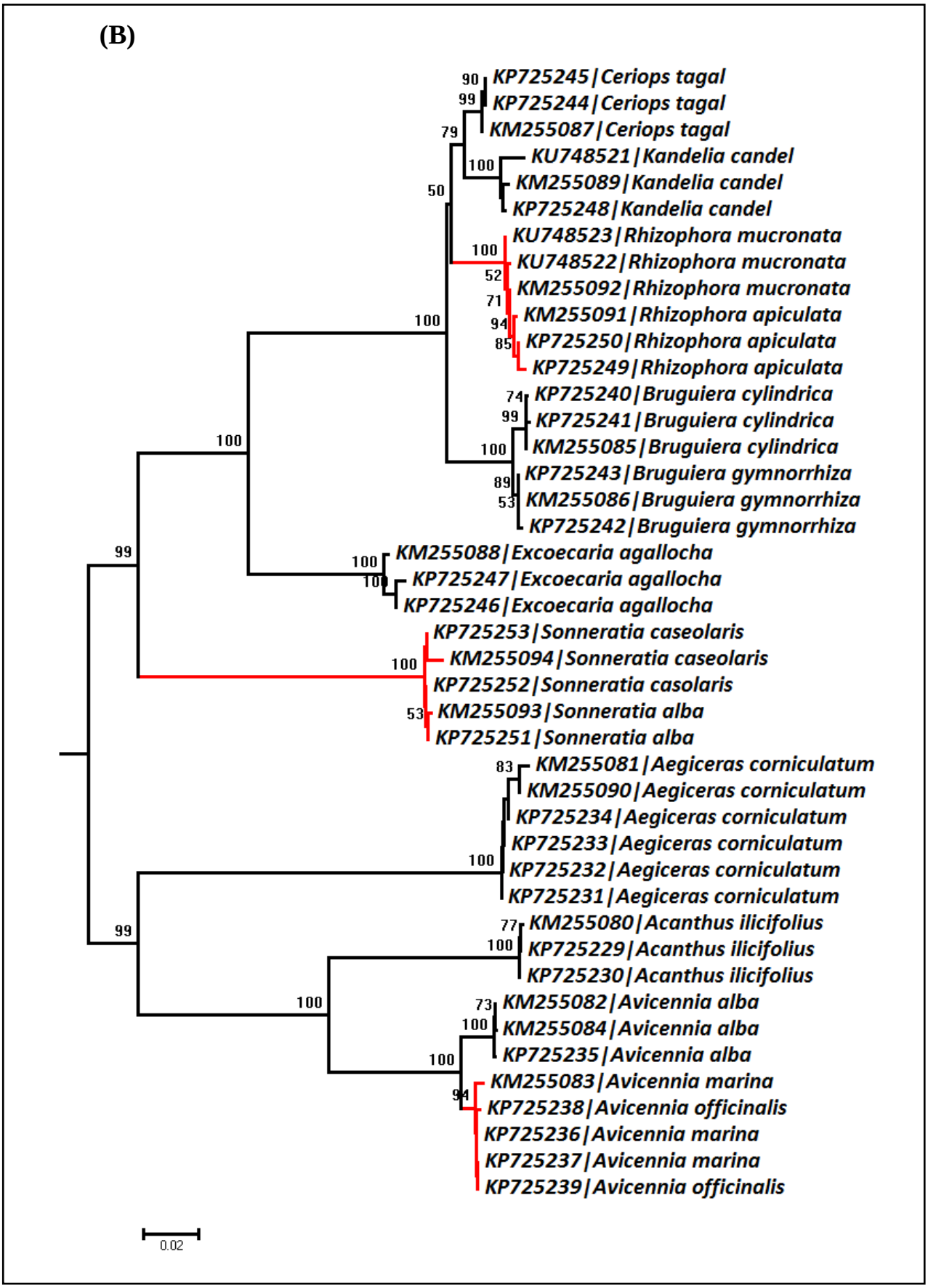


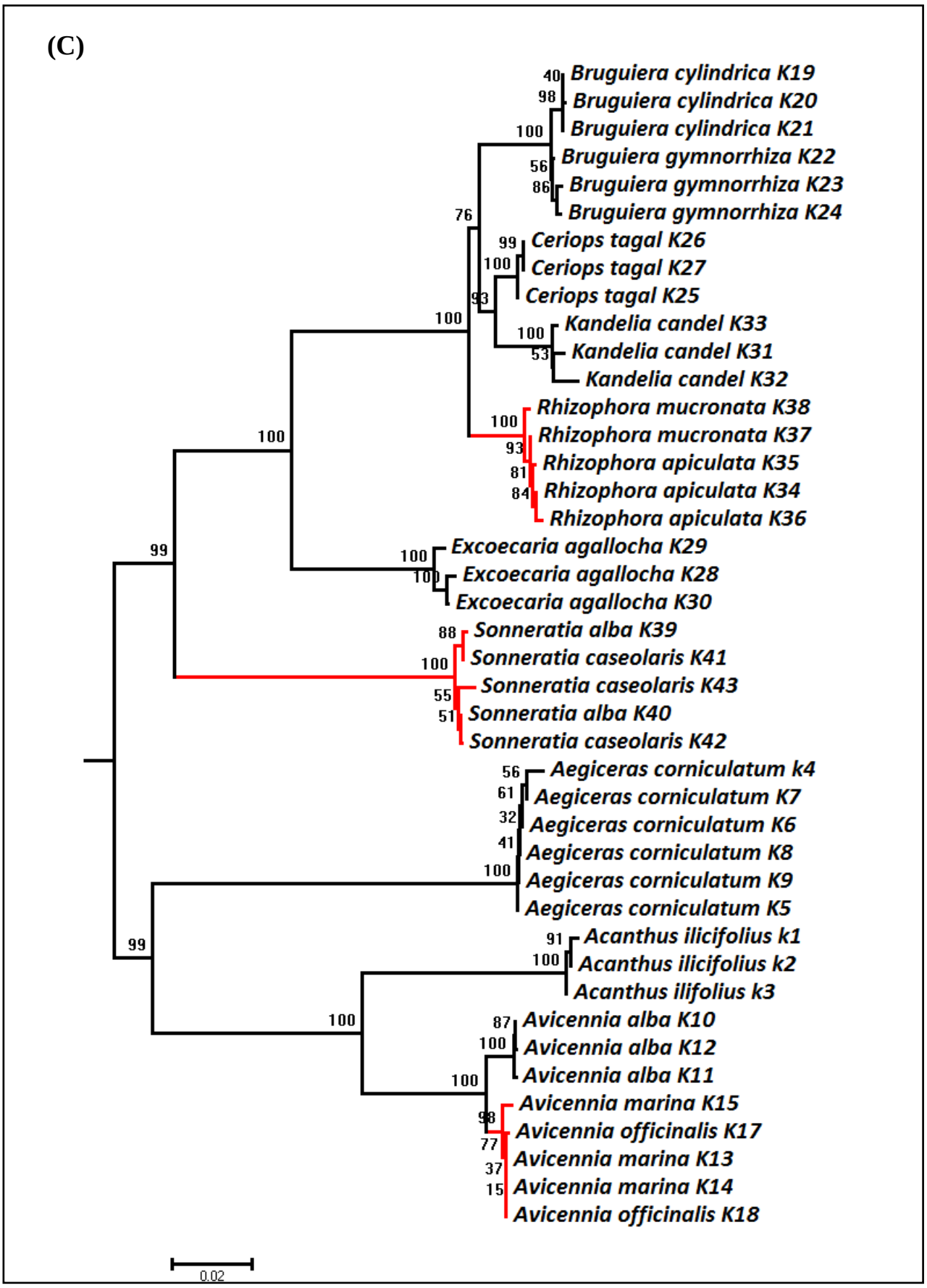
Figure 2.3 The scatter plots confirmed the existence and magnitude of the barcode gap. The given scatter plots showed the overlap of the max and mean intra-specific distances vs. the inter-specific (nearest neighbor) distances. (A) *rbcL* (B) *matK* (C) ITS2 (D) *atpF-atpH* (E) *rpoC1* and (F) *psbK-psbI*.

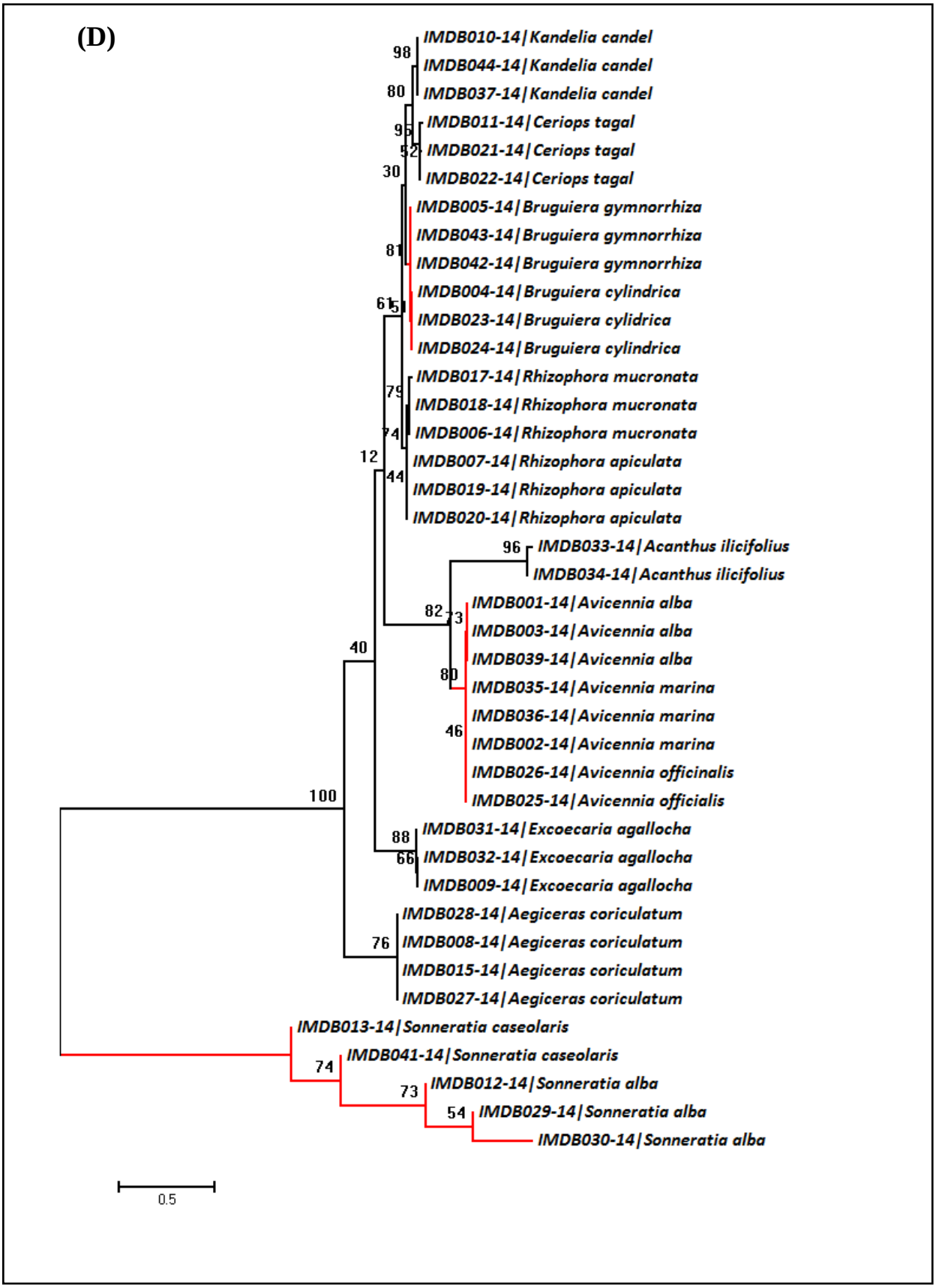
2.3.3 Taxonomic assignment of mangrove species using NJ tree with K2P

The Neighbor-Joining (NJ) is one of the commonly used phylogenetic methods with Kimura 2-parameter (K2P) distance matrix. Species with multiple individuals forming a monophyletic clade in NJ trees with a bootstrap value greater than 60% were considered as successful identifications (Kress et al. 2010). NJ trees were constructed for individual barcode such as *rbcL, matK*, ITS2 and concatenated barcode such as *rbcL*+*matK* and *matK*+ITS2. Barcode *rbcL* formed 9 cluster and *K. candel, C. tagal, A. ilicifolius, E. agallocha, A. corniculatum* were discriminated. Similarly, *matK* also formed 9 clades which discriminated same mangrove species as discriminated by *rbcL*. The *matK* and *rbcL*+*matK* discriminated mangrove species in NJ model test method, while *rbcL* alone failed to identify those species (Figure 2.4 A, B, C). Further analysis revealed similar rates of species resolution using both methods for *matK* as well as *rbcL*. *Rhizophora*, *Sonneratia* and *Avicennia* genera were failed to discriminate their species using plastid markers *rbcL, matK* and *rbcL*+*matK*. Altogether 40 DNA barcodes from ITS2 and *matK*+ITS2 were used for species delineation. The NJ (K2P) trees constructed with bootstrap support (1000) for individual ITS2 and concatenated with ITS2+*matK*. ITS2 barcode formed 9 clusters and failed to discriminate *Avicennia, Sonneratia*, *Rhizophora,* and *Bruguiera* genus (Figure 2.4 C). Concatenated *matK*+ITS2 exhibited substantial resolution among the operational taxonomic units (OTUs) corresponding to their genera except for *A. marina* and *A. officinalis* (Figure 2.4 D and E).


(A)
KM255070|Bruguiera cylindrica
KM255071|Bruguiera gymnorrhiza
KP697353|Bruguiera cylindrica
KP697354|Bruguiera cylindrica
KP697356|Bruguiera gymnorrhiza
KP697355|Bruguiera gymnorrhiza
KP697358|Ceriops tagal
KP697357|Ceriops tagal
KM255072|Ceriops tagal
KM255074|Kandelia candel
KU748518|Kandelia candel
KP697361|Kandelia candel
KU748519|Rhizophora mucronata
KP697362|Rhizophora apiculata
KP697363|Rhizophora apiculata
KM255076|Rhizophora apiculata
KM255077|Rhizophora mucronata
KM255073|Excoecaria agallocha
KP697360|Excoecaria agallocha
KP697359|Excoecaria agallocha
KM255079|Sonneratia caseolaris
KU748520|Sonneratia alba
KP697365|Sonneratia caseolaris
KM255078|Sonneratia alba
KP697364|Sonneratia alba
KP697366|Sonneratia caseolaris
KM255066|Aegiceras corniculatum
KM255075|Aegiceras corniculatum
KP697345|Aegiceras corniculatum
KP697347|Aegiceras corniculatum
KP697346|Aegiceras corniculatum
KP697344|Aegiceras corniculatum
KM255065|Acanthus ilicifolius
KP697342|Acanthus ilicifolius
KP697343|Acanthus ilicifolius
KM255067|Avicennia alba
KM255069|Avicennia alba
KP697348|Avicennia alba
KU748517|Avicennia officinalis
KM255068|Avicennia marina
KP697349|Avicennia marina
KP697350|Avicennia marina
KP697351|Avicennia officinalis
KP697352|Avicennia officinalis
0.01


(B)
KP725245|Ceriops tagal
KP725244|Ceriops tagal
KM255087|Ceriops tagal
KU748521|Kandelia candel
KM255089|Kandelia candel
KP725248|Kandelia candel
KU748523|Rhizophora mucronata
KU748522|Rhizophora mucronata
KM255092|Rhizophora mucronata
KM255091|Rhizophora apiculata
KP725250|Rhizophora apiculata
KP725249|Rhizophora apiculata
KP725240|Bruguiera cylindrica
KP725241|Bruguiera cylindrica
KM255085|Bruguiera cylindrica
KP725243|Bruguiera gymnorrhiza
KM255086|Bruguiera gymnorrhiza
KP725242|Bruguiera gymnorrhiza
KM255088|Excoecaria agallocha
KP725247|Excoecaria agallocha
KP725246|Excoecaria agallocha
KP725253|Sonneratia caseolaris
KM255094|Sonneratia caseolaris
KP725252|Sonneratia casolaris
KM255093|Sonneratia alba
KP725251|Sonneratia alba
KM255081|Aegiceras corniculatum
KM255090|Aegiceras corniculatum
KP725234|Aegiceras corniculatum
KP725233|Aegiceras corniculatum
KP725232|Aegiceras corniculatum
KP725231|Aegiceras corniculatum
KM255080|Acanthus ilicifolius
KP725229|Acanthus ilicifolius
KP725230|Acanthus ilicifolius
KM255082|Avicennia alba
KM255084|Avicennia alba
KP725235|Avicennia alba
KM255083|Avicennia marina
KP725238|Avicennia officinalis
KP725236|Avicennia marina
KP725237|Avicennia marina
KP725239|Avicennia officinalis
0.02


(C)
Bruguiera cylindrica K19
Bruguiera cylindrica K20
Bruguiera cylindrica K21
Bruguiera gymnorrhiza K22
Bruguiera gymnorrhiza K23
Bruguiera gymnorrhiza K24
Ceriops tagal K26
Ceriops tagal K27
Ceriops tagal K25
Kandelia candel K33
Kandelia candel K31
Kandelia candel K32
Rhizophora mucronata K38
Rhizophora mucronata K37
Rhizophora apiculata K35
Rhizophora apiculata K34
Rhizophora apiculata K36
Excoecaria agallocha K29
Excoecaria agallocha K28
Excoecaria agallocha K30
Sonneratia alba K39
Sonneratia caseolaris K41
Sonneratia caseolaris K43
Sonneratia alba K40
Sonneratia caseolaris K42
Aegiceras corniculatum k4
Aegiceras corniculatum K7
Aegiceras corniculatum K6
Aegiceras corniculatum K8
Aegiceras corniculatum K9
Aegiceras corniculatum K5
Acanthus ilicifolius k1
Acanthus ilicifolius k2
Acanthus ilifolius k3
Avicennia alba K10
Avicennia alba K12
Avicennia alba K11
Avicennia marina K15
Avicennia officinalis K17
Avicennia marina K13
Avicennia marina K14
Avicennia officinalis K18
0.02


(D)
IMDB010-14|Kandelia candel
IMDB044-14|Kandelia candel
IMDB037-14|Kandelia candel
IMDB011-14|Ceriops tagal
IMDB021-14|Ceriops tagal
IMDB022-14|Ceriops tagal
IMDB005-14|Bruguiera gymnorrhiza
IMDB043-14|Bruguiera gymnorrhiza
IMDB042-14|Bruguiera gymnorrhiza
IMDB004-14|Bruguiera cylindrica
IMDB023-14|Bruguiera cylidrica
IMDB024-14|Bruguiera cylindrica
IMDB017-14|Rhizophora mucronata
IMDB018-14|Rhizophora mucronata
IMDB006-14|Rhizophora mucronata
IMDB007-14|Rhizophora apiculata
IMDB019-14|Rhizophora apiculata
IMDB020-14|Rhizophora apiculata
IMDB033-14|Acanthus ilicifolius
IMDB034-14|Acanthus ilicifolius
IMDB001-14|Avicennia alba
IMDB003-14|Avicennia alba
IMDB039-14|Avicennia alba
IMDB035-14|Avicennia marina
IMDB036-14|Avicennia marina
IMDB002-14|Avicennia marina
IMDB026-14|Avicennia officinalis
IMDB025-14|Avicennia officialis
IMDB031-14|Excoecaria agallocha
IMDB032-14|Excoecaria agallocha
IMDB009-14|Excoecaria agallocha
IMDB028-14|Aegiceras coriculatum
IMDB008-14|Aegiceras coriculatum
IMDB015-14|Aegiceras coriculatum
IMDB027-14|Aegiceras coriculatum
IMDB013-14|Sonneratia caseolaris
IMDB041-14|Sonneratia caseolaris
IMDB012-14|Sonneratia alba
IMDB029-14|Sonneratia alba
IMDB030-14|Sonneratia alba
0.5

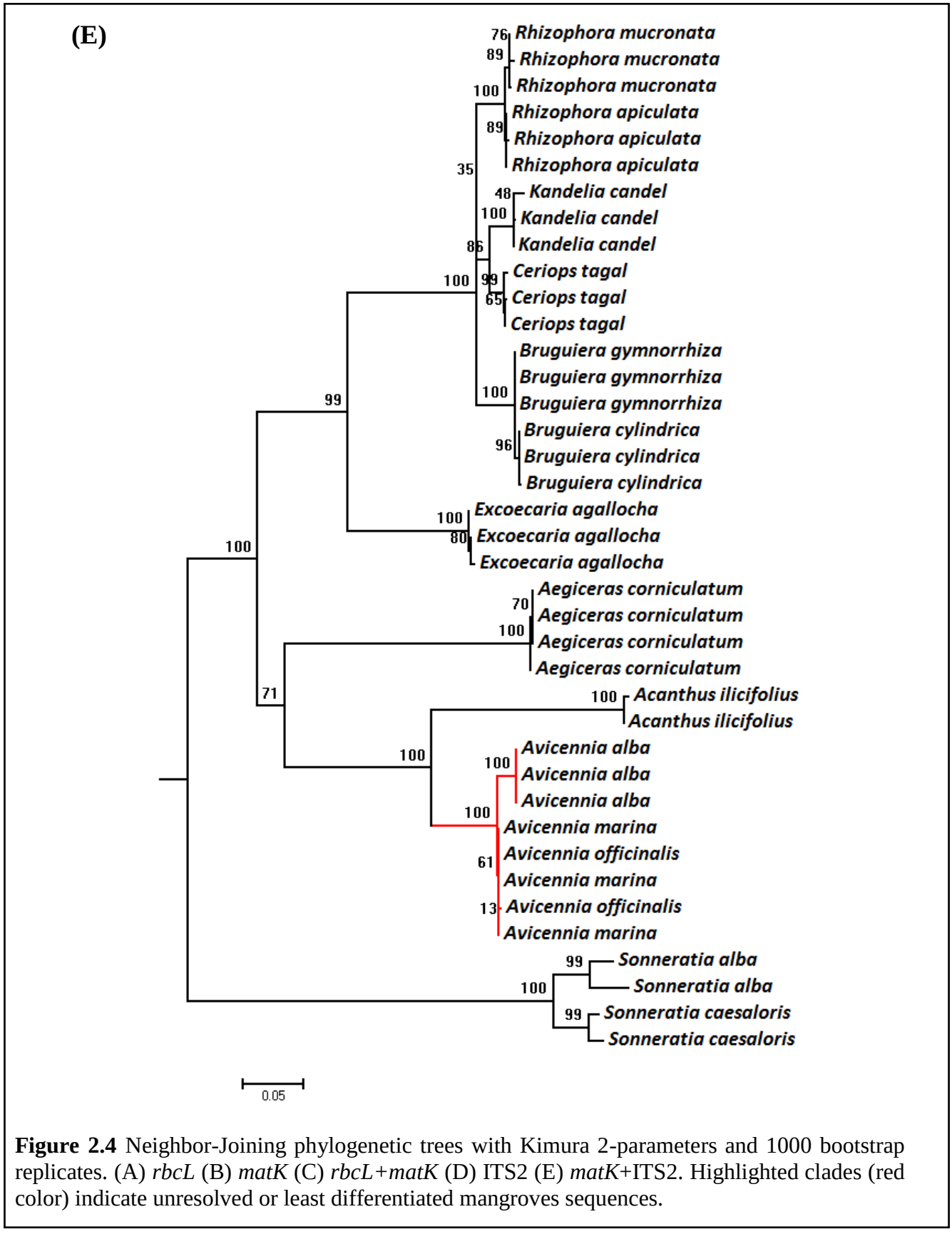


Figure 2.4 Neighbor-Joining phylogenetic trees with Kimura 2-parameters and 1000 bootstrap replicates. (A) *rbcL* (B) *matK* (C) *rbcL+matK* (D) ITS2 (E) *matK*+ITS2. Highlighted clades (red color) indicate unresolved or least differentiated mangroves sequences.

2.3.4 Diagnostic characteristics

The diagnostic character analysis determines the nucleotide or amino acid polymorphism between sets of sequences that are grouped by taxonomic or geographic labels. More specifically, this tool identifies consensus bases from each group, compares them to those from the remaining sequences in other groups and characterize how unique each base. The diagnostics characters are defined as the bases found only in one group based on the multiple sequence alignment. Partial diagnostics characters are defined as same base character appears in some but not all sequences in other groups, this base may be classified as partial characters. Diagnostic character based delineation of mangrove species was done using six barcode markers (*rbcL*, *matK*, ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1*) along with concatenated *matK*+ITS2. In the diagnostic characteristics, mangroves species with minimum 3 specimens per species were considered for analysis. For *rbcL*, 13 mangroves species were analysed and highest diagnostics characters were observed in *A. corniculatum* (15), followed by *A. ilicifolius* (8), and *E. agallocha* (7). Similarly, for *matK*, 9 mangroves species showed diagnostic characters and *Rhizophora* as well as *Bruguiera* species failed to show diagnostic characters. Interestingly, *A. corniculatum* showed 50 diagnostics characters followed by *S. caseolaris* (39), *A. ilicifolius* (34), and *E. agallocha* (22). In ITS2, highest diagnostic characters were recorded in *E. agallocha* (34) and *A. corniculatum* (35), whereas single diagnostic character was observed in the species of *Avicennia* genera followed by *B. cylindrica* (Table 2.5). In concatenated *matK*+ITS2, highest diagnostic characters were observed in *A. corniculatum* (96) and *E. agallocha* (60). However, all species of *Avicennia* genera revealed diagnostic characters, but *B. gymnorrhiza* failed to exhibit any diagnostic character. The supplementary marker *rpoC1* failed to show any diagnostic character in

Avicennia, while *atpF-atpH* and *psbK-psbI* exhibited diagnostic characters (Table 2.5). Identification of diagnostic nucleotides for each of the 14 mangrove taxa recovered from the BOLD system. Based on the nucleotide polymorphism in a sequences are referred as diagnostic characters, diagnostic or partial character, partial characters and partial or uninformative characters.

Table 2.5 Diagnostic characteristics of mangrove taxa. Identification of diagnostic nucleotides for each of the 14 mangrove taxa recovered from the BOLD system. Based on their utility for mangrove taxa delineating referred as diagnostic characters, diagnostic or partial character, partial characters and partial or uninformative characters.

Barcodes	Group name (sequences)	Diagnostic characters	Diagnostic or Partial characters	Partial characters	Un-informative characters
rbcL	*Avicennia alba (3)*	1	0	0	0
	Avicennia marina (3)	0	0	0	0
	Bruguiera cylindrica (3)	0	0	0	0
	Bruguiera gymnorrhiza (3)	0	0	0	0
	Rhizophora apiculata (3)	0	0	0	0
	Aegiceras corniculatum (6)	15	0	0	0
	Excoecaria agallocha (3)	7	0	0	0
	Kandelia candel (3)	4	1	0	0
	Ceriops tagal (3)	2	0	0	0
	Sonneratia alba (3)	0	0	1	0
	Sonneratia caseolaris (3)	0	0	0	0
	Acanthus ilicifolius (3)	8	0	0	0
	Avicennia officinalis (3)	0	0	0	0
matK	*Avicennia alba (3)*	6	0	0	1
	Avicennia marina (3)	2	0	0	1
	Bruguiera cylindrica (3)	1	0	0	0
	Bruguiera gymnorrhiza (3)	0	0	0	0
	Rhizophora mucronata (3)	0	0	0	0
	Rhizophora apiculata (3)	0	0	1	22
	Aegiceras corniculatum (6)	50	2	0	1
	Excoecaria agallocha (3)	22	1	0	1
	Kandelia candel (3)	7	0	0	0
	Ceriops tagal (3)	1	1	0	0

	Sonneratia caseolaris (3)	39	4	1	1
	Acanthus ilicifolius (3)	34	7	1	1
ITS2	*Aegiceras corniculatum (4)*	35	4	0	0
	Avicennia alba (3)	1	0	1	0
	Avicennia marina (3)	1	0	1	0
	Avicennia officinalis (3)	0	0	0	0
	Bruguiera cylindrica (3)	1	0	0	0
	Bruguiera gymnorrhiza (3)	0	0	0	0
	Ceriops tagal (3)	4	1	0	0
	Excoecaria agallocha (3)	34	2	0	1
	Kandelia candel (3)	5	0	1	1
	Rhizophora apiculata (3)	2	0	0	1
	Rhizophora mucronata (3)	6	1	0	0
matK+ ITS2	*Aegiceras corniculatum (6)*	96	3	0	1
	Avicennia alba (3)	8	0	0	1
	Avicennia marina (3)	5	0	1	1
	Bruguiera cylindrica (3)	2	0	0	0
	Bruguiera gymnorrhiza (3)	0	1	0	0
	Ceriops tagal (3)	5	2	0	0
	Excoecaria agallocha (3)	60	3	0	3
	Kandelia candel (3)	12	0	1	1
	Rhizophora apiculata (3)	2	0	1	23
	Rhizophora mucronata (3)	6	0	0	0
atpF-atpH	*Avicennia alba (3)*	0	0	0	0
	Avicennia marina (3)	4	0	0	0
	Avicennia officinalis (3)	2	0	0	0
psbK-psbI	*Avicennia marina (3)*	3	0	5	40
	Avicennia officinalis (3)	3	0	1	13
rpoC1	*Avicennia marina (3)*	0	0	1	0
	Avicennia officinalis (3)	0	0	0	0

2.3.5 Species identification and assignment based on TaxonDNA

To assess the species assignment of single region and multiple regions, we used the 'Best Match' (BM) and 'Best Closest Match' (BCM) criteria from TaxonDNA. For TaxonDNA analysis, we need to set threshold (T) below which 95% of all intraspecific distances were found. All the results above the threshold (T) were treated as 'incorrect'. Similarly, if all matches of the query sequence were below threshold (T), the barcode assignment was

considered to be correct identification. The matches of the query sequence were equally good, but correspond to a mixture of species, and then the test was treated as ambiguous identification. For the single barcode region, *matK* had the highest rate of correct identification using BM (72.09%) and BCM (39.53%) than *rbcL* with BM (47.72%), BCM (31.81%) (Table 2.6). The concatenated regions (*rbcL+matK*) demonstrated to resolve species at the level of 66.6% using BM and BCM criteria (Table 2.6). The single barcode marker ITS2 produced a moderate rate of correct identification using BM (87.8%) and BCM (75.6%) than the concatenated *matK*+ITS2 using BM (89.74%), and BCM (84.61%) (Table 2.6). ITS2 barcode produced 13 clusters at 3% threshold, of which 5 species (*A. corniculatum, A. ilicifolius, E. agallocha, K. candel* and *C. tagal*) were the perfect match. Whereas, *Avicennia, Rhizophora* and *Bruguiera* species were clumped into 3 clusters, while *S. alba* and *S. caseolaris* were split into 5 clusters. As compared to single barcode marker (ITS2), concatenated (*matK*+ITS2) markers at 3% threshold produced 11 clusters, where *S. caseolaris* was successfully resolved. Single barcode *atpF-atpH* demonstrated 100% correct identification in both BM and BCM method for *Avicennia* genera with 3 clusters. *psbK-psbI* locus identified 50% *Avicennia* species in BM and BCM methods, however, *rpoC1* showed lowest identification rate of about 33.33% (Table 2.6).

Table 2.6 TaxonDNA analysis. Identification success rates using TaxonDNA (Species Identifier) program under 'Best Match' and 'Best Closest Match' methods. N-Number of sequences, C-correct, A-ambiguous, Inc-Incorrect, T-Threshold value, M-match, MM-mismatch

Barcode	N	Best Match (%)			Best Closest match (%)				T (%)	No of Cluster	M/ MM
		C	A	Inc	C	A	Inc	NM			
RbcL	44	47.7	36.36	15.9	31.8	27.27	11.3	13	0	23	6/8
matK	43	72.0	25.58	2.32	39.5	13.95	2.32	44.18	0.11	24	10/4
rbcL+ matK	42	66.6	16.6	16.6	66.6	16.66	16.6	0	0.2	21	8/6
ITS2	40	87.8	2.43	9.75	75.6	2.43	9.75	12.19	3	14	10/4
ITS2 + *matK*	39	89.7	2.56	7.6	84.6	2.56	7.6	5.12	3	11	6/8
atpF-atpH	9	100	0	0	100	0	0	0	0.3	3	3/0
psbK-psbI	6	50	0	50	50	0	50	0	0.8	4	1/1
rpoC1	6	33.3	66.66	0	33.3	66.6	0	0	3	1	0/2

2.3.6 Species identification based on Automated Barcode Gap Discovery tool (ABGD)

ABGD tool clustered the given sequences into various hypothetical species based on the three different distance matrices such as JC, K2P and p-distance. Initially, it finds the barcode gap in the sequences and further uses it to partition the data. All the obtained mangroves sequences of *rbcL*, *matK*, ITS2, *atpF-atpH*, *rpoC1* and *psbK-psbI* were analyzed using ABGD tool with three distance matrix (JC69, K2P and p-distance) and two relative gap width (X=1 and X=1.5). The input sequences were grouped into initial and recursive partition based on the distance matrices and relative gap width. For *rbcL*, based on all three distance metrics (JC69, K2P and p-distance) at X=1 and prior intraspecific distance 0.001, gave 9 partition for an initial partition. However, at X=1.5, all distance metrics gave only 4

partitions for *P*=0.001. Similarly, for *matK*, three distances metrics produced minimum 8 initial partition (P=0.001) at X=1 and 6 initial partition at X=1.5. But recursive partition reached up to 15 partitions for JC69 and K2P. Surprisingly, concatenated *rbcL+matK* produced the same number of partition as produced by *rbcL* and *matK*.

The initial partition of ITS2, K2P with X=1.0, prior maximal distance *P*=0.021 produced consistent 12 operational taxonomic units (OTUs). *S. alba* was split into 3 groups, while members of *Rhizophora* and *Avicennia* were merged (Table 2.7). Whereas, recursive partitioning with *P*=0.00167, produced inconsistently 18 OTUs, of which *A. alba, A. officinalis*, and *B. cylindrica* showed split, while *B. Gymnorrhiza* was clustered perfectly (Table 2.7). In concatenated *matK*+ITS2, at X=1.0 for all three metrics, OTUs ranged from 4-11 in the initial partition, but recursive partition tends to exhibit inconsistent OTUs (Table 2.7). When relative gap width was increased from X=1.0 to X=1.5, suddenly OTUs in ITS2 for initial partition was dropped to maximum 7, while recursive partition showed an increase in OTUs, upto 16 at *P*=0.001. The initial partition for *matK*+ITS2, with X=1, *P*=0.0129 produced 11 OTUs. *Avicennia* and *Bruguiera* members were merged, while *S. alba* showed split. In recursive partition, with *P*=0.001, *A. alba*, *B. cylindrica*, *B. gymnorrhiza* were resolved perfectly, while *A. officinalis*, *A. marina* along with *R. apiculata* and *R. mucronata* remained merged.

The initial partition with an *atpF-atpH* barcode, JC and K2P metrics with (X=1, 1.5) showed 3 OTUs (*P*=0.0027) without any recursive partition except (X=1.5, *P*=0.00278, 1 OTU). With *atpF-atpH*, at X=1.5 initial partition with *P*=0.00278, 3 OTUs were produced in *A. alba*, *A. officinalis*, and *A. marina.* Similarly, *psbK-psbI* showed 4 OTUs (*P*=0.001) in an initial partition for JC and K2P metrics at X=1 and p-distance had only 2 OTUs with 1 OTU

in the recursive partition. At X=1.5, only JC and p-distance were able to partition data. JC the initial partition at *P*=0.001 produced 4 OTUs, while at *P*=0.0046, produced 2 OTUs. Metrics p-distance predicted 2 OTUs in an initial partition and 1 OTU in the recursive partition. Barcode locus *rpoC1* at X=1 with JC and K2P metrics showed initial partition of 2 OTUs and there cursive partition at *P*=0.00278 predicted 1 OTU.

Table 2.7 Automated Barcode Gap Discovery (ABGD) web server based analysis of all barcodes (*rbcL*, *matK*, *rbcL+matK*, ITS2, *matK*+ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* using two relative gap width (X=1 and 1.5) and three different matrices such as JC, K2P and p-simple distance. IP- Initial Partition, RP- Recursive Partition

Barcode	Relative gap width	Prior intraspecific distance	Jukes-Cantor (JC69)		Kimura -2 parameter		p-distance	
			IP	RP	IP	RP	IP	RP
rbcL	X=1.0	0.359	0	1	0	1	0	1
		0.021544	4	5	4	5	4	0
		0.01291	4	5	4	5	4	0
		0.007743	4	5	4	5	4	6
		0.004642	9	0	9	0	9	0
		0.002783	9	14	9	14	9	0
		0.001668	9	16	9	15	9	0
		0.001000	9	16	9	15	9	0
	X=1.5	0.359	0	1	0	1	0	1
		0.021544	4	5	4	5	4	0
		0.01291	4	5	4	5	4	0
		0.007743	4	5	4	5	4	6
		0.004642	4	11	4	9	4	9
		0.002783	4	14	4	14	4	9
		0.001668	4	17	4	15	4	9
		0.001000	4	17	4	15	4	9
matK	X=1.0	0.1	4	5	4	5	4	5
		0.059948	4	5	4	5	4	5
		0.035938	6	6	6	6	6	6
		0.021544	6	6	6	6	6	6
		0.012915	8	9	8	9	8	8
		0.007743	8	10	8	10	8	9
		0.004642	8	11	8	10	8	10
		0.002783	8	11	8	11	8	10
		0.001668	8	11	8	11	8	10
		0.001	8	15	8	15	8	10
	X=1.5	0.1	0	1	0	0	0	1

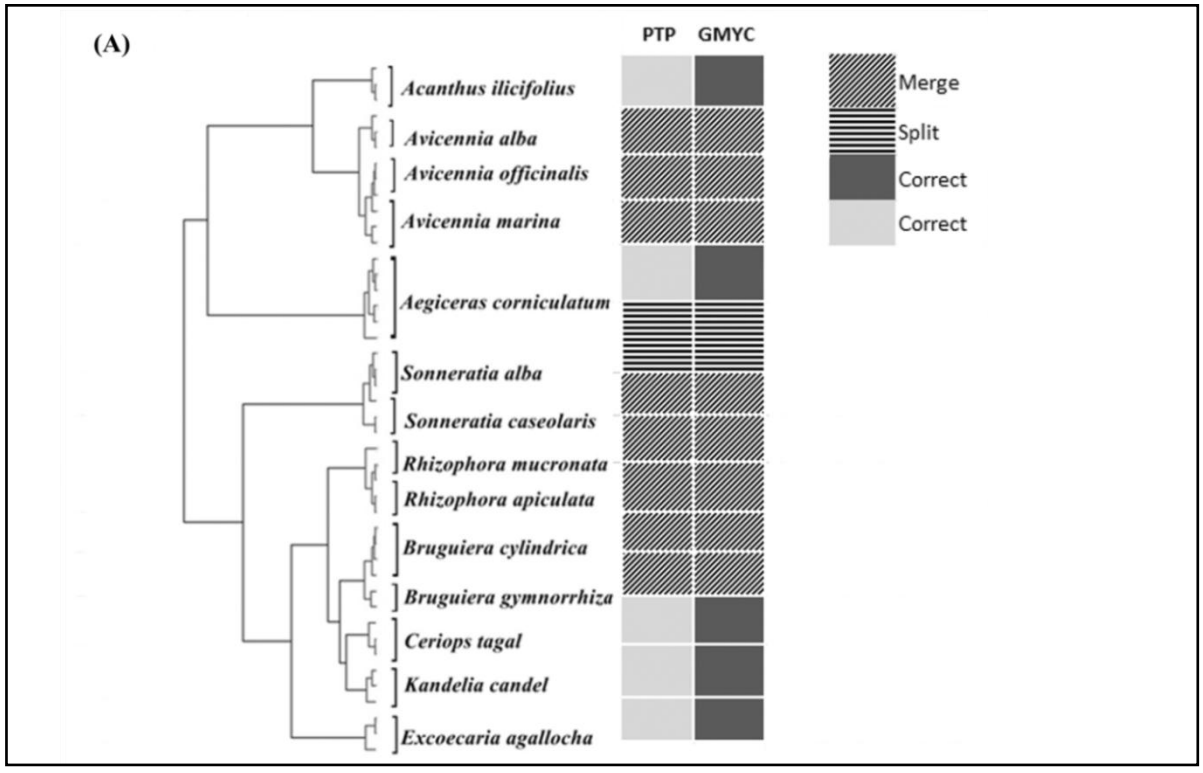
		0.059948	4	5	0	1	4	5
		0.035938	6	6	6	6	6	6
		0.021544	6	6	6	6	6	6
		0.012915	6	7	6	7	6	6
		0.007743	6	10	6	10	6	9
		0.004642	6	11	6	10	6	10
		0.002783	6	11	6	11	6	10
		0.001668	6	11	6	11	6	10
		0.001	6	25	6	16	6	12
rbcL+ matK	X=1.0	0.1	0	1	0	1	0	1
		0.059948	4	5	4	5	4	5
		0.035938	4	5	4	5	4	5
		0.021544	6	0	6	0	6	0
		0.012915	9	0	9	0	9	0
		0.007743	9	10	9	10	9	0
		0.004642	9	10	9	10	9	0
		0.002783	9	10	9	10	9	10
		0.001668	9	10	9	10	9	10
		0.001	9	13	9	12	9	10
	X=1.5	0.059948	0	1	0	1	0	1
		0.035938	4	5	4	5	4	5
		0.021544	6	0	6	0	6	
		0.012915	6	0	6	0	6	
		0.007743	6	10	6	7	6	8
		0.004642	6	10	6	10	6	9
		0.002783	6	10	6	10	6	10
		0.001668	6	10	6	10	6	10
		0.001	6	11	6	12	6	10
ITS2	X=1.0	0.1	2	2	2	2	2	2
		0.059948	7	7	7	7	7	7
		0.035938	13	13	7	7	7	7
		0.021544	13	13	12	12	13	13
		0.012915	13	13	12	12	13	13
		0.007743	13	13	12	12	13	13
		0.004642	13	14	12	13	13	14
		0.002783	13	14	12	13	13	14
		0.001668	13	19	12	18	13	14
		0.001	13	19	12	18	13	14
	X=1.5	0.1	2	2	2	2	2	2
		0.059948	7	7	7	7	7	7
		0.035938	7	10	7	7	7	7
		0.021544	7	10	7	10	7	10
		0.012915	7	10	7	10	7	10
		0.007743	7	10	7	10	7	10
		0.004642	7	11	7	11	7	11
		0.002783	7	11	7	11	7	11
		0.001668	7	16	7	16	7	11

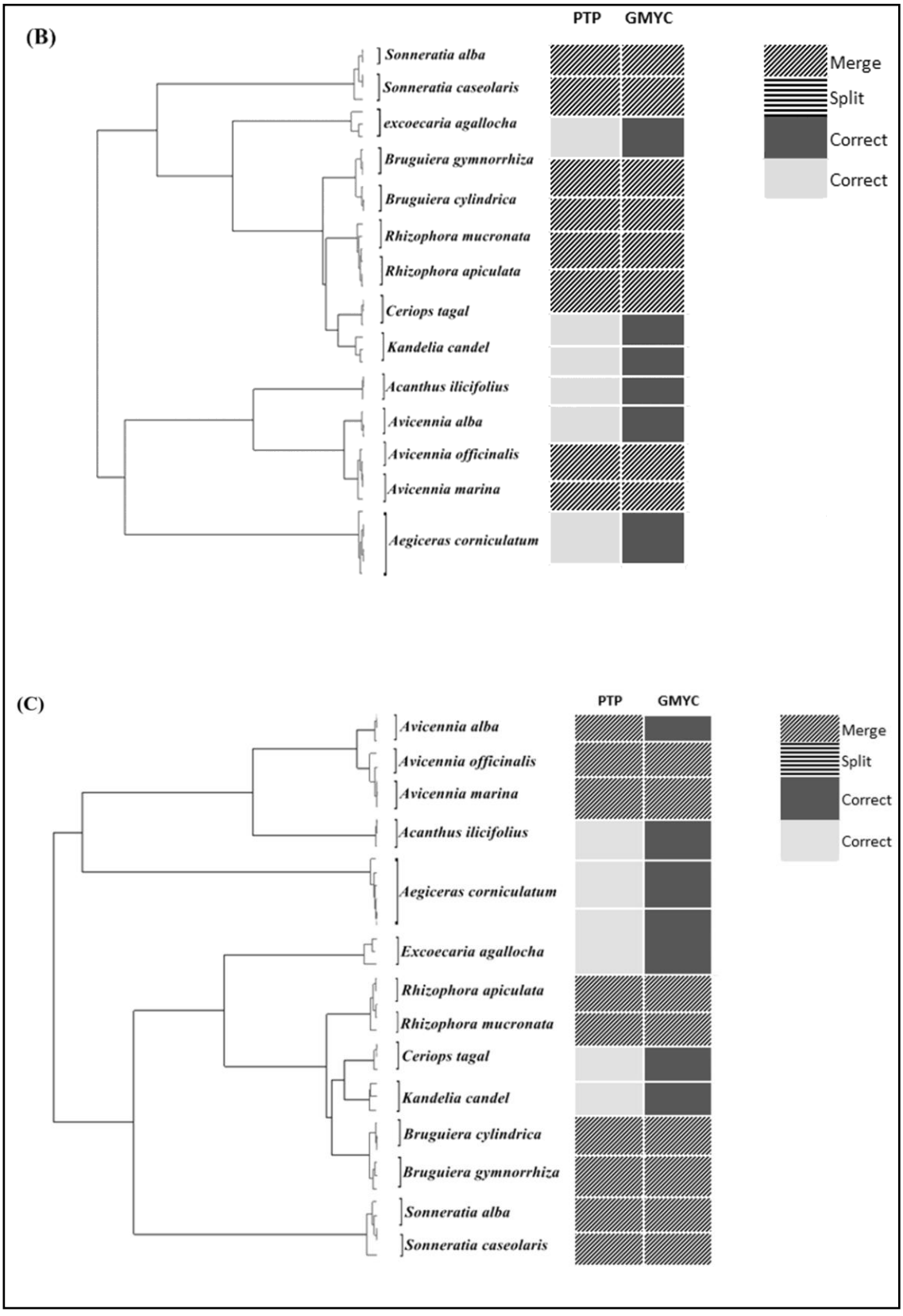
		0.001	7	16	7	16	7	11
matK+ ITS2	X=1.0	0.1		1		1		1
		0.059948	4	5	4	5	4	5
		0.035938	6		6		6	
		0.021544	6		6		6	
		0.012915	11		11		11	
		0.007743	11		11		11	
		0.004642	11	12	11		11	
		0.002783	11	12	11		11	
		0.001668	11	13	11		11	
		0.001	11	13	11		11	
matK+ ITS2	X=1.5	0.1		1		1		1
		0.059948	4	5	4	5	4	5
		0.035938	6		6		6	
		0.021544	6		6		6	
		0.012915	6	8	6	8	6	8
		0.007743	6	9	6	9	6	8
		0.004642	6	10	6	10	6	9
		0.002783	6	10	6	10	6	10
		0.001668	6	11	6	10	6	11
		0.001	6	11	6	11	6	11
atpF-atpH	X=1.0	0.0046	1		1			
		0.002783	3		3			
		0.001668	3		3			
		0.001	3		3			
	X=1.5	0.0046		1		1		
		0.002783	3		3			
		0.001668	3		3			
		0.001	3		3			
psbK-psbI	X=1.0	0.0129	0	1			0	1
		0.0077	2				2	
		0.0046	2		1		2	
		0.002783	4		4		2	
		0.001668	4		4		2	
		0.001	4		4		2	
	X=1.5	0.0129	0	1			0	1
		0.0077	2				2	
		0.0046	2				2	
		0.002783	4				2	
		0.001668	4				2	
		0.001	4				2	
rpoC1	X=1.0	0.002783		1		1		
		0.001668	2		2			
		0.001	2		2			

2.3.7 Species identification and assignment based on GMYC and PTP methods

Mangroves species were delimiting based on the coalescent theory such as generalized mixed Yule-Coalescent (GMYC) and Poisson Tree Processes model (PTP). For *rbcL* GMYC analysis, single threshold time of -0.005375214 with confidence interval of 7-12 exhibited for 9 Maximum Likelihood (ML) clusters and entities respectively (Maximum Likelihood ratio: 8.022772) (Figure 2.5 A). For *matK* single threshold time of -0.007376445 with confidence interval of 2-10 exhibited for 9 ML clusters and 2-13 for 9 ML entities (Likelihood ratio: 8.022772) (Figure 2.5 B). For concatenated *rbcL+matK* barcode, single threshold time of -0.006304029 with confidence interval of 2-10 exhibited for 10 ML clusters and 2-12 for 10 ML entities (Likelihood ratio: 9.957527) (Figure 2.5 C). The single threshold GMYC (sGMYC) model generated through BEAST using the ultrametric phylogenetic tree resulted in an identification of 9 Maximum Likelihood (ML) clusters for ITS2 with confidence interval (CI) of 4-9 and 14 ML entities with CI of 4-18 (Threshold time: -0.013035). Similarly, with *matK*+ITS2, 10 ML clusters with CI of 4-10 and 14 ML entities with CI of 4-16 (Threshold time: -0.005793) were identified. The resulting ML entities in ITS2 exhibited 5 species merged in 2 OTUs, while in *matK*+ITS2 only 4 species were merged with exception of *A. alba*. Also, splitting of two species (*S. alba* and *S. caseolaris*) formed additional OTUs (Figure 2.5 D and E). The multiple threshold methods (mGMYC) gave two threshold time for ITS2 (-0.013035 and -0.005441) resulting into 9 clusters (CI: 4-9) and 17 ML entities (CI: 4-17). *matK*+ITS2 gave threshold time of -0.010859 and -0.004847, resulting into 9 clusters (CI: 5-11) and 13 ML entities (CI: 5-16). However, multiple thresholds overestimated the number of species, so we took a more conservative approach to consider only the results obtained from the single threshold (sGMYC) method.

In GMYC, apart from other metrics, three unresolved species *R. apiculata, R. mucronata* and *A. alba* were distinctly resolved. In addition to the above methods used for taxonomic evaluation, maximum likelihood (ML) based approach was added to get an additional perspective towards the species delineation through PTP. Similarly, individual *rbcL* barcode PTP analysis formed 9 clusters of 14 mangroves species. The *Avicennia, Bruguiera, Rhizophora* and *Sonneratia* genus were clumped together and failed to discriminate at species level. In PTP analysis, single *matK* barcode formed 9 ML partition and *Avicennia, Bruguiera, Rhizophora* and *Sonneratia* genus failed to discriminate. PTP analysis for concatenated *rbcL+matK* could form 9 OUT's which discriminate only 5 species and failed to discriminate 4 genera. The ML analysis exhibited 10 OTUs with ITS2, where *Avicennia, Bruguiera, Rhizophora, Ceriops,* and *Kandelia* genera were merged while *S. alba* and *S. caseolaris* were split. With *matK*+ITS2, OTUs were formed by merging of *Avicennia, Bruguiera* and *Rhizophora* genera and *S. alba* was splitted.




(B)
PTP
GMYC
Sonneratia alba
Sonneratia caseolaris
excoecaria agallocha
Bruguiera gymnorrhiza
Bruguiera cylindrica
Rhizophora mucronata
Rhizophora apiculata
Ceriops tagal
Kandelia candel
Acanthus ilicifolius
Avicennia alba
Avicennia officinalis
Avicennia marina
Aegiceras corniculatum
Merge
Split
Correct
Correct
(C)
PTP
GMYC
Avicennia alba
Avicennia officinalis
Avicennia marina
Acanthus ilicifolius
Aegiceras corniculatum
Excoecaria agallocha
Rhizophora apiculata
Rhizophora mucronata
Ceriops tagal
Kandelia candel
Bruguiera cylindrica
Bruguiera gymnorrhiza
Sonneratia alba
Sonneratia caseolaris
Merge
Split
Correct
Correct

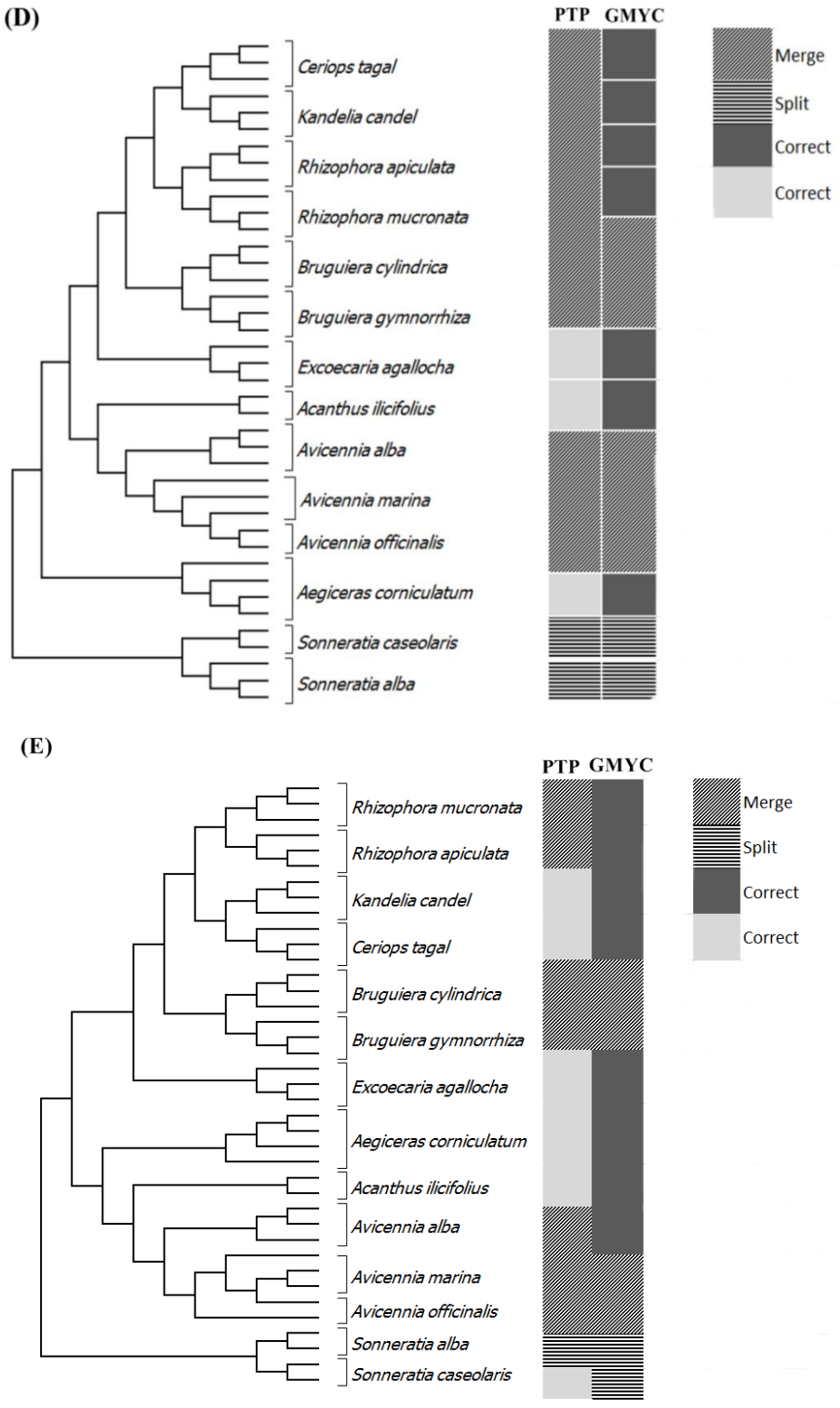


Figure 2.5 Bayesian phylogenetic trees. Bayesian phylogenetic trees were represented with GMYC and PTP methods which categorized correct identification, ambiguous as a merge and spit species. (A) *rbcL* (B) *matK* (C) *rbcL*+*matK* (D) ITS2 and (E) *matK*+ITS2 gene. Vertical boxes on the right indicate the clades detected by the coalescent-based PTP and GMYC methods.

2.4 Discussion

To the best of our knowledge, current study is the first attempt of performing DNA barcoding based assessment of mangroves from Goa using plastid as well as nuclear markers such as *rbcL, matK,* ITS2, *rpoC1*, *atpF-atpH*, and *psbK-psbI*. Some countable reports based on molecular taxonomy and phylogeny of Indian mangroves are available using nuclear, mitochondrial, and plastid markers (ITS, *rbcL*, RFLP, RAPD, PCR-RAPD and AFLP) (Parani et al., 1997a, b; Lakshmi et al., 1997, 2000; Setoguchi et al., 1999; Schwarzbach and Ricklefs, 2000). Besides this there are many reports on mangroves identification based on morphological characters (Untawale, 1985; Tomlinson, 1986; Untawale and Jagtap, 1992). There is no consensus regarding perfect plant DNA barcode, however few of plastid and nuclear coding (*rbcL, matK, rpoB,* and *rpoC1*) and non-coding (*trnH-psbA,* ITS2, *psbK-psbI* and *atpF-atpH*) marker fulfilled the required criteria (Fazekas et al., 2008; Kress et al., 2005; Pennisi, 2007). The *rbcL* and *matK* are considered as core barcode, which can be further complemented with *trnH-psbA* and ITS2 as plant barcode (CBOL, 2009; China Plant BOL, 2011). We employed these markers for molecular identification of mangrove plant species. We have tested potential barcode candidate's *rbcL, matK* and ITS2 individual as well as concatenated *rbcL+matK,* which demarcated all mangrove species except *Avicennia* species. Further, *Avicennia* species such as *A. alba, A. marina,* and *A. officinalis* were discriminated based on *atpF-atpH* marker.

An analysis was performed based on traditional (Barcode gap analysis and NJ tree with the K2P method) as well as advanced barcode methods (ABGD, TaxonDNA, GYMC and PTP). Individual, as well as concatenated *rbcL* and *matK* barcode demarcated almost all mangroves species except for *Rhizophora*, *Sonneratia* and *Avicennia* genera. The CBOL

plant working group (2009) reported that only 72% species were resolved using combined *rbcL* and *matK*. A similar result was observed after combining *rbcL* and *matK* from closely related species of *Curcuma* (Chen et al., 2015). Moreover, *Avicennia* genera with three species, of which *A. alba*, was resolved appropriately using *matK*, but *A. officinalis* and *A. marina* lumped together and unable to resolve at the species level. Low resolution using DNA barcode regions has been documented in many other plants such as the genus *Araucaria* (32%), *Solidago* (17%) and *Quercus* (0%) (Pennisi, 2007).

A high percentage of bidirectional reads were critical for a successful plant barcoding system, given the low amount of variation that separates many plant species (Fazekas et al., 2008; CBOL, 2009). The risk of misassignment can be anticipated due to sequencing error or incomplete bidirectional reads. We observed the significant quality of PCR amplification and sequencing ranged from 95% to 100% in all tested markers. However, for ITS2 barcode, we performed many amplifications and sequencing attempt for *S. alba*, *S. caseolaris*, and *A. ilicifolius* mangroves taxa. Sequencing of *S. alba* and *S. caseolaris* resulted in a mixed and low-quality chromatogram with unidirectional success. The possible explanation for this kind of situation can be underscored by the presence of either ITS as multiple copies or pseudogene or/and fungal ITS contamination in plant (Alvarez et al., 2003). Species identification success rate using *rbcL+matK* is higher, where as *rbcL* sequence recovery ranged from 90-100% (Little and Stevenson, 2007; Ross et al., 2008; CBOL, 2009). Hence, CBOL plant working group recommends *rbcL* primers to possess universality for land plants. As reported by CBOL, the *matK* region showed sequencing success of 90% (CBOL, 2009). The *matK* marker provided 88% sequencing success, with the use of 10 primer pair combinations (Fazekas et al., 2008).

Very few reports are available on the DNA barcoding of the mangrove taxa (Sahu et al., 2016). Lower genetic distances were observed based on K2P among mangrove taxa particularly genera *Rhizophora*, *Sonneratia, Avicennia*, and *Bruguiera* based on *rbcL*, *matK* and ITS barcode (Sahu et al., 2016). Genetic distance ranged from 0.01 to 0.25 for *rbcL* gene, 0.01 to 0.89 for *matK* and 0.01-0.508 for ITS locus (Sahu et al., 2016). Similar results were observed in our studies, for *rbcL* and *matK* the genetic distance ranged from 0-0.68% and 0-1.32% respectively. The discrimination power of proposed DNA barcode by the CBOL Plant working group may vary in different plant group (Hollingsworth et al., 2009; Li et al., 2015; Vinitha et al., 2014). Depending on the taxon, the use of additional markers may be needed for discrimination (CBOL, 2009).

For single barcode ITS2, ABGD (K2P, X=1), TaxonDNA (T=3%) and GMYC produced consistent OTUs with corresponding results. Additionally, GMYC resolved *R. apiculata*, *R. mucronata,* and *A. alba* species. Overall highest taxon assignment was observed as 57.14% in GMYC and taxon resolution was upto 42.85% in ABGD, TaxonDNA, and PTP barcoding methods. However, the resolution of Chlorella-like species (microalgae) produced by GMYC, PTP, ABGD and character-based barcoding methods were variables based on several marker studies such as *rbcL,* ITS, and *tufA* (Zou et al., 2016). Single ITS2 with PTP analysis was not able to resolve *C. tagal* and *K. candel*, which was further improved in the *matK*+ITS2 analysis. Analysis following the above methods, species delimitation through PTP and GMYC was utilized, due to their robustness in the absence of barcoding gap (Tang et al., 2014). Even though they are based on different algorithms, both methods calculated the point of transition between species and population (Zou et al., 2016). The GMYC method has a theoretically strong background and requires ultrametric gene tree that takes more time to

analyse data. In contrast, the PTP is a recently developed method as an alternative to GMYC which requires non-ultrametric gene tree and consumes less time (Tang et al., 2014; Dumas et al., 2015). Both the methods revealed sort of identical results, however, the two analyses differed in resolution. In both the methods, five species (*B. cylindrica*, *B. gymnorrhiza*, *A. officinalis* and *A. marina*) in GMYC and seven species (*B. cylindrica*, *B. gymnorrhiza*, *A. alba*, *A. officinalis*, *A. marina*, *R. apiculata* and *R. mucronata*) in PTP were merged into single OTUs, potentially indicating low intraspecific diversity. It reflected that there are many overlooked/cryptic species present within the mangroves. When we performed ABGD with relative gap width X=1.5 for K2P method, *S. alba*, and *S. caseolaris* species were demarcated, while rest of the mangrove species were split. At a relative gap width (X=1) about seven species of the mangroves were merged into single OTU and observed that the ABGD tends to lump species by increasing the number of merged OTUs (Yang et al., 2016). Beside this, we also observed inconsistency of OTUs count during initial partition to recursive partition. Recursive partitioning recognizes more OTUs than initial ones, showing their superior capability to deal with variation in sample sizes of the species under study (Yang et al., 2016). Moreover, TaxonDNA with a lower threshold value (0.3%) demarcated *B. cylindrica* and *B. gymnorrhiza*. The possible explanation for this might be due to lack of barcode gap resulting in merged OTUs, which can be optimized by analyzing more than 5 sequences per species, and we have used 3 sequences per species (Puillandre et al., 2012). In TaxonDNA analysis, for *rbcL* threshold (T) was observed 0%, a similar result was recorded for *rbcL* in the Zingiberaceae family (Li et al., 2016). However, the threshold (T) for Indian Zingiberaceae family members was recorded as 0.20% for *rbcL* and 0% for *rpoB* and *accD* (Vinitha et al., 2014).

Avicennia is the most diverse mangrove genus, comprising eight species, out of which three are endemic to the Atlantic-East Pacific (AEP) region and five are endemic in the Indo-West Pacific region (IWP) (Li et al., 2016). A recent systematic revision of *Avicennia* based on morphological characters formed three groups: (1) *A. marina*; (2) *A. officinalis* and *A. integra*; and (3) *A. rumphiana* and *A. alba* (Li et al., 2016). In the current study, we have included *A. marina*, *A. officinalis*, and *A. alba* species, which were resolved with other barcode markers. Two plastid spacers such as *psbK-psbI* and *atpF-atpH* are recommended as potential plant DNA barcodes based on the flora of the Kruger National Park South Africa as a model system (Lahaye et al., 2008). Similarly, we used three markers (*atpF-atpH, psbK-psbI* and *rpoC1*) for cryptic genera *Avicennia* and further evaluated with ABGD and TaxonDNA barcode methods. Both the methods consistently resolved all three *Avicennia* species using an *atpF-atpH* marker. Similarly, phylogenetic reconstruction of *Avicennia* genera based on *trnT-trnD* intergenic spacer region and the *psbA* gene revealed that *A. marina* is sister to the *A. officinalis*/*A. integra* and *A. alba* is genetically distinct (Li et al., 2016).

2.5 Conclusions

In the present study, we tested core DNA barcode *rbcL, matK,* ITS2, and their combinations. Moreover, for *Avicennia* species three additional markers such as *atpF-atpH, psbK-psbI* and *rpoC1* were used to resolve species. Individual, as well as concatenated *matK*+ITS2 are helpful to demarcate mangroves at the species level. Single barcode *matK* is sufficient to resolve *A. ilicifolius, A. corniculatum, E. agallocha, Ceriops tagal, K. candel, B. cylindrica and B. gymnorrhiza.* ITS2 was able to discriminate *R. apiculata* and *R. mucronata* species

based on GMYC method, while *A. alba* was resolved by concatenation of *matK*+ITS2. A cryptic genus *Avicennia* was delimitated based on the *atpF-atpH* single barcode. In the present work, the foundation work was done towards DNA barcoding of mangroves plant genera, such as *Rhizophora*, *Avicennia*, *Acanthus*, *Kandelia*, *Ceriops*, *Bruguiera*, *Aegiceras* and *Excoecaria*. Compiled mangroves barcoding result had some limitations, most of which are due to the low mangrove taxa sample coverage. Further, there is a need to explore additional mangroves taxa from other geographical sources which will improve mangrove species identification for practical conservation.

CHAPTER 3

Selection of reference genes for qRT-PCR

CHAPTER 3

Selection and validation of reference genes for qRT-PCR analysis in *Rhizophora apiculata*

3.1 Introduction

Several techniques are available to investigate gene expression analysis including, semi-quantitative reverse transcription polymerase chain reaction (RT-PCR), northern blotting, *in-situ* hybridization, and quantitative real-time PCR (qRT-PCR). The qRT-PCR is a reliable, sensitive, and wide quantification range gene expression analysis technique (Bustin, 2002). Moreover, reference gene for qRT-PCR normalization is not universally standardized and it varies according to plant tissue material and experimental conditions (Bustin et al., 2009). For precise quantification and reproducible profiling, selection and validation of stable candidate reference genes are crucial steps prior to qRT-PCR for data normalization. Some commonly used reference genes include Actin (*ACT*), β-tubulin (*β-TUB*), Ubiquitin (*UBQ*), Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), elongation factor 1α (*EF1α*), and 18S ribosomal RNA (18S) are preferred to normalize the expression profiles of candidate reference genes. The reference genes are involved in basic cellular functions, maintaining cell size and shape, and cellular metabolism (Bustin, 2002). However, several reports have shown that the level of reference genes expression varies in different cultivars, tissues, and stress conditions (Sinha et al., 2015; Reddy et al., 2015; Nikalje et al., 2018). Hence, it is very important to select and validate most appropriate reference genes involved in various experimental conditions before proceeding to gene expression analysis. Various web-based tools and algorithms are available to address validation of candidate reference genes

including, comparative ΔCt (cycle thresholds) (Silver et al., 2006), NormFinder (Andersen et al., 2004), BestKeeper (Pfaff et al., 2004), and geNorm algorithm (Vandesompele et al., 2002). RefFinder is a web-based program which provides a comprehensive ranking of reference genes (Xie et al., 2012). Based on the literature survey, there were no reports available on evaluation of candidate reference genes for qRT-PCR in *R. apiculata*. In the present study, we aim to evaluate the most stable candidate reference gene for qRT-PCR gene expression analysis in *R. apiculata* physiological tissues and in salt-stressed leaf samples. The current study will promote the gene expression analysis in the *R. apiculata*, especially when studied under salinity stress.

3.2 Material and Methods

3.2.1 Plant materials

In the present study, we collected three month old *R. apiculata* seedlings located in the West Coast of India with the geographical latitude of 15.5256° N and longitude of 73.8753° E, with the permission from the Principal Chief Conservator of Forest, Goa Forest Department, Goa, India. Mangrove species identification was performed based on morphological characteristics using a comparative guide to the mangroves of Goa (Naskar and Mandal, 1999). All seedlings were acclimatized and maintained in half-strength Hoagland solution at a temperature regime of 24-30°C, 40-50% relative humidity. Various physiological tissues such as leaves, stems, roots and flower samples were collected. To imitate salt stress conditions, young seedlings of *R. apiculata* were exposed to Hoagland nutrient solution supplemented with 250 mM sodium chloride (NaCl) continuously and leaf samples were harvested at different time-course such as 0, 6, 12 and 24 h.

3.2.2 RNA isolation and cDNA synthesis

Total RNA was extracted using modified cetyl-trimethyl ammonium bromide (CTAB) protocol with 2% polyvinyl pyrrolidone (PVP-30) and 10% β-mercaptoethanol (Fu et al., 2004). Freshly collected tissues were immediately pulverized into 2 ml of pre-warmed CTAB buffer and incubated at 60°C for 30 m. The suspension was gently mixed and centrifuged at 14,000 rpm for 10 m at room temperature with an equal volume of chloroform: isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and RNA was precipitated with a $1/3^{rd}$ volume of 8 M lithium chloride (LiCl) and incubated at -20°C for 1 h followed by adding an equal volume of chilled isopropanol (-20°C). The RNA was precipitated by centrifugation at 14,000 rpm for 10 m at room temperature followed by washing with 70% ethanol. RNA was finally dissolved in 0.1% DEPC treated water and its quantity and quality were confirmed by spectrophotometer (Nanodrop, Thermo Fisher Scientific, Waltham, MA, USA).

Genomic DNA contamination was removed by DNase I enzyme (Thermo Fisher Scientific, Waltham, MA, USA) treatment at 37°C for 30 m and heat inactivated at 65°C for 10 m with 50 mM EDTA. The cDNA synthesis was performed in 20 µl reaction volume using the RevertAid Reverse Transcriptase (Thermo Fisher Scientific, Watham, MA, USA), 0.1-5 mg RNA sample and oligo $(dT)_{18}$ primer, as per manufacturer's instructions .

3.2.3 Selection of reference genes and primer designing

Nine housekeeping genes such as *ACT*, *α-TUB*, *β-TUB*, *GAPDH*, *UBQ*, 18S rRNA, *rbcL*, *Histone H3*, and *EF1α* used in qRT-PCR along with one target gene sodium/proton antiporter (NHX) were selected. There is no genome sequence available publicly for *R. apiculata*

hence, homologous candidate reference gene sequences were retrieved from the model plants such as *Arabidopsis thaliana* and *Oryza sativa* from Gramene and NCBI databases. A full-length gene sequences were used for primer designing using PrimerQuest (Integrated DNA Technologies) with given parameters: melting temperature (T_m) of 55-65°C, primer length of 17-25 bp, and amplicon length of 100-500 bp (Table 3.1). The amplicon was sequenced and annotated based on the sequence similarity-based search tool. Further, all the confirmed sequences were submitted to the GenBank for accession numbers. After primer specificity analysis *α-Tub* and *Histone H3* were removed from further analysis. The primer sequences, accession numbers, and their efficiency were given in Table 3.1. For qRT-PCR, primer specificity was determined using melting curve analysis and the PCR products were checked on 2% agarose gel. The primer efficiency of all candidate reference genes was calculated based on the standard curve generated from a 10-fold serial dilution of cDNA (10^0, 10^{-1}, 10^{-2}, and 10^{-3}) and regression coefficient (R^2) values. Primer efficiency was calculated using the given formula [$E = (10^{(-1/slope)}-1) \times 100$], where E = 2 and corresponds to 100% efficiency; high/acceptable amplification efficiency equals 90-110% (Sinha et al., 2015).

Table 3.1 Details of candidate reference genes, Accession number, primer sequences, amplicon size, PCR efficiency (%) and regression coefficient (R^2) for each candidate reference gene selected in this study.

	Gene	Accession no.	Primer sequence (5'-3')	Size (bp)	PCR efficiency (%)	R^2
1	18S	MH277331	F-CCGTCCTAGTCTCAACCATAAAC R-GCTCTCAGTCTGTCAATCCTTG	189	102.30	1.00
2	*ACT*	MH279969	F-ATCACACCTTCTACAACGAGC R-CAGAGTCCAACACGATACCAG	207	92.03	0.99
3	*EF1α*	MH310424	F-AGCGTGTGATTGAGAGGTTC R-AGATACCAGCCTCAAAACCAC	53	98.60	0.99
4	*UBQ*	MH310425	F-CACTTCGACCGCCACTAC R-AGGGCATCACAATCTTCACAG	60	90.54	0.99
5	*RbcL*	KP697362	F-ATGTCACCACAAACAGAGACTAAAGC R-GTAAAATCAAGTCCACCRCG	530	97.69	0.99
6	*β-TUB*	MH310423	F-ACCTCCATCCAGGAGATGTT R-GTGAACTCCATCTCGTCCATTC	60	94.08	0.99
7	*GAPDH*	MH279970	F-ACCACAGTCCATGCCATCAC R-TCCACCACCCTGTTGCTGTA	264	96.78	0.99
8	*NHX*	KU525079	F-TGCTAGCTCTTGTCCTGATTG R-ATTGACACAGCACCTCTCATTA	120	103.70	0.99

3.2.4 Quantitative RT-PCR analysis

The quantitative RT-PCR analysis was carried out using SYBR green master mix (2X Brilliant III SYBR Green QPCR; Agilent Technologies, Santa Clara, CA, USA), on AriaMx Agilent system (AriaMx; Agilent Technologies, Santa Clara, CA, USA) with the following reaction conditions: initial denaturation at 95°C for 3 min, 40 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 45 s extension, and a melt-curve program (65-95°C with a temperature increase of 0.5°C after every 5 s). The melting curve was generated to determine the amplicon specificity. The qRT-PCR experiments were performed using three biological and two technical replicates. A reaction with no template control and a reverse transcription negative control were performed to check the potential reagents and genomic DNA contamination.

3.2.5 Analysis of gene expression stability

The candidate reference gene ranking was analyzed using five different algorithms such as geNorm, NormFinder, Bestkeeper, ΔCt, and comprehensive ranking analysis by RefFinder.

3.2.6 geNorm analysis

The geNorm determines the most stable reference genes based on the gene expression stability value (M) for a reference gene. It also calculates the minimum number of candidate reference genes required for normalization of target genes. It requires calculated Cq values in relative quantities using the given formula: $Q = E^{(minCq-Cq)}$, where Q represents sample quantity relative to sample with the highest expression, E is amplification efficiency and min Cq is the lowest Cq values. The stability value (M) is defined as an average pairwise variation (V) of the gene compared with all other tested reference genes and the cut-off is 1.5 (Vandesompele et al., 2002). If M value is lower than 1.5, it represents stable candidate reference gene and higher values reflect least stable.

3.2.7 NormFinder

NormFinder calculates expression stability values for candidate reference genes and evaluates the most stable reference gene pairs. It also calculates intra and intergroup variation using a direct comparison between genes. It uses same input calculation files which are required for geNorm with a little variation such as the first row represents a sample, the first column represents genes and the last row represents a group of samples. NormFinder is available with Excel spreadsheet add-in (https://moma.dk/normfinder-software). It ranks candidate reference genes based on M value. Lowest M value represents

most stable reference gene and higher the value, least stable are the genes (Andersen et al., 2004).

3.2.8 BestKeeper analysis

BestKeeper determines the best reference gene based on the normalization factor (also called Bestkeeper index) and pairwise correlation analysis. It requires raw Cq values as an input data to select most stable and least stable candidate reference genes. It is available in MS Excel spreadsheet file (http://www.gene-quantification. de/bestkeeper.html) and in RefFinder (http://150.216.56.64/referencegene.php/type =reference) as well. It evaluates the candidate reference gene stability by comparing the standard deviation of each gene and averages of these values. It also calculates the coefficient of variance, Pearson correlation coefficient (r) values, geometric mean (GM) and arithmetic mean (AM).

3.2.9 ΔCt method

This tool is available in an MS Excel spreadsheet as well as in RefFinder (http: //150.216.56.64/referencegene.php/type=reference). It calculates the stable candidate reference gene based on standard deviation and pairwise comparison with other genes. ΔCt requires raw Cq values as an input data. It considers a pair of gene for calculations and compares ΔCt values among genes (Silver et al., 2006).

3.2.10 RefFinder analysis

RefFinder is a web-based comprehensive tool developed for evaluating and screening

reference genes from extensive experimental datasets. RefFinder was used to generate comprehensive stability rankings (Xie et al., 2012). Comprehensive ranking of seven candidate reference genes was analyzed using RefFinder.

3.2.11 Validation of candidate reference genes

The reliability of highly stable candidate reference genes identified in the current study was validated using sodium/proton antiporter (*NHX*) as a salt stress target gene. The differential gene expression profiles of *NHX* under salt stress at 0, 6, 12 and 24 h were normalized using *EF1α*, *ACT*, 18S and *UBQ* along with the combination of *EF1α+ACT* genes. The input values for *EF1α+ACT* were calculated using the geometric mean formula given below to normalize gene of interest Geometric Mean $= \sqrt[n]{\times 1, \times 2, \times 3, \ldots \times n,}$ where the n= number of times (Vandesompele et al., 2002). The average Cq values from three biological replicates were used for relative expression analysis and the relative gene expression level calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Statistical analysis was performed using SPSS 15.0 for windows evaluation version to verify the significant difference between relative gene expressions. One-way Analysis of variance (ANOVA) with Tukey's Honest Significant Difference (HSD) test was performed for comparison between reference genes and target genes. A p-value < 0.05 was considered statistically significant.

3.2.12 Minimum Information for publication of qRT-PCR experiments guidelines

All the qRT-PCR experiments and data analysis in the present study were performed in accordance with the MIQE guidelines (Bustin, 2002).

3.3 Results

3.3.1 Expression profiling of selected reference genes

In order to select stable reference genes, transcript levels in tissues such as leaf, root, stem, and flower as well as salt stress samples were quantified based on their cDNA concentration. The primer specificity was determined by PCR products wherein single, expected amplicon size was obtained (Figure 3.1).

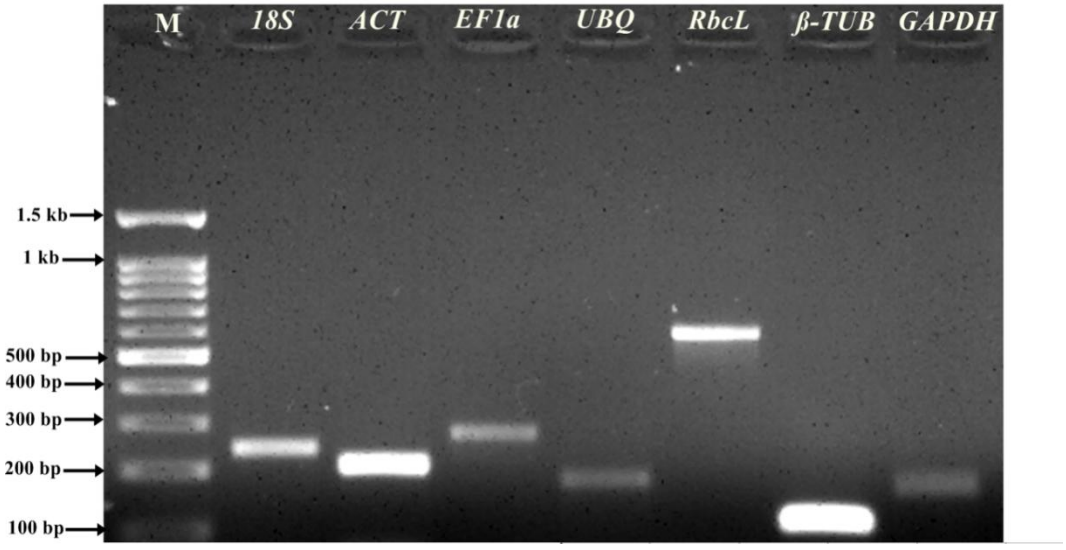


Figure 3.1 Amplification of candidate reference genes product. PCR products on 2% agarose gel stained with EtBr. Amplification products of seven candidate reference genes selected for gene validation of *R. apiculata* samples. M: 100 bp DNA ladder. Lanes 1, 2, 3, 4, 5, 6 and 7 were the gene products of 18S, *ACT, EF1α, UBQ, rbcL, β-TUB,* and *GAPDH* respectively.

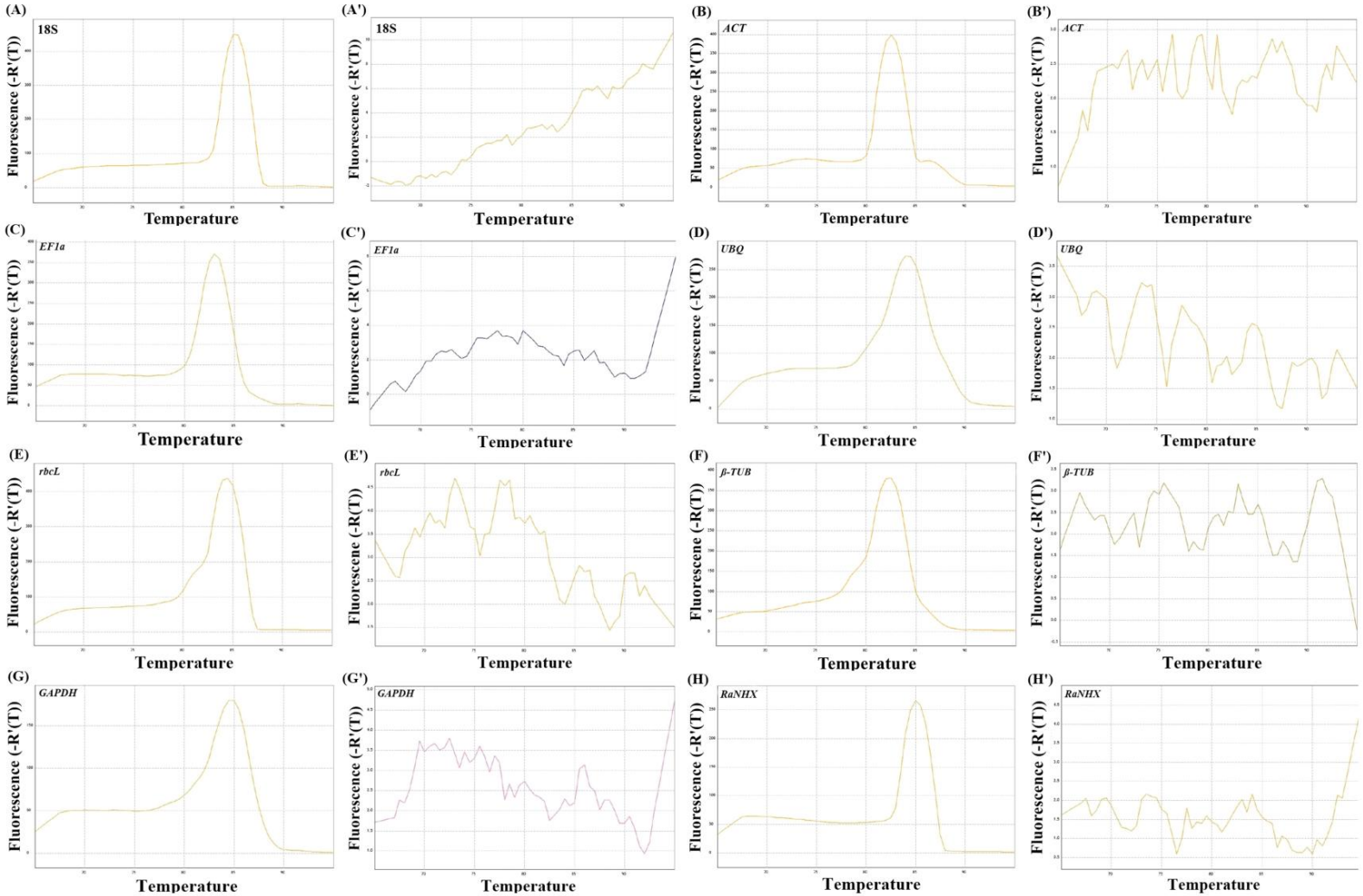


Figure 3.2 Melting curve analysis for the qRT-PCR template and negative control (NTC). Template melting curves (A) 18S, (B) *ACT*, (C) *EF1*α, (D) *UBQ*, (E) *RbcL*, (F) β-*TUB* (G) *GAPDH* and (H) *NHX*. Negative control samples without template (NTC) melting curve (A') 18S, (B') *ACT*, (C') *EF1*α, (D') *UBQ*, (E') r*bcL*, (F') β-*TUB* (G') *GAPDH* and (H') *NHX*.

The qRT-PCR melting curve for template test and negative control (NTC) without template were analyzed for primer-dimer and reagents contamination (Figure 3.2). Further, NTC samples were confirmed by running 2% agarose gel electrophoresis. The amplified PCR products were sequenced and submitted to GenBank for accession numbers. All the sequenced PCR products were identified and annotated based on BLAST search. The primer efficiency (%) ranged from 103.70%, (R^2 =0.997) for *NHX* to 90.54% (R^2=0.98) for *UBQ* including 18S (102.30, R^2 =1), *ACT* (92.03%, R^2 =0.994), EF1α (98.60%, R^2=0.99), *β-TUB* (94.08%, R^2 =0.996), *GAPDH* (96.78%, R^2 =0.992) and *RbcL* (97.69%, R^2 =0.996) (Table 3.1).

The mean cycle threshold (Cq) values of the seven selected reference genes for different tissue samples ranged from 14.16 for 18S to 21.77 for *GAPDH* (Figure 3.3 A). Similarly, for the salt stress samples, the mean Cq values ranged from 13.96 for 18S to 24.23 for *UBQ* (Figure 3.3 B). Mean Cq values gave insight into approximate gene expression data. Negative control showed higher Cq values indicating no product amplification which was further checked on a 2% agarose gel. Moreover, negative control without reverse transcriptase did not show any product amplification, thus indicating no gDNA contamination.

3.3.2 geNorm analysis

For physiological tissues, seven candidate reference genes showed average expression stability value (M) less than 1.5. ACT (M =0.721) was most stable reference gene followed by *EF1α* (M =0.761), and *β-TUB* (M =0.763) (Table 3.2 A; Figure 3.4 A). *GAPDH* was the least stable candidate reference gene with M value 1.599.

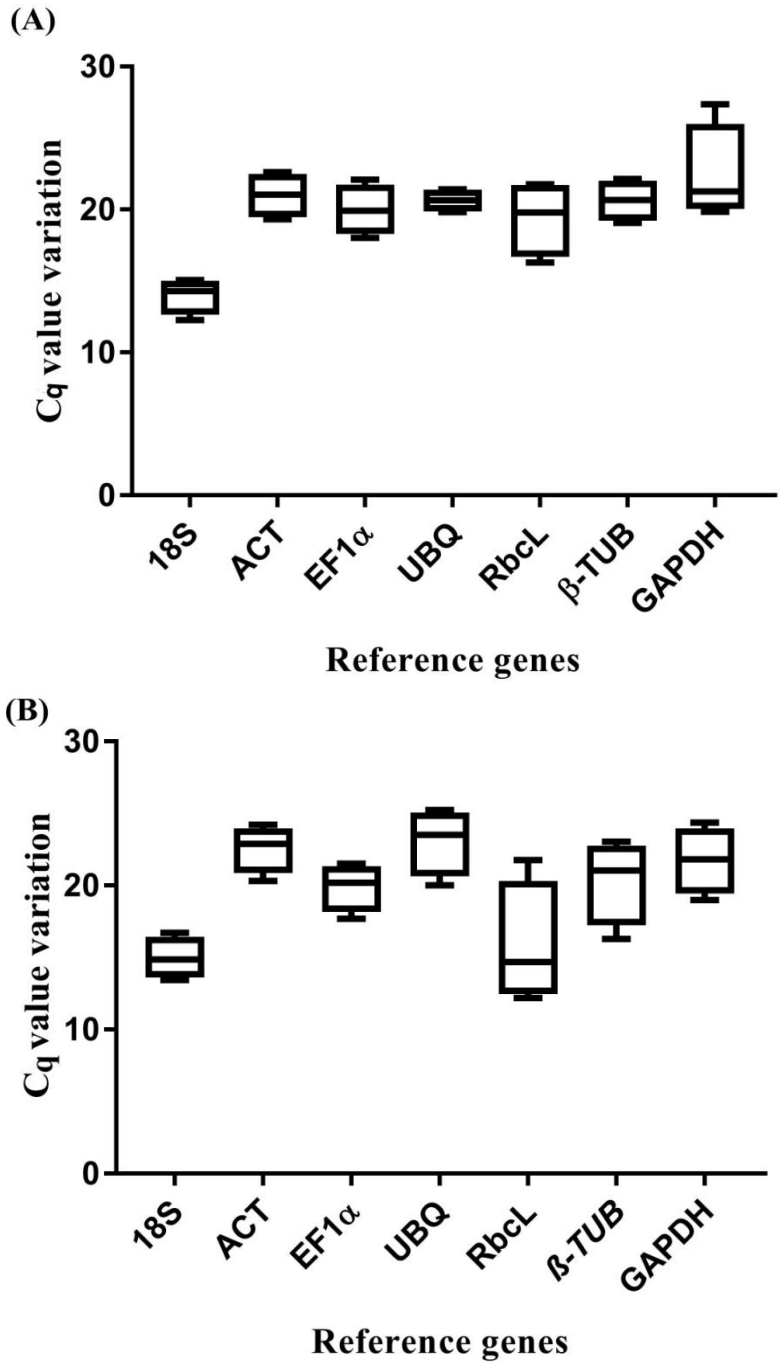


Figure 3.3 Quantification cycle (Cq) values of seven candidate reference genes. (A) Tissue-specific box plot for the Cq values of seven candidate reference genes from the qRT-PCR analysis. (B) Salt stress box plot for the Cq values of seven candidate reference genes from the qRT-PCR analysis. For each reference gene, the line inside the box is the median. The top and bottom lines of the box are the first and third quartiles, respectively. The top and bottom whiskers are the 5^{th} and 95^{th} percentiles, respectively.

The geNorm also determines an optimum number of candidate reference gene for normalization based on the calculation of pairwise variation (V_n/V_{n+1}) between sequential normalization factor (NF_n and NF_{n+1}). To select the best pair for normalization, the threshold value is 0.15. If pairwise variation value is lower than 0.15, there is no need to add more candidate reference gene. Moreover, the best pairwise variation value 0.382 was observed for a combination of *ACT* and *EF1α* and comprehensively recommended for normalization (Table 3.2B; Figure 3.4 A).

Table 3.2 geNorm **a**nalysis and ranking of candidate reference genes based on stability value (*M*). Lower *M* value represents most stable reference genes and Higher *M* value showed least stable reference genes.

A. geNorm analysis for individual candidate reference genes					
Sr. No	**Reference genes**	**Physiological tissue samples**		**Salt stress samples**	
		Stability Value (M <1.5)	Ranking	Stability Value (M <1.5)	Ranking
1	18S	0.880	4	1.020	2
2	*ACT*	0.721	1	0.927	1
3	*EF1α*	0.761	2	1.021	3
4	*UBQ*	0.928	5	1.399	7
5	*RbcL*	1.225	6	1.357	6
6	*B-TUB*	0.763	3	1.351	5
7	*GAPDH*	1.599	7	1.257	4
B. Best pair of reference genes based on geNorm analysis					
1	*ACT+EF1 α*	0.382	Most stable pair of reference genes in physiological tissue		
2	*ACT+EF1 α*	0.462	Most stable pair of reference genes under salt stress		

Based on the observation, there were no effects on an addition of the third reference gene in the combination of *ACT* and *EF1α* which showed pairwise variation value below 0.15 (Figure 3.5 A). Under salinity stress, *ACT* was most stable candidate reference gene with M value 0.927, followed by 18S and *EF1α* showing same stability M value 1.02 (Table 3.2A).

Moreover, *rbcL* and *UBQ* performed least stable candidate reference gene with M value 1.357 and 1.399 respectively. In salt stress, *ACT+EF1α* were the most suitable combination for normalization of the gene of interest with pairwise variation value of 0.462 (Table 3.2B; Figure 3.4 B). According to pairwise variation analysis, if the third gene was added in the *ACT+EF1α*, it showed higher pairwise variation value of 0.215 (Figure 3.5 B).

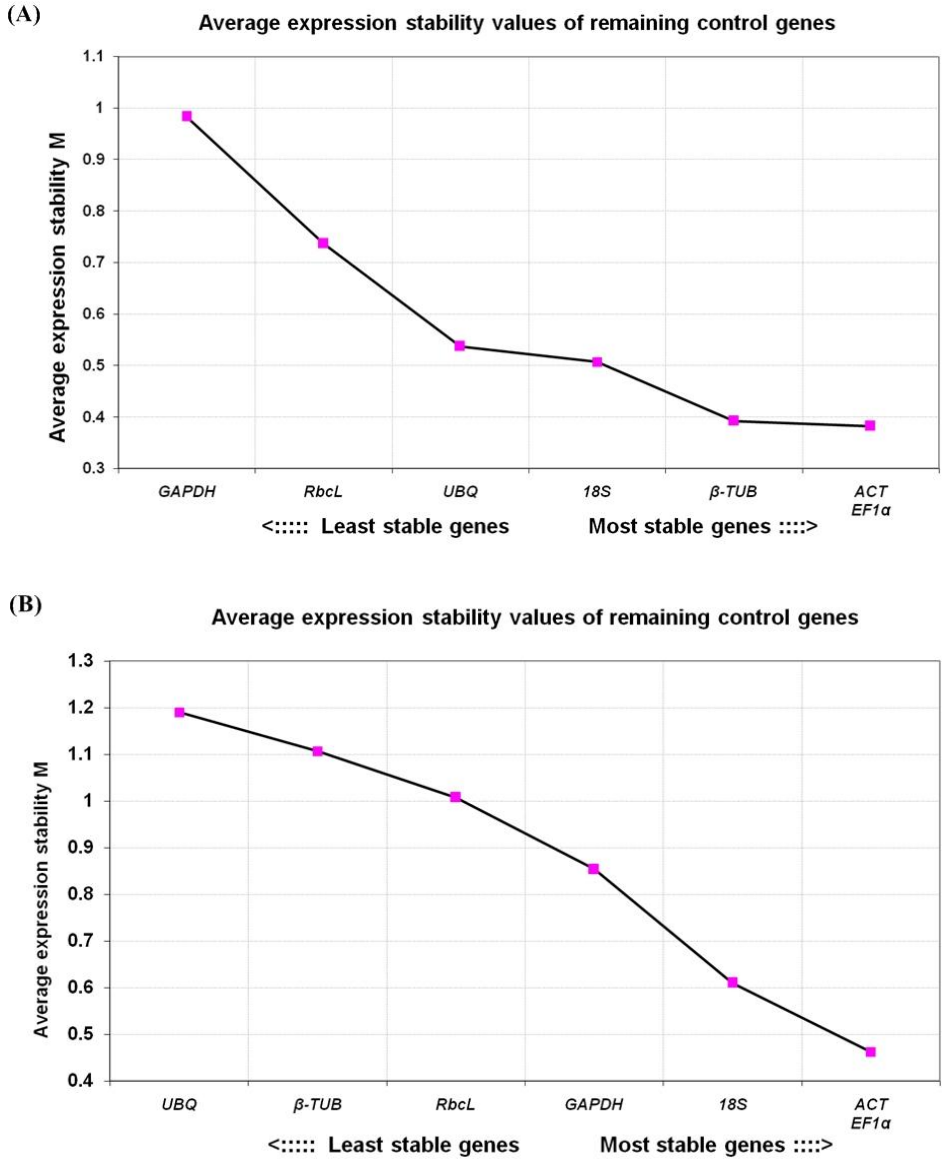


Figure 3.4 Average expression stability (M) of the candidate reference genes are analysed using geNorm tool. The X-axis represents the most stable and least stable candidate reference genes. (A) Physiological tissue samples (B) Salt stress samples.

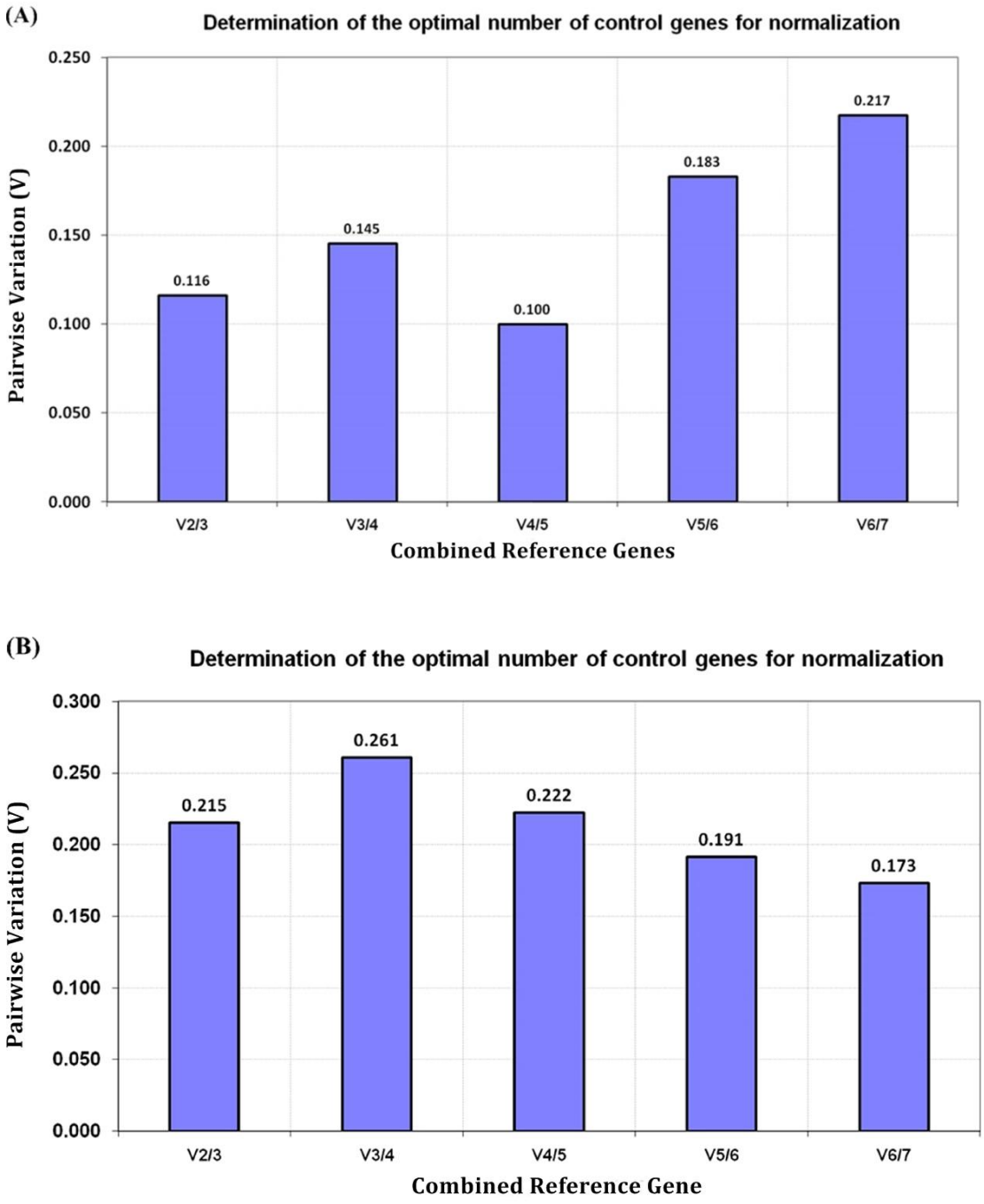


Figure 3.5 geNorm pairwise variation (V) analyses to determine minimum number of candidate reference genes required for normalization in qRT-PCR of *R. apiculata* (A) Pairwise variation analysis for physiological tissue samples (B) Salt stress leaf samples. V1 to V7 stand for the variation in candidate reference genes ranked based on their stability, where V1 is the variation for the most stable and V7 is the variation for the least stable gene.

3.3.3 NormFinder

In *R. apiculata* physiological tissue samples *EF1α* was most stable with stability value of 0.085. The *β-TUB* (0.135) was the second most stable candidate reference gene followed by *ACT* (0.164) (Table 3.3A). *EF1α* and *β-TUB* (0.070) showed the most stable combination for the pair of candidate genes for normalization (Table 3.3B). Overall, *GAPDH*, *UBQ*, and *RbcL* were least stable reference genes. In salt stress, *ACT* was most stable reference gene with a stability value of 0.196. *EF1α* and 18S were second and third most stable candidate reference genes with stability value 0.257 and 0.273 respectively. *ACT* and *EF1α* showed the best pair of reference genes with stability value 0.183 (Table 3.3B). Under salt stress, geNorm and Normfinder showed almost similar results for a selection of candidate reference gene.

Table 3.3 NormFinder analysis and ranking of candidate reference genes based on stability value. Lower stability value represents most stable reference genes and higher value showed least stable reference genes.

A. NormFinder analysis for individual candidate reference genes					
Sr. No	Reference genes	Physiological tissue samples		Salt stress samples	
		Stability value	Ranking	Stability value	Ranking
1	18S	0.410	4	0.273	3
2	*ACT*	0.164	3	0.196	1
3	*EF1α*	0.085	1	0.257	2
4	*UBQ*	0.463	5	0.518	6
5	*RbcL*	0.500	6	0.499	5
6	*B-TUB*	0.135	2	0.533	7
7	*GAPDH*	0.568	7	0.483	4

B. Best pair of candidate reference genes based on geNorm analysis			
Sr. No.	Best pair of genes	Stability value	
1	*EF1 α+B-TUB*	0.070	Most stable pair of candidate reference genes in physiological tissue
2	*ACT+EF1 α*	0.183	Most stable pair of candidate reference genes in salt stress

3.3.4 BestKeeper

In the BestKeeper analysis, standard deviation (SD) and coefficient of correlation (r) value were the criteria used for comparison. Highest r value represents the most stable candidate reference genes and lower r value represents the least stable genes. Here, we considered r value for evaluation, showing *EF1α* as the most stable reference gene followed by *ACT* with r value 0.987 and 0.966 respectively. *GAPDH* was ranked as the least stable candidate reference gene with lower r value (Table 3.4). The result is consistent with geNorm and NormFinder analysis. In salt stress, *ACT* (r = 0.638) showed most stability followed by *β-TUB* (r = 0.625) and *EF1α* (r = 0.523) (Table 3.4). Under salt stress, similar results were observed with little variation in BestKeeper. BestKeeper determined *β-TUB* second most stable candidate reference gene in salt stress.

Table 3.4 Candidates reference gene stability and ranking were analyzed by BestKeeper (Coefficient of correlation r), Ct (Mean STDEV) ranking of genes. Coeff of corr- Coefficient of correlation, RG- reference genes.

Sr. No.	RG	**BestKeeper**				**ΔCt Analysis**			
		Physiological Tissue samples		**Salt stress samples**		**Physiological Tissue samples**		**Salt stress samples**	
		(r)	Rank	(r)	Rank	Mean SD	Rank	Mean SD	Rank
1	18S	0.935	4	0.507	4	0.88	3	1.02	2
2	*ACT*	0.966	2	0.638	1	0.76	2	0.93	1
3	*EF1α*	0.987	1	0.523	3	0.72	1	1.02	2
4	*UBQ*	0.964	3	0.001	7	0.93	4	1.40	6
5	*RbcL*	0.958	5	0.310	6	1.22	5	1.36	5
6	*B-TUB*	0.964	3	0.625	2	0.76	2	1.35	4
7	*GAPDH*	0.850	6	0.435	5	1.60	6	1.26	3

3.3.5 ΔCt analysis

According to ΔCt analysis, *EF1α* was the most stable candidate reference gene followed by *ACT* and *β-TUB* in physiological tissue (Table 3.4). 18S was ranked as an average or moderately stable reference gene. The results were consistent with earlier analysis. *GAPDH*, *RbcL*, and *UBQ* were the least stable. Under salt stress, *ACT* was most stable candidate reference gene followed by *EF1α* and18S (Table 3.4).

3.3.6 Comprehensive ranking of candidate reference genes

Based on the Geomean value, a comprehensive ranking of all candidate reference genes showed *EF1α* (1.32) was the most stable followed by *ACT* (2.34) and *β-TUB* (2.91) (Figure 3.6 A). Moreover, *rbcL* and *GAPDH* performed as a least stable candidate reference genes. In salt stress, *ACT* was the most stable with Geomean value 1. Moreover, 18S was second most stable candidate reference gene with Geomean value 2.29 (Figure 3.6 B). Here, *UBQ* and *rbcL* performed as least stable candidate reference genes.

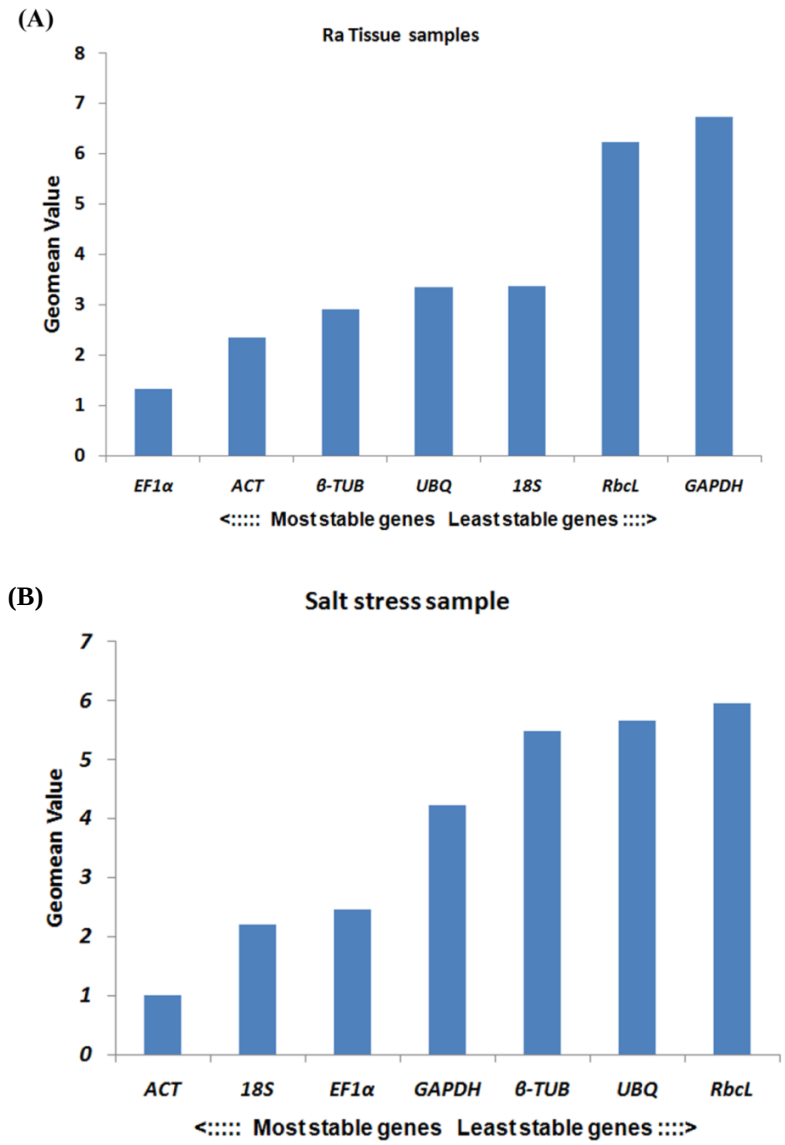


Figure 3.6 Comprehensive ranking of candidate reference genes in *R. apiculata* based on the rankings from each algorithms using RefFinder (A) Overall ranking of candidate reference gene in physiological tissues (B) Overall ranking of candidate reference gene in salt stress leaf samples.

3.3.7 Validation of stable candidate reference genes under salt stress

To validate the efficacy of candidate reference genes, *ACT*, *EF1α*, 18S and *UBQ* were used to normalize the expression levels of NHX in salt stress at four different time course. Set of the most stable candidate reference genes such as *ACT*, *EF1α*, 18S and the least stable

candidate reference gene *UBQ* were used as internal controls. While using *EF1α*, *ACT*, and 18S alone for normalization, *NHX* showed significant upregulation expression pattern in salt stress at 12 h. However, with *UBQ* as an internal control, *NHX* expression was upregulated in salt stress after 6 h of salt stress (Figure 3.7).

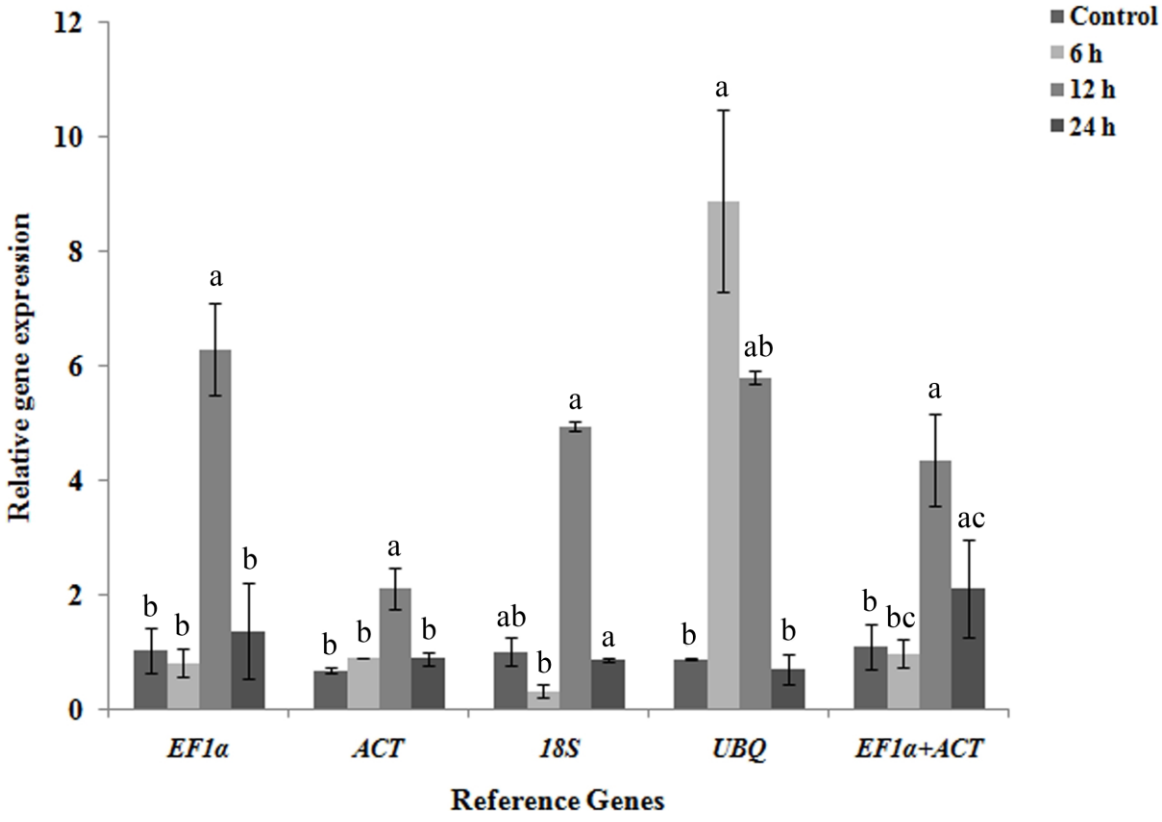


Figure 3.7 Validation and normalization of target NHX gene of *R. apiculata* under salt stress at four different time-courses such as 0, 6, 12 and 24 h using *EF1α*, *ACT*, 18S and *UBQ* reference genes. Normalization of *NHX* using *EF1α, ACT, 18S, UBQ* and combined *EF1α+ACT*, Error bars represent the mean standard error of relative abundance of three biological replicates. The bars having different superscript letter are significantly different at $p<0.05$.

3.4 Discussion

The present study is the first systematic assessment of candidate reference gene in *R. apiculata* physiological tissues as well as under salt stress. The MIQE guidelines give a

framework for good experimental practice and transparent results (Bustin et al., 2009). The results were in accordance with the MIQE guidelines, where the ideal PCR efficiency is 100%, while the acceptable range is from 90 to 110% (Bustin et al., 2009). In the present study, we designed the primers based on homologous genes of *Arabidopsis thaliana* because genome sequence of *R. apiculata* is not available. To check the designed primers specificity, we tested PCR and confirmed on 2% agarose gel for single desired size bands. Further, amplified PCR products were confirmed through sequencing and identified by BLASTN tool. All the sequences were submitted to GenBank for accession numbers. In the present work, the primer efficiency ranged from 90-103% and most of the study reported primer efficiency ranging from 90-110%. The primer efficiency was recorded between 92 to 98.6% in *Sesuvium portulacastrum*, 92.89-98.76% in *Suaeda aralocaspica*, 81-100.88% in *Halostachys caspica*, 90.5-104.43% in *Cajanus cajan* (Sinha et al., 2015; Cao et al., 2016; Zhang et al., 2016; Nikalje et al., 2018). Selection of unstable reference gene can lead to fallacious relative gene expression result and errors in normalization (Dheda et al., 2005). Besides the selection of suitable genes, it is equally important to select more than one candidate reference gene which improves the gene expression analysis (Vandesompele et al., 2002).

The geNorm algorithm evaluates single as well as best pair stable candidate reference genes for normalization. In the current study, a comprehensive ranking of candidate reference genes was evaluated; *EF1α* being the most stable candidate reference gene in physiological tissues and *ACT* in salt stress. The geNorm algorithm gave a consistent result with a comprehensive ranking which showed the most stable candidate reference gene as *EF1α* in physiological tissues and *ACT* in salt stress tissue samples. A similar observation was

reported in the *Halostachys caspica* halophyte species, which showed that *EF1α* and *TUB3* was the most stable under salt and drought stress (Zhang et al., 2016). Under salt stress, most stable reference genes in *S. portulacastrum* shoot tissue were *β-TUB*, *EIF4a* and *EF1α*, while *UCE 2*, *TBP* and *EF1α* in the root tissue (Nikalje et al., 2018). This result reflects that a reference gene is not universal and altered according to plant species and stress conditions. So it is always recommended to select and validate the commonly used candidate reference genes. One of the possible reasons might include the differential expression patterns under unstressed and stressed conditions and a difference in response to the particular stress. We observed a little variation in assessed best pair candidate reference genes between geNorm (*EF1α+ACT*) and NormFinder (*EF1α+β-TUB*) analysis. The possible explanation is subtle differences between their algorithm methods. Similar results were observed in earlier studies during evaluation of candidate reference genes, wherein a little variation in geNorm and NormFinder was reported, which leads to minute variation in candidate reference gene ranking, as reflected in the current study (Cruz et al., 2009; Pellino et al., 2011).

The geNorm calculates candidate reference genes to normalize target gene based on their average stability value (M) and also determines the optimum number of candidate reference genes required for normalization. Although NormFinder calculates stability values for each gene and BestKeeper ranks the genes according to r values, these algorithms do not determine the minimum number of reference genes required for normalization (Kozera and Rapacz, 2013). We have performed target gene validation using geNorm analyzed data because it ranks candidate reference genes based on their stability and also evaluate the minimum number of reference genes required for normalization. We used individual candidate genes as well as a combination of *EF1α+ACT*. We found that *EF1α*, *ACT*, and 18S

had given significant upregulation of *NHX* gene, while using least stable candidate reference gene *UBQ* showed different expression pattern after normalization. We observed that relative gene expression of *NHX* showed significant transcript accumulation pattern at 12 h. It was earlier reported that most significant expression patterns were observed in *R. apiculata* after 12 h time-course (Menon and Soniya, 2014). The geNorm data suggests the use of two reference genes for normalization of gene of interest. Moreover, most of the previous study underscored the use of more than one reference gene to improve the relative gene expression (Bustin, 2002). In summary, we have successfully evaluated and validated stable reference genes in *R. apiculata* physiological tissues and under salt stress. This analysis revealed that the suitable reference genes differ between physiological tissues and in salt stress tissues. We found that commonly used reference genes such as *EF1α* and *ACT* are the most useful references in an individual as well as in combined form.

3.5 Conclusions

The current study examined the most stable candidate reference gene for the normalization of relative gene expression in *R apiculata* physiological tissue and under salt stress. We strongly recommend *EF1α* followed by *ACT* and *β-TUB* as the best stable candidate reference genes for normalization in *R. apiculata* physiological tissue gene expression analysis. Under salt stress, *EF1α* followed by *ACT* and 18S are the most suitable candidate reference genes for normalization. In conclusion, *EF1α* and *ACT* can be used as candidate reference genes for the study of *R. apiculata*.

CHAPTER 4

Cloning and characterization of *RaNHX*

CHAPTER 4

Cloning and functional characterization of a novel sodium/hydrogen antiporter from *Rhizophora apiculata*

4.1 Introduction

Salinity is one of the untoward environmental factors which affect growth and development of plants (Kumar et al., 2013). In salinity stress, Na^+ and Cl^- are main culprits which cause ion imbalance, oxidative damage; disturb metabolic and physiological processes (Kumar et al., 2013; Gupta and Huang, 2014; Negrão et al., 2017). The Na^+ enters into the cell through various membrane transporters families such as HKT family and plasma-membrane non-selective cation channels (NSCCs) (Gupta and Huang, 2014). However, plants are evolved with several counter mechanisms to cope with salinity stress including minimum uptake of Na^+, compartmentalization into the vacuoles and effluxing surplus Na^+ out of the cell (Gupta and Huang, 2014; Sandhu et al., 2017). Discovery of a novel salt overly sensitive (SOS) pathway in plants opened the new horizon of salt tolerance mechanism (Ji et al., 2013; Sandhu et al., 2017). Currently, there are three key components of SOS pathways, which regulates the SOS activity in salt stress such as SOS1 (NHX), SOS2 (protein kinase) and SOS3 (SCaBP8, calcium sensor) (Ji et al., 2013; Sandhu et al., 2017). Interestingly, the important role of SOS1 in halophytic plant *Thellungiella halophila* was demonstrated by knockdown studies of SOS1, which led to the loss of halophytic characteristics (Oh et al., 2009).

The prevalent function of SOS1 (NHX) under salt stress is an extrusion of Na^+ out of the cell. However, the sequestration of excess Na^+ into vacuole is another crucial process to

minimize cellular Na^+ toxicity and also helped to build osmotic potential inside the cell to facilitate water uptake into the cell (Gupta and Huang, 2014). Many reports are available on functional characterization of plant NHX under salinity stress and also their involvement in several physiological processes such as transport of Na^+ as well as K^+, osmotic adjustment and water uptake, growth and development of cell, vesicular trafficking and protein targeting, calcium signaling, stomatal movements as well as flowering (Bassil et al., 2011; Pittman, 2012; Andrés et al., 2014; Bassil and Blumwald, 2014; Reguera et al., 2014).

Mangroves are natural salt tolerant plant species which is evolutionarily adapted to the intertidal coastal ecosystem (Kathiresan, 2018). However, the halophytes rely on the unique mechanism to cope with salinity stresses such as ultrafiltration, minimum uptake and compartmentalization of Na^+, K^+ and Cl^-, synthesis of an osmoprotectant, and excess ions secreted through salt glands (Menon and Soniya, 2014). *Rhizophora apiculata* is a small mangrove tree, distributed along the coastal area of the tropical and subtropical region of the world. They are perennial halophytic species belongs to the Rhizophoraceae family and provides natural tolerance to salinity stress. Till now, there are no reports available on functional characterization of NHX members from mangroves species. The aim of the present study is cloning, transcript analysis and functional characterization of *R. apiculata* NHX member by heterologous expression in a yeast mutant lacking NHX.

4.2 Materials and methods

4.2.1 *R. apiculata* growth and stress treatments

Three month-old seedlings of *R. apiculata* were collected from the mangroves plant nursery, Chorao Island, maintained by Goa Forest Department, Goa, India. *R. apiculata*

species were identified based on molecular as well as morphological features. Three-month-old seedlings were uprooted carefully and transferred to hydroponic condition at 25°C for two weeks. Plants were treated in the presence of 250 mM NaCl for various time-courses (0, 1, 3, 6, 12, 24, 48 h) followed by harvesting of leaves, shoots and roots separately. For elemental analysis leaves, stems, and roots samples were collected separately and stored at room temperature. All tissue samples for qRT-PCR and elemental analyses were collected in biological replicates.

4.2.2 RNA isolation and qRT-PCR analysis

Total RNA isolation was successfully performed using a modification of CTAB method (Fu et al., 2004). A freshly collected plant samples were directly homogenized into CTAB buffer with 2% PVP-30 and 10% β-ME in sterile mortar and pestle. The suspension was incubated at 60°C for 60 m with gentle mixing and centrifuged at 14,000 rpm for 10 m at room temperature (RT) with an equal volume of chloroform: isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and RNA was precipitated with 0.6 volume of cold isopropanol (-20°C) and 8 M lithium chloride followed by incubation at -20°C for 1 h. The precipitated RNA was centrifuged at 14,000 rpm for 10 min at RT followed by washing with 70% ethanol. RNA was finally dissolved in 0.1% DEPC treated sterile water. Genomic DNA contamination was removed by DNase I enzyme (Thermo Scientific USA) treatment at 37°C for 30 m and heat inactivated at 65°C for 10 min with 50 mM EDTA. The cDNA synthesis was performed in 20 µl reaction volume using the RevertAid Reverse Transcriptase (Thermo Scientific USA), 0.1-5 µg RNA sample and oligo $(dT)_{18}$ primer, as per the manufacturer's instructions. Quantity and quality of RNA were confirmed on the agarose gel electrophoresis

and the spectrophotometer (Nanodrop, Thermo Scientific, USA). *RaNHX1* specific primer for the qRT-PCR was designed using the PrimerQuest (Integrated DNA Technologies) (Table 4.1). The experiments were performed on AriaMx qRT-PCR system (AriaMx Agilent Technologies) using SYBR green master mix (2X Brilliant III SYBR® Green QPCR, Agilent Technologies). The relative *RaNHX1* expression was quantified using the $2^{-\Delta\Delta C}{}_{T}$ method using 18S rRNA as a reference gene (Livak and Schmittgen, 2001). Three biological replicates of each sample were used for qRT-PCR analysis.

Table 4.1 List of primers used in the present study. List of *RaNHX1* degenerate, full-length primers, cloning primers for yeast and plant, and qRT-PCR primers were given in the table below.

S N	Primers Name	Primer purpose	5'-3' Primer sequences
1	NHX_D_F	NHX amplification	TATWATATTCAATGCMGGGTTTCARGTR
2	NHX_D_R	NHX amplification	GCATTRTGCCARGTRTAATGWGACATVAC
3	FLnhx_F	Full length NHX	ATGGATTCGTACGTTAGCTC
4	FLnhx_R	Full length NHX	TCATTGCAATTGATTGTGAACGC
5	RaC_F	NHX cloning in yeast	TACGTCGTCGACATGGATTCGTACGTTAGCTC
6	RaC_R	NHX cloning in yeast	TAG CTGTCTAGATCATTGCAATTGATTGTGA
7	RaL_F	NHX localization	TACGTCCCATTGATGGATTCGTACGTTAGCTC
8	RaL_R	NHX localization	TAGCTGACTAGTTTGCAATTGATTGTGAACGC
9	qNHX_F	qRT-PCR primer	TGCTAGCTCTTGTCCTGATTG
10	qNHX_R	qRT-PCR primer	ATTGACACAGCACCTCTCATTA

4.2.3 *In-silico* analysis of RaNHX sequence

The full-length *RaNHX* nucleotide was translated protein sequence was analyzed for the presence of motifs and domains using Interproscan 5 (http://www.ebi.ac.uk/Tools/pfa/iprscan). The transmembrane regions were predicted using the TMpred online tool. Physiochemical data were generated from the ExPASy ProtParam server (http://web.expasy.org/protparam/), including sequence length and molecular weight. RaNHX1 protein secondary structure was predicted by the Psipred server

(http://bioinf.cs.ucl.ac.uk/psipred/), phosphorylation sites (http://www.cbs.dtu.dk/services/NetPhos/**)**, lipid modification and SUMO modifications (http://gps.biocuckoo.org/) were predicted. A phylogenetic tree was constructed based on the Neighbor-Joining (NJ) tree based on the bootstrap value of 1000 replicates using MEGA 7 tool. Threading of RaNHX1 3D model was performed on the I-TASSER web server with default parameters (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The secondary structure of the target protein RaNHX1 was predicted using the position-specific iterated prediction (PSI-PRED) and position specific iterated-BLAST (PSI-BLAST). An evaluation of the protein structure was done using the Support Vector Machine SVMSEQ and SPICKER programs. The quality of the model was estimated based on C-score, cluster density, TM-score, and root mean squared deviation (RMSD). The PyMOL (Schrodinger Inc) was used to visualize the model and generate publishable images.

4.2.4 Cloning of *RaNHX* in pYES2.1 and pCAMBIA1302

The full length cds of *RaNHX1* was amplified from *R. apiculata* using degenerate primers designed using highly conserved sequences of vacuolar NHX regions. Amplified full-length cds was cloned into pGEM-TEasy vector system (Promega, UK) as per manufacturer's instruction. The *RaNHX1* clone was sequenced and submitted to GenBank. Full-length fragment of *RaNHX1* was cloned into the GAL (galactose) promoter-driven expression cassette of pYES2.1 using *XbaI/XhoI* restriction sites for heterologous expression in a yeast system. Initially construct transformed in competent *E. coli* cells selected on the Luria-Bertani agar plates supplemented with amphicilin. Further, six positive constructs confirmed through colony PCR and restriction digestion of *RaNHX1*. Purified plasmid pYES2.1 was

transformed in the AXT3 mutant strain using lithium chloride method and positive colony was screened on the SD uracil dropout media. Empty pYES2.1 vector was transformed in the AXT3 mutant strain and used for comparative phenotypic study with AXT3-pYES-*RaNHX1*.

For subcellular localization, full-length *RaNHX1* without stop codon sequence was cloned into pGEM-TEasy vector system (Promega). Further, the *RaNHX1* cloned into the pCAMBIA1302 binary vector using *NcoI/SpeI* restriction sites with green fluorescence protein (GFP) reporter gene to create a recombinant cassette, *RaNHX1*: GFP. A recombinant cassette was transformed in to *E. coli* and positives clones were screened on the LB agar plates supplemented with kanamycin. After positive clone selection, pCAMBIA1302 vector with *RaNHX1:gfp* was purified using alkaline lysis method and confirmed the presence of *RaNHX1* insert through PCR and restriction digestion. Further, plasmid transformants were purified and checked in the *Agrobacterium* strain EHA105 by $CaCl_2$ method and positive clones selected on yeast extract and peptone (YEP) media supplemented with rifampicin and kanamycin antibiotics (Holster et al., 1978). YEP plates were incubated at 30 °C for two days. For subcellular localization in plant, the pCAMBIA1302 vector with *RaNHX1:gfp* construct was transformed into tobacco leaves using *Agrobacterium* strain EHA105 by agro-infiltration method (Zhao et al., 2017). The transformed leaf incubated up to 48 hrs and images were captured by an Olympus IX81/FV500 confocal microscopy using argon laser (488 nm) and a green helium/neon laser (543 nm).

4.2.5 Measurement of total Na^+ and K^+ ions in salt-stressed *R. apiculata* tissues

R. apiculata seedlings stressed tissues such as leaves, stems, and roots were harvested at different time-courses (0, 6, 12, and 24 h). The samples were dried, digested with 1.5 ml

concentrated HNO_3 at 90°C for 1 h and centrifuged at 12,000 rpm for 10 min at room temp. (Mishra et al., 2014). The suspension was diluted with 13.5 ml of sterile milliQ water and analyzed for Na^+ and K^+ content in the flame photometer. Elemental analysis experiments for *R. apiculata* were performed with three biological replicates.

4.2.6 Yeast strains, media, and growth conditions

Saccharomyces cerevisiae strain W303-1A (*MATa ade2-1 can1-100 his3-11, 15 leu2-3,112 trp1-1 ura3-1*), and AXT3K (*ena1-4D::HIS3, nha1D::LEU2, nhx1D::TRP1*) was a gift from Dr. Kees Venema. Untransformed strains were grown at 30°C in 1% yeast extract, 2% peptone and 2% glucose medium (YPD). Yeast cells were transformed with pYES2.1 empty as well as pYES-*RaNHX1* vectors using the lithium acetate method and positive transformants were selected on synthetic defined uracil dropout (SD-URA^-) media. Hygromycin-B antibiotic sensitivity was performed on YPD media supplemented with different concentration of hygromycin-B such as 75, 100 and 150 $\mu g.ml^{-1}$ and spotted on plate's 10-fold serially diluted culture (0, 10^{-1}, 10^{-2} and 10^{-3}). Complementation assay was performed by culturing yeast cells with OD_{600} of 0.1 and 10-fold serially diluted (1, 10^{-1}, 10^{-2} and 10^{-3}) and spotting 5 µl on SD-Gal media supplemented with or without different concentrations of NaCl (50, 100, 150, and 200 mM) and KCl (500 mM, 750 mM, and 1 M). SD Plates were incubated at 30 °C and growth was observed after 2 days. Effect of pH on AXT3 growth under NaCl 150 mM and KCl 750 mM concentrations were performed at three different pH range such as 3.0, 4.5 and 7.0. SD media pH was adjusted with 8 M arginine for alkaline and phosphoric acid for acidic conditions. Yeast growth performed at 30 °C in liquid SD-Gal media supplemented with different concentration of NaCl (50, 100, 150, and 200

mM) and KCl (250, 500, 750, and 1000 mM) and the growth was observed at OD_{600} after 48 h. All the results were confirmed by performing experiments three times.

4.2.7 Measurement of Na^+ and K^+ in yeast mutant

Total ion contents were measured in W303-1A, AXT3-pYES2.1 and AXT3-*RaNHX1* yeast strains by growing the strains at 30 °C in liquid YPGal media, pH 4.6 supplemented without or with 75 mM NaCl. Briefly, cells were harvested at an OD_{600}, centrifuged at 3000g at 3 min, washed twice in ice-cold 10 mM $MgCl_2$, 10 mM $CaCl_2$ and 1 mM HEPES buffer and re-suspended in the same buffer. The cell density and yeast dry weight was determined. Yeast cells were subjected to acid digestion using hydrochloric acid (HCl) with a final working concentration of 0.4% and incubated at 90°C till complete digestion. The procedure was followed by centrifugation at 3000g at 3 min in order to remove cell debris and supernatant reconstituted with sterile milliQ water. The total Na^+ and K^+ were measured in the yeast cells using flame photometer (Systronics, MP, India).

4.2.8 Subcellular localization of RaNHX1 in Plant

The pCAMBIA1302 vector with *RaNHX1*:*gfp* construct was transformed into tobacco leaves using *Agrobacterium* strain *EHA105* by agro-infiltration method (Zhao et al., 2017). The transformed leaves incubated up to 48 h and images were captured by an Olympus IX81/FV500 confocal microscopy using an argon laser (488 nm) and a green helium/neon laser (543 nm).

4.2.9 Statistical analysis

All the experiments were performed using biological replicates and statistical significance was performed using a one-way analysis of variance (One-way ANOVA) followed by post-hoc t-test. The $p \leq 0.01$ is considered as highly significant labeled as *.

4.3 Results

4.3.1 *In-silico* analysis, phylogeny and homology modeling of RaNHX

In the present study, we cloned full-length putative Na^+, K^+/H^+ antiporter from *R. apiculata.* The full-length vacuolar NHX was isolated from *R. apiculata* root cDNA library using a degenerate primer. *RaNHX1* nucleotide sequence deposited to GenBank and retrieved accession no. KU525079. The domains and motif prediction underscored the presence of cation/proton antiporter 1 (CPA) family (IPR018422), Na^+/H^+ exchanger domain (IPR004709), and Na^+/H^+ exchanger NHX type plants (IPR029872) which was similar to plant NHX group (Figure 4.1). RaNHX1 sequence analysis showed that it shared very close identity about 86% with AtNHX1 and AtNHX2 antiporters, henceforth we designated it as '*RaNHX1*'.

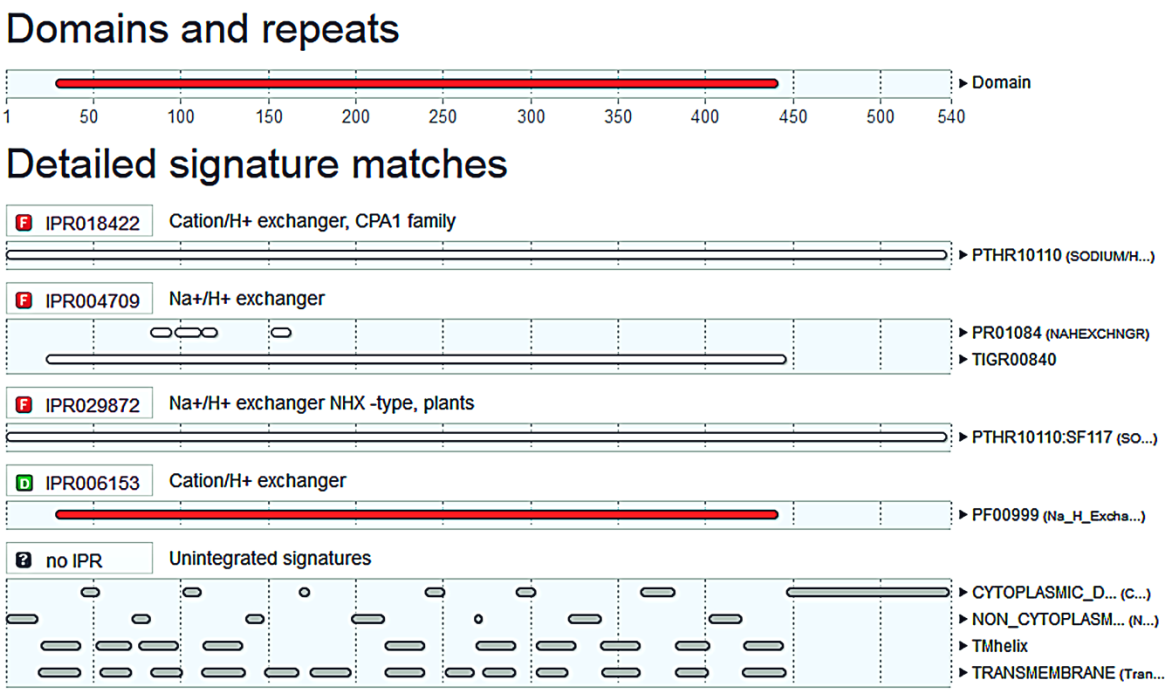


Figure 4.1 *In-silico* RaNHX1 sequence analysis using Interproscan 5 web based tool. It predicted three families such as cation/H^+ exchanger (CPA1 family), Na^+/H^+ exchanger plant family, and Na^+/H^+ exchanger NHX-type, plants. It also predicted the domain such as cation/H^+ exchanger domain.

RaNHX1 has 12 transmembrane helices, amiloride motif 'LFFIYLLPPI' in transmembrane 3 helices (TM3), NES and NVT motifs in its sequence. RaNHX1 has Asparagine (N) and Aspartic acid (D) at position 187 and at 188 positions respectively, which form conserved 'ND' motif in the TM5 region. Highly conserved four residues in RaNHX1 sequence such as Y at 149 in TM4, N at 187 in TM5, D at 188 positions in TM5, and R at 356 positions were present in TM10 region. A highly variable C-terminal domain present in the cytoplasm regulates antiporter activity of NHX. Gene ontology predicted *RaNHX1* involved in the cation/proton biological process, sodium/proton molecular function and belongs to the vacuolar membrane as a cellular component (Figure 4.2).

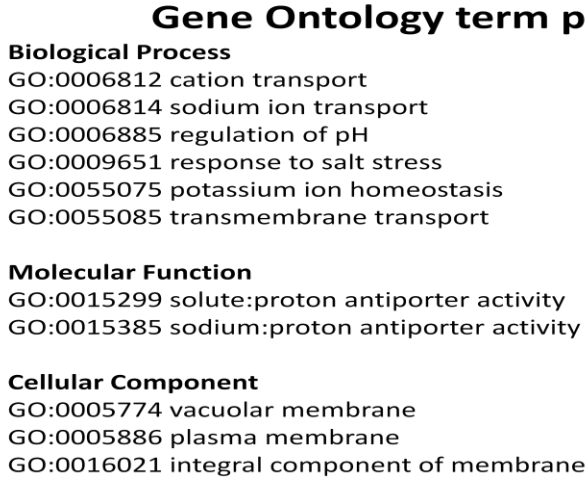
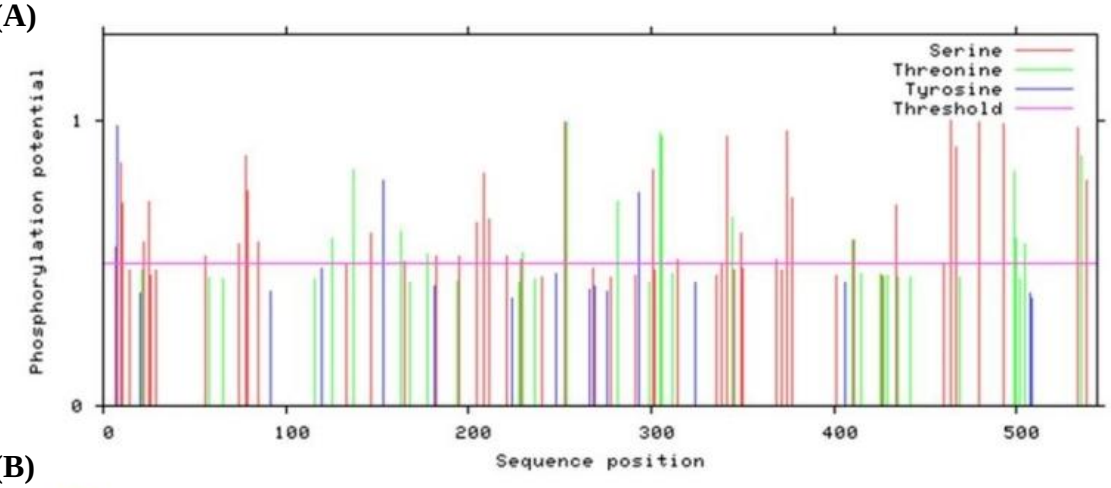


Figure 4.2 Gene ontology predicted *Rhizophora apiculata* NHX gene function based on biological process, molecular function and cellular component.

Post-translational modifications were predicted in RaNHX such as phosphorylation sites, lipid modification sites, and Small Ubiquitin-like Modifier (SUMO) modifications (Figure 4.3). Overall more than 5 serine phosphorylation sites were predicted but most of the modification sites were observed at C-terminal serine residues. Similarly, at 36 and 125 position s-palmitoylation sites were predicted. SUMO interaction sites were predicted at 57-61 and 441-445 positions. The RaNHX1 sequence has 8 Cysteine residues highlighted by a blue circle, out of those 124th residues probably involved in the C-C bond formation (Figure 4.4A).

(A)



(B)

ID	Position	Peptide	Score	Cutoff	Type
FLRa	36	FVALLCACIVIGHLL	4.861	3.076	S-Palmitoylation: Cluster B
FLRa	125	AVGTLICCCVISVGA	2.59	1.396	S-Palmitoylation: Cluster C

(C)

ID	Position	Peptide	Score	Cutoff	P-value	Type
FLRa	57 - 61	NESITALIIGVCTGVVILL	36.926	35.23	0.098	SUMO Interaction
FLRa	441 - 445	FGLLTKPLIILLLPHAKQP	46.979	35.23	0.021	SUMO Interaction

Figure 4.3 Prediction of post translational modifications in RaNHX protein (A) NetPhos 3 predicted Phosphorylation sites in the RaNHX sequence at Ser, Thre, and Tyr residues (B) GPS lipid modification prediction predicted at the position 36 and 125 are probable sites of lipid modification (C) GPS-SUMO modification prediction in the RaNHX predicted protein sequence.

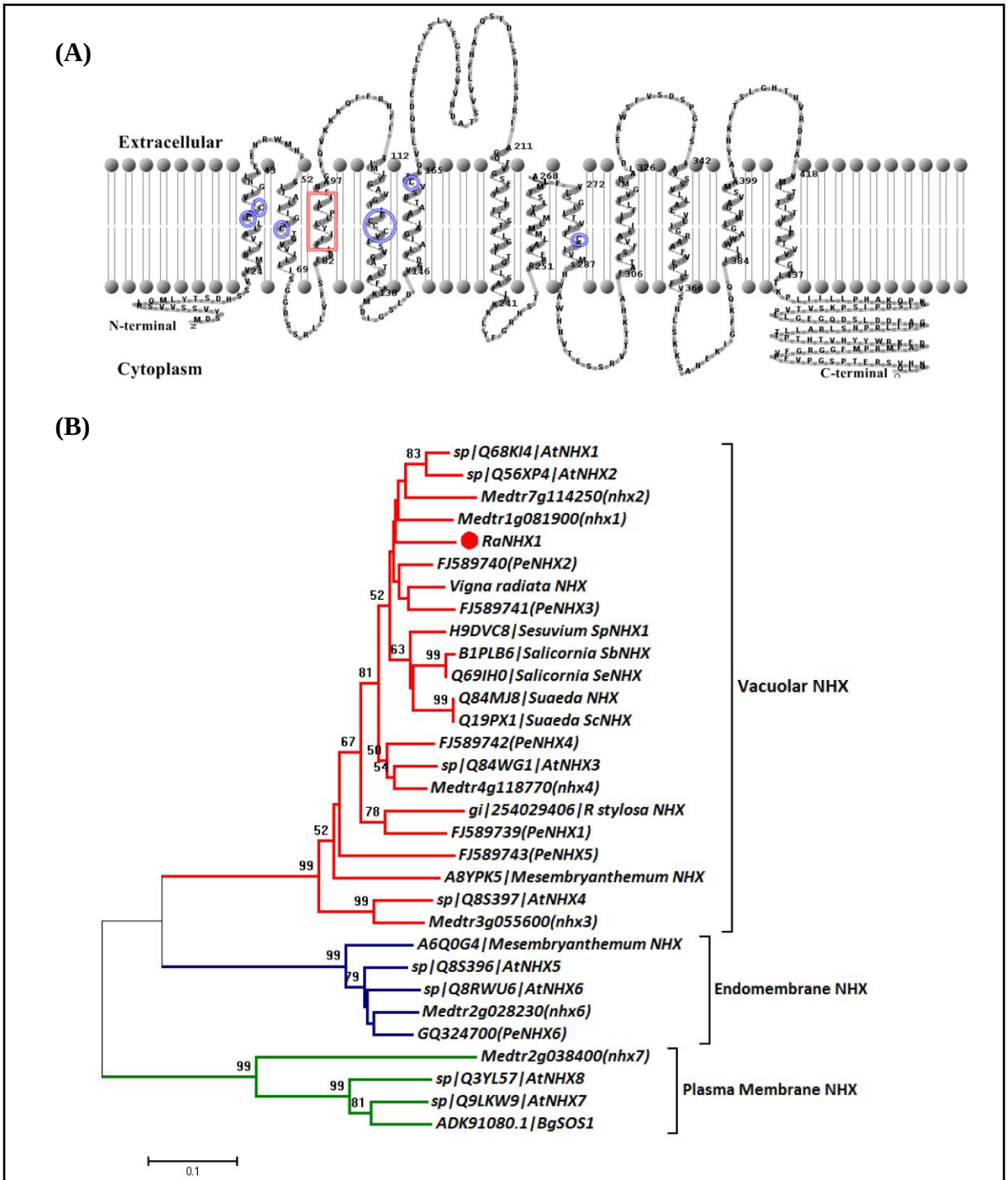


Figure 4.4 *In-silico* analysis of RaNHX1 transmembrane helices and phylogenetic analysis (A) Transmembrane cartoon image constructed by softberry software, showed amiloride motif, and probable cysteine residues by forming C-C bonds (B) Phylogenetic analysis were performed using NJ tree between *RaNHX1* and *Populus* NHX, *Medicago* NHX, *Salicornia* NHX, *Mesembryanthemum* NHX, *Suaeda* NHX, *Sesuvium* NHX, *Vigna radiata* NHX, *R. stylosa* NHX, *Bruguiera gymnorrhiza* SOS1 and *Arabidopsis* NHX family.

Phylogenetic relationship of NHX members (*Medicago truncutula, S. brachiata, M. crystallinum, Suaeda salsa, Sesuvium portulacastrum, V. radiata, R. stylosa, Bruguiera gymnorrhiza,* and *Arabidopsis*) with cloned RaNHX1, clustered into 3 different clades (group I‑III) (Figure 4.4 B). Clade I consist of all the vacuolar NHX members along with RaNHX1. RaNHX1 shared >90% bootstrap values with AtNHX1 and AtNHX2. Similarly, clade II clustered endosomal NHX members such as AtNHX5 and AtNHX6, while Clade III members, AtNHX7 and AtNHX8 were localized to the plasma membrane (SOS1) (Figure 4.4 B).

The PSIPRED server predicted secondary structure such as helix regions with pink residues, and β-sheets highlighted with yellow color (Figure 4.5 A). The top-ranked threading fold of the *E. coli* Na^+/H^+ antiporter NhaA (PDB ID-1ZCD) shared significant identity with the RaNHX sequence. I-TASSER server generated five top-ranked 3D homology models using *E. coli* NhaA as a template. After structural evaluation, the model 1 with the highest C-score (-0.99) considered a model of best quality with a structural similarity between the predicted model and the native structure (TM-score 0.59 ± 0.14 and RMSD 9.8 ± 4.6 Å). The RaNHX1 model showed the presence of twelve transmembrane helices with the large cytoplasmic C-terminal region. Amiloride motif at TM3 was highlighted with red color. Four crucial residues such as tyrosine 149, Asparagine 187, Aspartic acid 188 and Arginine 356 were shown in the predicted model (Figure 4.5 B and C). Moreover, crucial residues of TM4 and TM11 alongwith ND motif were shown in the figure 4.5C.

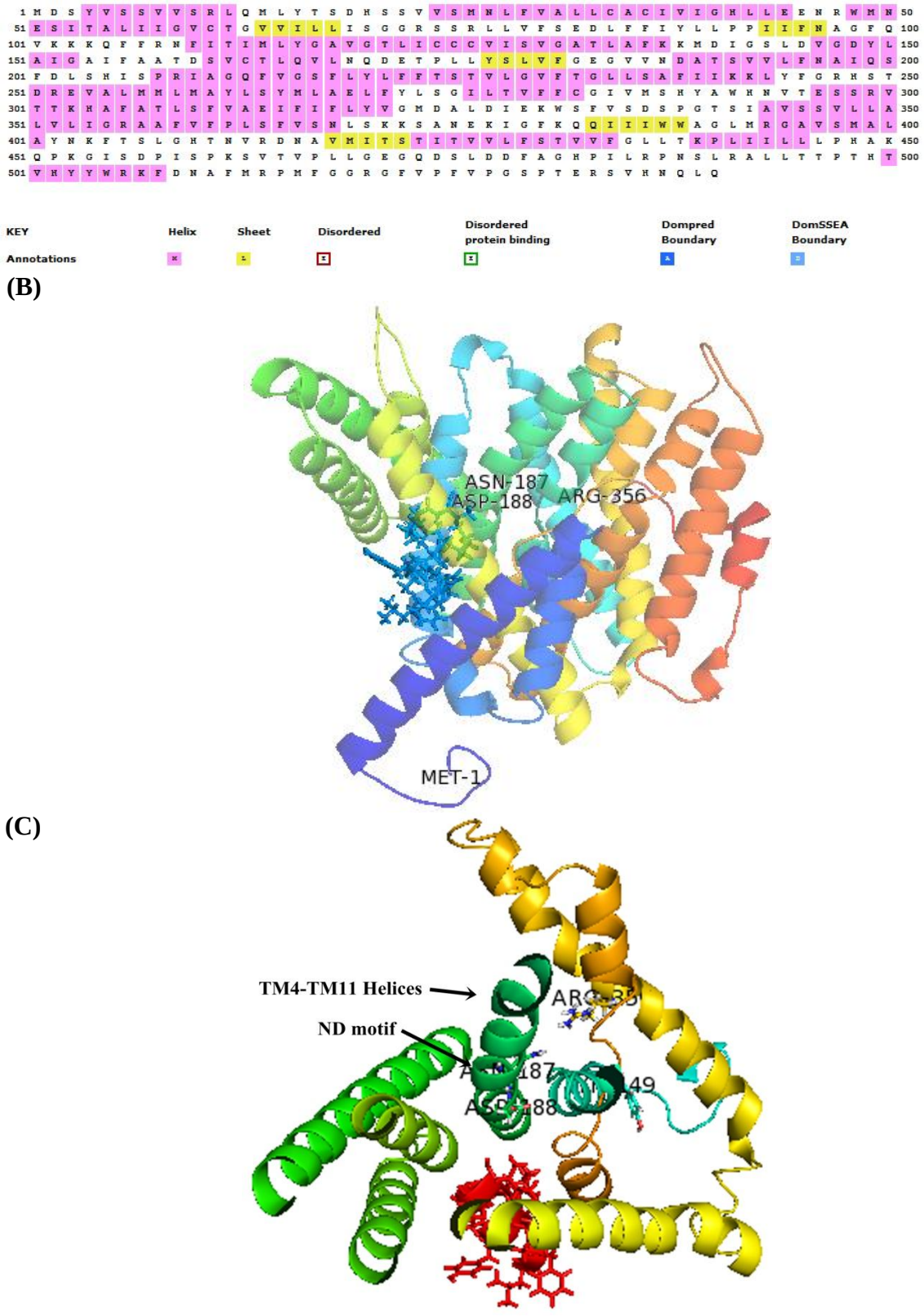


Figure 4.5 RaNHX1 secondary and 3D structure predictions (A) Secondary structure of RanHX1 (B and C) 3D model of *RaNHX1* showed amiloride motif, ND motif TM4-TM11 and conserved residues Y149, N187, D188 and R356.

4.3.2 *RaNHX1* is differentially regulated in leaves, stems, and roots under salt stress

Transcript accumulation of *RaNHX1* was performed to understand the response and regulation in different physiological tissues such as leaves, stems, anchor root, primary roots, and flowers. Moreover, the *RaNHX1* transcript response was analyzed under 250 mM salt stress at different time points (0, 1, 3, 6, 12, 24 and 48 h) in leaves, roots, and stems tissues by qRT-PCR. Tissue-specific expression of *RaNHX1* showed highest about 14-fold transcript accumulations in the stem. Young leaves showed approximately 5-fold transcript accumulation of *RaNHX1*. Primary roots, and anchor root were showed approximately 5-fold *RaNHX1* expression at physiological conditions. Moreover, matured leaves exhibited 2-fold transcript accumulation in comparison to flowers expression. Flower expression data was used for normalization of gene expression and recorded least transcript accumulation of *RaNHX1* in the flower (Figure 4.6 A). Overall this data suggested that RaNHX1 expressed in the all vegetative tissues at physiological condition and low expression detected in reproductive organ such as flower.

In salinity stress, leaves showed 3-fold up-regulation patterns of *RaNHX1* transcript at 1 h and further, downregulated at 3 h and again increased up to 5-fold at 12 h. Interestingly, transcript accumulation attained basal expression at 24 and 48 h (Figure 4.6 B). In roots tissue, initially at 1, 3 and 6 h *RaNHX1* transcript level was 9-fold down-regulated in salinity stress, while at 12 h transcript accumulation was increased to 4-fold. However, at 24 and 48 h *RaNHX1* level reached the basal level (Figure 4.6 C). Interestingly, in stems under salt stress *RaNHX1* was up-regulated up to 2-fold at 1, 3 and 6 h. Further, it was attained basal transcript accumulation at 12, 24, and 48 h (Figure 4.6 D).

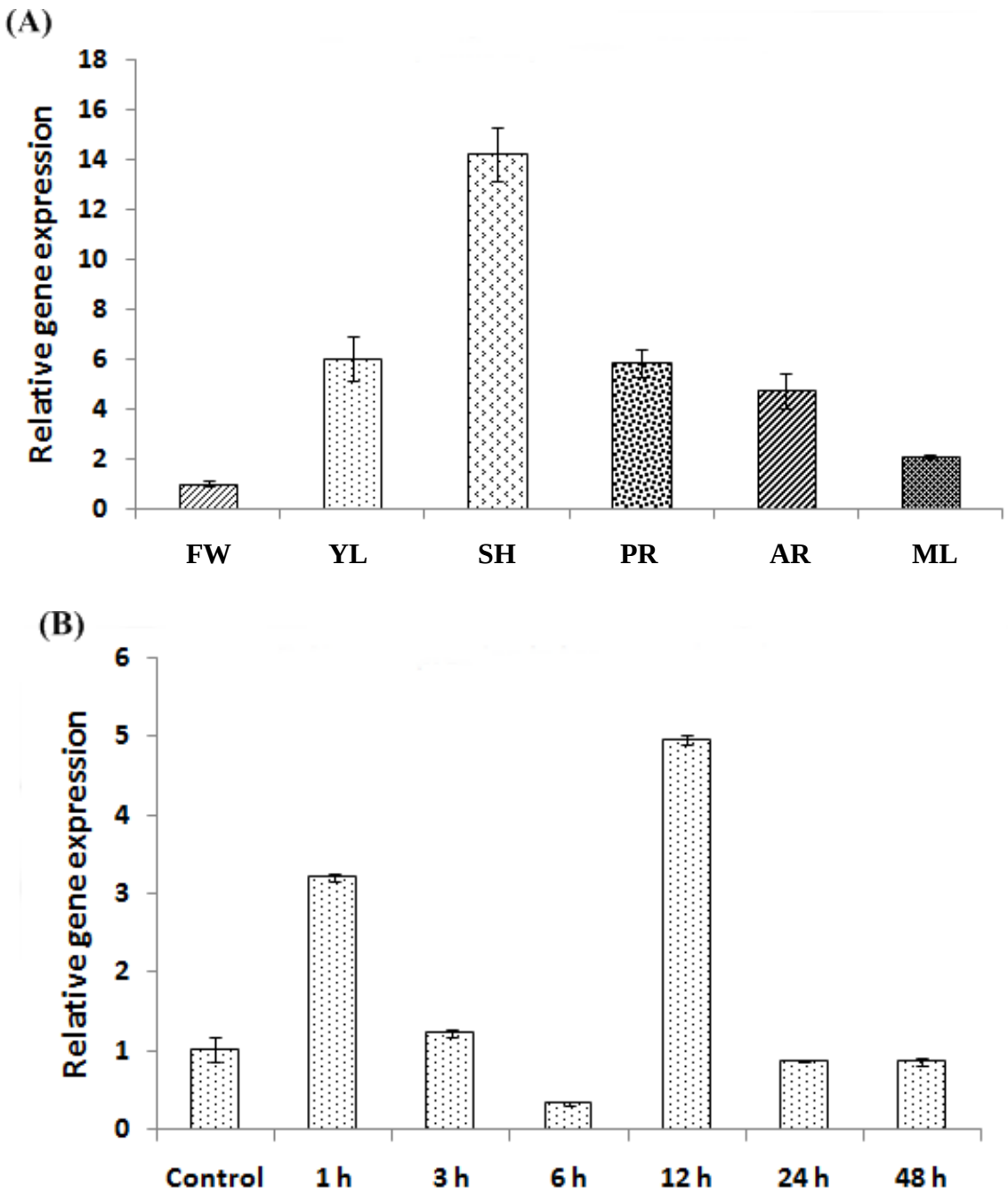


Figure 4.6 Relative gene expression of *RaNHX1* using qRT-**PCR** (A) Tissue-specific expression of *RaNHX1* in young leaves (YL), matured leaves (ML), primary roots (PR), anchor root (AR), stem (SH), and Flower (FW). The qRT-PCR data was normalized using 18S rRNA gene and are shown relative to flower (B) *RaNHX1* expression under 250 mM salt stress in leaves. Error bars represent the mean ± standard error of relative abundance of three biological replicates.

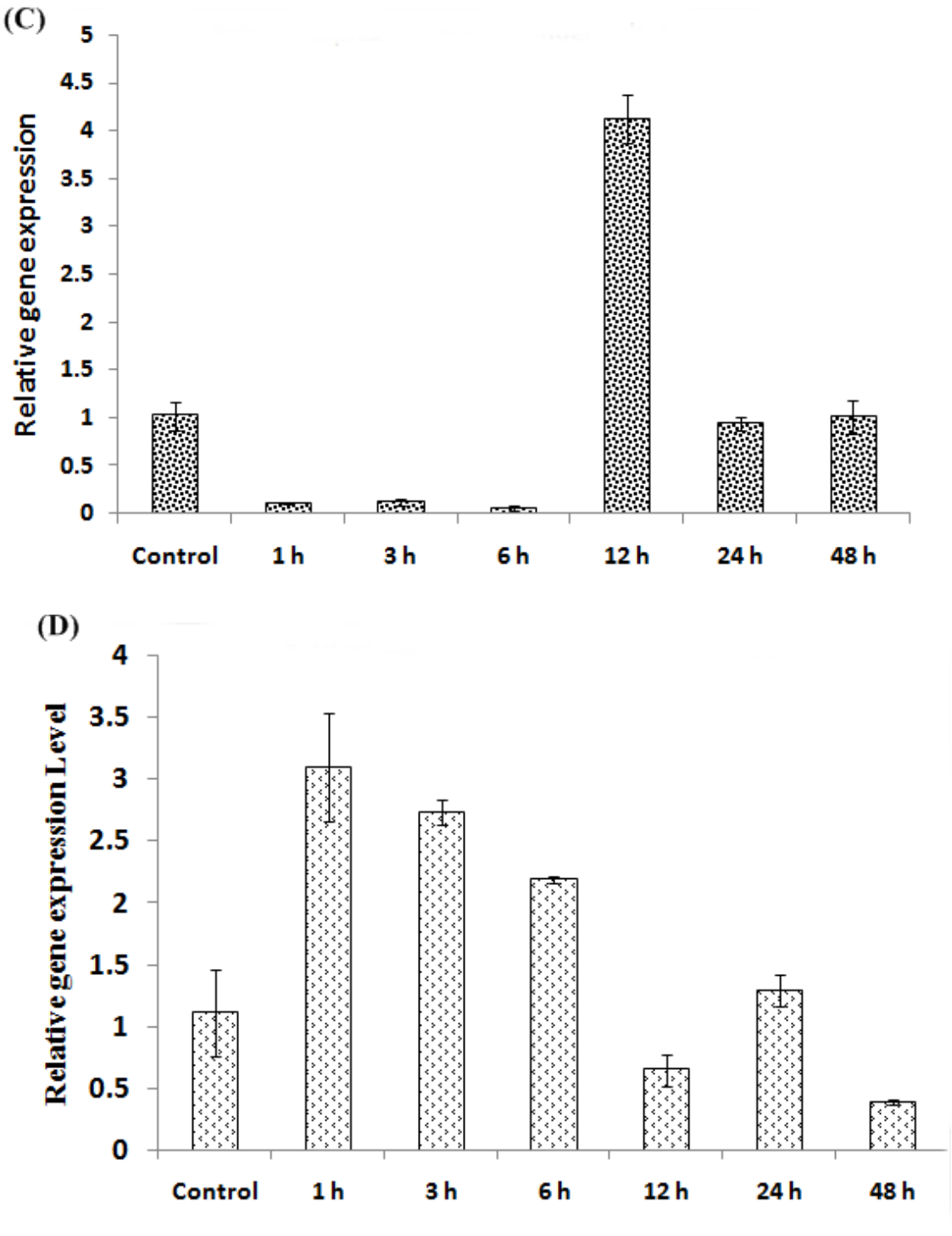


Figure 4.6 Relative gene expression of *RaNHX1* using qRT-PCR (C) *RaNHX1* expression under 250 mM salt stress in roots (D) stems. QRT-PCR data was normalized using 18S rRNA gene and are shown relative to 0 h. Error bars represent the mean ± standard error of relative abundance of three biological replicates.

4.3.3 Higher accumulation of Na^+ in *R. apiculata* leaves

Total Na^+ and K^+ concentration were analyzed in *R. apiculata* tissues under 250 mM salt stress at different time point such as 0, 6, 12 and 24 h. In leaves, the basal Na^+ concentration was approximately 0.16 µmol.mg^{-1} of dry weight without any stress at 0 h and it increased to 0.170 µmol.mg^{-1} at 24 h in salt stress condition (Figure 4.6 A). In roots at 0 h, Na^+ concentration was recorded 0.110 µmol.mg^{-1} which rapidly increased at 6 h and 24 h up to 0.126 µmol.mg^{-1}. In stem, least Na^+ (0.08 µmol.mg^{-1}) accumulation was observed at 0 h and it decreased to 0.059 µmol.mg^{-1} with increased time points. At 24 h of salt stress, 0.077 µmol.$mg^{-1}$of Na^+ accumulations were observed in the stem (Figure 4.7 A). The total K^+ analysis was performed in the tissues of leaves, stems and roots of *R. apiculata* under salt stress. Overall, maximum K^+ accumulation was observed at 6 and 12 h in all tissue and reduced at 24 h under salt stress (Figure 4.7 B). In leaves, basal level of K^+ concentration was 0.0138 µmol.mg^{-1} and it increased up to 0.014 µmol.mg^{-1} at 6 and 12 h of salt stress. Interestingly, at 24 h of salt stress, the K^+ concentration was decreased upto 0.009 µmol.mg^{-1} (Figure 4.7 B). In root tissues, the basal K^+ concentration was 0.0057 µmol.mg^{-1}, while K^+ concentration at 6 h of salt stress sharply increased up to 0.007 µmol.mg^{-1}. In contrast, at 12 and 24 h of salt stress, K^+ level decreased up to 0.0049 and 0.0027 µmol.mg^{-1} respectively. In stems, there was no significant change in K^+ level at 0 and 6 h. However, at 12 h of salt stress, K^+ concentration increased up to 0.0051 µmol.mg^{-1} and at 24 h increased up to 0.0027 µmol.mg^{-1} (Figure 4.7 B).

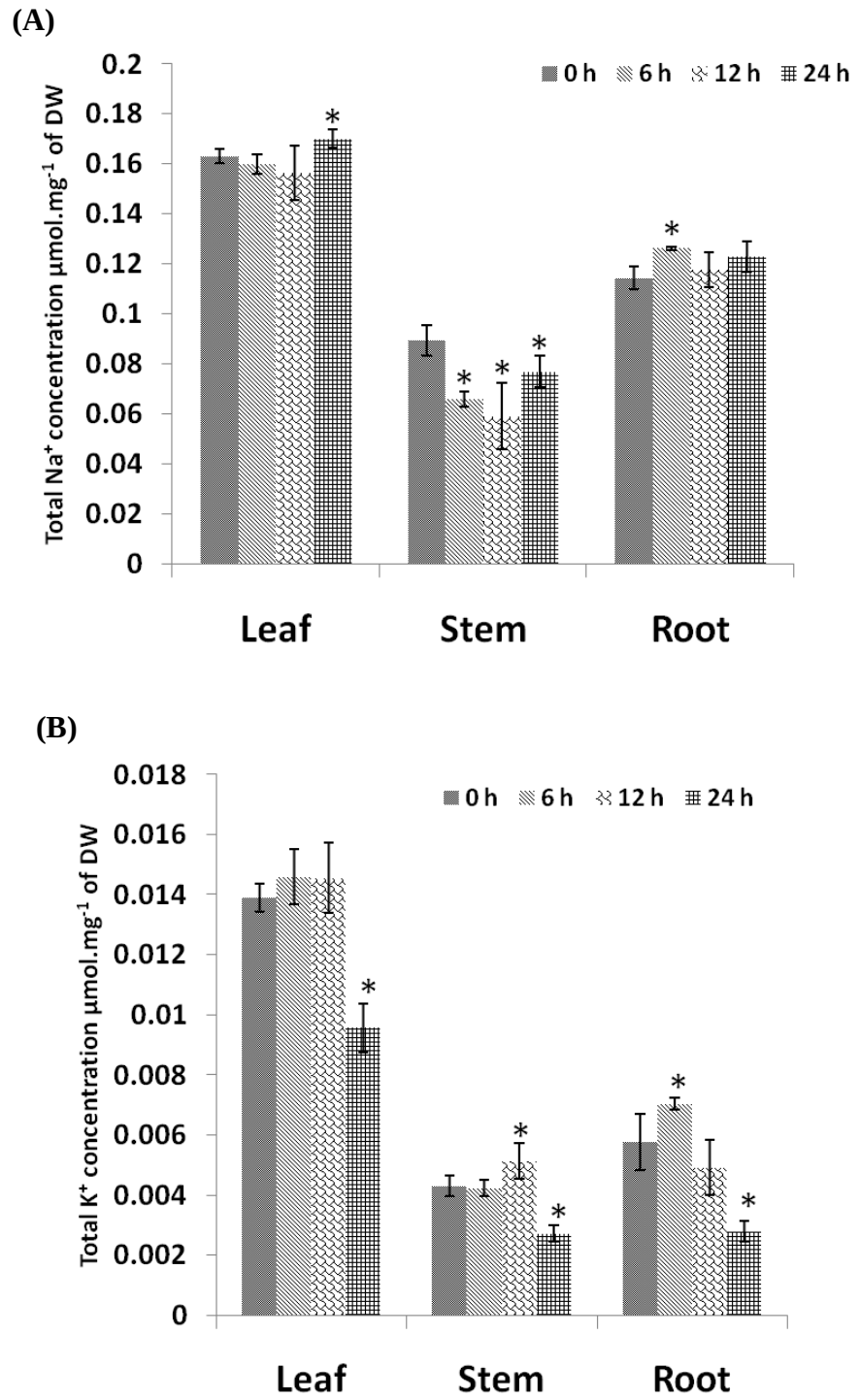


Figure 4.7 Total Na^+ and K^+ analysis in *R. apiculata* tissues under salt stress (A)Total Na^+ content and (B) Total K^+ content analyzed in leaves, roots and stems tissue under 250 mM salinity. The Representative histograms with mean ± standard error of three biological replicates are represented. The asterisks represent the statistically significant values for Na^+ and K^+ accumulations in tissues under salinity compared with 0 h tissue samples.

4.3.4 Expression of *RaNHX1* in AXT3 complements Na^+ sensitive phenotypes

The AXT3 strain lacks functional vacuolar and plasma membrane localized Na^+/H^+ antiporters such as Δ*ena1-4, Δnha1, and Δnhx1*. This strain showed hypersensitivity against hygromycin-B, a cationic aminoglycoside antibiotic, which was unable to sequester into vacuoles due to lack of vacuolar NHX transporter. *RaNHX1* was cloned with GAL promoter using *XbaI/XhoI* restriction sites (Figure 4.8 A). The hypersensitivity of W303-1A, AXT3-pYES2.1, and AXT3-*RaNHX1* yeast strains were tested using different concentration of hygromycin-B (75, 100 and 150 $\mu g.ml^{-1}$). AXT3-*RaNHX1* strain showed most tolerant phenotypes at 100 and 150 $\mu g.ml^{-1}$ of hygromycin-B compared to mutant AXT3-pYES2.1 (Figure 4.8 B). Overall, AXT3-*RaNHX1* yeast strain showed increased tolerance against hygromycin, Na^+, and K^+ ions. Different strains including mutant and AXT3-*RaNHX1* complementation activity were assessed in KCl stress using various concentration of KCl such as 500 mM, 750 mM, and 1 M (Figure 4.8 C). Similarly, AXT3-*RaNHX1* showed improved phenotypes in NaCl stress compared to AXT3-pYES2.1 mutant at a different NaCl concentration such as 50, 100, and 150 mM were supplemented with the SD media (Figure 4.8 D).

Similar results were obtained, when W303-1A, AXT3-pYES2.1 and AXT3-*RaNHX1* strains cultured in the liquid SD-Gal media supplemented with various concentrations of NaCl (50, 100, 150 and 200 mM) and KCl (250, 500, 750 and 1000 mM). The yeast strains growth were recorded at 600nm after 48 h (Figure 4.9 A and B). The significant growth differences were observed at 100, 150 and 200 mM NaCl concentration. Further, KCl 750 and 1000 mM supplemented exhibit growth difference between yeast strains as shown in Figure 4.9 B. The NHX transport activity is regulated by cellular pH. At acidic pH, its

activity sharply decreased compared to the alkaline pH. Interestingly, AXT3-*RaNHX1* showed enhanced growth on slightly alkaline media supplemented with 150mM NaCl and 750 mM KCl compared to AXT3-pYES2.1 mutant. However, at acidic pH 3.0, AXT3-*RaNHX1* and AXT3-pYES2.1 didn't show significant phenotype variation even in the presence of 150 mM NaCl and 750 mM KCl (Figure 4.10 A, B and C). These results suggested that *RaNHX1* have ability to function at a slightly alkaline pH.

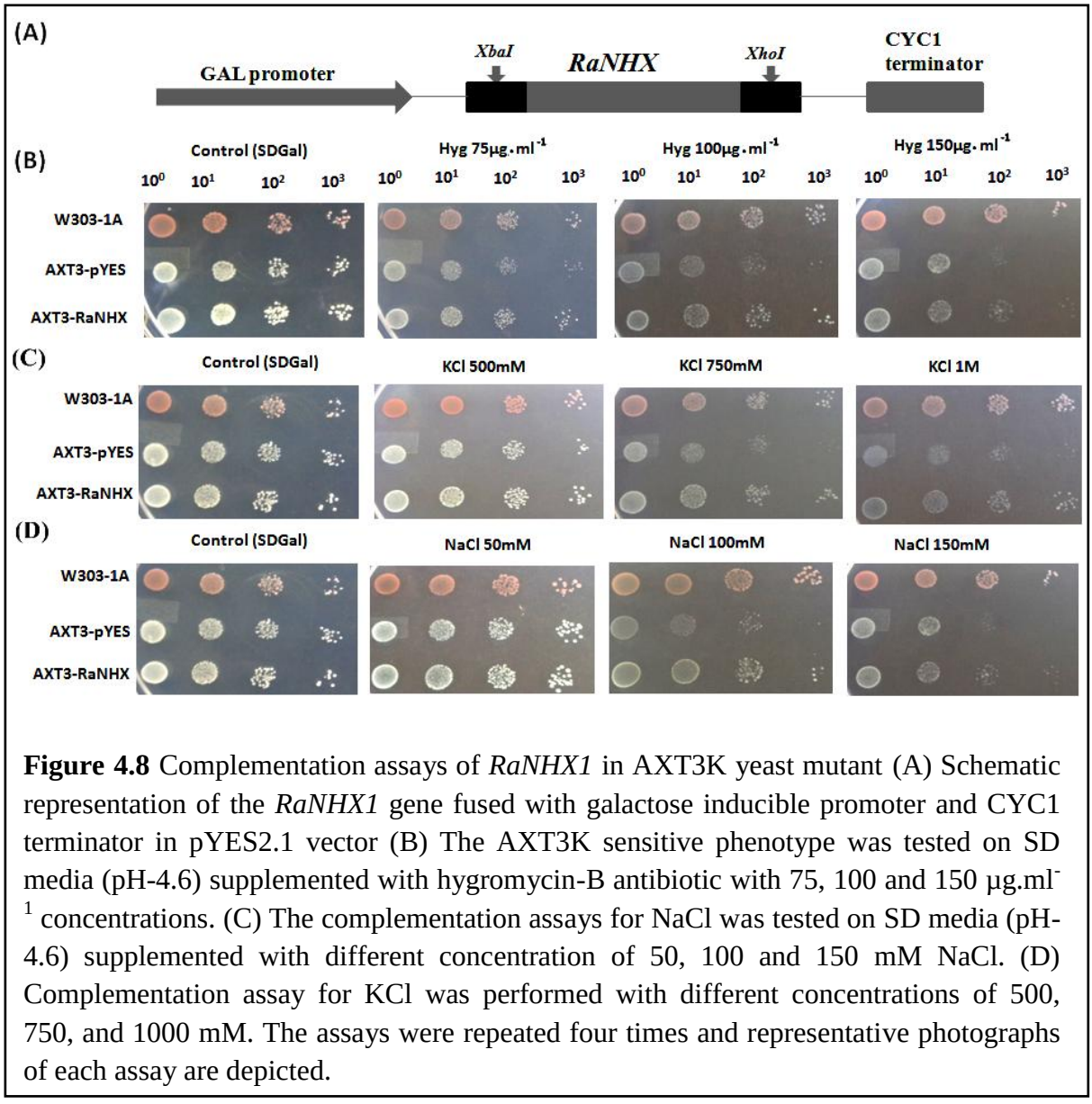


Figure 4.8 Complementation assays of *RaNHX1* in AXT3K yeast mutant (A) Schematic representation of the *RaNHX1* gene fused with galactose inducible promoter and CYC1 terminator in pYES2.1 vector (B) The AXT3K sensitive phenotype was tested on SD media (pH-4.6) supplemented with hygromycin-B antibiotic with 75, 100 and 150 µg.ml^{-1} concentrations. (C) The complementation assays for NaCl was tested on SD media (pH-4.6) supplemented with different concentration of 50, 100 and 150 mM NaCl. (D) Complementation assay for KCl was performed with different concentrations of 500, 750, and 1000 mM. The assays were repeated four times and representative photographs of each assay are depicted.

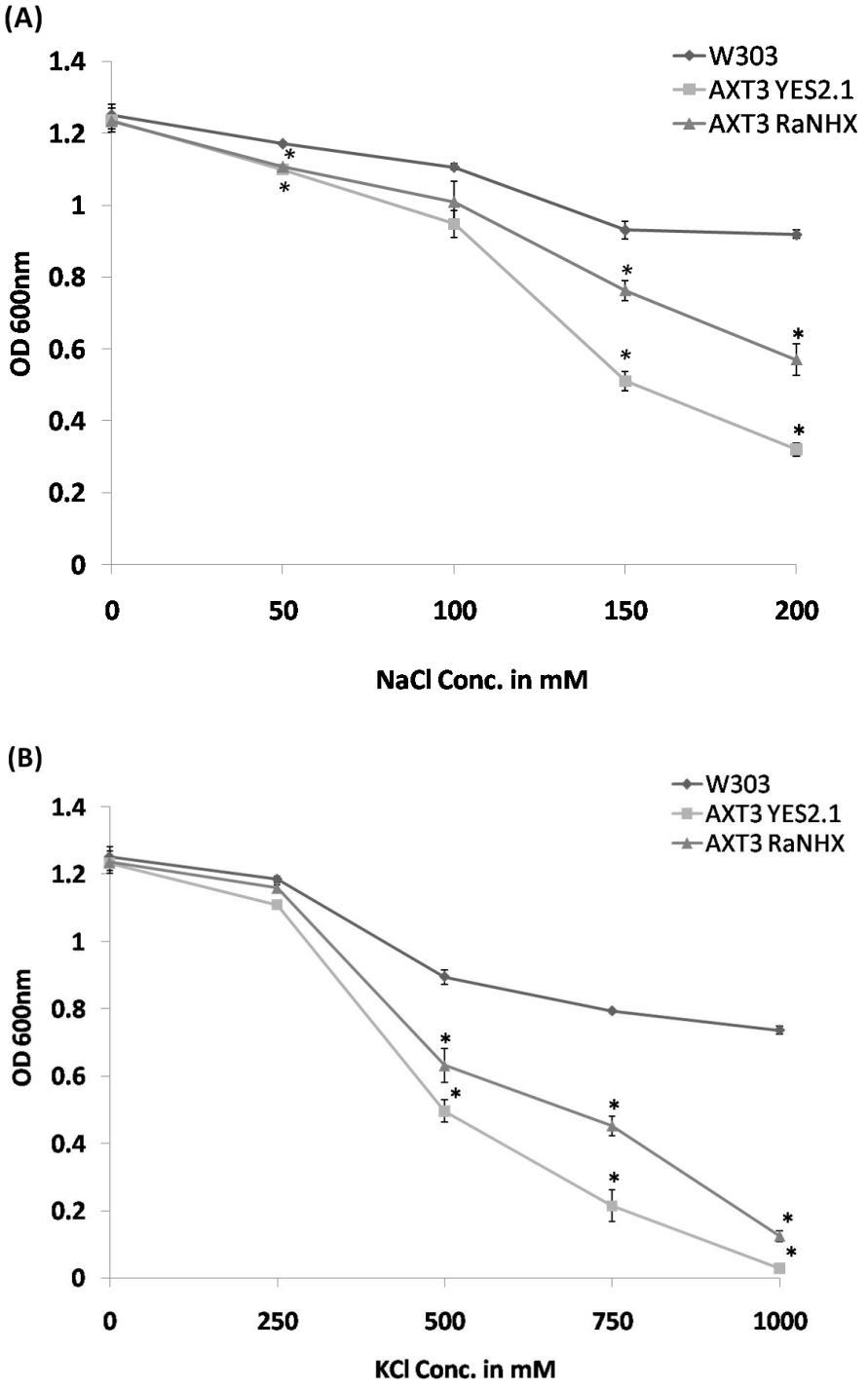


Figure 4.9 Yeast growth curve was performed in liquid SD media supplement with different concentrations of (A) NaCl and (B) KCl absorbance were measured at OD_{600} after 48 h. The asterisks represent the significant difference between growth rate under NaCl and KCl compared with W303-1A strain (*) $p < 0.01$.

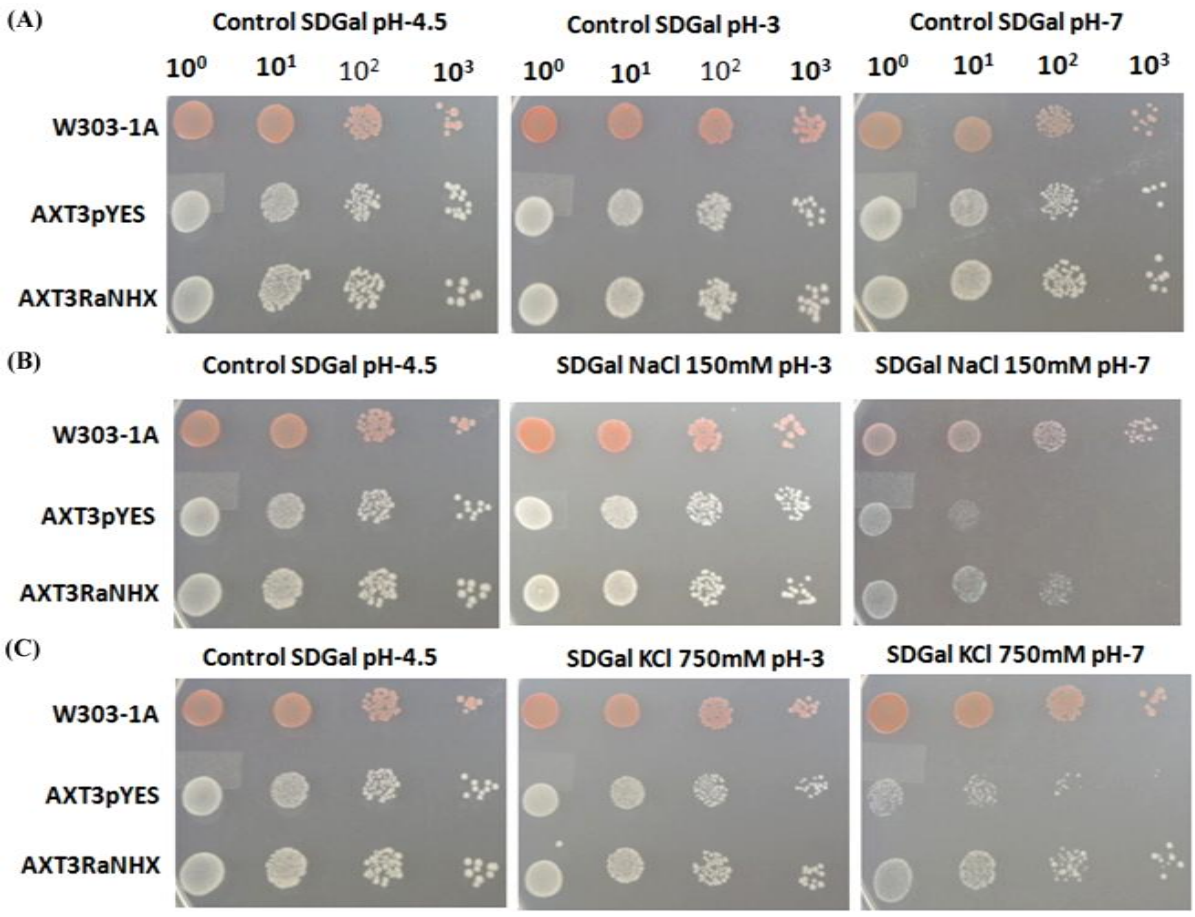


Figure 4.10 Effect of pH on AXT3 mutant and transformed cells under 150 mM NaCl and 750 mM KCl (A) SD media plates with three different pH 3, 4.6 and 7 without any salt stress were considered as control **(**B**)** The effect of pH on AXT3K phenotype was tested on SD media supplemented with 150mM NaCl concentration at pH 3, 4.6 and 7 were maintained by arginine and phosphoric acid (C) The effect of pH on AXT3 cells in the presence of KCl 750mM concentration. The assays were repeated three times and representative photographs of each assay are depicted.

4.3.5 Total Na^+ and K^+ increased in stressed condition

The AXT3 mutant strain with empty pYES2.1 vector showed lower Na^+ and K^+ accumulation than AXT3-*RaNHX1* in the unstressed condition (Figure 4.11 A and B). Total Na^+ content in unstressed W303, AXT3-pYES2.1 and AXT3-*RaNHX1* cells were ranged from 0.04 to 0.057

µmol.mg^{-1} of dry weight (Figure 4.11 A). Moreover, in 75 mM NaCl stress condition Na^+ level significantly increased in all the yeast strains such as W303, AXT3-pYES2.1 and AXT3-*RaNHX1*. Similarly, the total K^+ content observed in all three strains ranged from 0.06 to 0.1µmol.mg^{-1}. The K^+ level in W303, AXT3-pYES2.1 and in AXT3-*RaNHX1* were significantly decreased in stress condition (Figure 4.11 B).

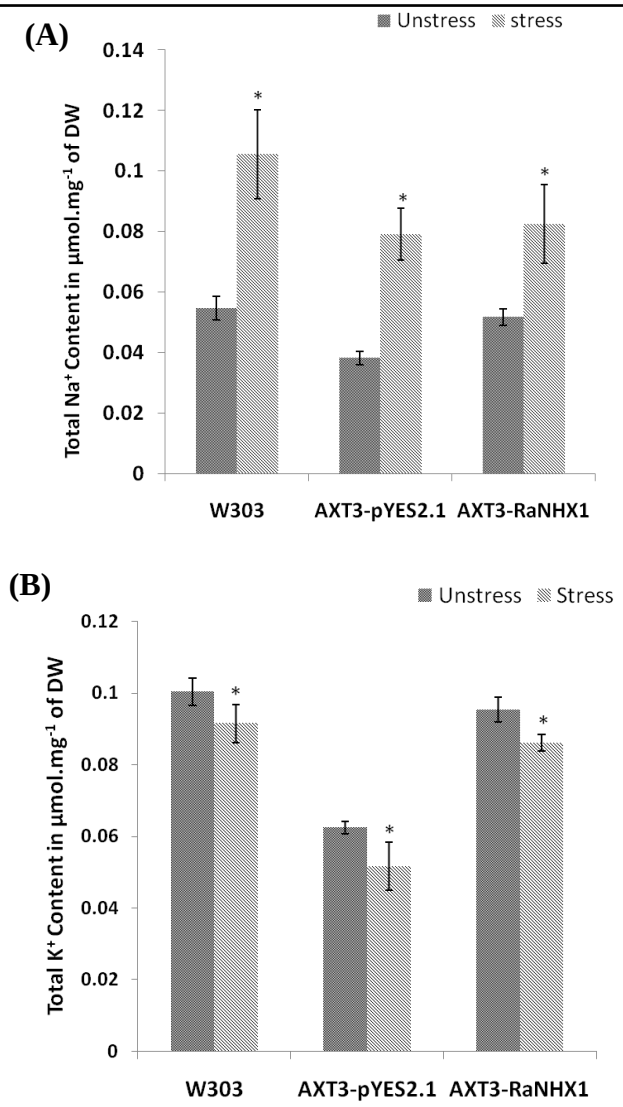


Figure 4.11 Total Na^+ and K^+ analysis (A) Total Na^+ was analyzed in yeast cells under 75 mM NaCl stressed and unstressed conditions (B) Total K^+ was analyzed in yeast cells under 75 mM NaCl stressed and unstressed conditions. Error bars represent the mean ± standard error. Statistical significance are represented with asterisk (*) $p < 0.01$.

4.3.6 RaNHX1 is localized in the stomatal subsidiary and guard cells

The subcellular localization of RaNHX1was determined by fusing *RaNHX1* in the frame to the coding region of green fluorescent protein (GFP) under the control of 35S promoter of CaMV and NOS terminator in the pCAMBIA1302 binary vector (Figure 4.12 A). The resulting pCAMBIA1302 binary construct was transformed into *Nicotiana tabaccum* leaves. The transient expression of a fusion protein RaNHX1 with GFP was observed in tobacco epidermal cells using confocal microscopy. The empty vector used as a control which was localized uniformly on the plasma membrane and the nucleus of tobacco cells (Figure 4.12 B). The GFP signals predominantly were found on the plasma membrane of and stomata of tobacco cells (Figure 4.12 C). Moreover, GFP signals were detected in the stomatal subsidiary cells as shown in the (Figure 4.12 D). Beside, subsidiary cells, GFP detected in the stomatal guard cells (Figure 4.12 E). Figure 4.12 F was the overlapping of above three images (Figure 4.12 C, D and E) where stomatal subsidiary and guard cells along with the plasma membrane all are visible in the same image (Figure 4.12 F). Overall, the localization RaNHX1 fusion protein with GFP in the tobacco cells suggesting a functional role in stomatal aperture movements.

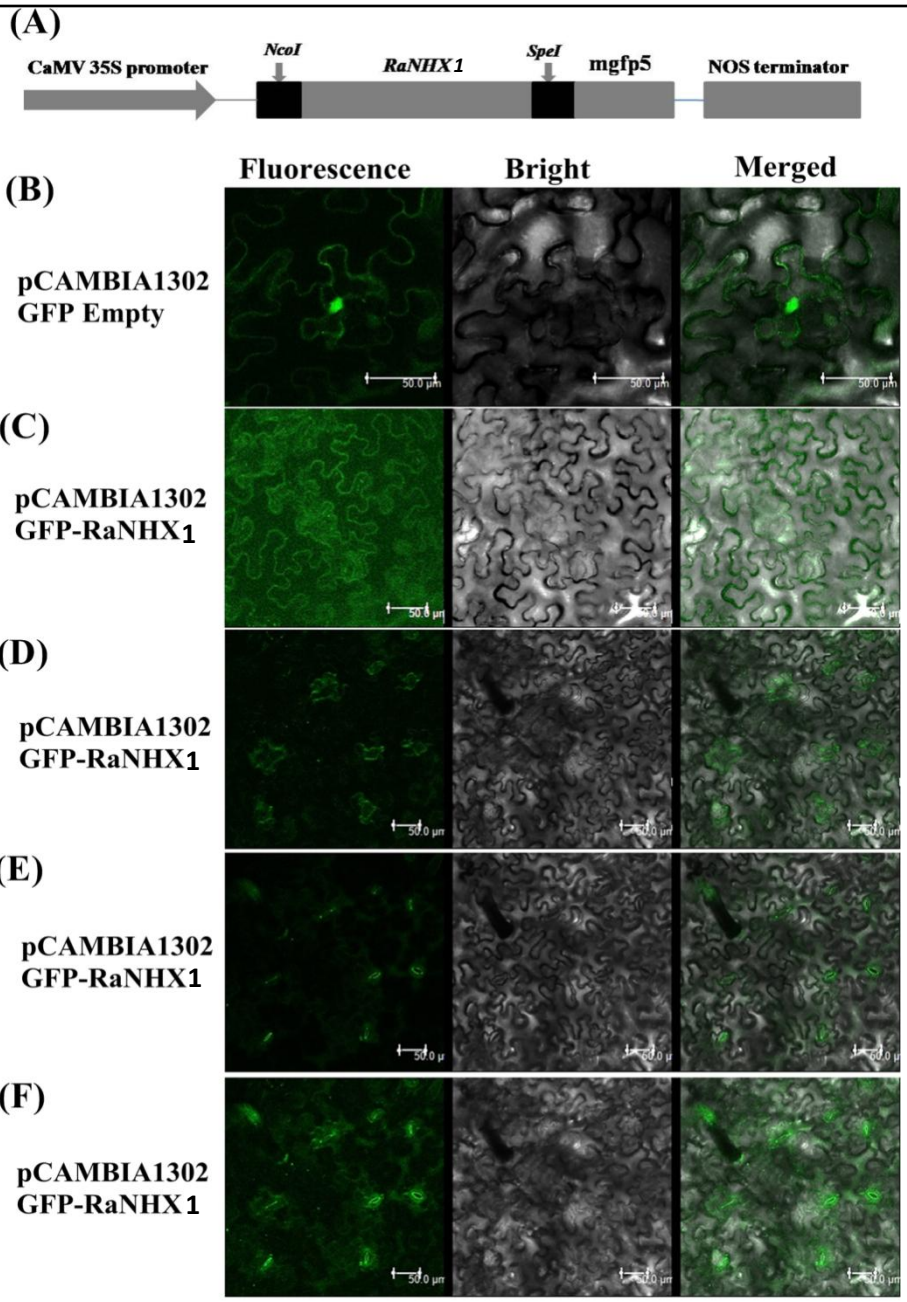


Figure 4.12 Subcellular localization of RaNHX1. A green fluorescence protein (GFP)-coding sequence was fused to *RaNHX1*gene sequence and performed agro-infiltration into tobacco leaves and images were captured after 2 days.(A) Schematic representation of the *RaNHX* gene fused with *mgfp5* and nopaline synthase terminator (NOS) in pCAMBIA 1302 vector (B) pCAMBIA1302 empty vector showed localization of GFP alone in the plasma membrane and in the nucleus (C) pCAMBIA1302::*RaNHX1* fused with *gfp* showed uniform green signals in plasma membrane and stomata (D) GFP is observed in subsidiary cells (E) Localization in the stomata guard cells and (F) This image is the overlapping of above three images where stomatal subsidiary and guard cells along with the plasma membrane are visible in the same image.

4.4 Discussion

The plant NHX members perform diverse roles in the growth and development, regulation of stomatal activity and modulate salinity stress. A series of NHX members were isolated and functionally characterized from glycophytes as well as halophytic plants such as *Arabidopsis*, rice, *Pennisetum*, *Populus*, *Medicago*, *Salicornia*, *Suaeda* and *Mesembryanthemum* species. However, there are no reports available on halophytic mangrove species such as *R. apiculata*. The predicted protein structure, bonafide conserved signature amiloride and ND motif, NHX domain and the presence of transmembrane helices in RaNHX1 strongly supported their identity as the members of NHX. Moreover, a phylogenetic relationship predicted RaNHX1 protein formed a clade with *Arabidopsis* vacuolar NHX members.

Till date, three models are available out of that, *E. coli* NhaA crystal structure is used as a template for homology modeling. Recently, 3D homology models of human NHA2 were generated using EcNhaA as a template (Schushan et al., 2010). Similarly, in the present study, we constructed a homology model of RaNHX1 based on EcNhaA template. The structural analysis of EcNhaA revealed the importance of DD motif (Asp163 and Asp164) in TM5 region, which is a part of the TM4-TM11 assembly and played a crucial role in binding, and transport of ions (Hunte et al., 2005). In *Populus euphratica* NHX3 predicted model had four conserved residues such as Tyr 149 in TM4, Asn 187 and Asp 188 in TM5, and Arg 356 in TM10 (Wang et al., 2014). In the present study we observed that RaNHX1 predicted structure has same conserved residues at the same positions as earlier reported in NHX structure. The homology model study revealed RaNHX1 is also involved in the ion binding, translocation, and transport of ions across the membrane.

R. apiculata developed a unique mechanism to cope with these harsh conditions such as ultrafiltration and accumulation of Na^+ in leaf tissues (Menon and Soniya, 2014). In this context, we tried to explore the gene regulation mechanism of *RaNHX1* in salt stress at the different time point. We observed that *RaNHX1* significantly expressed in all vegetative tissues such as leaves, stems, and roots while low expression was detected in the flower. Similarly, *Arabidopsis AtNHX1* transcript accumulation was observed in roots, stems, leaves and flower tissues (Apse et al., 1999). Mangrove associates halophytic wild rice species, *Porteresia coarctata PcNHX1* showed differential transcript abundance in leaves and roots (Kizhakkedath et al., 2015). Expression pattern of *Medicago* NHX members showed transcript abundance in all vegetative tissues such as root, stem, leaf, flower, and pod (Sandhu et al., 2017). Moreover, under salinity stress, *RaNHX1* showed differential expression pattern in tissue as well as time-dependent manner. *Oryza sativa NHX1* expression was induced by high concentrations of NaCl and KCl in roots and stem (Fukuda et al., 2004). Earlier reports on *R. apiculata* salt specific gene expressions were performed at 6 h time point in leaves (Menon and Soniya, 2014). Similarly, *Populus NHX* expression was significantly observed in roots compared to stems and leaves after 6 h of salt stress (Wang et al., 2014). Most of plant vacuolar NHX expression studies showed significant expression pattern under salt stress including *Arabidopsis AtNHX1*, *Oryza sativa OsNHX1*, *Medicago NHX*, *Populus NHX*, *Suaeda salsa NHX*, and *Atriplex gmelini NHX* (Apse et al., 1999; Hamada et al., 2001; Fukuda et al., 2004; Ma et al., 2004; Wang et al., 2014; Sandhu et al., 2017).

Total Na^+ and K^+ analysis in *R. apiculata* tissues at different time points were analyzed. As previously reported *R. apiculata* accumulates excess Na^+ in leaves and is the best-evolved

strategy for removing surplus ions (Menon and Soniya, 2014). Moreover, *R. apiculata* root allows restricted entry of Na^+ attributed to ultrafiltration mechanism (Menon and Soniya, 2014). Based on the present study, we found that leaves are major Na^+ storage organ in *R. apiculata* species and there were no significant Na^+ level fluctuation in other tissues. Moreover, the K^+ level didn't altered significantly at other time points except at 24 h. Interestingly, in the leaves higher Na^+ content was observed at 24 h but roots showed accumulation at early time-course about 6 h. It might be possible, as roots sense high Na^+ level in rhizosphere which enters into root through various transporters and further translocate towards the leaves. Beside this, the stem may act as Na^+ conduit from root to leaves through vascular tissues which might lead to low Na^+ accumulation in the stem. Yadav et al. (2012) reported that *S. brachiata SOS1* encodes a plasma membrane Na^+/H^+ antiporter, which enabled Na^+ loading to xylem from root and leaf tissues.

Many reports on the heterologous expression of several plants vacuolar NHX in yeast mutant strain AXT3 partially complementing the hygromycin-B antibiotic, NaCl, and KCl phenotypes were available (Mishra et al., 2014). The mungbean *NHX*, Populus *NHX3*, and *OsNHX* partially complemented AXT3 mutant strain (Kinclova-Zimmermannova et al., 2004; Mishra et al., 2014; Wang et al., 2014). In the current study, we used yeast strain AXT3K mutant (*ena1-4*Δ*, nha1*Δ*, nhx1*Δ) to partial complement *RaNHX1*. It partially complemented sensitive phenotypes of AXT3 under hygromycin-B, NaCl, and KCl stress. It helps to conclude that, *RaNHX1* has a dual role in Na^+ and K^+ transport activity on the vacuolar membrane. However, as we know that the established role of *AtNHX1* was regulation of Na^+ transport but at the same time, it also regulates K^+ homeostasis under normal conditions (Venema et al., 2002). The *Arabidopsis* double mutant *nhx1nhx2* line can

retain only 30% of the K^+ level in vacuoles as compared to the WT plant, indicated that both *NHX1* and *NHX2* are involved in K^+/H^+ exchange (Basil et al., 2011). *AtNHX1* mediates Na^+/H^+ and K^+/H^+ exchanges at similar rates, whereas *TaNHX2* exhibits a slight preference for K^+/H^+ exchanges over Na^+/H^+ exchanges (Bassil et al., 2011). We observed that slight alkaline condition showed significant phenotypes in AXT3-*RaNHX1* compared to AXT3-pYES2.1. We inferred that RaNHX may have better transport function in alkaline condition over acidic condition. Similar, results were observed where *Arabidopsis* cation/H^+ exchanger (CHX20) complemented into Δ*nhx1* Δ*nha1* Δ*kha1* and Δ*ena1-4* mutant yeast strain KTA40-2 maintained K^+ homeostasis and influences pH under certain conditions (Padmanaban et al., 2007).

Subcellular localization of RaNHX1 was shown to be localized in a stomatal subsidiary, guard cells, and plasma membrane. The physiological roles of *AtNHX1* and *AtNHX2* were elucidated using the double knockout mutant line, which showed that the vacuolar localized NHX is present in the stomatal guard cells and they are involved in the vacuolar accumulation of K^+. Moreover, the double mutant lines exhibited stomatal opening and closing dysfunction (Andrés et al., 2014). Compilation of these data helps to confirm successfully isolation of full-length Na^+, K^+/H^+ antiporter gene, which plays an important role in vacuolar sequestration of Na^+/K^+ and they are involved in a stomatal activity. Moreover, isolated NHX responded to salt stress by altering transcript accumulation in different tissues. This is the first report of *R. apiculata* NHX members which may further enable to understand the salt stress regulatory mechanism in halophytic plants.

4.5 Conclusion

A full-length *RaNHX1* gene from *R. apiculata* which encodes Na^+, K^+/H^+ antiporter were isolated. Conserved domain, motif and transmembrane helix confirmed their relationship with plant NHX family members. Phylogenetic relationship and 3D homology modeling of RaNHX1 substantiated close relation with AtNHX1 and AtNHX2. Transcript accumulations of *RaNHX1* were induced by 250 mM salt stress in leaves, roots, and stems. Elemental analysis revealed that leaves were a major Na^+ storage organ in *R. apiculata* and elements were significantly enhanced in salinity stress. Isolated gene also partial complements AXT3 sensitive phenotypes under hygromycin-B, NaCl and KCl treatments. Moreover, subcellular localization helps to elucidate their physiological function during stomatal movements, regulated by Na^+, K^+/H^+ antiporter. In future, it will be interesting to reveal more physiological functions of NHX members besides their roles in abiotic stresses.

CHAPTER 5

Summary and Future scope

CHAPTER 5

Summary

There are two main objectives of the present thesis, first to assign the Goan mangrove species based on the potential DNA barcode and the second is cloning and functional characterization of NHX from *Rhizophora apiculata*. To identify mangroves species based on the molecular markers, we used plastid and nuclear markers such as *rbcL*, *matK*, ITS2, *atpF-atpH*, *rpoC1*, and *psbK-psbI*. Moreover, we have assessed their efficacy individual as well as concatenated forms. We found *rbcL* and ITS2 barcodes were easy to amplify and higher sequencing recovery compared to *matK*. In identification context, *matK* and ITS2 are potential DNA barcode to demarcate mangroves at the species level. *RbcL* barcode offered ease of PCR amplification, and sequence recovery but low species discrimination. Besides this, single barcode *matK* is sufficient to resolve *A. ilicifolius*, *A. corniculatum*, *E. agallocha*, *Ceriops tagal*, *K. candel*, *B. cylindrica* and *B. gymnorrhiza*. ITS2 was able to discriminate *R. apiculata* and *R. mucronata* species based on GMYC method, while *A. alba* was resolved by concatenation of *matK*+ITS2. A cryptic genus *Avicennia* was delimitated based on the *atpF-atpH* single barcode. In the present work, the foundation work was done towards DNA barcoding of mangroves plant genera such as *Rhizophora*, *Avicennia*, *Acanthus*, *Kandelia*, *Ceriops*, *Bruguiera*, *Aegiceras* and *Excoecaria*. Compiled mangroves barcoding result had some limitations, most of which are due to the low mangrove taxa sample coverage.

Besides DNA barcode work, we also focused on cloning of RaNHX, gene expression analysis and heterologous characterization from *Rhizophora apiculata*. Before qRT-PCR analysis, we have performed selection and validation of candidate reference genes for gene

expression analysis. We strongly recommend *EF1α* followed by *ACT* and *β-TUB* as the most stable candidate reference genes for normalization in *R. apiculata* physiological tissue gene expression analysis. Under salt stress, *EF1α* followed by *ACT* and 18S are the most suitable candidate reference genes for normalization. A full-length *RaNHX1* gene from *R. apiculata* which encodes Na^+, K^+/H^+ antiporter were isolated. Conserved domain, motif and transmembrane helix confirmed their relationship with plant NHX family members. Phylogenetic relationship and 3D homology modeling of RaNHX1 substantiated close relation with AtNHX1 and AtNHX2. In order to understand the role of *RaNHX1* in salt stress, we have performed qRT-PCR and used 18S rRNA for normalization. We observed that *RaNHX1* was expressed in the almost all vegetative tissues such as leaves, stems and roots but low expression was detected in the flower. Salt stress triggered the transcript accumulations in leaves, roots, and stems which suggested the regulatory roles of *RaNHX1* during salinity stress. In order to understand the management of surplus Na^+ inside *R. apiculata*, we performed total Na^+ and K^+ analysis. Elemental analysis revealed that leaves were a major Na^+ storage organ in *R. apiculata* and significantly enhanced in salinity stress. Surprisingly, K^+ content is far lower than Na^+ content in all tissues.

A full-length *RaNHX1* gene was functionally characterized by heterologous expression in the AXT3 mutant. RaNHX1 partial complemented the AXT3 mutant sensitive phenotypes under hygromycin-B, NaCl and KCl treatments. Similar results were observed in liquid growth medium supplemented with various concentrations of NaCl and KCl and recorded OD at 600 nm. Moreover, we have observed the significant RaNHX1 activity in alkaline conditions but no effect in acidic conditions. Subcellular localization of RaNHX1 in tobacco leaves showed presence of GFP signal in plasma membrane, subsidiary and guard cells of

stomata. This suggested that RaNHX1 helps to elucidate their physiological function during stomatal movements, regulated by Na^+, K^+/H^+ antiporter.

Future scope of the study

1. This thesis mainly focused on the DNA barcoding of Goa mangroves which comprises 14 species, identified based on the core DNA barcode *rbcL, matK* and ITS2 with supplementary markers such as *atpF-atpH, rpoC1*, and *psbK-psbI* for *Avicennia* genus. This study paved the way for mangroves species identification based on DNA barcode, which would be utilized for evaluation of Indian and world mangrove vegetation.
2. The salt tolerance mechanism in *Rhizophora apiculata* is not fully understood, so it would be very important to explore the basic salt tolerance mechanism and related pathways in mangroves.
3. We have selected and validated the candidate reference genes for qRT-PCR which would help gene expression study in the *Rhizophora apiculata* species.
4. *RaNHX1* gene was cloned and functionally characterized in the AXT3 mutant yeast strain, which partially complemented NaCl and KCl sensitive phenotypes. Moreover, it would be very important to characterize *RaNHX1* function in plants by overexpression in the *Arabidopsis thaliana*. Moreover, it would be very helpful to explore the RaNHX1 regulatory mechanism through hunting interacting protein partners.
5. Subcellular localization of RaNHX1 on the plasma membrane, subsidiary and guard cells of stomatal cells. This study will open the new horizon and regulatory mechanism of RaNHX1 in the stomatal cells.

References

References

1. Agarwal, M., Shrivastava, N. and Padh, H., 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant cell reports*, *27*, 617-631.
2. Aharon, G.S., Apse, M.P., Duan, S., Hua, X. and Blumwald, E., 2003. Characterization of a family of vacuolar Na^+/H^+ antiporters in Arabidopsis thaliana. *Plant and Soil*, *253*, 245-256.
3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology*, *215*, 403-410.
4. Álvarez, I. and Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular phylogenetics and evolution*, *29*, 417-434.
5. Andersen, C.L., Jensen, J.L. and Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research*, *64*, 5245-5250.
6. Andrés, Z., Pérez-Hormaeche, J., Leidi, E.O., Schlücking, K., Steinhorst, L., McLachlan, D.H., Schumacher, K., Hetherington, A.M., Kudla, J., Cubero, B. and Pardo, J.M., 2014. Control of vacuolar dynamics and regulation of stomatal aperture by tonoplast potassium uptake. *Proceedings of the National Academy of Sciences*, *111*, E1806-E1814.
7. Apse, M.P., Aharon, G.S., Snedden, W.A. and Blumwald, E., 1999. Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in Arabidopsis. *Science*, *285*, 1256-1258.
8. Apse, M.P., Sottosanto, J.B. and Blumwald, E., 2003. Vacuolar cation/H^+ exchange, ion homeostasis, and leaf development are altered in a T-DNA insertional mutant of *AtNHX1*, the *Arabidopsis* vacuolar Na+/H+ antiporter. *The plant journal*, *36*, 229-239.
9. Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Veuille, M. and Laredo, C., 2009. DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC bioinformatics*, *10*, 10.
10. Bao, A.K., Du, B.Q., Touil, L., Kang, P., Wang, Q.L. and Wang, S.M., 2016. Co-expression of tonoplast Cation/H^+ antiporter and H^+-pyrophosphatase from xerophyte *Zygophyllum xanthoxylum* improves alfalfa plant growth under salinity, drought and field conditions. *Plant biotechnology journal*, *14*, 964-975.

11. Bao-Yan, A.N., Yan, L.U.O., Jia-Rui, L.I., Wei-Hua, Q.I.A.O., Zhang, X.S. and Xin-Qi, G.A.O., 2008. Expression of a vacuolar Na^+/H^+ antiporter gene of alfalfa enhances salinity tolerance in transgenic Arabidopsis. *Acta Agronomica Sinica*, *34*, 557-564.
12. Bassil, E. and Blumwald, E., 2014. The ins and outs of intracellular ion homeostasis: NHX-type cation/H^+ transporters. *Current opinion in plant biology*, *22*, 1-6.
13. Bassil, E., Tajima, H., Liang, Y.C., Ohto, M.A., Ushijima, K., Nakano, R., Esumi, T., Coku, A., Belmonte, M. and Blumwald, E., 2011. The Arabidopsis Na^+/H^+ antiporters *NHX1* and *NHX2* control vacuolar pH and K^+ homeostasis to regulate growth, flower development, and reproduction. *The Plant Cell*, *23*, 3482-3497.
14. Bhadalkar, A., Vyas, B. and Bhatt, S., 2014. Molecular Characterization of mangroves plants "A Review". *Life Sciences Leaflets*, *49*, 148-158.
15. Bhaskaran, S. and Savithramma, D.L., 2011. Co-expression of *Pennisetum glaucum* vacuolar Na^+/H^+ antiporter and *Arabidopsis* H^+-pyrophosphatase enhances salt tolerance in transgenic tomato. *Journal of experimental botany*, *62*, 5561-5570.
16. Brini, F., Hanin, M., Mezghani, I., Berkowitz, G.A. and Masmoudi, K., 2007. Overexpression of wheat Na^+/H^+ antiporter *TaNHX1* and H^+-pyrophosphatase TVP1 improve salt-and drought-stress tolerance in *Arabidopsis thaliana* plants. *Journal of experimental botany*, *58*, 301-308.
17. Bustin, S.A., 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of molecular endocrinology*, *29*, 23-39.
18. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. and Vandesompele, J., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, *55*, 611-622.
19. Caballero, A., Quesada, H. and Rolán-Alvarez, E., 2008. Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, *179*, 539-554.
20. Cao, J., Wang, L. and Lan, H., 2016. Validation of reference genes for quantitative RT-PCR normalization in *Suaeda aralocaspica*, an annual halophyte with heteromorphism and C4 pathway without Kranz anatomy. *Peer J*, *4*, e1697.

21. Cavanagh, C., Morell, M., Mackay, I. and Powell, W., 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current opinion in plant biology*, *11*, 215-221.
22. CBOL Plant Working Group, Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L. and Fazekas, A.J., 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*, 12794-12797.
23. Chalmers, K.J., Waugh, R., Sprent, J.I., Simons, A.J. and Powell, W., 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. *Heredity*, *69*, 465.
24. Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N. and Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*, 1889-1895.
25. Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.L. and Kron, K.A., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden*, *80*, 528-580.
26. Chauhan, S., Forsthoefel, N., Ran, Y., Quigley, F., Nelson, D.E. and Bohnert, H.J., 2000. Na^+/myo-inositol symporters and Na^+/H^+ antiporter in *Mesembryanthemum crystallinum*. *The Plant Journal*, *24*, 511-522.
27. Chen, J., Zhao, J., Erickson, D.L., Xia, N. and Kress, W.J., 2015. Testing DNA barcodes in closely related species of Curcuma (Zingiberaceae) from Myanmar and China. *Molecular ecology resources*, *15*, 337-348.
28. Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X. and Luo, K., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one*, *5*, e8613.
29. China BOL, Li, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., Chen, Z.D., Zhou, S.L., Chen, S.L. and Yang, J.B., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core

barcode for seed plants. *Proceedings of the National Academy of Sciences*, *108*, 19641-19646.

30. Chinnusamy, V. and Zhu, J.K., 2009. Epigenetic regulation of stress responses in plants. *Current opinion in plant biology*, *12*, 133-139.
31. Chitale, V.S., Behera, M.D. and Roy, P.S., 2014. Future of endemic flora of biodiversity hotspots in India. *PloS one*, *9*, e115264.
32. Cruz, F., Kalaoun, S., Nobile, P., Colombo, C., Almeida, J., Barros, L.M., Romano, E., Grossi-de-Sá, M.F., Vaslin, M. and Alves-Ferreira, M., 2009. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Molecular Breeding*, *23*, 607-616.
33. Cuénoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J. and Chase, M.W., 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *American Journal of Botany*, *89*(1), pp.132-144.
34. D'Souza, J. and Rodrigues, B.F., 2013. Biodiversity of Arbuscular Mycorrhizal (AM) fungi in mangroves of Goa in West India. *Journal of forestry research*, *24*, 515-523.
35. Das, S.S., Das, S. and Ghosh, P., 2014. Plylogenetic relationships among three species of the mangrove genus *Avicennia* found in Indian Sundarban as revealed by RAPD analysis. *Asian Journal of Plant Science Research*, *4*, 25-30.
36. Dhargalkar, V.K., D'Souza, R., Kavlekar, D.P., Untawale, A.G., 2014. Mangroves of Goa. Forest department, Government of Goa and Mangroves society of India, Goa, India, 12-31.
37. Dheda, K., Huggett, J.F., Chang, J.S., Kim, L.U., Bustin, S.A., Johnson, M.A., Rook, G.A.W. and Zumla, A., 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Analytical biochemistry*, *344*, 141-143.
38. Drager, R.G. and Hallick, R.B., 1993. A novel *Euglena gracilis* chloroplast operon encoding four ATP synthase subunits and two ribosomal proteins contains 17 introns. *Current genetics*, *23*, 271-280.
39. Drummond, A.J. and Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, *7*, 214.

40. Drummond, A.J., Ho, S.Y., Phillips, M.J. and Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS biology*, *4*, e88.
41. Duke CN (1991) A Systematic Revision of the Mangrove Genus *Avicennia* (Avicenniaceae) in Australasia. *Australian Systematic Botany*, *4*, 299-324
42. Duke, C.N., Bunt, S.J., 1979. The Genus *Rhizophora* (Rhizophoraceae) in North-eastern Australia. *Australian Journal of Botany*, *27*, 657-678.
43. Duke, C.N., Jackes, R.B., 1987. A systematic revision of the mangrove genus *Sonneratia* (Sonneratiaceae) in Australasia. Blumea, *32*, 277-302.
44. Duke, N.C. and Ge, X.J., 2011. *Bruguiera* (Rhizophoraceae) in the Indo-West Pacific: a morphometric assessment of hybridization within single-flowered taxa. *Blumea-Biodiversity, Evolution and Biogeography of Plants*, *56*, 36-48.
45. Duke, N.C. and Ge, X.J., 2011. *Bruguiera* (Rhizophoraceae) in the Indo-West Pacific: a morphometric assessment of hybridization within single-flowered taxa. *Blumea-Biodiversity, Evolution and Biogeography of Plants*, *56*, 36-48.
46. Duke, N.C., 1992. Tropical mangrove ecosystems (Coastal and Estuarine studies). In: Robertson AI, Alongi DM (eds) American geophysical union, Washington, 63–100.
47. Dumas, P., Barbut, J., Le Ru, B., Silvain, J.F., Clamens, A.L., d'Alençon, E. and Kergoat, G.J., 2015. Phylogenetic molecular species delimitations unravel potential new species in the pest genus *Spodoptera Guenée*, 1852 (Lepidoptera, Noctuidae). *PLoS one*, *10*, e0122407.
48. Ezard, T., Fujisawa, T. and Barraclough, T.G., splits: SPecies' LImits by Threshold Statistics. 2009. R package version. 1.0-14/r31.
49. Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M. and Barrett, S.C., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *Plos one*, *3*(7), e2802.
50. Fu, X., Deng, S., Su, G., Zeng, Q. and Shi, S., 2004. Isolating high-quality RNA from mangroves without liquid nitrogen. *Plant Molecular Biology Reporter*, *22*(2), 197a.
51. Fukuda, A., Nakamura, A. and Tanaka, Y., 1999. Molecular cloning and expression of the Na^+/H^+ exchanger gene in Oryza sativa. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1446*, 149-155.

52. Fukuda, A., Nakamura, A., Tagiri, A., Tanaka, H., Miyao, A., Hirochika, H. and Tanaka, Y., 2004. Function, intracellular localization and the importance in salt tolerance of a vacuolar Na^+/H^+ antiporter from rice. *Plant and Cell Physiology*, *45*, 146-159.
53. Gernhard, T., 2008. The conditioned reconstructed process. *Journal of theoretical biology*, *253*, 769-778.
54. Gouiaa, S., Khoudi, H., Leidi, E.O., Pardo, J.M. and Masmoudi, K., 2012. Expression of wheat Na^+/H^+ antiporter *TNHXS1* and H^+-pyrophosphatase *TVP1* genes in tobacco from a bicistronic transcriptional unit improves salt tolerance. *Plant molecular biology*, *79*, 137-155.
55. Guan, B., Hu, Y., Zeng, Y., Wang, Y. and Zhang, F., 2011. Molecular characterization and functional analysis of a vacuolar Na^+/H^+ antiporter gene (*HcNHX1*) from *Halostachys caspica*. *Molecular biology reports*, *38*, 1889-1899.
56. Guo, W., Wu, H., Zhang, Z., Yang, C., Hu, L., Shi, X., Jian, S., Shi, S. and Huang, Y., 2017. Comparative analysis of transcriptomes in rhizophoraceae provides insights into the origin and adaptive evolution of mangrove plants in intertidal environments. *Frontiers in plant science*, *8*, 795.
57. Gupta, B. and Huang, B., 2014. Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *International journal of genomics*, *2014*.
58. Hamada, A., Shono, M., Xia, T., Ohta, M., Hayashi, Y., Tanaka, A. and Hayakawa, T., 2001. Isolation and characterization of a Na^+/H^+ antiporter gene from the halophyte *Atriplex gmelini*. *Plant molecular biology*, *46*, 35-42.
59. Hebert, P.D., Cywinska, A. and Ball, S.L., 2003b. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*), 313-321.
60. Hebert, P.D., Ratnasingham, S. and de Waard, J.R., 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*, S96-S99.
61. Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R. and Sauquet, H., 2003. Angiosperm phylogeny based on the *matK* sequence information. *American journal of botany*, *90*, 1758-1776.

62. Hollingsworth, P.M., Graham, S.W. and Little, D.P., 2011. Choosing and using a plant DNA barcode. *PloS one*, *6*, e19254.
63. Holsters, M., De Waele, D., Depicker, A., Messens, E., Van Montagu, M. and Schell, J., 1978. Transfection and transformation of *Agrobacterium tumefaciens*. *Molecular and General Genetics*, *163*,181-187.
64. Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. and Michel, H., 2005. Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH. *Nature*, *435*, 1197.
65. Hutchings, P.A., Saenger, P., 1987. Ecology of mangroves. Australian Ecology Series University of Queensland Press, St Lucia, 14–44.
66. Jagtap, T.G., Chavan, V.S., Untawale, A.G., 1993. Mangrove ecosystems of India: A need for protection. *AMBIO*, *22*, 252-254.
67. Jha, A., Joshi, M., Yadav, N.S., Agarwal, P.K. and Jha, B., 2011. Cloning and characterization of the *Salicornia brachiata* Na^+/H^+ antiporter gene *SbNHX1* and its expression by abiotic stress. *Molecular biology reports*, *38*, 1965-1973.
68. Ji, H., Pardo, J.M., Batelli, G., Van Oosten, M.J., Bressan, R.A. and Li, X., 2013. The Salt Overly Sensitive (SOS) pathway: established and emerging roles. *Molecular plant*, *6*, 275-286.
69. Johnson, L.A. and Soltis, D.E., 1995. Phylogenetic inference in Saxifragaceae sensu stricto and Gilia (Polemoniaceae) using matK sequences. *Annals of the Missouri Botanical Garden*, 149-175.
70. Kathiresan, K. and Bingham, B.L., 2001. Biology of mangroves and mangrove ecosystems. *Advances in Marine Biology*, *40*, 81-251.
71. Kathiresan, K., 2018. Mangrove forests of India. *Current Science*, *114*, 976-981.
72. Kinclova-Zimmermannova, O., Flegelova, H. and Sychrova, H., 2004. Rice Na^+/H^+-antiporter Nhx1 partially complement the alkali-metal-cation sensitivity of yeast strains lacking three sodium transporters. *Folia Microbiology*, *49*, 519-525.
73. Kizhakkedath, P., Jegadeeson, V., Venkataraman, G. and Parida, A., 2015. A vacuolar antiporter is differentially regulated in leaves and roots of the halophytic wild rice *Porteresia coarctata* (Roxb.) Tateoka. *Molecular Biology report*, *42*, 1091-1105.

74. Knauf, U. and Hachtel, W., 2002. The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Molecular genetics and genomics*, *267*, 492-497.
75. Kozera, B. and Rapacz, M., 2013. Reference genes in real-time PCR. *Journal of applied genetics*, *54*, 391-406.
76. Kress, W.J. and Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS one*, *2*, e508.
77. Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences*, *102*(23), 8369-8374.
78. Kumar, K., Kumar, M., Kim, S.R., Ryu, H. and Cho, Y.G., 2013. Insights into genomics of salt stress response in rice. *Rice*, *6*, 27.
79. Kumar, S., Skjæveland, Å., Orr, R.J., Enger, P., Ruden, T., Mevik, B.H., Burki, F., Botnen, A. and Shalchian-Tabrizi, K., 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC bioinformatics*, *10*, 357.
80. Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*, 1870-1874.
81. Lahaye, R.R., Savolainen, V., Duthoit, S., Maurin, O. and Van der Bank, M., 2008. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nature Precedings*, 1-21.
82. Lakshmi, M., Parani, M., Ram, N. and Parida, A., 2000. Molecular phylogeny of mangroves VI. Intraspecific genetic variation in mangrove species *Excoecaria agallocha* L. (Euphorbiaceae). *Genome*, *43*, 110-115.
83. Lakshmi, M., Rajalakshmi, S., Parani, M., Anuratha, C.S. and Parida, A., 1997. Molecular phylogeny of mangroves I. Use of molecular markers in assessing the intraspecific genetic variability in the mangrove species *Acanthus ilicifolius* Linn. (Acanthaceae). *Theoretical and Applied Genetics*, *94*, 1121-1127.

84. Levin, R.A., Wagner, W.L., Hoch, P.C., Nepokroeff, M., Pires, J.C., Zimmer, E.A. and Sytsma, K.J., 2003. Family-level relationships of Onagraceae based on chloroplast *rbcL* and ndhF data. *American Journal of Botany*, *90*, 107-115.
85. Li, X., Duke, N.C., Yang, Y., Huang, L., Zhu, Y., Zhang, Z., Zhou, R., Zhong, C., Huang, Y. and Shi, S., 2016. Re-evaluation of phylogenetic relationships among species of the mangrove genus *Avicennia* from Indo-West Pacific based on multilocus analyses. *PloS one*, *11*, e0164453.
86. Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. and Chen, S., 2015. Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*, 157-166.
87. Little, D.P. and Stevenson, D.W., 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics*, *23*, 1-21.
88. Livak, K.J. and Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods*, *25*, 402-408.
89. Lo, E.Y.Y., 2010. Testing hybridization hypotheses and evaluating the evolutionary potential of hybrids in mangrove plant species. *Journal of evolutionary biology*, *23*, 2249-2261.
90. Lu, S.Y., Jing, Y.X., Shen, S.H., Zhao, H.Y., Ma, L.Q., Zhou, X.J., Ren, Q. and Li, Y.F., 2005. Antiporter gene from *Hordum brevisubulatum* (Trin.) link and its overexpression in transgenic tobaccos. *Journal of Integrative Plant Biology*, *47*, 343-349.
91. Ma, X.L., Zhang, Q., Shi, H.Z., Zhu, J.K., Zhao, Y.X., Ma, C.L. and Zhang, H., 2004. Molecular cloning and different expression of a vacuolar Na^+/H^+ antiporter gene in *Suaeda salsa* under salt stress. *Biologia Plantarum*, *48*, 219-225.
92. MacNae, M. 1968. A general account of the fauna and flora of mangrove swamps and forest in the Indo-west pacific region. *Advanced marine biology*, *6*, 73–270.
93. Maniatis, T., Fritsch, E.F. and Sambrook, J., 1982. *Molecular cloning: a laboratory manual* (Vol. I). Cold Spring Harbor, NY: Cold spring harbor laboratory. 5.1–5.17.
94. McNeal, J.R., Kuehl, J.V., Boore, J.L. and de Pamphilis, C.W., 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus Cuscuta. *BMC Plant Biology*, *7*, 57.

95. Meier, R., Shiyang, K., Vaidya, G. and Ng, P.K., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic biology*, *55*, 715-728.
96. Meng, B.Y., Wakasugi, T. and Sugiura, M., 1991. Two promoters within the psbK-psbI-trnG gene cluster in tobacco chloroplast DNA. *Current genetics*, *20*(3), 259-264.
97. Menon, T.G. and Soniya, E.V., 2014. Isolation and characterization of salt-induced genes from *Rhizophora apiculata* Blume, a true mangrove by suppression subtractive hybridization. *Current Science*, *107*, 650-655.
98. Mishra, S., Alavilli, H., Lee, B.H., Panda, S.K. and Sahoo, L., 2015. Cloning and characterization of a novel vacuolar Na+/H+ antiporter gene (VuNHX1) from drought hardy legume, cowpea for salt tolerance. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *120*, 19-33.
99. Mishra, S., Alavilli, H., Lee, B.H., Panda, S.K. and Sahoo, L., 2014. Cloning and functional characterization of a vacuolar Na^+/H^+ antiporter gene from mungbean (*VrNHX1*) and its ectopic expression enhanced salt tolerance in *Arabidopsis thaliana*. *PloS one*, *9*, e106678.
100. Mittler, R. and Blumwald, E., 2010. Genetic engineering for modern agriculture: challenges and perspectives. *Annual review of plant biology*, *61*, 443-462.
101. Mondini, L., Noorani, A. and Pagnotta, M.A., 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, *1*, 19-35.
102. Mukherjee, A.K., Acharya, L., Panda, P.C. and Mohapatra, T., 2006. Assessment of genetic diversity in 31 species of mangroves and their associates through RAPD and AFLP markers. *Zeitschrift für Naturforschung C*, *61*, 413-420.
103. Munns, R. and Tester, M., 2008. Mechanisms of salinity tolerance. *Annual Review of Plant Biology*, *59*, 651-681.
104. Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A. and Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853.
105. Naskar, K. and Mandal, R., 1999. *Ecology and biodiversity of Indian mangroves* (Vol. 1). Daya Books. New Delhi.
106. Negrão, S., Schmöckel, S.M. Tester, M., 2017.Evaluating physiological responses of plants to salinity stress. *Annals of Botany*, *119*, 1-11.

107. Newmaster, S.G., Fazekas, A.J. and Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Botany*, *84*, 335-341.
108. Newmaster, S.G., Fazekas, A.J., Steeves, R.A.D. and Janovec, J., 2008. Testing candidate plant barcode regions in the Myristicaceae. *Molecular ecology resources*, *8*(3), 480-490.
109. Nikalje, G.C., Srivastava, A.K., Sablok, G., Pandey, G.K., Nikam, T.D. and Suprasanna, P., 2018. Identification and validation of reference genes for quantitative real-time PCR under salt stress in a halophyte, *Sesuvium portulacastrum*. *Plant Gene*, *13*, 18-24.
110. Oh, D.H., Leidi, E., Zhang, Q., Hwang, S.M., Li, Y., Quintero, F.J., Jiang, X., D'Urzo, M.P., Lee, S.Y., Zhao, Y. and Bahk, J.D., 2009. Loss of halophytism by interference with SOS1 expression. *Plant physiology*, *151*, 210-222.
111. Padmanaban, S., Chanroj, S., Kwak, J.M., Li, X., Ward, J.M., Sze, H., 2007. Participation of endomembrane cation/H^+ exchanger *AtCHX20* in osmoregulation of guard cells. *Plant Physiology*, *144*, 82-93.
112. Pang, X., Song, J., Zhu, Y., Xie, C. and Chen, S., 2010. Using DNA barcoding to identify species within Euphorbiaceae. *Planta medica*, *76*, 1784-1786.
113. Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289-290.
114. Parani, M., Lakshmi, M., Elango, S., Ram, N., Anuratha, C.S. and Parida, A., 1997a. Molecular phylogeny of mangroves II. Intra-and inter-specific variation in *Avicennia* revealed by RAPD and RFLP markers. *Genome*, *40*, 487-495.
115. Parani, M., Lakshmi, M., Senthil kumar, P., Ram, N. and Parida, A., 1998. Molecular phylogeny of mangroves V. Analysis of genome relationships in mangrove species using RAPD and RFLP markers. *Theoretical and Applied Genetics*, *97*, 617-625.
116. Parani, M., Rao, C.S., Mathan, N., Anuratha, C.S., Narayanan, K.K. and Parida, A., 1997b. Molecular Phylogeny of mangroves III Parentage analysis of a Rhizophora hybrid using random amplified polymorphic DNA and restriction fragment length polymorphism markers. *Aquatic Botany*, *58*, 165-172.
117. Pehlivan, N., Sun, L., Jarrett, P., Yang, X., Mishra, N., Chen, L., Kadioglu, A., Shen, G. and Zhang, H., 2016. Co-overexpressing a plasma membrane and a vacuolar membrane sodium/proton antiporter significantly improves salt tolerance in transgenic Arabidopsis plants. *Plant and Cell Physiology*, *57*, 1069-1084.

118. Pellino, M., Sharbel, T.F., Mau, M., Amiteye, S. and Corral, J.M., 2011. Selection of reference genes for quantitative real-time PCR expression studies of microdissected reproductive tissues in apomictic and sexual Boechera. *BMC research notes*, *4*, 303.

119. Pennisi, E., 2007. Wanted: a barcode for plants. *Science*, *318*, 190-191.

120. Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology letters*, *26*, 509-515.

121. Pires, I.S., Negrão, S., Pentony, M.M., Abreu, I.A., Oliveira, M.M. and Purugganan, M.D., 2013. Different evolutionary histories of two cation/proton exchanger gene families in plants. *BMC plant biology*, *13*, 97.

122. Pittman, J., 2012. Multiple transport pathways for mediating intracellular pH homeostasis: the contribution of H^+/ion exchangers. *Frontiers in plant science*, *3*, 11.

123. Powell, W., 1992. Plant genomes, gene markers, and linkage maps. Biotechnology and crop improvement in Asia. Patancheru, India. International Crops Research Institute for the Semi-Arid Tropics, 297-322.

124. Puillandre, N., Lambert, A., Brouillet, S. and Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular ecology*, *21*, 1864-1877.

125. Ragavan, P., Saxena, A., Jayaraj, R.S.C., Mohan, P.M., Ravichandran, K., Saravanan, S. and Vijayaraghavan, A., 2016. A review of the mangrove floristics of India. *Taiwania*, *61*, 224-242.

126. Rajagopal, D., Agarwal, P., Tyagi, W., Singla-Pareek, S.L., Reddy, M. K., Sopory, S.K., 2007. *Pennisetum glaucum* Na^+/H^+ antiporter confer high level of salinity tolerance in transgenic *Brassica juncea*. *Molecular Breeding*, *19*, 137-151.

127. Ratnasingham, S. and Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes*, *7*(3), 355-364.

128. Reddy, P.S., Reddy, D.S., Sharma, K.K., Bhatnagar-Mathur, P. and Vadez, V., 2015. Cloning and validation of reference genes for normalization of gene expression studies in pearl millet [*Pennisetum glaucum* (L.) R. Br.] by quantitative real-time PCR. *Plant Gene*, *1*, 35-42.

129. Reguera, M., Bassil, E. and Blumwald, E., 2014. Intracellular NHX-type cation/H^+ antiporters in plants. *Molecular plant*, *7*, 261-263.
130. Robertson AI, Alongi DM. 1992. Tropical Mangrove ecosystems. Wiley Online Library DOI 10.1029/CE041.
131. Rodríguez-Rosales, M.P., Jiang, X., Gálvez, F.J., Aranda, M.N., Cubero, B. and Venema, K., 2008. Overexpression of the tomato K^+/H^+ antiporter *LeNHX2* confers salt tolerance by improving potassium compartmentalization. *New Phytologist*, *179*, 366-377.
132. Ross, H.A., Murugan, S. and Sibon Li, W.L., 2008. Testing the reliability of genetic methods of species identification via simulation. *Systematic biology*, *57*, 216-230.
133. Rozas, J., 2009. Polymorphism Analysis using DnaSP. In: Posada D (ed) Bioinformatics for DNA sequence analysis; methods in molecular biology series. Humana Press, New Jersey, *537*, 337–350.
134. Sahu, S.K., Singh, R. and Kathiresan, K., 2016. Multi-gene phylogenetic analysis reveals the multiple origin and evolution of mangrove physiological traits through exaptation. *Estuarine, Coastal and Shelf Science*, *183*, 41-51.
135. Sandhu, D., Pudussery, M.V., Kaundal, R., Suarez, D.L., Kaundal, A. and Sekhon, R.S., 2018. Molecular characterization and expression analysis of the Na^+/H^+ exchanger gene family in *Medicago truncatula*. *Functional & integrative genomics*, *18*, 141-153.
136. Sanyal, P., Mandal, R.N., Ghosh, D., Naskar, K.R., 1998. Studies on the mangrove patch at Subarnarekha river mouth of Orissa state. *Journal Interacademy*, *2*, 140-149.
137. Schushan, M., Xiang, M., Bogomiakov, P., Padan, E., Rao, R. and Ben-Tal, N., 2010. Model-guided mutagenesis drives functional studies of human NHA2, implicated in hypertension. *Journal of Molecular Biology*, *396*, 1181-1196.
138. Schwarzbach, A.E. and Ricklefs, R.E., 2000. Systematic affinities of Rhizophoraceae and Anisophylleaceae, and intergeneric relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. *American Journal of Botany*, *87*, 547-564.
139. Serino, G. and Maliga, P., 1998. RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiology*, *117*, 1165-1170.

140. Setoguchi, H., Kosuge, K. and Tobe, H., 1999. Molecular phylogeny of Rhizophoraceae based on *rbcL* gene sequences. *Journal of Plant Research*, *112*, 443-455.
141. Shen, G., Wei, J., Qiu, X., Hu, R., Kuppu, S., Auld, D., Blumwald, E., Gaxiola, R., Payton, P. and Zhang, H., 2015. Co-overexpression of *AVP1* and *AtNHX1* in cotton further improves drought and salt tolerance in transgenic cotton plants. *Plant Molecular Biology Reporter*, *33*,167-177.
142. Shi, H. and Zhu, J.K., 2002. Regulation of expression of the vacuolar Na^+/H^+ antiporter gene *AtNHX1* by salt stress and abscisic acid. *Plant molecular biology*, *50*, 543-550.
143. Shi, S., Huang, Y., Zeng, K., Tan, F., He, H., Huang, J. and Fu, Y., 2005. Molecular phylogenetic analysis of mangroves: independent evolutionary origins of vivipary and salt secretion. *Molecular Phylogenetics and Evolution*, *34*, 159-166.
144. Silver, N., Best, S., Jiang, J. and Thein, S.L., 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC molecular biology*, *7*, 33.
145. Singh, J.S. and Chaturvedi, R.K., 2017. Diversity of ecosystem types in India: a review. *Proceedings of the Indian National Science Academy*, *83*, 569-594.
146. Sinha, P., Singh, V.K., Suryanarayana, V., Krishnamurthy, L., Saxena, R.K. and Varshney, R.K., 2015. Evaluation and validation of housekeeping genes as reference for gene expression studies in pigeon pea (*Cajanus cajan*) under drought stress conditions. *PloS one*, *10*, e0122847.
147. Spalding M, Kainuma M, Collins L: World atlas of mangroves. Earthscan eBook, 2010.
148. Sun, M. and Lo, E.Y., 2011. Genomic markers reveal introgressive hybridization in the Indo-West Pacific mangroves: a case study. *PLoS One*, *6*, e19671.
149. Tang, C.Q., Humphreys, A.M., Fontaneto, D. and Barraclough, T.G., 2014. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*, *5*, 1086-1094.
150. Team, R.C., 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*. ISBN 3-900051-07-0.
151. Tomlinson, P.B., 1986. The botany of mangroves. Cambridge tropical biology series.
152. Untawale, A.G., 1985. Mangroves of India: present status and multiple use practices. UNDP.

153. Untawale, A.G., Jagtap, T.G., 1992. Floristic composition of the deltaic regions of India. *Memoirs geological society of India, 22*, 243-263.

154. Vandesompele, J., 2002. Accurate normalization of real-time RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*, 4-1.

155. Venema, K., Quintero, F.J., Pardo, J.M. and Donaire, J.P., 2002. The *Arabidopsis* Na^+/H^+ exchanger *AtNHX1* catalyzes low affinity Na^+ and K^+ transport in reconstituted liposomes. *Journal of Biological Chemistry*, 277, 2413-2418.

156. Verma, D., Singla-Pareek, S.L., Rajagopal, D., Reddy, M.K., Sopory, S.K., 2007. Functional validation of a novel isoform of Na^+/H^+ antiporter from *Pennisetum glaucum* for enhancing salinity tolerance in rice. *Journal of Bioscience*, 32, 621-628.

157. Vijayan, K. and Tsou, C.H., 2010. DNA barcoding in plants: taxonomy in a new perspective. *Current Science*, 1530-1541.

158. Vinitha, M.R., Kumar, U.S., Aishwarya, K., Sabu, M. and Thomas, G., 2014. Prospects for discriminating Zingiberaceae species in India using DNA barcodes. *Journal of integrative plant biology*, *56*, 760-773.

159. Wang, L., Feng, X., Zhao, H., Wang, L., An, L., Qiu, Q.S., 2014. Functional analysis of the Na_+, K_+/H_+ antiporter *PeNHX3* from the tree halophyte *Populus euphratica* in yeast by model-guided mutagenesis. *PloS one*, 9, e104147.

160. Waugh, R. and Power, W., 1992. Using RAPD markers for crop improvement. *Trends in Biotechnology*, *10*, 186-191.

161. White, T.J., Bruns, T., Lee, S.J.W.T. and Taylor, J.L., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*, 315-322.

162. Wu, C.A., Yang, G.D., Meng, Q.W. and Zheng, C.C., 2004. The cotton *GhNHX1* gene encoding a novel putative tonoplast Na^+/H^+ antiporter plays an important role in salt stress. *Plant and Cell Physiology*, *45*, 600-607.

163. Wu, G.Q., Feng, R.J., Wang, S.M., Wang, C.M., Bao, A.K., Wei, L. and Yuan, H.J., 2015. Co-expression of xerophyte Zygophyllum xanthoxylum ZxNHX and ZxVP1-1 confers enhanced salinity tolerance in chimeric sugar beet (Beta vulgaris L.). *Frontiers in plant science*, *6*, 581.

164. Xia, T., Apse, M.P., Aharon, G.S. and Blumwald, E., 2002. Identification and characterization of a NaCl-inducible vacuolar Na^+/H^+ antiporter in *Beta vulgaris*. *Physiologia Plantarum*, *116*, 206-212.

165. Xie, F., Xiao, P., Chen, D., Xu, L. and Zhang, B., 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology*, *80*, 75-84.

166. Xu, S., He, Z., Zhang, Z., Guo, Z., Guo, W., Lyu, H., Li, J., Yang, M., Du, Z., Huang, Y. and Zhou, R., 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *National Science Review*, *4*, 721-734.

167. Xue, Z.Y., Zhi, D.Y., Xue, G.P., Zhang, H., Zhao, Y.X. and Xia, G.M., 2004. Enhanced salt tolerance of transgenic wheat (*Tritivum aestivum L.*) expressing a vacuolar Na^+/H^+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na^+. *Plant Science*, *167*, 849-859.

168. Yadav, N.S., Shukla, P.S., Jha, A., Agarwal, P.K. and Jha, B., 2012. The *SbSOS1* gene from the extreme halophyte *Salicornia brachiata* enhances Na^+ loading in xylem and confers salt tolerance in transgenic tobacco. *BMC plant biology*, *12*,188.

169. Yang, Z., Landry, J.F. and Hebert, P.D., 2016. A DNA Barcode Library for North American Pyraustinae (Lepidoptera: Pyraloidea: Crambidae). *PloS one*, *11*,e0161449.

170. Ye, C.Y., Zhang, H.C., Chen, J.H., Xia, X.L. and Yin, W.L., 2009. Molecular characterization of putative vacuolar NHX-type Na^+/H^+ exchanger genes from the salt-resistant tree *Populus euphratica*. *Physiologia plantarum*, *137*, 166-174.

171. Yokoi, S., Quintero, F.J., Cubero, B., Ruiz, M.T., Bressan, R.A., Hasegawa, P.M. and Pardo, J.M., 2002. Differential expression and function of *Arabidopsis thaliana* NHX Na^+/H^+ antiporters in the salt stress response. *The Plant Journal*, *30*, 529-539.

172. Zhang, H.X., Blumwald, E., 2001. Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. *Nature biotechnology*, *19*, 765.

173. Zhang, H.X., Hodson, J.N., Williams, J.P., Blumwald, E., 2001. Engineering salt-tolerant Brassica plants: characterization of yield and seed oil quality in transgenic plants with increased vacuolar sodium accumulation. *Proceedings National Academy of the Science*, *98*, 12832-12836.

174. Zhang, S., Zeng, Y., Yi, X. and Zhang, Y., 2016. Selection of suitable reference genes for quantitative RT-PCR normalization in the halophyte *Halostachys caspica* under salt and drought stress. *Scientific reports*, *6*, 30363.

175. Zhao, H., Tan, Z., Wen, X. and Wang, Y., 2017. An improved syringe agroinfiltration protocol to enhance transformation efficiency by combinative use of 5-azacytidine, ascorbate acid and Tween-20. *Plants*, *6*, 9.

176. Zhou, R., Shi, S. and Wu, C.I., 2005. Molecular criteria for determining new hybrid species-an application to the Sonneratia hybrids. *Molecular phylogenetics and evolution*, *35*, 595-601.

177. Zörb, C., Noll, A., Karl, S., Leib, K., Yan, F. and Schubert, S., 2005. Molecular characterization of Na^+/H^+ antiporters (*ZmNHX*) of maize (*Zea mays L.*) and their expression under salt stress. *Journal of plant physiology*, *162*, 55-66.

178. Zou, S., Fei, C., Song, J., Bao, Y., He, M. and Wang, C., 2016. Combining and comparing coalescent, distance and character-based approaches for barcoding microalgae: A Test with Chlorella-like species (Chlorophyta). *PloS one*, *11*, p.e0153833.

Appendix

Appendix I - Reagent, Buffers, and Media compositions

A. TE Buffer (pH 8.0)

Ingredients	**gm.L^{-1}**
1. Tris-HCl (10 mM)	1.21
2. Na_2EDTA (1 mM)	0.372

B. CTAB Buffer in MilliQ water (DNA isolation)

Ingredients	**gm.L^{-1}**
1. CTAB	20
2. EDTA (25 mM)	7.3
3. NaCl (1.4 M)	81.8
4. PVP-30	20
5. β-mercaptoethanol	10
6. SDS	100

C. CTAB Buffer (RNA isolation) in 0.1% DEPC treated water

Ingredients	**gm.L^{-1}**
1. CTAB	20
2. EDTA (25 mM)	7.3
3. NaCl (1.4 M)	81.8
4. PVP-30	20
5. β-mercaptoethanol	10
6. Tris base (0.1 M)	12

D. Hoagland nutrient solution

	gm.L^{-1}
$Ca(NO_3)_2.4H_2O$	270
KCl	18.6
KNO_3	24.6
$Fe(NO_3)_3.9H_2O$	13.31
$ZnSO_4.7H_2O$	0.88
$Na_2MoO_4.2H_2O$	0.26

E. YPD Media

Ingredients	**gm.L^{-1}**
1. Yeast extract	10
2. Peptone	20
3. Dextrose	20
4. Agar	20

F. YP-Gal Media

Ingredients	**gm.L^{-1}**
1. Yeast extract	10
2. Peptone	20
3. Galactose	20
4. Agar	20

Appendix II

List of Publications

Thesis Publications:

1. Saddhe, A.A., Jamdade, R.A., and Kumar, K., 2016. Assessment of mangroves in Goa, west coast India using DNA barcode markers. SpringerPlus, 5:1554.
2. Saddhe, A.A., Jamdade, R.A., and Kumar, K., 2017. Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. PLoS ONE, 12:e0183245.
3. Saddhe, A.A., and Kumar, K., 2018. DNA barcoding of plants: Selection of core markers for taxonomic groups. Plant Science Today, 5:9-13.
4. Saddhe, A.A., Malvankar, M.R., and Kumar K., 2018. Selection of reference genes for quantitative real-time PCR analysis in halophytic plant *Rhizophora apiculata*. PeerJ, 6:e5226.
5. Saddhe, A.A., Kumar, K.. 2019. Molecular cloning, expression analysis and heterologous characterization of *Rhizophora apiculata NHX*. (Submitted)

Other publications:

6. Saddhe, A.A., and Kumar, K., 2015. *In-silico* identification and expression analysis of MscS like gene family in rice. Plant Gene, 1:8-17.
7. Manuka, R., Saddhe, A.A., and Kumar, K., 2015. Genome-wide identification and expression analysis of WNK kinase gene family in rice. Computational biology and chemistry, 59:56-66.
8. Manuka, R., Saddhe, A.A., and Kumar, K., 2018. Expression of *OsWNK9* conferred tolerance to salt and drought stresses in *Arabidopsis.* Plant Science, 270:58-71.
9. Dahibhate, N.L., Saddhe, A.A., and Kumar, K., 2018. Mangrove plant: A potential source of natural product and bioactive compounds. The Natural product Journal, (Accepted).
10. Saddhe, A.A., Karle, S., Malvankar, M.R., Kumar, K., 2018. Reactive Nitrogen Species: Paradigm of Cellular Signalling and Regulation of Salt Stress in Plants. Environmental and Experimental Botany.

Appendix III

Conferences and Workshop attended

1. Ankush Ashok Saddhe, Kundan Kumar. Evaluation of multilocus marker efficacy for delineating mangrove species of West coast India 7th International Barcode of Life Conference (dnabarcodes2017.org), 20– 24 November 2017, African center of DNA barcoding, University of Johannesburg, South Africa.
2. Ankush Ashok Saddhe**,** Kundan Kumar. Multilocus marker approach can discriminate mangrove plant species from Goa, West coast India. National conference of Young Researcher 2017 on New Frontier in life Sciences and Environment. 16-17 March 2017, Goa University Goa.
3. Ankush Ashok Saddhe, Kundan Kumar. *In-silico* analysis revealed 33 members of aquaporin gene family in *Brachypodium*. International conference on Trends in cell and Molecular Biology 19-21 Dec 2015. BITS Pilani, K. K Birla Goa campus, Goa, India
4. Ankush Ashok Saddhe, Kundan Kumar. Phylogenetic Assessment of Goan Mangroves along West Coast India using DNA Barcode Markers. Fifth International conference on Plants and Environmental pollution 24-27 Feb 2015. IBEB and CSIR-NBRI, Lucknow.
5. Ankush Ashok Saddhe, Kundan Kumar. In silico identification and expression analysis of MscS like gene family in rice. National Seminar on New Frontiers in plant Sciences and Biotechnology 29-30 Jan 2015 Dept of Botany, Goa University, Goa.
6. Ankush Ashok Saddhe**.** One day workshop on introduction to practical NMR spectroscopy, 1 Aug 2015. Organizer CSIR-National Chemical Laboratory, Pune and Venture centre, Pune.

Appendix IV

Brief Biography of the Candidate

Name	Ankush Ashok Saddhe
Date of Birth	16/03/1987
Education	M.Sc. (Biotechnology) Government Institute of Science, Aurangabad, MS, India B.Sc. (Botany, Zoology, Environment Science) Deogiri College, Aurangabad, Dr. BAMU (M.S.) India
Email ID	sadhyeankush@gmail.com, p2013403@goa.bits-pilani.ac.in

Working Experience

- August 2017 – present: Institute Fellowship, BITS Pilani, K. K. Birla, Goa campus, India.
- May 2016 – July 2017: UGC- SRF, BITS Pilani, K. K. Birla, Goa campus, India.
- Oct 2013 – April 2016: UGC- JRF, BITS Pilani, K.K. Birla, Goa campus, India.
- Aug 2012 -Aug 2013: UGC-JRF, Paul Hebert center for DNA Barcoding and Biodiversity studies, Dept of Zoology, Dr BAMU Aurangabad, India.

Research Publications

11 publications in International journal, 01 National journal, 02 book chapter

Awards / Fellowships

- DST-SERB International travel grant: to attend iBOL 2017 conference at Kruger National Park, South Africa.
- Research Fellowship 2017- 2018 BITS Pilani K.K. Birla Goa campus
- Research Fellowship: 2012-2017, JRF & SRF, Awarded by University Grant Commission (UGC) Govt. of India.
- GATE 2009 and 2010 (Life Sciences and Biotechnology)
- National Eligibility Test (NET) June 2009, Dec 2009 and June 2011: Conducted jointly by University Grants Commission (UGC) and Council of Scientific and Industrial Research (CSIR), Govt. of India.
- Department of Biotechnology: 2010, JRF, Govt. of India.

Appendix V

Brief Biography of the Supervisor

Name	Dr. Kundan Kumar
Education	Ph.D. (2009): National Institute of Plant Genome Research, New Delhi
Ph.D. Thesis Title	Investigation of the role of mitogen activated protein kinase kinase in *Oryza sativa* under abiotic stress conditions
Contact Details	Chamber No: B112, BITS PILANI, K.K. Birla Goa Campus NH17B, Zuarinagar, Goa, India
Email ID	kundan@goa.bits-pilani.ac.in
Phone	0832-2580196
Research Interest	Plant molecular biology and stress physiology, Plant signaling, Phytoremediation and Plant natural products

Professional Experience

May 2009-Oct 2011	Postdoctoral Research Associate, University of Massachusetts, Amherst, USA
Nov 2011-April 2012	Postdoctoral Fellow, McGill University, McDonald Campus, Sainte-Anne-de-Bellevue, Canada
May 2012-June 2018	Assistant Professor, Department of Biological Sciences, BITS Pilani K K Birla Goa Campus, Goa, India
July 2018-present	Associate Professor, Department of Biological Sciences, BITS Pilani K K Birla Goa Campus, Goa, India

Sponsored Research Project

Completed

1. Identification and characterization of abiotic stress responsive With No Lysine (WNK) kinase in rice (*Oryza sativa* L.): Science & Engineering Research Board) under DST Fast Track scheme. (2013-16)
2. DNA Barcoding of Goan Mangroves: Research Initiation Grant from BITS. (2013-15)

Ongoing

1. Screening, isolation and identification of novel antimicrobial compounds from potential mangroves of Goa. Funded by CSIR EMR-II (2016-19)
2. Study on the role of TIP family members in arsenite accumulation and their transport in rice, Funded by BRNS (2017-2020)

Publications

30 publications in International journals, 03 book chapters

Reviewer for international journals

Plant Molecular Biology, Scientific Reports, International Journal of Phytoremediation, Phytochemistry, Plant Science, PeerJ, and BMC Genomics etc.

No of Ph.D. Students

Registered 03

Member of professional body

Life member of Indian Science Congress Association and Life member of society of Plant Biochemistry and Biotechnology

Reprints

Saddhe *et al. SpringerPlus (2016) 5:1554*
DOI 10.1186/s40064-016-3191-4

SpringerPlus

RESEARCH Open Access

Assessment of mangroves from Goa, west coast India using DNA barcode

Ankush Ashok Saddhe[1], Rahul Arvind Jamdade[2] and Kundan Kumar[1*]

Abstract

Mangroves are salt-tolerant forest ecosystems of tropical and subtropical intertidal regions. They are among most productive, diverse, biologically important ecosystem and inclined toward threatened system. Identification of mangrove species is of critical importance in conserving and utilizing biodiversity, which apparently hindered by a lack of taxonomic expertise. In recent years, DNA barcoding using plastid markers *rbcL* and *matK* has been suggested as an effective method to enrich traditional taxonomic expertise for rapid species identification and biodiversity inventories. In the present study, we performed assessment of available 14 mangrove species of Goa, west coast India based on core DNA barcode markers, *rbcL* and *matK*. PCR amplification success rate, intra- and inter-specific genetic distance variation and the correct identification percentage were taken into account to assess candidate barcode regions. PCR and sequence success rate were high in *rbcL* (97.7 %) and *matK* (95.5 %) region. The two candidate chloroplast barcoding regions (*rbcL*, *matK*) yielded barcode gaps. Our results clearly demonstrated that *matK* locus assigned highest correct identification rates (72.09 %) based on TaxonDNA Best Match criteria. The concatenated *rbcL* + *matK* loci were able to adequately discriminate all mangrove genera and species to some extent except those in *Rhizophora, Sonneratia* and *Avicennia*. Our study provides the first endorsement of the species resolution among mangroves using plastid genes with few exceptions. Our future work will be focused on evaluation of other barcode markers to delineate complete resolution of mangrove species and identification of putative hybrids.

Keywords: Mangrove, Goa, DNA barcode, *rbcL*, *matK*

Background

Mangroves are unique ecosystem exist along the sheltered inter-tidal coastline, in the margin between the land and sea in tropical and subtropical areas. This ecosystem endowed with productive wetland having flora and fauna adapted to local environment such as fluctuated water level, salinity and anoxic condition (Tomlinson 1986; Hutchings and Saenger 1987). They are most productive and biologically important ecosystems of the world which provide goods and services to human society in coastal and marine systems (FAO 2007). They have unique features such as aerial breathing roots, extensive supporting roots, buttresses, salt-excreting leaves and viviparous propagules (Duke 1992; Shi et al. 2006). The term 'mangroves' are referred to either individual plant or intertidal ecosystem or both, as 'Mangrove plants' and 'Mangrove ecosystem' (MacNae 1968). However, in this context we used mangrove term as a mangrove plants. Anthropogenic activity and climate are responsible for destruction of coastal mangroves vegetation. Globally among 11 of the 70 mangrove species were listed threatened species by International Union for Conservation of Nature (IUCN) (Polidoro et al. 2010).

Mangrove species diversity and distribution reported existence of 34 major and 20 minor mangrove species belonging to 20 genera and 11 families across the world (Tomlinson 1986). Ricklefs and Latham (1993) reported the existence of 19 genera with 54 mangrove species including few hybrids. According to world atlas of mangroves database, 73 mangrove species along with few recognized hybrids are distributed in 123 countries with territorial coverage of 150,000 km^2 area globally (Spalding et al. 2010). Indian mangrove vegetation represents

*Correspondence: kundan@goa.bits-pilani.ac.in
[1] Department of Biological Sciences, Birla Institute of Technology and Science Pilani, K. K. Birla Goa Campus, Sancoale, Goa 403726, India
Full list of author information is available at the end of the article

fourth largest in the world, distributed along the coastline and occupies 8 % of the total world mangrove covering 6749 km^2 areas (Naskar and Mandal 1999). The entire mangrove habitats in India are situated in three zones: east coast (4700 km^2), west coast (850 km^2) and Andaman & Nicobar Islands (1190 km^2). East coast zone ranges from Sundarban forest of West Bengal to Cauvery estuary of Tamil Nadu and comprises 70 % mangrove (Untawale and Jagtap 1992; Jagtap et al. 1993; Sanyal et al. 1998). West coast region stretches from Bhavnagar estuary of Gujarat to Cochin estuary of Kerala and constitute 15 % mangrove (Mandal and Naskar 2008). Mangrove flora of India constitutes about 60 species belonging to 41 genera and 29 families (Untawale 1985). Along the west coast of India, 34 species of mangroves belonging to 25 genera and 21 families have been reported. There are about 11, 20, 14 and 10 species of mangroves reported along the coast of Gujarat, Maharashtra, Goa and Karnataka respectively in western India. Goa state is located in western coast of India and mangrove vegetation in Goa occupies 500 ha of area (Government of India, 1997). The Cumbarjua canal (15 km) links the two river channels of Mandovi and Zuari, forming an estuarine complex which supports a substantial mangrove extent. D'Souza and Rodrigues (2013) reported the presence of 17 mangrove species in Goa that include 14 true and 3 associated mangrove species.

DNA barcoding is currently used effective tool that enables rapid and accurate identification of plant (Li et al. 2015). The Consortium for the Barcode of Life (CBOL) recommended *rbcL* + *matK* as the core barcode. However, these core barcode further combined with the *psb*A-*trn*H intergenic non-coding spacer region which improved discrimination power of core barcode. The non-coding intergenic region *psb*A-*trn*H exhibits high rates of insertion/deletion and sequence divergence (Kress and Erickson 2007). These features make *trnH-psbA* highly suitable candidate plant barcode for species resolution. Later on, the nuclear ribosomal internal transcribed spacer (ITS) region considered as supplementary barcode, though China Plant Barcode of Life claimed ITS region had higher discriminatory power than plastid core barcodes (CBOL Plant Working Group 2009; Hollingsworth et al. 2011; China Plant BOL Group 2011). Hollingsworth et al. (2011) observed ITS region has some limitations which prevent it from being a core barcode such as incomplete concerted evolution, fungal contamination and difficulties of amplification and sequencing. Plastid gene large subunit of the ribulose-bisphosphate carboxylase gene (*rbcL*) is of 1350 bp in length and choice for DNA barcoding (Chase 1993).The maturase gene *matK* is about 1500 bp long and located within the *trnK* gene encoding the tRNALys (UUU). Substitution rate of the *matK* gene is highest among the plastid genes (Hilu et al. 2003). Plastid gene *matK* can discriminate more than 90 % of species in the Orchidaceae but less than 49 % in the nutmeg family (Kress and Erickson 2007; Newmaster et al. 2008). In another case, identification of 92 species from 32 genera using the *matK* barcode could achieve a success rate of 56 % (Fazekas et al. 2008). However, a recent study of the flora of Canada revealed 93 % success in species identification with *rbcL* and *matK*, while the addition of the trnH-psbA intergenic spacer achieved discrimination up to 95 % (Burgess et al. 2011). Gonzalez et al. (2009) reported that species discrimination was lower (<50 %) for *rbcL* + *matK* combination in the study of tropical tree species in French Guiana. Lower discrimination were reported in closest and complex taxa of *Lysimachia*, *Ficus*, *Holcoglossum* and *Curcuma* using *rbcL* and *matK* (Xiang et al. 2011; Zhang et al. 2012; Li et al. 2012; Chen et al. 2015). The lowest discriminatory power was observed in closely related groups of *Lysimachia* with *rbcL* (26.5–38.1 %), followed by *matK* (55.9–60.8 %) and combinations of core barcodes (*rbcL* + *matK*) had discrimination of 47.1–60.8 % (Zhang et al. 2012).

Delineating mangrove species from putative hybrids using morphological characters are always questionable. Putative hybrids were reported within the major genera of *Rhizophora*, *Sonneratia* and *Lumnitzera* and recently in *Bruguiera* (Tomlinson 1986; Duke and Ge 2011). In the present study, we assessed mangrove species using plastid coding loci viz. *rbcL* and *matK*. Mangroves from Goa are rich in diversity and accounted 14 species belonging to four order and five families. This is our first step towards DNA barcoding of mangroves based on plastid genes. Our study might be helpful in identification as well as developing various strategies towards mangrove conservation.

Methods

Sample collection

In the present study, leaf samples of 14 mangrove species were collected from Goa, located on the west coast of India with geographical latitude of 15.5256°N and longitude of 73.8753°E. Mangrove species identification was performed based on morphological characteristics using a comparative guide to the Asian mangroves and mangroves of Goa (Yong and Sheue 2014; Dhargalkar et al. 2014; Setyawan et al. 2014). Herbarium of these specimens was deposited at Botanical Survey of India, western regional centre, Pune, India. The morphology based identification keys used to authenticate the taxon identities of 14 mangroves species from Goa were listed in supplementary information (Additional file 1: Table S1). The well identified voucher specimens along with their taxonomic information and collection details are

listed (Table 1) with their photographs in supplementary information (Additional file 1: Fig. S1). The sequences obtained using barcode markers: *rbcL* and *matK* were submitted to the NCBI GenBank (Accession numbers indicated in Table 1), and publicly accessible through the dataset of project DNA Barcoding of Indian Mangroves (Project code: IMDB) in Barcode of Life Data systems (BOLD) (doi:10.5883/DS-IMDBNG) (Ratnasingham and Hebert 2007).

DNA extraction

High content of mucilage, latex, phenolics, secondary metabolites and polysaccharides in these plants make it a difficult system for protein and nucleic acid isolation from mangrove plants. Cetyl-trimethyl ammonium bromide (CTAB) protocol for DNA extraction from mangroves (Parani et al. 1997a) was modified. Leaf tissue was pulverized in liquid nitrogen and pulverized leaf sample (0.2 g) were mixed with CTAB buffer (20 mM EDTA; 1.4 M NaCl; 2 % PVP-30; 1 % β-mercaptoethanol; 10 % SDS and 10 mg/ml proteinase K). The suspension was incubated at 60 °C for 60 min with gentle mixing and centrifuged at 14,000 rpm for 10 min at room temperature with equal volume of chloroform: isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and DNA was precipitated with 0.6 volume of cold isopropanol (−20 °C) and chilled 7.5 M ammonium acetate followed by storing at −20 °C for 1 h. The precipitated DNA was centrifuged at 14,000 rpm for 10 min at 4 °C followed by washing with 70 % ethanol. DNA was finally dissolved in TE buffer (10 mM Tris–HCl, 1 mM Na_2EDTA, pH 8.0) and its quantity and quality was confirmed by agarose gel electrophoresis and nanodrop (Thermo Scientific, USA).

PCR and sequencing

Amplification of plastid genes (*rbcL* and *matK*) was carried out in 50-µl reaction mixture containing 10–20 ng of template DNA, 200 µM of dNTPs, 0.1 µM of each primers and 1 unit of Taq DNA polymerase (Thermo Scientific, USA). The reaction mixture was amplified in Bio-Rad (T100 model) thermal cycler with temperature profile for *rbcL* (94 °C for 4 min; 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min; repeated for 35 cycles, final extension 72 °C for 10 min) and for *matK* (94 °C for 1 min; 35 cycles of 94 °C for 30 s, 50 °C for 40 s, 72 °C for 40 s; repeated for 37 cycles, final extension 72 °C for 5 min). The amplified products were separated by agarose gel (1.2 %) electrophoresis and stained with ethidium bromide (Sambrook et al. 1989). Two pair

Table 1 Details of the mangrove species used in the present study with family, status, life form, voucher number and GenBank accession numbers obtained after sequence submission

S. No.	Specimen	Family	Status	Life form	Herbarium Voucher No.	Accession No. *rbcL*	Accession No. *matK*
1	*Avicennia officinalis*	Acanthaceae	TM	Tree	AAS-100-02	KP697351, KP697352, KU748517	KP725238, KP725239
2	*Avicennia marina*	Acanthaceae	TM	Tree	AAS-110-12	KP697349, KP697350, KM255068	KP725236, KM255083, KP725237
3	*Avicennia alba*	Acanthaceae	TM	Tree	AAS-120-22	KM255067, KM255069, KP697348	KM255082, KM255084, KP725235
4	*Bruguiera cylindrica*	Rhizophoraceae	TM	Tree	AAS-130-32	KP697354, KM255070, KP697353	KP725241, KM255085, KP725240
5	*Bruguiera gymnorrhiza*	Rhizophoraceae	TM	Tree	AAS-140-42	KM255071,KP697355,KP697356	KM255086,KP725242,KP725243
6	*Rhizophora mucronata*	Rhizophoraceae	TM	Tree	AAS-150-52	KM255077, KU748519	KM255092, KU748522, KU748523
7	*Rhizophora apiculata*	Rhizophoraceae	TM	Tree	AAS-160-62	KP697362, KP697363, KM255076	KP725249, KP725250, KM255091
8	*Aegiceras corniculatum*	Primulaceae	MM[T]	Tree/Shrub	AAS-170-72	KM255066, KP697344, KP697345, KM255075, KP697346, KP697347	KM255081, KP725231, KP725232, KM255090, KP725233, KP725234
9	*Excoecaria agallocha*	Euphorbiaceae	TM	Tree	AAS-180-82	KM255073, KP697360, KP697359	KM255088, KP725247, KP725246
10	*Kandelia candel*	Rhizophoraceae	TM	Tree	AAS-190-92	KP697361, KM255074, KU748518	KP725248, KM255089, KU748521
11	*Ceriops tagal*	Rhizophoraceae	TM	Tree	AAS-200-02	KM255072, KP697358, KP697357	KM255087, KP725244, KP725245
12	*Sonneratia alba*	Lythraceae	TM	Tree	AAS-210-12	KM255078, KP697364, KU748520	KM255093, KP725251
13	*Sonneratia caseolaris*	Lythraceae	TM	Tree	AAS-220-22	KP697365, KP697366, KM255079	KP725252, KP725253, KM255094
14	*Acanthus ilicifolius*	Acanthaceae	TM	Shrub	AAS-230-32	KM255065, KP697342, KP697343	KM255080, KP725229, KP725230

TM True Mangroves, *MM* Minor Mangroves, *T* Tomlinson (1986)

of universal primers *rbcL* (*rbcL*a_F and *rbcL*a_R) and *matK*_390f and *matK*_1326r were used for the amplification purpose (Kress and Erickson 2007; Vinitha et al. 2014; Chen et al. 2015). To amplify *R. apiculata matK* locus, we designed *matK*_RA reverse primer as follows: 5′-AAAGTTCGTTTGTGCCAATGA-3′. PCR products were purified according to manufacturer's instruction (Chromous Biotech) and further sequencing reactions were carried out using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on ABI 3500xL Genetic Analyzer (Applied Biosystems).

Data analysis

Sequence alignment and assembly was achieved in Codon code Aligner v.3.0.1 (Codon Code Corporation) and MEGA 6 (Tamura et al. 2013). The NCBI BLAST was performed to confirm identity of specimens (Altschul et al. 1990). All known mangroves sequences were searched with our sequenced samples using 'BLASTn' tool against NCBI database and highest-scoring hit from each query is taken as the mangrove identification. Intraspecific, interspecific and barcode gap analysis was performed at Barcode of Life Data systems web portal. Further, *rbcL* and *matK* sequences were concatenated using DNASP v5.10 and analyzed in MEGA 6 for their resolution inference (Rozas, 2009). The effectiveness of the analysed barcodes in *rbcL*, *matK* and *rbcL* + *matK* was evaluated using TaxonDNA v1.6.2, Species Identifier 1.8 (Meier et al. 2006) and BLASTClust (http://toolkit.tuebingen.mpg.de/blastclust). Neighbor-joining (NJ) trees were constructed using MEGA 6.0 and K2P genetic distance model, and node support was assessed based on 1000 bootstrap replicates. Species with multiple individuals forming a monophyletic clade in phylogenetic trees with a bootstrap value above 60 % were considered as successful identification.

Results

DNA barcode and sequence analysis

Mangroves belonging to 14 species, 9 genera and 5 families were collected. We acquired high quality DNA barcodes for 45 specimens belonging to 14 species, which were sequenced for *rbcL* and *matK*. The sequencing result of *rbcL* produced an average of 510 bp without any insertion, deletion and stop codon, whereas *matK* sequencing produced 712 bp with few insertion and deletions in the form of gaps without stop codon. Overall GC content observed in *rbcL* was 43.29 % (SE = 0.09), while in *matK* it was 33.18 % (SE = 0.18). The mean GC content of codon at positions 1-3 in *rbcL* was 54.66 % (SE = 0.1), 45.77 % (SE = 0.09) and 29.44 % (SE = 0.21), and in *matK*, it was 33.15 % (SE = 0.18), 30.92 % (SE = 0.36), 29.91 % (SE = 0.25) respectively. The specimen data, collection site details and sequences were submitted to BOLD database in form of project IMDB (doi:10.5883/DS-IMDBNG) (For details, Table 1). The specimens were verified from sequenced data by performing NCBI BLAST. This is performed for preliminary verification for all mangroves at species level but downside in our case study is limited reference data for comparison. The *rbcL* and *matK* correctly identified genera up to 100 %, while species identification with *rbcL* and *matK* leads to 64 and 85 % identification respectively.

Intraspecific and interspecific relationship

Barcoding of mangrove exhibited absolute average interspecific differentiation of 0.35 % and 0.9 % in *rbcL* and *matK* respectively, while for species average intraspecific variability was 0.24 % in *rbcL* and 0.20 % in *matK* (Table 2) with low species resolution in few taxa. The intraspecific and interspecific analysis for *rbcL* revealed largest average pairwise distance of 0.68, while in *matK* it was 2.05 and 2.32 respectively. The highest range of congeneric differentiation in *Bruguiera* and *Avicennia* were observed in *rbcL* from 0 to 0.68, whereas for *matK,* it ranged from 1.29 to 2.31 in *Avicennia,* further suggesting significant genetic divergence within *Avicennia.*

Barcode gap analysis

The barcode gap analysis revealed highest intraspecific distance (>2 %) in 9 specimens of *rbcL* and 6 specimens of *matK*, while low intraspecific distance (<2 %) in 11 specimens of *rbcL* and 9 specimens of *matK*. Here, low intraspecific distance (<2 %) suggests low species resolution, thus leading to species overlap.

With *rbcL* the largest nearest neighboring distance of 8.43 was observed in *Avicennia alba* with mean intraspecific distance of 0.11 (Fig. 1a). The maximum intraspecific distance of 0.68 was observed within three individuals of *Kandelia candel, Bruguiera gymnorrhiza, A. officinalis* and *Sonneratia caseolaris* (Fig. 1b). With *matK*, maximum intraspecific distance of 2.05 was observed in *Excoecaria agallocha* with three individuals per species (Fig. 1d), while largest distance to the nearest neighbor of 24.65 was observed in *A. officinalis* with mean intraspecific distance of 0.12 (Fig. 1c). Overall average nearest neighboring divergence observed among mangroves using *rbcL* was 1.39 % (S.E = 0.17) and *matK* was 4.07 % (S.E = 0.5) (Fig. 1a).

Species identification and assignment

The species were assigned to their taxa based on three methods, similarity based method using TaxonDNA, BLAST score based single linkage (BLASTClust) and tree based method (NJ). To assess the species assignment of single region and multi regions, we used the 'Best Match' (BM) and 'Best Closest Match' (BCM) criteria

Table 2 Genetic divergence of mangrove species based on Kimura 2 Parameter within species, genus and family levels

	No. of sequences	Taxa	Comparisons	Min Dist (%)	Mean Dist (%)	Max Dist (%)	SE Dist (%)
For *rbcL*							
Within species	44	14	53	0	0.24	0.68	0
Within genus	26	4	50	0	0.35	0.68	0
Within family	29	2	132	1.71	2.63	4.01	0
For *matK*							
Within species	43	14	50	0	0.2	1.32	0.01
Within genus	25	4	45	0	0.9	2.32	0.02
Within family	29	2	141	2.11	5.82	13.37	0.02

Min Dist Minimum distance, *Max Dist* Maximum distance, *SE Dist* Standard error distance

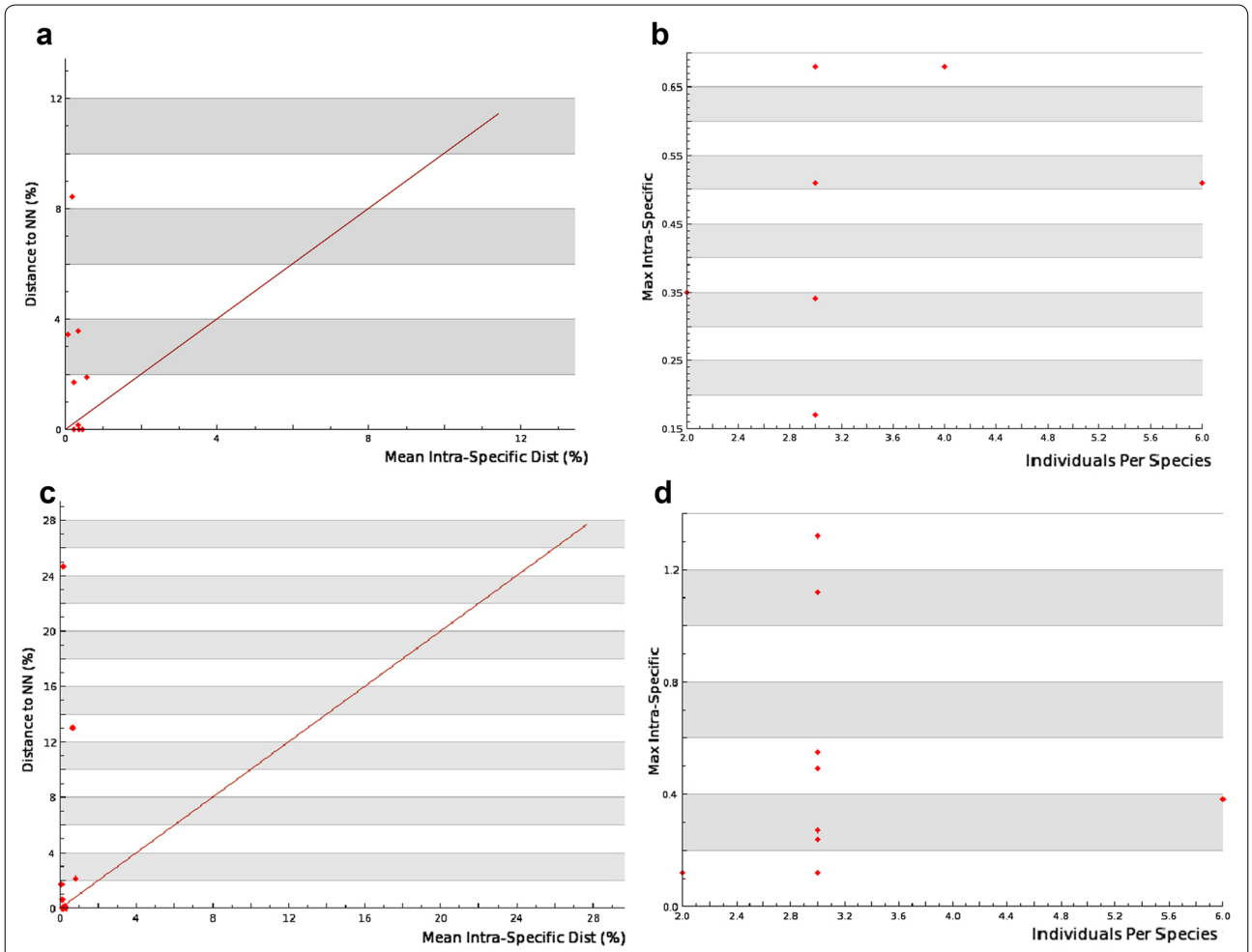


Fig. 1 Scatterplots confirming the existence and magnitude of the barcode gap. **a** For *rbcL* mean intra-specific versus Nearest Neighbor (NN). **b** For *rbcL* individuals per species vs. max intra-specific. **c** For *matK* mean intra-specific versus Nearest Neighbour. **d** For *matK* individuals per species versus max intra-specific

from TaxonDNA. For TaxonDNA analysis, we need to set threshold (T) below which 95 % of all intraspecific distances were found. All the results above the threshold (T) were treated as 'incorrect'. Similarly, if all matches of the query sequence were below threshold (T), the barcode assignment was considered to be correct identification. The matches of the query sequence were equally good, but correspond to a mixture of species, then test was

treated as ambiguous identification. For the single barcode region, *matK* had the highest rate of correct identification using BM (72.09 %) and BCM (39.53 %) than *rbcL* with (BM 47.72 %), BCM (31.81 %) (Table 3). The concatenated regions (*rbcL* + *matK*) demonstrated to resolve species at the level of 66.6 % using BM and BCM criteria (Table 3). The species specific clustering using match and mismatch criteria was evaluated in TaxonDNA and BLASTClust, where sequences with highest similarity and identity were considered as successfully identified. Those species with an identical barcode sequence to an individual of other species were considered as ambiguous, and sequences matching with different species names were treated as failure identifications. Species having single sample and unique sequence were considered as potentially distinguishable. The BLASTClust analysis revealed slightly different results than that of TaxonDNA, where the rate of species resolution and cluster formation was low as that of TaxonDNA (Table 4). Species with multiple individuals forming a monophyletic clade in NJ trees with a bootstrap value above 60 % were considered as successful identifications (Kress et al. 2010). The *matK* and *rbcL* + *matK* discriminated mangrove species in NJ model test method, while *rbcL* alone failed to identify those species (Fig. 2a–c). Further analysis revealed similar rates of species resolution using both methods for *matK* as well as *rbcL* (Table 5). *Rhizophora*, *Sonneratia* and *Avicennia* genera were failed to discriminate their species using plastid markers *rbcL, matK* and *rbcL* + *matK*.

Discussion

To the best of our knowledge, current study is the first attempt of performing DNA barcoding based assessment of mangroves from Goa using plastid core markers *rbcL* and *matK*. Some countable reports based on molecular taxonomy and phylogeny of Indian mangroves are available using nuclear, mitochondrial and plastid markers (ITS, *rbcL*, RFLP, RAPD, PCR-RAPD and AFLP) (Parani et al. 1997a, b; Lakshmi et al. 1997, 2000; Setoguchi et al. 1999; Schwarzbach and Ricklefs 2000). Besides this there are many reports of mangroves identification based on morphological characters (Untawale 1985; Tomlinson 1986; Untawale and Jagtap 1992). Present study revealed discrimination of mangroves based on DNA barcoding at species level excluding some taxa (*Rhizophora*, *Sonneratia* and *Avicennia*). Highest rate of PCR amplification and sequencing was observed in *rbcL* (97.7 %), while amplification as well as sequencing rate of *matK* was 95.5 %. Similarly, highest success rate of identification was observed with *matK* (80.5 %) in local temperate flora of Canada and in combination *rbcL* + *matK* identified 93 % flora (Burgess et al. 2011). Species identification success rate using *rbcL* seems to be higher, whereas *rbcL* recovery ranged from 90 to 100 % (Little and Stevenson 2007; Ross et al. 2008; CBOL Plant Working Group 2009). *matK* showed difficulties in PCR amplification and sequencing. Fazekas et al. (2008) showed that *matK* markers provide possibility of 88 % sequencing success, with the use of 10 primer pair combinations. Similarly, a

Table 3 Identification success rates using TaxonDNA (Species Identifier) program under 'Best Match' and 'Best Closest Match' methods

Barcodes	No. of Sequences	Best Match (%)			Best closest match (%)				T (%)	No. of clusters	Match/ mismatch
		Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect	No match			
rbcL	44	47.72	36.36	15.9	31.81	27.27	11.36	13	0	23	6/8
matK	43	72.09	25.58	2.32	39.53	13.95	2.32	44.18	0.11	24	10/4
rbcL + *matK*	42	66.66	16.66	16.66	66.66	16.66	16.66	0	0.2	21	8/6

TaxonDNA is an alignment-based method based on sequence distance matrices. Percentage of correct/incorrect/ambiguous assignment of a taxon is compared using molecular operating taxonomic unit (MOTU). The species specific clustering using match and mismatch criteria

T Threshold

Table 4 Identifications of all mangrove samples based on BLASTClust result

Barcode	No. of sequences	Average length of sequences	Number of species	Number of clusters	Match/ mismatch
rbcL	44	586	14	6	3/11
matK	43	818	14	8	3/11
rbcL + *matK*	42	1404	14	15	4/10

BLASTClust is a method based on blast similarity scores of unaligned sequences

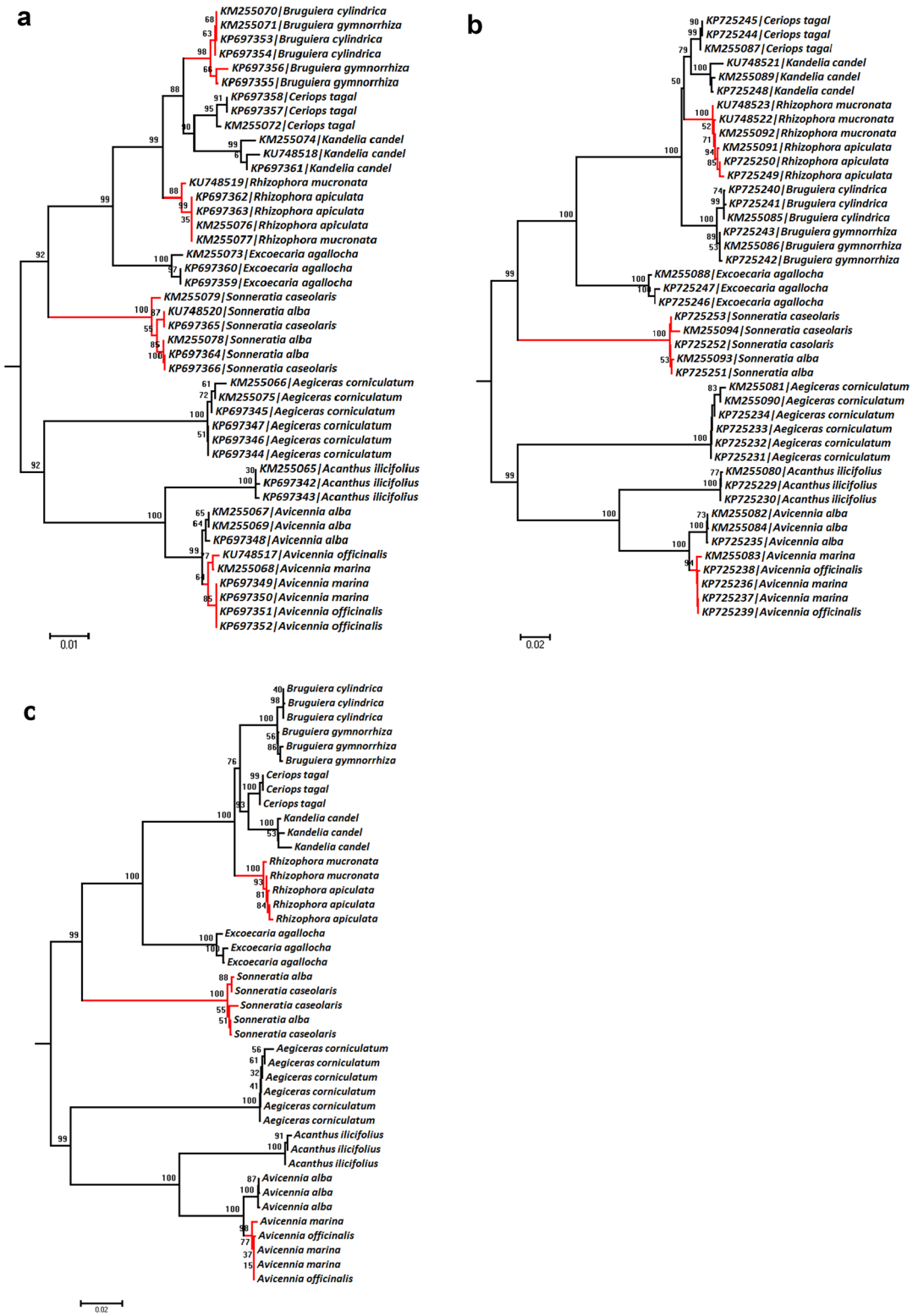


Fig. 2 Neighbor joining tree (Kimura 2 Parameter distance using bootstrap value of 1000 replicates). **a** *rbcL*, **b** *matK*, and **c** *rbcL* + *matK* concatenated NJ (K2P) trees. Highlighted clades (*red color*) indicate unresolved or least differentiated mangroves sequences

lower amplification and sequencing success of *matK* has been reported in several other studies and amplification ranges from 42 to 70 % (Ford et al. 2009; Gonzalez et al. 2009; Kress et al. 2010; Hollingsworth et al. 2011). In contrast, CBOL reported that single pair of *matK* primer was successfully amplified and sequenced 84 % angiosperm species (CBOL Plant Working Group, 2009). We faced many hindrances in amplification and sequencing of *Rhizophora* genera species *R. apiculata* using universal *matK* primers. *R. apiculata* was amplified and sequenced using universal *rbcL* marker but for *matK* amplificaiton, we designed a reverse primer. The possible explanation for the trouble could be due to secondary metabolite might hindered amplification of target genes or failure of primers to amplify genes.

Initially, species identification was performed by NCBI BLAST using *rbcL* and *matK* sequence data, the BLAST could yield accurate identifications results (Hollingsworth et al. 2009; Kress et al. 2010; Kuzmina et al. 2012). On a similar note BLAST was performed revealing its least efficacy in species identification. It has been used for verification purpose in recent years and comparisons based on test datasets (Ford et al. 2009). Parmentier et al. (2013) reported that species assignment using BLAST method was reliable for genus identification of African rainforest tree (95–100 % success), but less for species identification (71–88 %). Sometimes it gave erroneous identifications, most often due to the limited number of available reference sequences. In the present study, BLAST result with default parameter, for *rbcL* successfully identified genera (100 %) and species identification rate was 64.28 % for 14 mangroves species. *matK* was able to identify genera (100 %) and species identification up to 85.71 % successfully. The possible reason for this erroneous assignment in some taxa at species level due to availability of limited sequences in the BOLD or GenBank database (Parmentier et al. 2013). Our result underscored the importance of BLAST method to assigned correct mangroves genera identification (with *rbcL* and *matK*). Both *Sonneratia alba* and *Avicennia marina* were incorrectly identified at species level using *rbcL* and *matK*. Some mangrove species viz. *R. apiculata*, *B. cylindrica* and *A. alba* were misidentified at species level using *rbcL*.

The genetic divergence analysis exhibited highest divergence in *Avicennia* species, while barcode gap and nearest neighbor analysis revealed low species resolution and barcode gap with nearest neighboring distance (<2 %), further confirming species overlap in *Avicennia* (*A. officinalis* (*rbcL*:0; *matK*: 0–1.71) and *A. marina* (*rbcL*: 0–0.34; *matK*: 0), *Bruguiera* (*B. gymnorrhiza* (*rbcL*: 0; *matK*: 0.61) and *B. cylindrica* (*rbcL*: 0–1.71; *matK*: 0.61), *Rhizophora* (*R. mucronata* (*rbcL*: 0; *matK*: 0.14) and *R. apiculata* (*rbcL*: 0; *matK*: 0.14), *Sonneratia* (*S. caseolaris* (*rbcL*: 0; *matK*: 0) and *S. alba* (*rbcL*: 0; *matK*: 0). Low genetic distances between species was largely due to the presence of species-rich genera with low sequence variation for the plastid genome (Burgess et al. 2011).

The species identification and taxon assignment was evaluated using TaxonDNA and BLASTClust for *rbcL, matK* and *rbcL* + *matK*. Overall *matK* marker showed good performance at species and genus level (Tables 3, 4). In contrast to *matK*; *rbcL* alone showed poor performance at species level identification. Combined, *rbcL* + *matK* markers showed better performance at species and genus level identification (Tables 3, 4, 5). Accordingly, plant CBOL group (2009) reported only 72 % species level resolution using combined *rbcL* and *matK*. Similar result was observed after combined *rbcL* and *matK* at species level resolution (Chen et al. 2015). Lowest resolution was recorded in closely related groups of *Lysimachia* with combination of *rbcL* and *matK* universal markers (Zhang et al. 2012). However, the identification rates based on TaxonDNA and phylogenetic tree methods (Tables 3, 5) were significant with *matK* as compared to *rbcL*. Low resolution using DNA barcoding regions has been documented in many other plants such as the genus *Araucaria* (32 %), *Solidago* (17 %) and *Quercus* (0 %) (Little and Stevenson 2007; Leon-Romero et al. 2012). In TaxonDNA analysis, for *rbcL* threshold (T) was observed 0 %, similar result was recorded for *rbcL* in the Zingiberaceae family (Chen et al. 2015). However, threshold (T) for Indian Zingiberaceae family members were recorded as 0.20 % for *rbcL* and 0 % for *rpoB* and *accD* (Vinitha et al. 2014). In BLASTClust, the *rbcL* and *matK* regions showed similar identification rates, while concatenation of both these regions increased the efficiency of species resolution as well as cluster formation (Gonzalez et al. 2009; Blaalid et al. 2013). In case of closest taxa of mangroves viz. *Avicennia*, *Rhizophora* and *Sonneratia* species, there is a need to explore new DNA barcode markers, which may leads to species level resolution.

Table 5 Identification achieved by phylogenetic analysis using Neighbor Joining (NJ) and various methods, obtained from models test

Barcodes	Match/mismatch (NJ method)	Match/mismatch (Model test method)
rbcL	6/(8)	6/8 (K2 + G)
matK	8/(6)	8/6 (GTR + I)
rbcL + *matK*	8/(6)	8/6 (T92 + I)

For each, Bootstrap replicates = 1000

K2 + *G* Kimura 2 + Gamma distribution, *GTR* + *I* Generalised time reversible + proportion of invariable sites (I), *T92* + *I* Tamura 1992 Model + proportion of invariable sites (I)

Conclusions

DNA barcoding can be a very effective tool to identify mangroves. Here, we tested DNA barcodes of plant plastid DNA, *rbcL* and *matK* to resolve available mangrove species. For the single barcode region, *matK* had the highest rate of correct identification using BM and BCM than *rbcL*. When both regions were concatenated (*rbcL* + *matK*) their efficiency to resolve species was 66.6 % using BM and BCM criteria. In the present work, we lay the foundation towards DNA barcoding applications for mangroves plant genera viz. *Acanthus, Kandelia*, *Ceriops*, *Bruguiera*, *Aegiceras* and *Excoecaria*. *matK* is proposed to be a suitable candidate DNA barcode marker for mangrove species identification. Compiled mangroves barcoding result had some limitations, most of which are due to imperfect discrimination ability of the markers, natural hybridization and homoplasy. Further need to explore with additional markers which may improve mangrove species identification for practical conservation.

Additional file

Additional file 1. Figure S1. Photos of 14 mangroves species. **Table S1**. Morphological key features.

Abbreviations

RbcL: ribulose bisphosphate carboxylase large subunit; *matK*: maturase K.

Authors' contributions

AAS collected the samples, performed the experiments and drafted the manuscript. RAJ helped in data analysis. KK conceived of the study, participated in its design, coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Biological Sciences, Birla Institute of Technology and Science Pilani, K. K. Birla Goa Campus, Sancoale, Goa 403726, India. [2] Department of Zoology, Yashwantrao Chavan Institute of Science, Satara, Maharashtra 415001, India.

Acknowledgements

AAS acknowledge the Senior Research Fellowship provided by University Grant Commission, India. Authors thank Mr. Vinod Naik, Department of Forest, Goa for his kind help in mangrove sample collection. This work was supported by research initiation grant from the Institute.

Competing interests

The authors declare that there is no competing interests.

Received: 15 February 2016 Accepted: 1 September 2016
Published online: 13 September 2016

References

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

Blaalid R, Kumar S, Nilsson HR, Abarenkov K, Kirk MP, Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol Ecol Resour 13:218–224

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH (2011) Discriminating plants pecies in a local temperate flora using the *rbcL* + *matK* DNA barcode. Method Ecol Evol 2:333–340

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

Chase MW (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann Missouri Bot Gar 80:528–580

Chen J, Zhao J, Erickson DL, Xia N, Kress WJ (2015) Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. Mol Ecol Resour 15:337–348

China Plant BOL Group (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. PNAS USA 108:19641–19646

D'Souza J, Rodrigues BF (2013) Biodiversity of arbuscular mycorrhizal (AM) fungi in mangroves of Goa in West India. J For Res 24:515–523

Dhargalkar VK, D'Souza R, Kavlekar DP, Untawale AG (2014) Mangroves of Goa. Forest department, Government of Goa and Mangroves society of India, Goa, India, pp 12–31

Duke NC (1992)Tropical mangrove ecosystems (Coastal and Estuarine studies). In: Robertson AI, Alongi DM (eds) American geophysical union, Washington, pp 63–100

Duke NC, Ge XJ (2011) *Bruguiera* (Rhizophoraceae) in the Indo-West Pacific: a morphometric assessment of hybridization within single flowered taxa. Blumea Biodiver Evol Biogeogr Plants 56:36–48

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3:e2802

Food and Agriculture Organization (2007) The world's mangroves 1980–2005. United Nations, Rome, pp 1–6

Ford CS, Ayres KL, Toomey N, Haider N, Van Alphen Stahl J et al (2009) Selection of candidate coding DNA barcoding regions for use on land plants. Bot J Linean Soc 159:1–11

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

Government of India (1997) The State of Forest Report. Forest Survey of India, Ministry of Environment & Forests. New Delhi, vol 38, pp 5–6

Hilu KW, Borsch T, Muller K et al (2003) Angiosperm phylogeny based on *matK* sequence information. Am J Bot 90:1758–1776

Hollingsworth ML, Clark A, Forrest LL et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. Mol Ecol Resour 9:439–457

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6:e19254

Hutchings PA, Saenger P (1987) Ecology of mangroves. Australian Ecology Series University of Queensland Press, St Lucia, pp 14–44

Jagtap TG, Chavan VS, Untawale AG (1993) Mangrove ecosystems of India: a need for protection. Ambio 22:252–254

Kress J, Erickson DL (2007) A two locus global DNA barcode for land plants: the Coding *rbcL* Gene complements the non-coding trnH-psbA spacer region. PLoS ONE 2:e508

Kress WJ, Erickson DL, Swenson NG et al (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. PLoS ONE 5:e15409

Kuzmina ML, Johnson KL, Barron HR, Hebert PDN (2012) Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. BMC Ecol 12:25

Lakshmi M, Rajalakshmi S, Parani M, Anuratha CS, Parida A (1997) Molecular phylogeny of mangroves I. Use of molecular markers in assessing the intra-specific gene variability in the mangrove species, *Acanthus ilicifolius Linn.* (Acanthaceae). Theor Appl Genet 94:1121–1127

Lakshmi M, Parani M, Ram N, Parida A (2000) Molecular phylogeny of mangroves IV. Intraspecific genetic variation in mangrove species *Excoecaria agallocha L.* (Euphorbiaceae). Genome 43:110–115

Leon-Romero Y, Mejia O, Soto-Galera E (2012) DNA barcoding reveals taxonomic conflicts in the *Herichthys bartoni* species group (Pisces: Cichlidae). Mol Ecol Resour 12:1021–1026

Li HQ, Chen JY, Wang S, Xiong SZ (2012) Evaluation of six candidate DNA barcoding loci in Ficus (Moraceae)of China. Mol Ecol Resour 12:783–790

Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. Biol Rev 90:157–166

Little DP, Stevenson DW (2007) A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 23:1–21
MacNae M (1968) A general account of the fauna and flora of mangrove swamps and forest in the Indo-west pacific region. Adv mar biol 6:73–270
Mandal RN, Naskar KR (2008) Diversity and classification of Indian mangroves: a review. Trop Ecol 49:131–146
Meier R, Kwong S, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. Syst Biol 55:715–728
Naskar KR, Mandal NR (1999) Ecology and biodiversity of Indian mangroves. Daya Publishing House, New Delhi, pp 386–388
Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour 8:480–490
Parani M, Lakshmi M, Elango S, Ram N, Anuratha CS, Parida A (1997a) Molecular phylogeny of mangroves II. Intra and inter-specific variation in *Avicennia* revealed by RAPD and RFLP markers. Genome 40:487–495
Parani M, Rao CS, Mathan M, Anuratha CS, Narayanan KK, Parida A (1997b) Molecular phylogeny of mangroves III. Parentage analysis of Rhizophora hybrid using random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers. Aquat Bot 58:165–172
Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW et al (2013) How effective are DNA barcodes in the Identification of African Rainforest Trees? PLoS ONE 8:e54921
Polidoro BA, Carpenter KE, Collins L, Duke NC, Ellison AM et al (2010) The loss of species: mangrove extinction risk and geographic areas of global concern. PLoS ONE 5:e10095
Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). Mol Ecol Notes 7:355–364
Ricklefs RE, Latham RE (1993) Global patterns of diversity in mangrove floras. In: Ricklefs RE, Schluter D (eds) Species diversity in ecological communities: historical and geographical perspectives. University of Chicago Press, Chicago, pp 215–230
Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. Syst Biol 57:216–230
Rozas J (2009) Polymorphism Analysis using DnaSP. In: Posada D (ed) Bioinformatics for DNA sequence analysis; methods in molecular biology series. Humana Press, New Jersey, vol 537, pp 337–350
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual, vol I, Cold Spring Harbor Laboratory Press, New York, pp 5.1–5.17
Sanyal P, Mandal RN, Ghosh D, Naskar KR (1998) Studies on the mangrove patch at Subarnarekha river mouth of Orissa state. J Interacad 2:140–149
Schwarzbach AE, Ricklefs RE (2000) Systematic affinities of Rhizophoraceae and Anisophylleaceae and intergeneric relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. Am J Bot 87:547–564
Setoguchi H, Kosuge K, Tobe H (1999) Molecular phylogeny of Rhizophoraceae based on *rbcL* gene sequences. J Plant Res 112:443–455
Setyawan AD, Ulumuddin YI, Ragavan P (2014) Review: Mangrove hybrid of Rhizophora and its parental species in Indo-Malayan region. Nusantara Biosci 6:69–81
Shi S, Huang Y, Zeng K, Tan F, He H, Huang J, Fu Y (2006) Molecular phylogenetic analysis of mangroves: independent evolutionary origins of vivipary and salt secretion. Mol Phylogenet Evol 40:298–304
Spalding M, Kainuma M, Collins L (2010) World atlas of mangroves. Earthscan eBook, pp 1–11
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729
Tomlinson PB (1986) The botany of mangroves. Cambridge University Press, Cambridge, pp 1–22
Untawale AG (1985) Mangroves of India: present status and multiple use practices, UNDP/UNESCO Regional Mangrove Project, pp 67
Untawale AG, Jagtap TG (1992) Floristic composition of the deltaic regions of India. Memoirs Geol Soc India 22:243–263
Vinitha RM, Kumar SU, Aishwarya K, Sabu M, Thomas G (2014) Prospects for discriminating Zingiberaceae species in India using DNA barcodes. J Integr Plant Biol 56:760–773
Xiang XG, Hu H, Wang W, Jin XH (2011) DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Mol Ecol Resour 11:1012–1021
Yong WHJ, Sheue CR (2014) A comparative guide to the asian mangroves. doi:10.13140/2.1.3765.1843
Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). Mol Ecol Resour 12:98–108

PLOS ONE

RESEARCH ARTICLE

Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India

Ankush Ashok Saddhe[1], Rahul Arvind Jamdade[2], Kundan Kumar[1]*

1 Department of Biological Sciences, Birla Institute of Technology & Science Pilani, K. K. Birla Goa Campus, Goa, India, **2** Sharjah Research Academy, University of Sharjah, Sharjah, United Arab Emirates

* kundan@goa.bits-pilani.ac.in

OPEN ACCESS

Citation: Saddhe AA, Jamdade RA, Kumar K (2017) Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. PLoS ONE 12(8): e0183245. https://doi.org/10.1371/journal.pone.0183245

Editor: Manoj Prasad, National Institute of Plant Genome Research, INDIA

Received: April 2, 2017

Accepted: August 1, 2017

Published: August 17, 2017

Data Availability Statement: All relevant data are within the paper and submitted to Barcode of Life Data Systems (BOLD) database under the project code IMDB with their taxonomic and sampling details (doi:10.5883/DS-IMDBNG).

Funding: This work was supported by financial assistance from the Council of Scientific and Industrial Research [38(1416)/16/EMR-II], India. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The plant DNA barcoding is a complex and requires more than one marker(s) as compared to animal barcoding. Mangroves are diverse estuarine ecosystem prevalent in the tropical and subtropical zone, but anthropogenic activity turned them into the vulnerable ecosystem. There is a need to build a molecular reference library of mangrove plant species based on molecular barcode marker along with morphological characteristics. In this study, we tested the core plant barcode (*rbcL* and *matK*) and four promising complementary barcodes (ITS2, *psbK-psbI*, *rpoC1* and *atpF-atpH*) in 14 mangroves species belonging to 5 families from West Coast India. Data analysis was performed based on barcode gap analysis, intra- and inter-specific genetic distance, Automated Barcode Gap Discovery (ABGD), TaxonDNA (BM, BCM), Poisson Tree Processes (PTP) and General Mixed Yule-coalescent (GMYC). *matK*+ITS2 marker based on GMYC method resolved 57.14% of mangroves species and TaxonDNA, ABGD, and PTP discriminated 42.85% of mangrove species. With a single locus analysis, ITS2 exhibited the higher discriminatory power (87.82%) and combinations of *matK* + ITS2 provided the highest discrimination success (89.74%) rate except for *Avicennia* genus. Further, we explored 3 additional markers (*psbK-psbI*, *rpoC1*, and *atpF-atpH*) for *Avicennia* genera (*A. alba*, *A. officinalis* and *A. marina*) and *atpF-atpH* locus was able to discriminate three species of *Avicennia* genera. Our analysis underscored the efficacy of *matK* + ITS2 markers along with *atpF-atpH* as the best combination for mangrove identification in West Coast India regions.

Introduction

Plant DNA barcoding is more complex than animal DNA barcoding and it often requires more than one locus approach. The mitochondrial cytochrome oxidase I (COI) gene fragment is considered as the universal animal barcode. Plant mitochondrial COI was excluded from the barcoding, due to the low substitution rates [1–3]. Later, the Consortium for the Barcode of Life (CBOL) evaluated 7 leading candidate DNA regions (*matK*, *rbcL*, *trnH–psbA* spacer, *atpF–atpH* spacer, *rpoB*, *rpoC1*, and *psbK–psbI* spacer) [4]. The CBOL recommended

Competing interests: The authors have declared that no competing interests exist.

two-locus combinations of *rbcL* and *matK* as the core plant barcode complemented with *trnH-psbA* intergenic spacer based on the parameters of recoverability, sequence quality, and levels of species discrimination, CBOL [4–6]. China Plant Barcode of Life recommended the internal transcribed spacer (ITS) as an additional candidate plant DNA barcode [7]. Comparative studies of seven markers *psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, ITS2, and ITS from medicinal plant species were performed. Authors recommended that ITS2 is the best potential marker which discriminated 92.7% plants at the species level in more than 6600 plant samples [8]. The potential discriminating DNA barcode varies from one botanical family to other. The plastid marker *matK* can differentiate more than 90% of species in the Orchidaceae (Orchid family) but less than 49% in the Myristicaceae (nutmeg family) [9–10]. However, identification of 92 species from 32 genera using multilocus markers (coding regions (*rpoB*, *rpoC1*, *rbcL*, *matK* and 23S rDNA) and non-coding (*trnH-psbA*, *atpF–atpH*, and *psbK–psbI*) could achieve 69%-71% with several combinations [3]. More than two loci can improve the plant identification success rate; a recent example of the flora of Canada revealed 93% success in species identification with *rbcL* and *matK*, while the addition of the *trnH-psbA* intergenic spacer achieved discrimination up to 95% [11]. *rbcL* and *matK* loci showed poor discrimination in species-rich genera and complex taxa of *Lysimachia*, *Ficus*, *Holcoglossum*, and *Curcuma* [12–15]. The lowest discriminatory power was observed in closely related groups of *Lysimachia* with *rbcL* (26.5–38.1%), followed by *matK* (55.9–60.8%) and combinations of core barcodes (*rbcL* + *matK*) had discrimination of 47.1–60.8% [15]. Beside all these markers, several plastid regions such as *ycf1*, *atpF-H*, *psbK-psbI*, *ropC1*, *rpoB*, and *trnL-trnF* were frequently evaluated as plant barcode. However, the application of DNA barcoding has been hindered owing to the difficulty in distinguishing closely related species, especially in recently diverged taxa.

Mangroves are unique component of the coastal ecosystem of the world with a niche distribution in tropical and subtropical climates [16]. They are adapted to the local environment like fluctuated water level, salinity and anoxic condition through special features such as aerial breathing and extensive supporting roots, buttresses, salt-excreting leaves and viviparous propagules [17–18]. Plant mangrove species comprise 70 species belonging to about 20 families and 27 genera [19–20]. The West Coast of India is more or less steeply shelved, lack major deltas, river estuaries and dominated by sandy and rocky substratum. The West Coast also harbors one of the world's biodiversity hotspot of Western Ghats in India. It includes the states of Gujarat, Maharashtra, Goa, Karnataka, and Kerala, which harbors 37 species (25 genera under 16 families). The most dominant mangrove species found along the West Coast of India are *Rhizophora mucronata*, *R. apiculata*, *Bruguiera gymnorrhiza*, *B. parviflora*, *Sonneratia alba*, *S. caseolaris*, *Cariops tagal*, *Heretiera littoralis*, *Xylocarpus granatum*, *X. molluscensis*, *Avicennia officinalis*, *A. marina*, *Excoecaria agallocha* and *Lumnitzera racemosa* [21].

In the previous study, we reported the efficacy of single and concatenation of *rbcL* and *matK* marker which resolved *Acanthus*, *Excoecaria*, *Aegiceras*, *Kandelia*, *Ceriops* and *Bruguiera* genus perfectly, but were unable to delimit species-rich genera such as *Rhizophora*, *Avicennia* and *Sonneratia* [17]. In the present work, we comprehensively evaluated the potential of ITS2, concatenated ITS2+*matK*, *atpF-atpH*, *psbK-psbI* and *ropC1*markers for 14 mangroves species. The evaluation was based on genetic distance, diagnostic nucleotide characters, Neighbour-joining (NJ) Kimura 2 Parameter (K2P) tree, TaxonDNA, Automated Barcode Gap Discovery (ABGD), Poisson tree process (PTP) and Generalized mixed Yule- Coalescent model (GMYC) analysis.

Material and methods

Ethics statement

The mangrove samples were collected from different parts of Goa, west coast region, with the permission from the Principal Chief Conservator of Forest, Goa Forest Department, Goa, India. Further, none of the species are endangered or protected species.

Mangrove plant sampling

In the present study, a total of 44 specimens of mangroves belonging to 14 species, 9 genera and 5 families were collected from Goa region, West Coast of India with geographical co-ordinates latitude of 15.5256˚ N and longitude of 73.8753˚ E. The selected genera of mangroves such as *Rhizophora*, *Bruguiera*, *Avicennia*, and *Sonneratia* each represented by at least two species and *Aegiceras*, *Excoecaria*, *Ceriops*, *Kandelia*, *Acanthus* genus were represented by single species. Mangrove species were identified based on morphological keys [22] and mounted on herbarium sheets, photographed and deposited at the Botanical Survey of India, Western Regional Centre, Pune, India as barcode vouchers [17]. The well-identified voucher specimens along with their taxonomic information, collection details, and GenBank accession numbers were described in Table 1. For each specimen, leaf tissue was collected in the field, labeled and stored in -80˚ C for further analysis.

DNA extraction

Genomic DNA was isolated from mangrove species by modified cetyl-trimethyl ammonium bromide (CTAB) protocol [17]. Leaf tissue was homogenized in liquid nitrogen and CTAB

Table 1. Details of the mangrove species.

S. No.	Specimen	Voucher No.	Accession No. of ITS2
A			
1	*Avicennia officinalis*	AAS-100-02	KU876892, KU876893
2	*Avicennia marina*	AAS-110-12	KU876889, KU876890, KU876891
3	*Avicennia alba*	AAS-120-22	KU876886, KU876887, KU876888
4	*Acanthus ilicifolius*	AAS-230-32	KY250442, KY250443
5	*Bruguiera cylindrica*	AAS-130-32	KU876894, KU876895, KU876896
6	*Bruguiera gymnorrhiza*	AAS-140-42	KU876897, KU876898, KU876899
7	*Rhizophora mucronata*	AAS-150-52	KU876910, KU876911, KY250446
8	*Rhizophora apiculata*	AAS-160-62	KU876908, KU876909, KY250445
9	*Kandelia candel*	AAS-190-92	KU876906, KU876907, KY250444
10	*Ceriops tagal*	AAS-200-02	KU876900, KU876901, KU876902
11	*Excoecaria agallocha*	AAS-180-82	KU876903, KU876904, KU876905
12	*Aegiceras corniculatum*	AAS-170-73	KU876881, KU876882, KU876883, KU876884
13	*Sonneratia caseolaris*	AAS-220-22	KY250450, KY250451
14	*Sonneratia alba*	AAS-210-12	KY250447, KY250448, KY250449

B				
S. No.	Specimen	*atpF-atpH*	*psbK-psbI*	*rpoC1*
1	*Avicennia officinalis*	KY754573, KY754574, KY754575	KY754564, KY754565, KY754566	KY754187, KY754188, KY754189
2	*Avicennia marina*	KY754570, KY754571, KY754572	KY754561, KY754562, KY754563	KY754184, KY754185, KY754186
3	*Avicennia alba*	KY754567, KY754568, KY754569,		

Details of the mangrove species with accession numbers used in the present study for ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* with voucher number and GenBank accession numbers.

https://doi.org/10.1371/journal.pone.0183245.t001

buffer containing 2% PVP-30 and 1% β-mercaptoethanol was mixed. The suspension was incubated at 60˚C for 60 min and centrifuged at 14000 rpm for 10 min at room temperature. It was further extracted with equal volume of chloroform: isoamyl alcohol (24:1) and precipitated with cold isopropanol (-20˚C) and ammonium acetate. The precipitated DNA was washed with 70% ethanol and finally dissolved in TE buffer. Quantity and quality of the DNA samples were confirmed by agarose gel electrophoresis and Nanodrop (Thermo Scientific, USA).

PCR and sequencing

PCR amplification of ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* were carried out in the 50-μl reaction mixture containing 10-20ng of template DNA, 200 μM of dNTPs, 0.1 μM of each primer and 1 unit of Taq DNA polymerase (Thermo Scientific, USA). The reaction mixture was amplified in Bio-Rad (T100 model) thermal cycler with temperature profile for ITS2 (94˚C for 4 min; 35 cycles of 94˚C for 30 sec, 56˚C for 30 sec, 72˚C for 1 min; final extension 72˚C for 10 min), *atpF-atpH* (94˚C for 1 min; 35 cycles of 94˚C for 30 sec, 50˚C for 40 sec, 72˚C for 40 sec; final extension 72˚C for 5 min), *psbK-psbI* (94˚C for 5 min; 35 cycles of 94˚C for 30 sec, 55˚C for 30 sec, 72˚C for 45 sec; final extension 72˚C for 10 min), *rpoC1* (94˚C for 5 min; 35 cycles of 94˚C for 30 sec, 55˚C for 30 sec, 72˚C for 45 sec; final extension 72˚C for 10 min). The amplified products were separated by agarose gel (1.2%) electrophoresis and stained with ethidium bromide. The primers used for amplification were listed (Supporting information S1 Table). PCR products were purified as per manufacturer's instruction (Chromous Biotech) and further sequencing reactions were carried out using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on ABI 3500xL Genetic Analyzer (Applied Biosystems).

Data analysis

Sequence assembly and alignment were performed in Codon Code Aligner v.3.0.1 (Codon Code Corporation) and MEGA 6.0.6 respectively [23]. All sequences were submitted to Barcode of Life Data Systems (BOLD) database under the project code IMDB with their taxonomic and sampling details (doi:10.5883/DS-IMDBNG) [24]. Nucleotide diagnostic characters of mangrove species were analyzed based on the BOLD system. Further, *matK* and ITS2 sequences were concatenated using DNASP v5.10 tool and analyzed in MEGA 6 [25]. NJ trees were constructed using MEGA 6.0 and Kimura 2 parameter (K2P) genetic distance model with node support based on 1000 bootstrap replicates.

TaxonDNA

TaxonDNA v1.6.2 analysis for species identification with 'Best Match' and 'Best Closest Match' method was performed [17, 26]. The threshold (T) was set at 95%. All the results above the threshold (T) were treated as 'incorrect'. Similarly, if all matches of the query sequence were below threshold (T), the barcode assignment was considered to be the 'correct' identification. If the matches of the query sequences were good and corresponded to a mixture of species, then it was treated as ambiguous identification.

Automated Barcode Gap Discovery (ABGD)

The ABGD, is a web server based distance method, which can partition the sequences into potential species based on the barcode gap whenever the divergence within the same species is smaller than organisms from different species [27–29]. The ABGD analysis was performed

with two relative gap width (X = 1.0, 1.5) and three distance metrics (Jukes-Cantor, K2P, and p-distance) with default parameters.

General Mixed Yule-coalescent (GMYC)

The GMYC method requires a fully resolved ultrametric tree for analysis. This Bayesian tree was built using BEAST v1.8 [30–31]. Input file (XML) for BEAST was compiled in BEAUti v1.83 with an HKY+G molecular evolutionary model for the ITS2 dataset and GTR+G for concatenated dataset of *matK*+ITS2. These models were derived using PartitionFinder V1.1.1. Tree prior was set to Yule Process and the length of Markov chain Monte Carlo (MCMC) chain was 40,000,000 generation and sampling was performed at every 4000 step. However, all other settings were kept as default. Convergence of the BEAST runs to the posterior distribution. The adequacy of sampling (based on the Effective Sample Size (ESS) diagnostic) was assessed with Tracer v1.4. After removing the first 20% of the samples as burn-in, all other runs were combined to generate posterior probabilities of nodes from the sampled trees using TreeAnnotator v1.7.4. Estimation of the number of species included in the tree was analyzed using GMYC with single and multiple thresholds in R by the APE and SPLITS packages [27, 30–36].

Poisson Tree Process model (PTP)

The PTP model is a tree-based method that differentiates specimen into populations and species level using coalescence theory [27–29] The RaxML tree was constructed using CIPRES portal and input data was generated for bPTP analysis. The calculations were conducted on the bPTP web server (http://species.h-its.org), with the following parameters (500,000 MCMC generations, thinning 100 and burn-in 25%).

Results

Sequence analysis

A total of 148 sequences (44 *rbcL*, 43 *matK*, 40 ITS2, 9 *atpF-atpH*, 6 *psbK-psbI* and 6 *rpoC1*) were acquired from 44 specimens of mangrove belonging to 14 species, 9 genera, and 5 families. The sequences (*rbcL*: 510bp, *matK*: 712bp, ITS2: 445bp, *atpF-atpH*: 511bp, *psbK-psbI*: 360bp and *rpoC1*: 451bp) with few insertions and deletions, without stop codon, along with the specimen collection details were submitted to the Barcode of Life Data Systems (BOLD) in form of a project 'IMDB' (dx.doi.org/10.5883/DS-IMDBNG). These sequences were submitted to the NCBI GenBank through BOLD systems and their accession numbers were obtained (Table 1). The scatter plot represented the number of individuals in each species against their maximum intra-specific distances, as a test for sampling bias (Fig 1). Previous evaluation of DNA barcode using *rbcL* and *matK* demonstrated 47.72% and 72.09% efficiency in resolving mangrove taxa respectively. The *matK* sequence region showed better efficiency than the *rbcL* for resolution of mangrove taxa [17]. In the present study, *matK* along with ITS2 and few supplementary markers (*atpF-atpH*, *psbK-psbI* and *rpoC1*) were used for the species identification of the cryptic mangrove taxa. Sequence analysis was performed to estimate the average GC content of the corresponding locus. The average GC content observed was 63.11%, 42.7%, 35.18%, 31.22% and 44.6% for ITS2, *matK*+ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* locus respectively.

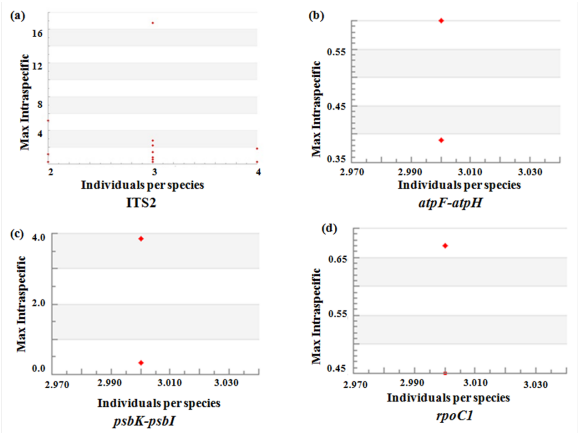


Fig 1. Scatter plot. The scatter plot represents the number of individuals in each species against their maximum intra-specific distances, as a test for sampling bias. (a) ITS2 locus (b) *atpF-atpH* locus (c) *psbK-psbI* locus and (d) *rpoC1* locus.

https://doi.org/10.1371/journal.pone.0183245.g001

Genetic divergence analysis

The genetic distances were calculated for individual barcode marker by K2P model on the BOLD system. The mean intraspecific distance for ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* was calculated as 1.85%, 0.11%, 1.63% and 0.37% respectively. While mean intrageneric distance for ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1*was calculated as 5.8%, 1.03%, 2.16% and 0.3% respectively (Table 2). Higher intraspecific distances (>2%) for ITS2 were observed in 19.51% individuals and *S. alba* exhibited highest intraspecific distance of 16.75%. While lower intrageneric distances (<2%) for ITS2 were observed in 50.98% individuals and *A. marina* showed the lowest intrageneric distance of 0%. Higher intraspecific distances for *matK*+ITS2 were observed in 9.30% individuals and *S. alba* exhibited the highest distance of 4.01%. While lower intrageneric distances were observed in almost 90.69% individuals (Table 2). In some species intraspecific distance was higher than the intrageneric distance (Fig 2A and 2B). Six species (*A.*

Table 2. Distance summary.

Barcode	Level	N	Taxa	Comparisons	Min Dist (%)	Mean Dist (%)	Max Dist (%)	SE Dist (%)
ITS2	Species	40	14	39	0	1.85	16.75	0.1
	Genus	25	4	45	0	5.8	35.14	0.25
	Family	28	2	133	5.72	12.35	40.26	0.08
***matK* + ITS2**	Species	39	14	37	0	0.51	4.02	0.02
	Genus	24	4	43	0	1.76	7.84	0.05
	Family	28	2	133	3.35	7.39	19.89	0.03
atpF-atpH	Species	9	3	9	0	0.11	0.6	0.02
	Genus	9	1	27	0.39	1.03	1.62	0.02
psbK-psbI	Species	6	2	6	0	1.63	3.85	0.27
	Genus	6	1	9	0.96	2.16	4.94	0.14
rpoC1	*Species*	6	2	6	0.22	0.37	0.67	0.03
	Genus	6	1	9	0	0.3	0.67	0.02

Summary distribution of sequence divergence at the species, genus and family level is summarized (Distance summary result—BOLD system). N—Number of sequences.

https://doi.org/10.1371/journal.pone.0183245.t002

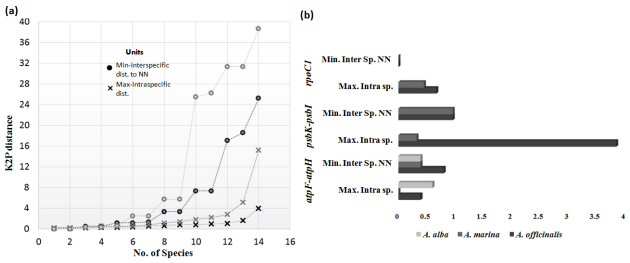


Fig 2. Genetic distance. Distribution of intra and inter specific K2P mean divergence (arranged in ascending order). (a) ITS2 and ITS2+*matK* (concatenated) are represented by grey and black colors respectively. (b) For *atpF-atpH*, *psbK-psbI and rpoC1* markers maximum intraspecific distance and minimum interspecific distance to nearest neighbor are represented by a bar.

https://doi.org/10.1371/journal.pone.0183245.g002

alba, *A. officinalis*, *A. marina*, *B. cylindrica*, *B. gymnorrhiza* and *R. mucronata*) were resolved with ITS2, while in concatenation of *matK*+ITS2, error rates were minimized in two species (*A. officinalis* and *A. marina*). *Avicennia* genus in the former and current analysis has revealed low resolution. To resolve this cryptic genus, we used few supplementary markers such as *atpF-atpH*, *psbK-psbI* and *rpoC1*. *Avicennia* genus showed intraspecific distance ranging from 0%-1.0% with almost all barcode markers, with highest intraspecific distance (>2%) was observed in *psbK-psbI* (3.85%) (Fig 2B, Table 3). While lower intrageneric distance (<2%) was observed in nearly all barcode markers, except for *psbK-psbI* (4.94%).

Diagnostic character based delineation of mangrove species was done using four barcode markers (ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1*) along with concatenated *matK*+ITS2 with minimum 3 specimens per species. Highest diagnostic characters were observed in ITS2 for *Excoecaria agallocha* (34) and *Aegiceras corniculatum* (35), whereas single diagnostic character was observed in the species of *Avicennia* genera followed by *Bruguiera cylindrica* (Table 4). In concatenated *matK*+ITS2, highest diagnostic characters were observed in *Aegiceras corniculatum* (96) and *Excoecaria agallocha* (60). However, all species of *Avicennia* genera revealed diagnostic characters, but *Bruguiera gymnorrhiza* failed to exhibit any diagnostic character. The supplementary marker *rpoC1* failed to show any diagnostic character in *Avicennia*, while *atpF-atpH* and *psbK-psbI* exhibited diagnostic characters (Table 4).

Taxonomic assignment

Altogether 40 DNA barcodes from ITS2 and *matK*+ITS2 were used for species delineation. The Neighbour-Joining (K2P) trees constructed with bootstrap support (1000) and bootstrap values of >60 exhibited substantial resolution among the OTUs corresponding to their genera except for *A. marina* and *A. officinalis* (Supporting information S1 Fig).

Table 3. Mean divergence of *Avicennia* genus.

	atpF-atpH		*psbK-psbI*		*rpoC1*	
	Max. Intraspecific	**Min Interspecific NN**	**Max. Intraspecific**	**Min Interspecific NN**	**Max. Intraspecific**	**Min Interspecific NN**
A. officinalis	0.39	0.8	3.85	0.96	0.67	0
A. marina	0	0.39	0.32	0.96	0.45	0
A. alba	0.6	0.39	NA	NA	NA	NA

Distribution of intra and inter specific K2P mean divergence for *atpF-atpH*, *psbK-psbI and rpoC1* are represented in table for *Avicennia* genus. NN-Nearest Neighbor, Max-Maximum, Min-Minimum.

https://doi.org/10.1371/journal.pone.0183245.t003

Table 4. Diagnostic characters of mangrove taxa.

Barcode	Group Name (sequences)	Diagnostic Characters	Diagnostic or Partial Characters	Partial Characters	Partial or Uninformative Characters
***matK* + ITS2**	*Aegiceras corniculatum* (6)	96	3	0	1
	Avicennia alba (3)	8	0	0	1
	Avicennia marina (3)	5	0	1	1
	Bruguiera cylindrica (3)	2	0	0	0
	Bruguiera gymnorrhiza (3)	0	1	0	0
	Ceriops tagal (3)	5	2	0	0
	Excoecaria agallocha (3)	60	3	0	3
	Kandelia candel (3)	12	0	1	1
	Rhizophora apiculata (3)	2	0	1	23
	Rhizophora mucronata (3)	6	0	0	0
ITS2	*Aegiceras corniculatum (4)*	35	4	0	0
	Avicennia alba (3)	1	0	1	0
	Avicennia marina (3)	1	0	1	0
	Avicennia officinalis (3)	0	0	0	0
	Bruguiera cylindrica (3)	1	0	0	0
	Bruguiera gymnorrhiza (3)	0	0	0	0
	Ceriops tagal (3)	4	1	0	0
	Excoecaria agallocha (3)	34	2	0	1
	Kandelia candel (3)	5	0	1	1
	Rhizophora apiculata (3)	2	0	0	1
	Rhizophora mucronata (3)	6	1	0	0
atpF-atpH	*Avicennia alba (3)*	0	0	0	0
	Avicennia marina (3)	4	0	0	0
	Avicennia officinalis (3)	2	0	0	0
psbK-psbI	*Avicennia marina (3)*	3	0	5	40
	Avicennia officinalis (3)	3	0	1	13
rpoC1	*Avicennia marina (3)*	0	0	1	0
	Avicennia officinalis (3)	0	0	0	0

Identification of diagnostic nucleotides for each of the 14 mangrove taxa recovered from the BOLD system. Based on their utility for mangrove taxa delineating referred as diagnostic characters, diagnostic or partial character, partial characters and partial or uninformative characters.

https://doi.org/10.1371/journal.pone.0183245.t004

Species identification based on barcoding gap

The initial partition of ITS2, K2P with X = 1.0, prior maximal distance $P = 0.021$ produced consistent 12 operational taxonomic units (OTUs). *S. alba* was split into 3 groups, while members of *Rhizophora* and *Avicennia* were merged (Fig 3; Supporting information S2 Table). Whereas, recursive partitioning with $P = 0.00167$, produced inconsistently18 OTUs, of which *A. alba*, *A. officinalis*, and *B. cylindrica* showed split, while *B. gymnorrhiza* was clustered perfectly (Fig 4A). In concatenated *matK*+ITS2, at X = 1.0 for all three metrics, OTUs ranged from 4–11 in the initial partition, but recursive partition tends to exhibit inconsistent OTUs (Fig 4B).

When relative gap width was increased from X = 1.0 to X = 1.5, suddenly OTUs in ITS2 for initial partition was dropped to maximum 7, while recursive partition showed an increase in OTUs, up to 16 at $P = 0.001$. The initial partition for *matK*+ITS2, with X = 1, $P = 0.0129$ produced 11 OTUs. *Avicennia* and *Bruguiera* members were merged, while *S. alba* showed split. In recursive partition, with $P = 0.001$, *A. alba*, *B. cylindrica*, *B. gymnorrhiza* were resolved

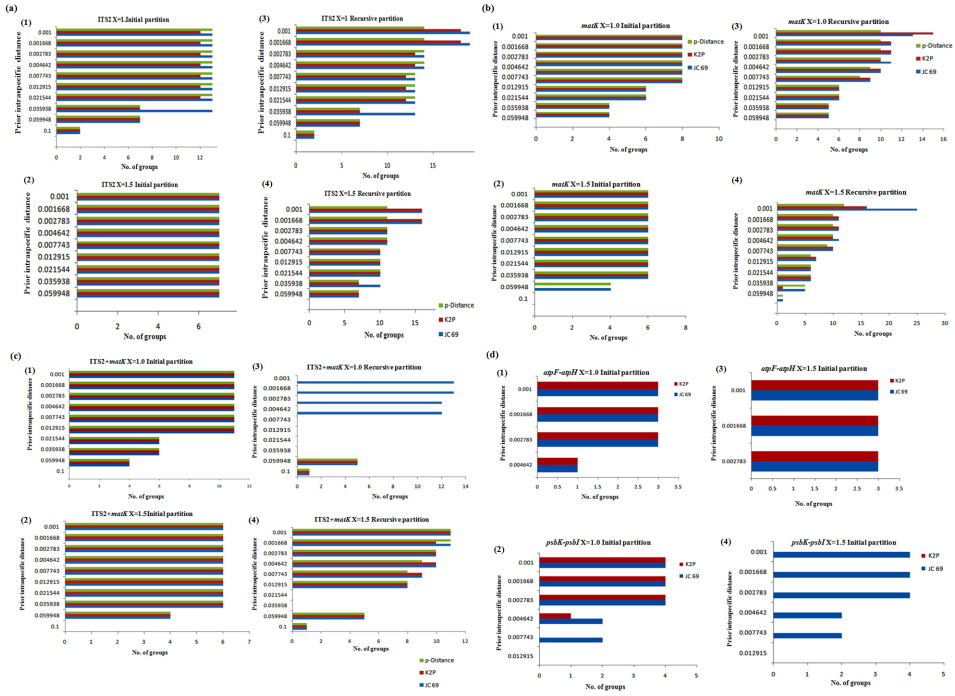


Fig 3. Automated partition. The automatic partition by ABGD with three metrics (JC69, K2P and p-distance) and two X-values (X = 1, 1.5) for (a) ITS2 (initial partition 1,2 and Recursive partition 3 and 4); (b) *matK* (initial partition 1,2 and Recursive partition 3 and 4); (c) ITS2+*matK* (initial partition 1,2 and Recursive partition 3 and 4);(d) *atpF-atpH* and *psbK-psbI* (initial partition 1,2 for *atpF-atpH* and Initial partition 3 and 4 for *psbK-psbI*).

https://doi.org/10.1371/journal.pone.0183245.g003

perfectly, while *A*. *officinalis*, *A*. *marina* along with *R*. *apiculata and R*. *mucronata* remained merged.

The initial partition with an *atpF-atpH* barcode, JC and K2P metrics with (X = 1, 1.5) showed 3 OTUs ($P = 0.0027$) without any recursive partition except (X = 1.5, $P = 0.00278$, 1 OTU). With *atpF-atpH*, at X = 1.5 initial partition with $P = 0.00278$, 3 OTUs were produced in *A*. *alba*, *A*. *officinalis*, and *A*. *marina*. Similarly, *psbK-psbI* showed 4 OTUs ($P = 0.001$) in an initial partition for JC and K2P metrics at X = 1 and p-distance had only 2 OTUs with 1 OTU in the recursive partition. At X = 1.5, only JC and p-distance were able to partition data. JC the initial partition at $P = 0.001$ produced 4 OTUs, while at $P = 0.0046$, produced 2 OTUs. Metrics p-distance predicted 2 OTUs in an initial partition and 1 OTU in the recursive partition. Barcode locus *rpoC1* at X = 1 with JC and K2P metrics showed initial partition of 2 OTUs and the recursive partition at $P = 0.00278$ predicted 1 OTU.

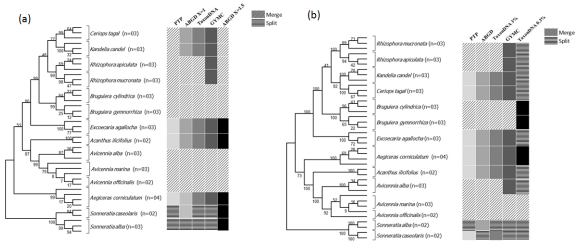


Fig 4. Bayesian phylogenetic tree. Bayesian phylogenetic tree of (a) ITS2 and (b) *matK*+ITS2 gene. Vertical boxes on the right indicate the clades detected by the coalescent-based GMYC, PTP, the distance-based ABGD and TaxonDNA methods.

https://doi.org/10.1371/journal.pone.0183245.g004

Table 5. TaxonDNA analysis.

Barcodes	No. of Sequences	Best Match (%)			Best Closest match (%)				T (%)	No of Cluster	Match / Mismatch
		C	A	Inc	C	A	Inc	No match			
ITS2	40	87.8	2.43	9.75	75.6	2.43	9.75	12.19	3	14	10/4
ITS2 + *matK*	39	89.7	2.56	7.6	84.6	2.56	7.6	5.12	3	11	6/8
atpF-atpH	9	100	0	0	100	0	0	0	0.3	3	3/0
psbK-psbI	6	50	0	50	50	0	50	0	0.8	4	1/1
rpoC1	6	33.33	66.66	0	33.33	66.6	0	0	3	1	0/2

TaxonDNA is an alignment-based method based on sequence distance matrices. Percentage of correct/incorrect/ambiguous assignment of a taxon is compared using the molecular operating taxonomic unit (MOTU). The species-specific clustering was performed using match and mismatch criteria. T -Threshold; C–Correct; A–Ambiguous; Inc–Incorrect.

https://doi.org/10.1371/journal.pone.0183245.t005

Species identification and assignment based on TaxonDNA

The single barcode marker ITS2 produced a moderate rate of correct identification using BM (87.8%) and BCM (75.6%) than the concatenated *matK*+ITS2 using BM (89.74%), and BCM (84.61%) (Table 5). ITS2 barcode produced 13 clusters at 3% threshold, of which 5 species (*A. corniculatum*, *A. ilicifolius*, *E. agallocha*, *K. candel* and *C. tagal*) were the perfect match. Whereas, *Avicennia*, *Rhizophora* and *Bruguiera* species were clumped into 3 clusters, while *S. alba* and *S. caseolaris* were split into 5 clusters. As compared to single barcode marker (ITS2), concatenated (*matK*+ITS2) markers at 3% threshold produced 11 clusters, where *S. caseolaris* was successfully resolved. Single barcode *atpF-atpH* demonstrated 100% correct identification in both BM and BCM method for *Avicennia* genera with 3 clusters. *psbK-psbI* locus identified 50% *Avicennia* species in BM and BCM methods, however, *rpoC1* showed lowest identification rate of about 33.33% (Table 5).

Species identification and assignment based on GMYC and PTP

The single threshold GMYC (sGMYC) model generated through BEAST using the ultrametric phylogenetic tree resulted in an identification of 9 Maximum Likelihood (ML) clusters for ITS2 with confidence interval (CI) of 4–9 and 14 ML entities with CI of 4–18 (Threshold time: -0.013035). Similarly, with *matK*+ITS2, 10 ML clusters with CI of 4–10 and 14 ML entities with CI of 4–16 (Threshold time: -0.005793) were identified. The resulting ML entities in ITS2 exhibited 5 species merged in 2 OTUs, while in *matK*+ITS2 only 4 species were merged with exception of *A. alba*. Also, splitting of two species (*S. alba* and *S. caseolaris*) formed additional OTUs (Fig 4A and 4B). The multiple threshold methods (mGMYC) gave two threshold time for ITS2 (-0.013035 and -0.005441) resulting into 9 clusters (CI:4–9) and 17 ML entities (CI:4–17). *matK*+ITS2 gave threshold time of -0.010859 and -0.004847, resulting into 9 clusters (CI:5–11) and 13 ML entities (CI:5–16). However, multiple thresholds overestimated the number of species, so we took a more conservative approach to consider only the results obtained from the single threshold (sGMYC) method. In GMYC, apart from other metrics, three unresolved species *R. apiculata*, *R. mucronata* and *A. alba* were distinctly resolved.

In addition to the above methods used for taxonomic evaluation, maximum likelihood (ML) based approach was added to get an additional perspective towards the species delineation through Poissons Tree Process (PTP). The ML analysis exhibited 10 OTUs with ITS2, where *Avicennia*, *Bruguiera*, *Rhizophora*, *Ceriops*, and *Kandelia* genera were merged while *S. alba* and *S. caseolaris* were split (Fig 4A and 4B). With *matK*+ITS2, 11 OTUs were formed by merging of *Avicennia*, *Bruguiera* and *Rhizophora* genera and *S. alba* was split.

Discussion

There is no consensus regarding perfect plant DNA barcode, however few of plastid and nuclear coding (*rbcL*, *matK*, *rpoB*, and *rpoC1*) and non-coding (*trnH-psbA*, ITS2, *psbK-psbI* and *atpF-atpH*) marker fulfilled the required criteria [3, 9, 37]. The *rbcL* and *matK* are considered as core barcode, which can be further complemented with *trnH-psbA* and ITS2 as plant barcode suggested by China Plant BOL [4, 7]. We employed these markers for molecular identification of mangrove plant species. In our earlier report, we have tested potential barcode candidates *rbcL* and *matK* individual as well as concatenated *rbcL+matK*, which demarcated most of the species such as *A*. *ilicifolius*, *E*. *agallocha*, *A*. *corniculatum*, *K*. *candel*, *C*. *tagal*, *B*. *cylindrica* and *B*. *gymnorrhiza*. An initial analysis was performed based on traditional barcode methods (Barcode gap analysis and NJ tree with the K2P method) [17]. Individual, as well as concatenated *rbcL* and *matK* barcode demarcated almost all mangroves species except for *Rhizophora*, *Sonneratia* and *Avicennia* genera [17]. The Plant CBOL group (2009) reported that only 72% species were resolved using combined *rbcL* and *matK*. A similar result was observed after combining *rbcL* and *matK* from closely related species of *Curcuma* [13]. Moreover, *Avicennia* genera with three species, of which *A*. *alba*, was resolved perfectly using *matK* but *A*. *officinalis* and *A*. *marina* lumped together and unable to resolve at the species level. Low resolution using DNA barcode regions has been documented in many other plants such as the genus *Araucaria* (32%), *Solidago* (17%) and *Quercus* (0%) [38].

A high percentage of bidirectional reads were critical for a successful plant barcoding system, given the low amount of variation that separates many plant species [3–4]. The risk of misassignment can be anticipated due to sequencing error or incomplete bidirectional reads. We observed the significant quality of PCR amplification and sequencing ranged from 95% to 100% in all tested markers. However, for ITS2 barcode, we performed many amplifications and sequencing attempt for *S*. *alba*, *S*. *caseolaris*, and *A*. *ilicifolius* mangroves taxa. Sequencing of *S*. *alba* and *S*. *caseolaris* resulted in a mixed and low-quality chromatogram with unidirectional success. The possible explanation for this kind of situation can be underscored by the presence of either ITS as multiple copies or pseudogene or/and fungal ITS contamination in plant [39]. Species identification success rate using *rbcL+matK* is higher, whereas *rbcL* sequence recovery ranged from 90–100% [4, 38, 40]. Hence, CBOL group recommends *rbcL* primers to possess universality for land plants. As reported by CBOL, the *matK* region showed sequencing success of 90% [4]. The *matK* marker provided 88% sequencing success, with the use of 10 primer pair combinations [3].

Very few reports are available on the DNA barcoding of the mangrove taxa [17, 41]. Lower genetic distances were observed based on K2P among mangrove taxa particularly genera *Rhizophora*, *Sonneratia*, *Avicennia*, and *Bruguiera* based on *rbcL*, *matK* and ITS barcode [41]. Genetic distance ranged from 0.01 to 0.25 for *rbcL* gene, 0.01 to 0.89 for *matK* and 0.01–0.508 for ITS locus [41]. Similar results were observed in our studies, for *rbcL* and *matK* the genetic distance ranged from 0–0.68% and 0–1.32% respectively [17]. The discrimination power of proposed DNA barcode by CBOL Plant Working Group may vary in different plant group [12, 42–43]. Depending on the taxon, the use of additional markers may be needed for discrimination [4].

For single barcode ITS2, ABGD (K2P, X = 1), Taxon DNA (T = 3%) and GMYC produced consistent OTUs with corresponding results. Additionally, GMYC resolved *R*. *apiculata*, *R*. *mucronata*, and *A*. *alba* species. Overall highest taxon assignment was observed as 57.14% in GMYC and taxon resolution was up to 42.85% in ABGD, TaxonDNA, and PTP barcoding methods. However, the resolution of Chlorella-like species (microalgae) produced by GMYC, PTP, ABGD and character-based barcoding methods were variables based on several marker

studies such as *rbcL*, ITS, and *tufA* [27]. Single ITS2 with PTP analysis was not able to resolve *C. tagal* and *K. candel*, which was further improved in the *matK*+ITS2 analysis. Analysis following the above methods, species delimitation through PTP and GMYC was utilized, due to their robustness in the absence of barcoding gap [44]. Even though they are based on different algorithms, both methods calculated the point of transition between species and population [27]. The GMYC method has a theoretically strong background and requires ultrametric gene tree that takes more time to analyse data. In contrast, the PTP is a recently developed method as an alternative to GMYC which requires non- ultrametric gene tree and consumes less time [44–45]. Both the methods revealed sort of identical results, however, the two analyses differed in resolution. In both the methods, five species (*B. cylindrica*, *B. gymnorrhiza*, *A. officinalis* and *A. marina*) in GMYC and seven species (*B. cylindrica*, *B. gymnorrhiza*, *A. alba*, *A. officinalis*, *A. marina*, *R. apiculata* and *R. mucronata*) in PTP were merged into single OTUs, potentially indicating low intraspecific diversity. It reflected that there are many overlooked/cryptic species present within the mangroves. When we performed ABGD with relative gap width X = 1.5 for K2P method, *S. alba*, and *S. caseolaris* species were demarcated, while rest of the mangrove species were split. At a relative gap width (X = 1) about seven species of the mangrove s were merged into single OTU and observed that the ABGD tends to lump species by increasing the number of merged OTUs [29]. Beside this, we also observed inconsistency of OTUs count during initial partition to recursive partition. Recursive partitioning recognizes more OTUs than initial ones, showing their superior capability to deal with variation in sample sizes of the species under study [29]. Moreover, TaxonDNA with a lower threshold value (0.3%) demarcated *B. cylindrica* and *B. gymnorrhiza*. The possible explanation for this might be due to lack of barcode gap resulting in merged OTUs, which can be optimized by analyzing more than 5 sequences per species, and we have used 3 sequences per species [28]. In TaxonDNA analysis, for *rbcL* threshold (T) was observed 0%, a similar result was recorded for *rbcL* in the Zingiberaceae family [46]. However, the threshold (T) for Indian Zingiberaceae family members was recorded as 0.20% for *rbcL* and 0% for *rpoB* and *accD* [43].

Avicennia is the most diverse mangrove genus, comprising eight species, out of which three are endemic to the Atlantic-East Pacific (AEP) region and five are endemic in the Indo-West Pacific region (IWP) [47]. A recent systematic revision of *Avicennia* based on morphological characters formed three groups: (1) *A. marina*; (2) *A. officinalis* and *A. integra*; and (3) *A. rumphiana* and *A. alba* [47]. In the current study, we have included *A. marina*, *A. officinalis*, and *A. alba* species, which were resolved with other barcode markers. Two plastid spacers such as *psbK-psbI* and *atpF-atpH* are recommended as potential plant DNA barcodes based on the flora of the Kruger National Park South Africa as a model system [48]. Similarly, we used three markers (*atpF-atpH*, *psbK-psbI* and *rpoC1*) for cryptic genera *Avicennia* and further evaluated with ABGD and TaxonDNA barcode methods. Both the methods consistently resolved all three *Avicennia* species using an *atpF-atpH* marker. Similarly, phylogenetic reconstruction of *Avicennia* genera based on *trnT-trnD* intergenic spacer region and the *psbA* gene revealed that *A. marina* is sister to the *A. officinalis*/*A. integra* and *A. alba* is genetically distinct [47].

Conclusions

In the present study, we tested core DNA barcode *rbcL*, *mat*K, ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* to resolve mangroves species. Individual, as well as concatenated *matK*+ITS2 are helpful to demarcate mangroves at the species level. Single barcode *matK* is sufficient to resolve *A. ilicifolius*, *A. corniculatum*, *E. agallocha*, *Ceriops tagal*, *K. candel*, *B. cylindrica and B. gymnorrhiza*. ITS2 was able to discriminate *R. apiculata* and *R. mucronata* species based on GMYC method, while *A. alba* was resolved by concatenation of *matK*+ITS2. A cryptic genus *Avicennia* was

delimitated based on the *atpF-atpH* single barcode. In the present work, the foundation work was done towards DNA barcoding of mangroves plant genera, such as *Rhizophora*, *Avicennia*, *Acanthus*, *Kandelia*, *Ceriops*, *Bruguiera*, *Aegiceras* and *Excoecaria*. Compiled mangroves barcoding result had some limitations, most of which are due to the low mangrove taxa sample coverage. Further, there is a need to explore additional mangroves taxa which will improve mangrove species identification for practical conservation.

Supporting information

S1 Table. List of primers used in the current study.
(DOCX)

S2 Table. Automated Barcode Gap Discovery web server based analysis of all barcodes (*matK*, ITS2, *matK*+ITS2, *atpF-atpH*, *psbK-psbI* and *rpoC1* using two relative gap width (X = 1 and 1.5) and three different matrices such as JC, K2P, and p-simple distance.
(DOCX)

S1 Fig. Neighbor-joining tree (Kimura 2 Parameter distance using bootstrap value of 1000 replicates) *matK*+ITS2 concatenated NJ (K2P).
(DOCX)

Acknowledgments

Authors are thankful to the Council of Scientific and Industrial Research, India for providing the financial assistance (38(1416)/16/EMR-II). AAS acknowledge the Senior Research Fellowship provided by University Grant Commission, India. Authors are grateful to Dr. V. Arunachalam and Dr. Shalini Upadhyay for critical reading and editing of the manuscript.

Author Contributions

Conceptualization: Kundan Kumar.

Data curation: Ankush Ashok Saddhe.

Funding acquisition: Kundan Kumar.

Investigation: Ankush Ashok Saddhe, Rahul Arvind Jamdade, Kundan Kumar.

Methodology: Ankush Ashok Saddhe, Rahul Arvind Jamdade.

Project administration: Kundan Kumar.

Resources: Kundan Kumar.

Supervision: Kundan Kumar.

Validation: Ankush Ashok Saddhe, Rahul Arvind Jamdade, Kundan Kumar.

Writing – original draft: Ankush Ashok Saddhe, Rahul Arvind Jamdade.

Writing – review & editing: Kundan Kumar.

References

1. Hebert PDN, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Biol Sci SerB. 2003; 270: S96–S99.
2. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc Biol Sci SerB. 2003; 270: 313–321.

3. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One. 2008; 3: e2802. https://doi.org/10.1371/journal.pone.0002802 PMID: 18665273
4. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009; 106: 12794–12797. https://doi.org/10.1073/pnas.0905845106 PMID: 19666622
5. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA. 2005; 102: 8369–8374. https://doi.org/10.1073/pnas.0503123102 PMID: 15928076
6. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS One. 2011; 6: e19254. https://doi.org/10.1371/journal.pone.0019254 PMID: 21637336
7. China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci USA. 2011; 108: 19641–19646. https://doi.org/10.1073/pnas.1104551108 PMID: 22100737
8. Chen S, Yao H, Han J, Liu C, Song J, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 2010; 5: e8613. https://doi.org/10.1371/journal.pone.0008613 PMID: 20062805
9. Kress J, Erickson DL. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS One. 2007; 2: e508. https://doi.org/10.1371/journal.pone.0000508 PMID: 17551588
10. Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour. 2008; 8: 480–490. https://doi.org/10.1111/j.1471-8286.2007.02002.x PMID: 21585825
11. Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, et al. Discriminating plants species in a local temperate flora using the *rbcL+matK* DNA barcode. Method Ecol Evol. 2011; 2: 333–340.
12. Li HQ, Chen JY, Wang S, Xiong SZ. Evaluation of six candidate DNA barcoding loci in Ficus (Moraceae) of China. Mol Ecol Resour. 2012; 12: 783–790. https://doi.org/10.1111/j.1755-0998.2012.03147.x PMID: 22537273
13. Chen J, Zhao J, Erickson DL, Xia N, Kress WJ. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. Mol Ecol Resour. 2015; 15: 337–348. https://doi.org/10.1111/1755-0998.12319 PMID: 25158042
14. Xiang XG, Hu H, Wang W, Jin XH. DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Mol Ecol Resour. 2011; 11: 1012–1021. https://doi.org/10.1111/j.1755-0998.2011.03044.x PMID: 21722327
15. Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, et al. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). Mol Ecol Resour. 2012; 12: 98–108. https://doi.org/10.1111/j.1755-0998.2011.03076.x PMID: 21967641
16. Tomlinson PB. The botany of mangroves. 2nd ed. Cambridge; Cambridge University Press; 1986.
17. Saddhe AA, Jamdade RA, Kumar K. Assessment of mangroves from Goa, west coast India using DNA barcode. SpringerPlus. 2016; 5: 1554. https://doi.org/10.1186/s40064-016-3191-4 PMID: 27652127
18. Shi S, Huang Y, Zeng K, Tan F, He H, et al. Molecular phylogenetic analysis of mangroves: independent evolutionary origins of vivipary and salt secretion. Mol Phylogenet Evol. 2006; 40: 298–304.
19. Spalding M, Kainuma M, Collins L. World atlas of mangroves. Earthscan eBook; 2010.
20. Polidoro BA, Carpenter KE, Collins L, Duke NC, Ellison AM, et al. The loss of species: mangrove extinction risk and geographic areas of global concern. PLoS One. 2010; 5: e10095. https://doi.org/10.1371/journal.pone.0010095 PMID: 20386710
21. Kathiresan K, Bingham BL. Biology of mangroves and mangrove ecosystems. Advances in Marine Biology. 2001; 40: 25–81.
22. Naskar K, Mandal R. Ecology and Biodiversity of Indian Mangroves. Daya Publishing House New Delhi; 1999.
23. Tamura K, Stecher G, Peterson D, Filipski A.Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evo. 2013; l30: 2725–2729.
24. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (www.barcodingoflife.org). Mol Ecol Notes. 2007; 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x PMID: 18784790
25. Rozas J. Polymorphism Analysis using DnaSP. In Posada D, editors. Bioinformatics for DNA Sequence Analysis: Methods in Molecular Biology Series, Humana Press NJ USA. 2009. pp. 337–350.

26. Meier R, Kwong S, Vaidya G, Ng PKL. DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. Syst Biol. 2006; 55: 715–728. https://doi.org/10.1080/10635150600969864 PMID: 17060194

27. Zou S, Fei C, Song J, Bao Y, He M, Wang C. Combining and comparing coalescent, distance and character-based approaches for barcoding microalgaes: A Test with Chlorella-like species (Chlorophyta). PLoS One. 2016; 11: e0153833. https://doi.org/10.1371/journal.pone.0153833 PMID: 27092945

28. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol ecol. 2012; 21: 1864–77. https://doi.org/10.1111/j.1365-294X.2011.05239.x PMID: 21883587

29. Yang Z, Landry J-F, Hebert PDN. A DNA Barcode Library for North American Pyraustinae (Lepidoptera: Pyraloidea: Crambidae). PLoS One. 2016; 11: e0161449. https://doi.org/10.1371/journal.pone.0161449 PMID: 27736878

30. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7: 214. https://doi.org/10.1186/1471-2148-7-214 PMID: 17996036

31. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4: e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

32. Gernhard T. The conditioned reconstructed process. J Theoret Biol. 2008; 253: 769–778.

33. Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, et al. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinfo. 2009; 10: 357.

34. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2012.

35. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinfo. 2004; 20: 289–290.

36. Ezard T, Fujisawa T, Barraclough TG. splits: SPecies' LImits by Threshold Statistics. 2009. R package version. 1.0-14/r31. (http://R-Forge.R-project.org/projects/splits/).

37. Pennisi E. Taxonomy. Wanted: A barcode for plants. Science. 2007; 318: 190–191. https://doi.org/10.1126/science.318.5848.190 PMID: 17932267

38. Little DP, Stevenson DW. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics. 2007; 23: 1–21.

39. Álvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. Mol phyl evol. 2003; 29: 417–34.

40. Ross HA, Murugan S, Li WLS. Testing the reliability of genetic methods of species identification via simulation. Syst Biol. 2008; 57: 216–230. https://doi.org/10.1080/10635150802032990 PMID: 18398767

41. Sahu SK, Singh R, Kathiresan K. Multi-gene phylogenetic analysis reveals the multiple origin and evolution of mangrove physiological traits through exaptation. Estuarine, Coastal and Shelf Science. 2016; 183: 41–51.

42. Hollingsworth ML, ANDRA CLARK AL, Forrest LL, Richardson J, Pennington R, et al. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Res. 2009; 9: 439–457.

43. Vinitha RM, Kumar SU, Aishwarya K, Sabu M, Thomas G. Prospects for discriminating Zingiberaceae species in India using DNA barcodes. J. Integr Plant Biol. 2014; 56: 760–773. https://doi.org/10.1111/jipb.12189 PMID: 24612741

44. Tang CQ, Humphreys AM, Fontaneto D, Barraclough TG. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. Methods in Ecol Evol. 2014; 5: 1086–1094.

45. Dumas P, Barbut J, Le Ru B, Silvain JF, Clamens AL, et al. Phylogenetic molecular species delimitations unravel potential new species in the pest genus Spodoptera Guenée, 1852 (Lepidoptera, Noctuidae). PLoS one. 2015; 10: e0122407. https://doi.org/10.1371/journal.pone.0122407 PMID: 25853412

46. Chen J, Zhao J, Erickson DL, Xia N, Kress WJ. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. Mol Ecol Resour. 2015; 15: 337–348. https://doi.org/10.1111/1755-0998.12319 PMID: 25158042

47. Li X, Duke NC, Yang Y, Huang L, Zhu Y, et al. Re-Evaluation of phylogenetic relationships among species of the mangrove genus *Avicennia* from Indo-West Pacific based on multilocus analyses. PLoS One. 2016; 11: e0164453. https://doi.org/10.1371/journal.pone.0164453 PMID: 27716800

48. Lahaye R, Savolainen V, Duthoit S, Maurin O, van der Bank M. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National park as a model system (South Africa). Nature Precedings. 2008; 1–21.

Selection of reference genes for quantitative real-time PCR analysis in halophytic plant *Rhizophora apiculata*

Ankush Ashok Saddhe, Manali Ramakant Malvankar and Kundan Kumar

Department of Biological Sciences, Birla Institute of Technology & Science Pilani, K K Birla Goa Campus, Zuarinagar, Goa, India

ABSTRACT

Rhizophora apiculata is a halophytic, small mangrove tree distributed along the coastal regions of the tropical and subtropical areas of the world. They are natural genetic reservoirs of salt adaptation genes and offer a unique system to explore adaptive mechanisms under salinity stress. However, there are no reliable studies available on selection and validation of reference genes for quantitative real-time polymerase chain reaction (qRT-PCR) in *R. apiculata* physiological tissues and in salt stress conditions. The selection of appropriate candidate reference gene for normalization of qRT-PCR data is a crucial step towards relative analysis of gene expression. In the current study, seven genes such as elongation factor 1α (*EF1α*), Ubiquitin (*UBQ*), β-tubulin (β-*TUB*), Actin (*ACT*), Ribulose1,5-bisphosphate carboxylase/oxygenase (*rbcL*), Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), and 18S rRNA (18S) were selected and analyzed for their expression stability. Physiological tissues such as leaf, root, stem, and flower along with salt stress leaf samples were used for selection of candidate reference genes. The high-quality expression data was obtained from biological replicates and further analyzed using five different programs such as geNorm, NormFinder, BestKeeper, Delta Ct and RefFinder. All algorithms comprehensively ranked *EF1α* followed by *ACT* as the most stable candidate reference genes in *R. apiculata* physiological tissues. Moreover, β-*TUB* and 18S were ranked as moderately stable candidate reference genes, while GAPDH and *rbcL* were least stable reference genes. Under salt stress, *EF1α* was comprehensively recommended top-ranked candidate reference gene followed by *ACT* and 18S. In order to validate the identified most stable candidate reference genes, *EF*1α, *ACT*, 18S and *UBQ* were used for relative gene expression level of sodium/proton antiporter (*NHX*) gene under salt stress. The expression level of *NHX* varied according to the internal control which showed the importance of selection of appropriate reference gene. Taken together, this is the first ever systematic attempt of selection and validation of reference gene for qRT-PCR in *R. apiculata* physiological tissues and in salt stress. This study would promote gene expression profiling of salt stress tolerance related genes in *R. apiculata*.

Subjects Molecular Biology, Plant Science

Keywords Reference gene, *Rhizophora apiculata*, Quantitative RT-PCR, Mangrove, Salt stress

Submitted 10 February 2018
Accepted 22 June 2018
Published 12 July 2018

Corresponding author
Kundan Kumar,
kundan@goa.bits-pilani.ac.in

Academic editor
Ann Cuypers

Additional Information and Declarations can be found on page 14

DOI **10.7717/peerj.5226**

INTRODUCTION

Mangroves are a unique intertidal ecosystem and evolutionarily adapted to the interface between land and water environments (*Saddhe, Jamdade & Kumar, 2017*). They are distributed along the tropical and subtropical part of the world and consist of 73 mangrove species with few recognized hybrids in 123 countries covering of 150,000 km^2 globally (*Spalding, Kainuma & Collins, 2010*). *Rhizophora apiculata* is a hardy woody fast growing mangrove tree. They are distributed throughout the Indian coastal region but the dominant population is on the southern coast of India (*Menon & Soniya, 2014*). They can tolerate salinity up to 65 parts per thousand (ppt) and show optimum growth at 8–15 ppt salinity (*Robertson & Alongi, 1992*). Mangrove plants are always exposed to the local hostile environments such as fluctuated water level, marshy land with anoxic conditions, hypersalinity and high UV light exposure (*Tomlinson, 1986*; *Hutchings & Saenger, 1987*). In order to survive in harsh conditions, they have developed some specialized traits such as viviparous propagules, aerial extensive supporting roots and high content of secondary metabolites. They are non-secretors and store surplus salt that enters through the transpiration stream into their leaves (*Menon & Soniya, 2014*). Mangroves are natural salt tolerant plant species but there are very few reports available on salt tolerance mechanism and salt stress associated genes. Several salt-induced genes were isolated and characterized from *R. apiculata* using suppression subtractive hybridization technique (*Menon & Soniya, 2014*). All salt-induced genes were highly upregulated at 12 h and further confirmed by qRT-PCR analysis using Actin (*ACT*) as a reference gene (*Menon & Soniya, 2014*). Recently *de novo* genome assembly of *R. apiculata* was reported (*Xu et al., 2017*), but the sequence is not accessible. Moreover, comparative transcriptome analysis was performed in mangroves species such as *Bruguiera gymnorrhiza, Kandelia obovata, R. apiculata*, and *Ceriops tagal* to understand adaptive evolution in the harsh intertidal habitats (*Xu et al., 2017*; *Guo et al., 2017*). However, there are no systematic reports available on selection and validation of reference gene for qRT-PCR in *R. apiculata* species.

Several techniques are available to investigate gene expression analysis including, semi-quantitative reverse transcription polymerase chain reaction, northern blotting, *in situ* hybridization, and quantitative real-time PCR (qRT-PCR). The qRT-PCR is a reliable, sensitive, and wide quantification range gene expression analysis technique (*Bustin, 2002*). Moreover, reference gene for qRT-PCR normalization is not universally standardized and it varies according to plant tissue material and experimental conditions (*Bustin et al., 2009*). For precise quantification and reproducible profiling, selection and validation of stable candidate reference genes are crucial steps prior to qRT-PCR for data normalization. Some commonly used reference genes include *Actin* (*ACT*), β-*tubulin (TUB), Ubiquitin* (*UBQ*), *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*), elongation factor 1α (*EF1α*) and 18S ribosomal RNA (18S) that are preferred to normalize the expression profiles of candidate genes. These reference genes are involved in basic cellular functions, maintaining cell size and shape, and cellular metabolism (*Bustin, 2002*). However, several reports have shown that the level of reference genes expression varies in different cultivars, tissues, and stress conditions (*Sinha et al., 2015*; *Reddy et al., 2015*;

Nikalje et al., 2018). Hence, it is very important to select and validate most appropriate reference genes involved in various experimental conditions before proceeding to gene expression analysis. Various web-based tools and algorithms are available to address validation of candidate reference genes including, comparative ΔCt (cycle thresholds) (*Silver et al., 2006*), NormFinder (*Andersen, Jensen & Orntoft, 2004*), BestKeeper (*Pfaffl et al., 2004*), and geNorm algorithm (*Vandesompele et al., 2002*). RefFinder, a web-based program, which provides a comprehensive ranking of reference genes (*Xie et al., 2012*).

Based on the literature survey, there were no reports available on evaluation of candidate reference genes for qRT-PCR in *R. apiculata*. In the present study, we aim to evaluate the most stable candidate reference gene for qRT-PCR gene expression analysis in *R. apiculata* physiological tissues and in salt-stressed leaf samples. The current study will promote the gene expression analysis in the *R. apiculata,* especially when studied under salinity stress.

MATERIALS & METHODS

Plant materials

In the present study, we collected three month old *R. apiculata* seedlings located in the west coast of India with the geographical latitude of 15.5256°N and longitude of 73.8753°E, with the permission from the Principal Chief Conservator of Forest, Goa Forest Department, Goa, India. Mangrove species identification was performed based on morphological characteristics using a comparative guide to the mangroves of Goa (*Naskar & Mandal, 1999*). All seedlings were acclimatized and maintained in half-strength Hoagland solution at a temperature regime of 24–30 °C, 40–50% relative humidity. Various physiological tissues such as leaves, stems, roots and flower samples were collected. To imitate salt stress conditions, young seedlings of *R. apiculata* were exposed to Hoagland nutrient solution supplemented with 250 mM sodium chloride (NaCl) continuously and leaf samples were harvested at different time-course such as 0, 6, 12 and 24 h.

RNA isolation and cDNA synthesis

Total RNA was extracted using modified cetyl-trimethyl ammonium bromide (CTAB) protocol with 2% polyvinylpyrrolidone and 10% β-mercaptoethanol (*Fu et al., 2004*). Freshly collected tissues were immediately pulverized into 2 ml of pre-warmed CTAB buffer and incubated at 60 °C for 30 min. The suspension was gently mixed and centrifuged at 14,000 rpm for 10 min at room temperature with an equal volume of chloroform: isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and RNA was precipitated with a 1/3rd volume of 8M lithium chloride (LiCl) and incubated at −20 °C for 1 h followed by adding an equal volume of chilled isopropanol (−20 °C). The RNA was precipitated by centrifugation at 14,000 rpm for 10 min at room temperature followed by washing with 70% ethanol. RNA was finally dissolved in 0.1% DEPC treated water and its quantity and quality were confirmed by Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA contamination was removed by DNase I enzyme (Thermo Fisher Scientific, Waltham, MA, USA) treatment at 37 °C for 30 min and heat inactivated at 65 °C for 10 min with 50 mM EDTA. The cDNA synthesis was performed in 20 μl reaction volume

Table 1 Details of candidate reference genes, Accession number, primer sequences, amplicon size, PCR efficiency (%) and regression coefficient (R^2) for each candidate reference gene selected in this study.

Sr. no.	Gene label	Accession No.	Gene description	Primer sequence (5′-3′)	Amplicon size (bp)	PCR efficiency (%)	R^2
1	18S	MH277331	18S rRNA	F-CCGTCCTAGTCTCAACCATAAAC R-GCTCTCAGTCTGTCAATCCTTG	189	102.30%	1
2	*ACT*	MH279969	Actin	F-ATCACACCTTCTACAACGAGC R-CAGAGTCCAACACGATACCAG	207	92.03%	0.994
3	*EF1α*	MH310424	Elongation Factor 1 α	F-AGCGTGTGATTGAGAGGTTC R-AGATACCAGCCTCAAAACCAC	53	98.60%	0.99
4	*UBQ*	MH310425	Ubiquitin	F-CACTTCGACCGCCACTAC R-AGGGCATCACAATCTTCACAG	60	90.54%	0.992
5	*RbcL*	KP697362	Ribulose 1,5-Bisphosphate Oxygenase/Carboxylase Large	F-ATGTCACCACAAACAGAGACTAAAGC R-GTAAAATCAAGTCCACCRCG	530	97.69%	0.996
6	β-*TUB*	MH310423	β-tubulin	F-ACCTCCATCCAGGAGATGTT R-GTGAACTCCATCTCGTCCATTC	60	94.08%	0.996
7	*GAPDH*	MH279970	Glyceraldehyde3-phosphate dehydrogenase	F-ACCACAGTCCATGCCATCAC R-TCCACCACCCTGTTGCTGTA	264	96.78%	0.99
8	*NHX*	KU525079	Sodium/proton antiporter	F-TGCTAGCTCTTGTCCTGATTG R-ATTGACACAGCACCTCTCATTA	120	103.70%	0.997

using the RevertAid Reverse Transcriptase (Thermo Fisher Scientific, Watham, MA, USA), 0.1–5 μg RNA sample and oligo $d(T)_{18}$ primer, as per manufacturer's instructions.

Selection of reference genes and primer designing

Nine housekeeping genes such as *ACT*, α-*TUB*, β-*TUB*, *GAPDH*, *UBQ*, 18S rRNA, *rbcL*, Histone H3 and *EF1α* used in qRT-PCR along with one target gene sodium/proton antiporter (*NHX*) were selected. There is no genome sequence available publicly for *R. apiculata* hence, homologous candidate reference gene sequences were retrieved from model plants such as *Arabidopsis thaliana* and *Oryza sativa* from Gramene and NCBI databases. Full-length candidate reference gene sequences were used for primer designing using PrimerQuest (Integrated DNA Technologies) with given parameters: melting temperature (T_m) of 55–65 °C, primer length of 17–25 bp, and amplicon length of 100–500 bp (Table 1). The amplicon was sequenced and annotated based on the sequence similarity-based search tool. Further, all the confirmed sequences were submitted to GenBank for accession numbers. After primer specificity analysis α-*Tub* and Histone H3 were removed from further analysis. The primer sequences, accession numbers, and their efficiency were given in Table 1.

For qRT-PCR, primer specificity was determined using melting curve analysis and the PCR products were checked on 2% agarose gel. The primer efficiency of all candidate reference genes was calculated based on the standard curve generated from a 10-fold serial dilution of cDNA (10^0, 10^{-1}, 10^{-2}, and 10^{-3}) and regression coefficient (R^2) values. Primer efficiency was calculated using the given formula $[E = (10^{(-1/\text{slope})} - 1) \times 100]$,

where $E = 2$ and corresponds to 100% efficiency; high/acceptable amplification efficiency equals 90–110% (*Sinha et al., 2015*).

Quantitative RT-PCR analysis

The quantitative RT-PCR analysis was carried out using SYBR green master mix (2X Brilliant III SYBR® Green QPCR; Agilent Technologies, Santa Clara, CA, USA), on AriaMx Agilent system (AriaMx; Agilent Technologies, Santa Clara, CA, USA) with the following reaction conditions: initial denaturation at 95 °C for 3 min, 40 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 45 s extension, and a melt-curve program (65–95 °C with a temperature increase of 0.5 °C after every 5 s). The melting curve was generated to determine the amplicon specificity. The qRT-PCR experiments were performed using three biological and two technical replicates. A reaction with no template control and a reverse transcription negative control were performed to check the potential reagents and genomic DNA contamination.

Analysis of gene expression stability

The candidate reference gene ranking was analyzed using five different algorithms such as geNorm, NormFinder, Bestkeeper, Δct, and comprehensive ranking analysis by RefFinder.

geNorm analysis

The geNorm determines the most stable reference genes based on the gene expression stability value (M) for a reference gene. It also calculates the minimum number of candidate reference genes required for normalization of target genes. It requires calculated Cq values into relative quantities using the given formula: $Q = E^{(\text{minCq}-\text{Cq})}$, where Q represents sample quantity relative to sample with the highest expression, E is amplification efficiency and min Cq is the lowest Cq values. The stability value (M) is defined as an average pairwise variation (V) of the gene compared with all other tested reference genes and the cut-off is 1.5 (*Vandesompele et al., 2002*). If M value is lower than 1.5, it represents stable candidate reference gene and higher values reflect least stable.

NormFinder

NormFinder calculates expression stability values for candidate reference genes and evaluates the most stable reference gene pairs. It also calculates intra and intergroup variation using a direct comparison between genes. It uses same input calculation files which are required for geNorm with a little variation such as the first row represents a sample, the first column represents genes and the last row represents a group of samples. NormFinder is available with Excel spreadsheet add-in (https://moma.dk/normfinder-software). It ranks candidate reference genes based on expression stability value. Lowest M value represents most stable reference gene and higher the value, least stable are the genes (*Andersen, Jensen & Orntoft, 2004*).

BestKeeper analysis

BestKeeper determines the best reference gene based on the normalization factor (also called Bestkeeper index) and pairwise correlation analysis. It requires raw Cq values as an input data to select most stable and least stable candidate reference genes. It is available

in an MS Excel spreadsheet file (http://www.gene-quantification.de/bestkeeper.html) and in RefFinder (http://150.216.56.64/referencegene.php?type=reference) as well. It evaluates the candidate reference gene stability by comparing the standard deviation of each gene and averages of these values. It also calculates the coefficient of variance, Pearson correlation coefficient (r) values, geometric mean (GM) and arithmetic mean (AM).

ΔCt method

This tool is available in an MS Excel spreadsheet as well as in RefFinder (http://150.216.56.64/referencegene.php?type=reference). It calculates the stable candidate reference gene based on standard deviation and pairwise comparison with other genes. ΔCt requires raw Cq values as an input data. It considers a pair of gene for calculations and compares ΔCt values among genes (*Silver et al., 2006*).

RefFinder analysis

RefFinder is a web-based comprehensive tool developed for evaluating and screening reference genes from extensive experimental datasets. RefFinder was used to generate comprehensive stability rankings (*Xie et al., 2012*). Comprehensive ranking of seven candidate reference genes was analyzed using RefFinder.

Validation of candidate reference genes

The reliability of highly stable candidate reference genes identified in the current study was validated using sodium/proton antiporter (NHX) as a salt stress target gene. The differential gene expression profiles of NHX under salt stress at 0, 6, 12 and 24 h were normalized using *EF1α*, *ACT*, 18S and *UBQ* along with the combination of *EF1α* + *ACT* genes. The input values for *EF1α* + *ACT* were calculated using the geometric mean formula given below to normalize gene of interest Geometric $Mean = \sqrt[n]{\times 1, \times 2, \times 3 \ldots \times n}$, where the $n =$ number of times (*Vandesompele et al., 2002*). The average Cq values from three biological replicates were used for relative expression analysis and the relative gene expression level calculated using the $2^{-\Delta\Delta^{CT}}$ method (*Livak & Schmittgen, 2001*; *Manuka, Saddhe & Kumar, 2018*). Statistical analysis was performed using SPSS 15.0 for Windows evaluation version to verify the significant difference between relative gene expressions. One-way Analysis of variance (ANOVA) with Tukey's Honest Significant Difference (HSD) test was performed for comparison between reference genes and target genes. A p-value <0.05 was considered statistically significant.

Minimum Information for publication of qRT-PCR experiments guidelines (MIQE)

All the qRT-PCR experiments and data analysis in the present study were performed in accordance with the MIQE guidelines (*Bustin, 2002*).

RESULTS

Expression profiling of selected reference genes

In order to select stable reference genes, transcript levels in tissues such as leaf, root, stem, and flower as well as salt stress samples were quantified based on their cDNA

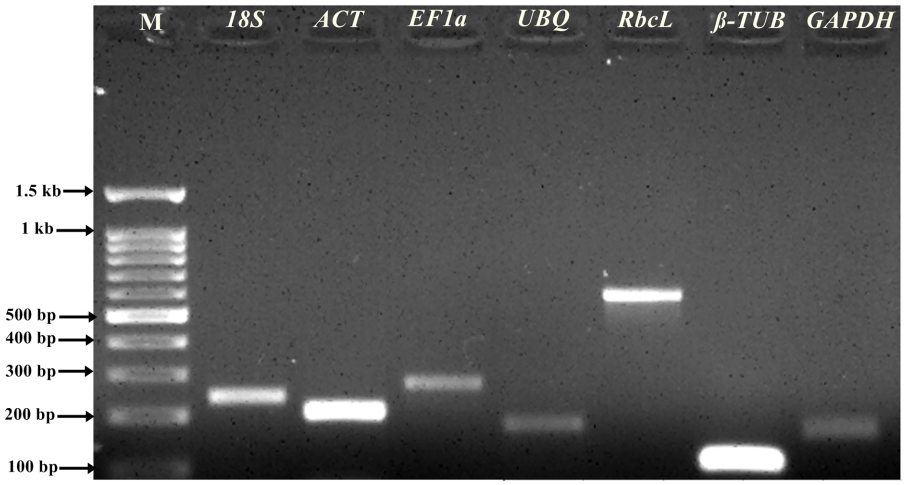


Figure 1 Amplification product of genes. PCR products on 2% agarose gel stained with ethidium bromide. Amplification products of seven candidate reference genes selected for gene validation of *R. apiculata* samples. M: 100 bp DNA ladder. Lanes 1, 2, 3, 4, 5, 6 and 7 were the gene products of 18S, *ACT*, *EF1α*, *UBQ*, *RbcL*, β-*TUB,* and *GAPDH*, respectively.

Full-size DOI: 10.7717/peerj.5226/fig-1

concentration. The primer specificity was determined by PCR products wherein single, expected amplicon size was obtained (Fig. 1). The qRT-PCR melting curve for template test and negative control (NTC) without template were analyzed for primer-dimer and reagents contamination (Fig. S1). Further, NTC samples were confirmed by running 2% agarose gel electrophoresis. The amplified PCR products were sequenced and submitted to GenBank for accession numbers. All the sequenced PCR products were identified and annotated based on BLAST search. The primer efficiency (%) ranged from 103.70%, ($R^2 = 0.997$) for *NHX* to 90.54% ($R^2 = 0.98$) for *UBQ* including 18S (102.30, $R^2 = 1$), *ACT* (92.03%, $R^2 = 0.994$), *EF1α* (98.60%, $R^2 = 0.99$), β-*TUB* (94.08%, $R^2 = 0.996$), *GAPDH* (96.78%, $R^2 = 0.992$) and *RbcL* (97.69%, $R^2 = 0.996$) (Fig. S2; Table 1).The mean cycle threshold (Cq) values of the seven selected reference genes for different tissue samples ranged from 14.16 for18S to 21.77 for *GAPDH* (Fig. 2A). Similarly, for the salt stress samples, the mean Cq values ranged from 13.96 for 18*S* to 24.23 for *UBQ* (Fig. 2B). Mean Cq values gave insight into approximate gene expression data. Negative control showed higher Cq values indicating no product amplification which was further checked on a 2% agarose gel. Moreover, negative control without reverse transcriptase did not show any product amplification, thus indicating no gDNA contamination.

geNorm analysis

For physiological tissues, seven candidate reference genes showed average expression stability value (M) less than 1.5. *ACT* ($M = 0.721$) was most stable reference gene followed by *EF1α* ($M = 0.761$), and β-*TUB* ($M = 0.763$) (Table 2A; Fig. S3A). *GAPDH* was the least stable candidate reference gene with M value 1.599. The geNorm also determines an optimum number of candidate reference gene for normalization based on the calculation of pairwise variation (V_n/V_{n+1}) between sequential normalization factor (NF_n and NF_{n+1}).

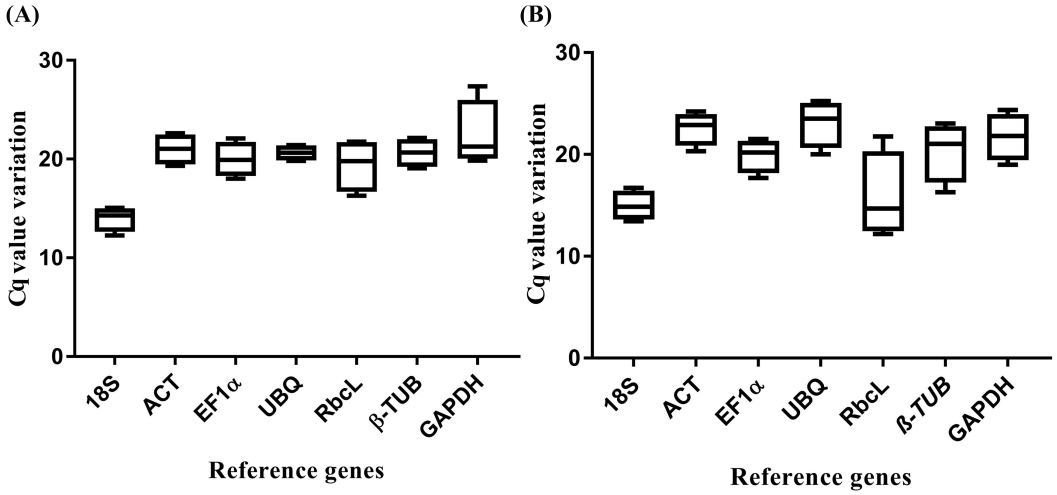


Figure 2 Threshold cycle (Ctq) values of seven candidate reference genes. (A) Tissue-specific box plot for the Cq values of seven candidate reference genes from the qRT-PCR analysis. For each reference gene, the line inside the box is the median. The top and bottom lines of the box are the first and third quartiles, respectively. The top and bottom whiskers represent the 5th and 95th percentiles. (B) Salt stress box plot for the Cq values of seven candidate reference genes from the qRT-PCR analysis. For each reference gene, the line inside the box is the median. The top and bottom lines of the box are the first and third quartiles, respectively. The top and bottom whiskers are the 5th and 95th percentiles, respectively.

Full-size DOI: 10.7717/peerj.5226/fig-2

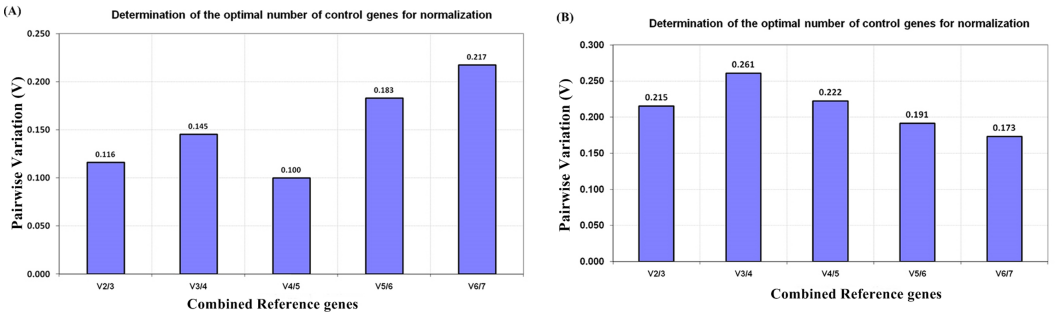


Figure 3 geNorm pairwise variation (V) analysis to determine minimum number of candidate reference genes required for normalization in qRT-PCR of *R. apiculata* (A) Pairwise variation analysis for physiological tissue samples (B) Pairwise variation analyzed for salt stress leaf samples. V1 to V7 stand for the variation in candidate reference genes ranked based on their stability, where V1 is the variation for the most stable and V7 is the variation for the least stable gene.

Full-size DOI: 10.7717/peerj.5226/fig-3

To select the best pair for normalization, the threshold value is 0.15. If pairwise variation value is lower than 0.15, there is no need to add more candidate reference gene. Moreover, the best pairwise variation value 0.382 was observed for a combination of *ACT* and *EF1α* and comprehensively recommended for normalization (Table 2B; Fig. S3A). Based on the observation, there were no effects on an addition of the third gene in the combination of *ACT* and *EF1α* which showed pairwise variation value below 0.15 (Fig. 3A).

Table 2 geNorm analysis and ranking of candidate reference genes based on stability value (*M*). Lower *M* value represents most stable reference genes and higher *M* value showed least stable reference genes.

A. geNorm analysis for individual candidate reference genes

Sr. No	Reference genes	Physiological tissue samples		Salt stress samples	
		Stability value ($M < 1.5$)	Ranking	Stability value ($M < 1.5$)	Ranking
1	*18S*	0.880	4	1.020	2
2	*ACT*	0.721	1	0.927	1
3	*EF1α*	0.761	2	1.021	3
4	*UBQ*	0.928	5	1.399	7
5	*RbcL*	1.225	6	1.357	6
6	*B-TUB*	0.763	3	1.351	5
7	*GAPDH*	1.599	7	1.257	4

B. Best pair of reference genes based on geNorm analysis

1	*ACT*+*EF1α*	0.382	Most stable pair of reference genes in physiological tissue samples
2	*ACT*+*EF1α*	0.462	Most stable pair of reference genes under salt stress samples

Under salinity stress, *ACT* was most stable candidate reference gene with *M* value 0.927, followed by 18S and *EF1α* showing same stability *M* value 1.02 (Table 2A). Moreover, *rbcL* and *UBQ* performed least stable candidate reference gene with *M* value 1.357 and 1.399 respectively. In salt stress, *ACT* + *EF1α* were the most suitable combination for normalization of the gene of interest with pairwise variation value of 0.462 (Table 2B; Fig. S3B). According to pairwise variation analysis, if the third gene was added in the *ACT* + *EF1α*, it showed higher pairwise variation value of 0.215 (Fig. 3B).

NormFinder

In *R. apiculata* physiological tissue samples *EF1α* was most stable with stability value of 0.085. β-*TUB* (0.135) was the second most stable candidate reference gene followed by *ACT* (0.164) (Table 3A). *EF1α* and β-*TUB* (0.070) showed the most stable combination for the pair of candidate genes for normalization (Table 3B). Overall, *GAPDH, UBQ,* and *RbcL* were least stable reference genes. In salt stress, *ACT* was most stable reference gene with a stability value of 0.196. *EF1α* and 18S were second and third most stable candidate reference genes with stability value 0.257 and 0.273 respectively. *ACT* and *EF1α* showed the best pair of reference genes with stability value 0.183 (Table 3B). Under salt stress, geNorm and Normfinder showed almost similar results for a selection of candidate reference gene.

BestKeeper

In the BestKeeper analysis, standard deviation (SD) and coefficient of correlation (*r*) value were the criteria used for comparison. Highest *r* value represents the most stable candidate reference genes and lower *r* value represents the least stable genes. Here, we considered *r* value for evaluation, showing *EF1α* as the most stable reference gene followed by *ACT* with *r* value 0.987 and 0.966 respectively. *GAPDH* was ranked as the least stable candidate reference gene with lower *r* value (Table 4). The result is consistent with geNorm and

Table 3 NormFinder analysis and ranking of candidate reference genes based on stability value. Lower stability value represents most stable reference genes and higher value showed least stable reference genes. Ra-*Rhizophora apiculata*.

A. NormFinder analysis for individual candidate reference genes

Sr. No	Reference genes	Physiological Tissue samples		Salt stress samples	
		Stability value	Ranking	Stability value	Ranking
1	18S	0.410	4	0.273	3
2	*ACT*	0.164	3	0.196	1
3	*EF1α*	0.085	1	0.257	2
4	*UBQ*	0.463	5	0.518	6
5	*RbcL*	0.500	6	0.499	5
6	*B-TUB*	0.135	2	0.533	7
7	*GAPDH*	0.568	7	0.483	4

B. Best pair of candidate reference genes based on geNorm analysis

Sr. No.	Best pair of genes	Stability value	
1	*EF1 α + B-TUB*	0.070	Most stable pair of candidate reference genes in Ra tissue samples
2	*ACT+EF1α*	0.183	Most stable pair of candidate reference genes in salt stress samples

Table 4 Candidates reference gene stability and ranking analyzed by BestKeeper (Coefficient of correlation, r), Ct (Mean, STDEV) ranking of genes. Coeff. of corr, Coefficient of correlation; RG, reference gene.

Sr. No	RG	BestKeeper				ΔCt analysis			
		Physiological Tissue samples		Salt stress samples		Physiological Tissue samples		Salt stress samples	
		Coeff. of corr. (r)	Rank	Coeff. of corr. (r)	Rank	Mean SD	Rank	Mean SD	Rank
1	*18S*	0.935	4	0.507	4	0.88	3	1.02	2
2	*ACT*	0.966	2	0.638	1	0.76	2	0.93	1
3	*EF1α*	0.987	1	0.523	3	0.72	1	1.02	2
4	*UBQ*	0.964	3	0.001	7	0.93	4	1.40	6
5	*RbcL*	0.958	5	0.310	6	1.22	5	1.36	5
6	*B-TUB*	0.964	3	0.625	2	0.76	2	1.35	4
7	*GAPDH*	0.850	6	0.435	5	1.60	6	1.26	3

NormFinder analysis. In salt stress, *ACT* ($r = 0.638$) showed most stability followed by β-*TUB* ($r = 0.625$) and *EF1α* ($r = 0.523$) (Table 4). Under salt stress, similar results were observed with little variation in BestKeeper. BestKeeper determined *β-TUB* second most stable candidate reference gene in salt stress.

ΔCt analysis

According to ΔCt analysis, *EF1α* was the most stable candidate reference gene followed by *ACT* and β-*TUB* in physiological tissue (Table 4). 18S was ranked as an average or moderately stable reference gene. The results were consistent with earlier analysis. *GAPDH, RbcL,* and *UBQ* were the least stable. Under salt stress, *ACT* was most stable candidate reference gene followed by *EF1α* and18S (Table 4).

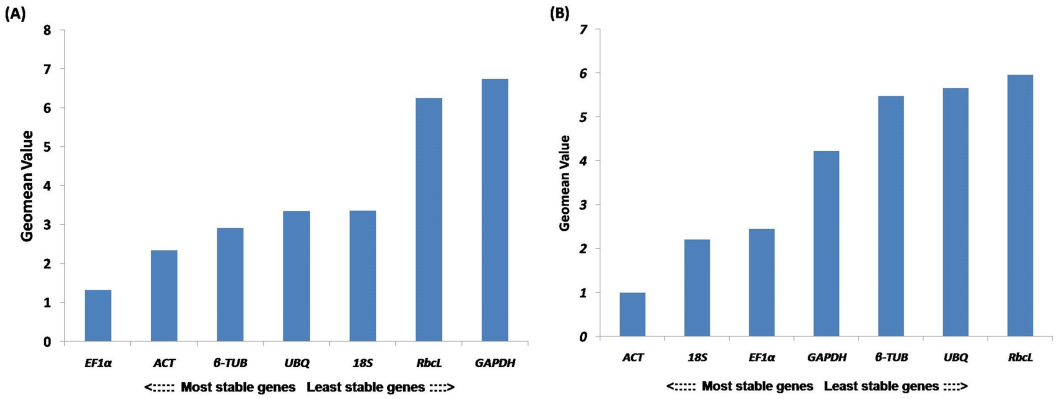


Figure 4 Ranking of reference genes. Comprehensive ranking of candidate reference genes in *R. apiculata* based on the rankings from each algorithms using RefFinder (A) Overall ranking of candidate reference gene in physiological tissues (B) Overall ranking of candidate reference gene in salt stress leaf samples.

Full-size DOI: 10.7717/peerj.5226/fig-4

Comprehensive ranking of candidate reference genes

Based on the Geomean value, a comprehensive ranking of all candidate reference genes showed *EF1α* (1.32) was the most stable followed by *ACT* (2.34) and β-*TUB* (2.91) (Fig. 4A). Moreover, *rbcL* and *GAPDH* performed as a least stable candidate reference genes. In salt stress, *ACT* was the most stable with Geomean value 1. Moreover, 18S was second most stable candidate reference gene with Geomean value 2.29 (Fig. 4B). Here, *UBQ* and *rbcL* performed as least stable candidate reference genes.

Validation of stable candidate reference genes under salt stress

To validate the efficacy of candidate reference genes, *ACT*, *EF1α*, 18S and *UBQ* were used to normalize the expression levels of NHX in salt stress at four different time course. Set of the most stable candidate reference genes such as *ACT*, *EF1α*, 18S and the least stable candidate reference gene *UBQ* were used as internal controls. While using EF1α, *ACT*, and 18S alone for normalization, *NHX* showed significant upregulation expression pattern in salt stress at 12 h. However, with *UBQ* as an internal control, NHX expression was upregulated in salt stress after 6 h of salt stress (Fig. 5).

DISCUSSION

The present study is the first systematic assessment of candidate reference gene in *R. apiculata* physiological tissues as well as under salt stress. The MIQE guidelines gives a framework for good experimental practice and transparent results (*Bustin et al., 2009*). The results were in accordance with the MIQE guidelines, where the ideal PCR efficiency is 100%, while the acceptable range is from 90 to 110% (*Bustin et al., 2009*). In the present study, we designed the primers based on homologous genes of *Arabidopsis thaliana* because genome sequence of *R. apiculata* is not available. To check the designed primers specificity, we tested PCR and confirmed on 2% agarose gel for single desired size bands. Further, amplified PCR products were confirmed through sequencing and identified by BLASTN

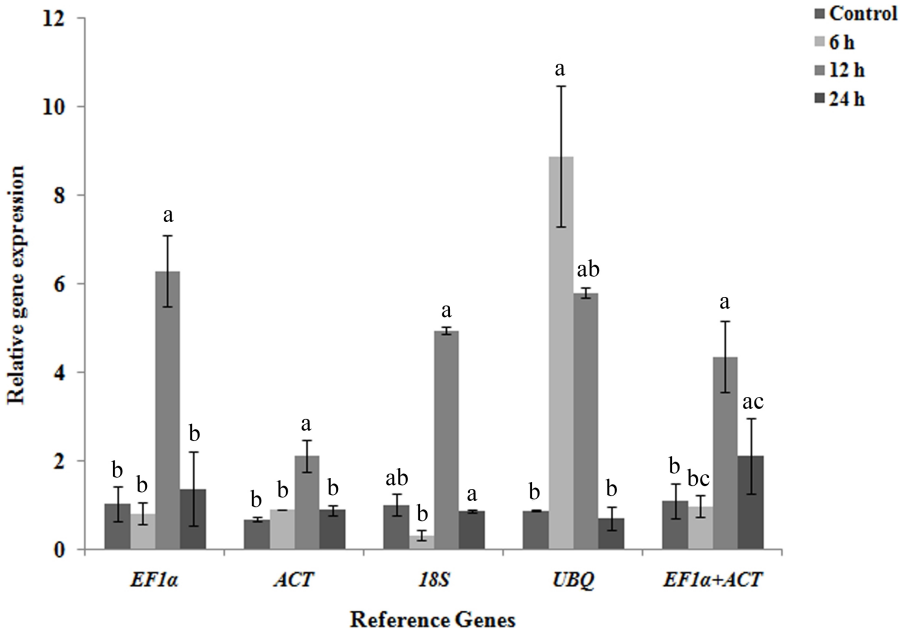


Figure 5 Validation and normalization of target NHX gene of *R. apiculata* under salt stress at four different time-course such as 0, 6, 12 and 24 h using EF1 α, ACT, 18S and *UBQ* reference genes. Normalization of NHX using *EF1 α*, *ACT*, 18S, *UBQ* and combined *EF1α*+ *ACT*. Error bars represent the mean ± standard error of relative abundance of three biological replicates. The bars having different superscript letter are significantly different at $p < 0.05$.

Full-size DOI: 10.7717/peerj.5226/fig-5

tool. All the sequences were submitted to GenBank for accession numbers. In the present work, the primer efficiency ranged from 90–103% and most of the study reported primer efficiency ranging from 90–110%. The primer efficiency was recorded between 92 to 98.6% in *Sesuvium portulacastrum*, 92.89–98.76% in *Suaeda aralocaspica*, 81–100.88% in *Halostachys caspica*, 90.5–104.43% in *Cajanus cajan* (*Nikalje et al., 2018*; *Cao, Wang & Lan, 2016*; *Zhang et al., 2016*; *Sinha et al., 2015*).

Selection of unstable reference gene can lead to fallacious relative gene expression result and errors in normalization (*Dheda et al., 2005*). Besides the selection of suitable genes, it is equally important to select more than one candidate reference gene which improves the gene expression analysis (*Vandesompele et al., 2002*). The geNorm algorithm evaluates single as well as best pair stable candidate reference genes for normalization. In the current study, a comprehensive ranking of candidate reference genes was evaluated; *EF1α* being the most stable candidate reference gene in physiological tissues and *ACT* in salt stress. The geNorm algorithm gave a consistent result with a comprehensive ranking which showed the most stable candidate reference gene as *EF1α* in physiological tissues and *ACT* in salt stress tissue samples. A similar observation was reported in the *Halostachys caspica* halophyte species, which showed that *EF1α* and *TUB3* was the most stable under salt and drought stress (*Zhang et al., 2016*). Under salt stress, most stable reference genes in *S. portulacastrum* shoot tissue were α-*TUB*, *EIF4a* and *EF1α*, while *UCE 2*, *TBP* and *EF1α* in the root tissue (*Nikalje et al., 2018*). This result reflects that a reference gene is not universal and altered

according to plant species and stress conditions. So it is always recommended to select and validate the commonly used candidate reference genes. One of the possible reasons might include the differential expression patterns under unstressed and stressed conditions and a difference in response to the particular stress.

We observed a little variation in assessed best pair candidate reference genes between geNorm (*EF1α+ACT)* and NormFinder (*EF1α+β-TUB*) analysis. The possible explanation is subtle differences between their algorithm methods. Similar results were observed in earlier studies during evaluation of candidate reference genes, wherein a little variation in geNorm and NormFinder was reported, which leads to minute variation in candidate reference gene ranking, as reflected in the current study (*Cruz et al., 2009*; *Pellino et al., 2011*).

The geNorm calculates candidate reference genes to normalize target gene based on their average stability value (*M*) and also determines the optimum number of candidate reference genes required for normalization. Although NormFinder calculates stability values for each gene and BestKeeper ranks the genes according to *r* values, these algorithms do not determine the minimum number of reference genes required for normalization (*Kozera & Rapacz, 2013*). We have performed target gene validation using geNorm analyzed data because it ranks candidate reference genes based on their stability and also evaluates the minimum number of reference genes required for normalization. We used individual candidate genes as well as a combination of *EF1α+ACT*. We found that *EF1α, ACT,* and18S had given significant upregulation of *NHX* gene, while using least stable candidate reference gene *UBQ* showed different expression pattern after normalization. We observed that relative gene expression of NHX showed significant transcript accumulation pattern at 12 h. It was earlier reported that most significant expression patterns were observed in *R. apiculata* after 12 h time-course (*Menon & Soniya, 2014*). The geNorm data suggests the use of two reference genes for normalization of gene of interest. Moreover, most of the previous study underscored the use of more than one reference gene to improve the relative gene expression (*Bustin, 2002*).

In summary, we have successfully evaluated and validated stable reference genes in *R. apiculata* physiological tissues and under salt stress. This analysis revealed that the suitable reference genes differ between physiological tissues and in salt stress tissues. We found that commonly used reference genes such as *EF1α* and *ACT* are most useful reference in an individual as well as in combined form.

CONCLUSION

The current study examined the most stable candidate reference gene for the normalization of relative gene expression in *R apiculata* physiological tissue and under salt stress. We strongly recommend *EF1α* followed by *ACT* and *β-TUB* as the best stable candidate reference genes for normalization in *R. apiculata* physiological tissue gene expression analysis. Under salt stress, *EF1α* followed by *ACT* and 18S are the most suitable candidate reference genes for normalization. In conclusion, *EF1α* and *ACT* can be used as candidate reference genes for the study of *R. apiculata*.

ACKNOWLEDGEMENTS

The authors thank BITS Pilani, K. K. Birla Goa Campus, for providing the necessary support and the anonymous reviewers for their constructive suggestions.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Council of Scientific and Industrial Research, India (No. 38(1416)/16/EMR-II). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Council of Scientific and Industrial Research, India: 38(1416)/16/EMR-II.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Ankush Ashok Saddhe and Manali Ramakant Malvankar performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Kundan Kumar conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are provided as Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5226#supplemental-information.

REFERENCES

Andersen CL, Jensen JL, Orntoft TF. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**:5245–5250 DOI 10.1158/0008-5472.CAN-04-0496.

Bustin SA. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**:23–39 DOI 10.1677/jme.0.0290023.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines:

minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**:611–622 DOI 10.1373/clinchem.2008.112797.

Cao J, Wang L, Lan H. 2016. Validation of reference genes for quantitative RT-PCR normalization in *Suaeda aralocaspica*, an annual halophyte with heteromorphism and C4 pathway without Kranzanatomy. *PeerJ* **4**:e1697 DOI 10.7717/peerj.1697.

Cruz F, Kalaoun S, Nobile P, Colombo C, Almeida J, Barros LMG, Alves-Ferreira M. 2009. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Molecular Breeding* **23**:607–616 DOI 10.1007/s11032-009-9259-x.

Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, Zumla A. 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Analytical Biochemistry* **344**:141–143 DOI 10.1016/j.ab.2005.05.022.

Fu X, Deng S, Su G, Zeng Q, Shi S. 2004. Isolating high-quality RNA from mangroves without liquid nitrogen. *Plant Molecular Biology Reporter* **22**:197 DOI 10.1007/BF02772728.

Guo W, Wu H, Zhang Z, Yang C, Hu L, Shi X, Jian S, Shi S, Huang Y. 2017. Comparative analysis of transcriptomes in rhizophoraceae provides insights into the origin and adaptive evolution of mangrove plants in intertidal environments. *Frontiers in Plant Science* **8**:795 DOI 10.3389/fpls.2017.00795.

Hutchings PA, Saenger P. 1987. *Ecology of mangroves*. St Lucia: Australian Ecology Series University of Queensland Press, 14–44.

Kozera B, Rapacz M. 2013. Reference genes in real-time PCR. *Journal of Applied Genetics* **54**:391–406 DOI 10.1007/s13353-013-0173-x.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\Delta\Delta Ct)}$ method. *Methods* **25**:402–408 DOI 10.1006/meth.2001.1262.

Manuka R, Saddhe AA, Kumar K. 2018. Expression of *OsWNK9* in Arabidopsis conferred tolerance to salt and drought stress. *Plant Science* **270**:58–71 DOI 10.1016/j.plantsci.2018.02.008.

Menon TG, Soniya EV. 2014. Isolation and characterization of salt-induced genes from *Rhizophora apiculata* Blume, a true mangrove by suppression subtractive hybridization. *Current Science* **107**:650–655.

Naskar K, Mandal R. 1999. *Ecology and biodiversity of Indian mangroves*. India: Daya Publishing House New Delhi.

Nikalje GC, Srivastava AK, Sablok G, Pandey GK, Nikam TD, Suprasanna P. 2018. Identification and validation of reference genes for quantitative real-time PCR under salt stress in a halophyte, *Sesuvium portulacastrum*. *Plant Gene* **13**:18–24 DOI 10.1016/j.plgene.2017.11.003.

Pellino M, Sharbel TF, Mau M, Amiteye S, Corral JM. 2011. Selection of reference genes for quantitative real-time PCR expression studies of microdissected reproductive tissues in apomictic and sexual boechera. *BMC Research Notes* **4**:303 DOI 10.1186/1756-0500-4-303.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**:509–515 DOI 10.1023/B:BILE.0000019559.84305.47.

Reddy PS, Reddy DS, Bhatnagar-Mathur P, Sharma KK, Vadez V. 2015. Cloning and validation of reference genes for normalization of gene expression studies in pearl millet (*Pennisetum glaucum* L.) by quantitative real-time PCR. *Plant Gene* **1**:35–42 DOI 10.1016/j.plgene.2015.02.001.

Robertson AI, Alongi DM. 1992. *Tropical Mangrove ecosystems*. Wiley Online Library DOI 10.1029/CE041.

Saddhe AA, Jamdade RA, Kumar K. 2017. Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. *PLOS ONE* **12(8)**:e0183245 DOI 10.1371/journal.pone.0183245.

Silver N, Best S, Jiang J, Thein SL. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* **7**:33 DOI 10.1186/1471-2199-7-33.

Sinha P, Singh VK, Suryanarayana V, Krishnamurthy L, Saxena RK, Varshney RK. 2015. Evaluation and validation of housekeeping genes as reference for gene expression studies in pigeon pea (*Cajanus cajan*) under drought stress conditions. *PLOS ONE* **10**:e0122847 DOI 10.1371/journal.pone.0122847.

Spalding M, Kainuma M, Collins L. 2010. *World atlas of mangroves*. London: Earthscan eBook, 1–11.

Tomlinson PB. 1986. *The botany of mangroves*. Second edition. Cambridge: Cambridge University Press.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**:research0034.1–research0034.11.

Xie FL, Xiao P, Chen DL, Xu L, Zhang BH. 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology* **80**:75–84 DOI 10.1007/s11103-012-9885-2.

Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y, Zhou R. 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoraceae) revealed by whole-genome sequencing. *National Science Review* **4**:721–734.

Zhang S, Zeng Y, Yi X, Zhang Y. 2016. Selection of suitable reference genes for quantitative RT-PCR normalization in the halophyte *Halostachys caspica* under salt and drought stress. *Scientific Reports* **6**:30363 DOI 10.1038/srep30363.

ISSN: 2348-1900
Plant Science Today
http://www.plantsciencetoday.online

Mini Review

DNA barcoding of plants: Selection of core markers for taxonomic groups

Ankush Ashok Saddhe and Kundan Kumar*

Department of Biological Sciences, Birla Institute of Technology & Science Pilani, K. K. Birla Goa Campus, Goa 403726, India

Article history
Received: 16 October 2017
Accepted: 17 November 2017
Published: 01 January 2018

Editor
K K Sabu

Publisher
Horizon e-Publishing Group

Correspondence
Kundan Kumar
✉ kundan@goa.bits-pilani.ac.in

Abstract

Plant identification is a crucial and routine taxonomic procedure in order to understand and conserve the biodiversity. Anthropogenic activity, pollution, deforestation, and exploitation of natural resources have been threatening to the plant biodiversity. Unfortunately, the major concern of traditional identification of plants is the gradual declined number of taxonomic expertise and lack of tools which accurately discriminate plant seeds, plant parts and seedling, and herbal adulterant. Presently, it is of utmost importance that plant biodiversity to be preserved. To overcome this issues the advent of molecular marker based technique which utilized short fragment of DNA and correctly assign plant taxa to their taxonomic group, called as DNA barcoding. First time, single marker based taxon identification successfully implemented to an animal taxa using mitochondrial cytochrome I (COI) gene fragment. However, Plant DNA barcoding is more complex and it often requires more than one set of DNA markers. In the present review, we have compiled the recent progress of plant DNA barcoding in various taxonomic groups and utility of plastids and nuclear DNA based markers for plant identification.

Keywords

DNA barcoding; rbcL; matK; ITS2

Citation

Saddhe A A, Kumar K. DNA barcoding of plants: Selection of core markers for taxonomic groups. Plant Science Today 2018;5(1):9-13. doi: https://dx.doi.org/10.14719/pst.2018.5.1.356

Introduction

Plant biodiversity is an essential and irreplaceable component of the ecosystem. In the present scenario, biodiversity hotspots are vulnerable due to habitat fragmentation, introduction of exotic species, overexploitation of species and anthropogenic activity. In order to identification, classification and conservation of plant species, present traditional taxonomic expertise is inadequate. Recently, the alternative revolutionary approach based on DNA marker was successfully introduced for an animal taxa using mitochondrial COI gene (1, 2). In contrast, plant DNA barcoding is a more complex and it often requires multiple loci. The Consortium for the Barcode of Life (CBOL) plant working group evaluated the efficacy of maturase K (*matK)* and ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and recommended two-locus based approach with *trnH-psbA* intergenic spacer as a supplementary marker (3). China Plant Barcode of Life recommended the internal transcribed spacer (ITS) as additional candidate plant DNA barcode. Comparative studies of seven markers *trnH-psbA*, *matK*, *rbcL*, chloroplast RNA polymerase subunit (*rpoC1*), *ycf5*, ITS2, and ITS from medicinal plant species were performed (4). Authors recommended that ITS2 is the best potential marker which discriminated 92.7% plants at the species level in more than 6600 plant samples (5). However, most of plant taxonomists have suggested that a multi-locus approach may be essential to resolve plant species (6). Beside all these markers, several plastid regions such as *ycf1*, *atpF-H*, *psbK-psbI*, *ropC1*, *rpoB*, and

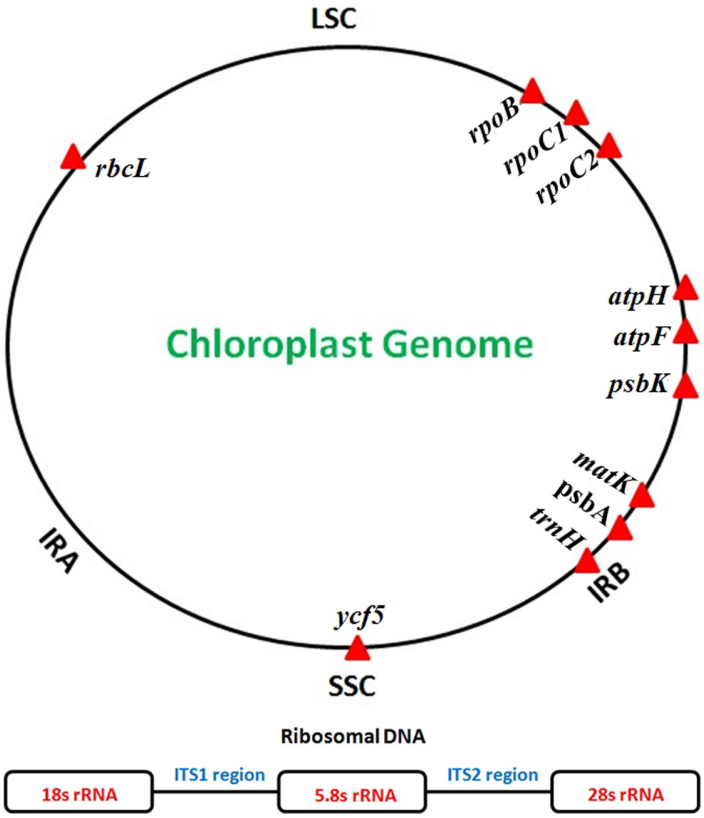


Fig. 1. Schematic representation of plastid (A) and nuclear (B) markers commonly used in plant DNA barcoding. Abbreviations used: LSC-large single copy region, SSC-small single-copy region, IR-Large Inverted repeat (IRA, IRB), *rbcL*-Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, matK- Maturase K, *rpoB* and *rpoC1* codes for chloroplast RNA polymerase subunit, *trnH-psb*- intergenic spacer, *atpF* and *atpH* encode ATP synthase subunits CFO I and CFO III respectively, *psbK* and *psbI* genes encode two polypeptides K and I, *ycf1* gene encodes Tic214 complex, ITS - Internal Transcribed Spacer.

trnL-trnF were frequently evaluated as plant barcode. However, the application of DNA barcoding has been hindered owing to the difficulty in distinguishing closely related species, especially in recently diverged taxa. The plastid markers *rbcL* and *matK* loci exhibited poor resolution in species-rich genera and complex taxa of *Lysimachia*, *Ficus*, *Holcoglossum*, and *Curcuma* (7-10). However, DNA barcoding has significant impact on various research areas such as molecular phylogeny, population genetics, evolution and ecology, biosecurity and food product regulation (6, 11, 12). It helps to detect adulterant in food and medicinal product (6, 11). In recent years, identification and authentication of medicinal plants using DNA barcode markers have made significant progress (6, 11).

Here, we have discussed recent progress of plant DNA barcoding and evaluation of the potential new DNA candidate markers for plant identification. Most of the DNA barcoding works mainly focused on angiosperm, however very few reports are available on DNA barcoding of algae, bryophytes, pteridophytes and gymnosperms. Most commonly used DNA barcode markers utilized in plant identification is depicted in Fig 1. The complete list of DNA barcodes markers used for taxonomic identification is given in Table 1. CBOL recommended two marker based approach for plant identification but still in some group additional group specific markers need to be incorporated. We summarized current update of plant DNA barcoding according to groups such as algae, bryophytes, pteridophytes, gymnosperms and angiosperms.

DNA barcoding of algae

Algae are highly diverse group of organisms and classified into six major groups comprised of Chlorophyta (green algae), Rhodophyta (red algae), Phaeophyta (brown algae), Chrysophyta (golden algae), Bacillariophyta (diatoms), and Ulvophyceae (green algae). Their diversity is reflected at the morphological, structural, genetic, biochemical, physiological and ecological level (13). In addition, there is increased commercial importance of algae group such as ecological bioindicator, production of biofuel, food and fodder for animals (14).The algae taxonomy is a more tedious and difficult to identify microscopic and cryptic species. However, DNA barcoding opened the new alternative and confined ways to identify algal species regardless of life stage. Many DNA markers were evaluated including chloroplast (*rbcL*, *tufA* and 23S), mitochondrial (*COI*) and nuclear genes (18S rDNA, nuITS1 and nuITS2) (15-18). The protist working group of the CBOL

Table 1. List of DNA barcodes markers used in various plant division identification with the references cited.

Plant Division	DNA Barcode	References
Algae	COI, *rbcL*, *matK*, tufA, 23S, 18S rDNA, nuITS1 and nuITS2	Hall et al 2010; Buchheim et al 2011; Caisová et al 2011; Pawlowski et al 2012; Hadi et al 2016
Bryophytes	*rbcL*, *matK*, *rpoB*, *trnH-psbA, trnL-trnF, rps4-trnT, rps19-rpl2*, ITS, *atpF-atpH, psbK-psbI*, and *rpoC1*	Lang et al 2014; Hofbauer et al 2016
Pteridophytes	*rbcL, matK, trnH-psbA, trnL-trnF*, *rpoB, rpoC1*, *atpA*, *atpB*, *rps4-trnS*, and ITS2	Ebihara et al 2010; Ma et al 2010; Li et al 2011, Gu et al 2013; Wang et al 2017
Gymnosperm	*rbcL, matK, ndhJ, rpoB, accD, YCF5 and rpoC1*	Sass et al 2007; Li et al 2011
Angiosperm	*rbcL , matK, trnH-psbA*, ITS2, *trnL-trnF, rpoB, rpoC1, accD, YCF5 , atpF-atpH, trnfM-trnT, trnD-psbM*, *petNtrnC*, *rps16, psaI*	CBOL 2009; Chen et al 2010; China Plant BOL Group, 2011; Saddhe et al 2016; Awad et al 2017; Saddhe et al 2017

recommended two step barcoding in which a universal barcode marker should be used first, followed by the use of a group-specific second barcode (19).

DNA barcoding of bryophytes

Bryophytes comprise three different phylogenetic lineages such as liverworts, hornworts, and mosses. They are the oldest land plants on earth and play an essential ecological role in various ecosystems. However, conservation strategies of bryophytes are always overlooked because of inadequate taxonomic expertise due to miniature size and small distinguish features. The development of new molecular identification tools for bryophytes would improve the ecological studies and help in investigating the impact of global climate change. Recently the closely related *Dicranum scoparium* species were collected from the high Arctic Archipelago of Svalbard resolved by combining five plastid regions (*rpoB*, *trnH-psbA*, *trnL-trnF*, *rps4-trnT*, *rps19-rpl2*) and the nuclear ribosomal ITS region (20).DNA barcoding of moss species diversity such as *Schistidium* species colonizing modern building surfaces showed morphological differences, and suggested cryptic taxa (21). Total 10 DNA barcode markers (*atpF-atpH*, ITS2, *matK, psbK-psbI, rbcL, rpoB, rpoC1*, and *trnH-psbA*) and two popular phylogenetic markers (*rps4* and *trnL-trnF*) were tested in 49 moss species and 9 liverwort species (22).

DNA barcoding of pteridophytes

Pteridophytes comprised ferns and lycophytes which are seedless vascular land plants possessing distinct, free-living sporophyte (2n) and gametophyte (1n) generations (23). Japanese pteridophytes were resolved based on traditional as well as DNA barcode approach and the efficacy of two proposed plastid barcode markers such as *rbcL* and *trnH-psbA* were tested (23). The discriminatory power of the core DNA barcode (*rbcL* and *matK*), and supplementary proposed fern barcodes (*trnH-psbA* and *trnL-F*), were tested across two genera in the hyper diverse polypod clade *Deparia* (Woodsiaceae) and the *Cheilanthes marginata* group (24). Some of the pteridophytes have medicinal value in Chinese medicine and the same plants were tested using six chloroplast DNA barcode such as *psbA-trnH*, *rbcL*, *rpoB*, *rpoC1*, and *matK*) and found that *psbA-trnH* intergenic region was best candidate marker for pteridophytes authentication (25). Pteridophyte genus *Selaginella* is a non-seed bearing plant which was effectively resolved using ITS2 barcode (26). *Adiantum* L. genus was discriminated using morphological characteristic and six plastid markers such as *atpA*, *atpB*, *rbcL*, *trnL-F*, *rps4-trnS* and *matK* (27).

DNA barcoding of gymnosperm

Gymnosperms are seed bearing plants comprises an important four subclasses such as cycadidae, Gingoidae, Gnetidae and Pinidae, representing 12 families, 83 genera and about 990 species (28). Some gymnosperms are considered as 'living fossils' such as Cycads, *Ginkgo biloba*, *Metasequoia glyptostroboides* and *Glyptostrobus pensilis*. However, very few reports are available on gymnosperm DNA barcoding and assessment of potential DNA barcodes in this division. An ancient gymnosperm order Cycadales members were tested using universal DNA barcode markers such as *ndhJ, rpoB, matK, accD, YCF5* and *rpoC1* (29). Recently universality of 9 potential *matK* and 1 *rbcL* primers were assessed for barcoding gymnosperms (30).

DNA barcoding of angiosperm

Angiosperms are an economically important group of flowering plants including 416 families, about 13,164 genera and 295,383 known species (28). The efficacy of most of DNA barcode markers were evaluated using angiosperm plants as a case study. As CBOL recommended *rbcL* and *matK* as core barcode with few supporting markers such as ITS2, *trnH-psbA* was successfully implemented into angiosperm groups. Some inherent problems in plant taxa such as cryptic and closely related taxa, genotypic and phenotypic variability, and natural hybridization which hide the success rate of DNA

barcoding in some plant taxa (31). To overcome this issue, multiple and enormous DNA markers with different combinations were evaluated ranged from plastid coding (*rbcL*, *matK*) to non-coding regions (*trnH-psbA*), nuclear spacer (ITS) (31). The plastid and nuclear markers commonly used in plant DNA barcoding is shown (Fig. 1). The plastid marker *matK* can differentiate more than 90% of species in the Orchidaceae (Orchid family) but less than 49% in the Myristicaceae (nutmeg family) (32-33). The plastid markers such as *rbcL* and *matK* exhibited low resolution in species-rich genera and complex taxa such as *Lysimachia*, *Ficus*, *Holcoglossum*, and *Curcuma* (7-10). The lowest discriminatory power was observed in closely related groups of *Lysimachia* with *rbcL* (26.5-38.1%), followed by *matK* (55.9-60.8%) and combinations of core barcodes (*rbcL* + *matK*) had discrimination of 47.1-60.8% (10). Mangroves identification based on core DNA barcode exhibited *rbcL* 47.72%, *matK* locus assigned (72.09%), ITS2 (87.82%) and combinations of *matK* + ITS2 resolved (89.74%) species however *Avicennia* species required additional *atpF-atpH* marker (34-35). Identification of *Triticum* plants using chloroplast genome-wide analysis revealed combination of the intergenic region (*trnfM-trnT*) with either (*trnD-psbM*), cytochrome b6-f complex subunit 8 (*petN) with trnC*, (*matK-rps16*) or (*rbcL-psaI*) demonstrated a very high discrimination capacity (36).

Future Perspective

Besides the core DNA barcode *rbcL*, *matK*, plant barcoding needs some supplementary markers such as *trnH-psbA* and ITS. Moreover, in closely related and cryptic taxa DNA barcoding is always ambiguous and demands more group specific markers. However, DNA barcoding has significant impact on molecular phylogeny, population genetics, evolution and ecology, biosecurity and food product regulation. Recently developed tools such as metabarcoding coupled with high-throughput sequencing (HTS) are rapid, accurate, and cost-effective alternative to resolve cryptic taxa. Moreover, environmental DNA (eDNA) metabarcoding, which includes universal DNA barcodes and HTS to characterize biological communities from terrestrial and aquatic environmental samples.

Competing interests: The authors have declared that no competing interests exist.

Acknowledgements

Authors thank BITS Pilani K. K. Birla Goa Campus, for providing the necessary support. AAS is gratefully acknowledged to University Grant Commission (UGC), India for senior research fellowship.

References

1. Hebert PDN, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Biol Sci SerB. 2003; 270: S96-S99. https://doi.org/10.1098/rsbl.2003.0025
2. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc Biol Sci SerB. 2003; 270: 313-321. https://doi.org/10.1098/rspb.2002.2218
3. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009; 106: 12794-12797. https://doi.org/10.1073/pnas.0905845106
4. China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci USA. 2011; 108: 19641-19646. https://doi.org/10.1073/pnas.1104551108
5. Chen S, Yao H, Han J, Liu C, Song J, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. Plos One. 2010; 5: e8613. https://doi.org/10.1371/journal.pone.0008613
6. Mohammed Abubakar B, MohdSalleh F, Shamsir Omar MS, Wagiran A. DNA Barcoding and Chromatography Fingerprints for the Authentication of Botanicals in Herbal Medicinal Products. Evid Based Complement Alternat Med. 2017.1352948.
7. Li HQ, Chen JY, Wang S, Xiong SZ. Evaluation of six candidate DNA barcoding loci in Ficus (Moraceae) of China. Mol Ecol Resour. 2012; 12: 783-790. https://doi.org/10.1111/j.1755-0998.2012.03147.x
8. Chen J, Zhao J, Erickson DL, Xia N, Kress WJ. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. Mol Ecol Resour. 2015; 15: 337-348. https://doi.org/10.1111/1755-0998.12319
9. Xiang XG, Hu H, Wang W, Jin XH. DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Mol Ecol Resour. 2011; 11: 1012-1021. https://doi.org/10.1111/j.1755-0998.2011.03044.x
10. Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, et al. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). Mol Ecol Resour. 2012; 12: 98-108. https://doi.org/10.1111/j.1755-0998.2011.03076.x
11. Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant biotechnol J. 2016; 14: 8-21. https://doi.org/10.1111/pbi.12419
12. Janjua S, Fakhar-I-Abbas, William K, Malik IU, Mehr J. DNA Mini-barcoding for wildlife trade control: a case study on identification of highly processed animal materials. Mitochondrial DNA. 2017; 28: 544-6. https://doi.org/10.3109/24701394.2016.1155051
13. Manoylov KM. Taxonomic identification of algae (morphological and molecular): species concepts, methodologies, and their implications for ecological bioassessment. J phycol. 2014; 50: 409-24. https://doi.org/10.1111/jpy.12183
14. Subhadra B. Algal biorefinery-based industry: an approach to address fuel and food insecurity for a

carbon-smart world. J Sci Food Agric. 201; 91: 2-13. https://doi.org/10.1002/jsfa.4207

15. Hadi SIIA, Santana H, Brunale PPM, Gomes TG, Oliveira MD, Matthiensen A, et al. (2016) DNA barcoding green microalgae isolated from neotropical inland waters. Plos One 11(2): e0149284. https://doi.org/10.1371/journal.pone.0149284
16. Hall JD, Fucikova K, Lo C, Lewis LA, Karol KG. An assessment of proposed DNA barcodes in freshwater green algae. Cryptogam, Algol. 2010; 31(4):529-55.
17. Buchheim MA, Keller A, Koetschan C, Forster F, Merget B, Wolf M. Internal transcribed spacer 2 (nu ITS2 rRNA) sequence-structure phylogenetics: towards an automated reconstruction of the green algal tree of life. PLoS One. 2011; 6(2):e16931. https://doi.org/10.1371/journal.pone.0016931
18. Caisová L, Marin B, Melkonian M. A close-up view on ITS2 evolution and speciation-a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol Biol. 2011; 11: 262. https://doi.org/10.1186/1471-2148-11-262
19. Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 2012; 10: e1001419. https://doi.org/10.1371/journal.pbio.1001419
20. Lang AS, Kruijer JD, Stech M. DNA barcoding of Arctic bryophytes: an example from the moss genus *Dicranum* (Dicranaceae, Bryophyta). Polar Biol. 2014; 37: 1157-69. https://doi.org/10.1007/s00300-014-1509-7
21. Hofbauer WK, Forrest LL, Hollingsworth PM, Hart ML. Preliminary insights from DNA barcoding into the diversity of mosses colonising modern building surfaces. Bryophyte Diversity Evol. 2016; 38:1-22. https://doi.org/10.11646/bde.38.1.1
22. Liu Y, YAN HF, Cao T, GE XJ. Evaluation of 10 plant barcodes in bryophyta (Mosses). J Syst Evol. 2010; 48: 36-46. https://doi.org/10.1111/j.1759-6831.2009.00063.x
23. Ebihara A, Nitta JH, Ito M. Molecular species identification with rich floristic sampling: DNA barcoding the pteridophyte flora of Japan. PLoS One 2010; 5: e15136. https://doi.org/10.1371/journal.pone.0015136
24. Li F-W, Kuo L-Y, Rothfels CJ, Ebihara A, Chiou W-L, Windham MD, et al. *rbcL* and *matK* earn two thumbs up as the core DNA barcode for Ferns. PLoS One. 2011; 6: e26597. https://doi.org/10.1371/journal.pone.0026597
25. Ma XY, Xie CX, Liu C, Song JY, Yao H, Luo K, Zhu YJ, Gao T, Pang XH, Qian J, Chen SL. Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast psbA-trnH intergenic region. Biol Pharm Bull. 2010; 33: 1919-24. https://doi.org/10.1248/bpb.33.1919
26. Gu W, Song J, Cao Y, Sun Q, Yao H, Wu Q, Chao J, Zhou J, Xue W, Duan J. Application of the ITS2 region for barcoding medicinal plants of Selaginellaceae in Pteridophyta. PloS one. 2013; 8(6): e67818. https://doi.org/10.1371/journal.pone.0067818
27. Wang AH, Wang FG, Zhang WW, Ma XD, Li XW, Yi QF, Li DL, Duan L, Yan YH, Xing FW. Revision of series Gravesiana (*Adiantum* L.) based on morphological characteristics, spores and phylogenetic analyses. PloS one. 2017; 12(4): e0172729. https://doi.org/10.1371/journal.pone.0172729
28. Christenhusz MJ, Byng JW. The number of known plants species in the world and its annual increase. Phytotaxa. 2016; 261:201-17. https://doi.org/10.11646/phytotaxa.261.3.1
29. Sass C, Little DP, Stevenson DW, Specht CD. DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of Cycads. PloS One 2007; 2: e1154. https://doi.org/10.1371/journal.pone.0001154
30. Li Y, Gao LM, Poudel RC, Li DZ, Forrest A. High universality of matK primers for barcoding gymnosperms. J Syst Evol. 2011; 49: 169-75. https://doi.org/10.1111/j.1759-6831.2011.00128.x
31. Vijayan K, Tsou CH. DNA barcoding in plants: taxonomy in a new perspective. Curr. Sci. 2010:1530-41.
32. Kress J, Erickson DL. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS One. 2007; 2: e508. https://doi.org/10.1371/journal.pone.0000508
33. Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour. 2008; 8: 480-490. https://doi.org/10.1111/j.1471-8286.2007.02002.x
34. Saddhe AA, Jamdade RA, Kumar K. Assessment of mangroves from Goa, west coast India using DNA barcode. SpringerPlus. 2016; 5:1554. https://doi.org/10.1186/s40064-016-3191-4
35. Saddhe AA, Jamdade RA, Kumar K. Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. PLoS One. 2017; 12: e0183245. https://doi.org/10.1371/journal.pone.0183245
36. Awad M, Fahmy RM, Mosa KA, Helmy M, El-Feky FA. Identification of effective DNA barcodes for Triticum plants through chloroplast genome-wide analysis. Comput Biol Chem. 2017; 71: 20-31. https://doi.org/10.1016/j.compbiolchem.2017.09.003