DNA BARCODING FOR IDENTIFICATION OF RICE (*Oryza sativa* L.) VARIETIES DEVELOPED BY Dr. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI.

By

Miss. MHATRE DHANASHREE PRASHANT

B. Sc. (Ag. Biotechnology)

PLANT BIOTECHNOLOGY CENTRE,

FACULTY OF AGRICULTURE,

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH,

DAPOLI - 415 712, DIST. RATNAGIRI (M.S.)

OCTOBER, 2020

DNA BARCODING FOR IDENTIFICATION OF RICE (*Oryza sativa* L.) VARIETIES DEVELOPED BY Dr. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI.

A thesis submitted to the

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI
(AGRICULTURAL UNIVERSITY)

DIST. RATNAGIRI (MAHARASHTRA STATE), INDIA

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE (AGRICULTURE)

In

AGRICULTURE BIOTECHNOLOGY

By

Miss. MHATRE DHANASHREE PRASHANT
B. Sc. (Ag. Biotechnology)

PLANT BIOTECHNOLOGY CENTRE,
FACULTY OF AGRICULTURE,

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH,
DAPOLI - 415 712, DIST. RATNAGIRI (M.S.)

OCTOBER, 2020

Dr. S.V. SAWARDEKAR
Professor and Incharge,
Plant Biotechnology Centre,
College of Agriculture, Dapoli.
Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth
Dapoli – 415 712, Dist. Ratnagiri (M.S.)

CERTIFICATE

This is to certify that the thesis entitled, **"DNA BARCODING FOR IDENTIFICATION OF RICE (*Oryza sativa* L.) VARIETIES DEVELOPED BY DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI."** submitted to the Plant Biotechnology Centre, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli, Dist. Ratnagiri, Maharashtra State, in the partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE (AGRICULTURE) in AGRICULTURE BIOTECHNOLOGY,** embodies the results of a piece of bona-fide research carried out by **Miss. MHATRE DHANASHREE PRASHANT** (Reg. No. 0031) under my guidance and supervision and that no part of this thesis has been submitted for any other degree or diploma or published in other form. All the assistance and help received during the course of investigation and the sources of literature have been duly acknowledged by her.

Place: Dapoli

Date :

(S.V. Sawardekar)
Chairman,
Advisory Committee
and
Research Guide

DNA BARCODING FOR IDENTIFICATION OF RICE (*Oryza sativa* L.) VARIETIES DEVELOPED BY

Dr. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI.

A thesis submitted to the

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI
(AGRICULTURAL UNIVERSITY)

DIST. RATNAGIRI (MAHARASHTRA STATE), INDIA

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE (AGRICULTURE)
In
AGRICULTURE BIOTECHNOLOGY
By

Miss. MHATRE DHANASHREE PRASHANT
B. Sc. (Ag. Biotechnology)

Approved by the Advisory Committee:
Chairman and Research Guide:

(S.V. Sawardekar)
In-charge
Plant Biotechnology Centre,
College of Agriculture, Dapoli.

Members:

R. S. Deshpande)
Associate Professor (CAS)
College of Agriculture, Dapoli.

(A.P. Chavan)
Head and Professor (CAS)
Department of Agronomy,

(S. S. Desai)
Associate Professor (CAS)
Department of Botany

CONTENTS

Acknowledgement

RESEARCH IS CREATING NEW KNOWLEGDE –NEIL ARMSTRONG

With lot of guidance, motivation and concurrent support from various people involved in this 2 year journey of undergoing Post-Graduation studies and as I am in consummation of this academic phase wish to acknowledge those without whom it would be very difficult for me to accomplish and execute the task.

*I am obliged and words are not enough to describe the deep affection, respect and hearted gratitude for my Chairman and Research Guide **Dr. S.V. Sawardekar**, In Charge, Plant Biotechnology Centre, College of Agriculture, Dapoli whose profound interest in the research, inspiring guidance, constructive criticism, ever willing help throughout the course of my post graduate studies and experience given while this research study and preparation of this manuscript with scholarly suggestions will be a treasure to me forever and he has been a source of constant inspiration and support.*

*With endless pleasure, I express my deep sense of gratitude to the members of my Advisory committee, **Prof. R.S. Deshpande** Associate Professor, Plant Biotechnology Centre, College of Agriculture, Dapoli, Associate Professor, Department of Ag. Botany, **Dr. A. P. Chavan**, Head and Professor (CAS),Department of Agronomy, College of Agriculture, Dapoli and **Dr. S. S. Desai** ,Associate Professor (CAS),Department of Botany, College of Agriculture, Dapoli for their kind and helpful suggestion, valuable advice during the present investigation.*

*A special thanks to **Reliance Industries Limited, Mumbai** for the outsourcing work and heartfelt thanks to **Dr. Shantanu Dasgupta**, Senior Vice President, RIL, **Dr. Vyankatesh Prasad**, Head, Molecular Biology Function, RIL, conferring an opportunity to carry a part of research aspect at Reliance Industries Limited, Mumbai. I am grateful to be polished and furnished in my practical knowledge and skills by **Dr. Neeta Madan**, Senior Scientist, RIL. I would express my gratitude and vote of thanks to **Dr. Bhaskar Bhadra**, Technical Team Leader, RIL for his valuable guidance, research aid and technical know-how to complete the project work and also thank **Dr. Dhananjay Dhokane,** Scientist, RIL for guiding me with the lab work and technical skills.*

*I am highly obliged to **Dr. Sanjay D. Sawant**, Honorable Vice Chancellor, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli, **Dr. S. S. Narkhede**, Dean, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli and **Dr. U.V. Mahadkar** Associate Dean, College of Agriculture, Dapoli for providing necessary facilities during the entire course of study.*

Here I would also like put an account of respect and deep hearted gratitude towards **Prof. R.S. Deshpande** Associate Professor, Plant Biotechnology Centre, College of Agriculture, Dapoli, Associate Professor, Department of Ag. Botany, throwing light on 'Work is Worship' and constantly persuading to gain knowledge not only related to studies but also life lessons from various collections in his literary.

I would also like to thank **Mrs. S.S. Sawant**, Senior Research Assistant, Plant Biotechnology Centre , College of Agriculture, Dapoli. I feel privileged to express my heartier gratitude to all the teachers and staff members of College of Agriculture, Dapoli for their immense help and constant encouragement in successful completion of my M.Sc degree programme.

I express my sincere thanks to **Mr. Vipul Kelkar,** PhD Fellow, KAU, Thissur for essential information regarding the topic with his kind co-operation, timely help, valuable suggestions during the completion of my research work. A heartfelt to teachers who have been guiding me since my budding stage **Mr. Saurabh Kadam**, PhD Fellow, Amity University, Panvel, **Mr. Devendra Rasam**, **Miss. Kirti Joshi** and my gratitude towards always helpful and resourceful with knowledge **Dr. Ajinkya Ambavane**, Breeder, Ankur Seeds.

I also thank **Mr. A. B. Kamble, Sawant Kaki**, **Dakshata** , **Vishwas P** and **Shigwan P.** for their generous help and co-operation during the course of my research work and study period. I convey my thanks to all the staff members of my department who helped me regularly.

I please to express my sincere heartfelt thanks to my dear seniors **Vinayak Patil, Vishal Patil, Sumitra Sartape** for their constant advice and friendly nature towards me. I also would specially like to mention thank my juniors **Shubham Patil, Rohit Shinge and Pranit Karanje** for the lovely time and very kind co-operation.

Last but not the least; I would infinitely like to thank my special lab mates **Pravin Pawar, Pooja Bhosale and Tushar Gajre** for their help in lab work and also for sharing all the wonderful moments during the M.Sc. degree programme. I avail this opportunity to thank **Nilesh Jorvekar** and all batchmates for their co-operation and for good time and company during the M. Sc. degree Programme.

Remembering my grandfather, **Late Divakar Mhatre** would like to dedicate this work to this for constantly emphasizing on importance of education and whatever I am today is all because of my strong pillars of support **Mrs. Salonee Mhatre and Mr.Prashant Mhatre** and care and love from grandmother, brother **Shreerang** and family. Life is always merrier with a great company of friends, I am happy to have you **Sumedha, Tanvi, Akshay, Miheer, Sneha, Priyankita, Shubham** besides me always.

I acknowledge DBSKKV for all the various facilities and campus to acquire knowledge and make its best use for the course of study.

Lastly, I would like to acknowledge and thank all those whom I might have missed out unknowingly.

Seeking wisdom finally thanking the Almighty for this wonderful life…….

Place : Dapoli.

Date: ***(Mhatre D.P.)***

APPENDIX I

ABBREVIATIONS

AFLP	:	Amplified Fragment Length Polymorphism
BSA	:	Bulk Segregant Analysis
bp	:	Base Pair
CBOL	:	The Consortium for the Barcode of Life (BOLD)
cM	:	Centi Morgan
CIA	:	Chloroform-Isoamyl Alcohol
DNA	:	Deoxyribose Nucleic Acid
dNTPs	:	Deoxyribo nucleotide tri-phosphates
EDTA	:	Ethylene Diamine Tetra Acetic Acid
EtBr	:	Ethidium Bromide
et al.,	:	Co-workers
GD	:	Genetic Distance
i.e.	:	That is
SSR	:	Simple Sequence Repeats
MVSP	:	Multivariate Statistical Package
mM	:	Milli Molar
$MgCl_2$	:	Magnesium Chloride
NaCl		Sodium Chloride
ng μl^{-1}		Nano gram per micro litre

nm	:	Nano meter
OD	:	Optical Density
PAGE	:	Polyacrylamide Gel Electrophoresis
PCR	:	Polymerase Chain Reaction
PIC	:	Polymorphic Information Content
RM	:	Rice Microsatellite
RF	:	Resolution Factor
RAPD	:	Random Amplified Polymorphic DNA
RFLPs	:	Restriction Fragment Length Polymorphisms
RNA	:	Ribose Nucleic Acid
RNase	:	Ribonuclease enzyme
rpm	:	revolution per minute
SI	:	Similarity Index
SSRs	:	Simple Sequence Repeats
SDS	:	Sodium Dodecyl Sulphate
PVP	:	Polyvinyl Pyrrolidone

TAE	:	Tris-Acetate EDTA
TE	:	Tris Buffer
Tris HCl	:	Tris-Hydrochloride
UV	:	Ultra Violet
U µl^{-1}	:	Units per micro litre
UPGMA	:	Unweighted Pair Group Method with Arithmetic
viz.,	:	Mean
µl	:	Namely
MT	:	Micro litre
MSL	:	Metric tonnes
%	:	Mean sea level
AtpA	:	Per cent
atpI-	:	ATP synthase alpha subunit
atpH	:	ATP synthase subunits CFO IV- CFO III
BLAST	:	Basic Local Alignment Search Tool
ITS	:	Internal transcribed spacer
ITS1	:	Internal transcribed spacer 1
ITS2		Internal transcribe spacer 2

APPENDIX II

COMPOSITION OF CHEMICALS

CIA (25 ml)

Chloroform	24 ml
Isoamyl alcohol	1 ml

Ethidium bromide (10 ml)

EtBr	0.01 mg
Distilled water	10 ml

Stored at 4^0C

6x Gel Loading dye (50 ml)

0.25% Bromophenol blue	125 mg
40% sucrose in water	20 g

Stored at 4^0C

5% Sarcosyl (10 ml)

Sarcosyl	0.5 g
Distilled water	10 ml

1x TE buffer (10 ml)

10x TE	1 ml
Sterile water	9 ml

50x TAE (1 lit.)

Tris Base	242 g
Glacial acetic acid	57.1 ml
0.25 M EDTA (pH 8.0)	200 ml
Final volume	1000 ml

APPENDIX III
SEQUENCES USED IN ANALYSIS

- **Sequences obtained from primer *matK***

>KARJAT-1---1M1_Contig1

GCGACAAAACCCTCATATTTCCTTTTTAGAAGACAAATTTTGCACTTACA
TTATCTAGCACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGC
TTCAACTCCTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTG
CGATTCTTTCTCAACTACTACTCGAATTGGAATAGTTTTATTACTTCAAT
GAAATCCATTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCC
TATATAACTCTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTA
AACAATCTTCTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACG
AATCATCTTTTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCT
TTTTTTCGGAAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGT
TCGATATCAAGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTG
AAGAAGAAATGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCT
CTTTTTGGACTCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTC
TTGCTTCGATTTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGT
TAGTAAGGAATCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAAT
GAAAAAATTCGATACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTA
GCAAAAGCCCAATTTTGTACTGGATCGGGGCATCCTATTAGTAAACCAA
TTTGGACCGATTTATCGGATTGGGATATTCTTGATCGGTTTGGTCGGAT
ATGTAGAAATCTTTTTCATTATCACAGTGGATCTTCAAAAAAGAAGACTT
TGTATCGACTAAAGTATATACTCGACTCTCGGG

>KARJAT-2---2M1_Contig1

CATATTCCTTTTTAGAAGACAAATTTTTGCACTTACATTATCTAGCACAT
ATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAACTCCTTCA
ATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTTTCTCA
ACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATTTTT
ATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTCTTA
TGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTCTTG
CTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTTCTA
GGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAAAAC
CATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGATATCAAGGA
AAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAATGGA
AATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGACTCAA
CCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTTCGATTTTC
TGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAATCAA
ATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAAATTCGATAC
CAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCAATTT
TGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCGATTTAT
CGGATTGGGATATTCATGATCGGTTTGGTCGGATATGTAGAAATCTTTT
TCATTAGCACAGTGGATCGTCAGAACAGAAGACTTTGTATCGACTAAAG
TATATAGTCGACTTCCG

>KARJAT-3---3M1_Contig1
TATTTCCTCATATTCCTTTTTAGAAGACAAATTTTTGCACTTACATTATCT
AGCACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAAC
TCCTTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATT
CTTTCTCAACTACTACTCGAATTGGAATAGTTTATTACTTCAATGAAAT
CCATTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATAT
AACTCTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAA
TCTTCTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCA
TCTTTTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTT
CGGAAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGAT
ATCAAGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAA
GAAATGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTT
GGACTCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTT
CGATTTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAA
GGAATCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAA
ATTCGATACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAA
GCCCAATTTTGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGA
CCGATTTATCGGATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAG
AAATCTTTTTCATTATCACAGTGGATCTTCAAAAAAGAAGACTTTGTATC
GACTAAAGTATATACTTCGACTTCGCGGNNNNNR

>KARJAT-4---4M1_Contig1
TATCTACTCATATTCCTTTTTTAGAAGACAAATTTTTGCACTTACATTATC
TAGCACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAA
CTCCTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGAT
TCTTTCTCAACTACTACTCGAATTGGAATAGTTTATTACTTCAATGAAA
TCCATTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATAT
AACTCTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAA
TCTTCTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCA
TCTTTTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTT
CGGAAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGAT
ATCAAGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAA
GAAATGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTT
GGACTCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTT
CGATTTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAA
GGAATCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAA
ATTCGATACCAA
AGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCAATTTTGT
ACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCGATTTATCG
GATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAGAAATCTTTTTC
ATTATCACAGTGGATCTTCAAAAAAGAAGACTTTGTATCGACTAAAGTA
TATACTTCGACTTCGCGGG

>KARJAT-5---5M1_Contig1
CATCAATATTTCCTTTTTTAGAAGACAAATTTTTGCACTTACATTATCTAG
CACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAACTC
CTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTT
TCTCAACTACTACTCGAATTGGAATAGTTTATTACTTCAATGAAATCCA
TTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACT
CTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTT
CTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTT
TTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGG
AAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGATATCA

AGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAA
TGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGAC
TCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTTCGAT
TTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAA
TCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAAATTCG
ATACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCA
ATTTTGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCGAT
TTATCGGATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAGAAATC
TTTTTCATTATCACAGTGGATCTTCAAAAAAGAAGACTTTGTATCGACTA
AAGTATATACTCGACTCTTCG

>RATNAGIRI-1--8M1_Contig1
CCACATATTCCTTTTTTAGAAGACAAATTTTTGCACTTACATTATCTAGC
ACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAACTCC
TTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTT
TCTCAACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAATCCA
TTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACT
CTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTT
CTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTT
TTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGG
AAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGATATCA
AGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAA
TGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGAC
TCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTTCGAT
TTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAA
TCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAAATTCG
ATACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCA
ATTTTGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCGAT
TTATCGGATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAG
AAATCTTTTTCATTATCACAGTGGATCTTCAAAAAAGAAGACTTTGTATC
GACTAAAGTATATACGTCGACTCTC

>RATNAGIRI-2---9M1_Contig1
GCTCTATTATCTCCACAATTTCCTTTTAGAAGACAAATTTGCACTACATA
TCTAGCACATATAGAAATACCTATCTATCATTGATATCTGCTCACTCTCA
TACCGATCAGATGTTCCATCTGCATTTATGCGATCTTTCTCAACTACTAC
TCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATTTTTATTTTGAAA
AAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTCTTATGTATCAGA
ATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTCTTGCTTGCGATT
AACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTTCTAGGAAGATG
GAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAAAACCATATGGTT
CGTTATGGATCCCCTTATG
CATTATGTTCGATATCAAGGAAAGGCAATTCTTGCATCAAAAGGAACTC
TTCTTTTGAAGAAGAAATGGAAATGTTACCTTGTCCGTTTGTGGCAATA
TTCTTTCTCTTTTTGGACTCAACCGCAAAGGATCCATCTAAACCAATTAG
AAAACTCTTGCTTCGATTTTCTGGGGTACTTTTCAAGTGTACCAATAAAT
TCTTTGTTAGTAAGGAATCAAATGCTGGAGAATTCATTTCTAATAGATAC
TCAAATGAAAAAATTCGATACCAAAGTCCCTGTTACTCCTCTCATTGGA
TCTTTAGCAAAAGCCCAATTTTGTACTGGATCGGGGCATCCTATTAGTA
AACCAATTTGGACCGATTTATCGGATTGGGATATTCTTGATCGGTTTGG
TCGGATATGTAGAAATCTTTTTCATTATCACAGTGGATCTTCAAAAAAGA
AGACTTTGTATCGACTAAAGTATATACTTCGACTCCGGNNNNNR

>RATNAGIRI-3---10M1_Contig1
CACATATTCCTTTTTTAGAAGACAAATTTTTGCACTTACATTATCTAGCA
CATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAACTCCT
TCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTTTC
TCAACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATT
TTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTC
TTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTC
TTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTT
CTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAA
AACCATATGGTTCGTTATGG
ATCCCCTTATGCATTATGTTCGATATCAAGGAAAGGCAATTCTTGCATC
AAAAGGAACTCTTCTTTTGAAGAAGAAATGGAAATGTTACCTTGTCCGT
TTGTGGCAATATTCTTTCTCTTTTTGGACTCAACCGCAAAGGATCCATC
TAAACCAATTAGAAAACTCTTGCTTCGATTTTCTGGGGTACTTTTCAAGT
GTACCAATAAATTCTTTGTTAGTAAGGAATCAAATGCTGGAGAATTCATT
TCTAATAGATACTCAAATGAAAAAATTCGATACCAAAGTCCCTGTTACTC
CTCTCATTGGATCTTTAGCAAAAGCCCAATTTTGTACTGGATCGGGGCA
TCCTATTAGTAAACCAATTTGGACCGATTTATCGGATTGGGATATTCTT
GATCGGTTTGGTCGGATATGTAGAAATCTTTTTCATTATCACAGTGGAT
CTTCAAAAAAGAAGACTTTGTATCGACTAAAGTATATACTTCGACTTCG
CNNNNNNR

>RATNAGIRI-4--11M1_Contig1
ACATATTCCTTTTTAGAAGACAAATTTTTGCACTTACATTATCTAGCACA
TATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAACTCCTTC
AATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTTTCTC
AACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATTTT
TATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTCTT
ATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTCTT
GCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTTC
TAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAAA
ACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGATATCAAG
GAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAATG
GAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGACTC
AACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTTCGATTT
TCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAATC
AAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAAATTCGAT
ACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCAAT
TTTGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCGATTT
ATCGGATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAGAAATCTT
TTTCATTATCACAGTGGATCTTCAAAAAAGAAGACTTTGTATCGACTAAA
GTATATACTCACTCTTGT

>PANVEL-1---17M1_Contig1
ATTTTCTTCCATATTCCTTTTTAGAAGACAAATTTTTGCACTTACATTATC
TAGCACATATAGAAATACCCTATCCTATCCATTTGGATATCTTGCTTCAA
CTCCTTCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGAT
TCTTTCTCAACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAA
TCCATTTTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATAT
AACTCTTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAA
TCTTCTTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCA
TCTTTTCTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTT

CGGAAAACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGAT
ATCAAGGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAA
GAAATGGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTT
GGACTCAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTT
CG
ATTTTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGG
AATCAAATGCTGGAGAATTCATTTCTAATAGATACTCAAATGAAAAAATT
CGATACCAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCC
CAATTTTGTACTGGATCGGGGCATCCTATTAGTAAACCAATTTGGACCG
ATTTATCGGATTGGGATATTCTTGATCGGTTTGGTCGGATATGTAGAAA
TCTTTTTCATTATCACAGTGGATCTTCAAAAAGAAGACTTTGTTTCGACT
AAAGTAATTCGTCGACTTCTGTG

- **Sequences obtained from primer *rbcL***

>KARJAT-3---3M2_Contig1

CCCCCAAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTA
TAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGACACTGATATC
TTGGCAGCATTCCGAGTAACTCCTCAGCCGGGGGTTCCGCCCGAAGA
AGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGGACAAC
TGTTTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATG
CTATCACATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTAT
GTAGCTTATCCATTAGACCTATTTGAAGAGGGTTCTGTTACTAACATGT
TTACTTCCATTGTGGGTAACGTATTTGGTTTCAAAGCCCTACGCGCTCT
ACGTCTGGAGGATCTGCGAATTCCCCCTACTTATTCAAAAACTTTCCAA
GGTCCGCCTCATGGTATCCAAGTTGAAAGGGATAAGTTGAACAAATAC
GGTCGTCCTTTATTGGGATGTACTATTAAACCAAAATTGGGATTATCTG
CAAAAAATTATGGTAGAGCATGTTATGAGTGTCTACGCGGTGGACTTG
ATTTTACCAAAGATGATGAAAACGTAAACTCACAACCATTTATGCGTTG
GAGGGACCGTTTTGTCTTTTGTGCCGAAGCTATTTATAAATCACAGGCC
GAAACCGGTGAAATTAAGGGGCATTACTTGAATGCTACTGCAGTACTG
GTGCAAAAAAAAAA

>KARJAT-4---4M2_Contig1

TTTTTTTTTATATGCACCCCAACAGAAACTAAAGCAAGTGTTGGATTTAA
AGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAA
ACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCG
GGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTC
TACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGAT
CGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGAT
AATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGG
GTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTC
AAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACT
TATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGG
ATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACC
AAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGT
CTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCAC
AACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTAT
TTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAAT
GCTACTGCAGGCCTGGGGAAAAAAATT

>KARJAT-5---5M2_Contig1
TTTTTTTTTATTTGCCCCCCAAACAGAAACTAAAGCAAGTGTTGGATTTA
AAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGA
AACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCC
GGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTT
CTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGA
TCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGA
TAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGG
GTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTC
AAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACT
TATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGG
ATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACC
AAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGT
CTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCAC
AACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTAT
TTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAAT
GCTACTGTGGCAGCCAAAAAAACGT

>KARJAT-7--6M2_Contig1
TTTTTAAATGCCAACCCAAACAGAAACTAAAGCAAGTGTTGGATTTAAA
GCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAA
CCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCGG
GGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCT
ACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGATC
GTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGATA
ATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGT
TCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTCAA
AGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACTTA
TTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGGAT
AAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACCAA
AATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGTCT
ACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCACAA
CCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTATTT
ATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAATG
CTACTGCTAACCCATCGTGCGGACGGCA

>RATNAGIRI-1---8M2_Contig1
TTTTTTTTTATGATTTGTCAACCCCCCAAAACTAAAGCAAGTGTTGGATT
TAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTAC
GAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAG
CCGGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAAT
CTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCT
TGATCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGA
GGATAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAG
AGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGG
TTTCAAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCC
TACTTATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAA
AGGGATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTA
AACCAAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGA
GTGTCTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAAC
TCACAACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAG
CTATTTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTT
GAATGCTACTGTGGCAGCCAAAAAAACG

>RATNAGIRI-2---9M2_Contig1
ACGAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAA
TTGACTTACTACACCCCGGAGTACGAAACCAAGGACACTGATATCTTG
GCAGCATTCCGAGTAACTCCTCAGCCGGGGGTTCCGCCCGAAGAAGC
AGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGT
TTGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTAT
CACATCGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTA
GCTTATCCATTAGACCTATTTGAAGAGGGTTCTGTTACTAACATGTTTAC
TTCCATTGTGGGTAACGTATTTGGTTTCAAAGCCCTACGCGCTCTACGT
CTGGAGGATCTGCGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
CGCCTCATGGTATCCAAGTTGAAAGGGATAAGTTGAACAAATACGGTC
GTCCTTTATTGGGATGTACTATTAAACCAAAATTGGGATTATCTGCAAAA
AATTATGGTAGAGCATGTTATGAGTGTCTACGCGGTGGACTTGATTTTA
CCAAAGATGATGAAAACGTAAACTCACAACCATTTATGCGTTGGAGGG
ACCGTTTTGTCTTTTGTGCCGAAGCTATTTATAAATCACAGGCCGAAAC
CGGTGAAATTAAGGGGCATTACTTGAATGCTACTGAGGAATGGAAAAA
ATGG

>RATNAGIRI-3---10M2_Contig1
TTTTAAGTGCACAACCCCCACAGAAACTAAAGCAAGTGTTGGATTTAAA
GCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAA
CCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCGG
GGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCT
ACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGATC
GTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGATA
ATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGT
TCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTCAA
AGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACTTA
TTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGGAT
AAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACCAA
AATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGTCT
ACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCACAA
CCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTATTT
ATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAATG
CTACTGCAGGCATGGGAAAAAT

>RATNAGIRI-4---11M2_Contig1
TTTTTAGTGTCCCCCCCAGCAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGT
GTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAACCAAGGACACT
GATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCGGGGGTTCCGCCCGAAG
AAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTT
TGGACTGATGGACTTACCAGTCTTGATCGTTACAAAGGCCGATGCTATCACAT
CGAGCCCGTTGTTGGGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATT
AGACCTATTTGAAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAA
CGTATTTGGTTTCAAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTC
CCCCTACTTATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAA
GGGATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACCAA
AATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGTCTACGCG
GTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCACAACCATTTATGC
GTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTATTTATAAATCACAGGCC
GAAACCGGTGAAATTAAGGGGCATTACTTGAATGCTACTGCATTAACTATTCGG
GATAAAAA

>RATNAGIRI-24---12M2_Contig1
YYTTTTTTTTATGTGCGCCACACCCAGCAGAAACTAAAGCAAGTGTTGG
ATTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAG
TACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCT
CAGCCGGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCG
AATCTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAG
TCTTGATCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGG
GGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTG
AAGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATT
TGGTTTCAAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCC
CCCTACTTATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTT
GAAAGGGATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTA
TTAAACCAAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTAT
GAGTGTCTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTA
AACTCACAACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCG
AAGCTATTTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTA
CTTGAATGCTACTGAGGAAAGGCAAAAAAGG

>RATNAGIRI-7---15M2_Contig1
YYTTTTTAGTGTGCCCCCCCAACAGAAACTAAAGCAAGTGTTGGATTTA
AAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGA
AACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCC
GGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTT
CTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGA
TCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGA
TAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGG
GTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTC
AAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACT
TATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGG
ATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACC
AAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGT
CTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCAC
AACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTAT
TTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAAT
GCTACTGCAGTACCATGGCGAAAAY

>RATNAGIRI-8---16M2_Contig1
TTTTTTTTTTTGATTGCACACCCAGACAGAAACTAAAGCAAGTGTTGGAT
TTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTA
CGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCA
GCCGGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAA
TCTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTC
TTGATCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGG
AGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAA
GAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTG
GTTTCAAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCC
CTACTTATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGA
AAGGGATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATT
AAACCAAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATG
AGTGTCTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAA
CTCACAACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAA
GCTATTTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACT
TGAATGCTACTGCATTAACACATGCGAAAAGCYYYYY

>PANVEL-1---17M2_Contig1
TTAGGCTCCACCCCCACAGAAACTAAAGCAAGTGTTGGATTTAAAGCT
GGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGTACGAAACCA
AGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAGCCGGGG
GTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGAATCTTCTAC
TGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGTCTTGATCGT
TACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGGGAGGATAAT
CAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAGAGGGTT
CTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTCAAA
GCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACTTAT
TCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGGATA
AGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACCAAA
ATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTATGAGTGTCTA
CGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTAAACTCACAAC
CATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCGAAGCTATTTA
TAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTACTTGAATGC
TACTGCATAGCCAATGTGGCGACAA

>PANVEL-2---18M2_Contig1
TTTTTTTATATAGTAGGTCCCCCCAGCAGAAACTAAAGCAAGTGTTGGA
TTTAAAGCTGGTGTTAAGGATTATAAATTGACTTACTACACCCCGGAGT
ACGAAACCAAGGACACTGATATCTTGGCAGCATTCCGAGTAACTCCTC
AGCCGGGGGTTCCGCCCGAAGAAGCAGGGGCTGCAGTAGCTGCCGA
ATCTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACTTACCAGT
CTTGATCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTGGG
GAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGA
AGAGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTT
GGTTTCAAAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCC
CCTACTTATTCAAAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTG
AAAGGGATAAGTTGAACAAATACGGTCGTCCTTTATTGGGATGTACTAT
TAAACCAAAATTGGGATTATCTGCAAAAAATTATGGTAGAGCATGTTAT
GAGTGTCTACGCGGTGGACTTGATTTTACCAAAGATGATGAAAACGTA
AACTCACAACCATTTATGCGTTGGAGGGACCGTTTTGTCTTTTGTGCCG
AAGCTATTTATAAATCACAGGCCGAAACCGGTGAAATTAAGGGGCATTA
CTTGAATGCTACTGCTTTTGGGATTGTTGTGCTCATGGAGAAACCAAAA
GC
GAGGGAGGAGAATAGGGAA

>PALGHAR-2---20M2_Contig1
CCCACAGAAACTAAAGCAAGTGTTGGATTTAAAGCTGGTGTTAAGGATTATAAA
TTGACTTACTACACCCCGGAGTACGAAACCAAGGACACTGATATCTTGGCAGC
ATTCCGAGTAACTCCTCAGCCGGGGGTTCCGCCCGAAGAAGCAGGGGCTGCA
GTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTTTGGACTGATGGACT
TACCAGTCTTGATCGTTACAAAGGCCGATGCTATCACATCGAGCCCGTTGTTG
GGGAGGATAATCAATATATCGCTTATGTAGCTTATCCATTAGACCTATTTGAAG
AGGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAACGTATTTGGTTTCA
AAGCCCTACGCGCTCTACGTCTGGAGGATCTGCGAATTCCCCCTACTTATTCA
AAAACTTTCCAAGGTCCGCCTCATGGTATCCAAGTTGAAAGGGATAAGTTGAA
CAAATACGGTCGTCCTTTATTGGGATGTACTATTAAACCAAAATTGGGATTATC
TGCAAAAAATTATGGTAGAGCATGTTATGAGTGTCTACGCGGTGGACTTGATTT
TACCAAAGATGATGAAAACGTAAACTCACAACCATTTATGCGTTGGAGGGACC
GTTTTGTCTTTTGTGCCGAAGCTATTTATAAATCACAGGCCGAAACCGGTGAAA
TTAAGGGGCATTACTTGAATGCTATCTGTGTGTCTGCTTTTTCCCCCTGGGAAT
TGTCTCCCATTCTATTTTTTTCGTAACGATGGAAAGAGAAAAAGAGATG

- **Sequences obtained from primer *psbA-trnH***

>KARJAT-1---1M3_Contig1

TTTTTTGTTTTGGGTTCGTAATGCTCACAACTTCCCTCTAGACCTAGCT
GCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATA
GCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAG
CAAGATATTGGGTATTTCTTTTTTTTTTTATTTTGAATCTTTCTATTCTG
AATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTC
GTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATA
GGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGAT
TGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACG
TTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCA
ATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGA
TTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTC
GATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAA
AGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATC
AAGGCAGTGGATTGTGAATCACATGCTATATTGAGGGTTTTTCTAATT
TTTCCTCACATTAATTACGTTTTTTCTACGCACGACCTAGAACACTACCA

>KARJAT-2---2M3_Contig1

TTTTGTAGTGCCAAAGGGGGTTTTGTACCAACTGGCATTTCTCAATTTA
TACGTGCAGAAAGATATAAATCTTGTCGTTAACTAATTCTGTATAGAAG
ATTTCAAAATAAAAAAAAAAAGAGAAGTGCCCAATATCTTGTGAAGCAA
GATGTTGGGTATTTTCTAGCTTTCCTTTCTTCAAAAGATTGCTATATGTA
GCAGAAAAGCTTTATCCATTGAGAGATGGAACTTCAAGAGCAGCTAGG
TCCTAGTGGAGTTTTTGTGTTTGGCTCTAATGCTCACAACTTCCCTCTA
GACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTG
CTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATAT
CTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCT
TTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACT
TTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGA
CCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATG
AATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCG
GGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTA
TTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTC
ACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCC
AATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAA
TGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCA
AGTGGATCAAGGCAGTGGATTGTGATACCCC

>KARJAT-3---3M3_Contig1

ATTTTTGTTTTGGGTTCTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTT
GAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATTTTTGA
AGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGCAAGATATTGGGTATT
TCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAATTCAGTTAACGACGAGATT
TAGTATCCTTTCTTGCACTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATT
CCCCCAATTTGCGACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCT
TTCCATTATGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGAT
GCCCGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTA
TTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGC
TGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGA
TCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGAGAAA
ATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATCAAGGCAGTGGA
TTGTGATAC

>KARJAT-4---4M3_Contig1
CTTTTGTGTTTTGGGTTCGTAATGCTCACAACTTCCCTCTAGACCTAGC
TGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATAT
AGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCA
GCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCT
GAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACT
CGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCAT
AGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGA
TTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCAC
GTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCC
AATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTG
ATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCT
CGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAA
AAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGAT
CAAGGCAGTGGATTGTGAATCCCCCCCGTGGCGAGAATTA

>KARJAT-5---5M3_Contig1
ACCTCTTACCTCACACAATCCTTTGTGGGTCGCAAATGGGAGTTGTAC
CATCTCGCATTTCACGAGTTAGAGAGGCAGAAGGATAATAACGACTCT
CTTAGACGCTTTTCGAATAGTGGGATTCATAATAAAGAACAAAAAGAAA
TAGGCCAAATCTGCTGAAGCAAGATGTTGGTAGTCTCAGCTTTCCTTCT
TCAAAAATGTATATGTAGCGAAAAGCTTTATCCATTGAGAGAGGACTCA
ACAAAACAAACCCACTTTTTTTTTTTGCTCTATGCTCACAACTTCCCTC
TAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTC
TGCTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAAT
ATCTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAAT
CTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCA
CTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGC
GACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTA
TGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCC
CGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTC
TATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTG
TCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGT
CCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATG
AATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGC
CAAGTGGATCAAGGCAGTGGATTGTGATACCCAGTAGATTTTTCCACC
TTTTTTAATGTTTTAATTTATATTTGTTTACTTTTTTGACTGTTGTTTACGT
AAGTGTGATA

>KARJAT-7---6M3_Contig1
TATGTGGGGCGCTCGAATGGAGTTCGTTCACACTGCGCATAACGGAGT
AATCAAGCTGCGAGCACGATAAAAAACGCTTCTCTTTGTGACTTCTCAT
TTATGGCCTAAAATAGAAAAGGTCAATATGAATGCCGGAAAATGTTGCT
TCAGCAAGCGTGTTGGCCGGGCATAGCTCTCTTTTCTTCCCAAAAGTT
GTCCATGTCAGTCGAGAATCTTTATGCGAAAAGAGAAGAGGAACTCTA
CACCAACCCAAAAACATTTTTTTTTTTTTTTCCAAGCTCACAACTTCCC
TCTAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTT
TCTGCTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCA
ATATCTTGCTGAAGCAAGATATTGGGTATTTCTTTTTTTTTTTATTTTGAA
TCTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGC
ACTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTG
CGACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATT

ATGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGC
CCGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTT
CTATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGT
GTCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATG
TCCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGAT
GAATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAG
CCAAGTGGATCAAGGCAGTGGATTGTGATACCCC

>RATNAGIRI-1---8M3_Contig1
TTTTTTTTTTGGGTTCTAATGCTCACAACTTCCCTCTAGACCTAGCTGCT
CTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCA
ATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAGCAA
GATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAATT
CAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGTG
AAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGGA
TTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGT
ATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTA
CTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATCAAG
GCAGTGGATTGTGATAC

>RATNAGIRI-2---9M3_Contig1
TCTTGTGTGGTCGCAAATGGGGATTTCGTACCAACTCGCATTTCAGAG
TTATGAAAGTGCAAGAAAAGATATAAATCTGTCGTAACGATTCAGAATA
GAGAGATTTCAAATAAAAAAAAGAGAAATACCCAATATCTTGTGAAGCA
GGATATTGGGTATTTCTAGCTTTCCTTTCTTCAAAAAATTGCTATATGTT
AGCAGAAAAGCTTTTTCCATTAAGAGATGGAACTTTCTAGAGCAGCTAG
GTCCAATTTAAGGGGGCGCCCAATTTTTCTCATGCTCACAACTTCCCTC
TAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTC
TGCTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAAT
ATCTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAAT
CTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCA
CTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGC
GACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTA
TGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCC
CGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTC
TATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTG
TCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGT
CCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATG
AATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGC
CAAGTGGATCAAGGCAGTGGATTGTGATACCCNNNNNR

>RATNAGIRI-3---10M3_Contig1

CAGAAAGTGCGAGAGAGATATAAAATCTCGTTGTAATGATTCTGATAGA
GAGATTCAAAATAAAAAAATAAAAAGAAATACCCCAATATCTTGCTGAA
GCAAGATGTTGGGTATTTCTAGCTTTCCTTTTTCTCAAAAAATTGCTATT
TCTAGCAGAAAAGCTTTATCCATTAAGAGAGGAAATTCCAGAGCAGCTA
GGTCAAATTGAAGTTTGTGTTTTGGGCTCTAATGCTCACAACTTCCCTC
TAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTC
TGCTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAAT
ATCTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAAT
CTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCA
CTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGC
GACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTA
TGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCC
CGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTC
TATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTG
TCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGT
CCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATG
AATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGC
CAAGTGGATCAAGGCAGTGGATTGTGAATCCCCCNNNNNNR

>RATNAGIRI-4---11M3_Contig1

GGCAAAAATAAAAAAAAAAACAGAAATACGCCTAAGTCTGATGAAGTAA
GAACTGGGTGCTAGGCGGTTTCCTTTCTACAAAAAAGCGTCTGTGCAG
CGAGAAGCATTTATTCATGGAGAGAGGAATTTTCACAGTGACAAGACC
AAAATTTACTTTTTGTGTTTTTTTTTTTCTAATGCTCACAACTTCCCTCTA
GACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTG
CTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATAT
CTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCT
TTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACT
TTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGA
CCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATG
AATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCG
GGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTA
TTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTC
ACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCC
AATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAA
TGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCA
AGTGGATCAAGGCAGTGGATTGTGAATCCCCCCGGGCGCCGAA

>RATNAGIRI-24---12M3_Contig1

TTGTTGTTTTTGGGTTCGTAATGCTCACAACTTCCCTCTAGACCTAGCT
GCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATA
GCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAG
CAAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTG
AATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTC
GTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATA
GGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGAT
TGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACG
TTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCA
ATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGA

TTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTC
GATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAA
AGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATC
AAGGCAGTGGATTGTGATAC

>RATNAGIRI-7---15M3_Contig1

TTGTTGTTATGGGTTACGTAATGCTCACAACTTCCCTCTAGACCTAGCT
GCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATA
GCAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAG
CAAGATATTGGGTATTTCTTTTTTTTTTTATTTTGAATCTTTCTATTCTG
AATTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTC
GTGAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATA
GGATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGAT
TGTATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACG
TTACTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCA
ATAAATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGA
TTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTC
GATCGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAA
AGAGAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATC
AAGGCAGTGGATTGTGAATCCCC

>RATNAGIRI-8---16M3_Contig1

CATCTTACATAACATGTCTTTGTAGGTCGCAAGACGGGATGTTGTACCA
ACTGGCATTTTCTGAGTATGAAGTGCAGCAAGATAACAAATCTCGTCGT
AAGCAGTTCCGTATGAGAGAGTCTAGATAGGACATAGAGAGATGCCCA
AGATCTGTTGAGCGAGATGTTGGGTATTATCTAGCTTTCCTTTCTTCAA
AAATGCTATATGTTAGCGGAAAGCTTATCCATTGAGAGAGGACTCCAA
GACAGCCAACCCCAATTTTTGTGTTTTTTTTTCTATGCTCACAACTTCC
CTCTAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGCTT
TTCTGCTAACATATAGCAATTTTTGAAGAAAGGAAAGCTAGAAATACCC
AATATCTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTATTTTG
AATCTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTCTT
GCACTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAATT
TGCGACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTCCA
TTATGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAGAT
GCCCGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGACCT
TTTCTATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTTTT
CGTGTCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATTC
GATGTCCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATAA
TGATGAATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGAT
GTAGCCAAGTGGATCAAGGCAGTGGATTGTGATAAATTAAGTGCAGTA
ATTCTCCCCTTGTGAAGGTTTAAGTGATTTCTTTAACTGGATAAAGGGG
GGAAAGTTACCCAT

>PANVEL-1---17M3_Contig1

TGCGCCGCCTACGGGCCTCGCTCGCTAATCCGGATACCAGGGCTGTA
ACAGCGGCGATGAGAATATCTTCTCCTTGTTGTACTCTTTCGATTAGTG
CCGGTTAAGGAACAGCGGACATACTGAACGGGGGGATAGTGCTCACG
AACCCAGCTGGTGCGAGCGACATGCTCCCCTTTCTGCTAAAAATGGCG
TGAGCTATGCAGAAAGCCTTTTTTTACTGAAGGGTGAAATTCCACAGG
GATCCGGTAAAAATCACATTTTTTTTTTTTTCCCTAATGCTCACAACTT

CCCTCTAGACCTAGCTGCTCTTGAAGTTCCATCTCTTAATGGATAAGGC
TTTTCTGCTAACATATAGCAATTTTTTGAAGAAAGGAAAGCTAGAAATAC
CCAATATCTTGCTTCAGCAAGATATTGGGTATTTCTTTTTTTTTTTTATTT
TGAATCTTTCTATTCTGAATTCAGTTAACGACGAGATTTAGTATCCTTTC
TTGCACTTTCATAACTCGTGAAATGCCGAGTTGGTACGAATTCCCCCAA
TTTGCGACCTACCATAGGATTTGTTATGTAAATAGGTATATGTTCCTTTC
CATTATGAATCGCGATTGTATGGCCAACCATTGCGGGTAGAATGCTAG
ATGCCCGGGACCACGTTACTATTGTTTCTTTCTCCTCCTTCATATTGAC
CTTTTCTATTTTTGCCAATAAATGATGAGCTACAAAAGGATTCGTTTTTT
TTCGTGTCACAGCTGATTACTCCTTTTTTTCCATTTTAAAGAGTGGCATT
CGATGTCCAATATCTCGATCGAAGTATGGAGGTCAGAATAAATAGAATA
ATGATGAATGGAAAAAAGAGAAAATCCTTTAGCTGGATAAGGGGCGGA
TGTAGCCAAGTGGATCAAGGCAGTGGATTGTGAACGGGTACTAGCGA
GAATATATTGCTTGCCTTTAATTCTGATACGGGGCCAACTATCAACACC
CTTGCAAGGCTTGGGGGTGG

>PANVEL-2---18M3_Contig1

TTTTTGTTTATGGGTTTATAATGCTCACAACTTCCCTCTAGACCTAGCTG
CTCTTGAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAG
CAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAGC
AAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAA
TTCAGTTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGT
GAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGG
ATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTG
TATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTA
CTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGGGGCGGATGTAGCCAAGTGGATCAAG
GCAGTGGATTGTGAATCACCCAAGGGCCAGAA

>PALGHAR-2---20M3_Contig1

TTTTTTTTTTTGGGTTTCTAATGCTCACAACTTCCCTCTAGACCTAGCTG
CTCTTGAAGTTCCATCTTTTAATGGATAAGGCTTTTCTGCTAACATATAG
CAATTTTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGC
AAGATATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAA
TTCAGTTAACGACAAAATTTAGTATCCTTTCTTGCACTTTCATAACTCGT
GAAATGCCAAGTTGGTACAAATTCCCCCAATTTGCGACCTACCATAGG
ATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTG
TATGGCCAACCATTGCGGGTAAAAGGCTAAATGCCCGGGACCACGTTA
CTATGGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAAAGGGGCATTCAATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAAAATAAATAAAATAATGATGAAGGGAAAAAAAA
AAATCCTTTAGCTGATTAGGGCGAATGTAGCCAGTGAATCAAGGCAGT
GATGGATCCC

- **Sequences obtained from primer *trnH-psbA***

>RATNAGIRI-3---10M4-Contig1

GTTACAGATTTGGTCAAGAGGAAGAGACTTATAATATTGTGGCCGCTCA
TGGTTATTTTGGCCGATTAATCTTCCAAAATGCTAGTTTTAACAACTCTC
GTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGTAGGGATTTGGTT
CACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTAAACGGATTCAATT
TCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTATTAATACTTGGGC
TGATATCATCAACCGTGCTAATCTTGGTATGGAAGTAATGCACGAACGT
AATGCTCACAACTTCCCTCTAGACCTAGCTGCGCTTGAAGTTCCATCTT
TTAATGGATAAGGCTTTTTTGCTAACATATAGCAATTTTTGAAGAAAGGA
AAGCTAGAAATACCCAATATCTTGCTTCAGCAAGATATTGGGTATTTCTT
TTTTTTTTTTATTTTGAATCTTTCTATTCTGAATTCAGTTAACGACGAGAT
TTAGTATCCTTTCTTGCACTTTCATAACTCGTGAAATGCCGAGT
TGGTACGAATTCCCCCAATTTGCGACCTACCATAGGATTTGTTATGTAA
ATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGTATGGCCAACCA
TTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTACTATTGTTTCTTT
CTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATAAATGATGAGCTA
CAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTACTCCTTTTTTTCC
ATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGATCGAAGTATGGAG
GTCAGAATAAATAGAATAATGATGAATGGAAAAAAGAGAAAATCCTTTA
GCTGGATAAGGGGCG

>RATNAGIRI-4---11M4_Contig1

CTGCTAATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATAT
TGTGGCCGCTCATGGTTATTTTGGCCGATTAATCTTCCAATATGCTAGT
TTTAACAACTCTCGTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGT
AGGGATTTGGTTCACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTA
AACGGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTA
TTAATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGT
AATGCACGAACGTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTT
GAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATT
TTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGCAAGAT
ATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAATTCAG
TTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGT
GAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGG
ATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTG
TATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTA
CTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGGGGCG

>RATNAGIRI-7---15M4_Contig1

AATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATATTGTGG
CCGCTCATGGTTATTTTGGCCGATTAATTCTTCCAATATGCTAGTTTTAA
CAACTCTCGTTCTTTACACTTCTTCTTGGGCTGCTTGGCCTGTAGTAGG
GATTTGGTTCACTGCTTTAGGTATTAGTACTAGGGCTTTCAATCTAAAC
GGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTATTA

ATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGTAAT
GCACGAACGTAATGCTCACAACTTCCCTCAAGACCTAGCTGCTCTTGA
AGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATTTT
TGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAGCAAGATAT
TGGGTATTTCTTTTTTTTTTTATTTGAATCTTTCTATTCTGAATTCAGTT
AACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGTGAAATG
CCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGGATTTGTT
ATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGTATGGC
CAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTACTATTG
TTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATAAATGAT
GAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTACTCCTTT
TTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGATCGAAGT
ATGGAGGTCAGATCATAAAATAATGATGAATGGAAAAAAGAGAAAATCC

TTTAGCTGGATA AGGGGCGG

>RATNAGIRI-8---16M4_Contig1

CTGCTAATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATAT
TGTGGCCGCTCATGGTTATTTTGGCCGATTAATCTTCCAATATGCTAGT
TTTAACAACTCTCGTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGT
AGGGATTTGGTTCACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTA
AACGGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTA
TTAATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGT
AATGCACGAACGTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTT
GAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATT
TTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGCAAGAT
ATTGGGTATTTCTTTTTTTTTTTATTTGAATCTTTCTATTCTGAATTCAG
TTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGTGAAAT
GCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGGATTTG
TTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGTATGG
CCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTACTATT
GTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATAAATGA
TGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTACTCCTT
TTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGATCGAAGT
ATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGAGAAAAT
CCTTTAGCTGGATAAGGGG

>PANVEL-1---17M4_Contig1

TGCTAATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATATT
GTGGCCGCTCATGGTTATTTTGGCCGATTAATCTTCCAATATGCTAGTT
TTAACAACTCTCGTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGTA
GGGATTTGGTTCACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTAA
ACGGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTAT
TAATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGTA
ATGCACGAACGTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTTG
AAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATTT
TTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGCAAGATA
TTGGGTATTTCTTTTTTTTTTTATTTGAATCTTTCTATTCTGAATTCAGT
TAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGTG
AAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGGA
TTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGT

ATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTA
CTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGGG

>PANVEL-2---18M4_Contig1

GCTAATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATATTG
TGGCCGCTCATGGTTATTTTGGCCGATTAATCTTCCAATATGCTAGTTT
TAACAACTCTCGTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGTA
GGGATTTGGTTCACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTAA
ACGGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTAT
TAATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGTA
ATGCACGAACGTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTTG
AAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATTT
TTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTGAAGCAAGATA
TTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAATTCAGT
TAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGTGA
AATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGGAT
TTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTGTA
TGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTAC
TATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATAA
ATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTAC
TCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGATC
GAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGGGG

>PALGHAR-2---20M4_Contig1

CTGCTAATGAGGGTTACAGATTTGGTCAAGAGGAAGAGACTTATAATAT
TGTGGCCGCTCATGGTTATTTTGGCCGATTAATCTTCCAATATGCTAGT
TTTAACAACTCTCGTTCTTTACACTTCTTCTTGGCTGCTTGGCCTGTAGT
AGGGATTTGGTTCACTGCTTTAGGTATTAGTACTATGGCTTTCAATCTA
AACGGATTCAATTTCAACCAATCTGTAGTTGATAGCCAAGGTCGCGTTA
TTAATACTTGGGCTGATATCATCAACCGTGCTAATCTTGGTATGGAAGT
AATGCACGAACGTAATGCTCACAACTTCCCTCTAGACCTAGCTGCTCTT
GAAGTTCCATCTCTTAATGGATAAGGCTTTTCTGCTAACATATAGCAATT
TTTGAAGAAAGGAAAGCTAGAAATACCCAATATCTTGCTTCAGCAAGAT
ATTGGGTATTTCTTTTTTTTTTTTATTTTGAATCTTTCTATTCTGAATTCAG
TTAACGACGAGATTTAGTATCCTTTCTTGCACTTTCATAACTCGT
GAAATGCCGAGTTGGTACGAATTCCCCCAATTTGCGACCTACCATAGG
ATTTGTTATGTAAATAGGTATATGTTCCTTTCCATTATGAATCGCGATTG
TATGGCCAACCATTGCGGGTAGAATGCTAGATGCCCGGGACCACGTTA
CTATTGTTTCTTTCTCCTCCTTCATATTGACCTTTTCTATTTTTGCCAATA
AATGATGAGCTACAAAAGGATTCGTTTTTTTTCGTGTCACAGCTGATTA
CTCCTTTTTTTCCATTTTAAAGAGTGGCATTCGATGTCCAATATCTCGAT
CGAAGTATGGAGGTCAGAATAAATAGAATAATGATGAATGGAAAAAAGA
GAAAATCCTTTAGCTGGATAAGG

- **Sequences obtained from primer *atpH-atpI***
  >KARJAT 2---2M5_Contig1

  TATCCATTGGTTATTTACAAGCGGTATTCAAGCTCTTATTTTGCAACGT
  TAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTG
  ACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGG
  TTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTAT
  GAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAA
  TTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
  GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
  CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
  CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
  AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
  TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
  GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
  TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
  GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
  AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
  AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
  TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
  TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
  CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
  GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
  GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
  AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGGTTAACAA
  A

  >KARJAT 3---3M5_Contig1

  TTTCTTGGTTATTTACAAGCGGTATTCAAGCTCTTATTTGTGCAACGTTA
  GCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTGAC
  TAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGGTT
  GCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTATGA
  ATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAATT
  GCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTAGA
  TCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTCCA
  CTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATACA
  CAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCAA
  GTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAAT
  TCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATTG
  GATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCTT
  TTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATCG
  AAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAGA
  AAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTAA
  AAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATTT
  GTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCTT
  GGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCATC
  ATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCCG
  TAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCAG
  GACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAAA
  ATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTAACAA

>KARJAT 4---4M5_Contig1
TGATTCTTGGTTATTTACAGCGGTATTCAAGCTCTTATTTTTGCTACGTT
AGCCGCAGCCTATATGGGTGAATCCATGGAGGGTCATCATTGAATTGA
CTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGGTT
GCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTATGA
ATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAATT
GCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTAGA
TCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTCCA
CTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATACA
CAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCAA
GTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAAT
TCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATTG
GATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCTT
TTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATCG
AAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAGA
AAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTAA
AAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATTT
GTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCTT
GGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCATC
ATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCCG
TAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCAG
GACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAAA
ATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTTAACAA

>KARJAT 5---5M5_Contig1

ATGATTTCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAA
CGTTAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAA
TTGACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCAT
GGTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGT
ATGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGG
AATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCT
AGATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTT
CCACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAAT
ACACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCC
AAGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCA
ATTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTAT
TGGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACC
TTTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATAT
CGAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAA
GAAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACT
AAAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAAT
TTGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTC
TTGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCA
TCATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGC
CGTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGC
AGGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTA
AAATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTT

>KARJAT 7---6M5_Contig1

GTTTCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTGCAACGT
TAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTG
ACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGG
TTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTAT
GAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAA
TTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTA

>RATNAGIRI 1---8M5_Contig1

TGATGATTCTCTTGTGATTATTTACAAGCGGTATTCAAGCTCTTATTTTT
GCAACGTTAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCAT
TGAATTGACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCAT
GCATGGTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATT
GCGTATGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATT
ACGGAATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTC
AGGCTAGATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATG
GGTTTCCACTTTAAGGAATTTTTTTGAATCCGATTCAATAGAAAATGAG
AAAATACACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATA
TTCCCAAGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCAT
ATCCAATTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATT
CCTATTGGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATT
TCACCTTTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAG
GATATCGAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTT
GGAAAGAAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAG
AAACTAAAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTT
TCAATTTGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGA
ATTTCTTGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGA
ACTCATCATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGA
TTGGCCGTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACT
GCTGCAGGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGA
AGGTAAAATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCT
TAACA.

>RATNAGIRI 2---9M5_Contig1
TCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAACGTTAG
CCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTGACT
AGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGGTTG
CGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTATGAA
TATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAATTG
CCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTAGAT
CTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTCCAC
TTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATACAC
AAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCAAGT
TAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAATTC
GATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATTGGA
TCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCTTTTA
TTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATCGAAG
AGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAGAAAG
AGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTAAAAA
GGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATTTGTA
CTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCTTGGT
TGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCATCATG
AATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCCGTAG
GTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCAGGAC
AAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAAAATAC
GCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTAACAA

>RATNAGRIRI 3---10M5_Contig1

TGTTTCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAACG
TTAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATT
GACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATG
GTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTA
TGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGA
ATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTAACAA
TT.

>RATNAGRIR 4---11M5_Contig1

TGACAGGTGCTGCGAGGCCCTCCTATCCAGGGCGGGTCCGCGGACA
CGAAGTCTAAACCCCGCGCCTGCCGGCGCGGAGGCTCGAACCTGCG
ATGACCAGCATAAGGGAGGCAGCACTAATTAGTGAATTCTCGTGAGTC
CCTAGGGTTAAAAACACCAAGGGGCTAAGGAGGCACAAAACCCCTAAT
GACCCATACTAACCTATATCGAACACAACTAACTGCAACCCCCAGACT
GGCACGACCACCCGATGACTCCAACTTGCCACTAAACGTCCGATAGAT
TTTAATCGTAATAGGTTAGGGTACCCAATATTCAATTAATTACCTTCGCA
ACAATCGGATGGTCCCTCCCACCGAATTTTTG

>RATNAGIRI 24---12M5_Contig1

TGTTTCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAACG
TTAGCCGCAGCCTATATGGGTGAATCCATGGAGGGTCATCATTGAATT
GACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATG
GTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTA
TGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGA
ATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTAACAA

>RATNAGIRI 7---15M5_Contig1

TATTTCTTGGTTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAACGT
TAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTG
ACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGG
TTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTAT
GAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAA
TTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA

AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGATAGTTGAACAT

>RATNAGIRI 8---16M5_Contig1
CATGATTCCCTTGGTTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCA
ACGTTAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGA
ATTGACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCA
TGGTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCG
TATGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACG
GAATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGG
CTAGATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGT
TTCCACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAA
ATACACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTC
CCAAGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATC
CAATTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCT
ATTGGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCA
CCTTTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGAT
ATCGAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGA
AAGAAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAA
CTAAAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCA
ATTTGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATT
TCTTGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACT
CATCATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTG
GCCGTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCT
GCAGGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGG
TAAAATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCGTTTA
A

>PANVEL 2---18M5_Contig1
TTATCCTTTTGTTAATTACAAGCGGTATTCAAGCTCTTATTTTTGCAACG
TTAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATT
GACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATG
GTTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTA
TGAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGA
ATTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA

GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTCCTTAGTCTAGCTTTCCGGATAGTTAC

>PALGHAR 2---20M5_Contig1
GTTTCTTGGATTATTTACAAGCGGTATTCAAGCTCTTATTTTTGCAACGT
TAGCCGCAGCCTATATAGGTGAATCCATGGAGGGTCATCATTGAATTG
ACTAGTTTTCAAAATAGTCTTTTTTTTAGCTTAACTCAATTCATGCATGG
TTGCGGAAAATTCGCTTGGTTGGAAAACAAAATAGTTAGAATTGCGTAT
GAATATACAATCTAGAGTTGTAGAAGAGAGAATAGGCTATATTACGGAA
TTGCCAAACAAAGTATATAGGCATTAGGGAGGGGCGGAGTCAGGCTA
GATCTATATCCTTTATGTCTATAAGTTCAGTCATCTTTTGTATGGGTTTC
CACTTTAAGGAATTTTTTTTGAATCCGATTCAATAGAAAATGAGAAAATA
CACAAACAAAATAGAAGAAACAAATTGATATGGGATATTATATATTCCCA
AGTTAGATTCATTATCTAATCCGATATATGGAATCGGATTCCATATCCAA
TTCGATGCAGCATATTGTTATCAATTGGATATCTTGATTTAATTCCTATT
GGATCTGGATTAGGTCGATTTCCATAGGGGTTCTTCCTCTATTTCACCT
TTTATTATGAATTAGATGATAGGGGAAAAAATAGAAACTCAAGGATATC
GAAGAGGAAAGAAAGAAGGATGGAATGAAAGATCAGTTGGTTGGAAAG
AAAGAGAAATAGAATAATGAGTACACAAACCTCTAATGATTAGAAACTA
AAAAGGAGATCTCGAAGCAGTTCGGAGAATTCAGATTATCGTTTCAATT
TGTACTTTTTAGTTACTTCTGTCCAATAGAGCTTAGAAATATGAATTTCT
TGGTTGATTGTATCCTTAACCATTTCTTTTTTTTGACACGAGGAACTCAT
CATGAATCCACTAATTGCTGCTGCTTCCGTTATTGCTGCTGGATTGGCC
GTAGGTCTTGCTTCTATTGGGCCTGGAGTTGGTCAAGGTACTGCTGCA
GGACAAGCTGTAGAAGGTATTGCGAGACAGCCAGAAGCAGAAGGTAA
AATACGCGGTACTTTATTGCTTAGTCTAGCTTTTATGGAAGCTTTAACAA
TT

- **Sequences obtained from primer *ITS***

>KARJAT-3---3M6Contig1
GCGATACTTTGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGA
ACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCCTG
GGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAGG
GCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGTG
GGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCATG
GTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCGC
GGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCGAGAGCGAAC
GCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTA
AGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAGT
AACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGCC
GCGCCGTCCGAATTGTAGTCTAAAAAAAACCGTCA

>KARJAT-4---4M6Contig1
GCGATACTTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTT
GAACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCC
TGGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCGAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG

TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTAAAAAAAACCGTCA

>KARJAT-5---5M6Contig1
GCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTG
AACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCCT
GGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCCAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG
TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTGAGAAGCGTCCACATCGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG
GCTGTGTGCACGAACCCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCG
CCACTGGCAGCAGCCACTGGTAACAGGATTAGC

>KARJAT-7---6M6Contig1
TGCGATACCTTTGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTT
GAACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCC
TGGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCCAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG
TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTGAGAAGCGTCA

>RATNAGIRI-3---10M6Contig1
GCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTG
AACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCCT
GGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCGAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG
TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTGAAAAACCGTCA

>RATNAGIRI-4---11M6Contig1

TGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTT
GAACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCC
TGGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCGAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG
TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTAAAAAAACCGTCA

>PALGHAR-2---20M6Contig1

TTGCGATACTTTGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTT
GAACGCAAGTTGCGCCCGAGGCCATCCGGCCGAGGGCACGCCTGCC
TGGGCGTCACGCCAAAAGACGCTCCACGCGCCCCCCCTATCCGGGAG
GGCGCGGGGACGCGGTGTCTGGCCCCCCGCGCCTCGCGGCGCGGT
GGGCCGAAGCTCGGGCTGCCGGCGAAGCGTGCCGGGCACAGCGCAT
GGTGGACAGCTCGCGCTGGCTCTAGGCCGCAGTGCACCCCGGCGCG
CGGCCGGCGCGATGGCCCCTCAGGACCCAAACGCACCGAGAGCGAA
CGCCTCGGACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTT
AAGCATATAAATAAGCGGAGGAGAAGAAACTTACGAGGATTCCCCTAG
TAACGGCGAGCGAACCGGGAGATGCCCAGCTTGAGAATCGGGCGGC
CGCGCCGTCCGAATTGTAGTCTAAAAAAGCGTCA

- **Sequences obtained from primer *petA-psbJ***

>KARJAT-1---1M7_Contig1

ATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTGG
ATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGAA
AAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCGG
AAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTATT
CCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAAGA
GAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTGTTT
GTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGATTT
ATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAGAA
AGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGTCT
CTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAATT
AAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAGAG
TATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATTGA
ACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCCGCCATA
TTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAAGA
GGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCCGAA
TAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAATCGACA
CAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATTAAG
AAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCTCGC
CCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAGGAA
TTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAAAAG
ATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATCGCG

GCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATTCATT
TCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAAAAAC
TTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGG

>KARJAT-2---2M7_Contig1

CGGTATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATT
CTGGATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAAC
ATGAAAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAG
AGCGGAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAA
TCTATTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAA
CAAAGAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGT
TTTGTTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGT
TGATTTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGA
GGAGAAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGG
AAGTCTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAAC
CTTAATTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATG
GAAGAGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACT
TAATTGAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCC
GCCATATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGC
AAAAGAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTT
GCCGAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAA
TCGACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAG
GATTAAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTG
TTCTCGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAA
TAAGGAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAA
AAAAAAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATA
AATCGCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGA
AATTCATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACC
AAAAAAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGC

>KARJAT-3---3M7_Contig1

TATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTG
GATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGA
AAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCG
GAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTA
TTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAA
GAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTG
TTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGAT
TTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAG
AAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGT
CTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAA
TTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAG
AGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATT
GAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCCGCCA
TATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAA
GAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCC
GAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAATCG
ACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATT
AAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCT
CGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAG
GAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAA

AAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATC
GCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATT
CATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAA
AAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGCC

>KARJAT-4---4M7_Contig1

GGTATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTC
TGGATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACAT
GAAAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAG
CGGAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATC
TATTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACA
AAGAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTT
TGTTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTT
GATTTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAG
GAGAAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGA
AGTCTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACC
TTAATTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGG
AAGAGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTT
AATTGAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCC
GCCATATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGC
AAAAGAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTT
GCCGAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAA
TCGACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAG
GATTAAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTG
TTCTCGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAA
TAAGGAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAA
AAAAAAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATA
AATCGCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGA
AATTCATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACC
AAAAAAACTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGC

>KARJAT-5---5M7_Contig1

TTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTGGA
TTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGAAA
AAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCGGA
AAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTATTC
CATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAAGAG
AAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTGTTTG
TTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGATTTAT
AACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAGAAAG
AAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGTCTCTT
GTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAATTAAA
ACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAGAGTAT
CCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATTGAAC
AAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCCGCCATATTT
TCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAAGAGG
TTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCCGAATA
GGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAATCGACACA
AGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATTAAGAA
GACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCTCGCCC

TTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAGGAATT
GCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAAAAGAT
TTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATCGCGGC
TTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATTCATTTC
GTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAAAAACTT
GTGCCAAAATAACGGATGCGAAGAAGAACAAAAGG

>KARJAT-7---6M7_Contig1

ATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCCCG
GATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGA
AAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCG
GAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTA
TTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAA
GAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTG
TTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGAT
TTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAG
AAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGT
CTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAA
TTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAG
AGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATT
GAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCACGCCA
TATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAA
GAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCC
GAATAGGAAGAAAAAAGATTCATATTTCGATACAAGAAGAATCGACACA
AGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATTAAGAA
GACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCTCGCCC
TTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAGGAATT
GCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAAAAGAT
TTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATCGCGGC
TTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATTCATTTC
GTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAAAAACTT
GTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGC

>RATNAGIRI-1---8M7_Contig1

TGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTGGAT
TGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGAAAA
AATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCGGAA
AGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTATTCC
ATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAAGAGA
AAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTGTTTGT
TAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGATTTATA
ACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAGAAAGA
AAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGTCTCTTG
TAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAATTAAAA
CTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAGAGTATC
CGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATTGAACAA
GTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCCGCCATATTTTC
TTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAAGAGGTT
CAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCCGAATAG
GAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAATCGACACAA
GAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATTAAGAAG
ACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCTCGCCCTT

GCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAGGAATTGC
AGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAAAAGATTT
CCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATCGCGGCTT
TTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATTCATTTCGT
ACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAAAAACTTGT
GCCAAAATAACGGATGCGAAGAAGAACAAAAGG

>RATNAGIRI-2---9M7_Contig1

AACCGGTATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCAT
ATTCTGGATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAA
ACATGAAAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATT
AGAGCGGAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTT
AATCTATTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAAT
AACAAAGAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTC
GTTTTGTTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCA
GTTGATTTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCG
AGGAGAAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAG
GAAGTCTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAA
CCTTAATTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAAT
GGAAGAGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAAC
TTAATTGAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCC
CGCCATATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAG
CAAAAGAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTT
TGCCGAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGA
ATCGACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAG
GATTAAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTG
TTCTCGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAA
TAAGGAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAA
AAAAAAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATA
AATCGCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGA
AATTCATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACC
AAAAAAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAATAGGCC
T

>RATNAGIRI-24---12M7_Contig1

ATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTGG
ATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGAA
AAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCGG
AAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTATT
CCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAAGA
GAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTGTTT
GTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGATTT
ATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAGAA
AGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGTCT
CTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAATT
AAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAGAG
TATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATTGA
ACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCACGCCATA
TTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAAGA
GGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCCGAA
TAGGAAGAAAAAAGATTCATATTTCGATACAAGAAGAATCGACACAAGA
AAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATTAAGAAGAC

CAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCTCGCCCTTG
CCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAGGAATTGCA
GACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAAAAGATTTC
CAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATCGCGGCTTT
TTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATTCATTTCGTA
CAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAAAAACTTGTG
CCAAAATAACGGATGCGAAGAAGAACAAAAGGC

>RATNAGIRI-7---15M7_Contig1

TATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCTG
GATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACATGA
AAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAGCG
GAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATCTA
TTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACAAA
GAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTTTG
TTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTTGAT
TTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAGGAG
AAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGAAGT
CTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACCTTAA
TTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGGAAG
AGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTTAATT
GAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCCGCCA
TATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGCAAAA
GAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTTGCC
GAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAATCG
ACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAGGATT
AAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTGTTCT
CGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAATAAG
GAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAAAAAA
AAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATAAATC
GCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGAAATT
CATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACCAAAA
AAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGCC

>RATNAGIRI-8---16M7_Contig1

CGGTATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATT
CTGGATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAAC
ATGAAAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAG
AGCGGAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAA
TCTATTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAA
CAAAGAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGT
TTTGTTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGT
TGATTTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGA
GGAGAAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGG
AAGTCTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAAC
CTTAATTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATG
GAAGAGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACT
TAATTGAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTTGTCCCCC
GCCATATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGC
AAAAGAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTT
GCCGAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAA
TCGACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAG

GATTAAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTG
TTCTCGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAA
TAAGGAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAA
AAAAAAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATA
AATCGCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGA
AATTCATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACC
AAAAAAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAATAGGCC

>PANVEL-2---18M7_Contig1

GTATTGCTGTGATTGGTTTAATAGGTGTTTTCTTTTACGGTTCATATTCT
GGATTGGGTTCATCTCTATAGTAATTGGAGGGACCAGGTTGTAAACAT
GAAAAAATAGGAACTTAGCGGGTCCTTACCCCCTTTTATCTGATTAGAG
CGGAAAGGACCCGCGGAATTCTTACTCTTATAACCCGACTTTTTTAATC
TATTCCATAACCTACAAACGGGTTTCCTATTTTGATTTGTAAAAATAACA
AAGAGAAAGCCTTTGCTCTGATGGTCAGAATCCCTCTATTTATTCGTTT
TGTTTGTTAATTTTGTTTGTTAATAATGTTAGTGAAGTAATCCATCAGTT
GATTTATAACCCCCTCGCCAATGAAATTAATTCACTTTTATGAAGCGAG
GAGAAAGAAAGGATCAATCGAGGATCCAATATTTTATTCCCCGGAGGA
AGTCTCTTGTAAATCATTGATTTAATTTAGAGTAATAAACTAATAAAACC
TTAATTAAAACTACTCAACTAGCCTAAAAATAAAAAACAGCCTTCAATGG
AAGAGTATCCGTTTTTTTTGCAAAATTTCTGTAAGTTCTAAATTGAACTT
AATTGAACAAGTCAAGCAATTCTATTTGGTCACAAAAAATTCGTCCCCC
GCCATATTTTCTTTCCCTCTTTTTGTCCACTACATCGCCTGAACCTAAGC
AAAAGAGGTTCAAGAGACAGACCCAAATAAAGAAGAAATAGTAATCTTT
GCCGAATAGGAAGAAAAAAGATTCATATTTCGATACAATACAAGAAGAA
TCGACACAAGAAAAAGGCAAAAAAGCCTTTTTCTTGTGCCCAATAGAG
GATTAAGAAGACCAGCAATCCCTCCTAAAAGGATTTGTTCCTTTTATTG
TTCTCGCCCTTGCCTTGCATTCTCCTCGTTTTGCATGAGATGATAGAAA
TAAGGAATTGCAGACAAGACATCTCGCAAACAAAAAATAGTCTAAACAA
AAAAAAGATTTCCAAAATGGATCATAGAGAATTCTAGCAATCGCCAATA
AATCGCGGCTTTTTACCAACTTGATGGTAAGAAATCCCGGGACCTAGA
AATTCATTTCGTACAATTGAACCTTTTCAAACTGTTTCTTTTTGAGAACC
AAAAAAACTTGTGCCAAAATAACGGATGCGAAGAAGAACAAAAGGCCT

- **Sequences btained from primer *trnK***

>KARJAT-1---1M8_CONTIG

ATCGATCTTGGTTACGTTGAGTAACCAAACCTCTCCCCAAATCCTGCTT
GCTTCTTTCGTTGCACTCGCCGTTCCCCATGTAGTGATACGCTATTTCT
TGTCCGTAGAGCCAGTCTTCTCGTCTTGTTCGGAGCCAGTTGATCCAC
TTTCTTGGGCTTACACTTCAGTGTACTCCTCAGAGTGGAGTCGGTGGA
CGAAATTTTGCCTTGGTCATTTATCTTTCTCTGCTTGTCCCTCGGTTTTG
TCCATTGGTTTTTCTACTGGATTCACCGAGAGAGCATTGTTTCGCGGTT
TATCCAATGGCCATACACGCTCACTGTGCGGTCCGCGCAGGGTGTTG
CATGTTGTTTTGGCGAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTA
AACGGTTCTTCTTCCGCTACCGTTCCCATCCATTGCGTTTATGCTCGTT
CCGTGCTCTTATGTGTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCT
CTTTTGACTTTGGTCGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGA
CACATGCAATCCATTTCATCCTTTTCCGTCCAAACCCATCAAGAATTTC
CCTTTCCGATAAATCGGGCCAAATTCAAAGGTCTACTCATAGGATGCC
CAGCGTTCCCCTACATCAGGCACTGGGCTTTTGCTAACGATCCAATGA

GATCAGGGAGTTCCTCCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTT
TCTACCAGAATTGGAGCCTCCCGCTTTCCATTCCTTCCTTACACCCAGT
TTATCGGTACTCTTGTATTCCACCCCAGAAATCCGAAGCAAGAGTTTTC
TAATTGGTTTTGTTGCGTCCTTTGCGGTTGTGTCCATTCAGTGAAAGAC
CCTTGCCACGATCGGTCGCGGTCTC

>KARJAT-2---2M8_CONTIG

AAATTAGTCCAAAACCTCTCCCCATACCGTCCTTGCATTTTCCGTTGCA
CACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGAG
GTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTACT
CCTGTACAGAGTACTGGGCAGACGGGATACTGTCTAGGTTTTACCTTG
AGCTTTTCTCAATCACCCTCGAGTTTTCCCGTTTGGTTTTCTGCTTGATT
CAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCATAC
ACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCGA
ACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCCG
CTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTGT
ACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTCG
ATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTT
CATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCG
GGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACAT
CAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTC
CGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGC
CAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTGG
TTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGCG
TTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGTATAGTCG
CCACCAACGGTCTGGTCGCATTTCC

>KARJAT-3---3M8_CONTIG

TTCCATAGTGATCCAACCTCTTCTCCCGATTCGTGTCGTAGCATTTTCC
GTTGCACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCC
GACGAGGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTA
TGTACTCCTGTACAGAGTACTGGGCAGACGGGATACTGTCTAGGTTTT
ACCTTGAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGC
TTGATTCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGG
CCATACACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTT
GGCGAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTT
CTTCCGCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTT
ATGTGTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTT
GGTCGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAAT
CCATTTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATA
AATCGGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCC
TACATCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGT
TCCTCCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATT
GGAGCCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTT
TTTGGTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTC
ATGCGTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGTAT
AGTCGCCACCAACGGTCTGGGCGA

>KARJAT-4---4M8_CONTIG

GACTCTTGGTAGAGTCTATTAGACCGCGACCGCTCCCCAAGGGTCGT
GCTTGGAACATTTTGCATGGCGCACGACTTTCCCTTTCTTGTTTTGGGC
TTTTGCTATGCCGTAGAGGAAGTCTGCTTTTCTGGGTCTTAGGAGTAAG
TGGATTCAGCCTGGCTCCATTTCTGGTAGTACAGAGAGCATCGTGCTG
AACGGCTTCTGTGGAAGTTCTCCCTTGATCATTTATCAGTCGTTTCTAG
TTGATTAGTTTTGCCTGATGTTGGGTTAAGCGGGTTCTCCTGTGCGTTG
ATCCGGTGATTTTCCCTTTTCCTTTTACGCTCCCTGTGCGTTCCTCGGT
TTGGATTGTAAAGGTTGTTTTGGCGTGCATGCGTAGAAGAAGCTGGCT
CTTCTTCCGGAGACTGTTCCCCTAAGTCTACCGAGCCCTTCCCCTGCA
TAAATGCTCGTTCCGTGCTCTTTTGTTATTCGAGCTAACGTTCTCGCGC
TGGATAGTCGAGGTGCGGACTTTGGTCCGTGGCTTAGTCTCCTTTTCT
GCCCCTCCGCTGTGTTTTTGAAAATGCTTCCCCCTGTTCCGACCATAC
CGTTCTAGTGTATCCCCATCCGATAAACCGGGCCATATTGCTCTCTCTC
TGATTCGAGGCCCCGTTCCAGTGCAAAATTGCGCGGGTGCTTCAGATC
CACTGAGAGGTCCTACACGGTCTTTCGTCTCCCCTCTTCTCTTTCGGG
TGTCCATTTGATGTGTTTGCCCCTGCTTTTGGTTCCTTACTGGCTCCGA
ACTTGTCGGTACTCTTGGTTCGTTCCCCAGAAATTCGGTGCACGACTT
GTCTAACTGGGTTTTGATGGTTCCCTTTGCGGTTGCGGCCGAATTGTG
GATGTATTTTGCCGCTATCGGCCAAGTAACATTTCCATTTCTT

>KARJAT-5---5M8_CONTIG

CCATAGTGATCCAACCTCTCCCCTGATACGTGCGTGTCATTTTCCGTTG
CACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACG
AGGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTA
CTCCTGTACAGAGTACTGGGCAGACGGGATACTGTCTAGGTTTTACCT
TGAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTTGA
TTCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCAT
ACACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGC
GAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTC
CGCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGT
GTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGT
CGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCA
TTTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAAT
CGGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTAC
ATCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCC
TCCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGA
GCCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTT
GGTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATG
CGTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGTATAGT
CGCCACCAACGGTCTGGTCGCATTT

>KARJAT-7---6M8_CONTIG

ATAGAGAAAAACCTCTCCCCTAATCGTGCTGGCTTGGTTCATTGCCCA
CGAATCCTTCCCTCTGCTGATACAGGCGATTCCTTTTCCGTTGCGCAA
GTCGTCTACTTTTTTTCGTTGTTAGTTGATCCACTTACTATTTCGTCTAG
TTCGGAGACCATTTCAGCCTGGATCCTGGGCAAGCTCTTCCTTGTTCT
CTGATCGAGTCGTGGCTAGTCGATTTGTTGTCTTGGATGTTTAGTTTTG
CGGATTCACCCGCGCATTGTTCCGGTGGTTTTCCTACTTCCATCTGCG
CGCGCGGTGCGTTCCTCGGTTTGTCAAGTAAGGTGTTTCGGCGTACAT
CATGGTCAGAACTGGCTGTTGCCTTCTTCCGTACAAACCCCTGCGGAA
GATACCGAGCCCAACCCTCGCATCTTTGCCCGCTCCGTGTCTTTCCTG
TTTGCGTGCGTTTGTTCTAGCTGCATGAGAGTACGAAGCCTTTACTTTT
GCCGGCCCATTGCTCTTCTGTCTTGGGGATCCACTCTCTTTATCAATAA
GATTTCTGCTTCTCCGACCCATCCGATCCAGAACATCCTTTTCAATCCG
ATAATCCGGTCCAATTTGGTTTTCTAATAGGATAAGCCCCGATCCAAGG
TCCACTCTTGGGCTTTCCCGCTTGTCCCTACGTCAGGCGCTACTCGGT
CTTTGGCGTCGAATTTGTTCAGGGGGGTCTTCCTATTAGGAAGGTATG
CTCGCTGCTTTTGGTTCCTTACCAGCATTGGAGCCATTGGCTCTCTCCA
TTTGTTCCCCAGACCCTCGAAGCCAGAGTTTTCTTATTCCTTTCCATGG
ATCCAAAATTTGCGGTTGAGTCCATTACGACATGCTATTTTGCCACATT
CGGTCCCCTTGCGGTTCCGTCGCGGT

>RATNAGIRI-1---8M8_CONTIG

TAGAGCCAAACCTCTCCCCAAATCCTGCTTGCTTCTTTCGTTGCACTCG
CCGTTCCCCATGTAGTGATACGCTATTTCTTGTCCGTAGAGCCAGTCTT
CTCGTCTTGTTCGGAGCCAGTTGATCCACTTTCTTGGGCTTACACTTCA
GTGTACTCCTCAGAGTGGAGTCGGTGGACGAAATTTTGCCTTGGTCAT
TTATCTTTCTCTGCTTGTCCCTCGGTTTTGTCCATTGGTTTTTCTACTGG
ATTCACCGAGAGAGCATTGTTTCGCGGTTTATCCAATGGCCATACACG
CTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCGAACC
CCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCCGCTA
CCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTGTACG
AGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTCGATA
CAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTTCAT
CCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCGGGC
CAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACATCAG
GCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTCCGA
CTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGCCTC
CCGCTTTCCATTCCTTCCTTACACCCAGTTTATCGGTACTCTTGTATTCC
ACCCCAGAAATCCGAAGCAAGAGTTTGCTTTTTGGTCGACATGCTTCCT
TTGCCGTTCAGTCCCCTTGGGGATCGGTCGCGGCCTCATACGCTCTAC
CTAGCGTCTCCACGTCTTCT

>RATNAGIRI-2---9M8_CONTIG

CATAGAGAACCACCTCTCTCCCAAATCCTGCTTGCTTCTTTCGTTGCAC
TCGCCGTTCCCCATGTAGTGATACGCTATTTCTTGTCCGTAGAGCCAG
TCTACTCGTCTTGTTCGGAGCCAGTTGATCCACTTTCTTGGGCTTACAC
TTCAGTGTACTCTGTCAGAGTGGTGGCTAGTCGAAGTTTTGCCTTGGT
CATTTATCTTTCTCTGCTTGTCCCTCGGTTTTGTCCATTGGTTTTTCTAC
TGGATTCACCGAGAGAGCATTGTTTCGCGGTTTATCCAATGGCCATAC
ACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCGA

ACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCCG
CTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTGT
ACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTCG
ATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTT
CATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCG
GGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACAT
CAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTC
CGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGC
CAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTGG
TTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGCG
TTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGATTATTGCC
TCTACCGGACAAGTGTCTGGTCCAT

>RATNAGIRI-3---10M8_CONTIG

ACATTGAGCCAAACCTCTCTCCCGATTCGTGCGTGTCATTTTCCGTTGC
ACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGA
GGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTAC
TCCTGTACAGAGTACTGGGCAGACGGGATACTGTCTAGGTTTTACCTT
GAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTTGAT
TCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCATA
CACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCG
AACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCC
GCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTG
TACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTC
GATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCAT
TTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATC
GGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACA
TCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCT
CCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAG
CCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTG
GTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGC
GTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGATTATTG
CCTCTACCGGACAAGTGTCTGGTCC

>RATNAGIRI-4---11M8_CONTIG

AACTTAGTCCAAAACCTCTCCCCATACCGTCCTTGCATTTTCCGTTGCACACGT
CGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGAGGTCGGCTACC
ATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTACTCCTGTACAGAGTACT
GGGCAGACGGGATACTGTCTAGGTTTTACCTTGAGCTTTTCTCAATCACCCTC
GAGTTTTTCCGTTTGGTTTTCTGCTTGATTCAGAGGGTTCACCAGAGCATTGTT
TCGCGGTTTATCCAATGGCCATACACGCTCACTGTGCGGTCCGCGCAGGGTG
TTGCATGTTGTTTTGGCGAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAA
CGGTTCTTCTTCCGCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCT
CTTATGTGTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTG
GTCGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTT
CATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCGGGCCA
AATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACATCAGGCACTGG
GCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTCCGACTTTGGTGTCGA
GCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGCCAGGCTCCAGCCTTTGATTC
CTTACTCCCAGAAAATCAGAATTTTTTGGTTCCCTTCCAAAGTTCCCCAGTAAG
TCGTTGCTTGTGTTTCCTCATGCGTTTTGTTCCATCCTTTGCCCTTGAGGCCCA
GTCGCGGTCGTATAGTCCCTCCCACGGGCCTGGTCGCATTTCC

>RATNAGIRI-24---12M8_CONTIG
ACATTGAGCCAAAACCTCTCCCCATACCGTCCTTGCATTTTCCGTTGCA
CACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGAG
GTCGGCTACCATTTTTTCTAGTAGCATGGGCACTCACTTACTATGTACT
CCTGTACAGAGTACTGGCAGACGGGATACTGTCTAGGTTTTACCTTG
AGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTTGATT
CAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCATAC
ACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCGA
ACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCCG
CTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTGT
ACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTCG
ATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTT
CATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCG
GGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACAT
CAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTC
CGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGC
CAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTGG
TTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGCG
TTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGTATAGTCG
CCACCAACGGTCTGGTCGCATTTCC

>RATNAGIRI-7---15M8_CONTIG
CATAGTGATCCAACCTCTCTCCCGATTCGTGCGTGTCATTTTCCGTTGC
ACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGA
GGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTAC
TCCTGTACAGAGTACTGGCAGACGGGATACTGTCTAGGTTTTACCTT
GAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTTGAT
TCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCATA
CACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCG
AACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCC
GCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTG
TACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTC
GATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCAT
TTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATC
GGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACA
TCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCT
CCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAG
CCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTG
GTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGC
GTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGATTATTG
CCTCTACCGGACAAGTGTCTGGTCC

>RATNAGIRI-8---16M8_CONTIG
CAAATTAGTCCAAAACCTCTCCCCATACCGTCCTTGCATTTTCCGTTGC
ACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGACGA
GGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATGTAC
TCCTGTACAGAGTACTGGCAGACGGGATACTGTCTAGGTTTTACCTT
GAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTTGAT
TCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCCATA
CACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCG
AACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCC
GCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTG
TACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTC
GATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCAT

TTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATC
GGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACA
TCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCT
CCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAG
CCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTTTG
GTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCATGC
GTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGATTATTG
CCTCTACCGGACAAGTGTCTGGTCC

>PANVEL-1--- 17M8_CONTIG
TAGAGCCAAAACCTCTCCCCAAATCCTGCTGGCATTCTTCGTTGCACTC
GCCGTTCCCCATGTAGTGATACGCTATTTCTTGTCCGTAGAGCCAGTC
TTCTCGTCTTGTTCGGAGCCAGTTGATCCACTTTCTTGGGCTTACACTT
CAGTGTACATTCCTGAATGGGTACTGGGTAGGCGTTACCTTGATCTTG
GATCTTTCAGTTCTTCTGTATTACCCTCGTTTAATGTTTCCTTTTGTGGT
TCCACCTGAATCAGTGAGACGGAGATATTTCTCCGTTTACCAAATGCG
CTCACTGTGCGCTCCACGGTCAGAACAGGGTGTTGCATTTTTTGGTAC
GAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTC
CGCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGT
GTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGT
CGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCA
TTTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAAT
CGGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTAC
ATCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCC
TCCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGA
GCCTCCCGCTTTCCATTCCTTCCTTACACCCAGTTTATCGGTACTCTTG
TATTCCACCCCAGAAATCCGAAGCAAGAGTTTGCTTTTTGGTCGACATG
CTTCCTTTGCCGTTCAGTCCCCTTGGGGATCGGTCGCGGCCTCATACG
GACTCGTCCCAGTGCCTGGTCGTCT

>PANVEL-2---18M8_CONTIG
AGAGACCAAACCTCTCCCCAAATCCTGCTTGCTTCTTTCGTTGCACTCG
CCGTTCCCCATGTAGTGATACGCTATTTCTTGTCCGTAGAGCCAGTCTT
CTCGTCTTGTTCGGAGCCAGTTGATCCACTTTCTTGGGCTTACACTTCA
GTGTACTCCTCAGAGTGGAGTCGGTGGACGAAATTTTGCCTTGGTCAT
TTATCTTTCTCTGCTTGTCCCTCGGTTTTGTCCATTGGTTTTTCTACTGG
ATTCACCGAGAGAGCATTGTTTCGCGGTTTATCCAATGGCCATACACG
CTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGGCGAACC
CCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTTCCGCTA
CCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATGTGTACG
AGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGGTCGATA
CAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCCATTTCAT
CCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAATCGGGC
CAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTACATCAG
GCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTCCTCCGA
CTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGGAGCCTC
CCGCTTTCCATTCCTTCCTTACACCCAGTTTATCGGTACTCTTGTATTCC
ACCCCAGAAATCCGAAGCAAGAGTTTTCTAATTGGTTTTGTTGCGTCCT
TTGCGGTTGTGTCCATTCAGTGACCCTTGAGGGCCACTCGCGGTCTTG
TAGCCTCTCCCTTGCGTCTGGACG

>PALGHAR-2---20M8_CONTIG
AACCAGTTGAGTCCAACCTCTCCCCTGATCCTGTCGTGTCATTTTCCGT
TGCACACGTCGTGCCCTAAATTTCTTGTTATTGGCTCCTTCTATTCCGA
CGAGGTCGGCTACCATTTTTTCCAGTAGCATGGGCATTCACTTACTATG
TACTCCTGTACAGAGTACTGGGCAGACGGGATACTGTCTAGGTTTTAC
CTTGAGCTTTTCTCAATCACCCTCGAGTTTTTCCGTTTGGTTTTCTGCTT
GATTCAGAGGGTTCACCAGAGCATTGTTTCGCGGTTTATCCAATGGCC
ATACACGCTCACTGTGCGGTCCGCGCAGGGTGTTGCATGTTGTTTTGG
CGAACCCCTGGGGAAGAACTGGCGCCTCTTCCGTAAACGGTTCTTCTT
CCGCTACCGTTCCCATCCATTGCGTTTATGCTCGTTCCGTGCTCTTATG
TGTACGAGCTCCTGTTCTTGCGCAGGCCAGTCGCTCTTTTGACTTTGG
TCGATACAGTCTCTTCTTCTTTGTACTGCCTCTGTGACACATGCAATCC
ATTTCATCCTTTTCCGTCCAAACCCATCAAGAATTTCCCTTTCCGATAAA
TCGGGCCAAATTCAAAGGTCTACTCATAGGATGCCCAGCGTTCCCCTA
CATCAGGCACTGGGCTTTTGCTAACGATCCAATGAGATCAGGGAGTTC
CTCCGACTTTGGTGTCGAGCTCGTTCCTTTGTGTTTCTACCAGAATTGG
AGCCAGGCTCCAGCCTTTGATTCCTTACTCCCAGAAAATCAGAATTTTT
TGGTTCCCTTCCAAAGTTCCCCAGTAAGTCGTTGCTTGTGTTTCCTCAT
GCGTTTTGTTCCATCCTTTGCCCTTGAGGCCCAGTCGCGGTCGATTAT
TGCCTCTACCGGACAAGTGTCTGG

- **Sequences obtained from primer *ndhA***
  >KARJAT-1---1M9_Contig1

TACTAGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTC
TGAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAG
TAAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTT
GTAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAG
TAAACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATT
AGTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGA
AAGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAA
CACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGAT
ATTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAAT
GAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTC
ACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTT
TTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGG
AATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCC
ATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATT
AATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAA
GAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTT
TTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAA
ATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGA
AATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAA
CTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATGTC

>KARJAT-2---2M9_Contig1

CAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGAAAT
TCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAAGTT
AAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTAGTA
AAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAAACAT
AAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGTCAT
CATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAGTTG
TAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACACTAA
AGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATTTTT
CCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGAGCA
AGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCACTTG
TTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTTCTT
TTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAATG
CAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCATATT
CATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTAATT
GTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAGAAA
AAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTTTAT
TATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAATCG
ATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAATG
AAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACTAA
GGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTCTCGATATC

>KARJAT-3---3M9_Contig1

TAGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGA
AATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAA
GTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTA
GTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAA
ACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGT
CATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAG
TTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACAC
TAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATT
TTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGA
GCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCAC
TTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTT
CTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAA
TGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCAT
ATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTA
ATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAG
AAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTT
TATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAAT
CGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAA
TGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACT
AAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGTATGTCG
T

>KARJAT-4---4M9_Contig1

GTACTAGACATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCT
CTGAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTA
GTAAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATT
TGTAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAA
GTAAACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGA
TTAGTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAG
AAAGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAA
ACACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGA
TATTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAA
TGAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATT
CACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTT
TTTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATG
GAATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATC
CATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTA
TTAATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACA
AAGAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGT
TTTTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTG
AAATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAAC
GAAATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGA
AACTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGCGATG
CT

>KARJAT-5---5M9_Contig1

GTACTAGACAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCC
TCTGAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTT
AGTAAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAAT
TTGTAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAA
GTAAACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGA
TTAGTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAG
AAAGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAA
ACACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGA
TATTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAA
TGAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATT
CACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTT
TTTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATG
GAATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATC
CATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTA
TTAATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACA
AAGAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGT
TTTTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTG
AAATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAAC
GAAATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGA
AACTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACCTGTGAT
TTTATCCGT

>KARJAT-7---6M9_Contig1

ACTAGACATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCT
GAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGT
AAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTG
TAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTA
AACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTA
GTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAA
AGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAAC
ACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATA
TTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAAT
GAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTC
ACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTT
TTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGG
AATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCC
ATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATT
AATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAA
GAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTT
TTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAA
ATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGA
AATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAA
CTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGCGATGCT
ATT

>RATNAGIRI-1---8M9_Contig1

GTGTGCTTAGACAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTT
TTCCTCTGAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCA
GGTTAGTAAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTA
TTAATTTGTAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAA
TGGAAGTAAACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTT
TTTGATTAGTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGA
TACAGAAAGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGAT
TAGAAACACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCA
TTAGATATTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGT
GGAAATGAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTC
CCATTCACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATT
CCTTTTTTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAG
ATATGGAATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGT
TTATCCATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTC
GTTTATTAATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCT
GAACAAAGAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGA
AGCGTTTTTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGA
CGTTGAAATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATA
ATAACGAAATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCG
TATGAAACTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTTC
GATGTCTTC

>RATNAGIRI-2---9M9_Contig1
AGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGAA
ATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAAG
TTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTAG
TAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAAAC
ATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGTCA
TCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAGTT
GTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACACTA
AAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATTTTT
CCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGAGCA
AGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCACTTG
TTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTTCTT
TTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAATG
CAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCATATT
CATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTAATT
GTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAGAAA
AAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTTTAT
TATTCTAGCAGACAGAATTCCTTTGGTCTAATTAAGGACGTTGAAATCG
ATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAATG
AAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACTAA
GGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATGTTATCC

>RATNAGIRI-24---12M9_Contig1
TAGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGA
AATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAA
GTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTA
GTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAA
ACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGT
CATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAG
TTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACAC
TAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATT
TTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGA
GCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCAC
TTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTT
CTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAA
TGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCAT
ATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTA
ATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAG
AAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTT
TATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAAT
CGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAA
TGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACT
AAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATGT

>RATNAGIRI-7---15M9_Contig1

AATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTCCTCTGAAATT
CATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAAGTTA
AACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTAGTAA
AAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAAACATA
AGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTGATTAGTCATC
ATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAGTTGTA
ATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACACTAAAG
TACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATTTTTCCT
CATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGAGCAAGT
AGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCACTTGTTA
AGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTTCTTTTT
AAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAATGCAA
TACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCATATTCAT
ACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTAATTGTT
ATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAGAAAAA
GTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTTTATTA
TTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAATCGAT
TTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAATGAA
ACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACTAAGG
TTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATGTT

>PANVEL-1---17M9_Contig1

TAGTGTACTAGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTT
TCCTCTGAAATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAG
GTTAGTAAGTTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATT
AATTTGTAGTAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATG
GAAGTAAACATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTT
TGATTAGTCATCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATA
CAGAAAGTTGTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTA
GAAACACTAAAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATT
AGATATTTTTCCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTG
GAAATGAGCAAGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCC
CATTCACTTGTTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTC
CTTTTTTCTTTTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGA
TATGGAATGCAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTT
TATCCATATTCATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCG
TTTATTAATTGTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGA
ACAAAGAAAAAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAG
CGTTTTTTATTATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACG
TTGAAATCGATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAAT
AACGAAATGAAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTA
TGAAACTAAGGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTG
ATGTC

>PANVEL-2---18M9_Contig1

AGCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGAA
ATTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAAG
TTAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTAG
TAAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAAAC
ATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGTCA
TCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAGTT
GTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACACTA
AAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATTTTT
CCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGAGCA
AGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCACTTG
TTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTTCTT
TTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAATG
CAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCATATT
CATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTAATT
GTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAGAAA
AAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTTTAT
TATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAATCG
ATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAATG
AAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACTAA
GGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATG

>PALGHAR-2---20M9_Contig1

GCAATATCTCTACGTGCGATTCGTTAAAATAGGATCTTTTTCCTCTGAAA
TTCATTGAATATTTATCCTCCTTTTCTTATTCTATATTCAGGTTAGTAAGT
TAAACTAGATAGCTATATGAGTGAAACAAAACTGCTTATTAATTTGTAGT
AAAAAGAAAAAATCTCATTTCCTATGTACAAGAAAAAATGGAAGTAAAC
ATAAGCGGTGTAAATTCTTTACCCCAAGGTTGAGATTTTTTGATTAGTCA
TCATATCTTGAAGCGGGCAATAATAAGGGATTCCCGATACAGAAAGTT
GTAATCATAAGGGAATCCATCAAAAGTTAGTGAGGGATTAGAAACACTA
AAGTACATAAAGGATTAGTAATGAGGGAATCTAAAGCATTAGATATTTTT
CCTCATAAAAGGAATCATAATAAGGACTTGAAATTGGTGGAAATGAGCA
AGTAGTACTTTCTTCGGATTCCGATCCAGAGTATGTTCCCATTCACTTG
TTAAGGAAATGGCTATCAAGAACGAATTAACCCTTTATTCCTTTTTTCTT
TTTAAATACCCCTCGGGGAAAGAAGAATAGGACAAAAGATATGGAATG
CAATACAAAAAAAAAGATCTTTATTTATTCTTTCCGTCGTTTATCCATATT
CATACAGAATTCTTCATGAACTAATACCCAACTCTTTTCGTTTATTAATT
GTTATAACGAGTGTTTTATTCCAAGATTAAGTTATTGCTGAACAAAGAAA
AAGTACCATTATATTATAAAGGATGAGATGAATTCCGAAGCGTTTTTTAT
TATTCTAGCAGACAGAATTCCTTTGGTCTAATTTAGGACGTTGAAATCG
ATTTTATCTTAGATAATTCTAACTACGCCCAAGAAGATAATAACGAAATG
AAACAGTACTTTCCTTTTTTCTGATCATAGAGGAGCCGTATGAAACTAA
GGTTTCATGTACGGTTTTGGAATAGCGGTGGGAACTGTGATGTT

- **Sequences obtained from primer *matK-1m-matK3RIM***
  >RATNAGIRI-7---15M10_Contig1

  TACCCAGTCCATCTGGAAATCTTGGTTCATATGTGCGTTGTTCAGGTAT
  TTTGAACAGTGAATCGAGCACTTGTTGTGGTGTCTTAGATATCGAGCTT
  GTCGATCTTGTAGATGATGTTGTTGAGTGGGGCTCCTCACATGCAACC
  TTCAAGTAGCTGTTGAGCAAGGATAACATAGAAACAGTCACATAAGTGC
  AAAAATAGAGTTGAACTTACAGATAGCAACTAACTAGTTATACAATCTG
  CGCAATTGAAGTAAATAAACAAAGTATATGCACCTACACTTAGCATAAA
  ATTTACAGGTTAAACTGGTAAGGTCAAAACATTGCAAGAGATACAAGCC
  TTGACTGCAAAAAATAAGCTCACATAATTTGTTTGTCCATTATAAAGACC
  ATATTTTAGAATTCCTAGGAAAGTTGCAACTTTAGTAATTAACACCATGC
  AATCTGATGTATCTAGACAACTTAAGTTTCTATCCGATCGACTGATAATA
  TGTCCCATGATCAAAGATTTTCTATTGTGCGTCTCGTAAACACAAAAGT
  ACTGTACG

  >PANVEL-2---18M10_Contig1

  TATACCAGTGCTCCCACTGAAATCTTGGTTCAACTCCTTCAATACCGGA
  TCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTTTCTCAACTACTAC
  TCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATTTTTATTTTGAAA
  AAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTCTTATGTATCAGA
  ATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTCTTGCTTGCGATT
  AACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTTCTAGGAAGATG
  GAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAAAACCATATGGTT
  CGTTATGGATCCCCTTATGCATTATGTTCGATATCAAGGAAAGGCAATT
  CTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAATGGAAATGTTACC
  TTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGACTCAACCGCAAAG
  GATCCATCTAAACCAATTAGAAAACTCTTGCTTCGATTTTCTGGGGTAC
  TTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAATCAAATGCTGGA
  GAATTCATTTCTAATAGATACTCAAATGAAAAAATTCGATACCAAAGTCC
  CTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCAATTTTGTACTGG
  ATCGGGGCATCCTATTAGTAAACCAATTTGGACCGATTTATCGGATTGG
  GATATTCTTGATCGGTTTGGTCGGATATGTAGAAATCTTTTTCATTATCA
  CAGTGGATCTTCAAAAAAGAAGACTTTGTATCGACTAAAGTATATACTT
  CGACTTTCATGCGCTAGAACTTTAGCTCGTAAACACAAAGTTACTGTAC
  GAAA

  >PALGHAR-2---20M10_Contig1

  CAAATCTGTATACTTAACCAGTCCCACTGGAAATCTTGGTTCAACTCCT
  TCAATACCGGATCCAAGATGTTCCATCCTTGCATTTATTGCGATTCTTTC
  TCAACTACTACTCGAATTGGAATAGTTTTATTACTTCAATGAAATCCATT
  TTTATTTTGAAAAAAGAAAATAAAAGACTATTTCGGTTCCTATATAACTC
  TTATGTATCAGAATATGAATTTTTCTTGTTGTTTCTTCGTAAACAATCTTC
  TTGCTTGCGATTAACTTCTTCCGGAACCTTTCTGGAACGAATCATCTTTT
  CTAGGAAGATGGAACATTTTGGGTTAATGTACCCTGCTTTTTTTCGGAA
  AACCATATGGTTCGTTATGGATCCCCTTATGCATTATGTTCGATATCAA
  GGAAAGGCAATTCTTGCATCAAAAGGAACTCTTCTTTTGAAGAAGAAAT
  GGAAATGTTACCTTGTCCGTTTGTGGCAATATTCTTTCTCTTTTTGGACT
  CAACCGCAAAGGATCCATCTAAACCAATTAGAAAACTCTTGCTTCGATT
  TTCTGGGGTACTTTTCAAGTGTACCAATAAATTCTTTGTTAGTAAGGAAT
  CAAATGCTGGAGAATTCATTTCTAAAAGATACTCAAATGAAAAAATTCG

AAACAAAAGTCCCTGTTACTCCTCTCATTGGATCTTTAGCAAAAGCCCA
ATTTTGTACTGCATCGGGGCATCATATTATAAGAACCGTTGGACTCGAT
TTATTTTCATTGTTATATACAAAAACGGATTGCTCTGAACTGATAAAACA
ATACCTCTTCGAGTAGAAGTTGAGCAAGAACAAAAAAATGCAACGATG
GAGCATCTAGGATCCGGTCTTGGCGTATGGGACAAAAACATAAAAAAC
AATGCCTCGTGTNNNNNNNNNNR

CHAPTER I

INTRODUCTION

Rice (*Oryza sativa L.*), is one of the most widely cultivated crops in the world and a staple food for one–third of the world's population. It belongs to family Graminae and subfamily Oryzoidea with a chromosomal compliment of 2n=24. *Oryza sativa* was domesticated from the wild grass *Oryza rufipogon* roughly ten-fourteen thousand years ago. Since then rice has been considered as major crop and food with all its ancient, asthetic, cultural and economical aspects. Being one of the world's largest cereal crop, occupies almost one-fifth of total area covered under cereals. In India, rice is one of the most important staple food for more than 2/3 rd of its population and is a primary source of food for more than 3 billion people of the world (Ma *et.al.* 2007).

In India, rice has unique position in the economy. Rice contributes more than 40 % of the country's total food grain production. Total production of rice during 2019-20 estimated at record 117.47 million tonnes. Rice is not only the staple food but also source of income and livelihood for millions of people around the globe. The 57th session of the United Nations General Assembly noted that rice is the staple food of more than half the world's population affirmed the need to heighten the awareness of the role of rice in alleviating poverty and malnutrition and reaffirmed the need to focus world attention on the role rice can play in providing food security and eradicating poverty and declared the year 2004, The International Year of Rice (FAO, 2002). Right from the rice seed to its stock has a great economic value in agriculture and food industries.

Rice is high in minerals such as calcium and iron and rich in vitamins namely niacin, Vitamin D, thiamine and riboflavin. Rice protein is biologically the richest and its digestibility is very high (88%) (Goplan *et.al.* 2007), provides almost 50-80 % of daily calorie intake amongst the poor class of the society. The chemical composition of rice grain varies considerably depending on the genetic factor of the rice varieties and environmental conditions such as location, season, fertilizer application and post-harvest operation. It is said that rice is primary a high energy calorie food. The protein of rice contains glutenin, also known as Oryzenin. The mineral content in rice comes from the pericarp and germ.

Many rice varieties are cultivated all around the world in wide and varied geographical locations. Two major subspecies of cultivated rice, indica and japonica have been brought under domestication since long time. To meet the demands of population many varieties with high yields and tolerance to abiotic and biotic stresses are developed, recorded, released and notified every year in India. Rice is highly diversified crop and assessment of genetic diversity is essential for further maintenance of varieties and pre-breeding programs. With recent advancements in technology and breeding program focusing on the genetic trait of the crop, development of the DNA markers has proved to be beneficial to a wider extent. Establishment of techniques and development of molecular markers has facilitated transfer of genes that confirm resistance to abiotic and biotic stresses. Tracking of genes responsible for a particular character/trait, genes linked for resistance and tolerance, mutation and variation have been located and studied using molecular markers. Different rice varieties of distinct genetic background are a good promise for the future of rice crop improvement programmes as genetic diversity helps in estimating and establishing of genetic relationship in germplasm collection, identifying diverse parental combinations to create segregating progenies with maximum genetic variability and superior recombination for further selection and introgression of desirable genes from diverse germplasm (Thompson *et.al.* 1996; Islam *et.al.* 2012). Molecular breeding offer the opportunity to increase the speed and efficiency of plants breeding (Whitford *et.al.*, 2010). Molecular markers can reveal abundant difference among genotypes at the DNA level, providing a more direct, reliable and efficient tool for germplasm characterization, conservation and management and untouched by environmental influence. Molecular markers are promising and effective tools for measuring genetic diversity in germplasm collection and elucidating their evolutionary relationships. Using Molecular marker technology, it is now feasible to analyze the quantitative traits and identify the chromosomal regions associating with such characters known as quantitative trait loci (QTLs) (Choudhary *et.al.,* 2008).

In rice crop seed is the basic input. Maintenance of seed purity is an important skill demanding aspect in seed production and multiplication. DNA fingerprinting allows identifications of plant genotypes with high precision and stability as it is not influenced by environment, epistatic interactions and pleiotrophic effects. During initial stages of breeding work on rice the *'Dee-Gee-Woo-Gene'* was incorporated in the rice cultivars and over the period of time there has been development of many dwarf varieties by the introgression of particular gene. Also a major parameter in rice breeding involves the adoption and cultivation of short bold varieties. The nature of the seed depends mainly also on the geographic locations from where the cultivars have been found. Unambiguous identifications of elite crops varieties and hybrids are essentials for their protection and prevention of unauthorized commercial use. In India, this is highly relevant especially in rice because the hybrid seed production and marketing of public sector bred hybrids is largely taken up by the private sector. Molecular markers, being based on DNA sequence variation, provide an unbiased means of identifying crop varieties (Nandakumar N. *et.al.,* 2004). DNA fingerprinting of plants has become an invaluable tool in forensic, scientific and industrial laboratories all over the world. PCR has become the basic requisite having its wide applications in the molecular biology field. Among various PCR based markers, SSR markers are more popular because they are highly informative, mostly monolocus, codominant, easily analysed, highly reproducible and cost effective (Gracia *et.al.,* 2004).SSR markers are able to detect high level of allelic diversity and they have been extensively used to identify genetic variation among rice subspecies (Ni and Mackill;2002). A more advance step of DNA fingerprinting comes the sequence generation of the amplicons which are called as DNA Barcodes and the process called as DNA Barcoding. In 2003, a new identification system, DNA barcoding, was developed by researchers at the University of Guelph (Canada). This approach is based on the analysis of the variability within a standard region of the genome called “DNA Barcode” (Herbert *et.al.,* 2003). Global DNA barcoding was initially regarded as a big science programme (Gregory, 2005) and even as the renaissance of taxonomy (Miller, 2007).

The major highlight of DNA barcoding is that it targets and detects the genes present in the plastids which are inherited from the maternal parent of the crop and are carried forward over many generations and they also exhibit variations at the gene level. The entire chloroplast genome of the two monocot rice (*Oryza sativa*) has been sequenced and comprises 134525 bp. The presence of DNA in the chloroplast was first identified in 1962. The size of its DNA is generally between 120 and 155 kb. Several evidences have been noted and confirmed that chloroplast DNA contains 45 genes coding for RNA and 27 genes coding for proteins. These proteins are mainly involved in the chloroplast gene expression. In the year 2005, it was found that *Oryza sativa indica* contains 134496 bp and that of *Oryza sativa japonica* contains 134551 bp genes. The single circular DNA molecule comprises with a quadripartite structure that includes two copies of an IR region that separate large and small single copy regions (LSC and SSC regions). At higher taxonomic levels (family levels), protein-coding regions and conserved sequences of the chloroplast genome can be used for the phylogenetic relationships between closely related taxa and for improving our understanding of the evolution of the plant species. A key application of the chloroplast genome in agriculture is the identification of commercial importance and the determination of their purity. DNA barcodes derived from the chloroplast genome can be used to identify varieties and in the conservation of breeding resources. The gene regions are chosen because they have less intraspecies (within species) variations than interspecies (between species) variations, which are known as the “Barcoding Gap”.

An ideal barcode should be such that it should be routinely retrievable with a single primer pair and should be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces (Vijayan *et.al.*, 2010). It can be concluded that the initial goal of the DNA barcoding process is to construct online libraries of barcode sequences for all known species that can serve as a standard to which DNA barcodes of any identified or unidentified specimens can be matched. DNA barcoding thus can provide a cost effective and efficient tool for the taxonomists, conservationists, and other ones who need the identification of species.

The major objective to study the DNA barcodes is to develop a "molecular tag" of the particular variety and to analyze the variations seen in the different sequences of different varieties. In India, Maharashtra being one of the major rice producing state in which Konkan Region is the major habitat of cultivation which lies near to coastline of the state. Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli has played a major role in developing new rice varieties and undertakes rice breeding and research programmes at different research stations located at Karjat, Panvel, Shirgaon. These varieties have been studied for genetic diversity analysis and screened for biotic and abiotic stresses. The present research entitled "DNA barcoding for identification of Rice (*Oryza sativa* L.) varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli" was undertaken with the following objectives:

i) To standardize comparative rice genotype profiles using molecular markers.
ii) DNA barcoding of different grain type variety of rice.

CHAPTER II
REVIEW OF LITERATURE

Recent advancements in the molecular biology and synthetic biology have led to study the genes and genomes at a very significant level. The advancement in finding out the gene sequence using sequencing methods and determine the activity or trait of the gene have proved beneficial in the breeding program at a molecular level for gene manipulations. Also, determining the DNA sequences have helped biologists, taxonomists and breeders to classify and add a molecular tag to the varieties of crop and construct phylogeny of the same. DNA barcoding is one such recent approach in the world of molecular biology. The present study aims at DNA barcoding of rice varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli and sequencing of the same.

2.1. DNA Barcoding

DNA barcoding is the technique which uses short small DNA segments from the genome to identify species and classify them based on the 'molecular tag' *viz.*' DNA BARCODE' which is unique for each species. Following are the research articles studied to get a introductory idea of the DNA barcoding concept.

Kress and Erickson (2007) proposed that a useful DNA barcode requires sufficient sequence variation to distinguish between species and ease of application across a broad range of taxa. Discovery of a DNA barcode for land plants has been limited by intrinsically lower rates of sequence evolution in plant genomes than that observed in animals. The principal findings in this study were a global plant DNA barcode system was evaluated by comparing universal application and degree of sequence divergence for nine putative barcode loci, including coding and non-coding regions, singly and in pairs across a phylogenetically diverse set of 48 genera (two species per genus). No single locus could discriminate among species in a pair in more than 79% of genera, whereas discrimination increased to nearly 88% when the non-coding *trnH-psbA* spacer was paired with one of three coding loci, including *rbcL*. In silico trials were also conducted in which DNA sequences from GenBank were used to further evaluate the discriminatory power of a subset of these loci. The study concluded that a combination of the non-coding *trnH-psbA* spacer region and a portion of the coding *rbcL* gene is recommended as a two-locus global land plant barcode that provides the necessary universality and species discrimination.

Chase *et. al.* (2007) proposed that use of three regions of plastid DNA as a standard protocol for barcoding all land plants. They reviewed and proposed that the low levels of variation in plastid DNA make three regions necessary; there are no plastid regions, codingor non-coding, that evolve as rapidly as mitochondrial DNA generally does in animals and also outlined two, three-region options, (1) *rpoC1*, *rpoB* and *matK* or (2) *rpoC1*, *matK* and *psbA*-*trnH* as viable markers for landplant barcoding.

Fazekas *et. al.* (2008) stated that a universal barcode system for land plants would be a valuable resource, with potential utility in fields as diverse as ecology, floristics, law enforcement and industry. In this study composition of eight candidate plant barcoding regions from the plastome and one from the mitochondrial genome for how well they discriminated the monophyly of 92 species in 32 diverse genera of land plants (N = 251 samples). The plastid markers comprised portions of five coding (*rpoB, rpoC1, rbcL, matK* and 23S rDNA) and three non-coding (*trnH-psbA, atpF–atpH, and psbK–psbI*) loci. Single locus resolution ranged from 7% (23S rDNA) to 59% (*trnH-psbA*) of species with well-supported monophyly. Sequence recovery rates were related primarily to amplification success (85–100% for plastid loci), with *matK* requiring the greatest effort to achieve reasonable recovery (88% using 10 primer pairs). Concluding the study it was recommended that resolution to the contentious debate on plant barcoding should therefore involve increased attention to practical issues related to the ease of sequence recovery, global alignability, and marker redundancy in multilocus plant DNA barcoding systems.

Lahaye *et. al.* (2007) undertook the intensive field collections in two biodiversity hotspots (Mesoamerica and southern Africa). Using>1,600 samples, comparison of eight potential barcodes was done and assessed to what extent a "DNA barcoding gap" is present between intra- and interspecific variations, using multiple accessions per species. Given its adequate rate of variation, easy amplification, and alignment, they identified a portion of the plastid *matK* gene as a universal DNA barcode for flowering plants. In addition, analyized>1,000 species of Mesoamerican orchids, DNA barcoding with *matK* alone revealed cryptic species and proved useful in identifying species listed in Convention on International Trade of Endangered Species (CITES) appendixes. They amplified and sequenced *accD*, *rpoC1*, *rpoB*, *ndhJ*, *matK*, and *ycf5*, following guidelines from the plant working group. For *matK*, additional primers390F and 1326R were used. Primers *trnH*f and *psbA*3f were used for *trnH-psbA.* DNA sequences were aligned in PAUP4b10. For genetic and phylogenetic analysis, inter- and intraspecific genetic divergences were calculated .Pairwise distances were calculated with PAUP4b10 and the best-fitting model as given by applying MODELTEST 3.7 was used. Wilcoxon signed rank tests were performed to compare intra- and interspecific variability for every pairs of barcodes. They evaluated DNA barcoding gaps by comparing the distribution of intra- versus interspecific divergences. To evaluate whether species were recovered as monophyletic with each barcode, they used standard phylogenetic techniques: MP, maximum likelihood (ML), neighbor joining (NJ), and UPGMA with PAUP4b10 .Bayesian statistical inferences (BI) were performed with MrBayers software, Version 3.1.2. The parsimony analysis of the large *matK* matrix of Mesoamerican orchids was performed by using the parsimony ratchet method. They identified genetic clusters by coalescence analyses.

Vijayan and Tsou (2010) mentioned that DNA barcoding is the process of identification of species based on nucleotide diversity of short DNA segments. It is well established in animals with the introduction of cytochrome *c* oxidase subunit 1 (COI) as a standard barcode. The article provided an overview of the technical details and merits and demerits of the loci as plant barcodes given by the Consortium for the Barcode of Life–Plant Working Group based on the relative efficacy testing has recently identified a few loci as potential barcode candidates and from them a two-locus standard barcode (*rbcL* + *matK*) recommending them for initiating the barcoding process of plant species.

Hollingsworth *et. al.* (2011) suggested that the main aim of DNA barcoding is to establish a shared community resource of DNA sequences that can be used for organismal identification and taxonomic clarification. In this paper, they reviewed the process of selecting and refining a plant barcode; evaluate the factors which influence the discriminatory power of the approach; describe some early applications of plant barcoding and summarise major emerging projects; and outlined tool development that would be necessary for plant DNA barcoding to advance. The paper focused on spelling out the challenges and difficulties for plant barcoding.

Vere *et.al.* (2015) put forth that DNA barcoding uses specific regions of DNA in order to identify species. The protocols mentioned in this paper described the whole DNA barcoding process, from the collection of plant material from the wild or from the herbarium how to extract and amplify the DNA, and how to check the quality of the data after sequencing.

Xiwen Li *et.al.* (2015) studied and put forth that DNA barcoding is currently a widely used and effective tool that enables rapid and accurate identification of plant species; however, none of the available loci work across all species. They reviewed the development of candidate barcodes and discuss the feasibility of using the chloroplast genome as a super-barcode. They advocated a new approach for DNA barcoding that, for selected groups of taxa, combines the best use of single-locus barcodes and super-barcodes for efficient plant identification.

2.2. Taxonomical Approach

In biological sciences, the basic principle step in studying a living entity is its identification which is called as Taxonomical Classification. With the advancement in biological sciences, the perspective of taxonomical classification right from Linnaeus's classification has changed its course reaching to the genetic and molecular level of the living entity. One such approach developed and is in recent trend is DNA barcoding which gives a complete data set of the designed living entity (plant, animal, bacteria, fungi) from a single specimen irrespective to morphological or life stage characters. Focusing on the taxonomical benefits from DNA barcoding which will benefit the course of this research studied are mentioned below.

Hajibabaei *et. al.*(2007) put forth in the article that DNA barcoding aims to provide an efficient method for species-level identifications and, as such, will contribute powerfully to taxonomic and biodiversity research. They compared the goals and methods of DNA barcoding with those of molecular phylogenetics and population genetics, and suggested that DNA barcoding can complement current research in these areas by providing background information that will be helpful in the selection of taxa for further analyses. Summarizing the article they mentioned that DNA barcoding is poised to contribute to taxonomic research and to population genetics and phylogenetics.

Hebert (2005) explained that if DNA barcoding proceeds on a large scale, it will generate important by-products for the scientific community and all DNA extracts produced during the barcode analysis of vouchered specimens would be stored, allowing future efforts to examine patterns of sequence diversity in other gene regions, and the collection programs instigated by DNA barcoding will expand the specimens available for morphological analysis. The barcode initiative would also create a Web-based system delivering not just automated identifications, but also providing a portal to biological information for all species included in the registry.

Valentini *et.al.* (2008) put forth that DNA barcoding – taxon identification using a standardized DNA region – has received much attention recently, and is being further developed through an international initiative. They reviewed the new avenues offered to ecologists by DNA barcoding, particularly in the context of new sequencing technologies. They highlighted the advantage of the DNA barcoding approach is that the basic data the sequences are not prone to subjectivity and can be reanalyzed in the future in accordance with improvements in taxonomic knowledge.

Ajmal *et.al.* (2014) in the research article mentioned that the discipline taxonomy (the science of naming and classifying organisms, the original bioinformatics and a basis for all biology) is fundamentally important in ensuring the quality of lifeof future human generation on the earth; yet over the past few decades, the teaching and research funding in taxonomy have declined because of its classical way of practice which lead the disciplinemany a times to a subject of opinion, and this ultimately gave birth to several problems and challenges,and therefore the taxonomist became an endangered race in the era of genomics. They suggested the sequences to be useful in DNA barcoding include cytoplasmic mitochondrial DNA (e.g. cox1) and chloroplast DNA (e.g. *rbcL, trnL-F, matK, ndhF*, and *atpB rbcL*), and nuclear DNA (*ITS*, and housekeeping genes e.g. *gapdh*). Highlighting the concept of the plant DNA barcoding they stated that it is now transitioning the epitome of species identification; and thus, ultimately helping in the molecularization of taxonomy. The ‘DNA barcodes’ show promise in providing a practical, standardized, species-level identification tool that can be used for biodiversity assessment, life history and ecological studies, forensic analysis, and many more.

Guo *et.al.* (2015) worked on diatoms which form an enormous group of photoautotrophic micro-eukaryotes and play a crucial role in marine ecology. In this study, they evaluated typical genes to determine whether they were effective at different levels of diatom clustering analysis to assess the potential of these regions for barcoding taxa. The test genes included nuclear rRNA genes (the nuclear small-subunit rRNA gene and the 5.8S rRNA gene+*ITS*-2), a mitochondrial gene (cytochrome c-oxidase subunit 1,COI), a chloroplast gene [ribulose-1,5-biphosphate carboxylase / oxygenase large subunit (*rbcL*)] and the universal plastid amplicon (UPA). Calculated genetic divergence was highest for the internal transcribed spacer (ITS; 5.8S+ITS-2) (p-distance of 1.569, 85.84% parsimony informative sites) and COI (6.084, 82.14 %), followed by the 18S rRNA gene (0.139, 57.69 %), *rbcL* (0.120, 42.01 %) and UPA (0.050, 14.97 %), which indicated that ITS and COI were highly divergent compared with the other tested genes, and that their nucleotide compositions were variable within the whole group of diatoms. Bayesian inference (BI) analysis showed that the phylogenetic trees generated from each gene clustered diatoms at different phylogenetic levels. The 18S rRNA gene was better than the other genes in clustering higher diatom taxa, and both the 18S rRNA gene and rbcL performed well in clustering some lower taxa. The COI region was able to barcode species of some genera within the Bacillariophyceae. ITS was a potential marker for DNA based-taxonomy and DNA barcoding of Thalassiosirales, while species of Cyclotella, Skeletonema and Stephanodiscus gathered in separate clades, and were paraphyletic with those of Thalassiosira. Finally, UPA was too conserved to serve as a diatom barcode.

Saddhe and Kumar (2018) remarked in the study that plant identification is a crucial and routine taxonomic procedure in order to understand and conserve the biodiversity. Anthropogenic activity, pollution, deforestation, and exploitation ofnatural 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. In the present review, they compiled the recent progress of plant DNA barcoding in various taxonomic groups and utility of plastids and nuclear DNA based markers for plant identification.Besides the core DNA barcode *rbcL* and *matK,* plant barcoding needs some supplementary markers such as *trnH-psbA* and ITS. Moreover, inclosely related and cryptic taxa DNA barcoding isalways 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 can be effectively used.

2.3. DNA Barcoding in Rice

The present study focuses on the DNA barcoding of rice (*Oryza sativa* L.) which forms the staple food of the nation and the work carried on rice for reference by scientists is as follows.

Singh *et. al.* (2017) remarked in the article DNA barcoding is a technique that makes use of short sequences from a standardized region of a genome to provide quick and reliable identification of species among all forms of life. Candidate loci belonging to chloroplast genome (CpG) and nuclear genome have been analyzed in various plants to identify universal barcoding loci capable of inter and intraspecies discrimination. In this study, relative potential of 24 candidate loci from plastid genome were validated on set of 231 diverse rice genotypes, for selection of suitable barcoding loci for DNA barcoding in rice. Results indicated that only one of the chloroplast CGS primer pair "*psbA-trnH*" showed (100%) amplification efficiency followed by "*rbcL*" (89.61%), "*atpH-atpl*" (68.39%), "*matK*" (66.2%) and "*petA-psbJ*" (62.33%). While 9 primers showed lower amplification efficiency between 5.19% and 52.81%. Based on amplification efficiency, reproducibility and amplicon size (as per Consortium for the Barcode of Life standard) five primers were selected for amplicon sequencing and further study of phylogenetic and phylogeographical relationships among above genotypes. The outcome of the study indicated that more standardization of universal primers is required to improve amplification efficiency and to get of higher number of informative loci.

Further designing of new primers from the specific site of rice chloroplast genome will help in precise amplification of reproducible chloroplast genome specific loci. As more loci will be identified and validated using sequencing informative data for analyzing intra species variation in rice will be achievable.

Singh and Banerjee (2018) focused on the diversity of rice germplasm of Chhattisgarh state which is well known for rice cultivation and called “rice bowl” of India. This genetic diversity among and between landraces, exists a wide scope for future crop improvement. In this study, *in situ* application of DNA barcodes was tested on selected diverse rice genotypes with the aim of contributing to the identification, conservation and protection of Intellectual Property Rights (IPR) of state. DNA Barcoding technique was successfully pioneered in animals using a portion of the *cytochrome oxidase 1* (CO1) mitochondrial gene. In plants, establishing a standardized DNA barcoding system has been more challenging. Thus potential of the *rbcl* and *matK* markers for the selection of barcoding loci of rice genotypes. The panel of 231 diverse rice genotypes including germplasm lines, elite, varieties and wild rice were used. The finding showed that amplification efficiency observed in panel of intraspecies of rice was in *rbcl*(89.6%) and *matK*(62.33%).On the basis of amplification efficiency panel of 24 rice genotypes selected for sequencing. The parsimony informative sites was estimated with maximum 305 sites recorded in *Matk,* followed by 264 sites *Rbcl* and number of variable sites reported highest in *rbc*l 672 followed by *MatK-f* 246. While nucleotide diversity per site pi(`) reported maximum in *rbcl* 0.21613(MEGA7.025). This scientific information data submitted to Barcoding of life database (BOLD) for generation of illustrative barcode. Chloroplast genome sequence specific loci were used in the study to assess their potential and identified candidate DNA Barcode loci for intraspecies discrimination in rice. Barcode regions must be relatively short in length to facilitate easy PCR amplification and DNA sequencing. The size of the amplification products obtained from other plant barcode loci ranged between700-1200bp. The *rbcL* and *matK* loci were not amplified in all the 231 genotypes amplification efficiency observed in panel of intraspecies of rice was in *rbcl* (89.6%) and *matK* (62.33%). The *rbcL* based DNA barcoding has been exhibited at efficiency of inter / intrageneric levels lived in cupressaceae, Cornaceae, Ericaceae, Graniaceae (Gilley*et al.*, 1994). Several studies have reported that *rbcL* proved to be the most promising locus in terms of amplification and sequencing success in plants followed by *rpoC1*, *rpoB*, *matK*, *ITS2*, *trnH-psbA*, *trnL-F* as well as *psbK*-I and lastly *atpF-H*. The panel representative of 24 genotypes consisted of landraces, wild rice, variety, elite lines along with aromatic genotypes. In conclusion, this study provided preliminary assessment data that will be useful for wider application of DNA barcoding, which will be useful to improve discrimination ability of intraspecies level in rice. On the basis of overall analysis, they found that *rbcL* and *matK* are useful for barcoding intrsapecies of rice genotypes belongs to Chhattisgarh, India.

2.4. DNA barcoding in various crops

DNA barcoding apart from focusing on just the taxonomical classification also focuses at the diversity pattern at genetic level, alterations and deletions in the plastid genome, etc.

The barcoding approach has also been applied in the research areas of different crops from varied species and genera. Some of the work recoginzed here for the study are put forth ahead.

Kress *et.al.* (2005) proposed the nuclear internal transcribed spacer region and the plastid *trnH-psbA* intergenic spacer as potentially usable DNA regions for applying barcoding to flowering plants. The internal transcribed spacer is the most commonly sequenced locus used in plant phylogenetic investigations at the species level and shows high levels of interspecific divergence. The *trnH-psbA* spacer, although short (450-bp), is the most variable plastid region in angiosperms and is easily amplified across a broad range of land plants. Comparison of the total plastid genomes of tobacco and deadly nightshade enhanced with trials on widely divergent angiosperm taxa, including closely related species in seven plant families and a group of species sampled from a local flora encompassing 50 plant families (for a total of 99 species, 80 genera, and 53 families), suggest that the sequences in this pair of loci have the potential to discriminate among the largest number of plant species for barcoding purposes. The ITS and *rbcL* loci provide a baseline against which to compare other genes and intergenic spacers in our directed search for sequences to use in plant DNA barcoding. Besides *ITS*, those single-copy nuclear genes or their introns that are gaining prominence in species-level molecular systematics studies (e.g., *leafy*, *waxy*, *pistillata*, and *RPB2*), also were considered. They focused attention to the plastid genome in search of the most variable sequences that would also meet the criteria needed for maximum utility (i.e., variability, universal primers, and short length) and that could be used in place of or in addition to the ITS region

CBOL (2009) suggested that DNA barcoding involves sequencing a standard region of DNA as a tool for species identification. However, there has been no agreement on which region(s) should be used for barcoding land plants. To provide a community recommendation on a standard plant barcode, they have compared the performance of 7 leading candidate plastid DNA regions (*atpF–atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK–psbI* spacer, and *trnH–psbA* spacer). Based on assessments of recoverability, sequence quality, and levels of species discrimination, they recommended the 2-locus combination of *rbcL_matK* as the plant barcode. This core 2-locus barcode would provide a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of land plants.

Gonzalez *et.al.* (2009) examined whether plant DNA barcoding could contribute to increasing the quality and the pace of tropical plant biodiversity surveys. Of the eight plant DNA markers we tested (rbcLa, rpoC1, rpoB, matK, ycf5, trnL, psbA-trnH, ITS), matK and ITS had a low rate of sequencing success. More critically, none of the plastid markers achieved a rate of correct plant identification greater than 70%, either alone or combined. The performance of all barcoding markers was noticeably low in few species-rich clades, such as the Laureae, and the Sapotaceae. A field test of the approach enabled us to detect 130 molecular operational taxonomic units in a sample of 252 juvenile trees. Including molecular markers increased the identification rate of juveniles from 72% (morphology alone) to 96% (morphology and molecular) of the individuals assigned to a known tree taxon. The study concluded that while DNA barcoding is an invaluable tool for detecting errors in identifications and for identifying plants at juvenile stages, its limited ability to identify collections will constrain the practical implementation of DNA-based tropical plant biodiversity programs.

Bafeel *et. al.* (2011) worked on the 2-locus combination of *rbcL* and *matK* as the standard plant barcode recommended by The Consortium for the Barcode of Life (CBOL) plant-working group. These two regions of chloroplast DNA were chosen due to efficient recovery of quality sequences and high levels of species discrimination. They evaluated the success rates of universal primers for amplification of *matK* and *rbcL* loci in 26 different plant species (covering 14 families) from Saudi Arabia. Success rate in PCR was higher for *rbcL* 88%) compared with *matK* (69%). The universal primers of both *matK* and *rbcL* failed to amplify the DNA form 3 plant species belonging to the family Asteraceae (*Anthemisdeserti, Pulicaria undulate,* and *Sonchusoleraceus*). Two plant species *Malvaparviflora* (Malvaceae) and *Salsola imbricate* (Chenopodiaceae) indicated different primer binding site (*matK*) as the amplified PCR products were of lower size than expected for these species. These findings indicated that although currently used universal primers of *rbcL* and *matK* were able to amplify many of the plant species they may fail in certain cases due to primer mismatch at the annealing site. The findings of this preliminary study indicated that currently used universal primers of *rbcL* and *matK* were able to amplify several of the plant species of arid origin with the amplification success rates of 88% and 69% respectively. All the primer pairs produced sharp bands that are suitable for gene sequencing.Not withstanding their failure in certain cases, the currently used primers for *rbcL* and *matK* are useful for plant barcoding until the discovery of more efficient and robust primers for a broader coverage of plant species. Concluding the study they suggested there is a need for protocol development to enhance the amplification strategies including the development of new primers or primer cocktails for enhanced success in barcoding of plant species of different regions.

Wang *et.al.* (2012) evaluated the four proposed barcoding loci (*matK, rbcL, trnH-psbA* and *ITS*) on nine species of Nyssaceae. The results showed that ITS was the best performing single locus, although *mat*K + *rbc*L might be used as the core barcodes for land plants. The chloroplast regions have low resolution compared with ITS. The low efficiency of these candidate barcodes in Nyssaceae might be caused by a poor taxonomy, especially within the genus *Nyssa*. The results also indicated that species status of *N. shangszeensis*, *N. sinensis*, *N. shweliensis*and *N. wenshanensis* requires to be re-evaluated based on more morphological characters combined with rapidly evolving loci. The data indicated that *Nyssa* in eastern Asia comprises three relatively old lineages, i.e. *N. javanica*, *N. yunnanensis*and all other species (bearing in mind that relationships of *N. leptophyll a*re main uncertain), and that recent diversification within the latter clade has given rise to much of the morphological diversity now evident in eastern Asian *Nyssa*. In-spite of potential taxonomical problems in *Nyssa*, among the studied loci, ITS appeared to be the best barcode candidate for Nyssaceae.

Purushothaman *et.al.*, (2014) in their work mentioned that DNA barcoding is a desirable tool for medicinal product authentication. DNA barcoding is a method for species identification using short DNA sequences that are conserved within species, but variable between species. Unlike animals, there is no single universal DNA barcode locus for plants. Coding markers, *matK* and *rbcL*, and noncoding markers, *trnH-psbA* (chloroplast) and *ITS2* (nuclear), have been reported to be suitable for the DNA barcoding of plants with varying degree of success. Sixty-four accessions from 20 species of the medicinal plant *Cassia* were collected, and analyzed for these 4 DNA barcoding markers.

PCR amplification was 100% successful for all 4 markers, while intra-species divergence was 0 for all 4 *Cassia* species in which multiple accessions were studied. Assuming 1.0% divergence as the minimum requirement for discriminating 2 species, the 4 markers could only differentiate 15 to 65% of the species studied when used separately. Adding indels to the divergence increased the percentage of species discrimination by *trnH-psbA* to 90%. In 2-locus barcoding, while *matK+rbcL*(which is recommended by Consortium for the Barcoding of Life) discriminated 90% of the species, the other combinations of *matK*+ITS and *rbcL+trnH-psbA* showed 100% species discrimination. However, matK is plagued with primer issues. The combination of *rbcL+trnH-psbA* provided the most accurate (100% species ID) and efficient tiered DNA barcoding tool for the authentication of *Cassia* medicinal products. In conclusion, the research provided a tiered barcode authentication tool to differentiate the medicinal *Cassia* species in India. The combination of *rbcL+trnH-psbA* provides the most accurate (100% species ID) and efficient multi-locus DNA barcoding tool for the authentication of *Cassia* medicinal products. Single-locus barcoding did not differentiate the studied 20 *Cassia* species, with this limitation being supported by research on many other species of plants. Although the addition of *indels* to the pair wise divergence in *trnH-psbA* seemed to increase the percentage of species discrimination, further confirmation is required after a more complete sampling of the genus. The multivariate analysis used here revealed considerable sequence variation that might easily differentiate all species of *Cassia*. Considering universal amplification and divergence, wherever necessary, *trnH-psbA* could serve as a supplementary marker to *rbcL* in a tiered barcode approach, in which other regions might be used as a second tier as needed. It is unlikely that more than 2 markers would be needed to sequence any specific group of plants.

Li *et.al.* (2017) aimed majorly to identify the variety and provenance of the tea plant. The present experiment used 113 tea plants [*Camellia sinensis* (L.) O. Kuntze] housed at the Tea Research and Extension Substation, from which 113 internal transcribed spacer 2(ITS2) fragments, 104 *trnL* intron, and 98 *trnL-trnF* intergenic sequence region DNA sequences were successfully sequenced. The similarity of the *ITS2* nucleotide sequences between tea plants housed at the Tea Research and Extension Substation was 0.379e0.994. In this polymerase chain reaction-amplified non coding region, no varieties possessed identical sequences. Compared with the *trnL* intron and *trnL-trnF* intergenic sequence fragments of chloroplast cpDNA, the proportion of *ITS2* nucleotide sequence variation was large and is more suitable for establishing a DNA barcode database to identify tea plant varieties. After establishing the database, 30 imported teas and 35 domestic made teas were used in this model system to explore the feasibility of using *ITS2* sequences to identify the varieties and provenances of made teas.

A phylogenetic tree was constructed using *ITS2* sequences with the un weighted pair group method with arithmetic mean, which indicated that the same variety of tea plant was likely to be successfully categorized into one cluster, but contamination from other tea plants was also detected. This result provided molecular evidence that the similarity between important tea varieties in Taiwan remains high. They suggested a direct, wide collection of made tea and original samples of tea plants to establish an *ITS2* sequence molecular barcode identification database to identify the varieties and provenances of tea plants. The DNA barcode comparison method could satisfy the need for a rapid, low-cost, frontline differentiation of the large amount of made teas from Taiwan and abroad, and could provide molecular evidence of their varieties and provenances. The present study successfully sequenced the *ITS2* sequence of 113 tea plant varieties. In this PCR-amplified non coding region fragment, no tea plant varieties share identical sequences. The molecular fingerprint library established using the *ITS2* sequences could successfully identify all tea plant varieties housed at the tea plantation, and they are good core molecular marker targets. They finally suggested that DNA barcode method could be used as a frontline inspection method to compare *ITS2* sequences across databases, to identify made tea sample variety and provenance.

Skuza *et.al.,* (2019) worked on plant material that consisted of 10 cultivated and non-cultivated species and subspecies of rye genus. Three chloroplast DNA regions (*rbcL, matK, trnH-psbA*) were tested for their suitability as DNA barcoding regions. Universal primers were used, and sequenced products were analyzed using Neighbour Joining and the

Maximum Likelihood in the MEGA 7.1 program. Only 2.2% of the sequences showed polymorphism in the *rbcL* region, while 6.5% in the *mat*K region. The most variable *trnH-p*sbA (15.6%) intergenic region was the most useful for rye barcoding. Individual application of the studied regions did not provide the expected results. None of the regions used in the study allowed the division of rye species and subspecies according to the adopted classification of the genus Secale. The results confirmed that the use of *matK* and *rbcL* is insufficient for DNA barcoding in rye species, and better discrimination within the genus Secale can be obtained only in combination with the non-coding *trnH-psbA* sequence. Our results also indicate the necessity of using a different region. All of the new sequences have been deposited in Genbank. The present study was the first to analyze selected rye species and subspecies, in which the usefulness of the combinations of the plastid *rbcL* and *matK* coding regions and intergenic *trnH-psbA* region for DNA barcoding was assessed. The results confirmed that the use of *matK* and *rbcL* is insufficient for DNA barcoding in rye species, and better discrimination within the genus Secale can be obtained only in combination with the non-coding *trnH-psbA* sequence.

2.5. Various Applied Areas

Many research articles focused on the application of DNA barcoding to achieve various research objectives. Some of the articles regarding the conceptual application have been given below.

Hebert *et. al.*, (2005) established that the mitochondrial gene cytochrome *c* oxidase I (COI) can serve as the core of a global bioidentification system for animals. First, they demonstrated that COI profiles, derived from the low-density sampling of higher taxonomic categories, ordinarily assign newly analyzed taxa to the appropriate phylum or order. Secondly, they demonstrated that species-level assignments can be obtained by creating comprehensive COI profiles. A model COI profile, based upon the analysis of a single individual from each of 200 closely allied species of lepidopterans, was 100% successful in correctly identifying subsequent specimens. When fully developed, a COI identification system will provide a reliable, cost-effective and accessible solution to the current problem of species identification. Its assembly would also generate important new insights into the diversification of life and the rules of molecular evolution.

Hajibabaei *et.al.,* (2005) proposed that the use of DNA barcodes, short DNA sequences from a standardized region of the genome as a tool to facilitate species identification and discovery. The study showed that cytochrome *c* oxidase I DNA barcodes effectively discriminate among species in three Lepidoptera families from Area de Conservacio´ n Guanacaste in north-western Costa Rica. They found that 97.9% of the 521 species recognized by prior taxonomic work possess distinctive cytochrome *c* oxidase I barcodes and that the few instances of inter specific sequence overlap involve very similar species. Also, two or more barcode clusters within each of 13 supposedly single species. Co variation between these clusters and morphological and or ecological traits indicated overlooked species complexes.

Chase *et.al.* (2005) assessed two of the markers commonly sequenced in land plant phylogenetic studies, plastid *rbcL* and internal transcribed spacers of the large subunits of nuclear ribosomal DNA (*ITS*). They concluded that DNA markers from uniparentally (usually maternally) inherited genomes can only provide half of the story required to improve taxonomic standards being used in DNA barcoding.

Kress *et.al,*(2005) presented results on the use of a phylogenetic constraint tree to generate a community phylogeny for a diverse, tropical forest dynamics plot in Puerto Rico. This enhanced method of phylogenetic reconstruction insures the congruence of the barcode phylogeny with broadly accepted hypotheses on the phylogeny of flowering plants (i.e., APG III) regardless of the number and taxonomic breadth of the taxa sampled. They also compare maximum parsimony versus maximum likelihood estimates of community phylogenetic relationships as well as evaluate the effectiveness of one- versus two- versus three-gene barcodes in resolving community evolutionary history. They reconstructed a community phylogeny for LDFP using maximum likelihood (ML) and maximum parsimony (MP) algorithms. Three different marker combinations were examined for performance in phylogenetic reconstruction: *rbcL+matK*, *rbcL+trnH-psbA*, and *rbcL+matK+trnH-psbA*. For all combinations of markers, 142 species were included, with six sequences of *rbcL* a obtained from GenBank and used in conjunction with our barcode sequences. ML analyses were conducted using RAxML via the CIPRES supercomputer cluster (www.phylo.org). The different locus combinations were partitioned for independent model assessment at each marker. For all combinations of markers a single most likely tree was estimated in addition to running 200–250 bootstrap replicates depending on the marker set.

Kress *et. al.,* (2009) in the study demonstrated that a three-locus DNA barcode when applied to 296 species of woody trees, shrubs, and palms found within the 50-ha Forest Dynamics Plot on Barro Colorado Island (BCI), Panama, resulted in >98% correct identifications. These DNA barcode sequences were also used to reconstruct a robust community phylogeny employing a super matrix method for 281 of the 296 plant species in the plot. The three-locus barcode data were sufficient to reliably reconstruct evolutionary relationships among the plant taxa in the plot that are congruent with the broadly accepted phylogeny of flowering plants (APG II). The results illustrated how highly resolved phylogenies based on DNA barcode sequence data will enhance research focused on the interface between community ecology and evolution.

Radulovici *et.al.* (2010) reviewed the role of DNA barcoding for the study of marine biodiversity at the species level. It will have multiple applications for marine life: identification of larvae, invasive species, cryptic species, new species, illegal trade of protected species, stock management, biodiversity assessments, ecosystem monitoring, revisions of certain taxa, inference of phylogenetic relationships, phylogeographic and speciation patterns.

Rydberg (2010) presented in this report to demonstrate the utility of DNA barcoding as a method for identifying ethnobotanical material. There is considerable potential for the use of barcoding in 19 applied cases, such as studying the interactions between market processes underlying the trade in medicinal plant products and the status of wild plant populations.

Bruni *et. al.,*(2012) investigated how the DNA barcoding can support the modern digital approaches to the identification of organisms, using as a case study a local flora, that of Mt. Valerio, a small hill near the centre of Trieste (NE Italy).

The core barcode markers (plastidial *rbcL* and *matK*), plus the additional *trnH-psbA* region, were used to identify vascular plants specimens. The usefulness of DNA barcoding data in enhancing the performance of a digital identification key was tested on three independent simulated scenarios. Results showed that the core barcode markers univocally identify most species of our local flora (96%). The *trnH-psbA* data improve the discriminating power of DNA barcoding among closely related plant taxa. In the multi parametric digital key, DNA barcoding data improves the identification success rate; in our simulation, DNA data overcame the absence of some morphological features, reaching a correct identification for 100% of the species. FRIDA, the software used to generate the digital key, has the potential to combine different data sources and they proposed to use this feature to include molecular data as well, creating an integrated identification system for plant biodiversity surveys.

Saarela *et.al.* (2013) generated DNA barcodes for the core plastid barcode loci (*rbcL* and *matK*) for 490 vascular plant species, representing nearly half of the Canadian Arctic flora and 93% of the flora of the Canadian Arctic Archipelago. Sequence recovery was higher for *rbcL* than *matK* (93% and 81%), and *rbcL* was easier to recover than *matK* from herbarium specimens (92% and 77%). Distance-based and sequence similarity analyses of combined *rbcL*+ *matK* data discriminate 97% of genera, 56% of species, and 7% of intraspecific taxa. We characterize barcode variation in detail in the ten largest genera sampled (*Carex*, *Draba*, *Festuca*, *Pedicularis*, *Poa*, *Potentilla*, *Puccinellia*, *Ranunculus*, *Salix*, and *Saxifraga*) in the context of their phylogenetic relationships and taxonomy. Discrimination with the core barcode loci in these genera ranges from 0% in *Salix* to 85% in *Carex*. Haplotype variation in multiple genera did not correspond to species boundaries, including *Taraxacum*, in which the distribution of plastid haplotypes among Arctic species was consistent with plastid variation documented in non-Arctic species. Introgression of *Poaglauca* plastid DNA into multiple individuals of *P. hartzii*is problematic for identification of these species with DNA barcodes. Of three supplementary barcode loci (*psbA–trnH*, *psbK–psbI*, *atpF–atpH*) collected for a subset of *Poa* and *Puccinellia* species, only *atpF–atpH* improved discrimination in *Puccinellia*, compared with *rbcL* and *matK*. Variation in *matK* in *Vaccinium uliginosum* and *rbcL* in *Saxifraga oppositifolia* corresponds to variation in other loci used to characterize the phylogeographic histories of these Arctic-alpine species. They have generated new plastid DNA barcodes for nearly half of the vascular plant species in the Canadian Arcticecozone, and 93% of the species in the Canadian Arctic Archipelago. Although the core plastid barcode loci distinguish just over 50% of the sampled Arctic species, this figure does not convey the full utility of a comprehensive DNA barcode reference library for facilitating species identification. The barcode data facilitated identification of taxonomic and identification problems, as the local placement of misidentified specimens in neighbour joining trees was often insightful in helping us make correct morphology-based identifications, even when several closely related taxa shared barcode haplotypes. Detailed examination of the barcode data in a subset of taxa indicated, not unexpectedly, that the ability of the core plastid barcode loci to discriminate species is closely related to the evolutionary relationships of sampled taxa.

Hollings worth *et.al.* (2016) provided a short summary of the strengths and limitations of plant DNA barcoding for addressing these issues and discussed options for enhancing current plant barcodes, focusing on increasing discriminatory power via either gene capture of nuclear markers or genome skimming. To achieve this, they advocated a twin-track approach of (i) continued construction of the reference library via large-scale sample sets and careful deployment of standard plant barcodes, while (ii) maintaining and enhancing international collaborative efforts to further develop plant barcode protocols to support the ultimate objective of establishing a workflow with the resolving power to uniquely discriminate the vast majority of the world's land plant species.

Galimberti *et.al.,* (2013) reviewed and analyzed the results of several researches in order to exploit the effectiveness of DNA barcoding in food traceability, and to delineate some best practices in the application of DNA barcoding throughout the industrial pipeline. The case studies and technical advancements clearly indicated that DNA barcoding is a sensitive, fast, cheap and reliable method for identifying and tracking a wide panel of raw materials and derived food commodities (even in the case of strongly processed food products), and for detecting allergens or poisonous components potentially occurring in food matrices.

Raja *et.al.* (2015) indicated that there is no single universal barcode candidate for identification of all plant groups. Hence, comparative analysis of plant barcode loci is essential for choosing a best candidate for authenticating particular medicinal plant genus/families. Currently, both chloroplast/ nuclear regions are used as universal barcodes for the authentication of phytomedicinals. A recent advance in genomics has further enhanced the progress in DNA barcoding of plants by the introduction of high-throughput techniques like next generation sequencing, which has paved the way for complete plastome sequencing that is now termed as super-barcodes. The approaches could improve the traditional ethno-botanical and scientific knowledge of phytomedicinals and their safe use. Hence, current study focused on the investigation of phytomedicinals and herbal product integrity and authenticity through DNA barcoding with the goal of protecting consumers from potential health risks associated with product substitution and contamination. DNA barcoding is being viewed as an integrated approach with classical taxonomy for species identification and authentication in the post genomics era. DNA barcoding has been employed effectively to identify the cryptic species, medicinal plants, species and biological authentication of materials, and plant biodiversity conservation that added value to both traditional and scientific knowledge. Contemporary approaches like ecological genomics along with the use of NGS could exploit and advance DNA barcoding research to the next level. The barcoding movement along with NGS approach could help to speed up the authentication of voucher specimens and herbal drugs.

Hogg and Hebert (2015) evaluated sequence diversity in the mitochondrial cytochrome-*c* oxidase I (COI; EC 1.9.3.1) gene as a tool for resolving differences among species of Arctic springtails. The Collembola examined in this analysis were collected from Igloolik, Cornwallis, and Somerset islands and included representatives from all major families found in the Arctic. Members of 13 genera and 19 species were examined, including

4 species of the genus *Folsomia* and 3 species of the genus *Hypogastrura*. In all cases, species were successfully discriminated. Sequence divergences within species were generally less than 1%, whereas divergences between species were greater than 8% in all cases. Divergences among individuals of one species of *Folsomia* were much higher (up to 13%), but this likely represents the presence of an undescribed sibling species. They concluded that DNA barcoding is a powerful tool for identifying species of Collembola and should regularly be useful as a complement to traditional, morphological taxonomy.

Mishra *et.al.* (2016) suggested that for effectively resolving authentication challenges associated with the herbal market, DNA barcoding must be used in conjunction with metabolomics along with need-based transcriptomics and proteomics. For herbal plant identification, *matK, rbcL, trnH-psbA, ITS, trnL-F*, 5S-rRNA and 18S-rRNA have been used as successful DNA barcodes.

2.6. Chloroplast Genome

The DNA segments from the genome which are used for barcoding. The regions mainly include *matK, rbcL, trnH-psbA* which cover most of the regions of the genome. To understand the regions more precisely some research articles reviewed have been notified here. The main aim of focusing on the plastidal genome (chloroplast in plants and mitochondria in animals) is they carry maternal genes which lack any changes in the gene and show good ancsetory which proves essential factor to study the phylogeny as well as to define the DNA barcode for that particular living entity.

Hiratsuka *et.al.,*(1989) studied the predicted genes that have been identified along with open reading frames (ORFs) conserved between rice and the previously sequenced chloroplast genomes, a dicot, tobacco*(Nicotiana tabacum),* and a liverwort *(Marchantia polymorpha).*The same complement of 30 tRNA and 4 rRNA genes has been conserved between rice and tobacco. Most ORFs extensively conserved between *N. tabacum* and *M. polymorpha* are also conserved intact in rice. However, several such ORFs are entirely absent in rice, or present only in severely truncated form. Structural changes are also apparent in the genome relative to tobacco. The inverted repeats, characteristic of chloroplast genome structure, have expanded outward to include several genes present only once per genome in tobacco and liverwort and the large single copy region has undergone a series of inversions which predate the divergence of the cereals. A chimerict RNA pseudogene overlapped an apparent endpoint of the largest inversion, and a model invoking illegitimate recombination between tRNA genes is proposed which accounts simultaneously for the origin of this pseudogene, the large inversion and the creation of repeated sequences near the inversion endpoints.

Ishii *et.al.,*(1988) established relationships between chloroplast genomes (=ctDNAs) from diploid, A genome species of rice (*Oryza spp.*) using length differences in restriction fragments of ctDNAs. Five out of 11 endonucleases used revealed differences in ctDNA fragment lengths among 19 accessions of the cultivated species, *0. sativa* and *0. glaberrima.* A larger sample of ctDNAs from 66 accessions of both wild and cultivated species were analyzed with three of those five endonucleases, resulting in the identification of nine types of chloroplast genome. In the wild species, the four geographical forms of *0.perennis* were found to contain one (African), one (Oceanian), three (Asian), and four (American)types, whereas *0. breviligulata* contained only one type of chloroplast genome.In the cultivated species, the three ecospecies of *0.sativa* contained one(Japonica and Javanica) and three (Indica) types, whereas *0. Glaberrima* contained one type of chloroplast genome. The latter type was shared exclusivelywith *0. breviligulata,* suggesting that the former species had been domesticatedfrom the latter. Two chloroplast genome types were shared by *0. Sativa* and the Asian form of *0.perennis*, suggesting that the former species has derived from the latter via two domestication events.

Shimada and Sugiura (1991) compared all the chloroplast genes, open reading frames and spacer regions in the plastid genomes of rice, tobacco and liverwort species in order to elucidate general structural features of the chloroplast genome. Analyses of homology, GC content and codon usage of the genes enabled to classify them into two groups: photosynthesis genes and genetic system genes. Based on comparisons of homology, GC content and codon usage, unidentified ORFs can also be assigned to each of these groups such that it is possible to speculate about the functions of products which may be produced by these ORFs. The spacer regions and intron sequences were compared- and found to have no obvious homology between rice and liverwort or between tobacco and liverwort.

Provan *et.al.*, (1996) identified short mononucleotide repeats analogous to nuclear microsatellites or simple sequence repeats (SSRs) in chloroplast genomes. Primers flanking mononucleotide repeats in the fully sequenced rice chloroplast genome were used in conjunction with PCR to amplify genomic DNA from 42 wild rice accessions. The amplification products exhibited length polymorphism, which allowed the levels of chloroplast variability detected to be quantified.

Seven primer pairs that amplified products from different regions of the rice chloroplast were used, five of which also amplified polymorphic products in cultivated rice (*Oryza sativa*). Diversity values ranged from 0.5224 - 0.0845 (SE) to 0.8298 + 0.0085 in the wild accessions, which was higher than that detected in the 0. sativa accessions. Both intra- and inter-specific polymorphism was detected, and the extent of chloroplast genomic differentiation based on chloroplast simple sequence repeat (cpSSR) assays was quantified.

Matsuoka *et.al.,* (2002) investigated the evolution of chloroplast genes and genomes in the grass family (Poaceae) by whole-genome comparison using the fully sequenced chloroplast genomes of maize (subfamily Panicoideae), rice (subfamily Bambusoideae), and wheat (subfamily Pooideae) .Analyses of nucleotide sequence variations in 106 cereal chloroplast genes with tobacco sequences suggested that (1) most of the genic regions of the chloroplast genomes of maize, rice, and wheat have evolved at similar rates; (2) RNA genes have highly conservative evolutionary rates relative to the other genes; (3) photosynthetic genes have been under strong purifying selection; (4) between the three cereals, 14 genes which account for about 28% of the genic region have evolved with heterogeneous nucleotide substitution rates; and (5) rice genes tend to have evolved more slowly than the others at loci where rate heterogeneity exists. Although the mechanism that underlies chloroplast gene diversification is complex, our analyses identified variation in non synonymous substitution rates as a genetic force that generates heterogeneity, which is evidence of selection in chloroplast gene diversification at the intrafamilial level. Phylogenetic trees constructed with the variable nucleotide sites of the chloroplast genes place maize basal to the rice wheat clade, revealing a close relationship between the Bambusoideae and Pooideae.

Tang *et.al,*(2004) assembled two chloroplast genome sequences from two rice (*Oryza sativa*) varieties, one from 93-11 (a typical indica variety) and the other from PA64S (an indica-like variety with maternal origin of japonica) using high quality sequence reads extracted from our whole genome shotgun repository, which are both parental varieties of the super-hybrid rice, LYP9. Based on the patterns of high sequence coverage, we partitioned chloroplast sequence variations into two classes, intra varietal and inter subspecific polymorphisms. Intra varietal polymorphisms refer to variations within 93-11 or PA64S. Inter sub specific polymorphisms were identified by comparing the major genotypes of the two subspecies represented by 93-11 and PA64S, respectively. They found that the inter sub specific variations of 93-11 (indica) and PA64S (japonica) chloroplast genomes consisted of 72 single nucleotide polymorphisms and 27 insertions or deletions. The inter sub specific polymorphism rates between 93-11 and PA64S were 0.05% for single nucleotide polymorphisms and 0.02% for insertions or deletions, nearly 8 and 10 times lower than their respective nuclear genomes. Based on the total number of nucleotide substitutions between the two chloroplast genomes, we dated the divergence of indica and japonica chloroplast genomes as occurring approximately 86,000 to 200,000 years ago.

Kumar *et.al.,* (2009) worked on *Parthenium argentatum* (guayule) which is an industrial crop that produces latex and was commercialized as a source of latex rubber safe for people with Type I latex allergy. The complete plastid genome of *P. argentatum* was sequenced. The sequence provides important information useful for genetic engineering strategies. Comparison to the sequences of plastid genomes from three other members of the Asteraceae, *Lactucasativa, Guitozia abyssinica* and *Helianthus annuus* revealed details of the evolution of the four genomes. Chloroplast-specific DNAbarcodes were developed for identification of *Parthenium* species and lines. The complete plastid genome of *P. argentatum*is 152,803 bp. Based on the overall comparison of individual protein coding genes with those in *L. sativa, G. abyssinica*and *H. annuus*, they demonstrated that the *P. argentatum* chloroplast genome sequence is most closely related to that of*H. annuus*. Similar to chloroplast genomes in *G. abyssinica, L. sativa* and *H. annuus*, the plastid genome of *P. argentatum* has a large 23 kb inversion with a smaller 3.4 kb inversion, within the large inversion. Using the *mat*K and *psb*A-*trn*H spacer chloroplast DNA barcodes, three of the four *Parthenium* species tested, *P. tomentosum*, *P. hysterophorus* and P. *schottii*, can be differentiated from*P. argentatum*. In addition, we identified lines within *P. argentatum*. The genome sequence of the *P. argentatum* chloroplast enriched the sequence resources of plastid genomes in commercial crops. The availability of the complete plastid genome sequence may facilitate transformation efficiency by using the precise sequence of endogenous flanking sequences and regulatory elements in chloroplast transformation vectors. The DNA barcoding study forms the foundation for genetic identification of commercially significant lines of *P. Argentatum* that are important for producing latex. The genome sequence of the *P. argentatum* chloroplast will enrich the sequence resources of plastid genomes in commercial crops. The availability of the complete plastid genome sequence may facilitate improved transformation efficiency by using the precise endogenous flanking sequences and regulatory elements in chloroplast transformation vectors. The DNA barcoding study forms the foundation for genetic identification of commercially important lines of *P. argentatum* that are producing natural rubber latex for biomedical applications.

Nock *et.al.,* (2010) proposed the analysis of chloroplast genome sequences from massively parallel sequencing (MPS) of total DNA as a simple and cost-effective option for plant barcoding, and analysis of plant relationships to guide gene discovery for biotechnology. They present chloroplast genome sequences of five grass species derived from MPS of total DNA. These data accurately established the phylogenetic relationships between the species, correcting an apparent error in the published rice sequence. The chloroplast genome may be the elusive single-locus DNA barcode for plants.

Waters *et. al.,*(2011) obtained the chloroplast genome sequence of *O. rufipogon* from Asia and Australia and *O. meridionalis* and *O. australiensis* (an Australian member of the genus very distant from *O. sativa*) by massively parallel sequencing and compared with the chloroplast genome sequence of domesticated *O. sativa. Oryza australiensis* differed in more than 850 sites single nucleotide polymorphism or indel from each of the other samples. The other wild rice species had only around 100 differences relative to cultivated rice.The chloroplast genomes of Australian *O.rufipogon* and *O.meridionalis* were closely related with only 32 differences. The Asian *O. rufipogon* chloroplast genome (with only 68 differences) was closer to *O. sativa* than the Australian taxa (both with more than 100 differences). The chloroplast sequences emphasized the genetic distinctness of the Australian populations and their potential as a source of novel rice germplasm.

Gao Li-Zhi *et.al.* (2019) reported the complete plastomes of 22 closely related *Oryza* species in chronologically ordered stages and generate the first precise map of genomic structural variation, to our knowledge. The occurrence rapidity was estimated on average to be ~7 insertions and ~15 deletions. Relatively fewer deletions than insertions result in an increased repeat density that causes the observed growth of *Oryza* chloroplast genome sizes. Genome-wide scanning identified 14 positively selected genes that are relevant to photosynthesis system, eight of which were found independently in shade tolerant or sun-loving rice species. *psaA* seemed positively selected in both shade-tolerant and sun-loving rice species. The results show that adaptive evolution of chloroplast genes makes rice species adapt to diverse ecological habitats related to sunlight preferences.

2.7. Bioinformatic Tools

The major task after PCR amplification is the sequencing and screening of the genome using bio informatics tools. Bioinformatics is the inter-disciplinary branch of biological science that uses mathematical and computational tools. These tools are available online for users. The major repository for the databases is the NCBI which has a good detailed information on genes and genome as well as contains several basic tools for use.

Chenna *et.al.,*(2003) threw light on the Clustal series of programs that are widely used in molecular biology for the multiple alignment of both nucleic acid and protein sequences and for preparing phylogenetic trees. The popularity of the programs depends on a number of factors, including not only the accuracy of the results, but also the robustness, portability and user-friendliness of the programs. New features included NEXUS and FASTA format output, printing range numbers and faster tree calculation. Although, Clustal was originally developed to run on a local computer, numerous Web servers have been set up, notably at the EBI (European Bioinformatics Institute) (http://www.ebi. ac.uk/clustalw/).

Daugelaite *et. al.,* (2013) explained that the multiple sequence alignment (MSA) of DNA, RNA, and protein sequences is one of the most essential techniques in the fields of molecular biology, computational biology, and bioinformatics. Next-generation sequencing technologies are changing the biology landscape, flooding the databases with massive amounts of raw sequence data. MSA of ever-increasing sequence data sets is becoming a significant bottleneck. In order to realise the promise of MSA for large-scale sequence data sets, it is necessary for existing MSA algorithms to be run in a parallelized fashion with the sequence data distributed over a computing cluster observer farm. Combining MSA algorithms with cloud computing technologies is therefore likely to improve the speed, quality, and capability for MSA to handle large numbers of sequences. They multiple sequence alignments were discussed, with a specific focus on the Clustal W and Clustal Omega algorithms.

2.8. DNA Fingerprinting

DNA fingerprinting has enabled to study the presence of various traits and locate them in the genome. The work recognized here from several articles is considered to set PCR temperature profile as well as given as information regarding the varieties of rice used for the present study.

Wu and Tanksley (1993) suggested that, among PCR based markers in rice, microsatellites are abundant and well distributed throughout the genome (Akagi*et al.*, 1996; McCouch *et al.*,1997). They are valuable as genetic markers because they are co-dominant, detect high levels of allelic diversity and are assayed efficiently by the PCR. Thus, there is the need to identify, evaluate and characterize the available rice genotypes at both morphological and molecular levels to diversify the genetic base of improved rice varieties (Ogunbayo *et al.*, 2005).

Gupta *et al.* (1999) revealed that, SSR markers are highly polymorphic even between closely related lines, requires low amounts of DNA, easily automated and allow high throughput screening, exchanged between laboratories and highly transferable between populations.

Chakarvarthi and Naravaneni (2006) investigated that, genetic diversity and DNA fingerprinting of 15 elite rice genotypes using 30 SSR primers on chromosome numbers 7-12. The results revealed that all the primers showed distinct polymorphism among the cultivars studied indicating the robust nature of microsatellites in revealing polymorphism.

Cluster analysis grouped the rice genotypes into 10 classes in which *Japonica* types DH-1 (Azucena) and Moroborekan clustered separately from *Indica* types. Principal component analysis was done t visualize genetic relationships among the elite breeding lines. The results were similar to UPGMA results. Based o this study, the larger range of similarity values for related cultivars using microsatellites provides greater confidence for the assessment of genetic diversity and relationships. The information obtained from the DNA fingerprinting studies helps to distinctly identify and characterize 9 varieties using 18 different RM primers. This information can be used in the background selections during backcross breeding programs.

Rahman (2009) reported on the utilization of a small set of three previously developed rice microsatellite markers for the identification and discrimination of 17HYVs and 17 local rice cultivars including two wild rice cultivars. All analyzed microsatellite markers were found to be polymorphic with an average number of 6.33 alleles per locus. These three markers were able to identify 15 local rice cultivars and 11 HYVs. A total of three variety specific alleles, RM-11/147,RM-151/289 and RM-153/178 were identified for BR-11,Badshabhog and BR-19 cultivars respectively.DNA fingerprints of rice cultivars by means of microsatellites provided meaningful data, which was extended by additional microsatellite markers. The data obtained was used for the protection of plant genetic resources.

Singh (2013) determined the allelic diversity and relationship among 32 rice genotypes by using the SSR marker analysis. A total of 78 alleles were detected at the 19 SSR loci, of which 78 (100 %) were polymorphic. The size of smallest an largest allele ranged from 1 (RM 264) to 9 (RM47).The size difference for a given SSR locus varied from 20 to 100 bp. Maximum variations in allele size was observed with the RM222 and RM547 markers. SSR loci with greater size difference between the alleles would be relatively better for fingerprinting and diversity analysis. Null alleles were observed in respect of RM19 and RM21 with all of the genotypes. Multiple alleles were observed at 1 of the 19 loci analyzed in varieties TWC and ACK 2. PIC values ranged between 0.775 (RM247) and 0.998 (RM544, RM561), with an average of 0.83 per marker. A number of SSRs were identified that could be utilized to differentiate between original and revised rice genotypes. Jaccard's similarity coefficients ranged from 0.65 to 0.97. These results could be useful for monitoring purity, genotype identification and for plant variety protection.

Chungada *et al.*, (2016) carried out study to screen the 55 rice germ plasm for biotic and abiotic stresses and their genetic diversity analysis by using 18 microsatellite markers pairs distributed throughout the genome. Blast resistance alleles were observed in the genotypes viz., Karjat-184, Karjat-1, IR 65598-112-2, PhuleMaval, Bhogavati, Ratnagiri-5, Phule Samruddhi, Ratnagiri-24, Karjat-7, RNT-55-3-2, RNT-1-1-2-1, RNT-66- 43-8, Karjat-3, ParasSona, Panvel-3, Jaya, BR-827 and Karjat-5. Salt tolerant genotypes include; Prabhavati, Ratnagiri-73, Karjat-1, Karjat-4, ParasSona, RTN Purple, BR 827, IR 65598-112-2, Karjat-5, Ratna, Abhaya, RNT-49-2-3-1-2, IR 68952-5-2-11-8-1, RNT 66-43-8, Jaya, Bhogavati, Phule Samruddhi, IR-46, Indrayani, Saysree, NPT-2, IR-54, RP-4-14, Panvel-1 and IR-44. Bacterial Blight resistant genotypes were; RNT 55-3-2, Ratnagiri-

5, Karjat-184, Karjat-1 and KJT-6. Two genotypes namely; Ratnagiri-60 and Ratnagiri-3 were observed tolerant to drought condition since it showed amplification at specific base pair. Gallmidge resistant genotypes are Pawana, IR-68, Karjat-4, Ratnagiri-711, Abhaya, Panvel-1, IR-44 and RP-4-14observed in the present studies. All the eighteen SSR primers used in this study were also subjected to genetic diversity study and primers were amplified, showed the higher level of polymorphism in rice germplasm. A total of 231 loci were generated by 18 primers. Each primer thus produced on an average 12.83 loci in the size ranging from 169.5 bp to 317.22 bp in the 55 rice genotypes in relation to diversity assessment. UPGMA grouped 55 rice genotypes into two main clusters which were further divided into two sub-clusters.

Okello *et al.*, (2017) carried out study to screen forty eight rice germplasm with 18 simple sequence repeat (SSR) markers to study their genetic diversity and phylogeny structure. Each primer showed 100% polymorphism. A total of 275 alleles were generated by 18 primers and each primer produced on an average 15.27 alleles of the size ranging from 172.22 bp (base pair) to 329.44 bp. The number of alleles amplified for each primer pair was ranged from 5 to 35. The markers pTA-248 generated a maximum number of alleles (35), while the primer RM-309 produced minimum number of alleles (5). The polymorphism information content (PIC) values of primers ranged from 0.58 (RM-206) to 0.85 (RM-140) with an average PIC value of 0.77. It was also observed that there was no correlation between percentage polymorphism and PIC value as SSR primer RM-206 showed minimum PIC value but were 100% polymorphic. The higher the PIC value, the more informative is the primer, thus, primers RM-140 and RM-122 were found to be more informative. All the 48 rice genotypes were separated into two major cluster patterns using Jaccards similarity coefficient matrices.

Vanisri *et al.,* (2018) developed DNA barcodes, the unique pattern of SSR polymorphism from the allelic variation data were developed for 14 visually similar varieties of medium grained rice and eight varieties of long slender grain types, having high market demand. Single-tube multiplex assays and DNA barcodes were generated using the available data on 32 markers for medium slender and 35 markers for long slender varieties to make variety identification easy and precise ,which can supplement traditional standard practices in determining purity and certification.

CHAPTER III

MATERIALS AND METHODS

The information regarding the materials used and the methods applied during the course of the present study is given further in this chapter. The laboratory work was carried out at Plant Biotechnology Centre, Dapoli and outsourcing work was carried out at Reliance Industries Limited, Mumbai, Maharashtra during the period November 2018 – December 2019.

3.1 Source of plant material:

For the present study, the seeds of 20 rice varieties were obtained from research stations working under Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli. The DNA was extracted from the 10-15 days old seedling of the varieties which were grown in the pots and used for further study. The seeds were collected from:

1. Agriculture Research Station, Shirgaon, Dist: Ratnagiri
2. Khar Land Research Station, Panvel, Dist: Raigad
3. Regional Agricultural Research Station, Karjat, Dist: Raigad

The list of varieties is given in Table.3.1 Details of the varieties under the study:

Table no.3.1: Details of rice varieties under study.

Sr. no.	Varieties	Sr. no.	Varieties
1.	KARJAT-1	11.	RATNAGIRI-4
2.	KARJAT-2	12.	RATNAGIRI-24
3.	KARJAT-3	13.	RATNAGIRI-73
4.	KARJAT-4	14.	RATNAGIRI-6
5.	KARJAT-5	15.	RATNAGIRI-7
6.	KARJAT-7	16.	RATNAGIRI-8
7.	KARJAT-184	17.	PANVEL-1
8.	RATNAGIRI-1	18.	PANVEL-2
9.	RATNAGIRI-2	19.	PANVEL-3
10.	RATNAGIRI-3	20.	PALGHAR-2

3.2. Extraction of genomic DNA

3.2.1. Plant material

For the present experimental study, all 20 varieties of rice were sown in the pots and kept in greenhouse. The leaf samples were collected from 10-12 days old seedlings for the extraction of genomic DNA.

3.2.2. DNA isolation kit:

For the isolation of DNA from rice leaves, In nitrogen (Thermo Fisher Scientific) Pure Link ® Genomic DNA Mini kit (K1820-02) was used. The contents of the kit are as follows:

- 50 ml PureLink™ Genomic Lysis/Binding Buffer
- 45 ml PureLink™ Genomic Digestion Buffer
- 50 ml PureLink™ Genomic Wash Buffer 1
- 37.5 ml PureLink™ Genomic Wash Buffer 2
- 50 ml PureLink™ Genomic Elution Buffer
- 5 ml RNase A (20 mg/ml)
- 5 ml Proteinase K (20 mg/ml)
- 5 × 50 each PureLink™ Spin Columns with Collection Tubes
- 5 × 100 PureLink™ Collecion Tubes (2.0 ml)

3.2.3. Protocol for DNA isolation :

The DNA was isolated using the protocol mentioned in the In nitrogen (Thermo Fisher Scientific) Pure Link ® Genomic DNA Mini kit. The newly grown leaf samples about 10-12 days old were taken for DNA isolation. The leafs were surface sterilized with 70% ethanol to prevent contamination. The following protocol was followed :-

- Leaf tissue of about 100mg was taken in 1.5 ml eppendorf tube.

- Crushed the samples with mricotarsals for about 15-20 seconds till leaf gets properly macerated.
- 400 µl digestion solution was added to the macerated leaf tissues followed by 20 µl of Protienase K and 10 µl of RNase.
- Slightly vortexed the tubes to mix the mixture evenly.
- Incubated the tubes at 65 ˚C for 2 hours.
- After incubation, 400 µl of binding buffer was added and inverted the tubes 2-3 times gently.
- Centrifuged the tubes at 10,000 rpm at 10 minutes in microcentrifuge.
- Carefully pipetted the supernatant in new eppendorf tube and added 250 µl of 100% ethanol mix completely by inversion.
- Added the lysate to spin column and centrifuged at 13,000 rpm for 1min. Discarded the collection tube.
- Added 600 µl of Wash Solution I and centrifuged at 13,000 rpm for 1 min.
- Added 600 µl of Wash Solution II and centrifuged at 13,000 rpm for 1 min.
- Dry spin for two minutes (max rpm around 14,0000 rpm).
- Eluted in 50 µl of molecular grade water.

3.2.4. DNA purification

Purification of DNA in samples was done to remove RNA, proteins and polysaccharides which were the major contaminants. RNA was removed by adding RNase to the DNA sample during the isolation process and incubated at 65 ˚C.

3.2.5.DNA quantification through agarose gel electrophoresis

The isolated DNA samples were stored at -20 ˚C. To determine the presence or absence of the DNA after isolation protocol, the samples were visualized on agarose gel of concentration 0.8 percent with standard DNA i.e. 1kb plus DNA gene ruler performing agarose gel electrophoresis keeping the voltage about 50 Volts by comparison of the intensity of staining with Gel Red.

3.2.6. DNA quantification using Nano Drop Spectrophotometer

The quality and quantity of DNA measured using a Nanodrop 2000 Spectrophotometer [2000/2000cc, Thermo Fisher Scientific] by calculating DNA concentration and absorbance ratio at both 260/280 nm and 230/260 nm (260/280 ratio ~1.8 and above considered as pure,25-50 ng/μl).

3.3. Introduction to DNA barcoding markers

Several attractive chloroplast encoded barcoding regions were proposed for plants (Pennisi,2007) including proteins of several protein - coding genes (mat k, rbcL rpoB and rpoc1) and intergenic spacers (atpf - atph, trnh-psbA, and psbK-psbI) (Kane,N.et.al.2012) In DNA barcoding, the unique nucleotide sequence patterns of small DNA fragments (400-800bp) are used as specific reference. Collections to identify specimens and to discover over looked species (Vijayan *et.al.* , 2010)

Important characteristics/features of DNA barcoding are its university, specificity on variation and easiness or employment. An ideal DNA barcode should also be rountinley retrievable with a single primer pair and should be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces (Vijayan K.et.al.2010). The primers used for the recent study have been briefly explained below.

- **matK :**

This chloroplast genes is one of the most rapidly evolving genes and has a length of about 1550 bp. It encodes the enzyme maturase which is involved in the splicing of Type-II introns from RNA transcripts. The main advantage of this marker is rapid evolution and the ubiquitous presence in plants and hence is used to construct plant phylogeneies. The primer set matK390F-matK1326R amplify a DNA fragment of ~930 bp between positions 429 and 1313 of matK gene and mostly suitable for all angiosperms. Observing the high universality and species discrimination, CBOL recommended matK in combination with rbcL as the standard two-locus barcode for plants.

- **rbcL:**

This was the first gene that was sequenced from the plants. Among plastid genes, rbcL is the best characterized gene sequence which encodes the large subunit of rubilose-1,5-biphosphate carboxylase / oxygenase (RUBISCO). GenBank already records for more than 10,000 rbcL segments. CBOL recognized rbcL as one of the most potential gene sequences for DNA barcoding in plants due to its ease in PCR amplification across wide plant groups and availability of sequence information in many plant groups.

- **trnH-psbA (intergenic sequences):**

This intergenic sequences is one of the most variable genome segments in the chloroplast of angiosperms. It has an average length of approximately 450 bp, but varying from 296 to 1120 bp based on available data. This marker has a highest species discrimatory power and hence is preferred as supplementary locus.

- **atpH:**

This gene encodes ATP synthase subunits CFO II. It is generally found as supplementary locus in combination with matK for barcoding in plants.

- **ndhA:**

The NAD(P)H dehydrogenise complex is encoded by 11 ndh genes in plant chloroplast (cp) genomes. This gene codes for functional respiratory protein complex of size ~550 kDa with the mature chloroplast.

- **ITS2:**

These gene is the internal spacer regions of nuclear ribosomal cistron. rDNA posses two internal transcribed spacers (ITS1 and ITS2), present on either side of 5.8S region. The location of the gene in the nuclear genome is Chromosome XII. The small size of ITS2 facilitates its amplification by universal primers, even in samples with partially degraded DNA.As ITS2 has evolved, which leads to a homogenization of all the copies of this gene throughout the genome and in most organisms ITS2 was treated as a single locus. Thus, the ITS2 region might be suitable marker for taxonomic classification.(Han, J.,2013).

3.3.1. Selection of markers for barcoding

For barcoding, short DNA sequences (700-800bp) are used. These sequences are the regions from chloroplast genomes of the plants or mitochondrial genomes in case of animals. To study plant barcoding generally the regions of rbcL, matK, trnH-psbA which cover most of the part of the chloroplast genome. Thus, the ideal DNA barcoding marker should be variable, standardized, phylogenetically informative, extremely robust and short. (Nielsen *et.al.*, 2006).

The ideal DNA barcoding system should meet the following criteria:

i. The gene region sequenced should be nearly identical among individuals of the same species, but different between species.

ii. It should be standardized, with the same DNA region used for different taxonomic groups.

iii. The target DNA region should contain enough phylogenetic information to easily assign unknown or not yet 'barcoded' species to their taxonomic group (genus, family, etc.).

iv. It should be extremely robust, with highly conserved priming sites and highly reliable DNA amplifications and sequencing. This is particularly important when using environmental samples, where each extract contains a mixture of many species to be identified at the same time.

v. The target DNA region should be short enough to allow amplification of degraded DNA. Usually, DNA regions longer than 150 bp are difficult to amplify from degraded DNA. (Taberlet, P. *et al.* 2007, Valentini *et.al.*2008)

For the following study, the markers selected are given below (**Table no.:3.2)**

Table no. 3.2: Primers used for this study.

Sr. no.	Locus	5'→3' Sequence (Forward Primer and Reverse Primer)	Genomic Source	Reference
1.	**matK**	matK-f: CGATCTATTCATTCAATATTTC matK-r: TCTAGCACACGAAAGTCGAAGT	Plastid	Singh,J. *et al.* (2017)
2.	**rbcL**	rbcL-f: ATGTCACCACAAACAGAAAC rbcL-r: TCGCATGTACCTGCAGTAGC	Plastid	Singh,J. *et al.* (2017)
3.	**psbA-trnH**	psbA-trnH-F: GTTATGCATGAACGTAATGCTC psbA-trnHR: CGCGCATGGTGGATTCACAATTC	Plastid	Singh,J. *et al.* (2017)
4.	**trnH - psbA**	trnH-f: CGCGCATGGTGGATTCACAAATC psbA-r: TGCATGGTTCCTTGGTAACTTC	Plastid	Singh,J. *et al.* (2017)
5.	**atpH-atpI**	atpH-f: AACAAAAGGATTCGCAAATAAAAG atpI-r: AGTTGTTGTTCTTGTTTCTTTAGT	Plastid	Singh,J. *et al.* (2017)
6.	**ITS**	ITS2-F: GCGATACTTGGTGTGAAT ITS2-R: GACGCTTCTCCAGACTACAAT	Nuclear	
7.	**petA-psbJ**	petA-f: GGATTTGGTCAGGGAGATGC psbJ-r :ATGGCCGATACTACTGGAAGG	Plastid	Singh,J. *et al.* (2017)
8.	**trnK**	trnK-f: GGGACTCGAACCCGGAACTA trnK-r: AGTACTCGGCTTTTAAGTGCG	Plastid	Singh,J. *et al.* (2017)
9.	**ndhA**	ndhA-f: TCAACTATATCAACTGTACTTGAAC ndhA-r: CGAGCTGCTGCTCAATCGAT	Plastid	Singh,J. *et al.* (2017)
10.	**matK-1RKIM & matK-3FKIM**	matK-1RKIM-f: ACCCAGTCCATCTGGAAATCTTGGTTC matK-3FKIM-r: CGTACAGTACTTTTGTGTTTACGAG	Plastid	Alphanso,P. (2014)

[Table reference : Singh,J. *et al.* (2017),Hollingsworth,P.M. *et.al.*(2011)]

3.3.2. Requirements

The standard requirements for the PCR amplification are given in detail in Table 3.3.

Table 3.3: PCR Stock Solutions and their Sources

Stock solutions	**Source**
Taq One Master Mix 2X MM w/std Buffer Quick Load	BioLabs, New England.
Primers (Fw./Rev.)	Sigma, Merck KGaA,United States of America
Phusion® High-Fidelity DNA Polymerase	BioLabs, New England .

Thermal cycler: T100™ Thermal Cycler by Bio-Rad [BioRad Laboratories, California, USA] was used for cyclic amplification of DNA.

3.3.2. Preparation of master mixture

Ready master mix was added with the primers and water to balance the mixture just prior before setting up a PCR reaction cycle. Since the pipetting of small volumes is difficult and often inaccurate, a master mix was prepared where constituents common to all the reactions were combined in one tube multiplying the volume for one reaction with total number of samples. Later, the appropriate amount of master mix was dispensed to each tube and template DNA was added separately in each tube. (Table no.3.4)

Table No. 3.4(A & B): Master mixture for polymerase chain reaction (PCR):

A) Using Taq One Master Mix

Components	**Vol. for one reaction/10 µl**
Taq One Master Mix	5 µl
Primer F	1 µl
Primer R	1 µl
DNA template	1
Molecular grade water /Distilled water	2 µl
Total	10

B) Using Phusion® High-Fidelity DNA Polymerase

Components	**Stock concentration**	**Vol. for one reaction per 25 µl**
Phusion buffer	5X	5 µl
dNTPs	10mM	0.5 µl
Primer F	10mM	1 µl
Primer R	10mM	1 µl
Phusion DNA polymerase	2,000 U/ml	0.25 µl
DMSO	-	0.75 µl
DNA template	50ng/ µl	2 µl
Molecular grade water /Distilled water	-	16.5 µl
Total		**25 µl**

For sequencing procedure, the concentration of the DNA in PCR product should be more 30 ng/ml. Hence, a reaction of 50 µl volume was amplified in PCR.

3.3.3 Thermal cycling

- Sterile micro centrifuge tubes were numbered from 1 to 20.
- 1.0 µl of template DNA from individual rice genotype was added to each tube.
- When master mix was made using Taq One ready mix 9 µl of was added whereas when master mix made using Phusion® High-Fidelity DNA Polymerase 23 µl of master mix was added to each tube and was given short spin to mix the contents.
- The tubes were placed in the thermal cycler for 35 cycles of PCR. Samples were held at 4^{0}C in the thermal cycler until the contents were loaded on to the gel for electrophoresis.

Table No. 3.5: Thermo Profile

Steps	**Temp. (ºC)**	**Period (min.)**
Initial Denaturation	94°C	5 min.
Denaturation	94°C	1 min
Annealing	48°C - 62°C*	30sec
Extension	72°C	1 min
Final extension	72°C	7 min.
Hold	4°C	-

(*Note: - Annealing temperature based on Tm value of each primer)

3.3.4. Optimization of annealing temperature for gene specific primers by PCR amplification studies

Gradient PCR amplification of gene specific primers was carried out so as to determine the annealing temperature of each primer. The PCR programme was set in Thermal Cycler BIO-RAD, Master Cycler Gradient, made in California, USA.

The annealing temperature of each primer was standardized based on Tm value of each forward and reverse primer and standardized temperature was used for further amplification (Table 3.6).

Table No. 3.6: Annealing temperatures of Barcoding markers used in the study.

Sr. No.	Primer	Tm value (^{0}C)		Temperature range (^{0}C)	Standardized Annealing temperature (^{0}C)
		Reverse primer	Forward primer		
1.	matK	62.9	54.5	48-55	51.5
2.	rbcL	64.0	57.4	55-60	60
3.	psbA-trnH	74.0	61.0	55-65	61
4.	trnH-psbA	64.9	74.0	55-65	60.9
5.	atpH-atpI	56.9	63.0	55-60	57
6.	ITS2	60.0	55.8	48-55	50
7.	petA-psbJ	63.9	65.7	55-65	60
8.	trnK	62.2	67.3	50-60	55
9.	ndhA	67.9	55.8	50-65	58
10.	matK1RIM - matK3RIM	60.6	72.1	48-55	48

3.3.5. Separation of amplified product by agarose gel electrophoresis

- **Requirements:**

  a) Electrophoretic unit (Gel casting tray, gel comb, power pack)
  b) Gel documentation system (Bio-Rad)
  c) Agarose
  d) Tracking dye (Bromophenol blue)
  e) Gel Red (Gel staining dye)
  f) 50 x TAE buffer.
  g) DNA ladder

- **Procedure**

The amplified products in SSR reaction were separated by electrophoresis in 1.5 per cent agarose gel (Sigma-Aldrich) containing Gel Red in 1x TAE Buffer (pH 8.0) and separation was carried out by applying constant voltage of 120 volts for 60 mins.

3.3.6 Photography and gel documentation

The gels were photographed under Trans-UV light using CCD camera. The images of gel was taken by the documentation systems (BIO-RAD Gel Doc XR^+ System, California, United States of America) and saved in computer for further analysis.

3.3.7. Molecular screening of varieties

The gels were carefully studied and amplicons which occurred only once for a particular genotype were marked which constituted the band for that particular genotype. Additionally, the amplicons size for the particular primers was noted and considering the amplicons size was future sent for sequencing.

3.4. PCR purification & Gel elution

The PCR product (amplicons) obtained after amplification were visualized on the gel and confirmed for a single band which is the prerequisite for sequencing to be done efficiently. To clean the PCR product, pcr purification protocol was carried out as per given in the kit. While some primers showed multiple banding pattern after amplification and hence to purify the pcr products gel elution was done.

A) PCR purification protocol:

For the purification, Zymo Research DNA Cleaner & Concentrator Kit was used.

Contents of the Kit:

- DNA Binding Buffer
- DNA Wash Buffer
- Collection tubes
- Zymo-Spin V columns w/v reservoir
- DNA Elution Buffer

Protocol for Purification:

- Take the PCR product amplified with primers in an Eppendorf tube and add 5 times Binding Buffer as that of volume of PCR product (250 µl) and mix carefully by pipetting.
- Add this mixture in spin columns with collection tubes and spin the columns in centrifuge at 12,000 rpm for 1 min.
- Discard the liquid in the collection tube after centrifugation.

- Add Wash Buffer to the spin columns and mix by gentle pipetting.
- Spin the columns in centrifuge at 12,000 rpm for 1min.
- Discard the liquid in the collection tube and dry spin at maximum rpm (~14,000 rpm) for 2 mins.
- Remove the spin column in a fresh eppendorf tube and add molecular grade water (preheated at 70°C) half the volume of the initial PCR product to the spin columns and incubate at room temperature for 5 mins.
- Eloute the DNA in water spinning the columns in centrifuge at 14,000 rpm for 1 min and store at -20 °C.

B) Gel Elution Protocol

The gel extraction procedure was carried out using QIAGEN PCR Gel Elution Kit.

Contents of kit:

- QIAquick Spin Columns
- Buffer QG
- Buffer PE (concentrate)
- Buffer EB
- Collection Tubes
- Loading Dye
- Quick-Start Protocol

Protocol for Gel Extraction:

The PCR products (50 μl) was loaded on 0.8% agarose gel and gel electrophoresis was carried out at 120 volts. The gel was observed under BIO-RAD Gel Doc XR$^+$ System and band were excised from the gel taking care of all the safety precautions. The protocol was followed as per the kit which is as follows.

- Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.
- Weigh the gel slice in a colorless tube. Add 3 volumes Buffer QG to 1 volume gel (100 mg gel ~100 µl). The maximum amount of gel per spin column is 400 mg. For >2% agarose gels, add 6 volumes Buffer QG.
- Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). Vortex the tube every 2–3 min to help dissolve gel. After the gel slice has dissolved completely, check that the color of the mixture is yellow (similar to Buffer QG without dissolved agarose). If the color of the mixture is orange or violet, add 10 µl 3 M sodium acetate, pH 5.0, and mix. The mixture turns yellow.
- Add 1 gel volume isopropanol to the sample and mix.
- Place a QIAquick spin column in a provided 2 ml collection tube or into a vacuum manifold. To bind DNA, apply the sample to the QIAquick column and centrifuge for 1 min or apply vacuum to the manifold until all the samples have passed through the column. Discard flow-through and place the QIAquick column back into the same tube. For sample volumes >800 µl, load and spin/apply vacuum again.
- If DNA will subsequently be used for sequencing, in vitro transcription, or microinjection, add 500 µl Buffer QG to the QIAquick column and centrifuge for 1 min or apply vacuum. Discard flow-through and place the QIAquick column back into the same tube.
- To wash, add 750 µl Buffer PE to QIAquick column and centrifuge for 1 min or apply vacuum. Discard flow-

through and place the QIAquick column back into the same tube. Centrifuge the QIAquick column in the provided 2 ml collection tube for 1 min to remove residual wash buffer.

- Place QIAquick column into a clean 1.5 ml microcentrifuge tube.
- To elute DNA, add 50 µl Buffer EB (10 mM Tris · Cl, pH 8.5) or water to the center of the QIAquick membrane and centrifuge the column for 1 min. For increased DNA concentration, add
  30 µl Buffer EB to the center of the QIAquick membrane, let the
  column stand for 1 min, and then centrifuge for 1 min. After the addition of Buffer EB to the QIAquick membrane, increasing the incubation time to up to 4 min can increase the yield of purified DNA.
- If purified DNA is to be analyzed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel.

3.5 Sequencing of the DNA samples

After the PCR product purification and the gel extraction protocols, the samples were gel visualized using gel electrophoresis loading 2 µl of purified product in 5 µl water and 3 µl loading dye in the gel at 120 volts. And later observed the gel in gel documentation system to verify the presence of single band in the purified samples. The purified samples were quantified using Nanodrop and the samples which should DNA concentration above 20ng/µl were outsourced for sequencing.

The sequencing was done by Sanger's Sequencing Method and the sequenced data was obtained in FASTA Format. The data was segregated and categorized as per the primers used.

3.5.1. Use of Bioinformatic tools:

All the forward and reverse sequences were grouped. For the purpose of multiple sequence alignment, reverse complimentary of the reverse sequences is the prerequisite. To make reverse complimentary of the reverse sequences the online tool **Reverse Complement (http://www.bioinformatics.org/sms/rev_comp.html)**

The consensus sequences were made using a online bioinformatics tool called **CAP3 Sequence Assembly Program-PRABI-Doua (http://doua.prabi.fr/software/cap3)**

The sequences obtained were blasted within NCBI BLAST to check whether the sequenced samples resemble to the species data already present with in repository.

The forward and reverse complimentary sequence data was entered in the **MEGA 7.0** software for Multiple Sequence Alignment and the sequences were aligned and analyized.

The phylogenetic tree was obtained by Neighbour Joining Phylogenetic tree in **MEGA 7.0** and **CLUSTAL OMEGA** online software tool.

DNAsp v6 was used to study the sequence polymorphism in the sequences under analysis. Nucleotide diversity, parsimony informative sites, variations and conserved region were analysed in through this software. NCBI BLAST was used to retrieve sequences for alignments. (Table no. 3.7).

Table no. 3.7: Bioinformatic tools used in the present study

Sr. no.	Bioinformatics tools	Use	References
1.	**Reverse Complement (http://www.bioinformatics.org/sms/rev_comp.html)**	To prepare reverse compliments of reverse sequences	-
2.	**CAP3 Sequence Assembly Program-PRABI-Doua (http://doua.prabi.fr/software/cap3)**	To prepare consensus sequences from forward and reverse sequences	-
3.	**MEGA 7.0.25. Software**	To analyize the sequence aligments	Singh, *et.al*,2017
4.	**CLUSTAL Omega**	To align sequences	Singh, *et.al*,2017
5.	**DNAsp v6**	To estimate nucleotide diversity (per site Pi)	-
6.	**NCBI BLAST**	For sequence database	NCBI portal
7.	**BIORAD DNA Generator**	To generate DNA barcodes	-

3.6 Scoring and Data Analysis

Each amplification product as a dominant expression was considered as barcode marker and was scored across all the samples. Bands were scored as present (+/1) or absent (-/0).The size of each allele was determined by running simultaneously a DNA ladder by using a software (Uvitec, Fire-reader software version 15.12). The data was used for similarity based analysis using the programme MVSP-A (Multivariate Statistical Package_5785 Version 3.1). Similarity coefficients were used to construct UPGMA (unweighted pair group method with average) to generate dendrogram.

Distance matrix and dendrogram was constructed based on diversity coefficient generated from pooled data by using Unweighted Pair Group Method of Arithmetic Means (UPGMA), a computer programme for distance estimation.

The polymorphism percentage of the obtained bands was calculated by using following formula,

$$\text{Per cent polymorphism} = \frac{\text{Total number of polymorphic bands}}{\text{Total number of bands}} \text{ X } 100$$

3.7 Polymorphism Information Content

Polymorphism Information Content (PIC) value were calculated as per formula developed by Powell *et al.,* (1996)

$$\mathbf{PIC = 1 - \sum P_{ij}^2}$$

Where,

P_{ij} is the frequency of i^{th} and j^{th} locus, summed across the entire locus over all lines.

PIC values range from 0 (monomorphic) to 1 (very highly discriminative, with many alleles each in equal and low frequency) were estimated for each profile generated across 50 rice genotypes.

CHAPTER IV

EXPERIMENTAL RESULTS

The present study entitled "DNA Barcoding for identification of Rice (*Oryza sativa* L.) Varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli." was carried out with view to identify the varieties by developing a barcode by sequencing and use of bioinformatics tools. The experimental results obtained in this study are presented in this chapter under different headings.

In this study, 10 barcoding specific markers were used to screen the set of 20 rice varieties to develop a molecular tag "barcode" as well as identify the barcoding gap. The polymorphic information produced by these markers between different rice varieties helps us in analysing the various nucleotide changes occurring at the molecular level focusing the plastid (cpDNA).

4.1. Morphological Characterization

Morphological characters were considered from the previous studies on the varieties mentioned in this present study. The information on this account was retrieved from the Rice varieties Report from University portal. The salient features of developed varieties are mentioned in the Table 4.1 and morphological variations in seed were put forth in Plate 1.

Characters:

- The varieties are mainly characterized on the basis of grain size and shape.
- They are long bold, short bold, long slender, short slender.
- They were developed considering the geographical conditions like high rainfall areas as well as salinity and drought tolerance.

- Two main leaf colours were observed light green and dark green.
- The seed colour varied from light ochre to dark ochre yellow.

4.2. Molecular Characterization Based on plastid genome primers (chloroplast genome)

4.2.1. DNA Quantification Values Obtained From Nanodrop :

After the isolation protocol was completed the DNA samples of the different varieties were gel electrophoresed on 0.8% agarose gel for DNA and 1.5% for PCR products and quantified using a Nanodrop 2000 Spectrophotometer [2000/2000cc, Thermo Fisher Scientific] by calculating DNA concentration and absorbance ratio at both 260/280 nm and 230/260 nm (260/280 ratio ~1.8 and above considered as pure, 25-50 ng/µl. Among the 20 rice varieties 16 varieties yielded desired quantity and quality of DNA and therefore only 16 varieties taken for further studies. The gel photograph of isolated DNA samples is mentioned in Plate no:2 and the quantified DNA samples of all 16 rice varieties is put forth in Table no:4.2.1.

Table 4.2.1.DNA Quantification of isolated DNA samples

Sr. No.	Sample	Concentration (ng/µl)	Ratio (Absorbance at 260/280 nm)
1	KARJAT-1	172.9	1.84
2	KARJAT-2	288.9	1.91
3	KARJAT-3	200.4	1.88
4	KARJAT-4	293.5	1.88
5	KARJAT-5	242.1	2.05
6	KARJAT-7	105.7	1.91
7	RATNAGIRI-1	118.5	1.95
8	RATNAGIRI-2	146.5	1.85
9	RATNAGIRI-3	50.7	1.77
10	RATNAGIRI-4	144.2	1.85
11	RATNAGIRI-24	76.0	1.79
12	RATNAGIRI-7	220.0	2.09
13	RATNAGIRI-8	30.8	1.77
14	PANVEL-1	106.2	1.65
15	PANVEL-2	64.9	1.88
16	PALGHAR-2	35.4	1.85

The concentration of the DNA samples was adjusted with water and diluted to a concentration of 25 ng/µl. The same dilutions were used for the PCR amplification process.

4.2.2. DNA Quantification Of The PCR Product Obtained From Amplification Of Different Primers Under Study:

The amplified products obtained from the different markers were quantified individually and the concentration standardized as a requirement for efficient sequencing was about equal to or more than 25 ng/µl. The range of concentration and ratio of absorbance at 260/280 nm was recorded on Nanodrop and is presented in Table no:4.2.2

Table. 4.2.2: Quantification of PCR products of 10 barcoding primers.

Sr. No.	Primers	Range of Concentration (ng/µl)	Range of Ratio (Absorbance at 260/280 nm)
1.	matK	22.2-43.1	1.56 - 1.79
2.	rbcL	20.5 – 30.9	1.79 -1.99
3.	*psbA-trnH*	14.9 -23.6	1.59 – 1.73
4.	trnH - psbA	20.3 -38.5	1.62 -1.78
5.	atpH-atpI	18.3 – 37.3	1.90 – 2.26
6.	ITS 2	15.0 – 22.6	1.22 – 2.0
7.	petA-psbJ	22.0 – 50.0	1.71 - 2.98
8.	*trnK*	51.9 – 98.8	1.84 – 1.95
9.	ndhA	23.2 – 73.2	1.82 – 1.96
10.	matK-1RKIM & matK-3FKIM	10.5 – 39.4	1.59 -1.82

4.2.3. PCR Amplification:

The PCR thermal cycler yielded amplicons as per the target site amplified by the primer which were further purified and gel eluted and further sent for sequencing. The PCR products were visualized over 1.5% agarose gel and the images of gel was taken by the documentation systems (BIO-RAD Gel Doc XR$^+$ System, California, United States of America).The gel photograph of all the PCR products of 10 barcoding primers is presented in Plate 3.1 to 3.10 and the range of simplified products is mentioned in Table 4.3.

Table 4.3: Range of amplified bands obtained from gel image anlaysis

Sr. no.	**Primers**	**Range of amplified bands (bp)**
1.	matK	560-600
2.	rbcL	831-884
3.	*psbA-trnH*	677-76
4.	trnH - psbA	359-417
5.	atpH-atpI	1350-1477
6.	ITS 2	625-725
7.	petA-psbJ	1240-1372
8.	*trnK*	2240-2811
9.	ndhA	1150-1360
10.	matK-1RKIM & matK-3FKIM	904-1069

4.3. DNA Sequencing:

Sanger's di-deoxy chain termination method produces reads of ~1000 bp long in a single reaction (Chan 2005). Therefore, in the present study, the DNA sequencer based on Sanger's di-deoxy chain termination method (Sanger *et al.* 1977) was used. The PCR amplified products of expected size were extracted from the gel then purified and then subjected to DNA sequencing. The amplified PCR products obtained from the markers under the study were sent for sequencing. A pair of forward and reverse sequences was obtained for each sample after sequencing.

4.4. Sequence Data Analysis

4.4.1. Preparation Of Reverse Complimentary Sequence From The Raw Data:

The sequences obtained after the sequencing by Sanger's sequencing were studied and the reverse compliments of the reverse sequences were prepared. The online bioinformatics tool named **Reverse Complement** (**http://www.bioinformatics.org/sms/rev_comp.html**) was used.

4.4.2. Preparation Of Consensus Sequence:

The forward sequence from the raw data and the reverse complimentary sequences were used to prepare a consensus sequence for each individual sample and for each individual marker set. The consensus sequences were made using a online bioinformatics tool called **CAP3 Sequence Assembly Program-PRABI-Doua** (**http://doua.prabi.fr/software/cap3**) and recorded as CONTIGs under each primer set. Range of forward sequences, reverse sequences and consensus sequences is presented in Table 4.4.

Table 4.4: Size range of forward, reverse and consensus sequence reads of all primers.

Sr. no.	Primers	Forward sequence (bp)	Reverse sequence (bp)	Consensus sequence (bp)
1	matK	874	856-874	909-928
2	rbcL	688-703	678-707	733-799
3	*psbA-trnH*	590- 681	616-874	654-995
4	trnH - psbA	500-874	571-874	908-912
5	atpH-atpI	874	874	1079-1086
6	ITS 2	448-455	414-456	500
7	petA-psbJ	874	874	1170-1850
8	*trnK*	874	874	900
9	ndhA	874	874	985-1000
10	matK-1RKIM & matK-3FKIM	830	839	900

4.5. Sequence Alignments of the varieties:

The consensus sequences obtained for a particular sample and of the particular marker were arranged and .txt file were prepared (Appendix III). All these sequences were then loaded in the CLUSTAL OMEGA software and aligned using software MEGA 7. The sequences aligned and the alignment indicated the places with similarities and point nucleotide differences. The analysis of DNA sequences was conducted by Neighbor-joining to assess topology with MEGA 7. The species identification and homology between the sequences was identified using BLAST method. The phylogenetic tree was developed using Neighbor-Joining (NJ) method which was tested with Kimura 2-parameter for evolutionary distances in MEGA 7.Pairwise distance, transitional/transpersonal substitutions, and the maximum likelihood substitution matrix were estimated using MEGA

7.0.25 software (Table 4.9). The images of alignment of each primer set for the samples under study is mentioned below (Appendix IV). Also phylogenetic trees were constructed for each marker set given as follows.

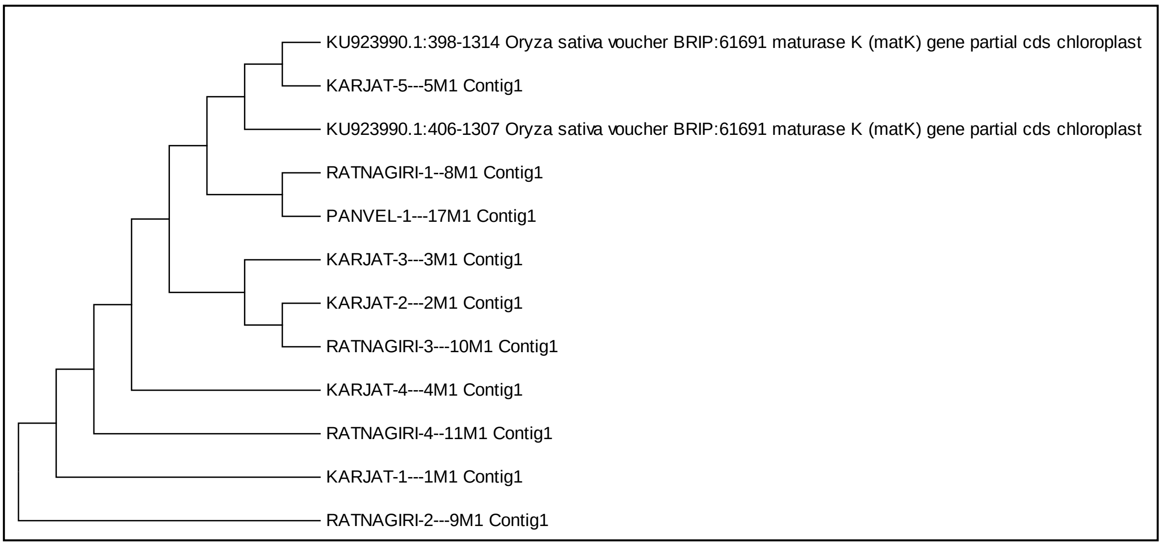
4.5.1. *matK* marker

PCR of *matK* primer amplified a band of 560 to 600 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *matK* is presented in Plate 3.1. The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Panvel-1, Panvel-2.The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length of 874 bp and that of the reverse sequences generated ranged from 856 to 874 bp in length. Around the range of 909-928 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7. Alignment of the sequences helped to visualize variable color coding (Appendix IV-Plate A). The specific variable regions were marked as probable DNA barcode reg ion. Transitions observed were a single bp change of C to T at 96 bp, T to C at 99 bp, a gap of 6 nucleotides at position 84 to 89 bp, another gap of 7 bp at 110 to 116 bp and a presence of A at a nucleotide 100 bp, G to T transition at 102 bp in sample Ratnagiri-2. And in sample Karjat-2 transition T to G at 877 and 889 bp, A to G at 893 bp, A to C at 896 bp. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00-0.021 (Table 4.5.1). P value of zero signifies nearest relative in the evolution where as the 0.021 value signifies distant in evolutionary ladder. K2P estimated overall average of 0.005.Overall average p-distance was 0.0047. Phylogenetic representation was done using Neighbour Joining Tree method (Fig. 4.1)

Table 4.5.1 : Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using matK

	1	2	3	4	5	6	7	8	9	10	11	12
1. KU923990.1:406-1307 Oryza sativa voucher BRIP:61691 maturase K (matK) gene partial cds chloroplast		0.000	0.002	0.003	0.001	0.001	0.000	0.002	0.004	0.001	0.001	0.002
2. KU923990.1:398-1314 Oryza sativa voucher BRIP:61691 maturase K (matK) gene partial cds chloroplast	-0.000		0.002	0.003	0.001	0.001	0.000	0.002	0.004	0.001	0.001	0.002
3. KARJAT-1---1M1 Contig1	0.002	0.002		0.003	0.001	0.001	0.002	0.002	0.004	0.001	0.001	0.002
4. KARJAT-2---2M1 Contig1	0.007	0.007	0.007		0.003	0.003	0.003	0.003	0.005	0.003	0.003	0.003
5. KARJAT-3---3M1 Contig1	0.001	0.001	0.001	0.006		0.000	0.001	0.001	0.004	0.000	0.000	0.002
6. KARJAT-4---4M1 Contig1	0.001	0.001	0.001	0.006	-0.000		0.001	0.001	0.004	0.000	0.000	0.002
7. KARJAT-5---5M1 Contig1	-0.000	-0.000	0.002	0.007	0.001	0.001		0.002	0.004	0.001	0.001	0.002
8. RATNAGIRI-1--8M1 Contig1	0.002	0.002	0.002	0.007	0.001	0.001	0.002		0.004	0.001	0.001	0.002
9. RATNAGIRI-2---9M1 Contig1	0.016	0.016	0.014	0.021	0.015	0.015	0.016	0.016		0.004	0.004	0.004
10. RATNAGIRI-3---10M1 Contig1	0.001	0.001	0.001	0.006	-0.000	-0.000	0.001	0.001	0.015		0.000	0.002
11. RATNAGIRI-4--11M1 Contig1	0.001	0.001	0.001	0.006	-0.000	-0.000	0.001	0.001	0.015	-0.000		0.002
12. PANVEL-1---17M1 Contig1	0.005	0.005	0.005	0.009	0.003	0.003	0.005	0.002	0.019	0.003	0.003	

Figure 4.1: Neighbor joining Phylogenetic Tree of samples sequences of matK primer.



The evolutionary history was inferred using the Neighbor-Joining method .The optimal tree with the sum of branch length = 0.02544554 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 12 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 874 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7 (Fig 4.1).

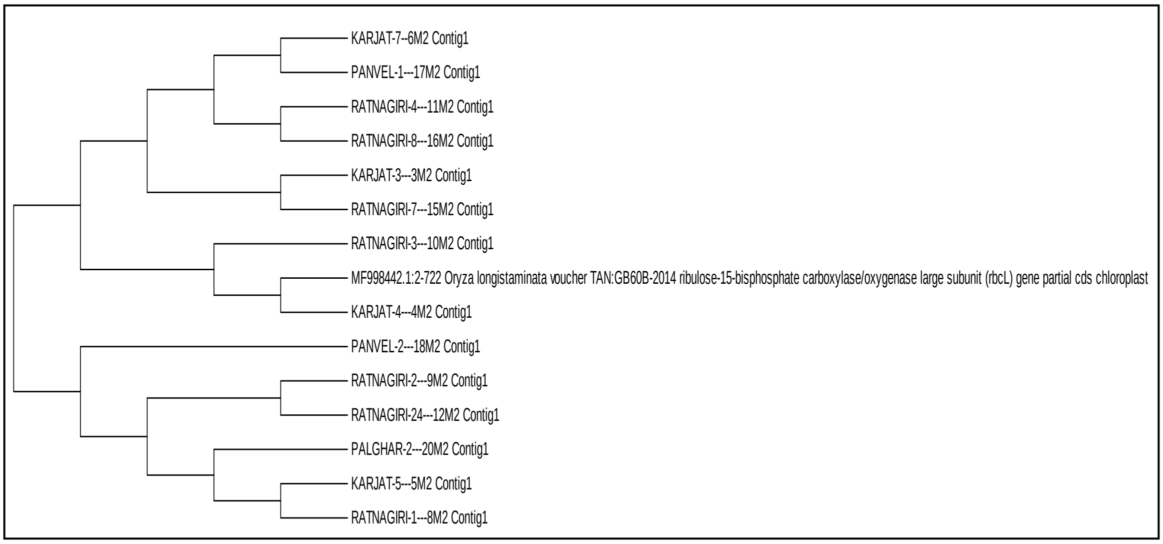
4.5.2. *rbcL*

PCR of *rcbL* primer amplified a band of 831- 884 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *rcbL* is presented in Plate 3.2. The samples amplified were Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3 Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from to 688-703 bp and that of the reverse sequences generated ranged from 678-707 bp in length. Around the range of 733-799 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7. Alignment of the sequences helped to visualize variable color coding (Appendix IV- Plate B). The sequences were aligned with one BLAST sequence obtained from NCBI Blast showing 99.72% identity with the sample sequence. The alignment result showed that the sample sequence got completely aligned with the reference sequence without any transitions or transversions expressing complete conservation of the *rcbL* region. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00- 0.007 (Table 4.5.2). P value of zero signifies nearest relative in the evolution where as the value 0.007 signifies distant in evolutionary ladder. K2P estimated overall average of 0.004.Overall average p-distance was 0.004. Phylogenetic representation was done using Neighbors Joining Tree method.(Fig.no:4.2).

Table 4.5.2: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *rcbL*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. MF998442.1:2-722 Oryza l		0.002	0.001	0.003	0.003	0.003	0.003	0.001	0.002	0.003	0.002	0.002	0.002	0.002	0.003
2. KARJAT-3---3M2 Contig1	0.003		0.001	0.002	0.002	0.002	0.003	0.001	0.001	0.002	0.000	0.001	0.002	0.002	0.003
3. KARJAT-4---4M2 Contig1	0.001	0.001		0.002	0.002	0.002	0.003	0.000	0.002	0.002	0.001	0.002	0.002	0.001	0.002
4. KARJAT-5---5M2 Contig1	0.006	0.004	0.004		0.003	0.000	0.002	0.002	0.003	0.002	0.002	0.003	0.003	0.002	0.002
5. KARJAT-7--6M2 Contig1	0.006	0.004	0.004	0.006		0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003
6. RATNAGIRI-1---8M2 Contig	0.006	0.004	0.004	0.000	0.006		0.002	0.002	0.003	0.002	0.002	0.003	0.003	0.002	0.002
7. RATNAGIRI-2---9M2 Contig	0.007	0.006	0.006	0.004	0.006	0.004		0.003	0.003	0.001	0.003	0.003	0.002	0.002	0.002
8. RATNAGIRI-3---10M2 Cont	0.001	0.001	0.000	0.004	0.004	0.004	0.006		0.002	0.002	0.001	0.002	0.002	0.001	0.002
9. RATNAGIRI-4---11M2 Cont	0.004	0.001	0.003	0.006	0.004	0.006	0.007	0.003		0.003	0.001	0.001	0.002	0.002	0.003
10. RATNAGIRI-24---12M2 Co	0.006	0.004	0.004	0.003	0.004	0.003	0.001	0.004	0.006		0.002	0.003	0.002	0.002	0.002
11. RATNAGIRI-7---15M2 Con	0.003	0.000	0.001	0.004	0.004	0.004	0.006	0.001	0.001	0.004		0.001	0.002	0.002	0.003
12. RATNAGIRI-8---16M2 Con	0.004	0.001	0.003	0.006	0.004	0.006	0.007	0.003	0.001	0.006	0.001		0.001	0.002	0.002
13. PANVEL-1---17M2 Contig1	0.004	0.003	0.003	0.006	0.003	0.006	0.006	0.003	0.003	0.004	0.003	0.001		0.002	0.002
14. PANVEL-2---18M2 Contig1	0.003	0.003	0.001	0.003	0.004	0.003	0.004	0.001	0.004	0.003	0.003	0.004	0.004		0.002
15. PALGHAR-2---20M2 Contig	0.006	0.006	0.004	0.003	0.006	0.003	0.006	0.004	0.006	0.004	0.006	0.004	0.004	0.003	

Figure 4.2 : Neighbor joining Phylogenetic Tree of samples sequences of *rcbL* primer.



The evolutionary history was inferred using the Neighbors-Joining method. The optimal tree with the sum of branch length = 0.01682649 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 15 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 718 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig. 4.2).

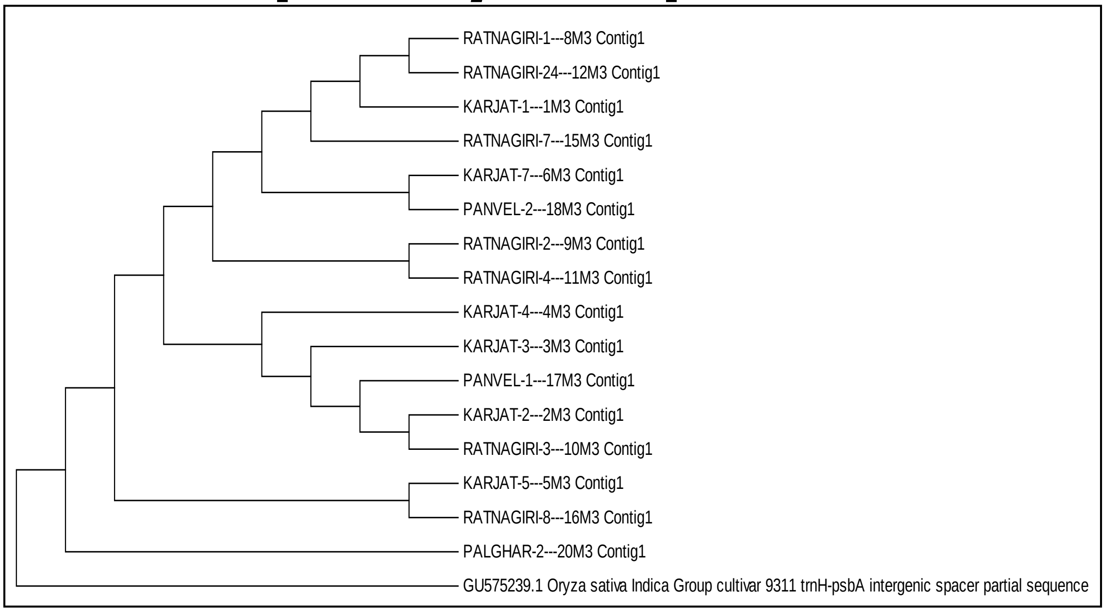
4.5.3. *psbA-trnH*

PCR of primer *psbA-trnH* amplified a band of 560 to 600 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *psbA-trnH* is presented in Plate 3.3. The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from 590-681 bp and that of the reverse sequences generated ranged from 616-874bp in length. Around the range of 654-995 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix 4 -Plate C) The alignment showed results with presence of AG at 325 – 326 bp position, TCTTGAAG at 337 – 334 bp throughout all samples. Absence of gap region TCT at 373 – 375 bp position and presence of GAAGAAAGGAAAG region in all samples which was absent in reference sequence at 398 to 410 bp. C to A transition at 435 bp in samples Karjat-1, Karjat-7, Ratnagiri-1, Ratnagiri-24, Ratnagiri-7, Panvel-2 was observed. Presence of TTTCTTTTTTTTTTTATT region at 453 – 471 bp and TTAGTATCCTTTCTTGCA at position 510 to 527 bp in all samples. A gap region at 596 – 598 bp, 605 – 619 bp, 650 – 655 bp in all samples was observed. Presence of region GGTAGA and GTTTCTTTCTCCTCC at 682 – 687 bp and 718 – 732 bp respectively. A larger gap at 967 – 1018 bp was observed in all samples. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00 – 0.007 (Table 4.5.3). P value of zero signifies nearest relative in the evolution where as the value 0.007 signifies distant in evolutionary ladder. K2P estimated overall average of 0.084. Overall average p- distance was 0.004. Phylogenetic representation was done using Neighbors Joining Tree method. (Fig. 4.3).

Table 4.5.3: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *psbA-trnH*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. GU575239.1 Oryza sativa Indica (		0.055	0.055	0.054	0.054	0.054	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.054	0.055	0.056	0.056
2. KARJAT-1---1M3 Contig1	0.630		0.003	0.003	0.003	0.004	0.004	0.000	0.004	0.003	0.003	0.000	0.002	0.004	0.003	0.002	0.009
3. KARJAT-2---2M3 Contig1	0.626	0.006		0.002	0.002	0.004	0.005	0.003	0.003	0.000	0.003	0.003	0.004	0.004	0.003	0.004	0.009
4. KARJAT-3---3M3 Contig1	0.621	0.004	0.002		0.000	0.003	0.005	0.003	0.003	0.002	0.002	0.003	0.003	0.003	0.002	0.003	0.009
5. KARJAT-4---4M3 Contig1	0.621	0.004	0.002	-0.000		0.003	0.005	0.003	0.003	0.002	0.002	0.003	0.003	0.003	0.002	0.003	0.009
6. KARJAT-5---5M3 Contig1	0.625	0.010	0.008	0.006	0.006		0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.005	0.010
7. KARJAT-7---6M3 Contig1	0.639	0.008	0.014	0.012	0.012	0.012		0.004	0.004	0.005	0.004	0.004	0.004	0.004	0.005	0.003	0.010
8. RATNAGIRI-1---8M3 Contig1	0.630	-0.000	0.006	0.004	0.004	0.010	0.008		0.004	0.003	0.003	0.000	0.002	0.004	0.003	0.002	0.009
9. RATNAGIRI-2---9M3 Contig1	0.631	0.008	0.006	0.004	0.004	0.008	0.008	0.008		0.003	0.002	0.004	0.004	0.003	0.003	0.003	0.009
10. RATNAGIRI-3---10M3 Contig1	0.626	0.006	-0.000	0.002	0.002	0.008	0.014	0.006	0.006		0.003	0.003	0.004	0.004	0.003	0.004	0.009
11. RATNAGIRI-4---11M3 Contig1	0.626	0.006	0.004	0.002	0.002	0.008	0.010	0.006	0.002	0.004		0.003	0.004	0.003	0.003	0.003	0.009
12. RATNAGIRI-24---12M3 Contig1	0.630	-0.000	0.006	0.004	0.004	0.010	0.008	-0.000	0.008	0.006	0.006		0.002	0.004	0.003	0.002	0.009
13. RATNAGIRI-7---15M3 Contig1	0.634	0.002	0.008	0.006	0.006	0.010	0.010	0.002	0.010	0.008	0.008	0.002		0.005	0.003	0.003	0.010
14. RATNAGIRI-8---16M3 Contig1	0.626	0.010	0.008	0.006	0.006	0.004	0.008	0.010	0.004	0.008	0.004	0.010	0.012		0.004	0.004	0.009
15. PANVEL-1---17M3 Contig1	0.626	0.006	0.004	0.002	0.002	0.006	0.014	0.006	0.006	0.004	0.004	0.006	0.006	0.008		0.004	0.009
16. PANVEL-2---18M3 Contig1	0.635	0.002	0.008	0.006	0.006	0.012	0.006	0.002	0.006	0.008	0.004	0.002	0.004	0.008	0.008		0.009
17. PALGHAR-2---20M3 Contig1	0.642	0.045	0.043	0.041	0.041	0.048	0.050	0.045	0.041	0.043	0.039	0.045	0.047	0.043	0.043	0.043	

Figure 4.3: Neighbor joining Phylogenetic Tree of samples sequences of *psbA-trnH* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.16966688 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 17 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 502 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig 4.3)

4.5.4 *trnH - psbA*

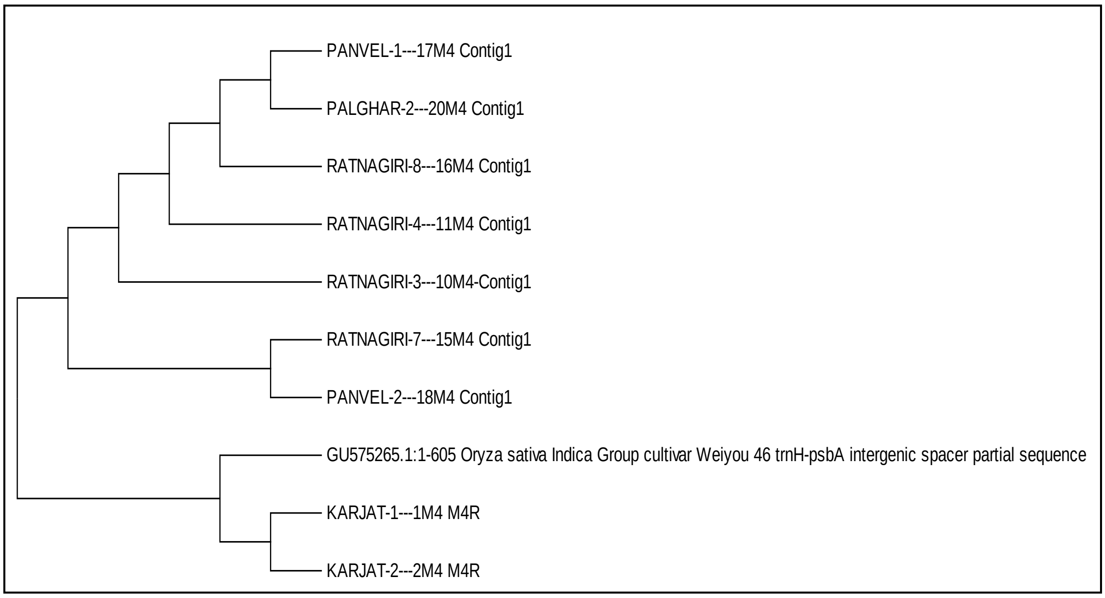
PCR of *trnH-psbA* primer amplified a band of 359- 417bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *trnH-psbA* is presented in Plate 3.4.The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Panvel-1, Panvel-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from 500-874 bp and that of the reverse sequences generated ranged from 571-874 bp in length. Around the range of 908-912 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7(Appendix IV-Plate D) The alignment showed presence of GGCCA region at 72- 76 bp, CCATATGCTAGTTTTA ACAACTCTC at 87 – 112 bp in all samples.

T to A transition at 91 bp position and C to T transition at 378 bp in DPL-10. A gap of 4 nucleotides at 420 – 423 bp and at 429 – 435 bp and the gap of two nucleotides at 457 and 458 bp position except of that in Ratnagiri-3. Presence of GATTTGTTA at 602 – 610 bp and a gap at 531 – 539 bp position was seen in all samples. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00 – 0.650 (Table 4.5.4). P value of zero signifies nearest relative in the evolution where as the value 0.650 signifies distant in evolutionary ladder. Overall average p- distance was 0.165. K2P estimated overall average of 0.0161. Phylogenetic representation was done using Neighbors Joining Tree method. (Fig. 4.4)

Table 4.5.4 Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *trnH-psbA*

	1	2	3	4	5	6	7	8	9	10
1. GU575265.1:1-605 Oryza sativa I		0.053	0.058	0.056	0.056	0.060	0.056	0.056	0.060	0.056
2. KARJAT-1---1M4 M4R	0.224		0.017	0.061	0.061	0.065	0.061	0.061	0.065	0.061
3. KARJAT-2---2M4 M4R	0.264	0.029		0.068	0.068	0.064	0.068	0.068	0.064	0.068
4. RATNAGIRI-3---10M4-Contig1	0.250	0.287	0.331		0.000	0.014	0.000	0.000	0.014	0.000
5. RATNAGIRI-4---11M4 Contig1	0.250	0.287	0.331	-0.000		0.014	0.000	0.000	0.014	0.000
6. RATNAGIRI-7---15M4 Contig1	0.276	0.316	0.302	0.019	0.019		0.014	0.014	0.000	0.014
7. RATNAGIRI-8---16M4 Contig1	0.250	0.287	0.331	-0.000	-0.000	0.019		0.000	0.014	0.000
8. PANVEL-1---17M4 Contig1	0.250	0.287	0.331	-0.000	-0.000	0.019	-0.000		0.014	0.000
9. PANVEL-2---18M4 Contig1	0.276	0.316	0.302	0.019	0.019	-0.000	0.019	0.019		0.014
10. PALGHAR-2---20M4 Contig1	0.250	0.287	0.331	-0.000	-0.000	0.019	-0.000	-0.000	0.019	

Figure 4.4: Neighbor joining Phylogenetic Tree of samples sequences of *trnH-psbA* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.65245219 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 8 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 586 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig. 4.4).

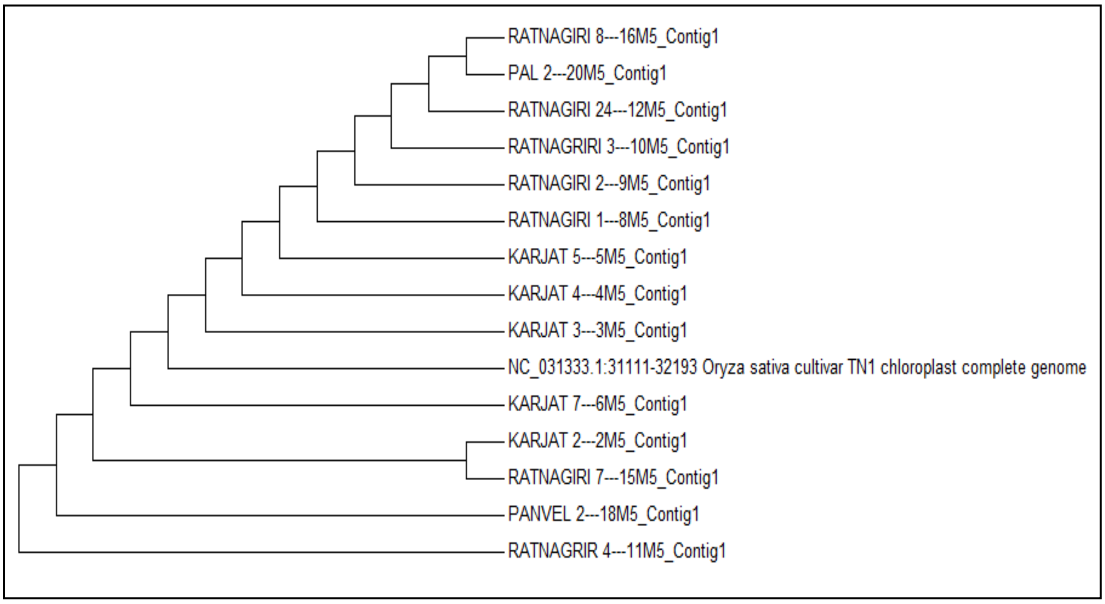
4.5.5 *atpH- atpI*

PCR of primer *atpH- atpI* amplified a band of 1350-1477 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *atpH- atpI* is presented in Plate 3.5.The samples amplified were Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length 874 bp and that of the reverse sequences generated were also 874 bp in length. Around the range of 1079-1086 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix IV-Plate E). The sample Ratnagiri-4 showed transitions T to C at 44 & 45 bp and a gap of 33 nucleotides from 47 bp to 79 bp. Many gaps in the same sample were seen at 161-208 bp, 215 – 237 bp, 259 – 283 bp, 295 – 368 bp, 394 – 470 bp, 553 – 570 bp position. Single nucleotide transitions of T to G at 875 bp, T to A at 876 bp, T to C at 877 bp, presence of CG at 896 and 897 bp in sample Ratnagiri-4 and was absent in other samples. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00 – 0.796 (Table 4.5.5). P value of zero signifies nearest relative in the evolution where as the value 0.796 signifies distant in evolutionary ladder. Overall average p- distance was 0.0109. K2P estimated overall average of 0.076.Phylogenetic representation was done using Neighbors Joining Tree method. (Fig.no4.5)

Table 4.5.5: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *atpH- atpI*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. NC_031333.1:31111-32193 Oryza sativa cultivar TN1 chloroplast comple		0.003	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.056	0.000	0.005	0.000	0.009	0.000
2. KARJAT 2---2M5_Contig1	0.003		0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.056	0.003	0.005	0.003	0.009	0.003
3. KARJAT 3---3M5_Contig1	-0.000	0.003		0.000	0.000	0.003	0.000	0.000	0.000	0.056	0.000	0.005	0.000	0.009	0.000
4. KARJAT 4---4M5_Contig1	-0.000	0.003	-0.000		0.000	0.003	0.000	0.000	0.000	0.056	0.000	0.005	0.000	0.009	0.000
5. KARJAT 5---5M5_Contig1	-0.000	0.003	-0.000	-0.000		0.003	0.000	0.000	0.000	0.056	0.000	0.005	0.000	0.009	0.000
6. KARJAT 7---6M5_Contig1	0.003	0.006	0.003	0.003	0.003		0.003	0.003	0.003	0.056	0.003	0.005	0.003	0.009	0.003
7. RATNAGIRI 1---8M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003		0.000	0.000	0.056	0.000	0.005	0.000	0.009	0.000
8. RATNAGIRI 2---9M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003	-0.000		0.000	0.056	0.000	0.005	0.000	0.009	0.000
9. RATNAGRIRI 3---10M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003	-0.000	-0.000		0.056	0.000	0.005	0.000	0.009	0.000
10. RATNAGRIR 4---11M5_Contig1	0.533	0.532	0.533	0.533	0.533	0.539	0.533	0.533	0.533		0.056	0.056	0.056	0.057	0.056
11. RATNAGIRI 24---12M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003	-0.000	-0.000	-0.000	0.533		0.005	0.000	0.009	0.000
12. RATNAGIRI 7---15M5_Contig1	0.011	0.008	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.543	0.011		0.005	0.009	0.005
13. RATNAGIRI 8---16M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003	-0.000	-0.000	-0.000	0.533	-0.000	0.011		0.009	0.000
14. PANVEL 2---18M5_Contig1	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.553	0.025	0.031	0.025		0.009
15. PAL 2---20M5_Contig1	-0.000	0.003	-0.000	-0.000	-0.000	0.003	-0.000	-0.000	-0.000	0.533	-0.000	0.011	-0.000	0.025	

Figure 4.5 : Neighbor joining Phylogenetic Tree of samples sequences of *atpH- atpI* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.80962785 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 15 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 361 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig 4.5)

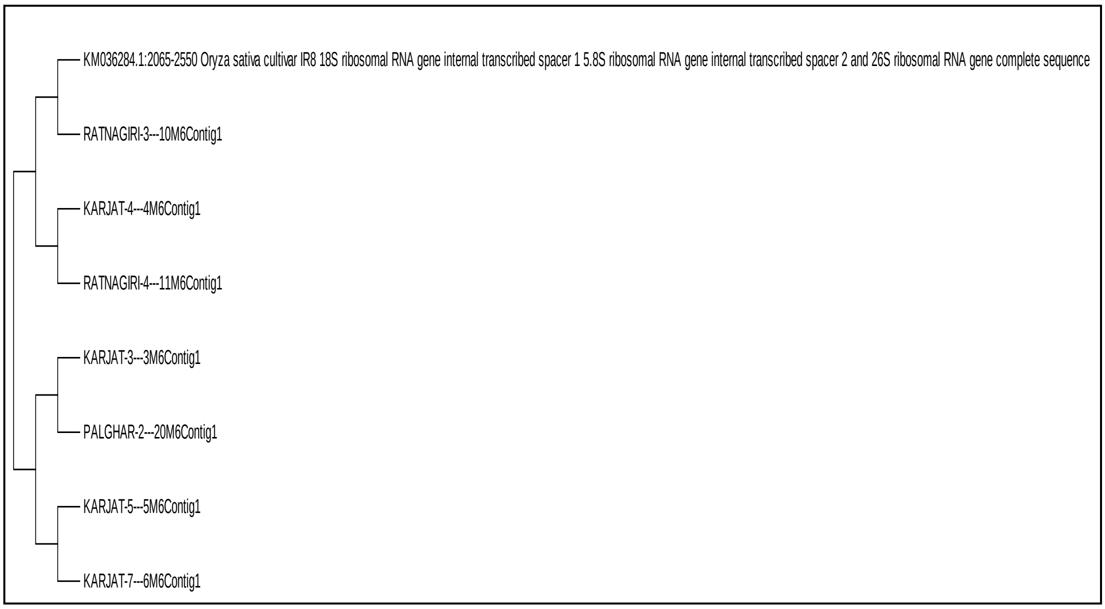
4.5.6. *ITS-2* (Inter- transcribed region 2)

PCR of primer *ITS-2* amplified a band of 625- 725 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *ITS-2* is presented in Plate 3.6. The samples amplified were Karjat-1,Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-3,Ratnagiri-4, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from 448-455 bp and that of the reverse sequences generated ranged from 414-456 bp in length. Around the range of 500 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix IV-Plate F). The result showed perfect alignment of sample sequences with a BLAST sequence with no variations. Only G to C transition was observed at 318 bp position in sample Karjat-5 & Karjat-7. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00 – 0.006 (Table 4.5.6). P value of zero signifies nearest relative in the evolution where as the value 0.006 signifies distant in evolutionary ladder. Overall average p- distance was 0.003. K2P estimated overall average of 0.003. Phylogenetic representation was done using Neighbour Joining Tree method. (Fig 4.6)

Table 4.5.6: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *ITS-2* primer

	1	2	3	4	5	6	7	8
1. KM036284.1:2065-2550 Oryza sativa cultivar I		0.003	0.003	0.003	0.003	0.003	0.002	0.002
2. KARJAT-3---3M6Contig1	0.004		0.003	0.003	0.002	0.000	0.002	0.002
3. KARJAT-4---4M6Contig1	0.004	0.004		0.003	0.004	0.003	0.002	0.002
4. KARJAT-5---5M6Contig1	0.004	0.004	0.004		0.002	0.003	0.002	0.002
5. KARJAT-7---6M6Contig1	0.006	0.002	0.006	0.002		0.002	0.003	0.003
6. PALGHAR-2---20M6Contig1	0.004	-0.000	0.004	0.004	0.002		0.002	0.002
7. RATNAGIRI-3---10M6Contig1	0.002	0.002	0.002	0.002	0.004	0.002		0.000
8. RATNAGIRI-4---11M6Contig1	0.002	0.002	0.002	0.002	0.004	0.002	-0.000	

Figure 4.6: Neighbor joining Phylogenetic Tree of samples sequences of *ITS-2* primer



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.00877021 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 8 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 486 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.(Fig 4.6)

4.5.7. ***petA- psbJ***

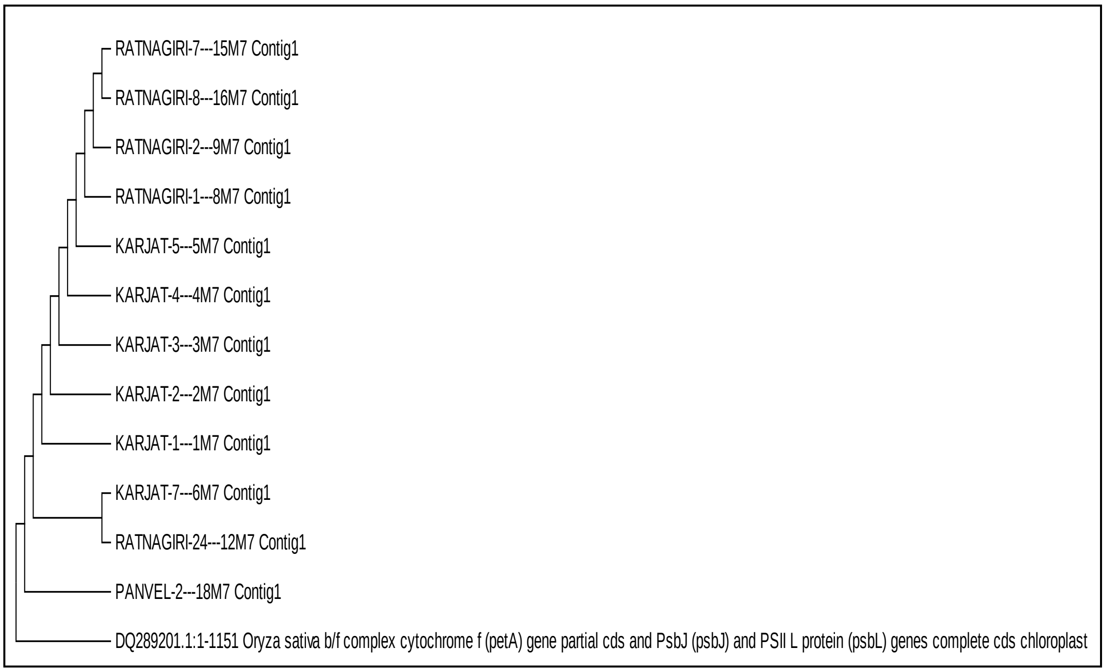
PCR of primer *petA- psbJ* amplified a band of 625-725 bp. The aga rose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *petA- psbJ* is presented in Plate 3.7. The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length 874 bp and that of the reverse sequences generated ranged 874 bp in length. Around the range of 1170-1850 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix IV-Plate G). The result showed a gap at 49 – 56 bp in all samples. G to C transition at 62 bp in Karjat-7 & other samples showed a single nucleotide gap at the same position. Presence of region AGAGCGGAAAGGACCCG at 177-193 bp.

In sample Karjat-7 & Ratnagiri-24 C to A transition at 815 bp and A to C transition at 542 & 560 bp in all samples. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- Values varied from 0.002.170 (Table 4.5.7). P value of zero signifies nearest relative in the evolution where as the value 2.170 signifies distant in evolutionary ladder. Overall average p-distance was 0.335. K2P estimated overall average of 0.095.Phylogenetic representation was done using Neighbour Joining Tree method. (Fig 4.7).

Table 4.5.7: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *petA- psbJ* .

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. DQ289201.1:1-1151 Oryza sativa		0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
2. KARJAT-1---1M7 Contig1	0.617		0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
3. KARJAT-2---2M7 Contig1	0.617	-0.000		0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
4. KARJAT-3---3M7 Contig1	0.617	-0.000	-0.000		0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
5. KARJAT-4---4M7 Contig1	0.617	-0.000	-0.000	-0.000		0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001
6. KARJAT-5---5M7 Contig1	0.617	-0.000	-0.000	-0.000	-0.000		0.001	0.000	0.000	0.001	0.000	0.000	0.001
7. KARJAT-7---6M7 Contig1	0.619	0.002	0.002	0.002	0.002	0.002		0.001	0.001	0.001	0.001	0.001	0.002
8. RATNAGIRI-1---8M7 Contig1	0.617	-0.000	-0.000	-0.000	-0.000	-0.000	0.002		0.000	0.001	0.000	0.000	0.001
9. RATNAGIRI-2---9M7 Contig1	0.617	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000		0.001	0.000	0.000	0.001
10. RATNAGIRI-24---12M7 Contig1	0.616	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001		0.001	0.001	0.001
11. RATNAGIRI-7---15M7 Contig1	0.617	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	-0.000	0.001		0.000	0.001
12. RATNAGIRI-8---16M7 Contig1	0.617	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	-0.000	0.001	-0.000		0.001
13. PANVEL-2---18M7 Contig1	0.617	0.001	0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.002	0.001	0.001	

Figure 4.7: Neighbor joining Phylogenetic Tree of samples sequences of *petA- psbJ* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 2.17282972 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 13 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 944 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig 4.7).

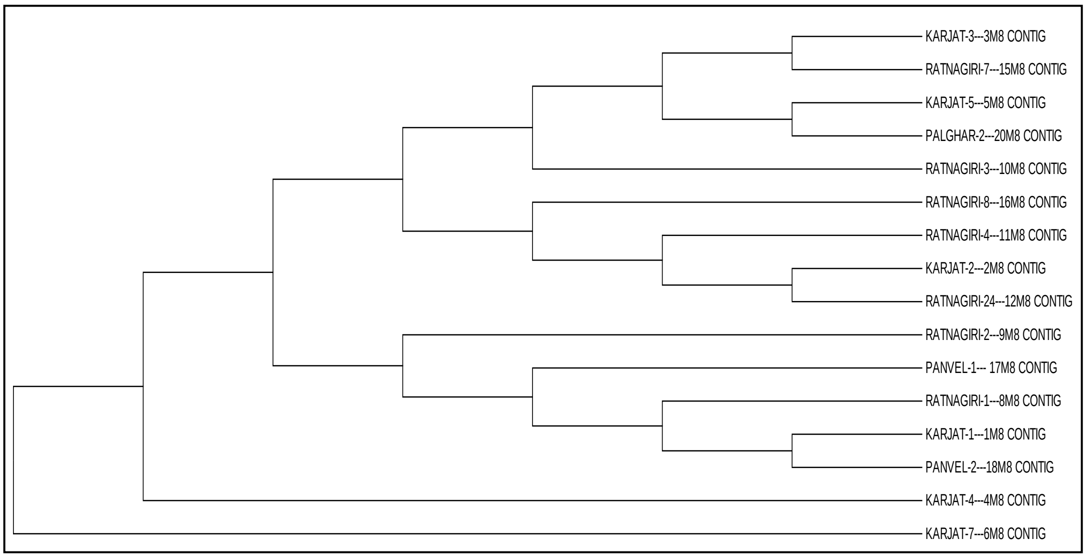
4.5.8. *trnK*

PCR of primer *trnK* amplified a band of 2204-2811 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *trnK* is presented in Plate 3.9 .The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length 874 bp and that of the reverse sequences generated ranged from 874 bp in length. Around the range of 900 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix IV-Plate H). Sample Karjat-4 showed many point nucleotide changes and did not get completely aligned with other sequences. Sample Karjat-7 contained an extra region at 537-564 bp where other samples did not contain the region and a gap was seen at the same position. All other samples were aligned showed similarity throughout. Estimation of evolutionary divergence between all 3 sequences was done by calculating p-values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.008-1.091 (Table 4.5.8). P value of zero signifies nearest relative in the evolution where as the value 0.271 signifies distant in evolutionary ladder. Overall average p- distance was 0.001. K2P estimated overall average of 0.184. Phylogenetic representation was done using Neighbour Joining Tree method. (Fig 4.8)

Table 4.5.8: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *trnK*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. KARJAT-1---1M8 CONTIG		0.032	0.036	0.006	0.010	0.011	0.000	0.015	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.016
2. KARJAT-4---4M8 CONTIG	0.474		0.047	0.032	0.032	0.034	0.032	0.030	0.029	0.030	0.030	0.030	0.030	0.030	0.030	0.031
3. KARJAT-7---6M8 CONTIG	0.564	0.753		0.037	0.034	0.040	0.036	0.033	0.033	0.034	0.033	0.033	0.033	0.033	0.033	0.034
4. RATNAGIRI-1---8M8 CONTIG	0.026	0.489	0.558		0.010	0.009	0.006	0.015	0.016	0.016	0.016	0.015	0.015	0.016	0.015	0.016
5. RATNAGIRI-2---9M8 CONTIG	0.067	0.481	0.530	0.069		0.014	0.010	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
6. PANVEL-1--- 17M8 CONTIG	0.091	0.531	0.601	0.063	0.130		0.011	0.016	0.016	0.017	0.016	0.016	0.016	0.016	0.016	0.017
7. PANVEL-2---18M8 CONTIG	-0.000	0.474	0.564	0.026	0.067	0.091		0.015	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.016
8. KARJAT-2---2M8 CONTIG	0.151	0.439	0.503	0.153	0.094	0.178	0.151		0.004	0.004	0.003	0.001	0.002	0.004	0.001	0.004
9. KARJAT-3---3M8 CONTIG	0.154	0.442	0.505	0.156	0.087	0.179	0.154	0.011		0.003	0.003	0.003	0.004	0.002	0.003	0.003
10. KARJAT-5---5M8 CONTIG	0.159	0.449	0.516	0.161	0.091	0.186	0.159	0.014	0.007		0.003	0.004	0.004	0.002	0.004	0.003
11. RATNAGIRI-3---10M8 CONTIG	0.153	0.442	0.505	0.154	0.091	0.179	0.153	0.009	0.007	0.008		0.003	0.004	0.002	0.003	0.004
12. RATNAGIRI-4---11M8 CONTIG	0.149	0.437	0.500	0.151	0.093	0.176	0.149	0.001	0.009	0.012	0.008		0.002	0.003	0.000	0.004
13. RATNAGIRI-24---12M8 CONTIG	0.153	0.442	0.500	0.154	0.096	0.179	0.153	0.004	0.012	0.015	0.011	0.003		0.004	0.002	0.004
14. RATNAGIRI-7---15M8 CONTIG	0.154	0.442	0.508	0.156	0.087	0.181	0.154	0.011	0.003	0.004	0.004	0.009	0.012		0.003	0.004
15. RATNAGIRI-8---16M8 CONTIG	0.149	0.437	0.500	0.151	0.093	0.176	0.149	0.001	0.009	0.012	0.008	-0.000	0.003	0.009		0.004
16. PALGHAR-2---20M8 CONTIG	0.164	0.456	0.519	0.166	0.096	0.188	0.164	0.014	0.009	0.008	0.015	0.012	0.015	0.011	0.012	

Figure 4.8 : Neighbor joining Phylogenetic Tree of samples sequences of *trnK* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.74894446 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 16 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 746 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Fig 4.8).

4.7.9. *ndhA*

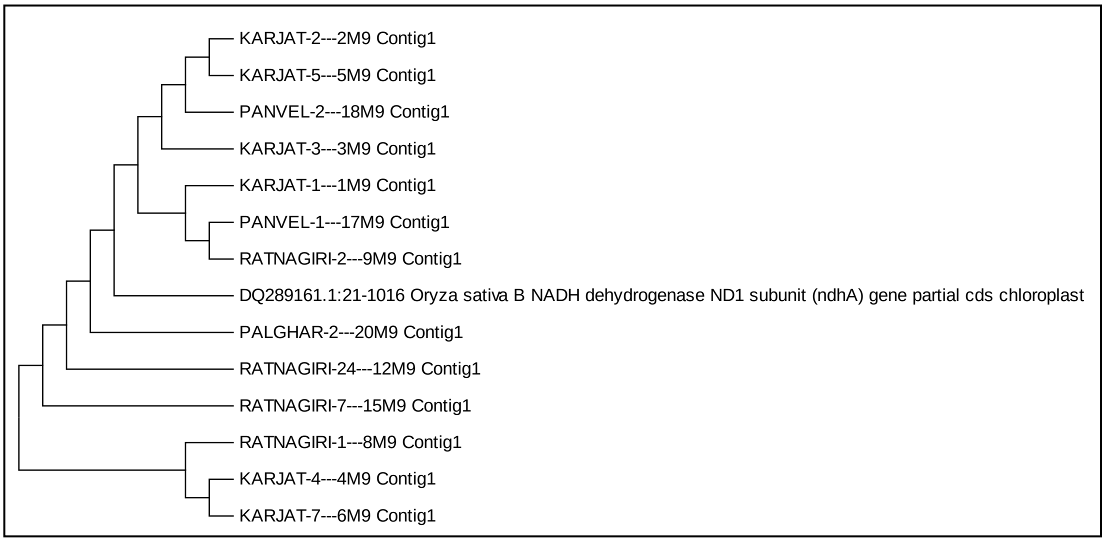
PCR of ndhA primer amplified a band of 1150- 1360 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer ndhA is presented in Plate 3.9. The samples amplified were Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from 874 bp and that of the reverse sequences generated ranged from 874bp in length. Around the range of 985-1000 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix 4-Plate I). The result showed that the sample sequences got completely aligned with the reference sequence from BLAST showing conserved region in the samples. Only a gap was observed at 866 – 871 bp.

Estimation of evolutionary divergence between all 3 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.00 – 0.003 (Table 4.5.9). P value of zero signifies nearest relative in the evolution where as the value 0.003 signifies distant in evolutionary ladder. Overall average p- distance was 0.001. K2P estimated overall average of 0.0010. Phylogenetic representation was done using Neighbour Joining Tree method. (Fig 4.9).

Table 4.5.9: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *ndhA*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. DQ289161.1:21-1016 Oryza sativa		0.000	0.001	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000
2. KARJAT-1---1M9 Contig1	-0.000		0.001	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000
3. KARJAT-2---2M9 Contig1	0.001	0.001		0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001
4. KARJAT-3---3M9 Contig1	-0.000	-0.000	0.001		0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000
5. KARJAT-4---4M9 Contig1	0.002	0.002	0.003	0.002		0.002	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.001
6. KARJAT-5---5M9 Contig1	0.001	0.001	0.001	0.001	0.003		0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001
7. KARJAT-7---6M9 Contig1	0.002	0.002	0.003	0.002	-0.000	0.003		0.001	0.001	0.001	0.001	0.001	0.002	0.001
8. PALGHAR-2---20M9 Contig1	-0.000	-0.000	0.001	-0.000	0.002	0.001	0.002		0.000	0.000	0.001	0.000	0.001	0.000
9. PANVEL-1---17M9 Contig1	-0.000	-0.000	0.001	-0.000	0.002	0.001	0.002	-0.000		0.000	0.001	0.000	0.001	0.000
10. PANVEL-2---18M9 Contig1	-0.000	-0.000	0.001	-0.000	0.002	0.001	0.002	-0.000	-0.000		0.001	0.000	0.001	0.000
11. RATNAGIRI-1---8M9 Contig1	0.002	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.002	0.002		0.001	0.002	0.001
12. RATNAGIRI-24---12M9 Contig1	-0.000	-0.000	0.001	-0.000	0.002	0.001	0.002	-0.000	-0.000	-0.000	0.002		0.001	0.000
13. RATNAGIRI-2---9M9 Contig1	0.001	0.001	0.002	0.001	0.003	0.002	0.003	0.001	0.001	0.001	0.003	0.001		0.001
14. RATNAGIRI-7---15M9 Contig1	-0.000	-0.000	0.001	-0.000	0.002	0.001	0.002	-0.000	-0.000	-0.000	0.002	-0.000	0.001	

Figure 4.9: Neighbor joining Phylogenetic Tree of samples sequences of *ndhA* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.00560946 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 14 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 982 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

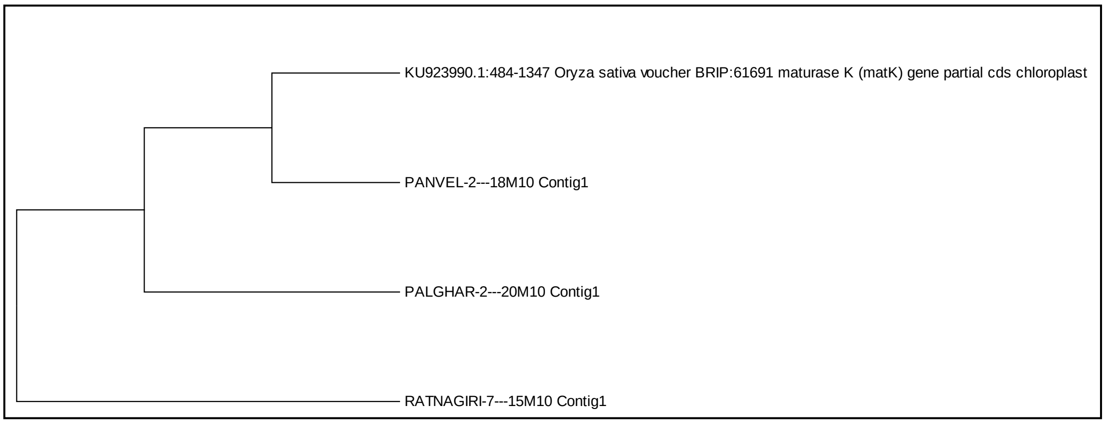
4.5.10. *matK-1M - matK3RIM*

PCR of primer *matK-1M - matK3RIM* amplified a band of 904-1069 bp. The agarose gel profile for the DNA amplification pattern observed in selected 16 rice varieties with primer *matK-1M - matK3RIM* is presented in Plate 3.10. The samples amplified were Karjat-5, Ratnagiri-2, Ratnagiri-7, Panvel-2, Palghar-2. The PCR band obtained on the gel was eluted by using gel elution kit and sent for sequencing. The sequences obtained were documented as FASTA files which were extracted for further use. Forward sequences generated were of length ranging from 830 bp and that of the reverse sequences generated ranged from 839 bp in length. Around the range of 900 bp long consensus sequences of samples were further use for blast analysis using NCBI, BLAST- N programme. Blast sequences of *Oryza sativa* species was download as FASTA format. Together with the blast sequence from NCBI and the consensus sequences of the samples were aligned and analyzed using MEGA 7 (Appendix IV- Plate J). The alignment showed that the sample sequences got aligned with the reference sequence showing conserved region. The variations observed were a gap at 47 – 110 bp and 122 – 136 bp, 142 – 153 bp in sample RATNAGIRI-7. Point variations of A to T at 169 bp, A to C at 174 bp, A to G at 175 & 177 bp, G to T at 178 bp in Ratnagiri-7. Larger gap at 322-351 bp & 359 to 397 bp was seen. A region of TAT was seen only in Ratnagiri-7 at 514 – 516 bp. Estimation of evolutionary divergence between all 16 sequences was done by calculating p- values using Bootstrap variance estimation method and Kimura-2- parameter of MEGA 7 software. P- values varied from 0.06 - 0.712 (Table 4.5.10). P value of zero signifies nearest relative in the evolution where as the value 0.712 signifies distant in evolutionary ladder. Overall average p- distance was 0.405. K2P estimated overall average of 0.351. Phylogenetic representation was done using Neighbour Joining Tree method. (Fig. 4.10)

Table 4.5.10: Estimates of Evolutionary Divergence between sequences blasted against NCBI sequences amplified using *matK-1M - matK3RIM* .

	1	2	3	4
1. KU923990.1:484-1347 Oryza sativa voucher BRIP:61691 maturase K (matK) gene pa		0.052	0.003	0.015
2. RATNAGIRI-7---15M10 Contig1	0.614		0.050	0.055
3. PANVEL-2---18M10 Contig1	0.006	0.600		0.015
4. PALGHAR-2---20M10 Contig1	0.114	0.663	0.111	

Figure 4.10: Neighbor joining Phylogenetic Tree of samples sequences of *matK-1M - matK3RIM* primer.



The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.80563158 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 4 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 512 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. (Fig. 4.10).

The overall estimates of Evolutionary Divergence by p-distance and K-2-Parameter are mentioned in the Table 4.6.1 and 4.6.2. The sequence polymorphism among 16 varieties representing the variable sites, parsimony informative sites and nucleotide diversity (Pi) is mentioned in the Table 4.6.3.

Table 4.6.1 : Estimates Of Evolutionary Divergence By P-Distance Using Bootstrap Variance

Sr. no.	Primers	Minimum	Maximum	Average	No. of varieties
1	matK	0.000	0.021	0.05	10
2	rbcL	0.000	0.007	0.004	14
3	*psbA-trnH*	0.000	0.007	0.004	16
4	trnH - psbA	0.000	0.650	0.165	7
5	atpH-atpI	0.000	0.796	0.011	14
6	ITS 2	0.000	0.006	0.003	7
7	petA-psbJ	0.000	2.170	0.335	12
8	*trnK*	0.000	-	0.271	16
9	ndhA	0.000	0.003	0.001	13
10	matK-1RKIM & matK-3FKIM	0.000	0.712	0.405	3

Table 4.6.2: Estimates of Evolutionary Divergence By Kimura-2-parameter

Sr. no.	Primers	Minimum	Maximum	Average	No. of varieties
1	matK	0.000	0.019	0.005	10
2	rbcL	0.000	0.007	0.004	14
3	*psbA-trnH*	0.000	0.642	0.084	16
4	trnH - psbA	0.000	0.633	0.161	7
5	atpH-atpI	0.000	0.543	0.076	14
6	ITS 2	0.000	0.006	0.003	7
7	petA-psbJ	0.000	0.619	0.095	12
8	*trnK*	0.000	0.531	0.184	16
9	ndhA	0.000	0.003	0.001	13
10	matK-1RKIM & matK-3FKIM	0.006	0.614	0.351	3

Table 4.6.3: Sequence polymorphism among 16 varieties based on the 10 barcoding loci

Sr. no.	Primers	Variable sites	Parsimony informative sites	Nucleotide diversity (per site Pi)
1	matK	47	21	-
2	rbcL	50	33	-
3	*psbA-trnH*	568	161	-
4	trnH - psbA	249	5	0.10824
5	atpH-atpI	13	2	0.05390
6	ITS 2	3	3	0.00294
7	petA-psbJ	390	1	0.06398
8	*trnK*	457	159	0.14555
9	ndhA	5	2	0.00121
10	matK-1RKIM & matK-3FKIM	285	13	-

4.6. BLAST analysis of the sequences:

The obtained sequences were individually analyzed by BLAST (Basic Local Alignment Sequence Tool) in NCBI database. The parameters studied were the score of alignment (S), E- value and the similar sequences showed colored appearance and could be identified by per cent efficiency. (Table.4.7)

Table no 4.7: BLAST Sequence Accession Details

Sr. No.	Primer	Description	E-value	Identity (%)	Accession
1	*matK*	Oryza sativa voucher BRIP:61691 maturase K (matK) gene, partial cds; chloroplast	0.0	99	KU923990.1: 406-1307
2	*rbcL*	Oryza sativa Indica Group cultivar DPA mitochondrion, complete genome	0.0	99	KY486275.1: 29348-30068
3	*psbA-trnH*	Oryza sativa Indica Group cultivar 9311 trnH-psbA intergenic spacer, partial sequence	0.0	99	GU575239.1
4	*trnH -psbA*	Oryza sativa Indica Group cultivar Weiyou 46 trnH-psbA intergenic spacer, partial sequence	0.0	99	GU575265.1: 1-605
5	*atpH-atpI*	Oryza sativa cultivar TN1 chloroplast, complete genome	0.0	99	NC_031333.1 :31111-32193
6	*ITS 2*	Oryza sativa cultivar IR8 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, and 26S ribosomal RNA gene, complete sequence	0.0	99	KM036284.1: 2065-2550
7	*petA-psbJ*	Oryza sativa b/f complex cytochrome f (petA) gene, partial cds; and PsbJ (psbJ) and PSII L protein (psbL) genes, complete cds; chloroplast	0.0	99	DQ289201.1: 1-1151
8	*ndhA*	Oryza sativa B NADH dehydrogenase ND1 subunit (ndhA) gene, partial cds; chloroplast	0.0	99	DQ289161.1: 21-1016
9	*matK-1RKIM & matK-3FKIM*	Oryza sativa voucher BRIP:61691 maturase K (matK) gene, partial cds; chloroplast	0.0	99	KU923990.1: 406-1313

4.6. Sequence polymorphism and genetic diversity among 16 varieties :

Genetic variation among the rice genotypes was estimated by calculating the number of haplotypes, haplotype diversity (HD), and parsimony informative sites using the DNAsp ver. 5.10 (Librado and Rozas 2009). Frequency of nucleotide substitution for all the ten loci among 16 varieties is mentioned in table. The estimate evolutionary divergence between sequences ranged from 0.000 to 2.170 (average 0.001 to 0.405). The number of base substitutions per site from between sequences were conducted using the Maximum Likelihood Substitution matrix. The parsimony informative sites was estimated with maximum 161 sites recorded in *psbA-trnH,* followed by 159 sites *trnK* and 33 *rbcL,* 21 in *matK,* 13 in *matK-1M-matK3RIM,* 5 in *trnH-psbA,* 2 in *atpH-atpI* and *ndhA,* 3 in *ITS2 and* only one site in *petA-psbJ* and number of variable sites reported highest in *psbA-trnH* 568, *trnK* 457 and *matK-1M-matK3RIM* 285.While nucleotide diversity (per site pi) reported maximum in *trnK* 0.1455. The sequence polymorphism data is mentioned in Table no 4.8 & 4.9. The sequence polymorphism among 16 varieties based on the 10 barcoding loci are represented in Table 4.10(A). The efficiency of Barcoding primers is represented in Table 4.10(B).

Table 4.8: Analysis of aligned sequences using DNAsp software

Parameters	**Barcoding Primers**									
	matK	**rbcL**	**psbA-trnH**	**trnH – psbA**	**atpH-atpI**	**ITS 2**	**petA-psbJ**	**trnK**	**ndhA**	**matK-1RKIM & matK-3FKIM**
No. of sequences in analysis	12	14	16	8	15	8	13	16	14	4
No.of sites	935	839	1154	930	1093	927	1388	1056	1014	933
Conserved region	881	734	466	665			789	380	983	607
Invariable sites	-	-	-	337	222	483	554	289	977	-
Variable sites	47	50	568	249	139	3	390	457	5	285
Singleton variable sites	24	9	396	244	137	0	389	298	3	271
Parsimony informative sites	21	33	161	5	2	3	1	159	2	13
Haplotype diversity	-	-	-	-	0.7857	0.5714	0.8571	0.9833	0.6813	-
Nucleotide diversity (Pi)	-	-	-	0.10824	0.05390	0.00294	0.06398	0.14555	0.00121	-

Table 4.9 : Maximum Likelihood Estimate Of Substitution Matrix Of 10 Barcoding Loci For 16 Varieties

matK					***rbcL***				
	A	T/U	C	G		A	T/U	C	G
A	-	*8.72*	*4.43*	**3.66**	A	-	*8.48*	*8.48*	**8.04**
T/U	*7.24*	-	**13.8**	*3.87*	T/U	*8.48*	-	**8.04**	*8.48*
C	*7.24*	**27.16**	-	*3.87*	C	*8.48*	**8.04**	-	*8.48*
G	**6.85**	*8.72*	*4.43*	-	G	**8.04**	*8.48*	*8.48*	-
psbA-trnH					***trnH - psbA***				
	A	T/U	C	G		A	T/U	C	G
A	-	*6.85*	*4.02*	**12.88**	A	-	*5.8*	*5.8*	**13.4**
T/U	*6.11*	-	**9.45**	*4.03*	T/U	*5.8*	-	**13.4**	*5.8*
C	*6.11*	**16.12**	-	*4.03*	C	*5.8*	**13.4**	-	*5.8*
G	**19.53**	*6.85*	*4.02*	-	G	**13.4**	*5.8*	*5.8*	-
atpH-atpI					***ITS 2***				
	A	T/U	C	G		A	T/U	C	G
A	-	*4.16*	*4.16*	**16.69**	A	-	*7.5*	*7.5*	**10**
T/U	*4.16*	-	**16.69**	*4.16*	T/U	*7.5*	-	**10**	*7.5*
C	*4.16*	**16.69**	-	*4.16*	C	*7.5*	**10**	-	*7.5*
G	**16.69**	*4.16*	*4.16*	-	G	**10**	*7.5*	*7.5*	-
petA-psbJ					***trnK***				
	A	T/U	C	G		A	T/U	C	G
A	-	*5.17*	*5.17*	**14.66**	A	-	*8.1*	*8.1*	**8.8**
T/U	*5.17*	-	**14.66**	*5.17*	T/U	*8.1*	-	**8.8**	*8.1*
C	*5.17*	**14.66**	-	*5.17*	C	*8.1*	**8.8**	-	*8.1*
G	**14.66**	*5.17*	*5.17*	-	G	**8.8**	*8.1*	*8.1*	-
ndhA					***matK-1RKIM & matK-3FKIM***				
	A	T/U	C	G		A	T/U	C	G
A	-	*10.4*	*10.4*	**4.16**	A	-	*6.28*	*6.28*	**12.43**
T/U	*10.4*	-	**4.16**	*10.4*	T/U	*6.28*	-	**12.43**	*6.28*
C	*10.4*	**4.16**	-	*10.4*	C	*6.28*	**12.43**	-	*6.28*
G	**4.16**	*10.4*	*10.4*	-	G	**12.43**	*6.28*	*6.28*	-

Table 4.10 (A) : Efficiency of Barcoding Primers

Parameters	Barcoding Primers									
	matK	**rbcL**	**psbA-trnH**	**trnH - psbA**	**atpH-atpI**	**ITS 2**	**petA-psbJ**	**trnK**	**ndhA**	**matK-1RKIM & matK-3FKIM**
Total no. of varieties taken	16	16	16	16	16	16	16	16	16	16
No. of varieties amplified	10	14	16	16	13	13	13	16	15	13
Amplification success (%)	62.5	87.5	100	100	81.25	81.25	81.25	100	93.75	81.25
No.of sequences of varieties obtained	10	14	16	7	14	7	12	16	13	3
Sequencing success (%)	62.5	87.5	100	43.75	87.5	43.75	75	100	81.25	18.75

Table 4.10 (B) : Banding pattern observed in 16 rice varieties for 10 loci

Sr. no.	Barcoding Primers	Monomorphic	Polymorphic	No amplification observed	Range of amplified bands (bp)
1.	matK	6	4	6	560-600
2.	rbcL	14	0	2	831-884
3.	psbA-trnH	12	4	-	677-76
4.	trnH - psbA	16	0	-	359-417
5.	atpH-atpI	13	0	3	1350-1477
6.	ITS 2	13	0	3	625-725
7.	petA-psbJ	13	0	3	1240-1372
8.	trnK	16	0	0	2240-2811
9.	ndhA	15	0	1	1150-1360
10.	matK-1RKIM & matK-3FKIM	13	0	3	904-1069

4.8 Barcode generation:

Online tool of BIO-RAD DNA Generator was used to develop barcodes for the samples. The aligned sequences were pasted in the tool and barcode was generated. The samples which were completely aligned with no variations got a single barcode with colour indications of by convention, **Adenine (A) is indicated by green fluorescence, Thymine(T) by red, Guanine (G) by black and Cytosine (C) by blue.**

The variable regions were selected and codes as barcode also the gaps were read and analyzed to read the barcode. The barcodes for different primers as per region (bp) is mentioned in Table 4.11.The barcodes of each of 16 rice varieties under study for the 10 barcoding loci has been represented in Plates 4 to 19.

Table 4.11: Region wise barcode generation in samples.

Sr. no.	Primer	Barcode region	Remarks
1.	matK	889 bp, 866 bp, 896 bp	**BARCODE1:**Karjat-2 **BARCODE 2:** Ratnagiri-2 **BARCODE 3:** Karjat-1, Karjat-3, Karjat-4, Karjat-5, Ratnagiri-1, Ratnagiri-3, Ratnagiri-4, Panvel-1.
2.	*rcbL*	715 bp	All samples except Karjat-1 & Karjat-2
3.	*psbA-trnH*	630 bp 628 bp	**BARCODE 1**: Karjat-2, Karjat-3, Karjat-4, Karjat-5, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-8, Panvel-1, Palghar-2 **BARCODE 2:** Karjat-1, Karjat-7, Ratnagiri-1.
4.	trnH - psbA	894 bp, 894 bp, 895 bp, 892 bp	**BARCODE 1:** Ratnagiri-3 **BARCODE 2:** Ratnagiri-4, Ratnagiri-8, Panvel-1, Palghar-2. **BARCODE 3:** Ratnagiri-7 **BARCODE 4**: Panvel-2

5.	atpH-atpI	357 bp 1057 bp	**BARCODE 1:** Ratnagiri-7 **BARCODE 2:** Karjat-2, Karjat-3, Karjat-4, Karjat-5, Ratnagiri-2, Ratnagiri-3, Ratnagiri-4, Ratnagiri-24, Ratnagiri-8, Panvel-2, Palghar-2.
6.	ITS 2	486 bp 487 bp	**BARCODE 1:** Karjat-2,Karjat-3, Karjat-4, Ratnagiri-3, Palghar-2 **BARCODE 2:** Karjat-5 & Karjat-7
7.	petA-psbJ	972 bp 1002 bp 1000 bp	**BARCODE 1:** Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Ratnagiri-1, Ratnagiri-2, Ratnagiri-4, Ratnagiri-7, Ratnagiri-8,Panvel-2. **BARCODE 2**: Karjat-7 **BARCODE 3:** Ratnagiri-24
8.	*trnK*	836 bp, 845 bp, 856bp, 885 bp.	**BARCODE 1:** Karjat-4 **BARCODE 2:** Karjat-7 **BARCODE 3:** Karjat-2, Karjat-3, Karjat-5, Ratnagiri-3,Ratnagiri-4,Ratnagiri-24, Ratnagiri-7,Ratnagiri-8 **BARCODE 4:** Ratnagiri-1, Ratnagiri-2, Panvel-1, Panvel-2
9.	ndhA	978 bp	ALL SAMPLES
10.	matK-1RKIM & matK-3FKIM	532 bp 864 bp 856 bp	**BARCDE 1:** Karjat-7 **BARCODE 2**: Panvel-2 **BARCODE 3:** Palghar-2

4.9 Molecular characterization based on SSR markers:

4.9.1. Standardization of annealing temperature:

The Tm value of reverse and forward primer taken into consideration and programme was set accordingly. The data obtained is presented in Table 4.12. Each primer has shown different annealing temperature. It has been observed that, the lowest annealing temperature was observed in primer RM-19 (57.5^{0}C) whereas, maximum annealing temperature was observed in primer RM- 181 (69.9^{0}C).

Table 4.12: Optimization of Annealing Temperature of Various SSR Markers.

Sr. No.	Name of primer	Tm value (^0C)		*Temperature range (^0C)*	*Standardized Annealing temperature (^0C)*
		Reverse primer	**Forward primer**		
1.	RM 338	56.0	57.0	52-62	58.7
2.	RM 335	55.0	55.2	52-62	58.7
3.	RM 112	66.3	63.6	57-68	65.7
4.	RM 223	56.5	56.3	6-66	63.2
5.	RM 201	52.5	50.0	53-63	61.9
6.	RM181	63.4	66.0	62-72	69.9
7.	RM140	57.9	65.6	55-68	62.1
8.	RM 19	54.7	49.1	50-58	57.5
9.	RM 309	63.3	61.2	58-68	66.9
10.	pTA 248	56.1	58.2	58-68	66.9

4.9.2. PCR amplification

The PCR amplification of 10 SSR markers was done following the temperature profiles from previous researches. The samples amplified as a single prominent band which was observed on agarose gel electrophoresis and visualized in gel documentation system. The range of amplified bands and the frequent band size observed in different rice varieties is mentioned in Table 4.13.

Table 4.13. Range of amplified bands and list of varieties with frequent observed band size

Sr. no.	SSR Marker	Range of amplified bands	Frequent band size	Variety Observed with frequent band size
1.	RM335	100-140 bp	164	Karjat-5, Karjat-7, Karjat-184, Ratnagiri-7 And Panvel-1 variety
2.	RM338	150-160 bp	163	Ratnagiri-3, Ratnagiri-4,Ratnagiri-73, Ratnagiri-6 and Ratnagiri-8
3.	RM112	127 bp	127	Karjat-3, Karjat-4,Karjat-5, Karjat-7, Karjat-184 and Panvel-2
4.	RM181	230- 240 bp	235	Karjat-1 Karjat-3, Karjat-5, Karjat-7, and Ratnagiri-2
5.	RM201	140-160 bp	149	Karjat-1, Karjat-3, Karjat-4, Karjat-7, Karjat-184 and Panvel-2.
6.	pTA 248	600-700 bp	711	Ratnagiri-4,Ratnagiri-24 and Ratnagiri-73
7.	RM309	170-180 bp	184	Karjat-1, Karjat-2, Karjat-3 and Panvel-1
8.	RM223	150-160 bp	150	Karjat-3, Karjat-184 and Ratnagiri-2

4.9.3. Primer wise PCR amplification:

The SSR primer, RM-335 produced 6 alleles and all were polymorphic. Among these one was unique size which was frequent was 164 bp and was present in Karjat-5, Karjat-7, Karjat-184, Ratnagiri-7 And Panvel-1 variety.

Primer RM-338 showed 5 alleles out of which all were polymorphic. The size of frequent band product was 163 bp and it was found in Ratnagiri-3, Ratnagiri-4, Ratnagiri-73, Ratnagiri-

6 and Ratnagiri-8 variety (Plate 21). In case of primer, RM-112 produce 3 loci and most observed band size was 127 bp found in Karjat-3, Karjat-4,Karjat-5, Karjat-7, Karjat-184 and Panvel-2 varieties (Plate 20).

Primer RM181 produced 3 alleles. Varieties Karjat-1 Karjat-3, Karjat-5, Karjat-7, and Ratnagiri-2 represented same band size of 235 bp (Plate 22). In primer RM 201, 5 alleles were observed with most frequent band size of 149 bp among in Karjat-1, Karjat-3, Karjat-4, Karjat-7, Karjat-184 and Panvel-2.

For RM 309, alleles observed were 4 and 184 bp band size observed in Karjat-1, Karjat-2, Karjat-3 and Panvel-1.Likely, the primer pTA-248 produced 6 alleles out of which all were polymorphic and 1 was unique frequent size was seen. The size of unique allele was 711 bp and found in Ratnagiri-4, Ratnagiri-24 and Ratnagiri-73 variety of rice.

Table 4.14 : Primers with allelic description

<table>
<tr><td colspan="3">RM 335</td></tr>
<tr><td rowspan="14">The primer RM 335 amplified a total of seventeen alleles ranging in size from 105 to 171 bp. These alleles were designated as A, B, C, D, E, F, G; H, I, J, K, L, M. Out of these seventeen alleles, allele D recorded the highest allele frequency.</td><td>Band designation</td><td>Band size</td></tr>
<tr><td>A</td><td>0.171</td></tr>
<tr><td>B</td><td>0.167</td></tr>
<tr><td>C</td><td>0.166</td></tr>
<tr><td>D</td><td>0.164</td></tr>
<tr><td>E</td><td>0.162</td></tr>
<tr><td>F</td><td>0.150</td></tr>
<tr><td>G</td><td>0.144</td></tr>
<tr><td>H</td><td>0.143</td></tr>
<tr><td>I</td><td>0.138</td></tr>
<tr><td>J</td><td>0.133</td></tr>
<tr><td>K</td><td>0.121</td></tr>
<tr><td>L</td><td>0.118</td></tr>
<tr><td>M</td><td>0.105</td></tr>
</table>

RM 338

The primer RM 338 amplified a total of twenty alleles ranging in size from 149 to 181 bp.These alleles were designated as A, B, C, D, E, F, G; H, I. Out of these alleles, allele D recorded the highest allele frequency. This marker showed a PIC value of 0.80.

Band designation	Band size
A	0.181
B	0.173
C	0.172
D	0.163
E	0.161
F	0.157
G	0.154
H	0.152
I	0.149

RM 181

The primer RM 181 amplified a total of nineteen alleles ranging in size from 213 to 245 bp. These alleles were designated as A, B, C, D, E, F, G, H, I, J, K. Out of these eight alleles, allele D recorded the highest allele frequency.

Band designation	Band size
A	0.245
B	0.238
C	0.237
D	0.235
E	0.233
F	0.231
G	0.228
H	0.219
I	0.217
J	0.215
K	0.213

RM 309		
The primer RM 309 amplified a total of seventeen alleles ranging in size from 162 to 218 bp. These alleles were designated as A, B, C, D, E, F, G; H, I. Out of these alleles, allele C & H recorded the highest allele frequency.	Band designation	Band size
	A	0.218
	B	0.186
	C	0.184
	D	0.182
	E	0.171
	F	0.167
	G	0.166
	H	0.164
	I	0.162

RM 201		
The primer RM 201 amplified a total of twenty alleles ranging in size from 142 to 286 bp.These alleles were designated as A, B, C, D, E, F, G; H, I, J, K, L. Out of these eight alleles, allele I recorded the highest allele frequency.	Band designation	Band size
	A	0.287
	B	0.286
	C	0.285
	D	0.284
	E	0.279
	F	0.277
	G	0.272
	H	0.155
	I	0.149
	J	0.146
	K	0.144
	L	0.142

RM 112

The primer RM 112 amplified a total of nineteen alleles ranging in size from 121 to 131 bp.These alleles were designated as A, B, C, D, E, Out of these alleles, allele B recorded the highest allele frequency.

Band designation	Band size
A	0.131
B	0.127
C	0.125
D	0.123
E	0.121

RM 223

The primer RM 223 amplified a total of thirty nine alleles ranging in size from 143 to 1000 bp. These alleles were designated as A, B, C, D, E, F, G; H, I, J, K, L, M, N, O, P, Q, R, S, T ,U ,V,W. Out of these alleles, allele S recorded the highest allele frequency.

Band designation	Band size
A	1.000
B	0.983
C	0.330
D	0.320
E	0.315
F	0.305
G	0.303
H	0.302
I	0.300
J	0.295
K	0.284
L	0.281
M	0.157
N	0.156
O	0.155
P	0.154
Q	0.153
R	0.151
S	0.150
T	0.147
U	0.146
V	0.145
W	0.143

<table>
<tr><th colspan="3">pTA 248</th></tr>
<tr><td rowspan="20">The primer pTA 248 amplified a total of twenty seven alleles ranging in size from 143 to 1000 bp. These alleles were designated as A, B, C, D, E, F, G; H, I, J, K, L, M, N, O, P, Q, R, S. Out of these alleles, allele E recorded the highest allele frequency.</td><td>Band designation</td><td>Band size</td></tr>
<tr><td>A</td><td>1.412</td></tr>
<tr><td>B</td><td>1.390</td></tr>
<tr><td>C</td><td>1.368</td></tr>
<tr><td>D</td><td>0.727</td></tr>
<tr><td>E</td><td>0.711</td></tr>
<tr><td>F</td><td>0.705</td></tr>
<tr><td>G</td><td>0.696</td></tr>
<tr><td>H</td><td>0.690</td></tr>
<tr><td>I</td><td>0.687</td></tr>
<tr><td>J</td><td>0.686</td></tr>
<tr><td>K</td><td>0.670</td></tr>
<tr><td>L</td><td>0.666</td></tr>
<tr><td>M</td><td>0.640</td></tr>
<tr><td>N</td><td>0.636</td></tr>
<tr><td>O</td><td>0.631</td></tr>
<tr><td>P</td><td>0.605</td></tr>
<tr><td>Q</td><td>0.600</td></tr>
<tr><td>R</td><td>0.586</td></tr>
<tr><td>S</td><td>0.324</td></tr>
</table>

4.9.4. Comparative analysis of SSR markers:

The analysis of eight primers out of ten was done, band size was observed and the following tables provides a comparative approach of the markers amplifying among the 20 varieties in the study (Table 4.15).

Table 4.15. Comparative Analysis of 20 varieties by SSR markers.

No.		**Band positions due to primer (bp)**																					
		RM338									**RM335**												
sr. no.	**Varieties**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
A) BOLD VARIETIES																							
I)LONG BOLD																							
1	**KARJAT-5**								152					164									
2	**RATNAGIRI- 1**	181																		133			
3	**RATNAGIRI-73**			172															138				
4	**PANVEL-3**						157																
II)SHORT BOLD																							
1	**KARJAT-3**							154															
2	**RATNAGIRI-2**			172																	121		
3	**PANVEL-1**								152					164									
4	**PALGHAR-2**			172																			
B) SLENDER VARIETIES																							
I)LONG SLENDER																							
1	**KARJAT-1**		173									167											
2	**KARJAT-2**					161					171												
3	**KARJAT-7**								152					164									
4	**KARJAT-184**									149				164									
5	**RATNAGIRI-3**				163												144						
6	**RATNAGIRI-4**				163											150							
7	**RATNAGIRI-24**				163													143					
8	**PANVEL-2**							154					166										
II)SHORT SLENDER																							
1	**KARJAT -4**						157								162								
	RATNAGIRI-6				163																		105
	RATNAGIRI-7			172										164									
	RATNAGIRI-8				163																	118	

No.		RM181											RM201											
sr. no.	**Varieties**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K`**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
A) BOLD VARIETIES																								
I)LONG BOLD																								
1	**KARJAT-5**				235																			142
2	**RATNAGIRI- 1**			237														277						
3	**RATNAGIRI-73**										215				285									
4	**PANVEL-3**	242																		155				
II)SHORT BOLD																								
1	**KARJAT-3**				235																149			
2	**RATNAGIRI-2**				235														272					
3	**PANVEL-1**						231															146		
4	**PALGHAR-2**											213			285									
B) SLENDER VARIETIES																								
I)LONG SLENDER																								
1	**KARJAT-1**				235																149			
2	**KARJAT-2**		238																				144	
3	**KARJAT-7**				235																149			
4	**KARJAT-184**					233															149			
5	**RATNAGIRI-3**																	277						
6	**RATNAGIRI-4**								219								279							
7	**RATNAGIRI-24**									217						284								
8	**PANVEL-2**					233															149			
II)SHORT SLENDER																								
1	**KARJAT -4**							228													149			
	RATNAGIRI-6								219				287											
	RATNAGIRI-7								219					286										
	RATNAGIRI-8											213			285									

No.		**Band positions due to primer bp**													
		RM122					**RM309**								
sr. no.	**varieties**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
A) BOLD VARIETIES															
I)LONG BOLD															
1	**KARJAT-5**		127							182					
2	**RATNAGIRI- 1**				123							167			
3	**RATNAGIRI-73**					121								164	
4	**PANVEL-3**	131						186							
II)SHORT BOLD															
1	**KARJAT-3**		127						184						
2	**RATNAGIRI-2**			125							171				
3	**PANVEL-1**			125					184						
4	**PALGHAR-2**					121									
B) SLENDER VARIETIES															
I)LONG SLENDER															
1	**KARJAT-1**				123				184						
2	**KARJAT-2**				123				184						
3	**KARJAT-7**		127												
4	**KARJAT-184**		127				218								
5	**RATNAGIRI-3**														
6	**RATNAGIRI-4**				123										162
7	**RATNAGIRI-24**					121								164	
8	**PANVEL-2**		127					186							
II)SHORT SLENDER															
1	**KARJAT -4**		127							182					
	RATNAGIRI-6					121								164	
	RATNAGIRI-7				123									164	
	RATNAGIRI-8					121							166		

4.10. Polymorphism percentage:

The polymorphism percentage for individual primer was calculated by the ratio of number of polymorphic bands obtained over the total number of bands produced across the 20 rice genotypes. A total of 10 SSR primer pairs distributed across the genome were used for molecular analysis of 20 rice varieties. Each of the primer found to be polymorphic, the details of which are depicted in materials and methods. All 10 microsatellite markers were found to be polymorphic. The overall average size of amplified products ranged from 121.86 bp to 217.14 bp. A total of 35 alleles were obtained using 10 SSR primer pairs with an average of 5 alleles per primer. The number of alleles amplified for each primer pair ranged from 3 to 6. The markers RM-335 and pTA 248 generated a maximum number of alleles (6). While the primer RM-112 and RM-181 produced minimum number of alleles (3).

4.11. Polymorphic Information Content (PIC):

The PIC values of primers ranged from 0.45 in SSR primer RM 338 to 0.65 in SSR pTA 248 primers with an average PIC value of 0.56. The molecular polymorphism and percent polymorphism is mentioned in the Table 4.16.

Table 4.16: Molecular Polymorphism, PIC Values, and Size of Loci Revealed by SSR Primers in Rice varieties.

Sr. No.	Primer	Range of Amplification (bp)	No. of Alleles	% Polymorphism	PIC
1.	RM335	105-171	6	100	0.45
2.	RM338	149-181	5	100	0.631
3.	RM112	121-131	3	100	0.45
4.	RM181	213-245	3	100	0.55
5.	RM201	142-286	5	100	0.650
6.	pTA 248	600-727	6	100	0.650
7.	RM309	162-218	4	100	0.55
	Total		**35**	**100**	
	Average		**5**	**100.00**	**0.56**

4.12. Genetic distance:

The diversity observed in the twenty rice varieties mainly attributed to the genetic dissimilarities. The Jaccard's similarity coefficient values among these rice varieties are presented in (Table 4.17). The pair wise similarity values ranged from 0.190 to 0.750. Maximum similarity value of 0.750 was noticed between Karjat-3, Karjat-4 And Karjat-5.Minimum similarity value of 0.016 was observed between Ratnagiri-2 and Karjat-184.

From Jaccard's similarity coefficient and consensus tree, it may be concluded that, crossing between rice varieties having Jaccard's similarity coefficient in the range of 0.190 to 0.750 and fall in different clusters and sub clusters may result in obtaining heterosis.

4.13. Cluster analysis and Dendrogram construction:

Marker alleles were converted to binary scores based on their presence or absence as 1 and 0, respectively. UPGMA cluster analysis was performed using Jaccard's similarity coefficient matrices calculated from SSR markers to generate a dendrogram for 20 rice varieties. A pairwise similarity index (SI) was calculated and the UPGMA based dendrogram of 20 rice varieties generated with Multivariate Statistical Package (MVSP) was presented in Fig. 4.11. It was observed that, there are two major clusters. The first major cluster consists of two sub clusters. The rice variety Ratnagiri-2 formed the independent first sub cluster. Second sub cluster consists of nine varieties *viz.,* Palghar -2, Ratnagiri-7, Ratnagiri- 73, Ratnagiri-24, Ratnagiri-4, Ratnagiri-8, Ratnagiri-6, Ratnagiri-3, Ratnagiri-1. The second major cluster consists of two sub cluster. First sub cluster consist of eight rice varieties *viz.,* Karjat-184, Panvel-3, Panvel-2, Karjat-4, Karjat-7, Karjat-5, Panvel-1, Karjat-3. Second sub cluster consist of rice varieties *viz.,* Karjat-1, Karjat-2. The cluster analysis data has been represented in Table 4.18.

Table 4.18: Clustering Pattern of 20 Rice Varieties.

Cluster			**No. of varieties**	**Name of the varieties**
I	IA		1	Ratnagiri-2
	IB	IBa	5	Palghar -2, Ratnagiri-7, Ratnagiri- 73, Ratnagiri-24, Ratnagiri-4
		IBb	4	Ratnagiri-8, Ratnagiri-6, Ratnagiri-3, Ratnagiri-1
II	IIA	IIAa	1	Karjat-184
		IIAb	7	Panvel-3, Panvel-2, Karjat-4, Karjat-7, Karjat-5, Panvel-1, Karjat-3
	IIB		2	Karjat-2, Karjat-1

Table 4.17: Genetic distance values based on Jaccard's similarity index values between 20 rice varieties

Similarity matrix																				
	KARJAT-1	**KARJAT-2**	**KARJAT-3**	**KARJAT-4**	**KARJAT-5**	**KARJAT-7**	**KARJAT-184**	**RATNAGIRI-1**	**RATNAGIRI-2**	**RATNAGIRI-3**	**RATNAGIRI-4**	**RATNAGIRI-24**	**RATNAGIRI-73**	**RATNAGIRI-6**	**RATNAGIRI-7**	**RATNAGIRI-8**	**PANVEL-1**	**PANVEL-2**	**PANVEL-3**	**PALGHAR-2**
KARJAT-1	1																			
KARJAT-2	0.364	1																		
KARJAT-3	0.4	0.364	1																	
KARJAT-4	0.273	0.154	0.556	1																
KARJAT-5	0.273	0.25	0.556	0.4	1															
KARJAT-7	0.273	0.25	0.556	0.4	0.75	1														
KARJAT-184	0.273	0.154	0.4	0.273	0.273	0.273	1													
RATNAGIRI-1	0.077	0.071	0	0	0	0	0.077	1												
RATNAGIRI-2	0.077	0	0.077	0.077	0.167	0.077	0.077	0.167	1											
RATNAGIRI-3	0.077	0.154	0	0	0	0	0.077	0.556	0.167	1										
RATNAGIRI-4	0.077	0.154	0	0.077	0.077	0	0	0.273	0.167	0.4	1									
RATNAGIRI-24	0.077	0.154	0	0	0.077	0	0	0.273	0.167	0.556	0.4	1								
RATNAGIRI-73	0.167	0.071	0	0	0.077	0	0	0.4	0.273	0.273	0.273	0.556	1							
RATNAGIRI-6	0.077	0.25	0.077	0	0	0	0	0.273	0.167	0.4	0.273	0.556	0.4	1						
RATNAGIRI-7	0.273	0.071	0.077	0.167	0.167	0.077	0.077	0.167	0.167	0.167	0.4	0.4	0.556	0.273	1					
RATNAGIRI-8	0.077	0.154	0	0	0	0	0.077	0.4	0.167	0.556	0.273	0.4	0.273	0.556	0.167	1				
PANVEL-1	0.273	0.364	0.75	0.4	0.75	0.75	0.273	0	0.077	0	0	0	0	0.077	0.077	0	1			
PANVEL-2	0.4	0.25	0.75	0.75	0.556	0.556	0.4	0	0.077	0	0	0	0	0	0.077	0	0.556	1		
PANVEL-3	0.4	0.25	0.556	0.556	0.4	0.4	0.273	0	0	0	0	0	0	0	0.077	0	0.4	0.75	1	
PALGHAR-2	0.273	0.071	0.077	0.077	0.167	0.077	0.077	0.273	0.273	0.273	0.273	0.4	0.556	0.273	0.556	0.4	0.077	0.077	0.077	1
	KARJAT-1	**KARJAT-2**	**KARJAT-3**	**KARJAT-4**	**KARJAT-5**	**KARJAT-7**	**KARJAT-184**	**RATNAGIRI-1**	**RATNAGIRI-2**	**RATNAGIRI-3**	**RATNAGIRI-4**	**RATNAGIRI-24**	**RATNAGIRI-73**	**RATNAGIRI-6**	**RATNAGIRI-7**	**RATNAGIRI-8**	**PANVEL-1**	**PANVEL-2**	**PANVEL-3**	**PALGHAR-2**

CHAPTER V

DISCUSSION

The present investigation entitled "DNA Barcoding for identification of rice (*Oryza sativa* L.) varieties developed by Dr. B. S. Konkan Krishi Vidyapeeth, Dapoli." is carried out with an objective to characterize rice varieties with molecular markers and to develop a molecular barcode for the same based on morphological aspect of grain type. Rice (*Oryza sativa L.*) is regarded as one of the major cereal crops with high agronomic and nutritional importance. Rice is first crop of which complete genome was sequenced and it has been ideal model plant for study due to its relatively small genome size of 430 Mb compared to other plants (Causse *et al.,* 1994). The genetic diversity available in rice crop has good scope for crop improvement aiming to meet the demand for productive crops has led to development of high yielding and stable varieties. Proliferation of rice varieties has narrowed down the number of combinations of morphological descriptors available to describe the uniqueness of a variety.

The basic objective of varietal characterization is to test the occurrence of traits that helps in identifying a particular variety. The characters that are used to distinguish cultivars should have the ability of precise description and recognition and is considered important only when they are not subjected to environmental influences. Thus, the ability to identify and distinguish between varieties is a fundamental component in seed quality programmes. This also benefits the seed production and certification authorities as well as the farmers in ensuring supply and distribution of genetically pure seeds. Therefore, characterization and varietal identification of available landraces and improved varieties have become important in modern day crop improvement (Vanniarajan *et al.*, 2012). With the parallel advances in next generation sequencing technology and bioinformatics tools for sequence characterization, DNA barcoding aims to develop an inexpensive, fast and precise method for species identification, especially useful for non-taxonomists.

DNA barcoding depends on short, highly variably regions of the genome. The unique nucleotide sequence patterns of small DNA fragments (400-800 bp) are used as specific reference collections to identify specimens and to discover over-looked species. In order to promote the use of DNA barcoding for all eukaryotic life on this planet, a Consortium for the Barcode of life (CBOL) was established in May 2004, which currently included more than 120 organizations from 45 nations (Vijayan *et. al.*, 2010).Sequences suggested to be useful in DNA barcoding include cytoplasmic mitochondrial DNA (eg. COX1) and chloroplast DNA (eg. rbcL, trnL-F, matK, ndhF and atpB rbcL) and nuclear DNA (ITS, and housekeeping genes eg. Gapdh). The plant DNA barcoding has now transitioned the epitome of species identification and thus, ultimately helping in the molecularization of taxonomy, a need of the hour (Ajmal M. *et.al.,*2014).

5.1. Morphological Characterization:

The morphological characters were observed for the varieties under this study. Mainly the grain type character was considered to differentiate varieties into two major catergories as bold and slender and then classified as long and short in both the catergories. Also the major agronomic aspects of these varieties highlighted in the Chapter 4.(Singh, *et.al.,*2017).

5.2. Molecular Characterization:

The 16 rice varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli were characterized using 10 barcoding loci focused on the chloroplast DNA. The assessment of species variation using the reported primers from conserved regions has been done. At present, *rbcL, matK, psbA-trnH, rpoC1, ITS2, atpF-atpH* spacer and *psbK-psbI* spacer have been popularly used as DNA barcodes in plant worldwide (Janzen *et.al.*, 2009) among that *matK* has been seen good for differentiation.

5.2.1. DNA isolation:

The isolation of high-quality DNA is important in any molecular biology work because contaminants such as proteins, polyphenols, and polysaccharides which may interfere with enzymes such as restriction enzymes (in blotting techniques) and *Taq* polymerase (in polymerase chain reaction [PCR]). For the present experimental study, tender leafs of about 10-12 days after sowing were collected from all the 20 varieties under study. The protocol for DNA isolation was followed (Doyle and Doyle, 1987) which were further modified to provide DNA suitable to several kinds of analysis. In the present study various concentrations of glucose, PVP, SDS, were tried. Among them, a combination of 0.900 g glucose, 0.300 g PVP, 0.040g sodium bisulphite, 0.050 g lauryl sulphate and 500 µl sarcosyl in 10 ml of extraction buffer (EB) yielded an appropriate quantity of DNA without any phenolic contamination. Also for best quality and quantity DNA, it was isolated using the Invitrogen (Thermo Fisher Scientific) Pure Link ® Genomic DNA Mini kit as per the protocol mentioned in the same.

5.2.2. DNA dilution and quantification:

DNA was visualized on 0.8percent agarose gel and PCR products were visualized on 1.5 per cent agarose gels under UV light by Ethidium bromide staining. Those samples with ratio between 1.8 and 2.0 were considered to be of high quality. Majority of samples recorded a ratio between 1.8 and 2.0.

5.2.3. PCR master mix:

PCR Master Mix is a ready-to-use solution containing all the necessary reaction components at optimal concentrations for efficient PCR amplification of DNA templates. The DNA template added in reaction mixture was 1 µl for SSR markers PCR and that for barcoding markers it was about 2 µl for a single PCR reaction.

5.3. DNA barcoding:

DNA barcoding depends on short, highly variably regions of the genome. The unique nucleotide sequence patterns of small DNA fragments (400-800 bp) are used as specific reference collections to identify specimens and to discover over-looked species. In order to promote the use of DNA barcoding for all eukaryotic life on this planet, a Consortium for the Barcode of life (CBOL) was established in May 2004, which currently included more than 120 organizations from 45 nations (Vijayan *et.al*, 2010). The plant DNA barcoding have now transitioned the epitome of species identification and thus, ultimately helping in the molecularization of taxonomy, a need of the hour. (Ajmal M. *et.al.,*2014)

Chloroplast DNA (cpDNA) possesses the most ideal DNA sequence for phylogenetic analysis. The reason behind this is they are relatively easy to purify, characterize, clone and sequence (Cleg *et al.*, 1990) and also endemic to plants. Thus Chloroplast DNA barcodes avoid the DNA contamination from other organisms without chloroplasts, such as animals and fungi. The chloroplast genome sequence of rice Nipponbare (*O. sativa L.ssp. japonica*) was reported to have a length of 134,525 bp (Hiratsuka *et al.,* 1989). Chloroplasts restrain both highly conserved genes important to plant life and more variable regions, which have been informative over broad time scales. Relative studies of the genomic structural design showed that the order of genes and the contents of essential genes are highly conserved among most chloroplast genomes (De Las Rivas *et al.*, 2002). Nevertheless, variations between different and closely related genomes have occurred during evolution (Tang *et al.*, 2004). Chloroplast genome sequence specific loci were used in the study to assess their potential and identified candidate DNA Barcode loci for intra species discrimination in rice. Barcode regions must be relatively short in length to facilitate easy PCR amplification and DNA sequencing.

The feasibility of using the chloroplast genome (CP genome) as a "Super Barcode" is evaluated and the concept of a 'specific barcode' derived from the comparison between plastid genome sequences from target group of taxa is presented as an effective option that might be widely applicable to plant identification studies (Xiwen Li. *et.al.,*2015). BOLD is a cloud based data storage and analysis platform developed at the centre for biodiversity genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINS (putative species) and a data collection and analysis work batch. The Consortium for the Barcode of Life – Plant Working Group (CBOL) recently recommended the two- locus combinations of matK + rbcL as the best plant barcode with a discriminatory efficiency of only 72% (CBOL plant working group, 2009).

Chloroplast markers are particularly used for revealing phylogenetic relationships among related groups. Being maternally inherited, they accumulate nucleotide substitutions occurring at neutral sites. The seven candidate plastid region *rpoB, rpoC, matK, rbcl, atpF-atpH and psbI and trnH-psbA* for groups of land plants (Hollingsworth *et al,* 2011) ; universal barcodes with combination of *matK+rpoc1+rpoB* (Chase *et al,* 2007) and *matK+rpoc1+trnH& psbA* out of which combination of *rbcl+matK* has been suggested for the terrestrial plants as the main barcode (CBOL,2009). The spacer between tRNA-His and photo system II protein D1 (trnH-psbA spacer) and the nuclear internal transcribed spacer 2 (ITS2) are also widely used (Chen *et.al.,* 2010, Fu *et.al.,* 2011; Gao *et.al.,* 2010). Barcoding success may be improved in some plants groups by careful choice of markers and appropriate sampling. The DNA barcode project aims to develop a simple diagnostic tool based on strong taxonomic data that is collected in the DNA barcode reference library (Schindel and Miller, 2005). According to Rubin off and Holland (2005), it was considered as a “tremendous tool” to speed up the species discovery and also to describe the new species, in addition it also re-opens the debate on species concepts (Fitzhugh, 2006; Rubinoff *et al.*, 2006b; Balakrishnan, 2007; Miller, 2007; Vogler and Monaghan, 2007). DNA barcode have applications in various fields like, ecology, biomedicine, epidemiology, evolutionary biology, biogeography, conservation biology and in bio-industry. The low cost and rapidity makes the process easier for enabling automated species identification especially in massive sampling campaigns (Rusch *et al.*, 2007).

5.3.1. PCR product sequencing

A total of 130 sequences of the 16 varieties were obtained from PCR amplified products using 10 barcoding primers and sequenced by outsourcing (Reliance Industries Limited, Mumbai). The forward and reverse sequencing was done for the samples, however some samples were unable to give good quality reads thus such sequences were eliminated and good quality reads were considered for further analysis. The average length of sequences was 600-2900 bp which shows good sequencing success.

5.3.2. Sequencing and Quality Check

The preliminary requirement for DNA barcode is expected to have a length of 500-800 bp. Sanger‘s di-deoxy chain termination method produces reads of ~1000 bp long in a single reaction (Chan 2005). Therefore, in the present investigation, the DNA sequencer based on Sanger‘s di - deoxy chain termination method (Sanger *et al.* 1977) was used. The PCR amplified products of expected size were extracted from the gel then purified and then subjected to DNA sequencing. Quality checking is one of the most crucial steps for the generation of DNA barcode sequences.

5.4. Sequence Data Analysis

The consensus sequences have been used for multiple sequence alignment in the present study. The Multiple Sequence Alignment using Clustal W was performed in the MEGA 7 software. The alignment gave colour codes as mentioned in the Appendix IV.

The aligned sequences indicated by "*" symbol were considered as conserved region. The analysis of DNA sequences was conducted by Neighbors - joining to assess topology with MEGA 7. Neighbors - Joining (NJ) method which was tested with Kimura 2-parameter for evolutionary distances in MEGA 7.0 and node support was assessed on 1000 bootstrap replicated. Pair wise distance, transitional/ transversional substitutions, and the maximum likelihood substitution matrix were estimated using MEGA 7.0.25 software (Tamura *et al.,* 1993-2017). Genetic variation among the rice genotypes was estimated by calculating the number of haplotypes, haplotype diversity (HD), and parsimony informative sites using the DNAsp ver. 5.10 (Librado and Rozas 2009).

5.4.1. Primer analysis

In this study the amplification and nucleotide sequencing of the individual barcode loci viz*, matK, rbcL, psbA-trnH, trnH-psbA, atpH-atpI, ITS2, petA-psbJ, trnK, ndhA, matK-1M-matK3RIM* was attempted for 16 varieties. The amplification efficiency for *psbA-trnH , trnH-psbA,* and *trnK* was 100% followed by *ndhA* (93.75%), *rbcL* (87.5 %)*, atpH-atpI, ITS2, petA-psbJ, matK-1M-matK3RIM* (81.25%) and *matK* (62.5 %). Both the *matK* primers sets showed presence of polymorphic bands hence gel elution process was followed to elute the target bands for sequencing.

It has been reported that genetic variation and phylogenetic relationship among the wild rice (*Oryza rufipogon Griff.*) and cultivate rice as revealed by chloroplast *matK* as promising loci (Roy, 2015), apart from *rbcL* and *mat*K barcode regions the most widely used plastid barcoding markers is the intergeneric spacer trnH-psbA (Shaw *et al.*, 2007) and the two New Potential Barcodes to discriminate Dalbergia Species highest rate of species identification (100%) with *matK,matK+rbcL* and *matK+trnH-ps*b loci (Bhagwat *et al.*, 2015).

5.4.2. BLAST (Basic Local Alignment Search Tool) method based evaluation of DNA barcodes

The sequencing data from 10 barcoding loci region was aligned and study were individually subjected to NCBI-Nucleotide (BLAST) server in order to find out their maximum level of taxonomic identification viz. species, genus, family, sub-tribe, tribe or class. The BLAST searches were carried out using one representative sequence for each locus with respect to the generated sequences of all the 16 varieties.

5.4.3. Sequence polymorphism among the 16 rice varieties

The estimate evolutionary divergence between sequences ranged from 0.000 to 2.170 (average 0.001 to 0.405). The number of base substitutions per site from between sequences were conducted using the Maximum Likelihood Substitution matrix. The parsimony informative sites was estimated with maximum 161 sites recorded in *psbA-trnH,* followed by 159 sites *trnK* and 33 *rbcL,* 21 in *matK,* 13 in *matK-1M-matK3RIM,* 5 in *trnH-psbA,* 2 in *atpH-atpI* and *ndhA,* 3 in *ITS2 and* only one site in *petA-psbJ* and number of variable sites reported highest in *psbA-trnH* 568, *trnK* 457 and *matK-1M-matK3RIM* 285.While nucleotide diversity (per site pi) reported maximum in *trnK* 0.1455.

In alignments of primer *psbA-trnH* it was observed that two main groups of varieties were seen showing similar alignments. In case of *trnH-psbA*, Ratnagiri-3 showed many variations at different sites in the sequences. For *atpH-atpI* , Ratnagiri-4 showed differences in sequence alignment whereas for *petA-psbJ* two varieties Karjat-7 and Ratnagiri-24 showed variations in sequences. In *trnK*, Karjat-4 had many variations at different sites and Karjat-7contained an extra region of nucleotides but other varieties did not show the presence of the same region. For *matK-1M-matK3RIM* alignments could be possible for only 3 varieties and for *matK* variations were seen in Ratnagiri-2. *rbcL* and *ndhA* showed complete alignment with the blast sequences with NCBI.

DNA barcoding is a molecular-based identification system, recently introduced in the scientific community. DNA barcoding is promising in providing a practical, standardized, species-level identification tool that can be used for different study including forensic analysis (Lahaye *et al,* 2008). The characterization of nucleotide and amino acid substitution along the gene may also provide information on site-dependent probabilities of nucleotide substitutions. Such information could provide a guide to the regions to be used in phylogenetic analysis since methods of phylogenetic inference assume the probabilities of replacements are independent of site (Clegg *et al.,* 1990).

5.5. Molecular Phylogenetic analysis

Phylogeny that analyses hereditary molecular differences, mainly in DNA sequences, to gain information on an organism's evolutionary relationships. The result of a molecular phylogenetic analysis is expressed in a phylogenetic tree. Molecular phylogenetic is one aspect of molecular systematic, a broader term that also includes the use of molecular data in taxonomy and biogeography. Molecular Phylogenetic analysis performed by Neighbor Joining Method. The results of each of the locus is reported and discussed below.

1. ***matK*:** The phylogenetic tree was constructed by Neighbor- Joining method in MEGA 7 software. The tree represented 10 varieties in 5 main clusters which showed Ratnagiri-2 , Karjat-1, Ratnagiri-4, Karjat-4 completely independent and Karjat-2, Karjat-3 and Ratnagiri-3 showed closed resemblance in genetic distance represented in one sub-cluster, Panvel-1, Ratnagiri-1 and Karjat-5 in another sub-clusters showed close genetic distance with blast sequence from NCBI. An average discrimination was seen for this loci.
2. ***rbcL* :** The tree represented 14 varieties separated out in two main clusters. First cluster contained Karjat-7, Panvel-1 Ratnagiri-4, Ratnagiri-8, Karjat-3, Ratnagiri-7, Ratnagiri-3, Karjat-4. Second cluster contained Panvel-2, Ratnagiri-2, Ratnagiri-24, Palghar-2, Karjat-5 and Ratnagiri-1 showed similarity which were placed in other sub-cluster. The two clusters showed a good differentiation between varieties.
3. ***psbA-trnH* :** The tree represented 14 varieties in the three main clusters, the longer arm represented the blast sequence separated out individually. The second cluster contained Palghar-2 independently and the third cluster had two sub-clusters of Karjat-5 and Ratnagiri-8 present in one and another sub-cluster was further bifurcated into minor sub clusters containing Ratnagiri-1, Ratnagiri-24, Karjat-1, Ratnagiri-7 Panvel-2, Karjat-7, Ratnagiri-2, Ratnagiri-4 in one of the minor sub-clusters and other contained Karjat-4, Karjat-3, Panvel-1, Karjat-2, Ratnagiri-3.

4. ***trnH-psbA* :** The phylogenetic tree represented 9 varieties in two major clusters for this loci, one cluster represented Karjat-1 and Karjat-2. Second cluster was divided into two sub-clusters, one sub-cluster having Ratnagiri-7 and Panvel-2 and another with Panvel-1, Palghar-2, Ratnagiri-8, Ratnagiri-4, Ratnagiri-3.
5. ***atpH-atpI* :** Two major clusters were seen representing Ratnagiri-4 in one and second cluster was represented with two sub-clusters one with Panvel-2 and all other 12 varieties closely related in another sub-cluster.
6. ***ITS2* :** The tree contained two clusters with each cluster divided into sub-clusters. First cluster represented two sub-clusters, one with Ratnagiri-3 and other with Karjat-4 and Ratnagiri-4. Second cluster presented one sub-cluster with Karjat-3 and Palghar-2 and another with Karjat-5 and Karjat-7.
7. ***petA-psbJ* :** Two main clusters were presented with blast sequence in one cluster, second cluster was bifurcated into two sub-clusters with Panvel-2 as one of it and rest all closely related varieties placed in the other.
8. ***trnK* :** The tree contained two main clusters, one represented Karjat-7 independently and other was divided into sub-clusters, one sub-cluster had Karjat-4 and other was again bifurcated into two groups of closely related varieties. First group showed Ratnagiri-2, Panvel-1, Ratnagiri-1, Karjat-1, Panvel-2 and second group with Karjat-3, Ratnagiri-7, Karjat-5, Palghar-2, Ratnagiri-3, Ratnagiri-8, Ratnagiri-4, Karjat-2, Ratnagiri-24.
9. ***ndhA* :** The tree depicted two major clusters with one containing Ratnagiri-1, Karjat-4, Karjat-7 and other

divided into sub-clusters. One sub-cluster with Ratnagiri-7 and all other varieties in the second sub-cluster. The tree contained of 13 varieties clustered as per genetic distance.

10. ***matK-1M-matK3RIM*** **:**Separate clusters were seen in the phylogenetic tree representing Ratnagiri-7 as distant in relation whereas Panvel-2 being closely related to the blast sequence from NCBI.

5.6. Barcode gap assessment and Barcode generation:

A robust DNA barcode should possess non-overlapping and separate variations or gaps or single nucleotide variations among the species. The changes in the sequences at nucleotide level and presence of extra sequence of nucleotides can be treated as successful discrimination where as identical sequences in all species fail to discriminate the same.

The sequences of amplified loci all 16 varieties with each primer set were aligned to check the inter and intraspecies similarities and differences in gene sequence. Among the ten different primers used in present studies, rbcL and ndhA showed 100% similarity indicating the region is highly conserved in rice varieties. The highest variation was observed in sequences of *psbA-trnH* and *trnH-psbA* followed by *petA-psbJ* and *trnK* , least variation was observed in sequences of *matK* and *atpH-atpI* primers.

The varieties which showed similar alignments were represented by a same barcode and those which showed variations are represented with individual barcodes. The natural barcodes usually consist of 400 to 800 bp DNA sequence in which each nitrogen base is represented by established colours (Adenine(A)=Green, Thymine(T)=Red, Cytosine(C)= Blue and Guanine=Black).The different barcoding loci exhibited different variations and gaps in the varieties and which were considered as a barcoding region.

5.7. Molecular Characterization by SSR markers:

Molecular markers play an important role and are an extremely powerful tool for varietal identification and determining variability between genetic materials of the cultivars. Varietal identification that directly utilizes DNA potentially addresses all the limitations associated with morphological and biochemical data. Characterization of cultivars using DNA profiles with highly polymorphic microsatellite markers has been used successfully for cultivar identification and verification in several crop species (Tautz, 1989).

SSR markers were highly informative and polymorphic as evident from its PIC value. The Polymorphism Information Content (PIC) value is a measure of polymorphism among varieties for a marker locus used in linkage analysis. The PIC value of each marker, which can be evaluated on the basis of its alleles, varied greatly for all tested SSR loci. The PIC values were calculated to find out the efficiency of primers in distinguishing individual genotypes. In the present study, 20 rice varieties were analyzed using 10 SSR primer pairs of which all 10 SSR primers were found to be polymorphic. A maximum number of six alleles were amplified by RM-335 and minimum 3 alleles by RM181.Earlier Yang *et al.,* (1994) observed an allelic range of 3-25 per SSR locus using different rice landraces and cultivars of both *Japonica* and *Indica* origin which is larger than that of Panaud *et al.,* (1996). Saghai *et* *al.,* (1994) observed 37 alleles in one SSR locus of barley, which was significantly a higher number when compared to the present study.

The Polymorphism Information Content (PIC) values were calculated to find out the efficiency of primers in distinguishing individual genotypes. The PIC values of primers ranged from 0.45 in SSR primer RM 338 to 0.65 in SSR pTA 248 primers with an average PIC value of 0.56. The higher the PIC value, the more informative is the SSR marker. Hence, primers RM 318, RM 276 and RM 343 were found to be highly informative. The PIC values are usually dependent on the genetic diversity of the accessions chosen (Garland, 1999) for the specific study. Markers with PIC values of 0.5 or higher are highly informative for genetic studies and are extremely useful in distinguishing the polymorphism rate of a marker at a specific locus (DeWoody *et al.,* 1995).

The pair wise similarity values ranged from 0.190 to 0.750. Maximum similarity value of 0.750 was noticed between Karjat-3, Karjat-4 And Karjat-5.Minimum similarity value of 0.016 was observed between Ratnagiri-2 and Karjat-184. From Jaccard's similarity coefficient and consensus tree, it may be concluded that, crossing between rice varieties having Jaccard's similarity coefficient in the range of 0.190 to 0.750 .Marker alleles were converted to binary scores based on their presence or absence as 1 and 0 respectively. UPGMA cluster analysis was performed using Jaccard's similarity coefficient Matrices calculated from SSR markers to generate a dendrogram for 23 rice varieties. A pairwise Similarity Index (SI) was calculated and the UPGMA based dendrogram of 23 rice varieties generated with Multivariate Statistical Package

(MVSP). It was observed that, two major clusters. The first major cluster consists of two sub clusters.

DNA based molecular markers have proven to be a powerful tool in identification of genetic variation and in the elucidation of genetic relationships within and among species, characterized by abundance and untouched by environmental influence (Powell, *et al.,* 1996). Thus, microsatellite markers provided more definitive separation of clusters indicating a higher level of efficiency for determining the relationship between closely related varieties.

CHAPTER VI

SUMMARY AND CONCLUSION

Rice holds a great gene pool which is studied by breeders in several crop improvement and breeding programs. Yearly, new high yielding varieties are developed by various organizations which needs identification and cataloguing as well as maintaining its purity. Hence, it is essential to assess the genetic diversity, DNA fingerprinting and DNA Barcoding of these varieties by using various biotechnological and molecular tools such as SSR markers and target oriented barcoding loci. Chloroplast DNA being maternally inherited has lower rate of variations and genetic alterations in it over the period of time and evolutionary aspects. Also the chloroplast genome sequences serves the purpose to locate the variations and nucleotide alterations occurred during the breeding and perpetuation of the plant species.

DNA barcoding is a fast, accurate and highly target oriented system for species identification which builds an ecological systems more approachable and easy user friendly by short DNA sequence instead of whole genome used for eukaryotes. An extensive application of DNA barcoding have been established in fields like preserving natural resources, protecting endangered species, identifying disease vectors, controlling agricultural pests, monitoring water quality, identification of medicinal plants, authentication of natural health products and food traceability. The most decisive factor of this technique is concerned with bio-security, bio-piracy and agricultural quarantine issues. It also would prove beneficial to create a unique identity of the varieties developed and established by the University.

The present study was conducted with an objective to characterize and add a molecular tag i.e. the DNA barcode for the 20 varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli. The study involved analysis of these varieties by study of 10 SSR markers and 10 barcoding loci markers. The barcoding primers focused at the chloroplast genome regions. The morphological characterization was based mainly on the basis of grain type. Molecular characterization was performed to get information on genetic variability at chloroplast genome sequence specific loci were used to analyize the varieties and generate barcodes for the same. The highly pure DNA was used for PCR amplification and sequencing process. Those primers which showed multiple banding, were gel eluted to obtain the target band for sequencing. The evolutionary analysis and phylogenetic relationship study was done using MEGA software. Several other online bioinformatics tools were used for sequence furnishing. CLUSTAL Omega online tool detected the conserved region in the varieties. The variations at nucleotide level in sequences was considered to construct the barcodes for the varieties.

The salient features of the study are mentioned as follows:

- DNA barcoding has a wider applications in the fields of taxonomy, conservation and identifications of different crops, trees and plants.
- In conclusion, this study provides a precursive assessment data that will be useful for extensive application of DNA Barcoding in not only rice varieties but other fruit crops,

ornamental and forestal plants of Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli.

- Two major types of seed grains were observed as bold and slender and categoried as short and long studied on the basis of morphological data.
- Amplification efficiency observed was about 100 % in 3 primers to 81.25 % in remaining primers.
- Fragment sizes range was 600-2900 bp in 10 loci.
- Sequencing efficiency of loci ranged from 100% to 18.75%.
- The estimated evolutionary divergence between sequences ranged from 0.000-2.170 (average 0.001-0.405).
- The Maximum Likelihood values of maximum transitional and transversional rate were observed.
- The parsimony informative sites was estimated with maximum 161 sites recorded in *psbA-trnH,* followed by 159 sites *trnK* and 33 *rbcL,* 21 in *matK* and number of variable sites reported highest in *psbA-trnH* 568, *trnK* 457 and *matK-1M-matK3RIM* 285.While nucleotide diversity (per site pi) reported maximum in *trnK* 0.1455.
- Phylogenetic relationship established using Neighbor Joining Method for all the 10 loci distinctly separated out Ratnagiri-2 for *matK,* Palghar-2 for *psbA-trnH* , Karjat-1 and Karjat-2 for *trnH-psbA* , Ratnagiri-4 for *atpH-atpI* , Panvel-2 for *petA-psbJ* , Karjat-7 for *trnK* and Ratnagiri-7 for *matK-1M-matK3RIM* .
- Both strand of *psbA-trnH* and *trnH-psbA* showed good discrimation power, also *petA-psbJ* showed higher discrimination. *trnK* distinctly seperated four of the 16 varieties in alignments. *rbcL* did not show discrimination but more of conserved region was seen whereas *matK* could give an average discrimination among the varieties.

- Barcodes were generated using online tool for the different loci considering the variations in sequence at nucleotide level.
- Comparative analysis of SSR markers was done with RM 338 highly informative with PIC value of 0.631 and observing the most frequent band size among 20 varieties for the microsatellites markers. Jaccard's similarity coefficient ranged from 0.190 to 0.750. Cluster analysis was established by UPGMA and similarity index by Jaccard's coefficient.

Conclusion:

The present study outlined the framework for the generation and development of DNA barcodes for rice crop. The study benefits to identify the varieties closely associated to each other on the basis of regions of chloroplast genome. The study adds additional information at genetic level about the varieties under study. DNA barcoding will benefit and open a new scope for molecular studies and also assist the DNA fingerprinting technique in molecular biology.

Future Prospects:

PCR amplification strategies, designing of new primers from the specific site of rice chloroplast genome will help in precise amplification of reproducible chloroplast genome specific loci. Validiation of *matK, rbcL, trnK,* and *psbA-trnH* in large groups with the availability of sequencing informative data more loci will be identified and would be helpful in analyzing other rice hybrids. The focused study of these loci will help to establish comparison and relationship with wild relatives as well as reflect the significance of these loci. Mainly the other cereal, ornamental and forest plants could be characterized and studied by the DNA barcoding technique in the future.

LITERATURE CITED

Ajmal , M.A., Gyulai, G., Hidvegi, N., Kerti, Al Hemaid, F.M.A., Pandey, A.K., and Lee, J.(2014). The changing epitone of species identification – DNA barcoding. *Saudi Journal of Biological Sciences*, **21**, 204-231.

Alphanso P. (2014). Molecular Characterization and DNA Barcoding of Zanthoxylum rhetsa. Kelkar Education Trust's Scientific Research Centre - Project Report.SRC/MB/6.

Bafeel, S. O., Arif, I. A., Bakir, M. A., Khan, H. A., AlFarhan, A. H., Al homaidan, A. A., Ahmed, A., and Thomas, J. (2011). Comparative evaluation of PCR success with universal primers of maturase k (matk) and ribulose – 1, 5 – biphosphate carboxylase oxygenase large subunit Crbcl for barcoding of same and plants. *Plant Omics Journal*, **4(4)** : 195 – 198.

Balakrishnan R. (2007). Species concepts, species boundaries and species identification: a view from the tropics. Syst. Biol. **54** (4), 689-693.

Bhagwat, R.M.(2015). DNA Barcoding of same forest tree species of Western Ghats. *Plant Molecular biology group Division of Biochemical Sciences, CSIR – national Chemical Laboratory, Pune, India.*

Bruni, I., De Mattia, F., martellas, S., Galimbrti, A., Savadori, P., Casiraghi, M., Nimis, P.L., and labra, M.(2012). DNA barcoding as an Effective Tool in Improving a Digital Plant Identification System : A case study for the Area of Mt. Valerio, Trieste (NE- Italy). *PLOS ONE*, **vol. 7**, Issue 9, e43256.

Causse , M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K ., Xiao, J. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics*, **138 (4)**, 1251-1274.

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceeding of the NationalAcademy of Sciences.*

Chakravarthi B.K. and Naravaneni R.,(2006).SSR marker based DNA fingerprinting and diversity study in Rice (*Oryza sativa.* L).*African J.of Biotechnology* **5(9)**, 684-688.

Chan E.Y. (2005). Advances in sequencing technology.Mutation Research/Fundamental and Molecular Meechanisms of Mutagenesis, **Vol.573**,13-40.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., haider, N., and Savolainen, V.(2005). Land plants and DNA Barcodes : short - term and long – term goals. *Phil. Trans. R. soc.8*, **360**, 1889 – 1895.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., vande Berg, C., Madrinam, S., Petersen, G., Seberg, O., Jorgsensen, T., Cameron , K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barradough, T.G., Kely, L. & Wilkinson, M., (2007). A proposal for a standardized protocol to barcode all land plants. *TAXON* **S6(2)**, 295 – 299.

Choudhary, K., Choudhary, O. P., and Shekhawat, N. S. (2008). Marker assisted selection: a novel approach for crop improvement. Am. *Eurasian J. Agric.* **1**:26–30.

Choudhury B., Khan M. L. and Selvadurai, D.(2013).Genetic structure and diversity of indigenous rice (*Oryza sativa*) varieties in the Eastern Himalyan region of Northeast India. *Spriger Plus* **2**:228.

Chungada A. S, Gokhale N. B, Patil D. M, Sawardekar S. V and Patil P. P (2016) Molecular Screening of Rice Germplasm for Biotic and Abiotic Stresses and Their Diversity Study by Using SSR Markers. *J. Indian Soc. Coastal Agric. Res.* **34**(2): 7-14

Clegg, M. T and M. L. Durbin. (1990). Molecular approaches to the study of plant biosystematics." Australian systematic botany **3.1**:1-8.

Daugelaite J, Driscoll A. and Sleator, R.D.(2013). An overview of Multiple Sequence Alignment and Cloud Computing in Bioinformatics. Hindawi Publishing Corporation, ISRN Biomathematics **vol, 2013**,1-14.

De Las Rivas, J., Lozano, J.J., and Ortiz, A.R. (2002). Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. Genome research **12**: 567-583

De Vere, N., Rich, T.C.G. trinder, S.A., and Long, C.(2015). DNA Barcoding for Plants. Plant Genotyping : Methods and Protocols, *Methods in Molecular Biology*, **vol. 1245**, 105 – 118

De Woody, J.A., Honeycutt, R.L., and Skow, L.C. (1995). Microsatellite Markers in white-tailed deer. *J.Heredity*, **86**: 317-319

Doyle J.J and Doyle J. L(1990).Isolation of plant DNA from fresh tissue.*Focus*,**12**:13-15.

Doyle, J. and J. Doyle (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull.*, **19 (11)** : 11-15.

FAO Rice Information, The Secretariat of the International Rice Commission Crop and Grassland Service Plant Protection and Protection Division Agricultural Department, vol 3, December 2002.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Gramah, S.W., Newmaster, S.G., Husband, B.C., Peray , D.M., Hajibabaei, M., and Barrett, S.C.H. (2008). Multiple Multilocus DNA barcodes from the Plstid Genome Discriminate Plant Species Equally well. *PLOS ONE*, **3(7)** e802

Felix A., Isralewitz B., Lutery Z. – Schulten, Sethi A., Pogordov T.(2005). Bioinformatics and sequence Alignment University of Illinois at Urbana – Champaign Luthey – Schulten Group Beckman institute for Advanced Science and Technology Theoretical and Computational Biophysics Group San Francisco Worshop.

Fitzhugh Kirk (2006).DNA Barcoding: An instance of technology-driven science?.*Bioscience*, **vol.56 (6)**, 462-463.

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Van Alphen stahl, J., Kelly, L.J., Wikstrom, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W., and Wilkinson, M.J.(2009). Selection of candidate coding DNA barcoding regions for use an land plants. *The Linnean society of London, Botanical Journal of the Linnean society,* **S9**, 1 – 11.

Fu, Y.X. and Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics **133**: 693-709.

Galimberti, A., De mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., martellos, S., labra,M.(2012). DNA BArcoding as a new tool for food traceability. *Food Research International*,**(50)** 55 – 63.

Gao Li - Zhi, Liu Yun – Long, Zhang, O., Li, W., Gao, J., Liu, Y., Li, K., Shi, C., Zhao, Y., Zhao, Y.J., Jiao, J.Y., Mao, S.Y., Gao, C.W., and Eichler. E.E. (2019). Evolution of Oryza chloroplast genome promoted adaption to diverse ecological habitats. *Communications Biology*. **2**:278 1 – 13.

Garland, S. M., Lewin, L. Abedinia, M. Henry, R. and Blakency, A. (1999). The use of microsatellite polymorphisms for the identification of Australian breeding lines of rice (*Oryza sativa* L.) *Euphytica,* **108**: 53-63.

Gonzlaez, M.A., Baraloto, C., Engel, J., Mori, S.A., Petronelli, P., et – al. (2009). Identification of Amazonian Trees with DNA Barcodes. *PLOS ONE*, **4(10)** : e7483.

Gopalan, C., Rama, B.V., Balasubramanian, S. (2007) Nutritive Value of Indian Foods. National Institute of Nutrition (NIN), ICMR, India.

Gracia, A. A. F., Benchimol, L. L., Antonica, M. M., Geraldi, I. O., Deuza, A. P. (2004).Comparison of RAPD, RFLP, AFLP and SSR marker for diversity studies in tropical maize inbred lines. *Euphytica*, **108**: 53-63.

Guo, L., Sui., Z., Zhang, S., Ren, Y., and Li, Y.(2015). Comparison of potential diatom 'barcode' genes (the 185, RNA gene and ITS, Col, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of systematic and Evolutionary Microbiology*, **65**.

Gupta, P. K. and Varshney, R. K. (1999). Molecular markers for genetic fidelity during micropropagation and germplasm conservation. *Curr. Sci.,* **76**: 1308-1310.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwaches, W., and Hebert, P.D.N. (2005). DNA Barcodes distinguish species of tropical Lepidoptera. *PNAS*, **vol. 103**, no.4, 968 – 971.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., and Hickey, D.A. (2007). DNA barcoding : how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics,* **vol. 23** no.4. 168 – 172.

Hebert, P.D.N. and Gregory, T.R.(2005). The promise of DNA barcoding for *Taxonomy. Syst.Biol.* **54(5**): 852 – 859.

Hebert, P.D.N., Cywinska, A., ball, S.L., and deklaard, J.R.(2003). Biological identifications through DNA barcodes, *Proc. R. Soc. Land.* B. **270**, 313 – 321.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome : Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Nol. Gen. Genet*, **217** : 185 – 194.

Hogg, I.D., and Hebert, P.D.N. (2015). Biological Identification of Springtails (Hexapoda: Collembola) from the Canadian Artic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology*, **82** : 749 – 754.

Hollingsworth, P.M., Graham, S.W., and Little, D.P. (2011). Choosing and using a Plant DNA Barcode. *PLOS ONE* **6(5)**: e19254.

Hollingsworth, P.M., De-Zhu Li, van der bank, M., and Twyford, A.D. (2016). Telling plant species apart with DNA : from barcodes to genomes. *Phil. Trans.r. soc.* B **371** : 20150338.

Ishii T., Terachi T. and Tsunewaki K. (1988). Restriction endonuclease analysis of Chloroplast DNA from A-genome diploid species of rice. Jpn. J. Genet. **63**, 523-536.

Islam, M. R., Gregorio, G. B., Salam, M. A., Collard, B. C. Y., Singh. R. K., Hassan, L. (2012). Validation of *SalTol* linked markers and haplotype diversity on chromosome 1 of rice. *Mol Plant Breed*, **3**:103–114.

Janzen, O.H., and CBOL Plant Working Group 1 (2009). A DNA barcode for Land plants. *PNAS*, **vol.106(31)**, 12794 – 12797.

Jurate Daugelaite, Aisling, O' Driscoll, and Roy D. Sleator, *Hndavi Publishing Corporation, ISRN Bioinformatics,* **vol.2013**, Article ID 615630, 14 pages.

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. *The National Academy of Sciences of the USA (PNAS)*, **vol. 102**, no.23, 8369 – 8374.

Kress, W.J., and Erickson, D.L. (2007). A Two- Locus Globle DNA Barcode for land plants : The Coding rbcl Gene Complements the Non – Coding trnH – psbA Spacer Region. *PLOS ONE* **2(6)** : e508

Kress, W.J., and Erickson, D.L. (2008). DNA barcodes: Genes, genomics, and bioinformations. *PNAS*, **vol.105**, no.8, 2761 – 2762.

Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjur, O., and Bermingham, E.(2009). Plant DNA Barcodes and a community phylogeny of a tropical forest dynamics plot in panama. *PNAS,* **vol. 104**, no.44, 18621 – 18626.

Kress , J.W., Erickson, D.L., Swenson, N.G., Thompson, J., Uriarte, M., Zimmerman, J.K.(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*, **vol.5**, Issue 11/ e 15409

Kumar,S., Hahn, F.M., McMahan, C.M., Cornish, K., and Whalen, M.C. (2009). Comparative analysis of the complete sequence of the plastid genome of Parthenium argentatum and identification of DNA barcodes to differentiate Parthenium Species and lines. *BMC plant Biology*,**9:131** doi : 10.1186/1471-2229/9/131.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., maurin, O., Duthoit, S., barradough, T.G., and Savolainen, V.(2007). DNA barcoding the floras of biodiversity hotspots. *PNAS*, **vol.105**, no.8, 2923 – 2928.

Lakshmana R.D.C., Kiran K., Srinivas R.S.H., Kanupriya A.C., Singh T. (2012) SSR-based DNA barcodes as a tool for identification of eggplant genotypes.*Int J Veg Sci* **8(3)**:260-271.

Li, S.C., Wang, C.H., yen, C.E., Chang, C.(2017). DNA barcodes and identification of the verieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. *Journal of Food & Drug Analysis*, **vol. 25**, 260 – 274

Li, Xi., Yang, Y., Henry, R.J., Rosseto, M., Wang, Y., and chen, S.(2015). Plant DNA barcoding : from gene to genome. *Biol. Rev.*, **90**,157 – 166.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

Ma, H. L., Zhu, J.G., Liu, G., Xie., Z. B, Wang, Y., Yang, L., Zeng, Q., (2007) Availability of soil nitrogen and phosphorus in a typical rice–wheat rotation system under elevated atmospheric [CO2]. *Field Crop Res,* **100**: 44–51.

Mahalingam., A. Saraswathi, R. and Ramalingam, J.(2013).Simple sequence repeat (SSR) markers for assessing genetic diversity among the parental lines of hybrid rice (*Oryza sativa* L.).*Afr.J.Biotechnology*. **12 (33)**:5105-5116.

Matsuoka, Y., Yamazaki, Y., Ogihara, Y., and Tsunewaki, K. (2002). Whole chloroplast Genome comparison of Rice, Maize, and Wheat : Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Mol. Biol. Evol.* **19(12)**: 2084 – 2091.

Miller, S.E. (2007). DNA barcoding and the renaissance of taxonomy. Proceedings of the National Academy of Sciences **104**: 4775-4776.

Mishra, P., kumar, A., nagireddy, A., mani, D.N.,Shukla, A.K., Tiwari, R., and sundaresan, V.(2016). DNA barcoding : an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology journal*, **(14)**, 8 -21.

Muhammad S.R., Md. Rezwan Molla, Md. Samsul Alam and Lutfur Rahman (2009).*Australian Journal of Crop Science* **3(3)**: 122-128.

Nandakumar, N., Singh, A. K., Sharma, R. K., Mohapatra, T., Prabhu, K. V., Zaman, F. U. (2004). Molecular fingerprinting of hybrids and assessment of genetic purity of hybrid seeds in rice using microsatellite markers. *Euphytica.* **136**: 257-264.

Nash JHE (1991) DNA frag, Version 3.03. Institute for Biological Sciences, 1991,National Research Council of Canada, Ottawa, Ontario, Canada.

Newmaster, S.G., Fazehas, A.J., and Ragupathy, S.(2006). DNA barcoding in land plants : evaluation of rbcl in a multigene tiered approach. *Can. J. Bot.* **84**: 335 – 341

Ni, J.P.M and Mackill, D. J. (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Science*, **42**: 601–607.

Nielsen, R. and Matz, M. (2006) Statistical approaches for DNA barcoding. Syst. Biol. **55**, 162–169.

Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, M., and Henry, R.J. (2011) . Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, 328 – 333.

Okello M, Sawardekar S. V, Ghokhale N. B , Waghmode B. D and Patil D. M (2017) Molecular Characterization of Rice Germplasm (*Oryza sativa* L.) using Simple Sequence Repeat (SSR) Markers. *Advanced Agricultural Research & Technology Journal.* **I**: 92-97

Panaud, O., Chen, X. and McCouch (1996). Development of microsatellite markers and characterization of Simple Sequence Length Polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet.,* **252:** 597–607.

Powell, W., Machray, G. C. and Provan, J. (1996). Polymorphism revealed by simple sequence repeats**.** *Trends Plant Sci.* **1**: 215-222.

Provan, J., Corbett, G., MvNicol, J.W., and Powell, W.(1996). Chloroplast DNA variability in wild and cultivated rice (*Oryza Spp.*) reveled by polymorphic chloroplast simple sequence repeats. *Genome*, **40**: 104 – 110.

Purshothaman , N., Newmaster, S.G., Ragupathy, S., Stalin, N., Suresh, D., Arunraj, D.R., Gnanasekaran, G., vassou, S.L., narasimhan, D., and Parani, M. (2014). A tiered barcode authentication tool to differentiate medicinal Cassia species in India. *Genetics and molecular research*, **13(2)** : 2959 – 2968.

Radulovici, A.E., Archambault, P., and Dufreshe, F. (2010).DNA Barcodes for Marine Biodiversity : Moving fast Forward? *Diversity*, **(2)**, 450 – 472.

Rahman M.S., Sohag M.K.H., and Rahman L. (2010). Microsatellite based DNA fingerprinting of 28 local rice varieties (*Oryza sativa* L.) of Bangladesh *J.Banglaesh Agril.Univ.***8(1)** :7-17.

Raja, K.S., Mohanasundaram B. S. and Ramalingam S. (2015). DNA Barcoding: a genomic-based tool for authentication of phytochemicls and its products. Botanics : Targets and Therpy. **5**, 77-84.

Rajaguru Bohra (2011).Study on awareness and willingness to adopt fast track DNA fingerprinting technology in seed certification for seed genetic purity analysis. M.Tech. thesis, Department of Plant molecular Biology and Biotechnology Centre for Plant Molecular Tamil Nadu Agricultural University, Coimbatore.

Ramakrishna, W., Lagu, M.D., Gupta, V.S., and Ranjekar, P.K. (1994).DNA fingerprinting in rice using oligonucleotide probes specific for simple repetitive DNA sequences. *Theo. Appl. Genet.*, **88**:402-406

Ramu Chenna, Sugawara H., Koike T., Lopez R., Toby J. Gibson, Desmond G.H. and Thompson J.D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, **Volume. 31**, No. 13, 3497 – 3500.

Robert C Edgar and Sersfim batzoglou, (2006), Multiple sequence alignment. *Current Opinion in Structural Biology*, **19** : 1 – 6.

Roy, S.C. (2015). Phylogenetic Relationship among the Wild Rice [Oryza rufipogon Griff.] of NBU Campus and Cultivated Rice as Revealed by Chloroplast matK Gene. International Journal of Agriculture Innovations and Research **3**: 1768-1774 .

Rubinoff, D and Holland, B.S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Systematic biology **54**: 952-961.

Rubinoff, D. (2006). Utility of mitochondrial DNA barcodes in species conservation. Conservation Biology **20**: 1026-1033.

Rusch D.B., Haplern A.L., Sutton G, Heidelberg K B , Williamson S, Yooseph S, WuD, *et. al.* (2007). The Sorcerer II Global Ocean Sampling expendition :Northwest Alantic through eastern tropical Pacific. PLOS Biol **5**:e77.

Rydberg, A.(2010). DNA Barcoding as a tool for the Identification of unknown plant material. A case study on medicinal roots traded in the medina of Marrakech.

Saarela, J.M., Sokoloff, P.C., Gillespie, L.J., consaul, L.L., and Bull, R.O.(2013). DNA Barcoding the Canadian Arctic Flora : Core Plastid barcodes (rbcl + matk) for 490 Vascular Plant Species. *PLOS ONE*, **vol. 8** (10) : e77982.

Saddhe, A.A., and Kumar,K.(2018). DNA Barcoding of Plants : Selection of core markers for taxonomic groups. *Plant science Today*, **5(1)** : 9-13.

Saghai, M.M.A,Biyashev, R.M., Yang, G.P., Zhang, Q., Allard, R.W. (1994). Extraordinary polymorphic microsatellite DNA in barley: species, diversity, chromosomal locations, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America.* **91(12)**; 5466.

Sanger F, Nicklen S and Coulson AR. (1977). DNA sequencing with chainterminating inhibitors. Proceedings of the National Academy of Sciences_**74**: 5463-5467.

Schindel D.E. and Miller S.E. (2005). DNA barcoding a useful tool for taxonomists.Nature. **vol.435**, no.7083.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E & Small, R.L (2005). The tortoise and the hare II: Relative utility of 21-noncodingchloroplasts. American Journal of Botany.**92**:142-166.

Shimada, H., and Sugiura, M.(1991). Fine structural features of the chloroplast genome : Comparison of the sequenced chloroplast genome. *Nucleic Acid Research*, **vol.19**, No.5, 983 – 995.

Singh Ambalika., (2013). Determination of Genetic Diversity among different Rice varieties (*Oryza sativa* L.) with SSR markers. *Bio-science Research Bulletin* **29**:39-51.

Singh, J., and Banerjee, S.(2018). Utility of DNA Barcoding Tool for conservation and Molecular Identification of Intraspecies of Rice genotypes belonging to Chhattisgarh using RBCL and matk Gene sequences. *Plant Archives*, **vol. 18**, Special Issue , 69-75.

Singh, J., Kakade, D.P., Wallalwar, M.R., Raghuvanshi, R., Kongbrailatpam, M., Verulkar, S.B., and Banerjee, S. (2017). Evaluation of Potential DNA Barcoding Loci from Plastid Genome : Intraspecies Discrimination in Rice (*Oryza* Species). *Int. J. Curr. Microbiol. App. Sci.* **6(5)** : 2746 – 2756.

Skuza, L., Szucko, I., Filip, E., and Adamczyk, A. (2019). DNA Barcoding in selected species and subspecies of Rye (Secale) Using Three Choloroplast Loci (matk, rbcl, trnH – psbA). *Not Bot Horti Agrobo*, **47(1)** : 54 – 62.

Sundari, R.K., Balachandran, Mohanasudaran, S., Ramalingam, S. (2015). DNA Barcoding a genomic based tool for authentication of phytochemicals and its products. *Botanics : Targets and Therapy*, **(5)** 77 – 84.

Taberlet, P. et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. **35**, **e14106** : 31 12794-12797.

Tang, J., Xia, H. a., Cao, M., Zhang, X., Zeng, W., hu, S., tong, W., Wang, J., Wang, J. and,Yu J.(2004). A comparison of rice chloroplast genomes. *Plant Physiol.*, **135(1)** : 412 – 420.

Tautz. (1989). Hyper variability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17:** 6463-6471.

Thompson, J. A., Nelson, R. L. And Vodkin, L. O. (1998). Identification of diverse soybean germplasm using RAPD markers. *Crop Sci*, **38**: 1348-1355.

Tom Madden. The BLAST sequence Analysis Tool. The NCBI handbook, Chapter 16.

Valentini, A., Pompanon, F., and Taberlet, P.(2008). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **vol.24,** No. 2, 110 – 117.

Vanisri S., Rani D., Gattu S., Jamal Md, Mulinti S., Nagireddy R.K., Eruvuri R., Yanda R. (2017).DNA Fingerprinting for Identification of Rice Varieties and seed Genetic Purity Assessment. *Agricultural Research.*

Vanniarajan, C., K. K. Vinod, and Pereira A. (2012).Molecular evaluation of genetic diversity and association studies in rice (*Oryza sativa L.*). Journal of genetics 91.1: **9**.

Vere , N., Rich T.C.G., Trinder S.A., and Long C. (2015). DNA Barcoding for Plants. Plant Genotyping: Methods and Protocols Methods in Molecular Biology, **vol.1245**,101-118

Vijayan, K. and Tsou, C.H.(2010). DNA barcoding in plants : taxonomy in a new perspective. *Current science*, **vol. 99**, No.11. 1530 – 1541.

Vogler, A. and Monaghan, M. (2007). Recent advances in DNA taxonomy. Journal of zoological systematics and evolutionary research **45**: 1-10

Wang, N., Jacques, F.M.B., Milne, R.I., Zhang, C.Q., and Yang, J.B. (2011). Dna Barcoding of Nyssaceae (cornales) and taxonomic issues. *Botanical Studies*, **53** : 265 – 274.

Waters, D.L.E., Nock, C.J., Ishikawa, R., Rice, N., and Henry, R.J.(2010). Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecology and Evolution*, **2(1)** : 211 - 217.

Whitford, R., Gilbert, M., Langridge, P. (2010). Biotechnology in agriculture. In: M.P. Reynolds, ed., Climate change and crop production, *CABI Series in Climate Change,* **1**: 219-244, CABI, UK.

Wu , K.S and Tanksley ,S.(1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics*. **241**:225-235

Xiwen Li, Yang Y., Henry, R.J., Rossetto, Wang Y., Chen S. (2015). Plant DNA Barcoding: from gene to genome.Wiley Online Library. Biological Reviwes, **vol.90** (1).

Yang, G. K., Saghai Maroof, M. A., Zang, C. G. Xu, Q. and Biyashev, R. M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol. Gen. Genet.*, **245**: 187-194.

DNA BARCODING FOR IDENTIFICATION OF RICE (*Oryza sativa* L.) VARIETIES DEVELOPED BY Dr. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH, DAPOLI.

2020

Student's Name: Mhatre Dhanashree Prashant **Research Guide:** Dr. S.V. Sawardekar

ABSTRACT

The present study was conducted with an objective to characterize and add a molecular tag i.e. the DNA barcode for the 20 varieties developed by Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli. The study involved analysis of these varieties by using 10 SSR markers and 10 barcoding loci markers. The evolutionary analysis and phylogenetic relationship study was done using MEGA software. CLUSTAL Omega online tool detected the conserved region in the varieties.

DNA barcodes derived from the chloroplast genome were used to identify varieties and in the conservation of breeding resources. The gene regions are chosen because they have less intraspecific (within species) variations than interspecific (between species) variations, which are known as the "Barcoding Gap". DNA barcoding has wider applications in the fields of taxonomy, conservation and identifications of different crops, trees and plants.

Two major types of seed grains were observed as bold and slender and categorized as short and long studied on the basis of morphological data. Amplification efficiency observed was about 100 % in 3 primers to 81.25 % in remaining primers. Fragment size range was 600-2900 bp in 10 loci. Sequencing efficiency of loci ranged from 100% to 18.75%. The estimated evolutionary divergence between sequences ranged from 0.000-2.170 (average 0.001-0.405).

The Maximum Likelihood values of maximum transitional and transversional rate were observed. The parsimony informative sites was estimated with maximum 161 sites recorded in *psbA-trnH*, followed by 159 sites *trnK* and 33 *rbcL*, 21 in *matK* and number of variable sites reported highest in *psbA-trnH* 568, *trnK* 457 and *matK-1M-matK3RIM* 285.While nucleotide diversity (per site pi) reported maximum in *trnK* 0.1455.

Phylogenetic relationship established using Neighbor Joining Method for all the 10 loci distinctly separated out Ratnagiri-2 for *matK,* Palghar-2 for *psbA-trnH* , Karjat-1 And Karjat-2 for *trnH-psbA* , Ratnagiri-4 for *atpH-atpI* , Panvel-2 for *petA-psbJ* , Karjat-7 for *trnK* and Ratnagiri-7 for *matK-1M-matK3RIM* . Both strand of *psbA-trnH* and *trnH-psbA* showed good discrimation power, also *petA-psbJ* showed higher discrimination. *trnK* distinctly separated four of the 16 varieties in alignments. *rbcL* did not show discrimination but more of conserved region was seen whereas *matK* could give an average discrimination among the varieties. Barcodes were generated using online tool for the different loci considering the variations in sequence at nucleotide level.

Comparative analysis of SSR markers was done with RM 338 highly informative with PIC value of 0.631 and observing the most frequent band size among 20 varieties for the microsatellites markers. Jaccard's similarity coefficient ranged from 0.190 to 0.750. Cluster analysis was established by UPGMA and similarity index by Jaccard's coefficient.

In conclusion, this study provides a precursive assessment data that will be useful for extensive application of DNA Barcoding in not only rice varieties but other fruit crops, ornamental and forestal plants of Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli.

Keywords: DNA barcoding, *matK, psbA-trnH, trnH-psbA, rbcL, trnK,* Parsimony informative sites, Neighbor Joining Method, CLUSTAL Omega, MEGA 7.0.

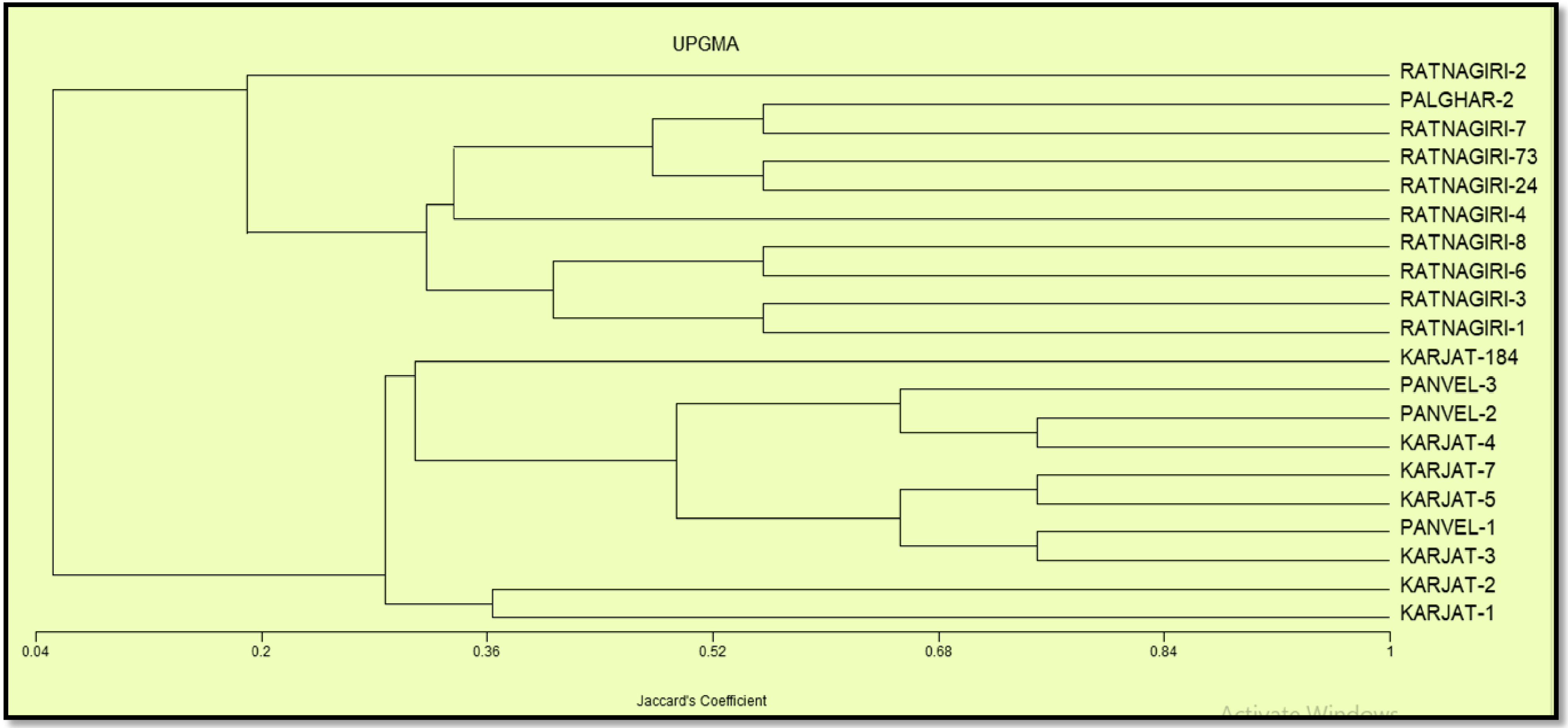


Fig.4.11: Dendrogram constructed using Jaccard's Similarity Coefficient.

LIST OF PLATES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF TABLES

Plate 1: Morphological Characterization of the 20 rice varieties

				
Karjat-1	Karjat-2	Karjat-3	Karjat-4	Karjat-5
				
Karjat-7	Karjat-184	Ratnagiri-1	Ratnagiri-2	Ratnagiri- 3
				
Ratnagiri-4	Ratnagiri-24	Ratnagiri-73	Ratnagiri-6	Ratnagiri-7

				
Ratnagiri-8	Panvel-1	Panvel-2	Panvel-3	Palghar-2

PLATE A: Mega alignment of *matK* sequences showing variable regions

PLATE B : Mega alignment of *rbcL* sequences showing variable regions

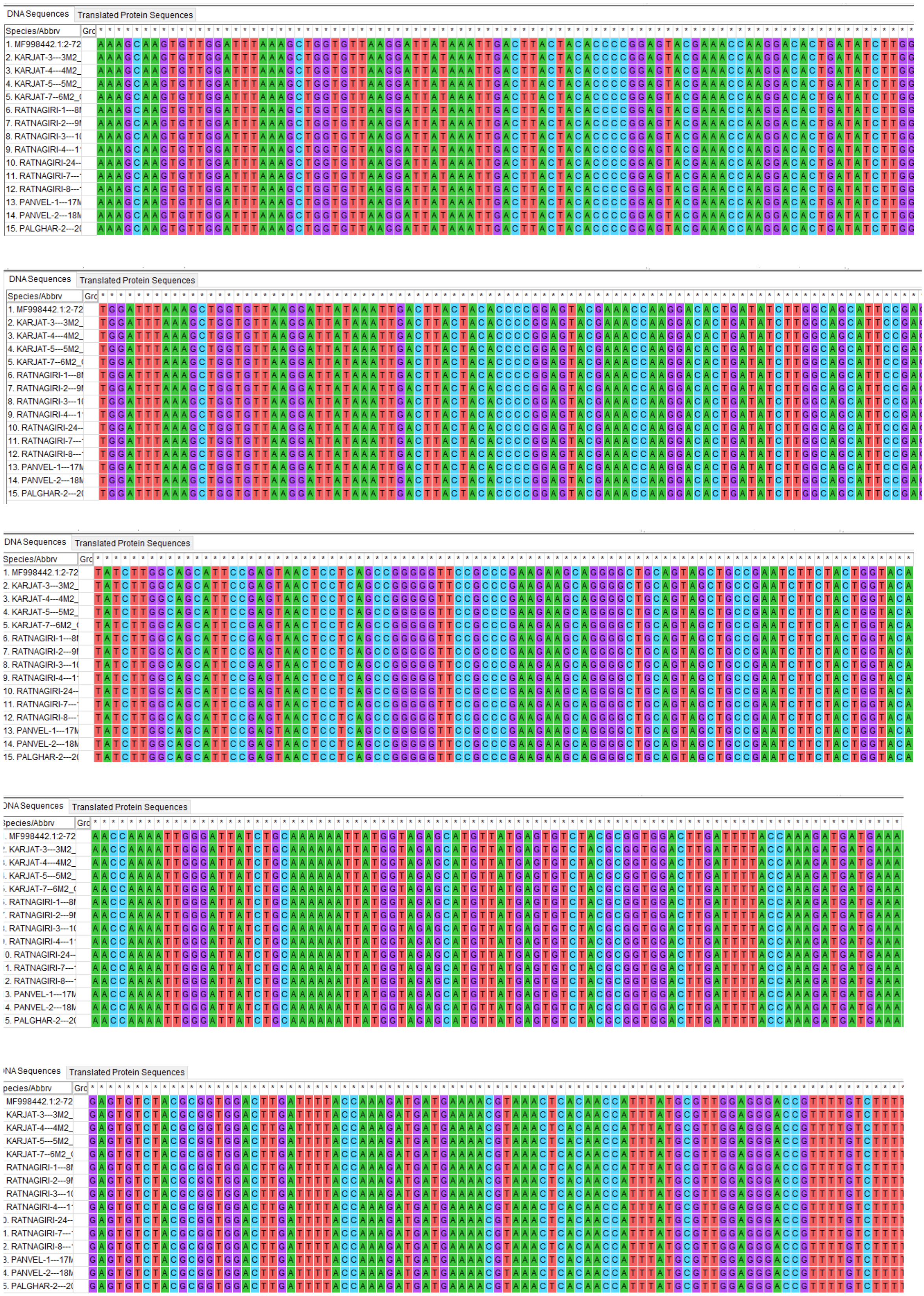
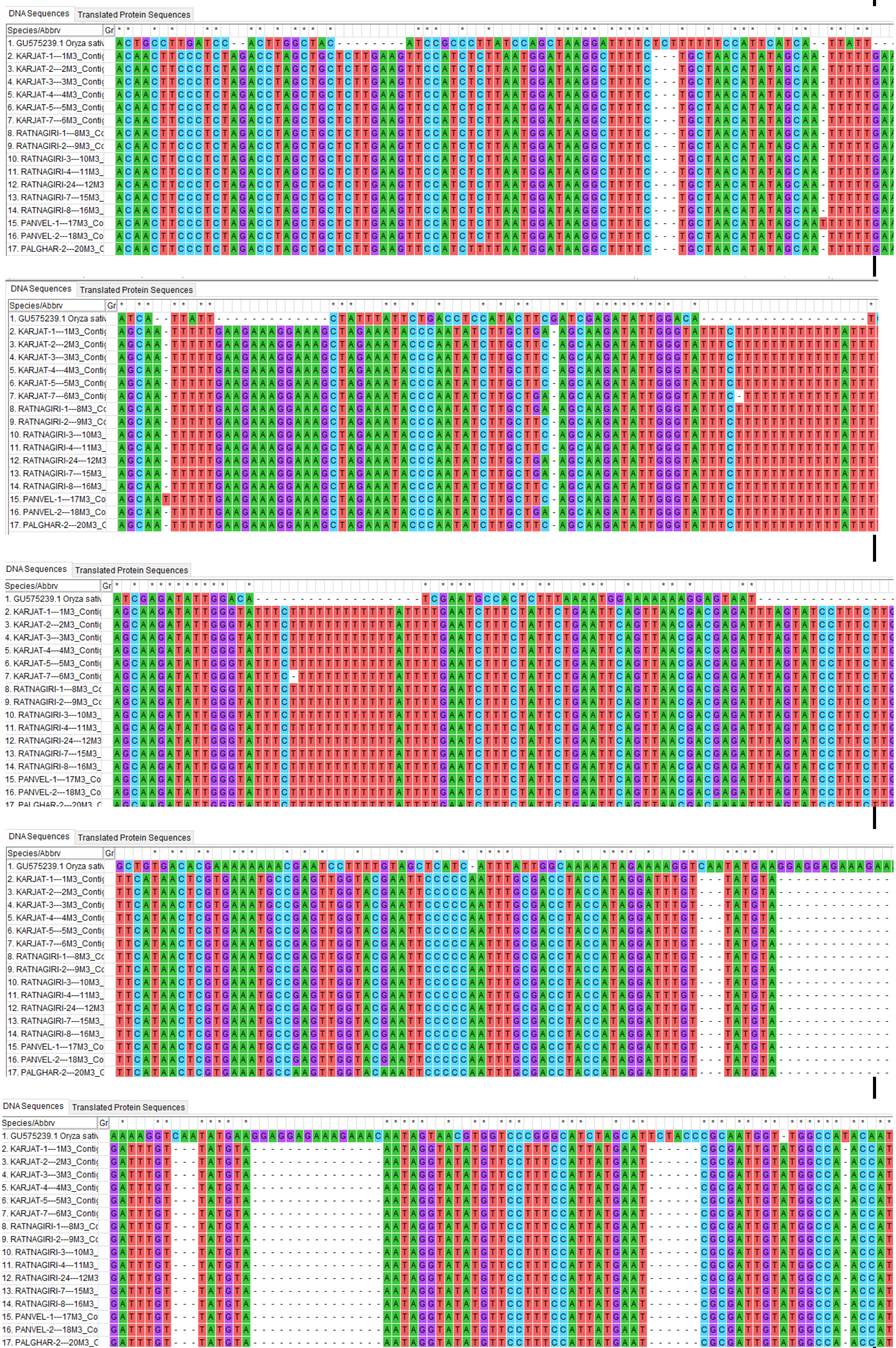
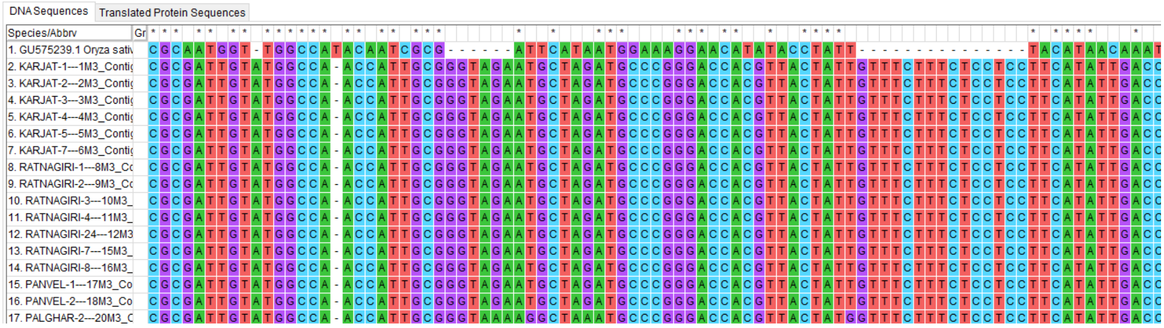
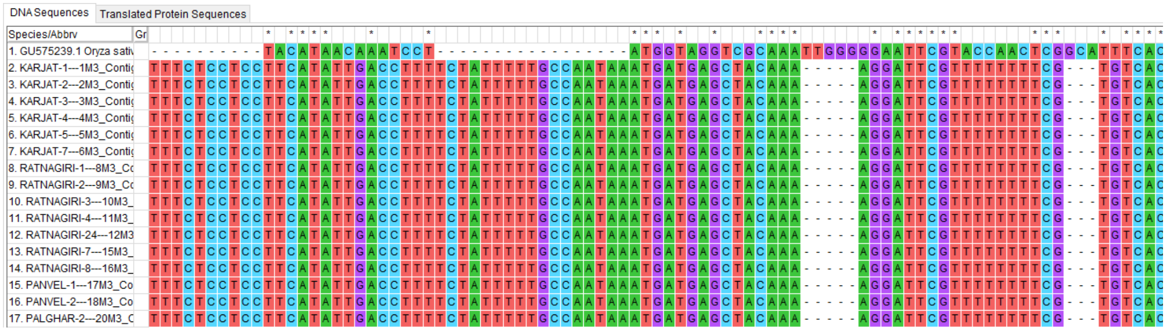
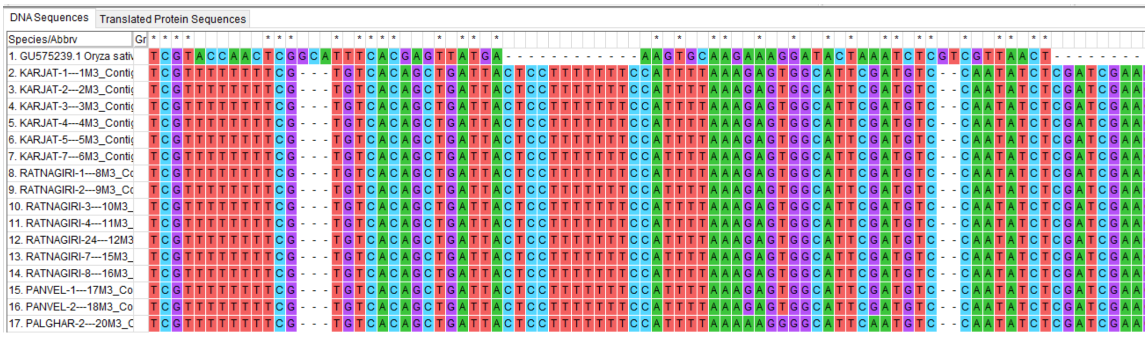


PLATE C: Mega alignment of *psbA-trnH*sequences showing variable regions




DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. GU575239.1 Oryza sati
2. KARJAT-1---1M3_Conti
3. KARJAT-2---2M3_Conti
4. KARJAT-3---3M3_Conti
5. KARJAT-4---4M3_Conti
6. KARJAT-5---5M3_Conti
7. KARJAT-7---6M3_Conti
8. RATNAGIRI-1---8M3_Cc
9. RATNAGIRI-2---9M3_Cc
10. RATNAGIRI-3---10M3_
11. RATNAGIRI-4---11M3_
12. RATNAGIRI-24---12M3
13. RATNAGIRI-7---15M3_
14. RATNAGIRI-8---16M3_
15. PANVEL-1---17M3_Co
16. PANVEL-2---18M3_Co
17. PALGHAR-2---20M3_C


DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. GU575239.1 Oryza sati
2. KARJAT-1---1M3_Conti
3. KARJAT-2---2M3_Conti
4. KARJAT-3---3M3_Conti
5. KARJAT-4---4M3_Conti
6. KARJAT-5---5M3_Conti
7. KARJAT-7---6M3_Conti
8. RATNAGIRI-1---8M3_Cc
9. RATNAGIRI-2---9M3_Cc
10. RATNAGIRI-3---10M3_
11. RATNAGIRI-4---11M3_
12. RATNAGIRI-24---12M3
13. RATNAGIRI-7---15M3_
14. RATNAGIRI-8---16M3_
15. PANVEL-1---17M3_Co
16. PANVEL-2---18M3_Co
17. PALGHAR-2---20M3_C


DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. GU575239.1 Oryza sati
2. KARJAT-1---1M3_Conti
3. KARJAT-2---2M3_Conti
4. KARJAT-3---3M3_Conti
5. KARJAT-4---4M3_Conti
6. KARJAT-5---5M3_Conti
7. KARJAT-7---6M3_Conti
8. RATNAGIRI-1---8M3_Cc
9. RATNAGIRI-2---9M3_Cc
10. RATNAGIRI-3---10M3_
11. RATNAGIRI-4---11M3_
12. RATNAGIRI-24---12M3
13. RATNAGIRI-7---15M3_
14. RATNAGIRI-8---16M3_
15. PANVEL-1---17M3_Co
16. PANVEL-2---18M3_Co
17. PALGHAR-2---20M3_C

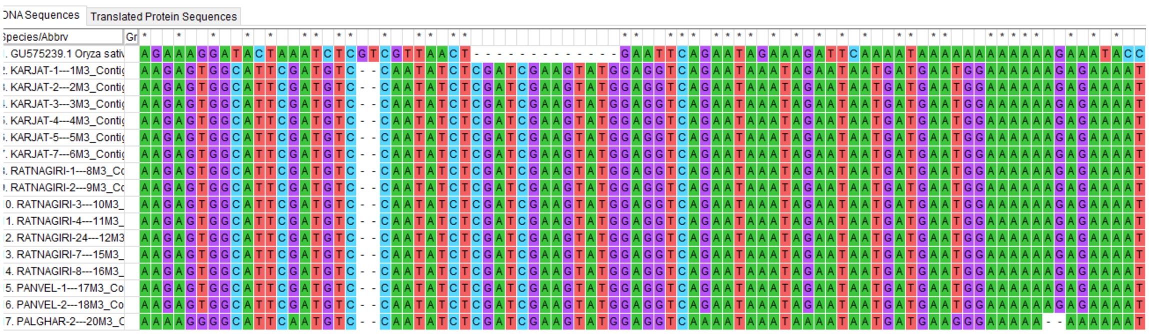

DNA Sequences
Translated Protein Sequences
Species/Abbrv
GU575239.1 Oryza sati
KARJAT-1---1M3_Conti
KARJAT-2---2M3_Conti
KARJAT-3---3M3_Conti
KARJAT-4---4M3_Conti
KARJAT-5---5M3_Conti
KARJAT-7---6M3_Conti
RATNAGIRI-1---8M3_Cc
RATNAGIRI-2---9M3_Cc
0. RATNAGIRI-3---10M3_
1. RATNAGIRI-4---11M3_
2. RATNAGIRI-24---12M3
3. RATNAGIRI-7---15M3_
4. RATNAGIRI-8---16M3_
5. PANVEL-1---17M3_Co
6. PANVEL-2---18M3_Co
7. PALGHAR-2---20M3_C

PLATE D: Mega alignment of*trnH - psbA* sequences showing variable regions

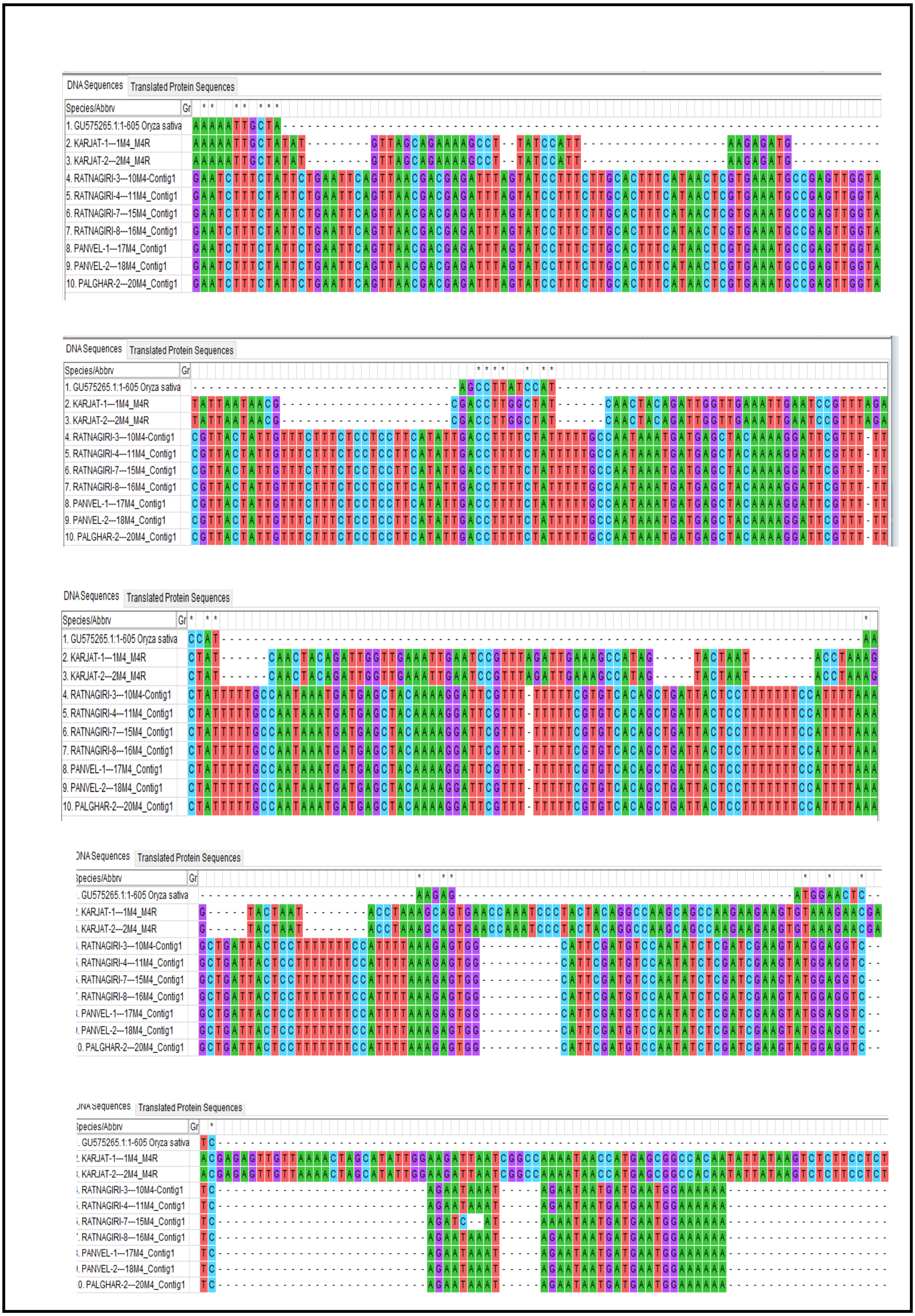
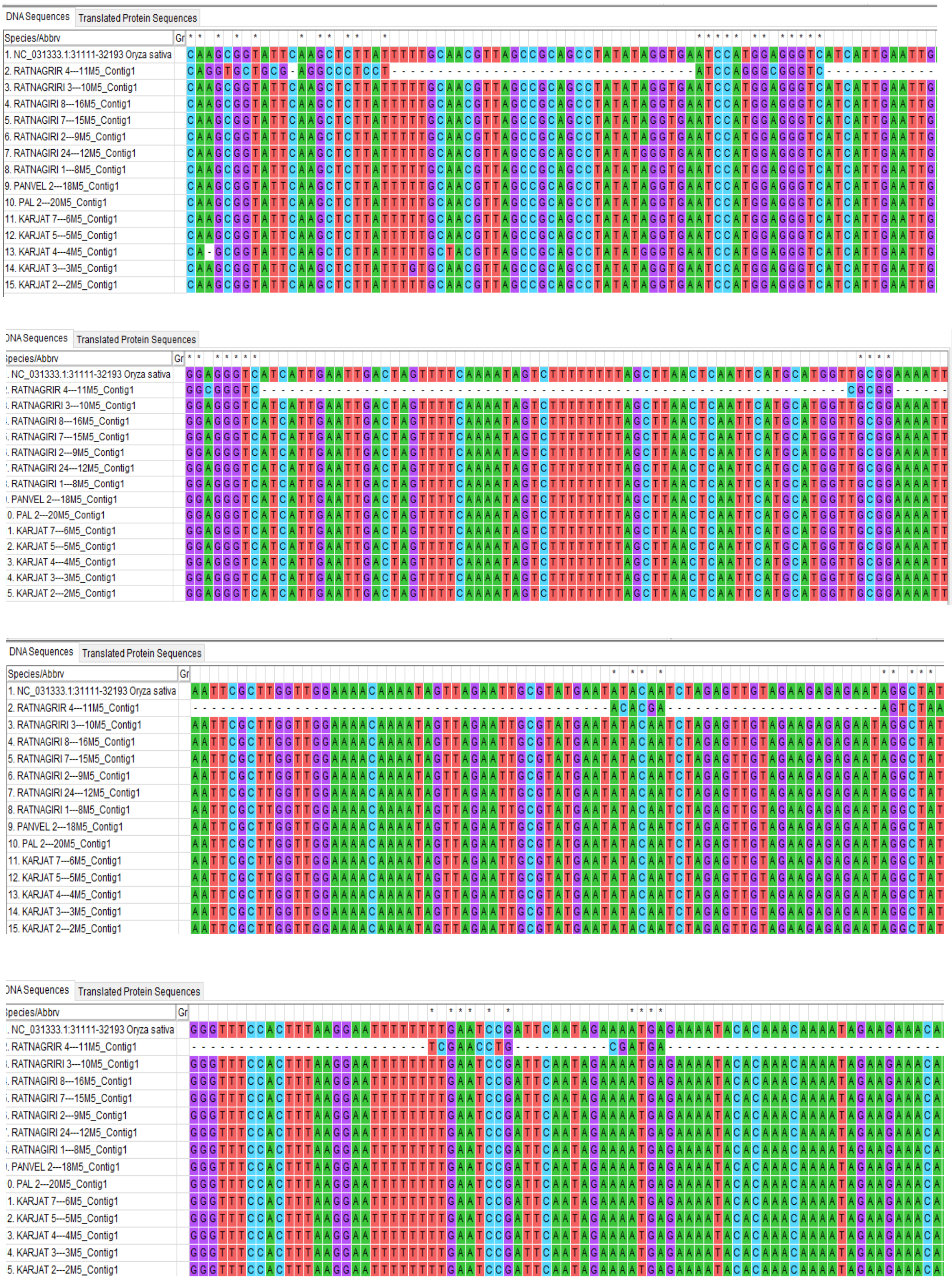

DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. GU575265.1:1-605 Oryza sativa
2. KARJAT-1---1M4_M4R
3. KARJAT-2---2M4_M4R
4. RATNAGIRI-3---10M4-Contig1
5. RATNAGIRI-4---11M4_Contig1
6. RATNAGIRI-7---15M4_Contig1
7. RATNAGIRI-8---16M4_Contig1
8. PANVEL-1---17M4_Contig1
9. PANVEL-2---18M4_Contig1
10. PALGHAR-2---20M4_Contig1

PLATE E: Mega alignment of *atpH-atpI*sequences showing variable regions




DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. NC_031333.1:31111-32193 Oryza sativa
2. RATNAGRIR 4---11M5_Contig1
3. RATNAGRIRI 3---10M5_Contig1
4. RATNAGIRI 8---16M5_Contig1
5. RATNAGIRI 7---15M5_Contig1
6. RATNAGIRI 2---9M5_Contig1
7. RATNAGIRI 24---12M5_Contig1
8. RATNAGIRI 1---8M5_Contig1
9. PANVEL 2---18M5_Contig1
10. PAL 2---20M5_Contig1
11. KARJAT 7---6M5_Contig1
12. KARJAT 5---5M5_Contig1
13. KARJAT 4---4M5_Contig1
14. KARJAT 3---3M5_Contig1
15. KARJAT 2---2M5_Contig1

PLATE F: Mega alignment of *ITS 2* sequences showing variable regions

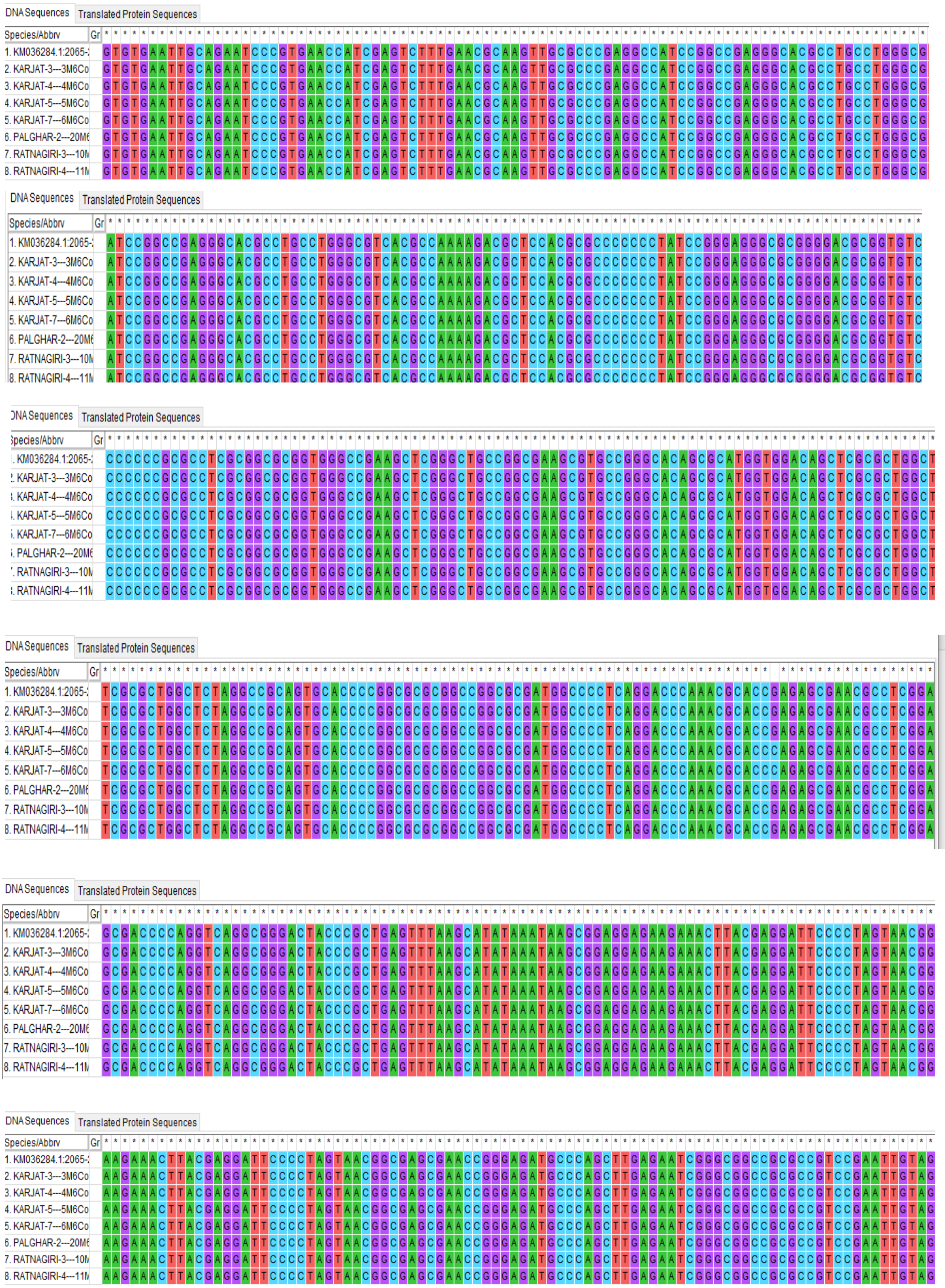
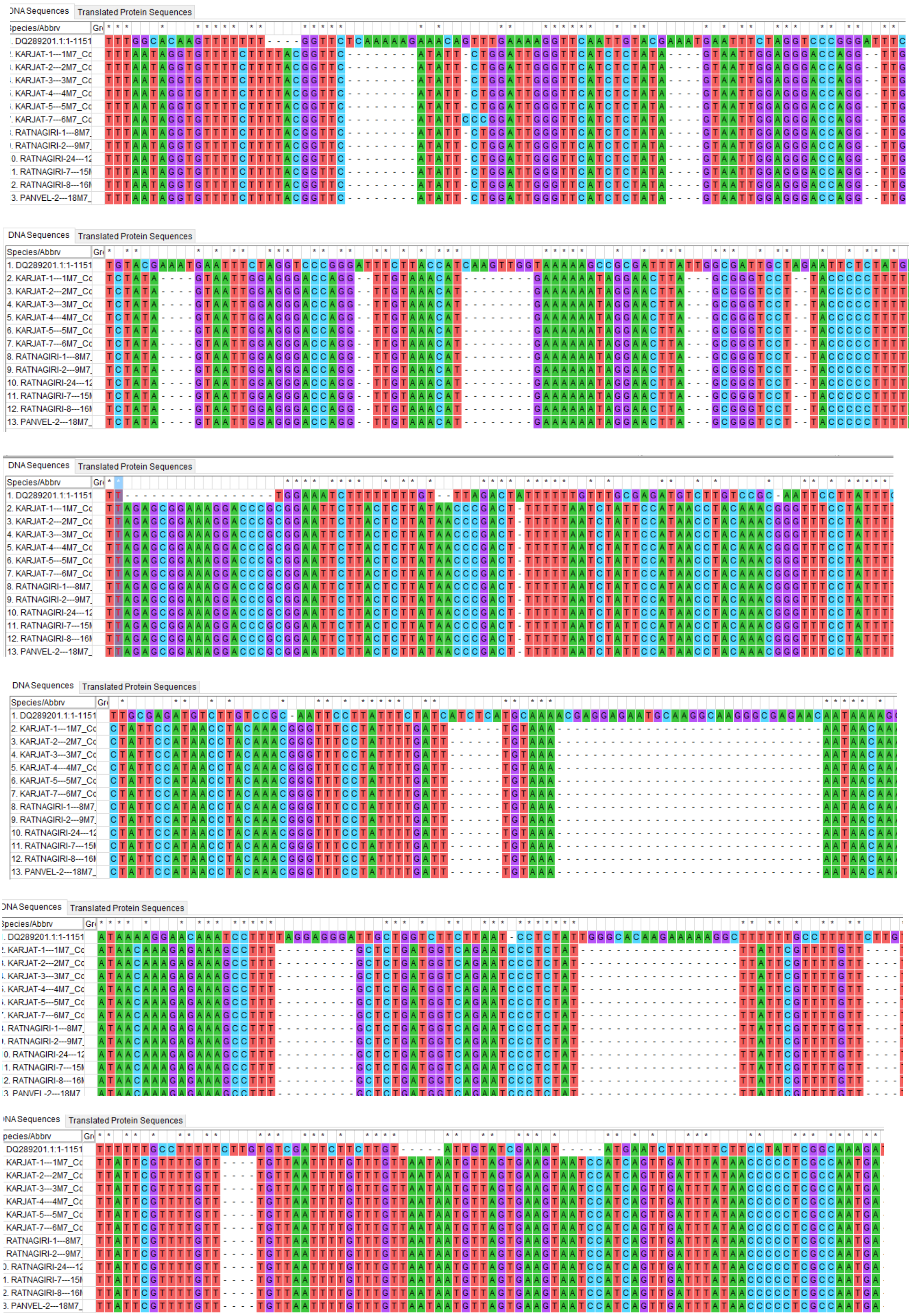
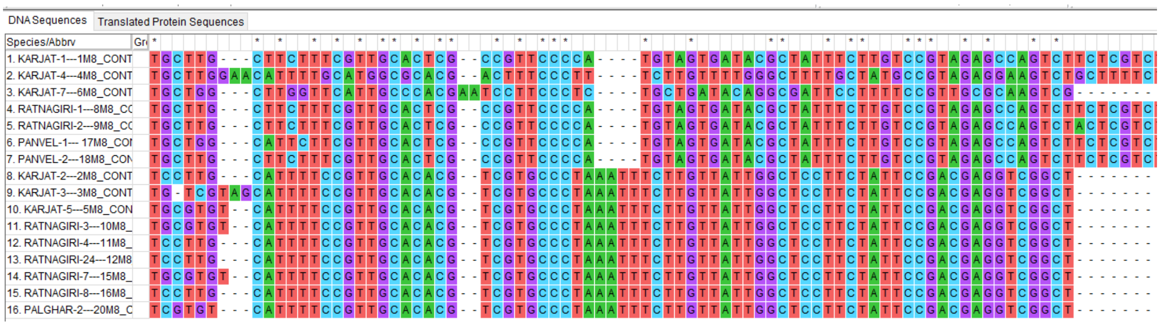


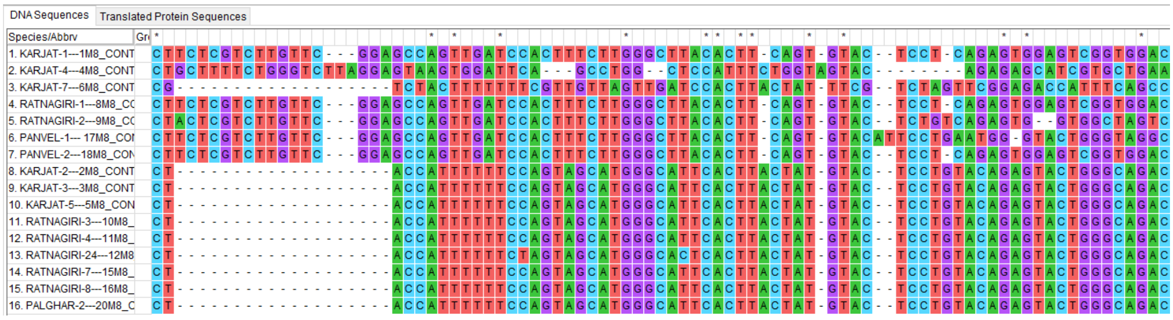
PLATE G: Mega alignment of *petA-psbJ* sequences showing variable regions

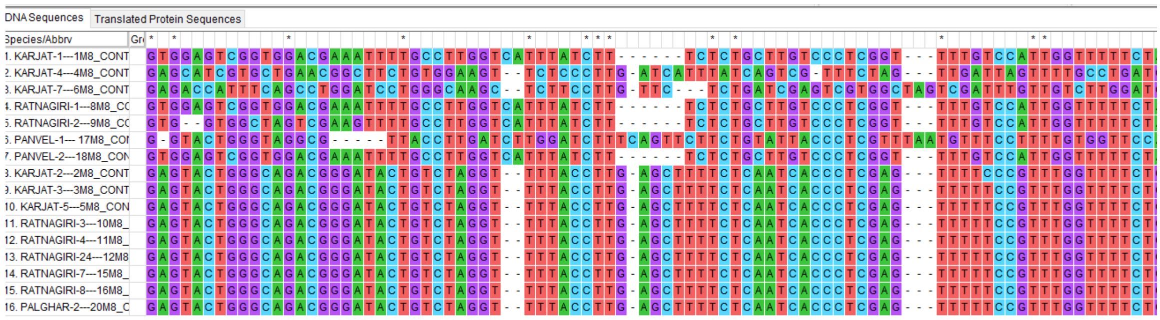


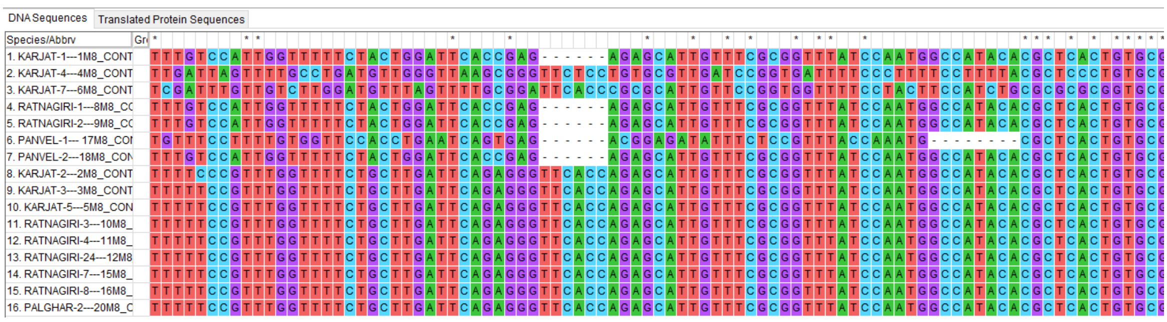

DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. DQ289201.1:1-1151
2. KARJAT-1---1M7_Cc
3. KARJAT-2---2M7_Cc
4. KARJAT-3---3M7_Cc
5. KARJAT-4---4M7_Cc
6. KARJAT-5---5M7_Cc
7. KARJAT-7---6M7_Cc
8. RATNAGIRI-1---8M7_
9. RATNAGIRI-2---9M7_
10. RATNAGIRI-24---12
11. RATNAGIRI-7---15I
12. RATNAGIRI-8---16I
13. PANVEL-2---18M7_

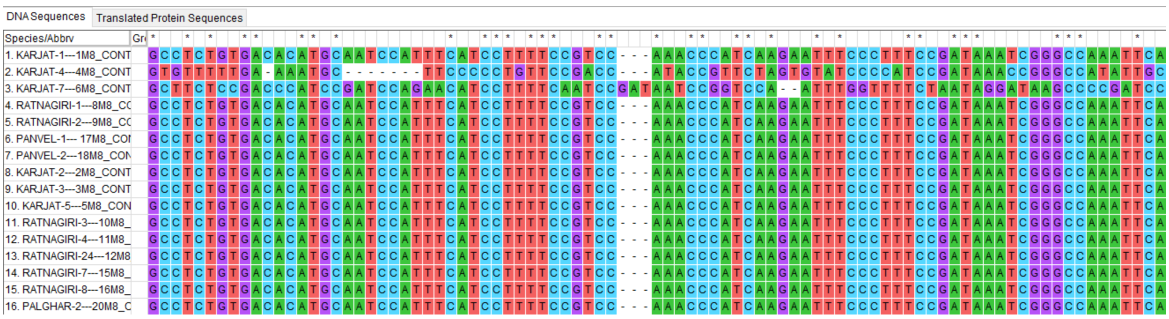
PLATE H: Mega alignment of *trnK* sequences showing variable regions

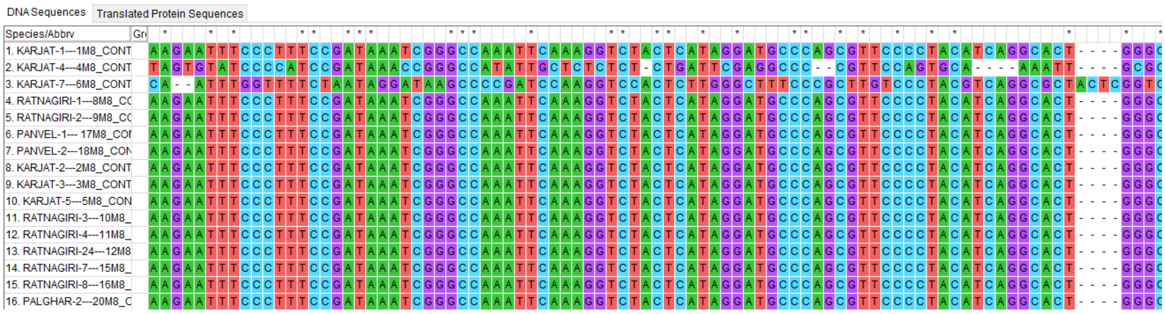


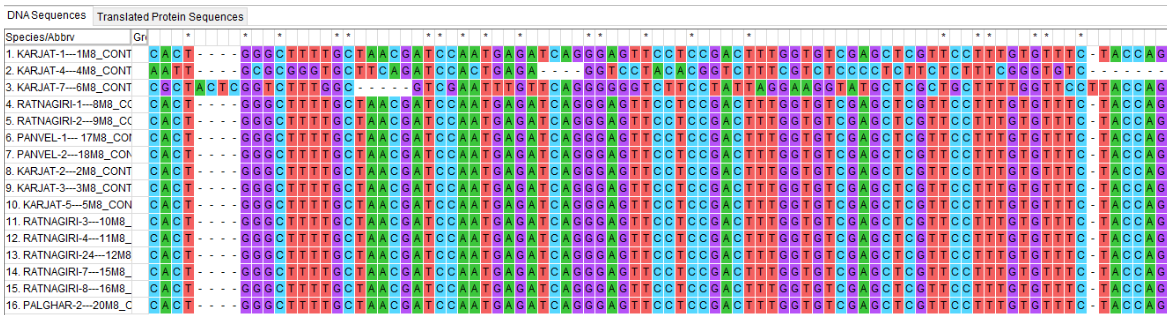


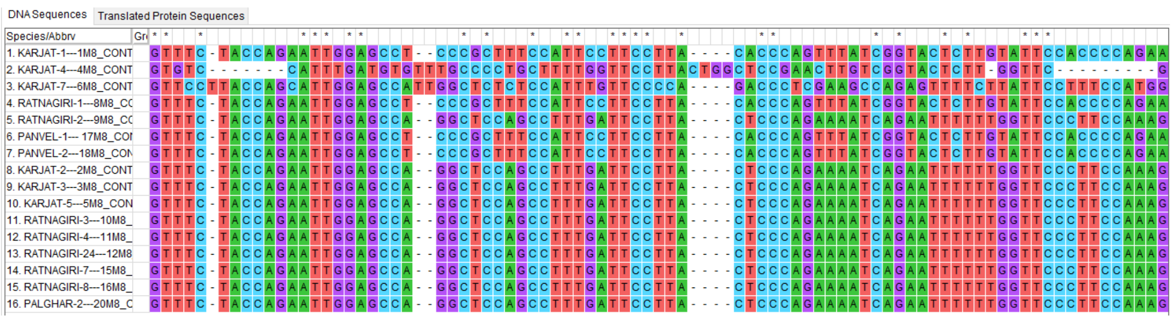







DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. KARJAT-1---1M8_CONT
2. KARJAT-4---4M8_CONT
3. KARJAT-7---6M8_CONT
4. RATNAGIRI-1---8M8_CC
5. RATNAGIRI-2---9M8_CC
6. PANVEL-1--- 17M8_COI
7. PANVEL-2---18M8_CON
8. KARJAT-2---2M8_CONT
9. KARJAT-3---3M8_CONT
10. KARJAT-5---5M8_CON
11. RATNAGIRI-3---10M8_
12. RATNAGIRI-4---11M8_
13. RATNAGIRI-24---12M8
14. RATNAGIRI-7---15M8_
15. RATNAGIRI-8---16M8_
16. PALGHAR-2---20M8_C


DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. KARJAT-1---1M8_CONT
2. KARJAT-4---4M8_CONT
3. KARJAT-7---6M8_CONT
4. RATNAGIRI-1---8M8_CC
5. RATNAGIRI-2---9M8_CC
6. PANVEL-1--- 17M8_COI
7. PANVEL-2---18M8_CON
8. KARJAT-2---2M8_CONT
9. KARJAT-3---3M8_CONT
10. KARJAT-5---5M8_CON
11. RATNAGIRI-3---10M8_
12. RATNAGIRI-4---11M8_
13. RATNAGIRI-24---12M8
14. RATNAGIRI-7---15M8_
15. RATNAGIRI-8---16M8_
16. PALGHAR-2---20M8_C


DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. KARJAT-1---1M8_CONT
2. KARJAT-4---4M8_CONT
3. KARJAT-7---6M8_CONT
4. RATNAGIRI-1---8M8_CC
5. RATNAGIRI-2---9M8_CC
6. PANVEL-1--- 17M8_COI
7. PANVEL-2---18M8_CON
8. KARJAT-2---2M8_CONT
9. KARJAT-3---3M8_CONT
10. KARJAT-5---5M8_CON
11. RATNAGIRI-3---10M8_
12. RATNAGIRI-4---11M8_
13. RATNAGIRI-24---12M8
14. RATNAGIRI-7---15M8_
15. RATNAGIRI-8---16M8_
16. PALGHAR-2---20M8_C

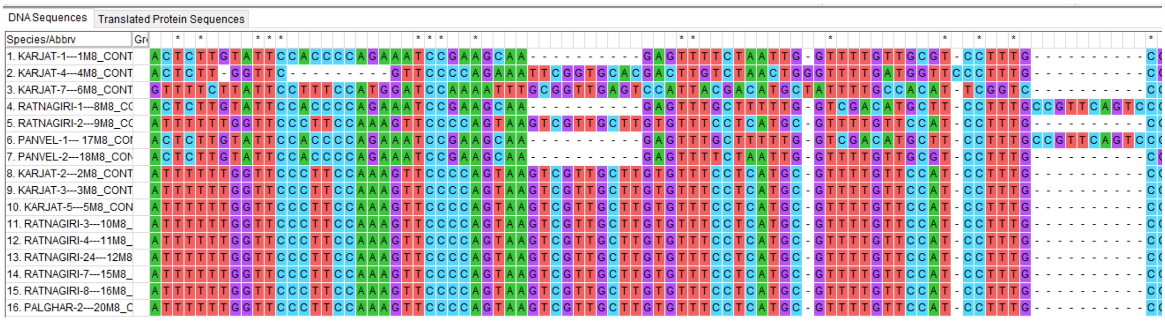
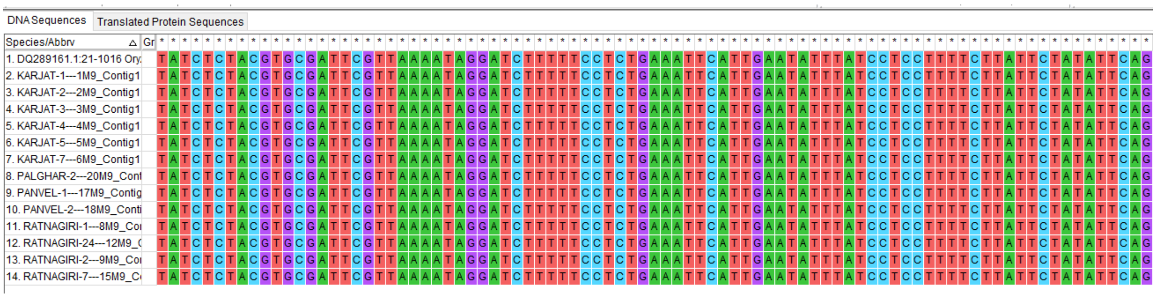
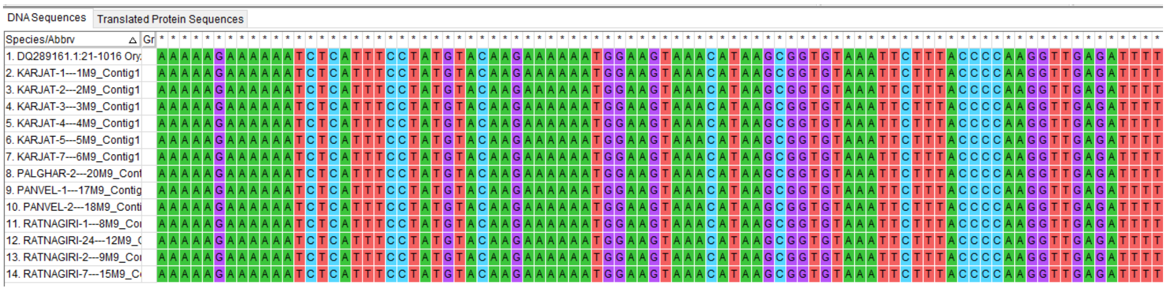
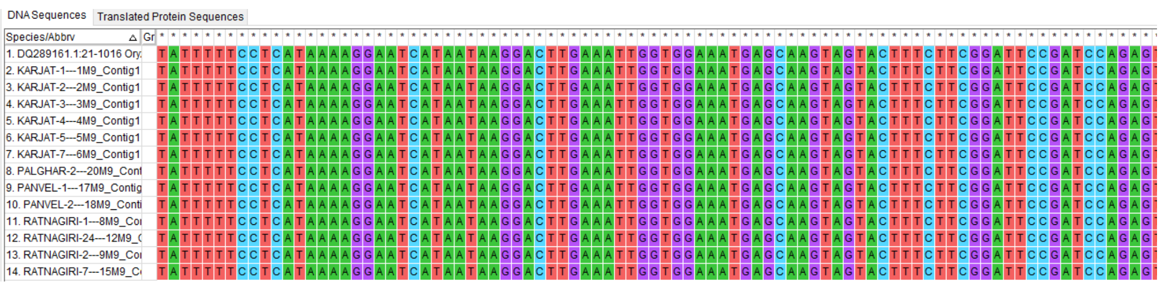

DNA Sequences
Translated Protein Sequences
Species/Abbrv
1. KARJAT-1---1M8_CONT
2. KARJAT-4---4M8_CONT
3. KARJAT-7---6M8_CONT
4. RATNAGIRI-1---8M8_CC
5. RATNAGIRI-2---9M8_CC
6. PANVEL-1--- 17M8_COI
7. PANVEL-2---18M8_CON
8. KARJAT-2---2M8_CONT
9. KARJAT-3---3M8_CONT
10. KARJAT-5---5M8_CON
11. RATNAGIRI-3---10M8_
12. RATNAGIRI-4---11M8_
13. RATNAGIRI-24---12M8
14. RATNAGIRI-7---15M8_
15. RATNAGIRI-8---16M8_
16. PALGHAR-2---20M8_C

PLATE I: Mega alignment of *ndhA* sequences showing variable regions







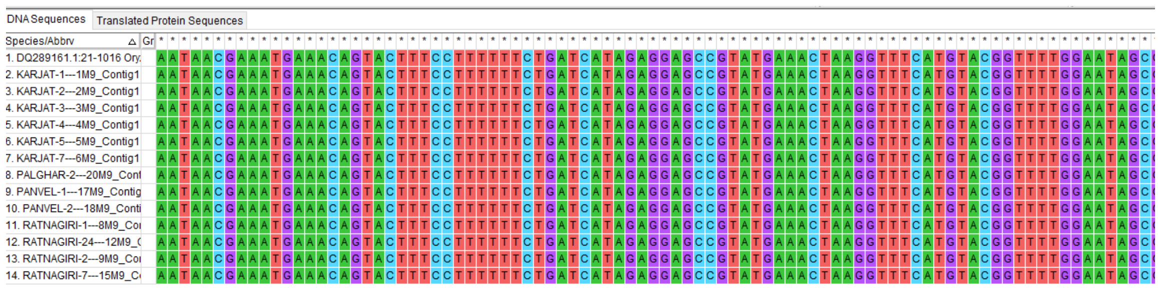


PLATE J: Mega alignment of *matK-1RKIM* & *matK-3FKIM*sequences showing variable regions

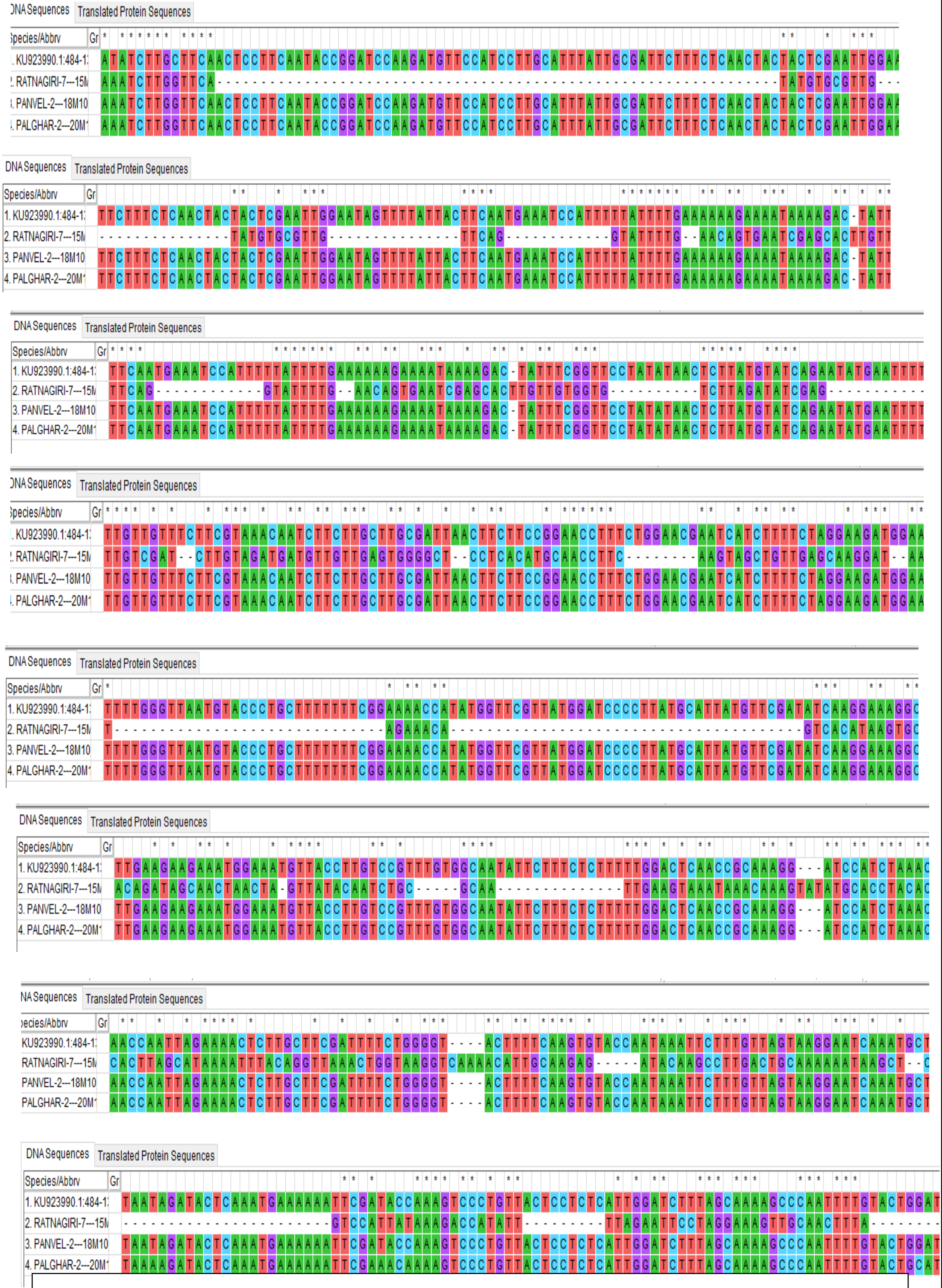
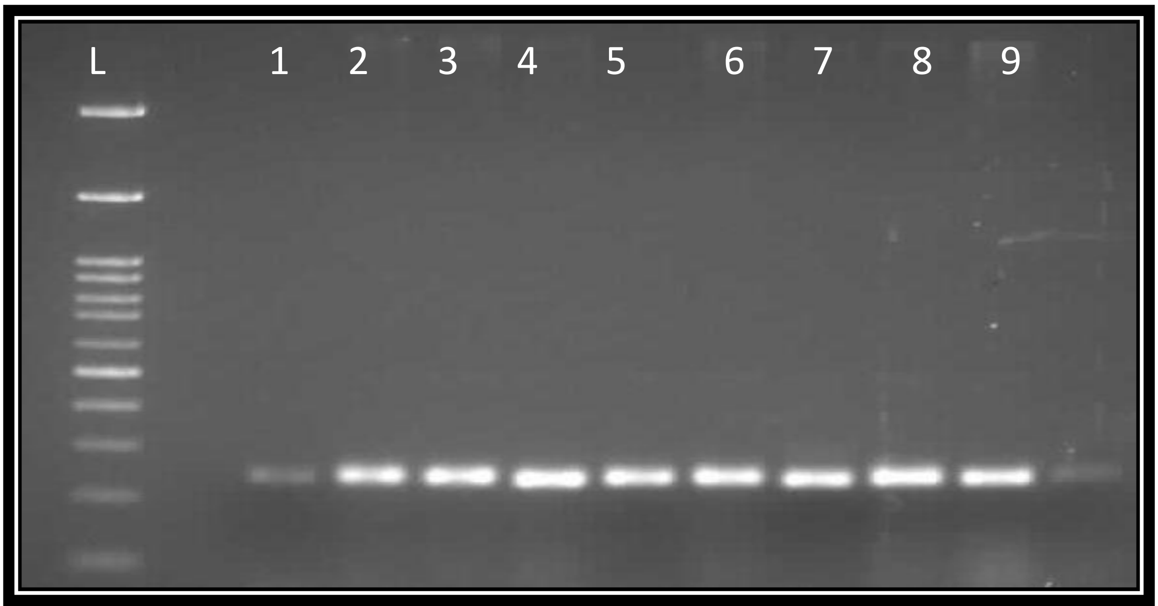
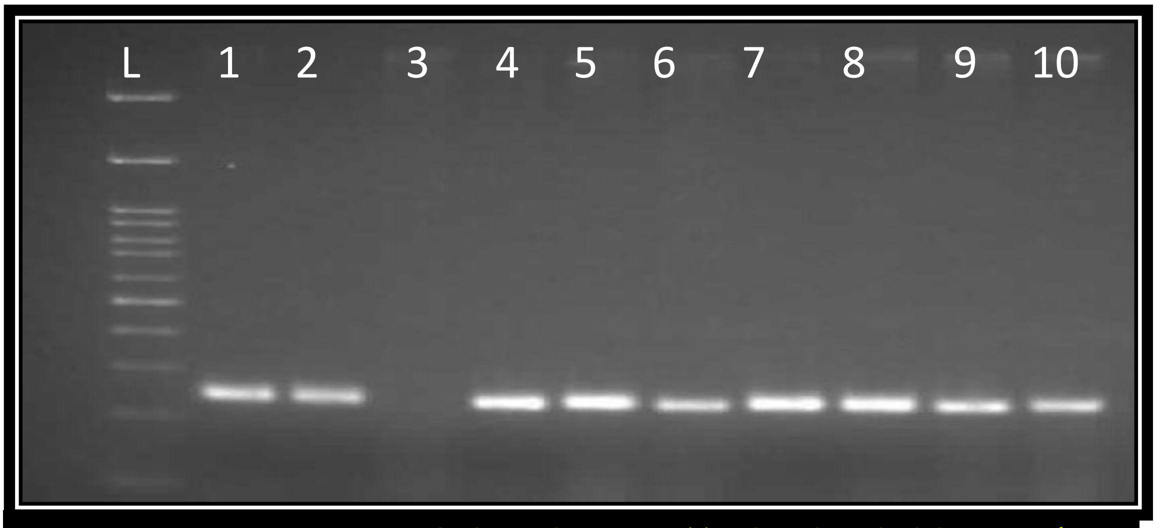


Plate 22: Gel photograph of SSR primer RM-181

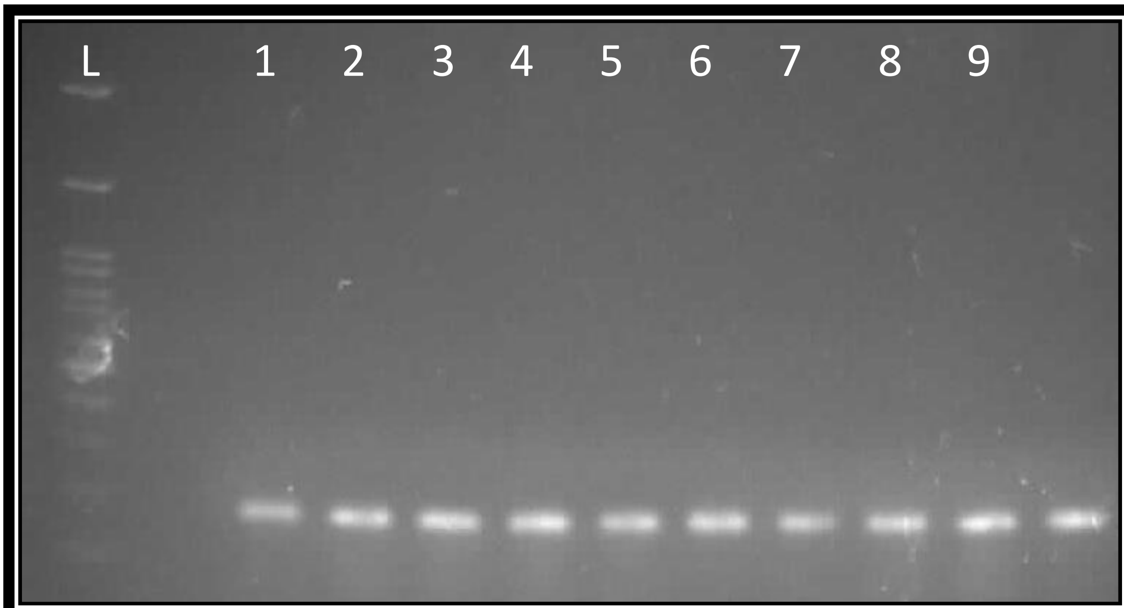


AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker
RM - 181
LADDER(L):100 bp
SAMPLES AMPLIFIED :Karjat-1, Karjat-2,Karjat-3,Karjat-4, Karjat-5,Karjat-7,Karjat-184,Panvel-1,Panvel-2,Panvel-3



AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker
RM -181
LADDER (L):100 bp
SAMPLES AMPLIFIED : Ratnagrir-1,Ratnagiri-2,Ratnagiri-4,Ratnagiri-24,Ratnagiri-73,Ratnagiri-6,Ratnagiri-7,Ratnagiri-8,Palghar-2

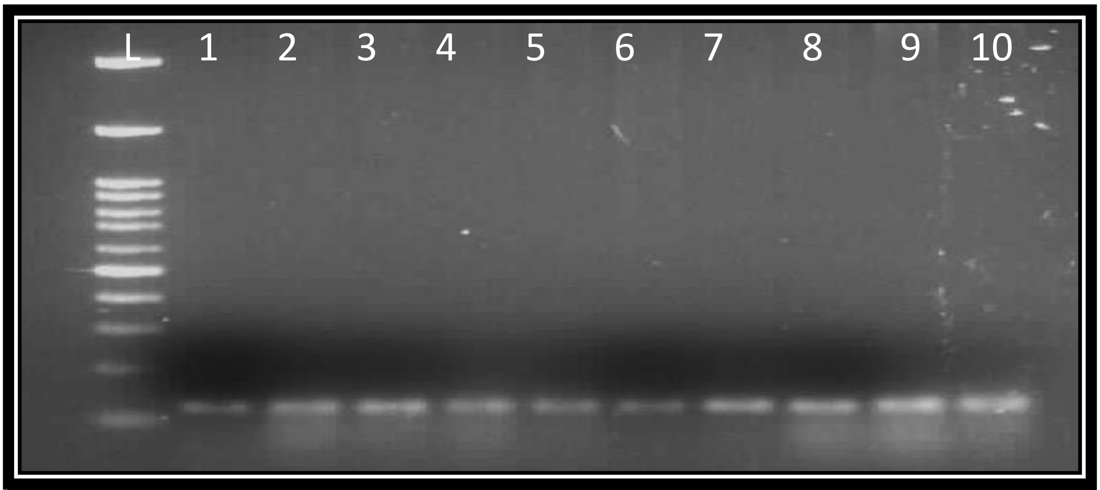
Plate 21: Gel photograph of SSR primer RM-338



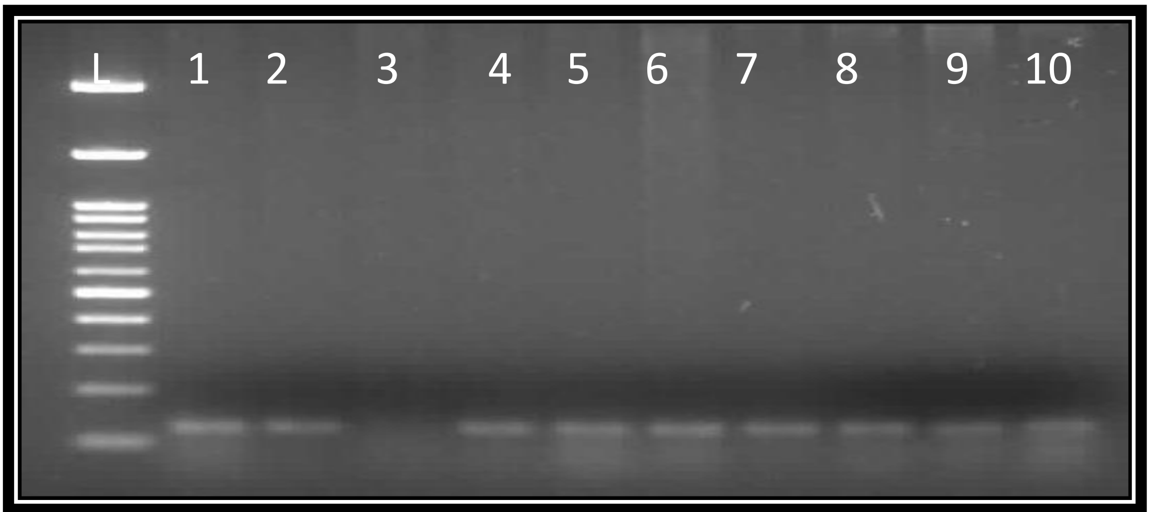
AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker RM -338
LADDER(L):100 bp
SAMPLES AMPLIFIED :Karjat-1, Karjat-2,Karjat-3,Karjat-4, Karjat-5,Karjat-7,Karjat-184,Panvel-1,Panvel-2,Panvel-3

AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker RM -338
LADDER(L):100 bp
SAMPLES AMPLIFIED : Ratnagir-1,Ratnagiri-2,Ratnagiri-3,Ratnagiri-4,Ratnagiri-24,Ratnagiri-73,Ratnagiri-6,Ratnagiri-7,Ratnagiri-8,Palghar-2

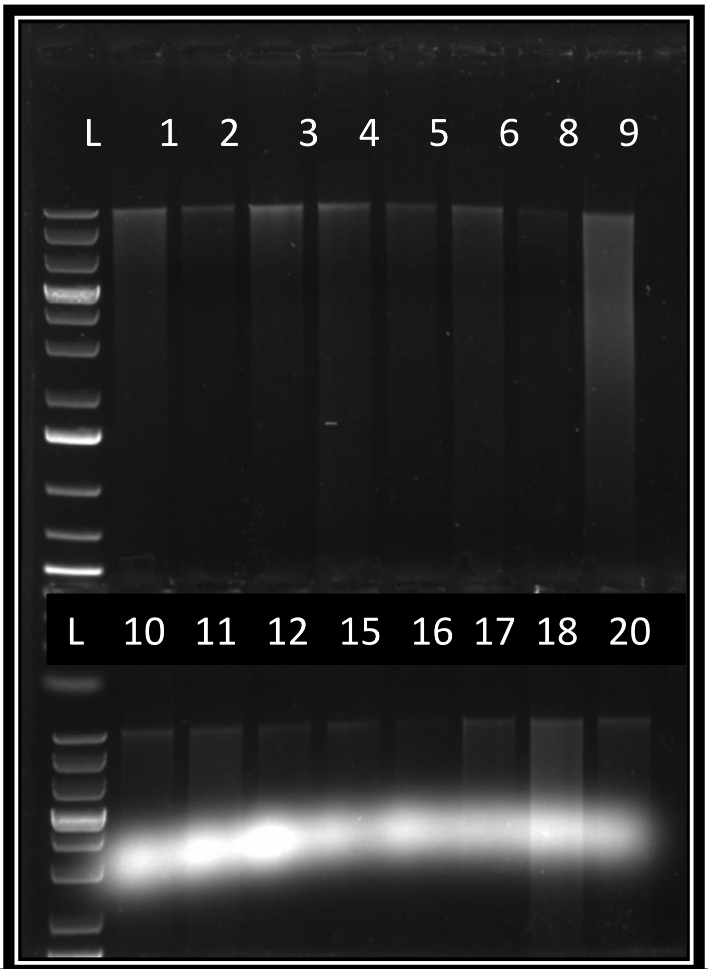
Plate 20: Gel photograph of SSR primer RM-112



AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker RM -112
LADDER(L):100 bp
SAMPLES AMPLIFIED :Karjat-1, Karjat-2,Karjat-3,Karjat-4, Karjat-5,Karjat-7,Karjat-184,Panvel-1,Panvel-2,Panvel-3

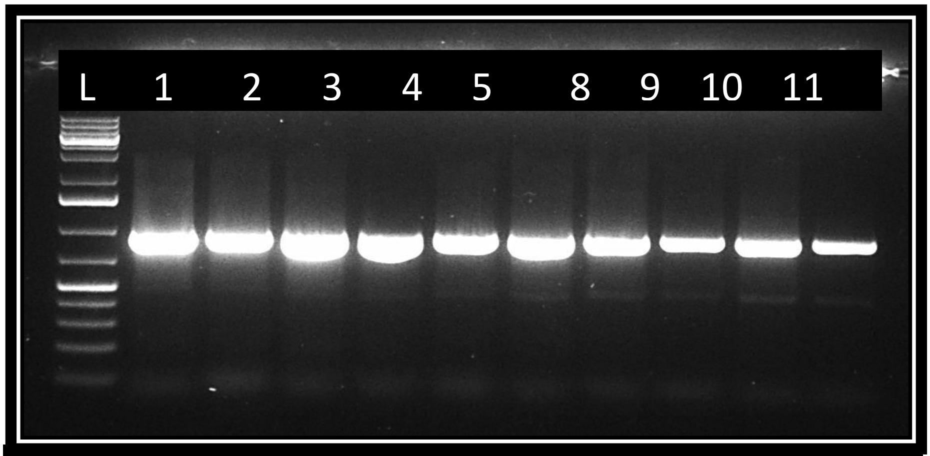


AMPLIFIED DNA BANDS OF 10 VARIETIES USING SSR marker RM -112
LADDER(L):100 bp
SAMPLES AMPLIFIED :: Ratnagrir-1,Ratnagiri-2,Ratnagiri-4,Ratnagiri-24,Ratnagiri-73,Ratnagiri-6, Ratnagiri-7, Ratnagiri-8,Palghar-2



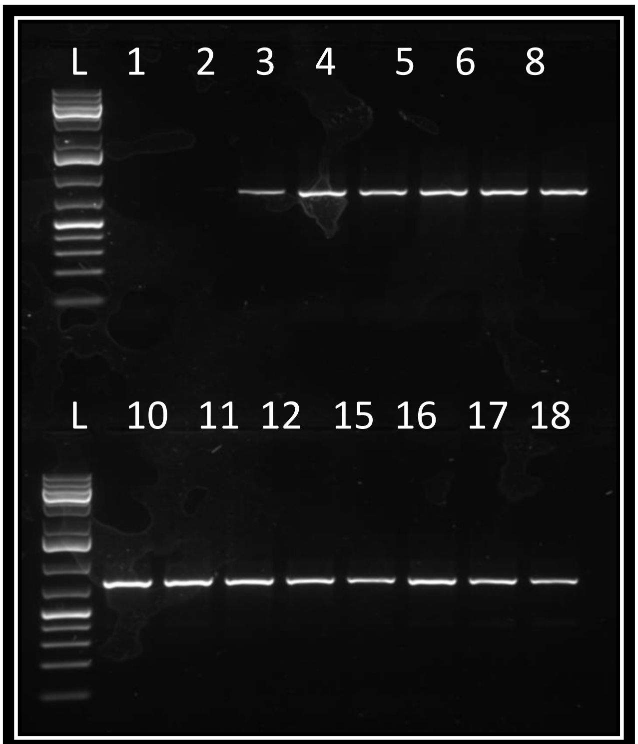
DNA images : L: ladder (1kb plus DNA ladder)
Samples: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2.

Plate2: Gel photograph of DNA isolated from 20 rice varieties



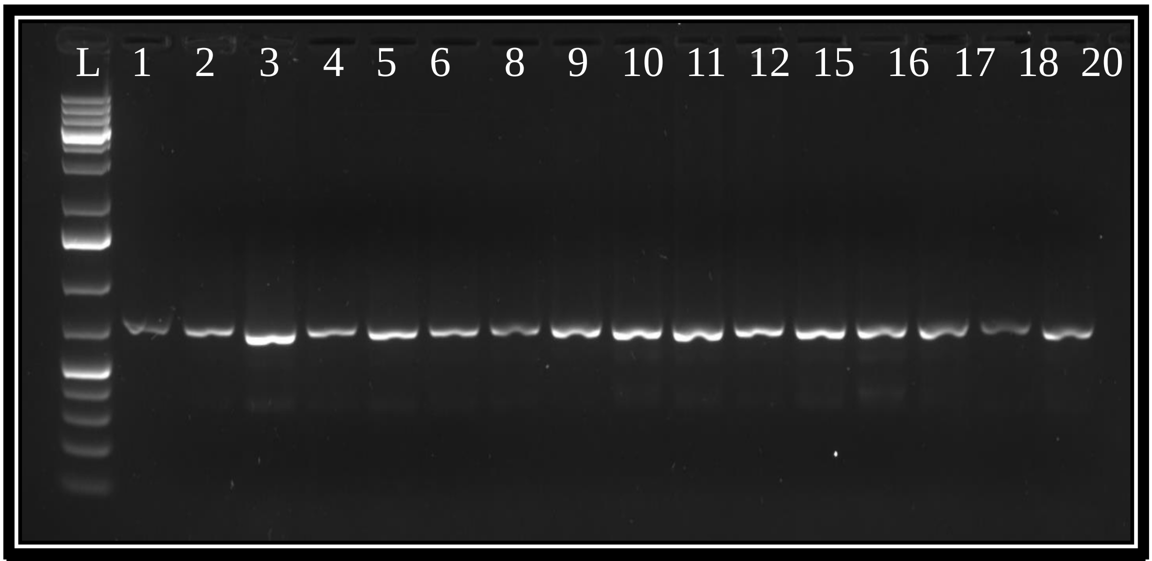
PCRAmplification By *matK*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Panvel-1

Plate 3.1: Gel photograph of barcoding primer *matK*



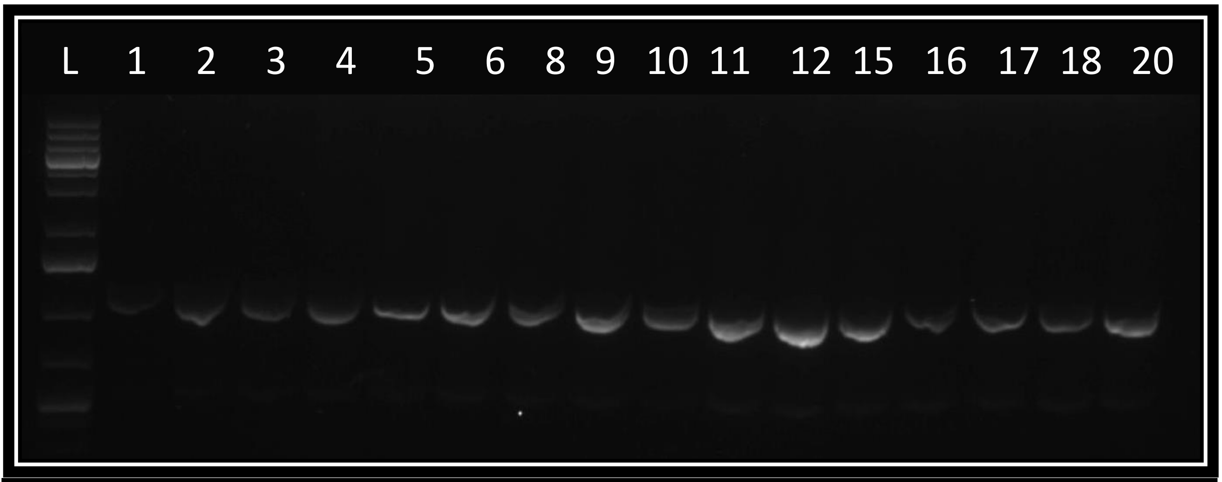
PCR Amplification by *rbcL*
Ladder(L): 1kb plus DNA ladder
Samples Amplified: Karjat-3, Karjat-4, Karjat-5, Karjat-7ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2.

Plate 3.2.Gel photograph of barcoding primer *rbcL*



PCR Amplification By *psbA - trnH*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2.

Plate 3.3.Gel photograph of barcoding primer *psbA-trnH*



PCR Amplification By *trnH-psbA*
Ladder(L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2.

Plate 3.4:Gel photograph of barcoding primer *trnH - psbA*

L 2 3 4 5 6 8 9 10 11 12 15 16 18

PCRAmplification By*atpH- atpI*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8,Panvel-2,

Plate 3.5.Gel photograph of barcoding primer*atpH-atpI*

L 1 2 3 4 5 6 8 9 10 11 12 15 16 17 18 20

PCR Amplification By*ITS2*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Panvel-1, Panvel-2, Palghar-2.

Plate 3.6.Gel photograph of barcoding primer*ITS 2*

L 1 2 3 4 5 6 8 9 11 12 15 16 18

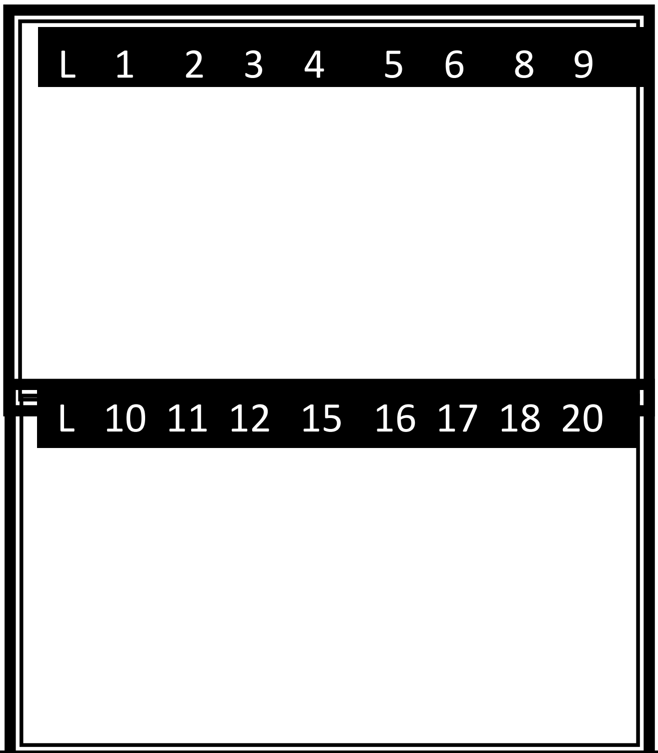
PCR Amplification By *petA- psbJ*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8,Panvel-2.

Plate 3.7.Gel photograph of barcoding primerpetA-psbJ

L 1 2 3 4 5 6 8 9 10 11 12 15 16 17 18 20

PCR Amplification By *trnK*
Ladder(L): 1kb plus DNA ladder
Samples Apmlified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5, Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Ratnagiri-8, Panvel-1, Panvel-2, Palghar-2.

Plate 3.8.Gel photograph of barcoding primer*trnK*



PCR Amplification By *ndhA*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-5,Karjat-7, Ratnagiri-1, Ratnagiri-2, Ratnagiri-3, Ratnagiri-4, Ratnagiri-24, Ratnagiri-7Panvel-1, Panvel-2, Palghar-2.

Plate 3.9.Gel photograph of barcoding primer *ndhA*

L 1 2 3 4 5 6 8 9

L 10 11 12 15 16 17 18 20

PCR Amplification by *matK1RIM-3RIM*
Ladder (L): 1kb plus DNA ladder
Samples Amplified: Karjat-1, Karjat-2, Karjat-3, Karjat-4, Karjat-7, Ratnagiri-1, Ratnagiri-2,Ratnagiri-4, Ratnagiri-24, Ratnagiri-7, Panvel-1, Panvel-2, Palghar-2.

Plate 3.10: Gel photograph of barcoding primer *matK-1RKIM & matK-3FKIM*

Fig.3.1: Rice (*Oryzasativa* L.) Chloroplast DNA

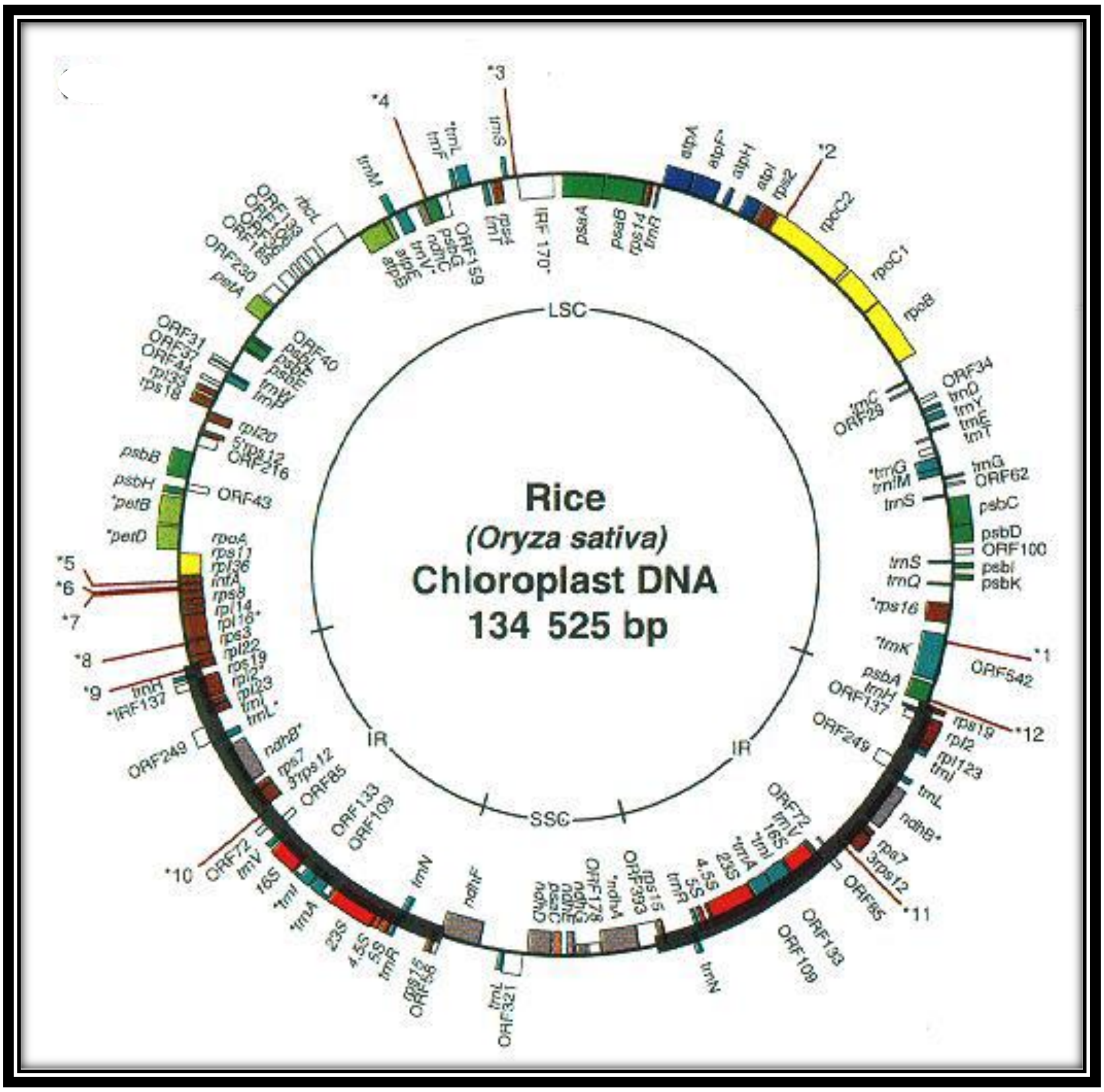


Table no 4.1. Morphological Characters of 20 rice varieties

Characters / Varieties	Year of release	Duration	Days for growth	Yield (q/ha.)	Grain type	Resistant/Tolerant to	Stature	Salient features
Karjat-1	1985	Early	110-115	35-40	Short bold	Resistant to bacterial blight ,blast and leaf scald	dwarf	Recommended for rainfed uplands
Karjat-2	1993	late	140-145	40-45	Long slender	Reistant ti blast & neck blast	Dwarf	-
Karjat-3	1994	Early	115-120 days	45-50	short bold grain	resistant to blast	Dwarf	suitable for rainfed uplands as well as irrigated areas for kharif & rabi season
Karjat-4	1996	Early	110-115 days	30-35	short slender superfine grain	resistant to leaf folder	Dwarf	suitable for rainfed upland as well as irrigated areas for kharif & rabi season
Karjat-5	2005	Midlate	125-130 days	50-55	, long bold	resistant to neck blast	Semi dwarf	suitable for midland under rainfed & irrigated conditions
Karjat-7	2007	Early	115-120 days	45-50	long slender	resistant to leaf folder, moderately to brown plant hopper, white brown plant hopper and blast and bacterial leaf blast	Dwarf	-
Karjat-184	2009	Early	100-105 days	30-35	medium slender	moderately resistant to blast and bacterial leaf blast tolerant to leaf folder and stem borer	Dwarf	recommended for rainfed uplands

Ratnagiri-1	1986	Early	115-120 days	35-40	Long bold grain	moderately resistant to blast and neck blast	-	internationally released in Zambia for irrigated ecology
Ratnagiri-2	1986	Late	145-150 days	35-40	Short bold	resistant to blast & neck blast	-	recommended for low land rice growing areas having assured rainfall in Maharashtra.
Ratnagiri- 3	1993	Late	140-145 days	40-45	Long bold grain	resistant to gall midge	-	suitable for shallow low land area

Characters Varieties	Year of release	Duration	Days for growth	Yield (q/ha.)	Grain type	Resistant/Tolerant to	Stature	Salient features
Ratnagiri-4	2009	Midlate	125-130 days	49	Long slender	Moderately resistant to neck blast, leaf blast & bacterial leaf blight	-	-
Ratnagiri-24	2009	Early	110-115 days	35-40	long slender	Tolerant to stem borer resistant to blast and moderately, susceptible to bacterial leaf blight	Semi dwarf	Suitable for both kharif and hot weather
Ratnagiri-73	1979	Very Early	90-100 days	35-40	short and bold	Moderately resistant to leaf blast, neck blast and bacterial leaf blight	Dwarf	-
Ratnagiri-6	2017	-	118-125 days	45-50	Medium slender	Moderately resistant to leaf blast and Macterial leaf blight diseases	-	Recommended for Uplands, midlands, Konkan region, MS
Ratnagiri-7	2017	-	122-125 days	40-45	Short bold	Moderately resistant to BLB and leaf blast	-	Red kernel rice with high zinc (30 ppm) and iron content (15.4 ppm) with low glycemic index (53) Recommended for Uplands, transplanted rice growing areas of Konkan region of MS
Ratnagiri-8	2018	-	135-138 days	55-58	Medium slender	Moderately resistant reactions to Leaf Blast, Bactrial Leaf Blight, Neck Blast, Rice Tungro diseases, Sheath root and Glum discoloration	-	BSKKV's first rice variety released through CVRC, Excellent cooking quality, heavy panicles, wider adaptability, high fertility(%)

Panvel-1	1984	Midlate	125-130 days	35-40	short bold	resistant to blast & moderately resistant to stem borer tolerant to salinity condition	Semi dwarf	recommended for coastal saline soils of Konkan region
Panvel-2	1987	Early	(110-115 days	40-45	long slender	resistant to blast & moderately, resistant to stem borer	Semi dwarf	recommended for coastal saline soils of Konkan region, tolerant to salinity condition
Panvel-3	1999	Early	110-115 days	40-45	short bold	, moderately resistant to blast highly salt tolerant	Mid tall	suitable for medium to high rainfall area in coastal saline soil condition
Palghar-2	2002	Midlate	125-130 days	30-35	Short slender	moderately resistant to stem borer and blast	Semi dwarf	-