A study on genetic diversity and marker trait association analysis in turmeric (*Curcuma longa* L.)

Thesis submitted to
University of Calicut

For the award of degree of
Doctor of Philosophy
(Botany)

By

ASWATHI A P

Under the guidance of
Dr. D. Prasath

ICAR-INDIAN INSTITUTE OF SPICES RESEARCH
Kozhikode-673 012, Kerala, India

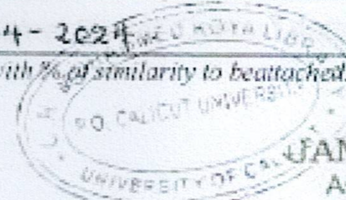
May 2024

UNIVERSITY OF CALICUT

CERTIFICATE ON PLAGIARISM CHECK

1.	Name of the Research Scholar	**ASWATHI A P**	
2.	Title of thesis / dissertation	**A study on genetic diversity and marker-trait association analysis in turmeric (*Curcuma longa* L.)**	
3.	Name of the Supervisor	**Dr. D. PRASATH**	
4.	Department/Institution	Division of crop improvement and Biotechnology, ICAR-Indian Institute of Spices Research, Kozhikode	
5.		Non Core	Core
		Introduction/ Theoretical overview/Review of literature/ Materials & Methods/ Methodology	**Analysis/Result/Discussion/ Summary/Conclusion/ Recommendations**
	Similar content (%) identified	1	0
	Acceptable maximum limit (%)	10	10
6.	Software used	iThenticate	
7.	Date of verification	18-04-2024	

**Report on plagiarism check, specifying included/excluded items with % of similarity to beattached.*



Checked by (with name, designation & signature)

JAMSHEER N. P.
Assistant Librarian
University of Calicut
Malappuram - 673 635

Name and signature of the Researcher: ASWATHI A P

Name and signature of the Supervisor: D. PRASATH

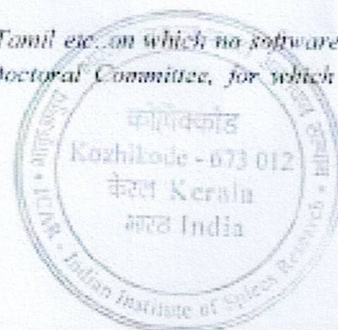
Dr. D. PRASATH, Ph.D.
Principal Scientist (Hort.)
Indian Institute of Spices Research
Kozhikode, Kerala 673012

The Doctoral Committee* has verified the report on plagiarism check with the contents of the thesis, as summarized above and appropriate measures have been taken to ensure originality of the Research accomplished herein.

Name & Signature of the HoD/HoI (Chairperson of the Doctoral Committee)

डा. आर. दिनेश Dr. R. Dines
निदेशक Director
ICAR - Indian Institute of Spices Rese

**In case of languages like Malayalam, Tamil etc. on which no software is available for plagiarism check, a manual check shall be made by the Doctoral Committee, for which an additional certificate has to be attached.*



iThenticate

Similarity Report ID: oid:3618:57848150

PAPER NAME

A study on genetic diversity and marker-trait association analysis in turmeric (Curcuma longa L.)

AUTHOR

Aswathi A P \ Non Core

WORD COUNT

17338 Words

CHARACTER COUNT

98834 Characters

PAGE COUNT

67 Pages

FILE SIZE

5.6MB

SUBMISSION DATE

Apr 17, 2024 3:18 PM GMT+5:30

REPORT DATE

Apr 17, 2024 3:20 PM GMT+5:30

1% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 1% Internet database
- 0% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

Excluded from Similarity Report

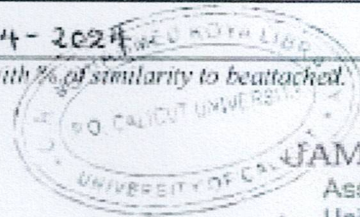
- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 12 words)
- Manually excluded sources
- Manually excluded text blocks

UNIVERSITY OF CALICUT

CERTIFICATE ON PLAGIARISM CHECK

1.	Name of the Research Scholar	**ASWATHI A P**	
2.	Title of thesis / dissertation	**A study on genetic diversity and marker-trait association analysis in turmeric (*Curcuma longa* L.)**	
3.	Name of the Supervisor	**Dr. D. PRASATH**	
4.	Department/Institution	Division of crop improvement and Biotechnology, ICAR-Indian Institute of Spices Research, Kozhikode	
5.		Non Core	Core
		Introduction/ Theoretical overview/Review of literature/ Materials & Methods/ Methodology	**Analysis/Result/Discussion/ Summary/Conclusion/ Recommendations**
	Similar content (%) identified	1	0
	Acceptable maximum limit (%)	10	10
6.	Software used	iThenticate	
7.	Date of verification	18-04-2024	

*Report on plagiarism check, specifying included/excluded items with % of similarity to beattached.



Checked by (with name, designation & signature)

JAMSHEER N. P.
Assistant Librarian
University of Calicut
Malappuram - 673 635

Name and signature of the Researcher: ASWATHI A P

Name and signature of the Supervisor: D PRASATH

Dr. D. PRASATH, Ph.D.
Principal Scientist (HoD)
Indian Institute of Spices Research
Kozhikode, Kerala-673012

The Doctoral Committee* has verified the report on plagiarism check with the contents of the thesis, as summarized above and appropriate measures have been taken to ensure originality of the Research accomplished herein.

Name & Signature of the HoD/HoI (Chairperson of the Doctoral Committee)

डॉ. आर. दिनेश Dr. R. Dinesh
निदेशक Director
ICAR - Indian Institute of Spices Research
Marikunnu P.O, Kozhikode-673 012

*In case of languages like Malayalam, Tamil etc. on which no software is available for plagiarism check, a manual check shall be made by the Doctoral Committee, for which an additional certificate has to be attached.

भाकृअनुप - भारतीय मसाला फसल अनुसंधान संस्थान
ICAR - INDIAN INSTITUTE OF SPICES RESEARCH
(भारतीय कृषि अनुसंधान परिषद *Indian Council of Agricultural Research)*
पोस्ट बैग संख्या: *Post Bag No: 1701,* मेरिकुन्नु पोस्ट *Marikunnu Post,*
कोषिक्कोड *Kozhikode-673012,* केरल, *Kerala,* भारत *India*
(ISO 9001: 2015 Certified Institute)

CERTIFICATE

This is to certify that the thesis entitled **'A study on genetic diversity and marker-trait association analysis in turmeric (*Curcuma longa* L.)'** submitted to the University of Calicut by **Ms. Aswathi A P** for the award of degree of **Doctor of Philosophy in Botany** is the result of research work carried out by her in the Division of Crop Improvement and Biotechnology, ICAR-Indian Institute of Spices Research, Kozhikode, Kerala, India during the period 2019 to 2024. This has not been submitted for the award of any other degree or diploma of this or any other university. The plagiarism has been checked at CMHK library, University of Calicut and the values are within the acceptable limits. The corrections and suggestions recommended by the adjudicators have been incorporated in the thesis.

Place: Kozhikode

(D Prasath)

Date: 19.12.2024

Dr.D.PRASATH. Ph.D
Senior Scientist (Hort.)
Indian Institute of Spices Research
Kozhikode, Kerala-673012

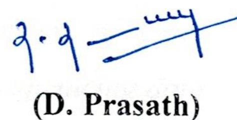
मसालों की महक है निराली, सेवन से होगा देश खुशहाली

PABX: 0495-2731410/2731753/2731345 निदेशक का कार्यालय Director's Office: 0495-2730294 परियोजना समन्वयक Project Coordinator: 0495-2731794, ATIC: 0495 - 2730704. आई आई एस आर प्रायोगिक प्रक्षेत्र, पेरुवण्णामुषि IISR Experimental Farm, Peruvannamuzhi: 0496 2249371 कृषि विज्ञान केन्द्र Krishi Vigyan Kendra, पेरुवण्णामुषि Peruvannamuzhi: 0496-2666041, फैक्स Fax: 0091-495-2731187 ई-मेल Email: director.spices@icar.gov.in वेबसाइट Website: www.spices.res.in

DECLARATION

I hereby declare that the work presented in this thesis entitled **"A study on genetic diversity and marker trait association analysis in turmeric *(Curcuma longa* L.)"** is based on the original work done by me under the guidance of Dr. D. Prasath and has not been included in any other thesis submitted previously for the award of any degree. The contents of the thesis have undergone plagiarism check using iThenticate® software at C.H.M.K. Library, University of Calicut, and the similarity index found is within the permissible limit. I also declare that the thesis is free from AI generated contents.

(Aswathi. A.P)



(D. Prasath)

Place : Kozhikode
Date : 19.12.2024

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to God almighty for serving as my constant source of inspiration, providing strength and guidance through the ebbs and flows of my research journey.

Throughout this academic pursuit, I have been privileged to receive support from numerous individuals in various capacities, and I'm sincerely thankful for their invaluable contributions, which have been instrumental in bringing this work to fruition.

I would like to express my sincere gratitude to Dr. D. Prasath, my guide, for all the support, valuable insights and guidance that have been offered throughout my doctoral journey. I'm deeply grateful for all his efforts for the completion of this work and thesis.

I'm grateful to Dr. R Dinesh, Director of ICAR-IISR, Kozhikode, for granting me permission to conduct my research and providing access to essential laboratory facilities. Additionally, I appreciate the support of Dr. C K Thankamani, Dr. J. Rema, Dr. Santhosh J. Eapen, and Dr. K. Nirmal Babu, former Directors of the institute, for providing me with the opportunities and institutional support.

Special thanks to Dr. T. E. Sheeja, Head of the Division of Crop Improvement and Biotechnology for providing me with the ddRAD sequencing data resources that have been crucial for my doctoral study. I also extend my sincere gratitude to Dr. K. V. Saji and Dr. J. Rema, former Heads of the Division, for their help support throughout my research journey.

I am immensely thankful to Dr. Mukesh Sankar S, Division of Crop improvement and Biotechnology for helping me with the population structure analysis and association mapping that was very crucial for the completion of my study. Without his valuable help, I must admit that I may not be able to complete my thesis within this time span. Additionally, I wish to express my appreciation to Dr.

Sona Charles for igniting a spark within me to begin my own path in bioinformatics. I must say that her promptness in action whenever I approached her with doubts and attitude towards her field of expertise has really inspired me. I'm also grateful to Dr. Aarthi.S and Dr. Gopu for their valuable suggestions and helping me out with NTSysPC dendrograms that was a significant result for the completion of my first research paper.

Working with the genomics of a polyploid crop presented unique challenges that tested my perseverance and problem-solving skills. The inherent complexity and variability of these species often led to uncertainty and a lack of clarity in my research, pushing me to explore more. In my days of shooting in the dark, I would like to sincerely acknowledge the affirmations and guidance I received from Dr. Marilyn L Warburton, Research Geneticist at USDA ARS and Dr. Taylor Chafin, Senior Bioinformatician at Biomathematics and Statistics Scotland (BioSS), Edinburg. Their guidance was far more equivalent than reading hundreds of research papers. I am extremely grateful for Mr. Abhinav of NCBS Bangalore for patiently explaining do's and don'ts in the ddRAD data analysis. I also thank Mr. Praveen Prakash, Research Scholar at NCBS Bangalore for helping me assimilate my understanding gained through literature survey and providing me with the much needed on clarity on ddRAD sequencing data pre-processing.

I am thankful to Dr. Hemant Agrawal, Flowsols pvt limited for performing the flow cytometric analysis of the core collection that was an important part of my morphological studies. I also would like to thank Dr. Himanshu for his guidance that had helped me to gain a good understanding of the flow cytometric technique. I also express my sincere thanks to Mr Akshay, Genespectrum pvt limited, Pune for helping me with the computationally intensive GATK based variant calling.

I am deeply thankful to my former labmates Anisha, Raghav, and Diana for their assistance during my biochemical work.

I am very grateful to Dr. C. Sarathambal, Division of Crop Protection for her unwavering support, love, and care throughout my research journey.

I am deeply indebted to all the scientists of the Division of Crop Improvement and Biotechnology at ICAR-Indian Institute of Spices Research for their concern, support, suggestions, and encouragement.

I express my sincere thanks to Ms. Shorna of Bionome pvt limited who taught me 3 months basic NGS based bioinformatic course that was worth it. I cannot describe in words the joy I had when I troubleshooted and successfully run my first NGS pipeline wherein Ms. Shorna patiently supported me online for a long 2 hours giving me necessary directions whenever I stumbled.

I express my gratitude to the Council of Scientific and Industrial Research for their financial assistance that supported my research journey. Additionally, I extend my thanks to the University of Calicut for granting me permission to enroll in the Ph.D. program and for their continuous support and assistance, from the initial registration to the successful completion of my Ph.D. I am deeply grateful to the staff and members of the doctoral committee at Calicut University for their invaluable assistance and valuable suggestions.

I extend my heartfelt gratitude to all the members of the HRD, PME, Library, Accounts, and Administration departments of ICAR-IISR for their invaluable assistance and support.

I am also grateful to Mr. A. Sudhakaran for his timely assistance with photography and artwork throughout the research process. I thank Mr. K. Jayarajan for his assistance in data analysis and statistical evaluation that was indispensable for the completion of all my research papers as well as thesis.

I express my special thanks to Mr. Harish B T, technical staff at ICAR-IISR experimental farm and to all supporting staff, viz., Ms. Raji and Ms. Reena for the help extended throughout this research work.

I would like to express my sincere gratitude to the supporting staff at the ICAR-IISR experimental farm in Peruvannamuzhi, including, Mr. Muralidharan, Mr. Padmanabhan, Mr. Gineesh, Mr. Ananthu, and Ms. Sheeba. Their unwavering assistance in all aspects of field trials gave me great relief and I deeply appreciate

it. The patterns that emerged from my phenotypic evaluation readings are imbued with the dedication and hard work of the field workers who tirelessly supported my field experiments in turmeric.

It brings me immense pleasure to express my gratitude to all my friends: Neenu, Theertha, Nazmin, Anusree, Raghuveer, Karthika, Zumaila, Aiswarya, Saljuna, Arya, Vijayasanthi, Dona, Fayad, Megha, Vidya, and Sreena, for their unwavering help and encouragement. Additionally, I extend my heartfelt thanks to my beloved seniors, Prashina, Aparna, Snigdha, Alka, Summayya and Subila for their enduring love and support. This journey has been truly remarkable, and I will forever treasure the memories of the love, assistance, and sincere support bestowed upon me by everyone.

I take this opportunity to express my sincere gratitude to all my colleagues at Food Safety department, Malappuram who had to bear additional charge due to my absence for thesis related work. I must say that I'm deeply touched and grateful for their willingness to take up additional responsibilities that was completely devoid of any extra remuneration, efforts that helped towards easing my workload and most importantly taking all those extra burdens without a word of complaining that demonstrated the true spirit of teamwork and cooperation during this crucial time that allowed me to focus on completing my doctoral thesis.

It would not be fair at all if I do not acknowledge my alma matter IISER Thiruvananthapuram for its significant role in deciding the course of my journey. It was an amazing journey. My heartfelt gratitude for IISER Thiruvananthapuram for providing us an academic environment of equality and openness where professors actively encouraged us to scrutinize their arguments and challenge assumptions, fostering an atmosphere where the freedom to question became our most cherished principle. This environment has made deep imprints in my mind that enabled me to grow as a researcher and I am indeed grateful for everything that IISER have given me. Additionally, I am deeply grateful for both my mentors at IISER Thiruvananthapuram, Dr. N. Sadanananda Singh, Assistant Professor in Biology whose patience and invaluable guidance have formed the researcher in me and Dr.

Rajendar Goreti, Assistant Professor in Organic Chemistry whose care and encouragement have also played a significant role in my journey. I thank all my professors and teachers who have contributed to my academic journey.

Last but not the least, I would like to express my deepest gratitude to my parents and my dear brother, whose unwavering love, support, and sacrifices have been the bedrock of my journey. Their encouragement, guidance, and endless sacrifices have shaped me into the person I am today. My heartfelt gratitude extends to my parents, who not only contributed to the meticulous data collection during my field experiments but also provided unwavering emotional support throughout this academic journey. My parents have helped me in all possible ways to the best of their abilities in this thesis completion. I am deeply grateful to my brother for his invaluable assistance during various stages of this research especially helping me working with Python environment in the bioinformatics analysis during my learning phase. His support, encouragement, and insightful feedback have been instrumental in completing my thesis. I am forever indebted to my family for their boundless love and unwavering belief in my abilities.

Finally, I extend my gratitude to everyone who has directly or indirectly contributed to this long journey of research. As this chapter of my academic journey concludes, I am reminded that this is just the beginning of my research journey. In the words of T.S. Eliot, "Every end is a beginning. "Thank you all for being part of this significant milestone of my journey.

Aswathi. A.P

Dedicated to My Family

CONTENTS

	Page No.
Chapter 1	**1-4**
INTRODUCTION	
1.1 Phenotypic variability in turmeric	2
1.2 Characterization of genetic diversity in turmeric	2
1.3 Trait marker association study	3
Chapter 2	**5-32**
REVIEW OF LITERATURE	
2.1 Taxonomy of turmeric	5
2.2 General plant morphology	5
2.3 Turmeric varieties in India	7
2.4 Turmeric cultivation and post-harvest management	7
2.4.1 Uses of turmeric	8
2.5 Genetic diversity and crop improvement in turmeric	12
2.5.1 Morphological studies	13
2.5.2 Cytogenetic studies	16
2.5.3 Biochemical studies	18
2.6 Crop improvement by biotechnology in turmeric	21
2.6.1 Molecular studies	22
2.6.2 Molecular marker-based trait improvement studies	28
Chapter 3	**33-70**
MATERIALS AND METHODS	
3.1 Field experiment: Experimental site	33
3.1.1 Climate	33
3.1.2 Soil	33
3.2 Experiment details	34
3.2.1 Land preparation, experiment design and maintenance	34
3.3 Plant materials	36
3.4 Phenotypic evaluation	38
3.4.1 General plant morphological characterization	39
3.4.2 Rhizome characterization	40
3.4.3 Yield related characters	42
3.4.4 Biochemical characterization	43

3.5 Determination of ploidy level using flow cytometry 44
3.6 Statistical analysis 45
3.6.1 Analysis of variance (ANOVA) 45
3.6.2 Duncan multiple range test (DMRT) 46
3.6.3 T test 46
3.6.4 Principal component analysis (PCA) 46
3.6.5 Estimation of genetic parameters 46
3.6.6 Phenotypic and genotypic variance 46
3.6.7 Phenotypic and genotypic coefficients of variation (%) 47
3.6.8 Heritability (h2) 47
3.6.9 Genetic advance (GA) 48
3.6.10 Genetic advance as per cent of mean (GAM) 48
3.6.11 Heritability in broad sense 48
3.7 Genotypic analysis 49
3.7.1 Genomic DNA isolation 49
3.7.2 PCR analysis 53
3.8 ddRAD based genotyping 54
3.8.1 Genotypic data 54
3.9 NGS data analysis: available tools and programs 55
3.9.1 Bioinformatics pipeline 56
3.10. Population structure and genetic diversity assessment 64
3.10.1 Admixture model 64
3.10.2 K value 64
3.10.3 Q matrix 64
3.10.4 Population structure analysis 65
3.11 Association analysis 66
3.11.1 Data preparation: Input files 67
3.11.2 Association analysis 68
3.12 Interpreting and reporting the findings 69
Chapter 4 71-153
RESULTS AND DISCUSSION
4.1 Overview 71
4.2 Objective 1. To study the level of phenotypic variability among core turmeric collections 72
4.2.1 Morphological characterization (pooled data over two years) 73
4.2.2 Rhizome characters 78

4.2.3 Yield related characters 88
4.3 Statistical analysis 94
4.3.1 ANOVA results 94
4.4 Flow cytometric analysis 96
4.5 Comparison between triploids and tetraploids 98
4.5.1 Principal component Analysis and genetic variability analysis 104
4.6 Discusssion on results of objective 1 105
4.6.1 Morphological, yield and curcuminoid content characterization in turmeric 105
4.6.2 Ploidy level variation in turmeric, implications in breeding 107
4.6.3 Ploidy level and agro-morphological traits 108
4.7 Objective 2: To genetically characterize the genotypes using available molecular markers and develop novel markers 111
4.7.1 Molecular characterization with available set of markers: preliminary screening of released varieties 111
4.7.2 Genetic similarity analysis 114
4.7.3 Identification of most unique or divergent genotypes 115
4.7.4 Screening of turmeric core collection with identified reliable and polymorphic SSR markers 116
4.8 Discussion on results of objective 2 118
4.8.1 Molecular characterization of released varieties 118
4.8.2 Genotyping of core collection 120
4.9 Objective 3: To determine the association of rhizome characters and quality traits with molecular markers in turmeric 122
4.9.1 Genotyping 122
4.10 Population structure analysis using STRUCTURE software 129
4.11 Phenotypic data for association analysis 133
4.11.1 BLUP value generation through Meta-analysis using R 133
4.12 Marker-trait association analysis using GAPIT and the BLINK model 135
4.13 BLAST analysis of GWAS-identified MTAs 142
4.14 Discussion of objective 146
Conclusion 152
Chapter 5 155-157
RECOMMENDATIONS
REFERENCES 159-179

LIST OF FIGURES

Figure No.	*Title*	*Page No.*
2.1	Illustration of *Curcuma longa* by Franz Eugen Kohler, from *Kohler's Medicinal Plants*, 1887.	6
2.2	2D molecular structures of curcumin	11
3.1	Land preparation, planting, weeding, and mulching	35
3.2	Turmeric plants after four months of planting	35
3.3	Fully grown turmeric plants after six months of planting	35
3.4	The characteristics of core collection of turmeric utilized in this study	36
3.5	ddRAD library preparation and sequencing workflow	55
3.6	NGS data analysis pipeline followed for SNP genotyping	57
3.7	Detailed workflow of bioinformatics pipeline for SNP discovery	58
3.8	Variant discovery pipeline using GATK	62
4.1	Morphological variation presents in the rhizome morphology of turmeric genotypes	72
4.2 a-c	Histograms showing G_0/G_1 peak of 3 representative turmeric genotypes	96
4.3 a-e	Density plot of phenotypic characters with significant difference between triploids and tetraploids	99
4.3 f-j	Density plot of phenotypic characters with significant difference between triploids and tetraploids	100
4.3 k-n	Density plot of phenotypic characters with significant difference between triploids and tetraploids	101
4.4	PCA Biplot based on phenotypic evaluation of triploids and tetraploids	104
4.5	PCR amplification profile of 18 improved varieties using CuMisat 19, ISSR-02, CuMisat 13 and CuMisat 03	112
4.6	Dendrogram obtained based on hierarchical clustering of 18 released turmeric genotypes	114
4.7	Improved turmeric genotypes are shown in x-axis and the total number of individual genotypes shared in the same cluster across all primer’s dendrograms by the respective	115

	genotype in x-axis is given in y-axis	
4.8	Dendrogram of genetic similarity between 93 turmeric genotypes	116
4.9 a	Agarose gel electrophoresis showing QC of samples	122
4.9b	Bioanalyzer image of sample	123
4.10	Per base sequence quality analysis, demonstrating consistently high-quality scores across the reads	123
4.11	Distribution of quality scores, confirming a prominent peak around 35, indicating the presence of superior quality bases	124
4.12	Sequence length distribution, illustrating consistent patterns aligning with anticipated read lengths for ddRAD-Seq	124
4.13	GC content distribution, revealing a normal distribution centered around the expected GC content	125
4.14	Adapter contamination analysis, indicating minimal adapter sequences across the samples	125
4.15	Overrepresented sequences, identifying specific sequences occurring at higher frequencies	126
4.16	BWA alignment statistics	127
4.17	VCF file obtained after GATK variant calling	128
4.18	VCF file sorted in hapmap format using TASSEL	128
4.19	Bar plot of population structure	129
4.20	ΔK plot showing the optimum number of subpopulation present	132
4.21a-b	Distribution of length of primary rhizomes across genotypes	134
4.22a-b	Distribution of girth of primary rhizomes across genotypes	134
4.23a-b	Distribution of curcuminoid content across genotypes	135
4.24	Manhattan plot of SNP associations for length of primary rhizome	138
4.25	Manhattan plot of SNP associations for girth of primary rhizome	139
4.26	Manhattan plot of SNP associations for curcuminoid content	140
4.27a-c	Q-Q plot for the trait curcuminoid content, length of primary rhizomes per plant and girth of primary rhizomes per plant	141

LIST OF TABLES

Table No.	*Title*	*Page No.*
2.1	Scientific classification of turmeric	5
2.2	Uses of turmeric	8
2.3	Nutritional profile of turmeric	10
2.4	Reported medicinal properties in turmeric	11
3.1	Soil profile of the experimental site	33
3.2	List of turmeric genotypes used in this study	36
3.3a	Characteristics of ISSR Primers used in the study	51
3.3b	Characteristics of SSR Primers used in the study	52
4.1	General plant morphological data of turmeric genotypes	75
4.2	Rhizome characterization data of turmeric genotypes	83
4.3	Yield related characterization and curcuminoid content of turmeric genotypes	90
4.4a	ANOVA of general plant morphological characters of turmeric genotypes	94
4.4b	ANOVA of rhizome and yield characters of turmeric genotypes	95
4.4 c	ANOVA of yield characters and curcuminoid content of turmeric genotypes	95
4.5	List of triploid and tetraploids turmeric genotypes based on flow cytometric analysis	97
4.6	Summary of ploidy level analysis with range of coefficient of variation (CV)	98
4.7	Results of t test for phenotypic characters between triploid and tetraploid genotypes with significant difference in means	102
4.8	GCV, PCV, h^2 and GA estimates for significant characters among genotypes	103
4.9	Important characters that can be considered during selection process	103
4.10	Characteristics of SSR and ISSR markers obtained after the PCR screening of varieties	112
4.11	Population membership matrix	130

4.12	Meta-analysis results - BLUP values	133
4.13	Significant marker trait associations	137
4.14	BLASTN similarity analysis results	143

ABBREVIATIONS

µg	-	Microgram
µl	-	Microlitre
µM	-	Micromolar
^{o}C	-	Degree celcius
BDMC	-	Bisdemethoxycurcumin
BIC	-	Bayesian information content
BLAST	-	Basic local alignment search tool
BLINK	-	Bayesian logistic interaction kernel
Bp	-	Base pair
BWA-mem	-	Burrows-wheeler alignment – maximal exact match
Cm	-	Centimeter
CGH	-	Comparative genomic hybridization
ClRSM	-	*Curcuma longa* resistance specific marker
ClSTS	-	*Curcuma longa* sequence-tagged site
Cur	-	Curcumin
Cv	-	Cultivar
DAP	-	Days after planting
ddRAD seq	-	Double digest restriction-site associated sequencing
DMC	-	Demethoxycurcumin
DNA	-	Deoxyribonucleic acid
DNase	-	Deoxyribonuclease
dNTP	-	Deoxynucleotide triphosphate
EST-SSR	-	Expressed sequence tag-derived simple sequence repeats
FCM	-	Flow cytometry
FISH	-	Fluorescence *in situ* hybridization
GA	-	Genetic advance
GAPIT	-	Genome association and prediction integrated tool
GATK	-	Genome analysis toolkit
Gb	-	Gigabase
Gbp	-	Gigabase pair
GCV	-	Genotypic coefficient of variation
GWAS	-	Genome wide association analysis
IISR	-	Indian Institute of Spices Research

ISSR	-	Inter simple sequence repeats
M	-	Molar
Mb	-	Megabase
Mg	-	Milligram
Min	-	Minute
Ml	-	Millilitre
mM	-	Millimolar
Mm	-	Millimeter
NCBI	-	National center for biotechnology information
Ng	-	Nanogram
NGS	-	Next generation sequencing
Nm	-	Nanometer
PCR	-	Polymerase chain reaction
PCV	-	Phenotypic coeficient of variance
pH	-	Potential of hydrogen
PIC	-	Polymorphic information content
PVP	-	Polyvinylpyrrolidone
RAPD	-	Random amplified polymorphic DNA
RNA	-	Ribonucleic acid
Rnase	-	Ribonuclease
Rpm	-	Revolutions per minute
S	-	Second
SNP	-	Single nucleotide polymorphism
SSR	-	Simple sequence repeats
STS	-	Sequence-tagged site
UPGMA	-	Unweighted pair group method with arithmetic mean
UV	-	Ultraviolet

// ABSTRACT

ABSTRACT

Turmeric (*Curcuma longa* L.) commonly known as golden spice is an economically important spice crop. Turmeric genotypes exhibit wide intraspecific variation for the biologically active principles coupled with morphological variation. Given the large influence of the environment on phenotypic variability, there is no clear understanding about the relation of genetic base with the observed phenotypic variability. Here, we have studied the genetic diversity in turmeric diversity panel comprising 93 genotypes by examining the phenotypic variability alongside the molecular genetic variability to find marker-trait associations.

Phenotypic characterization revealed significant genetic variation among 93 genotypes. Flow cytometric analysis revealed that among the 93 genotypes, majority were triploids (84) and nine were tetraploids. Most of the tetraploids have origins as seedling progenies while few are germplasm collection. It was found that triploids and tetraploids differ significantly in their plant height, number of shoots, number of leaves on main shoot, petiole length, leaf length, length of mother rhizome, length of primary rhizomes, length of secondary rhizomes, girth of secondary rhizomes, inner core diameter of primary rhizome, weight of mother rhizomes per plant, weight of primary rhizomes per plant, total rhizome weight per plant and dry rhizome weight per plant. Genetic variability analysis revealed that characters; number of shoots per plant, length of mother rhizome, total weight of mother rhizomes per plant, total weight of rhizome per plant and dry weight of rhizome per plant were the most important traits for selection across ploidy levels.

Here, we have utilized molecular marker based genotyping as well as NGS based genotyping to study the genetic relationship of the genotypes under study. Although, reliable and reproducible, the inherent limitations of molecular marker to achieve the requisite resolution for association studies led us to proceed with NGS based genotypic data. Here, we have optimized a bioinformatics pipeline or workflow for variant calling which have resulted in discovery of 30438 SNP variants. Further, we have analysed the population structure of genotypes under

study (to avoid false positive marker trait associations in downstream analysis) and have found that there are three subpopulations. Further the association mapping for three key traits has revealed nine significant marker trait association for : curcumin content (1), rhizome length (5), and rhizome girth (3). Most marker trait associations were from chromosome 4. *Zingiber officinale* MADS-box transcription factor 34-like (LOC122034397) can be a putative locus involved in rhizome length morphology. Thus, results obtained here will add onto the understanding of genetic base of these complex traits in turmeric and to further crop improvement efforts in turmeric.

Keywords: Turmeric, Phenotypic characterization, molecular markers, genotyping, ploidy analysis, association analysis, marker trait associations

സംഗ്രഹം

മഞ്ഞളിലെ ജനിതക വൈവിധ്യത്തെയും മാർക്കർ ട്രെയ്റ്റ് അസോസിയേഷൻ വിശകലനത്തെയും കുറിച്ചുള്ള ഒരു പഠനം (*കുർക്കുമ ലോംഗ എൽ.)*

സുവർണ്ണ സുഗന്ധവ്യഞ്ജനം എന്നറിയപ്പെടുന്ന മഞ്ഞൾ (*കുർക്കുമ ലോംഗ എൽ.)* വാണിജ്യപരമായി പ്രാധാന്യമുള്ള ഒരു സുഗന്ധവിളയാണ്. മഞ്ഞൾ ജൈവശാസ്ത്രപരമായി സജീവമായ തത്വങ്ങളിൽ വൈവിധ്യമാർന്ന ഇൻട്രാസ്പെസിഫിക് വ്യതിയാനവും രൂപാന്തര വ്യതിയാനവും കാണിക്കുന്നു. ഫിനോടൈപ്പിക് വേരിയബിലിറ്റിയിൽ പരിസ്ഥിതിയുടെ വലിയ സ്വാധീനം കണക്കിലെടുക്കുമ്പോൾ, നിരീക്ഷിച്ച ഫിനോടൈപ്പിക് വേരിയബിലിറ്റിയുമായുള്ള ജനിതക അടിത്തറയുടെ ബന്ധത്തെക്കുറിച്ച് വ്യക്തമായ ധാരണയില്ല. ഇവിടെ, *93* ജനിതകരൂപങ്ങൾ അടങ്ങിയ മഞ്ഞൾ വൈവിധ്യ പാനലിലെ ജനിതക വൈവിധ്യത്തെക്കുറിച്ച് ഞങ്ങൾ പഠിച്ചു.

93 ജനിതകരൂപങ്ങൾക്കിടയിൽ കാര്യമായ ജനിതക വ്യതിയാനം ഫിനോടൈപ്പിക് സ്വഭാവം വെളിപ്പെടുത്തി. *93* ജനിതകരൂപങ്ങളിൽ ഭൂരിഭാഗവും ട്രിപ്ലോയിഡുകളും *(84)* ഒമ്പതും ടെട്രാപ്ലോയിഡുകളുമാണെന്ന് ഫ്ലോ സൈറ്റോമെട്രിക് വിശകലനം വെളിപ്പെടുത്തി. ഭൂരിഭാഗം ടെട്രാപ്ലോയിഡുകൾക്കും തൈകളുടെ വംശജരായി ഉത്ഭവമുണ്ട്, ചിലത് ജേംപ്ലാസം ശേഖരണമാണ്.

ട്രൈപ്ലോയിഡുകളും ടെട്രാപ്ലോയിഡുകളും അവയുടെ ചെടികളുടെ ഉയരം, ചിനപ്പുപൊട്ടലിന്റെ എണ്ണം, പ്രധാന ചിനപ്പുപൊട്ടലിലെ ഇലകളുടെ എണ്ണം, ഇലഞെട്ടിന് നീളം, ഇലയുടെ നീളം, മദർ റൈസോമിൻ്റെ നീളം, പ്രാഥമിക റൈസോമുകളുടെ നീളം, ദ്വിതീയ റൈസോമുകളുടെ നീളം, ദ്വിതീയ റൈസോമുകളുടെ ചുറ്റളവ്, പ്രൈമറി റൈസോമിന്റെ അകത്തെ കാമ്പ് വ്യാസം, ഓരോ ചെടിയുടെയും മദർ റൈസോമുകളുടെ ഭാരം, ഓരോ ചെടിയുടെയും പ്രാഥമിക റൈസോമുകളുടെ ഭാരം, ഓരോ ചെടിയുടെയും ആകെ റൈസോമിന്റെ ഭാരം, ഓരോ ചെടിയുടെയും ഉണങ്ങിയ റൈസോമിന്റെ ഭാരം സ്വഭാവസവിശേഷതകളിൽ കാര്യമായ വ്യത്യാസമുണ്ടെന്ന് കണ്ടെത്തി കാര്യമായ വ്യത്യാസമുണ്ടെന്ന് കണ്ടെത്തി. ജനിതക വ്യതിയാന വിശകലനം വെളിപ്പെടുത്തിയത് പ്രതീകങ്ങൾ*;* ഒരു ചെടിയിലെ ചിനപ്പുപൊട്ടലിന്റെ എണ്ണം, മദർ റൈസോമിന്റെ നീളം, ഓരോ ചെടിയുടെയും അമ്മ റൈസോമുകളുടെ ആകെ ഭാരം, ഓരോ ചെടിയുടെയും

റൈസോമിന്റെ ആകെ ഭാരം, ഓരോ ചെടിയുടെയും റൈസോമിന്റെ ഉണങ്ങിയ ഭാരം എന്നിവ പ്ലോയ്ഡി തലങ്ങളിലുടനീളം മികച്ച വിളകൾ തിരഞ്ഞെടുക്കുന്നതിനുള്ള ഏറ്റവും പ്രധാനപ്പെട്ട സ്വഭാവസവിശേഷതകളാണ്.

ഇവിടെ, പഠനത്തിൻ കീഴിലുള്ള ജനിതകരൂപങ്ങളുടെ ജനിതകബന്ധം പഠിക്കാൻ ഞങ്ങൾ മോളിക്യുലാർ മാർക്കർ അടിസ്ഥാനമാക്കിയുള്ള ജനിതകരൂപീകരണവും *NGS* അടിസ്ഥാനമാക്കിയുള്ള ജനിതകരൂപീകരണവും ഉപയോഗിച്ചു. വിശ്വസനീയവും പുനരുൽപ്പാദിപ്പിക്കാവുന്നത് ആണെങ്കിലും അസോസിയേഷൻ പഠനങ്ങൾക്ക് ആവശ്യമായ മിഴിവ് നേടുന്നതിന് തന്മാത്രാ മാർക്കറിന്റെ അന്തർലീനമായ പരിമിതികൾ, *NGS* അടിസ്ഥാനമാക്കിയുള്ള ജനിതക ഡാറ്റയുമായി മുന്നോട്ട് പോകാൻ ഞങ്ങളെ പ്രേരിപ്പിച്ചു. ഇവിടെ, *30438 SNP* വേരിയന്റുകൾ കണ്ടെത്തുന്നതിന് കാരണമായ വേരിയൻ്റ് കോളിംഗിനായി ഞങ്ങൾ ഒരു ബയോ ഇൻഫോർമാറ്റിക്സ് പൈപ്പ്ലൈൻ അല്ലെങ്കിൽ വർക്ക്ഫ്ലോ ഒപ്റ്റിമൈസ് ചെയ്തിട്ടുണ്ട്. കൂടാതെ, പഠനത്തിൻ കീഴിലുള്ള ജനിതകരൂപങ്ങളുടെ ജനസംഖ്യാ ഘടന ഞങ്ങൾ വിശകലനം ചെയ്തു (പിന്നീടുള്ള വിശകലനത്തിൽ തെറ്റായ പോസിറ്റീവ് മാർക്കർ സ്വഭാവസവിശേഷതകൾ ഒഴിവാക്കുന്നതിന്) *3* ഉപജനസംഖ്യകൾ ഉണ്ടെന്ന് കണ്ടെത്തി. കൂടാതെ, *3* പ്രധാന സ്വഭാവസവിശേഷതകൾക്കായുള്ള അസ്സോസിയേഷൻ മാപ്പിംഗ് *9* പ്രധാന മാർക്കർ സ്വഭാവസവിശേഷതകൾ വെളിപ്പെടുത്തി : കുർക്കുമിൻ ഉള്ളടക്കം *(1)*, റൈസോമിന്റെ നീളം *(5)*, റൈസോം ചുറ്റളവ് *(3)*. മിക്ക മാർക്കർ ടെയ്റ്റ് അസോസിയേഷനുകളും ക്രോമസോം *4* ൽ നിന്നുള്ളവയാണ്. സിഞ്ചിബർ ഒഫീഷ്യനാലെ *MADS*-ബോക്സ് ട്രാൻസ്ക്രിപ്ഷൻ ഫാക്ടർ *34*-ലൈക്ക് *(LOC122034397)* റൈസോം ദൈർഘ്യത്തിന്റെ രൂപഘടനയിൽ ഉൾപ്പെട്ടിരിക്കുന്ന ഒരു പുറ്റേറ്റീവ് ലോക്കസ് ആകാം. അതിനാൽ, ഇവിടെ ലഭിച്ച ഫലങ്ങൾ മഞ്ഞളിലെ ഈ സങ്കീർണ്ണ സ്വഭാവങ്ങളുടെ ജനിതക അടിത്തറ മനസ്സിലാക്കുന്നതിനുള്ള ശ്രമങ്ങൾക്കും മഞ്ഞളിലെ വിള മെച്ചപ്പെടുത്തൽ ശ്രമങ്ങൾക്കും ആക്കം കൂട്ടും.

Chapter 1

INTRODUCTION

Turmeric (*Curcuma longa* L.) also known as 'Golden Spice' and 'Indian Saffron' is one of the important spice crops in the world (Prasad & Aggarwal, 2011). It is a rhizomatous herbaceous perennial plant belonging to the family Zingiberaceae, which is native to tropical South Asia (Kumar et al., 2017). *C. longa* is one species within the genus Curcuma, which is comprised of about 110 species distributed chiefly in tropical regions (Li et al., 2011). It grows in tropical and subtropical regions throughout the world and possesses high nutritional value (Li et al., 2011). India is the main producer and exporter of turmeric in the world (Aswathi et al., 2023). It is commonly used as spice for flavoring and as a natural colorant with a variety of applications such as its use as a condiment, in cosmetics and as a dye, besides being a potential source of therapeutically important molecules like curcumin (Prasad & Aggarwal, 2011). There are innumerable medicinal uses of the turmeric, varying from cosmetic face cream to the prevention of cox-2 inhibitors (Duke, 2007). So, crop improvement of turmeric has been always in focus due to its increasing economical demand (Prasad & Aggarwal, 2011).

Although India is the top producer of turmeric including some high yielding cultivars, the average productivity and quality of turmeric are not satisfactory (Ayer, 2017). Major problems are non-availability of requisite high yielding genotypes, slow multiplication rate, low curcumin and essential oil content in available cultivars and loss due to disease during cultivation and storage (Ayer, 2017). Most of crop improvement in turmeric relies on clonal selection of existing high yielding cultivars and recently on development of new varieties using biotechnological means (Usha et al., 2006). In any case, a clear understanding of existing genetic pool of turmeric is the need of the hour, both for its utilization and manipulation (Govindaraj et al., 2015). Phenotypic traits are the most direct visible indicator of diversity in the

genetic pool. But it is the evaluation at the DNA level that gives us an accurate and in-depth view of genetic diversity (Arif et al., 2010).

1.1 Phenotypic variability in turmeric

Phenotypic traits which include the morphological, physiological, or agronomic characteristics are a direct indicator to estimate intraspecific diversity (Bhanu, 2017). These traits involve the complex interaction between individual genotypes and their environments (Anandaraj et al., 2014). Therefore, phenotypic evaluation is fundamental to the genetic diversity studies. Turmeric exhibits wide intraspecific variation for the biologically active principles coupled with morphological variation with respect to above ground vegetative and floral characters as well as the below ground rhizome features besides for curcumin, oleoresin, and essential oil (Anindita et al., 2020). Most often cytological studies supplement morphological information. At the cytological level, studies have reported that majority of turmeric varieties is triploid with widely accepted chromosome number as 2n=63, while some are tetraploid with 2n=84 (Nair & Sasikumar, 2009). These form the major ploidy observed in turmeric. In general, the short duration types are tetraploids (2n=84) while medium and long duration types are triploids (2n=63) (Nair et al., 2010). As turmeric is a vegetatively propagated crop, the only way that it can have such a wide range of chromosomal counts is *via* means of sexual reproduction. Also, there is a lack of clear information regarding the ploidy level of different genotypes of turmeric. Hence ploidy analysis is useful as it serves as an anchor to study genetic diversity as change in ploidy level drives genomic variation contributing to genetic diversity.

1.2 Characterization of genetic diversity in turmeric

Understanding of genetic diversity is obscured without molecular characterization (Arif et al., 2010). Genetic diversity enables plants to combat wide varying environmental challenges and drives evolution and speciation (Ellegren & Galtier, 2016). It is the genetic variations that act as material for evolutionary forces resulting in genetic diversity which manifest grossly as phenotypic traits. Therefore, correlation of phenotypic traits and genetic variation serves to better understand the genetic pool in terms of identification and discrimination of genotypes and to

explore evolutionary tendencies (Gogoi et al., 2023). Genetic characterization based on molecular markers is a reliable tool as they are cost effective, PCR based and highly reproducible (Singh et al., 2018).

In case of turmeric, lack of clear cut morphological traits among some cultivars coupled with vernacular identity of germplasm results in accumulation of duplicates in germplasm accessions, taxing heavily on conservation cost and hampering the crop improvement work (Shamina et al., 1998). But the molecular characterization of turmeric germplasm is limited, and no clear marker system is identified for genotype discrimination. Also, genotypic base behind wide phenotypic traits observed in turmeric is not adequately explored.

So, after characterizing genetic diversity using molecular markers, the markers can be utilized for genome wide association studies to identify association between phenotypic traits and marker system.

1.3 Trait marker association study

Association mapping which is based on extensive historical recombination in a diverse natural population has been an efficient method to identify quantitative trait loci for important agronomic traits in model and non-model plants (Zhu et al., 2008). Given the fact that there is large influence of environment in phenotypic trait expression of turmeric, association mapping helps us to dissect out the environmental influence and study the genetic base of observed phenotypic traits as there is no clear understanding now about the relationship between phenotype and its genetic base.

So, the objectives of this study are:

(i) To study the level of phenotypic variability among core set of turmeric collections.

(ii) To genetically characterize the genotypes using molecular markers and develop novel markers.

(iii) To determine the association of rhizome characters and quality traits with molecular markers in turmeric.

Chapter 2

REVIEW OF LITERATURE

2.1 Taxonomy of turmeric

Turmeric is monocotyledonous, rhizomatous vascular green plant belonging to the family Zingiberaceae. Taxonomic hierarchy as well as the scientific classification of turmeric is given below (Table 2.1)

Table 2.1. Scientific classification of turmeric

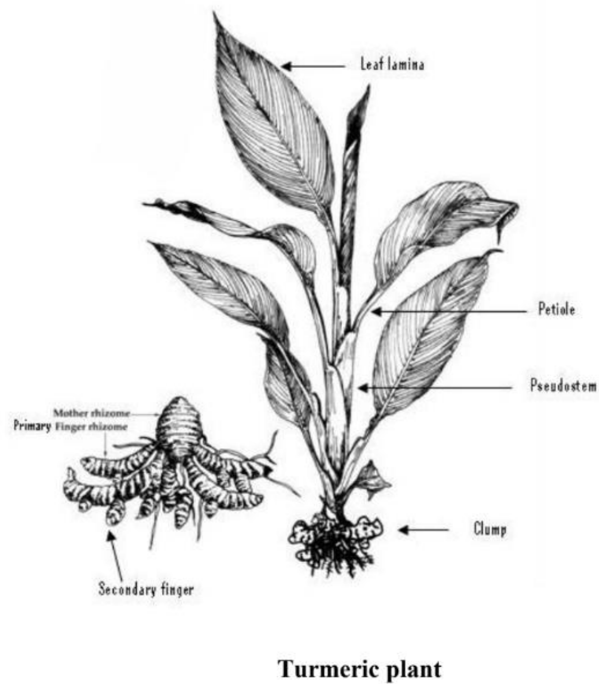
Rank	Scientific name
Kingdom	Plantae
Subkingdom	Viridiplantae
Superdivision	Embryophyta
Division	Tracheophyta
Class	Magnoliopsida
Superorder	Lilianae
Order	Zingiberales
Family	Zingiberaceae
Genus	Curcuma L.
Species	*Curcuma longa* L.

Source: Anon (2016)

2.2 General Plant Morphology

Turmeric is a perennial herbaceous plant that can grow up to a height of 100 cm. Leaves are alternate and comprise of leaf sheath, petiole, and leaf blade. Leaves are oblong to elliptical with narrow tips. Rhizomes are highly branched, with yellow to orange colour, cylindrical in shape and aromatic in nature. Rhizomes comprise of mother rhizomes that branches further to primary and secondary rhizomes. In East Asia, the turmeric flowers usually in August. Inflorescence stem which contains many flowers lies terminal to the false stem. The bracts are ovate to oblong with a

blunt upper end and light green colour (Prasath et al., 2019). Stem bracts, which are white to green in colour and occasionally stained reddish-purple, with tapered upper ends, are present at the top of the inflorescence but do not bear flowers. The blooms are threefold, zygomorphic, bilaterally symmetrical, and hermaphrodite (Prasath et al., 2019). The flowers have three long, fused, fluffy-haired white sepals and three bright yellow petals that combine to form a tube-shaped corolla (Fig.2.1).



Turmeric plant

Fig. 2.1. Illustration of *Curcuma longa* L. by Franz Eugen Kohler, from *Kohler's Medicinal Plants*, 1887.

2.3 Turmeric varieties in India

Different turmeric cultivars are grown across India, mostly known after the locality where they are grown. Furthermore, enhanced turmeric varieties have also been released as a result of research studies of several Indian research institutes. There are around 30 such released turmeric varieties (Aswathi et al., 2023). IISR Pratibha, IISR Alleppey Supreme, Suguna, IISR Pragati, Sudarsana, Varna, etc are some of the major released varieties of turmeric in India. The cultivated turmeric varieties exhibit considerable level of variation in the plant morphology, rhizome characters, size, and curcumin content (Aswathi et al., 2023). High curcumin containing yellow orange coloured turmeric powder are highly sought after in world market. 'Madras', and 'Alleppey', both named after the locality of cultivation are the two prominent turmeric types from India at the international market (Anne, 2004). Alleppey turmeric with bright orange rhizomes are mainly exported to the United States, wherein it is preferred as a spice and as condiment. The vibrant, light-yellow rhizomes of Madras turmeric are primarily sought after by the Middle Eastern and British markets. Alleppey turmeric has a constitution of 3.5 - 5.5% volatile oils, and 4 - 7% curcuminoid content. Whereas the Madras type has a lower volatile oil (2%) and curcuminoid content (2%) (Anne, 2004). The variety from Patna is known for its deep colour. Of the two commonly grown types of turmeric in Maharashtra, "Lokhandi" have brightly coloured firm rhizomes, while the other variety has soft, light-colored rhizomes. Apart from 'Kasturi Pasupa' of the Godavari Delta, 'Armoor' from the Nizamabad area, and 'Chaya Pasupa,' the main types from Andhra Pradesh are 'Duggirala' of Guntur and 'Tekurpeta', both of which have long, smooth, and hard fingers. The notable varieties from Odisha are Rasmi, Suroma, Ranga, and Roma. Meghalayan Lakadong cultivars are renowned for their high curcumin concentration (more than 5%) (Kumar et al., 2017).

2.4 Turmeric cultivation and post-harvest management

A warm, humid environment is required for growth and cultivation of turmeric. It grows at a height of 1500 metres above sea level in hilly regions. For turmeric production, an optimal temperature between 20 and 30°C and an annual rainfall

ranging from 1500 to 2250 mm are required (Prasath et al., 2019). Even though mostly cultivated as rainfed crop, the crop can also be cultivated under irrigation. Turmeric is harvested approximately after 7 to 10 months of planting when the plant starts drying, and this depends on cultivar, soil, and growing conditions. Before drying, turmeric rhizomes are cured first. Curing refers to the boiling of the rhizomes till it becomes tender. It is performed to gelatinize the starch for a more uniform drying, and to remove the fresh earthy odour (Prasath et al., 2019).

2.4.1 Uses of turmeric

Turmeric is widely utilized for a wide range of purposes as summarised in Table 2.2. The nutritive profile of turmeric is given in Table 2.3. The bioactive chemicals present in turmeric that causes its medicinal properties is shown in Fig.2.2. In India, turmeric was in use as a traditional medicine for various skin disorders, respiratory and digestive disease etc (Prasad & Aggarwal, 2011). Turmeric is used as a spice and condiment in modern times as well, in addition to its use as a nutritional supplement for a host of health issues, including depression, lung infections, liver illness, arthritis, and digestive issues (Hewlings & Kalman, 2017). Table 2.4 shows some of the medicinal properties reported in turmeric.

Table 2.2. Uses of turmeric

Product name	**Description**	**Uses**
Whole rhizome (dried form)	Appearance: pale yellow, reddish-brown, or orange-brown. Chemical composition: it could include 1.5–5% essential oils and 3–15% curcuminoids. Preparation: Typically, mother and finger rhizomes are boiled separately for 40 to 60 minutes in a slightly alkaline solution. To reduce the moisture level by roughly 10%, sun-drying should be done for ten to fifteen days afterward.	Medicinal uses
Ground turmeric	Appearance: either reddish yellow or yellow in hue. Chemical composition: when exposed to light and during the process, the concentration of the primary active components (curcuminoids and essential oils) may decrease. So, the powder is packed in a UV-resistant container. The powder is made by grinding the dried finger	Utilised as food additive, colourant, spice, and medication.

Product name	Description	Uses
	rhizomes. Preparation: dried finger rhizomes are grounded to produce its powder	
Turmeric oil	Appearance: Yellow to brown colored oil Chemical composition: monoterpenes typically make up the majority of the essential oils extracted from turmeric leaves. Sesquiterpenes make around 90% of the oil from turmeric rhizomes. Preparation: Steam distillation or supercritical CO_2 extraction to get the extract from dried rhizomes or leaves.	Utilised as a food additive, colourant, spice, and medication.
Turmeric oleoresins	Appearance: a thick, dark yellow, reddish-brown liquid Chemical composition: they are composed of 37–55% curcuminoids and up to 25% essential oil. Preparation: organic solvents (acetone, dichloromethane, 1,2-dichloroethane, methanol, ethanol, isopropanol, and light petroleum (hexanes)) or supercritical CO_2 utilized for extraction to oleoresin from dried rhizomes.	Utilised as a food additive, colourant, spice, and medication.
Curcumin	Appearance: crystalline powder with a yellow to orange-red hue Chemical composition: a combination of curcumin and its derivatives, bisdemethoxy and demethoxy, in variable ratios. Ninety percent of curcumin may be made up of the three main types of curcuminoids. Resins and oils make up the minority of the makeup. Preparation: It is made from pulverised turmeric rhizomes by solvent extraction, and then the extract is purified by crystallisation. Acetone, carbon dioxide, ethanol, ethyl acetate, hexane, methanol, and isopropanol are examples of organic solvents used in extraction processes.	Used as medicine and dietary supplement

Source: (Ahmad et al., 2020)

Table 2.3. Nutritional profile of turmeric

Principle constituents	Nutrient value (Kcal)	Percentage of RDA (%)
Energy	354	17
Carbohydrates	64.9	50
Total fat	9.88	33
Protein	7.83	14
Cholesterol	0	0
Dietary fiber	21	52.5
Vitamins		
Pyridoxine	1.80	138
Folates	39	10
Niacin	5.140	32
Riboflavin	0.233	18
Vitamin A	0	0
Vitamin C	25.9	43
Vitamin E	3.10	21
Vitamin K	13.4	11
Electrolytes		
Potassium	2525	54
Sodium	38	2.5
Minerals		
Manganese	7.83	340
Calcium	183	18
Copper	603	67
Iron	41.42	517
Magnesium	193	48
Phosphorus	268	38
Zinc	4.35	39.5

Source: (Ahmad et al., 2020)

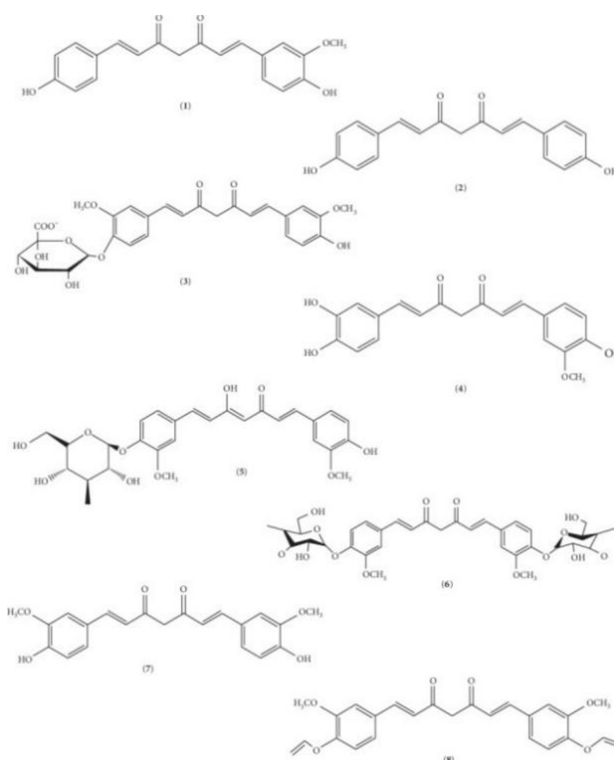


Fig. 2.2. 2D molecular structures of curcumin:

(1) demethoxycurcumin, (2) bisdemethoxycurcumin, (3) curcumin glucuronide, (4) monodemethylcurcumin, (5) curcumin monoglucoside, (6) curcumin diglucoside, (7) Keto-curcumin, and (8) allyl curcumin.

Source: (Ahmad et al., 2020)

Table 2.4. Reported medicinal properties in turmeric

Medicinal value	Reference
Antioxidant activity	Ammon & Wahl (1991)
Cardiovascular and antidiabetic effects	Khajehdehi et al. (2012)
Inflammatory and edematic disorders	Ramsewak et al. (2000)
Gastrointestinal effects	Davoodvandi et al.(2021), Rajasekaran (2011)
Anti-Cancer effects	Mansouri et al. (2020), Tomeh et al. (2019), Wilken et al.(2011)
Antimicrobial activity	Gul & Bakht (2015)
Hepatoprotective and renoprotective effects	Trujillo et al.(2013)
Alzemeirs and turmeric	Mishra & Palanivelu (2008)
Photoprotector activity	Deng et al. (2021)

2.5 Genetic diversity and crop improvement

Genetic diversity is the basic raw material for breeding and propagation of any crop plant (Bhanu, 2017). Exploration and utilization of genetic variation in plants dates to the very onset of agriculture in history of human civilization. Genetic diversity is an inexclusive part of farming, be it primitive subsistence agriculture or the modern-day commercial agriculture, ensuring food security by crop improvement and production (Govindaraj et al., 2015). Genetic diversity arises from genetic polymorphisms which can be present among species and within genomes which is also predictable from life history of organisms and can be used to understand the underlying evolutionary processes (Ellegren & Galtier, 2016). The goal of conservation genetics is to maintain prevailing genetic diversity at different levels and thus help in population monitoring and assessment that can be used for conservation planning (Ellegren & Galtier, 2016). Every individual has a unique genetic makeup. Even though conservation efforts and research mainly focus on the population level rather than the individuals, the assessment of genetic variation is always done at the individual level, and it can be determined for collections of individuals in a population/species (Govindaraj et al., 2015). Genetic variation at the individual level can manifest as morphological, biochemical, and cytological variation. Hence morphological characterization is useful to identify the genetic variation from variation either in quantitative traits (mainly polygenic traits that vary continuously, e.g., plant height) or discrete traits (e.g., flower colour) that fall into distinct categories (Gascon et al., 1996; Jenfaoui et al., 2021; Lahkar & Tanti, 2017; Ouaja et al., 2021; Petrova et al., 2014; Valuyskikh et al., 2019). Morphological characters studied include anatomical and structural characters as well in addition to the general plant morphology (quantitative and qualitative characters). Morphological markers together with biochemical and cytological markers/molecular markers provide a comprehensive view of genetic diversity and is used in diversity studies across plant species (Sammour, 2014). Thus, these studies encompass diverse fields of botany, bioanalytical chemistry, molecular biology as well as cytogenetics in agricultural research.

Turmeric varieties exhibit variation for morphological as well as biochemical characters (Bahadur et al., 2016; Roy et al., 2011.; Sandeep et al., 2016). Genetic factors may be responsible for variability seen in turmeric varieties under similar conditions. A single variety also shows variation according to changing environments. It has been found that agronomically relevant traits like yield and curcumin content is affected by agro-climatic interactions as well in addition to the genotype effect (Anandaraj et al., 2014). This necessitates the need of screening in multilocational trials and studies to develop stable and elite cultivars. Agro-morphological screening of turmeric varieties is an essential pre-requisite to find the variability in the traits and to further crop improvement efforts (Roy et al., 2011) .

2.5.1 Morphological studies in turmeric

Morphological studies in turmeric have dealt with the general morphological characterization of the crop in addition with economically relevant biochemical characters like curcumin content, oleoresin etc (Syamkumar, 2008). These morphological studies are fundamental to cultivar identification, selection, and improvement. These studies can be broadly based on either the agro-morphological examination of cultivars for the desirable traits of interest or screening of cultivars for their response to various stresses to identify or select best cultivars (Dev & Sharma, 2022). The agro-morphological studies in turmeric usually follows the DUS guidelines (PPV & FRA, 2009). In one such study, based on morphological evaluation of 15 genotypes of turmeric for 24 characters, the multiscale scores of DUS guidelines were used to analyse the cultivars (Aarthi et al., 2018a). It was found that for the 24 characters, 3 characters were monomorphic, 12 were dimorphic, and nine were polymorphic across the genotypes. Together with the DUS characters, this study has also found other significant characters that can serve as morphological markers for genotype identification, such as the collar girth of the pseudostem, the initial stage pigmentation in the leaf lamina that fades later, and the red pigmentation in the emerging shoot (at the initial stage of sprouting). There have been many studies in this direction wherein most of these studies aimed to understand the prevailing trend of genetic variability to enable selection of

desirable/superior genotypes along with correlated characters (Aarthi et al., 2018b; Anindita et al., 2020; Roy et al., 2011; Sivakumar et al., 2019). According to one such study (Roy et al., 2011), plant height, number of leaves, number of shoots, size of primary fingers size exhibited a significant positive association with yield of rhizomes. Another study done in Indonesia have reported an extensive phenotypic diversity among turmeric genotypes wherein the characters like plant height, number of shoots, number of leaves on the main shoot, length of petiole, length of leaf lamina, width of leaf lamina, number of mother rhizomes, total rhizome weight, weight per shoot, Pseudostem habit, leaf margin, and rhizome habit emerged as discerning traits for the collection under the study (Anindita et al., 2020) . Most of these studies have used clustering and principal component analysis for grouping turmeric accessions based on their morphological characteristics thereby to elucidate the pattern of genetic variation. One of the correlation studies reported a significant positive correlation of yield per plant with emergence, number of tillers and leaves per plant, leaf length, plant girth and height, length, core diameter and weight of mother, primary and secondary rhizomes, along with a significant negative correlation for incidence of rhizome rot with curcumin content (Dev & Sharma, 2022). The results of this study shows that the weight of the primary rhizome had the greatest direct beneficial impact on yield, followed by the length of the primary rhizome, the mother rhizome's core diameter, the length and weight of the secondary rhizome, emergence, plant girth, and height. In an earlier genetic study of turmeric germplasm based on 13 agro-morphological traits, it was found that a high heritability (h2BS%), genetic advance (GA), genotypic coefficient of variability (GCV) was recorded for dry and fresh weight of rhizomes (Gupta et al., 2016). This study indicates suitability of these characters as selective traits for good advance, as these were regulated by additive gene effects and also showed high co-heritability. This study also reported that leaf length and rhizome thickness had positive correlation with dry weight of rhizomes. Path coefficient analysis revealed that the fresh weight of rhizome had the largest direct contribution to dry weight of rhizome. There have been other studies as well that investigated pattern of morphological diversity in turmeric genotypes. One such study wherein 83 turmeric accessions

from five states in north-eastern India were examined using multivariate approaches, including clustering and principal component analysis, at a mid-altitude (1000 m elevation) in the high humid subtropics (Roy et al., 2011). Here, there was a positive and statistically significant association for the number of leaves, size of the primary fingers, and number of suckers with rhizome yield (Roy et al., 2011). The accessions were divided into four clusters, and morphological differences found within each cluster was of a very low order. This study also reiterates the utility of both clustering and principal component analysis as successful methodologies to categorise turmeric accessions according to their morphological traits and thus gain a better understanding of the existing diversity that can finally help in genotypic selection for crop development (Roy et al., 2011). Thus, studies like this not only help to connect characters to a particular trait (correlation analysis) that can be used for trait-specific selection but also to understand the overall pattern and extent of genetic variation. This in turn is essential to determine effective breeding programme and management strategies.

There had been studies as well that analysed morphology and physiology of turmeric in response to various environment stress (Roy et al., 2011). Various biotic and abiotic stresses are few of the major factors that limit turmeric production in India. One such study that investigated how the water shortage and reduced temperature affect growth of turmeric, it’s physiological adaptation as well as quantity and quality of rhizomes (Chintakovid et al., 2022). It was found that the water withholding, and controlled temperature acted as abiotic stresses and negatively affected turmeric plants by limiting the overall growth and curcuminoid yield of the plants. One of the main biotic pressures on turmeric cultivation is the root-knot nematode, *Meloidogyne incognita*. In a study which evaluated 70 cultivars for their resistance to *M. incognita*, it was found that ‘Dugirala’, ‘Ansitapani’, ‘PTS-31’, ‘PTS-42’, ‘PTS-47’ were fully resistant; ‘PTS-21’, ‘361 Gorakhpur’, ‘328 Sugandham’, were moderately resistant and the remaining cultivars were susceptible (Mohanta, 2015). This was further verified by DNA molecular marker-based studies. Investigation like this that help in understanding the degree and distribution

of genetic diversity among the cultivars based on response to biotic stresses would be a significant contribution to cultivar selection and improvement.

2.5.2 Cytogenetic studies in turmeric

Cytogenetics, an important part of cytology (study of cell structure and function) deals with chromosomes, their characteristics and segregation pattern during mitosis and meiosis in relation with cell behaviour. Techniques utilised include karyotyping, examination of G-banded chromosomes, various cytogenetic banding techniques, and molecular cytogenetics such as fluorescent in situ hybridization (FISH) and comparative genomic hybridization (CGH) (Bhadra & Bandyopadhyay, 2015). Ploidy level is an important cytogenetics statistic that refers to the number of sets of chromosomes in somatic cells in diplophase (2n) or gametophytic cells in haplophase (1n). It is usually represented by a number followed by the letter x (Crespel & Meynet, 2003). Karyological surveys based on chromosome counting provides the most efficient and direct method for ploidy level assessment (Maluszynska, 2003). This helps to determine the basic or fundamental chromosome number, a prominent karyotype feature that is useful in cytotaxonomy and evolutionary studies (Guerra, 2008).

In the genus Curcuma, the fundamental chromosomal number has been a source of ongoing controversy since the first karyological surveys. Previous investigations show a significant fluctuation in the somatic chromosomal counts in the genus Curcuma (Sugiura, 1931). In *C. longa*, numerous earlier cytogenetic studies have reported a chromosomal number of 2n=63 (Nair & Sasikumar, 2009). There have also been reports of deviations like 2n = 32 (Sato, 1948) 2n = 48 (Das et al., 1999), 2n = 62 (Raghavan & Venkatasubban, 1943; Sharma & Bhattacharya, 1959), 2n = 64 (Chakravorthy, 1948) and 2n = 84 (Renjith et al., 2001). The basic chromosome number of the genus Curcuma is x = 21, which was formed by secondary polyploidy or dibasic amphidiploidy from x = 9 and x = 12 (Nambiar, 1979; Ramachandran, 1961, 1969). Turmeric is commonly regarded as triploid with pollen fertilisation rate less than 60% (Nair et al., 2004; Nambiar, 1979). Although it was once thought that *C. longa* does not set seeds, in contrast to *C. aromatica* (Nambiar, 1979), seed

setting, germination and open pollinated progenies have been observed in the crop (Lad, 1993; Nair et al., 2004; Sasikumar et al., 1996).

Open pollinated progenies in turmeric improves the genetic variability and is important in breeding efforts in the crop as the yield and quality assessment of these open pollinated offspring have led to the development of high yielding cultivars (Sasikumar et al., 1996). But in most cases the information about the chromosome number of these seedling is limited. Also, only limited reports are available on the chromosome number in *C. longa* collections (Renjith et al., 2001). An earlier study has attempted the examination of the chromosome count in turmeric seedling descendants that were produced by open pollination and germplasm collections (Nair & Sasikumar, 2009). This study reports 2n = 63 as the most common chromosomal number found in the germplasm collections. The existence of genotypes with 2n = 84 (Acc. No. 300) confirms earlier studies that cultivars of *C. longa* are available with both 2n = 63 and 2n = 84 chromosome numbers (Nair & Sasikumar, 2009). Turmeric's chromosome number is more or less stable due to the vegetative propagation by rhizomes, unless affected by any sexual processes or uncommon somatic changes (Nair et al., 2010). But it is the reproductive process that produces offspring with different chromosome counts because of triploid segregation that makes variations in a triploid like turmeric possible. Therefore, it is hypothesised that 2n = 84 and 2n = 61 may have descended from common turmeric (2n = 63) by natural seed propagation (Nair et al., 2010). Considering the later reports, it can be affirmed that the earlier assertion that triploid *C. longa* will not set seed (Ramachandran, 1961) is unacceptable (Lad, 1993; Nair et al., 2004; Sasikumar et al., 1996). The analysis of other *C. longa* cultivars can further reveal the prevalence of various cytotypes in cultivated turmeric. Another recent study investigating the karyomorphological characters of some important plant species of Zingiberaceae have revealed insights into chromosomal morphology of turmeric as well (Bhadra & Bandyopadhyay, 2015). Here, they have analysed *C. longa* root tip cells which displayed 2n = 63 chromosomes as reported in most of past studies. So, it was reported that the somatic cells of *C. longa* contained 2n = 63 chromosomes, and the karyotype formula was 2n = 63= 11M + 48m + 4Sm. There were three sorts

of centromeric positions: median, almost median, and submedian. A 28% to 49.51% centromeric index was used. There were no chromosomes that had secondary constriction (Bhadra & Bandyopadhyay, 2015). Thus, all these previous works have helped to gain significant understanding of cytogenetics of *C. longa.*

2.5.3 Biochemical studies in turmeric

Studies with mice models have revealed the biological activities of curcumin or turmeric extracts (Sharifi-Rad et al., 2020). Particularly, it has been shown that taking curcumin/turmeric extracts orally in the right quantities can prevent the development of tumours in a variety of mouse and rat organs (Sharifi-Rad et al., 2020). In contrast, turmeric oleoresin administration to pigs has been linked to decreased food conversion efficiency (or weight gain), increased thyroid and liver weight, and histological changes in the kidney, liver, and urinary bladder (Tang et al., 2022). Additionally, the body weight gain, levels of red and white blood cells, and tissue weights were significantly affected in mice treated with large dosages of turmeric extract (Vo et al., 2021). Hence, these studies indicate the utility of curcumin /turmeric extracts as non-toxic and promising biologically active compounds at an appropriate dosage.

Biochemical activity of turmeric is attributed mainly to the total phenolic compounds present which includes polyphenols, tannins, and curcumin. Among these polyphenolic compounds, curcumin (1,7-bis-(4-hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5-dione) is of great importance because of its therapeutic or medicinal properties. Curcumin and its derivatives are ascribed with pharmacologically relevant anti-inflammatory, antifungal and antitumoral activities which adds to the therapeutic value of the crop (Ammon & Wahl, 1991; Oza et al., 2021; Tomeh et al., 2019) The anti-oxidant properties of curcuminoids are also widely studied and established (Jakubczyk et al., 2020; Selvam et al., 1995; Tanvir et al., 2017). As these therapeutic or pharmacological applications of turmeric are deeply intertwined with the biochemistry of the crop (especially the curcumin content) biochemical characterization hence becomes an inevitable practice in the turmeric crop improvement and breeding efforts. Many studies have explored the

biochemical attributes and variability in turmeric (Sahoo et al., 2017; Syamkumar, 2008; Thaikert & Paisooksantivatana, 2009).

Most of the studies exploring the biochemical variation among turmeric genotypes have focused on estimation of industrially relevant secondary metabolites like curcumin, essential oil content etc among genotypes for the identification of qualitatively superior genotypes (Singh et al., 2013). According to this study, the desired drug yielding qualitative traits are curcumin $> 5\%$, oleoresin $> 9\%$, leaf oil $> 0.5\%$ and rhizome oil $> 0.8\%$. This study also reported that among the genotypes under study, cv. Surama of Eastern ghat highland yielded high curcumin content (8.8%), cv. Lakadong of Eastern ghat highland and acc. Malkangiri of Southeastern ghat yielded high oleoresin content (15%) and also affirms the favourability of Eastern ghats and Southeastern ghats for turmeric cultivation (Singh et al., 2013).

This wide variation in biochemical composition along with yield exhibited by turmeric can be seen across genotypes and across agro-climatic conditions (Prasath et al., 2019). In other words, the variation observed in biochemical constituents (curcuminoids, oleoresin and essential oil) in turmeric can be attributed either to the genotype specific effect or to the environment effect i.e the changes in climate and soil conditions in different agro climatic zones or to the genotype and environment interaction effect (Anandaraj et al., 2014). In this study on environment-genotype interaction, IISR Kedaram performed consistently across different environments studied and was reported to be the most stable cultivar for curcumin production (Anandaraj et al., 2014). In another study, the total curcuminoid content (composed of curcumin (CUR), bisdemethoxycurcumin (BDMC) and demethoxycurcumin (DMC)) were analyzed in different genotypes of different agro-climatic origin and evaluated in different environments to understand the effect of environmental on these phytochemicals (Aarthi et al., 2020). The results revealed a higher curcumin content in IISR Prathibha that was comparable with curcumin content in SLP 389/1, Punjab Haldi 1, Duggirala Red, Rajendra Sonia, IISR Pragati, BSR 2 and CO-2. Thus, the remarkable effect of environment and its interaction with genotype

often result in different biochemical profile of the same cultivar in different environments limiting the commercial value of these cultivars and crop improvement efforts (Aarthi et al., 2020).

In another study (Kumari et al., 2022), the biochemical parameters like moisture content, total mineral content, carbohydrate content, crude fibre content, curcumin content, oleoresin content, and essential oil content were studied in the various varieties of turmeric (NDH-1, NDH-2, NDH-3, NDH-98, and IISR Prabha) that were obtained from the University of Agriculture and Technology Kumarganj in Ayodhya, Uttar Pradesh. The highest moisture concentration was recorded in NDH-1 (85.03%), with moisture contents ranging from 74.98% to 85.03%. The largest total mineral content was observed in NDH-1 (12 ± 0.34%), with total mineral content ranging from 7.01 ± 0.34 to12 ± 0.34%. The amount of crude fibre ranged between 4.64 and 5.65% and varied from variety to variety. NDH-2 has the greatest crude fibre content (5.6%) in fresh rhizome, followed by NDH-98 (5.6%), NDH-3 (4.75%), and NDH-1 (4.6%). IISR Prabha had the least amount of crude fibre (4.64%). The range of carbohydrate concentration was 61.50 to 70.40%; curcumin content was (2.22%) to (8.44%) and oleoresin ranged from 5.7% to 21.9%. NDH-2 had the highest carbohydrate content (70.40%), curcumin level (8.442%) and oleoresin content (21.9%). The percentage of essential oils varied from 1.63 to 3.91% and highest percentage of essential oils was observed in IISR Prabha (3.1%). In another study that screened the level of various primary phytochemical components of turmeric, the following were recorded; phenolic components (14.500 ± 0.70) mg/100gm, tannins (6.75 ± 360.75) mg/100gm, flavonoids (0.29 ± 8.873), curcumin in crude plant powder (2.490 ± 0.098 %), curcumin in etheric (1.150 ± 0.02 %) and ethanolic extract (24.680 ± 0.13) %, respectively (Essam & Ashraf, 2013). The following parameters were determined to be moisture (6.152 ± 0.27) %, total solid (93.848 ± 0.27) %, carbohydrate (36 ± 8.48) %, crude protein (7.737) %, total fat (2.637±0.22) %, total fibre (23.280 ± 1.23) % and ash (8.860 ± 0.03) %. With a significant number of vital elements (K, Mg, Na, Fe, Ca, Zn, Mn, Cr, Co, Ni and N), the energy value was determined to be 205.881 kcal/100gm (Essam & Ashraf, 2013).Thus, all these biochemical studies in turmeric help to gain

understanding about its biochemical profile as well as will aid in our efforts to improve the economically important biochemical traits in the crop (Kumari et al., 2022).

2.6 Crop improvement by biotechnology in turmeric

Crop improvement in turmeric have led to development of many improved varieties with desirable traits like increased curcumin and yield. Given the fact that India is the leading producer of turmeric, the average productivity and quality obtained is not up to the mark as most of crop improvement is by clonal selection (Ayer, 2017).This demands crop improvement efforts geared by latest biotechnological means to realise improved turmeric varieties in terms of yield, curcumin content and oleoresin. Some biotechnological practices like recombination by hybridization, micropropagation and invitro-pollination for production of disease-free plant materials, micro-rhizome production etc have resulted in an increased yield of turmeric crops which will help in production of disease-free seeds (Nirmal Babu et al., 2015). Some of these tissue culture techniques involving somatic embryogenesis, haploid technology, etc is important in production of disease-free planting material. Hence crop improvement efforts require a multifaceted approach both to improvement and selection of desirable traits as well as preservation of prevalent genetic diversity.

But the environmental influence on these characters renders them as unstable markers for marker-trait studies. However morphological characterization still has its prominence as they are basic features upon which any kind of selection operates (Ellegren & Galtier, 2016) . With the advent of molecular biotechnology, molecular characterization techniques have developed that enables us to assess the variation at DNA level and which is independent of environmental effects as well (Arif et al., 2010). So, both the phenotypic and molecular genotypic characterization reveals a fair understanding of diversity in the crop. This can be utilized in characterization of germplasm and the assessment of existing variability (Bhanu, 2017). It helps us to find if the genetic base is narrow or broad and to identify and develop improved variety as well as conserve the existing variation (Cooper et al., 2001). Molecular

markers like SSR, ISSR, RAPD which are stable and independent of environment is more reliable than morphological markers (Aswathi et al., 2023). Development of molecular markers as well as characterization of turmeric genotypes using them have provided insights into the genetic variation and relatedness among these genotypes (Siju, et al., 2010a).

2.6.1 Molecular studies in Turmeric

Molecular studies in turmeric mainly aimed to assess the existing genetic variation and to find distinct/divergent genotypes (Syamkumar, 2008). Most of the earliest works on molecular characterization of turmeric involves isozyme studies (Shamina et al., 1998) and RAPD/ISSR analysis of turmeric genotypes (Nayak et al., 2006; Syamkumar, 2008; Verma et al., 2015). These studies were mostly focused on assessment of genetic variation or relatedness among turmeric genotypes based on phylogenetic tree constructed from cluster analysis of molecular marker data (Singh et al., 2018). In one such study, the genetic diversity of turmeric accessions from 10 different agro-climatic regions were analyzed using RAPD and ISSR (Singh et al., 2012). Here, the dendrogram obtained based on marker data was used to study correlation between genetic similarity and geographical location as well as inferring genetic diversity based on the Nei's genetic diversity analysis. The Nei's genetic diversity (H) obtained from PopGene analysis reveals that the turmeric accessions from the Southeastern ghat and Western undulating zone exhibit relatively low genetic diversity, with values of 0.181 and 0.199 respectively . On the other hand, the Western central table land displays the highest genetic diversity, with a value of 0.257. Such information regarding the genetic diversity of turmeric in different agro-climatic regions based on marker data can be valuable for future breeding programs aimed at enhancing the production of curcumin, oleoresin, and essential oil (Singh et al., 2012). Another study has done a comparative analysis of clustering pattern of turmeric genotypes from morphological and molecular characterization involving RAPD markers (Vijayalatha & Chezhiyan, 2008). In this study, the genetic diversity of 30 turmeric accessions was evaluated using molecular techniques and compared to morphological traits to determine the extent of variation. The clustering patterns

obtained from quantitative data analysis using D^2, K means, and UPGMA methods showed some inconsistencies. However, when comparing the cluster profiles based on quantitative data and RAPD markers, there was a significant level of agreement between them. The analysis of RAPD profiles for the degree of divergence revealed 68.50% polymorphism across 21 primers. Among the primers investigated, OPB 08, OPC 20, OPE 09, and OPG 19 had a significant amount of polymorphism (> 90%). Based on this, they concluded that the observed disparities at both the morphological and molecular levels highlight the necessity for specific markers, both morphological and molecular, to accurately distinguish these turmeric accessions.

Although these phylogenetic trees/clustering based on RAPD/ISSR can group genotypes based on similarity, it cannot be effectively used in cultivar identification (Singh et al., 2012). Also, RAPD and ISSR markers are dominant markers with low reproducibility of banding pattern that makes them less informative and less reliable (Amom et al., 2020). Development and utilization of simple sequence repeat (SSR) markers helps to overcome the limitations of earlier marker systems. SSR markers are widely used in DNA profiling across species owing to their high polymorphism, co-dominant/multiallelic nature and better reproducibility (Vieira et al., 2016). Development and characterization of turmeric genotypes using SSR markers (genomic SSR and EST-SSR) have been reported in many studies mainly for assessment of genetic variation (Sigrist et al., 2010; Siju, et al., 2010a, b). In one such study, wherein a set of 18 genomic microsatellite markers was successfully developed from turmeric which was utilized to assess the genetic diversity of 20 turmeric accessions (Siju et al., 2010a). In this study, there were 103 alleles in total, with an average of 5.7 alleles per locus. The discriminating value of these markers ranged from 0.19 to 0.70, indicating that they were polymorphic to various degrees. These 20 turmeric accessions were divided into five main groups by UPGMA cluster analysis based on genetic distance values. The analysis of the samples revealed three sets of accessions that were genetically similar, raising the intriguing question of whether the technique for collecting germplasm based on vernacular identity should be revised. Interestingly, the observed grouping pattern was not influenced by the geographical origins of the accessions, indicating that the genetic

variation observed is independent of geographical origin of genotypes. The polymorphic SSR markers developed in this study hold great potential for population genetic studies and the management of turmeric germplasm. Another study from the same researcher, reiterates the usefulness of SSR markers in diversity studies in turmeric (Senan et al., 2013). Here, the study involved the isolation and characterisation of 21 polymorphic microsatellite loci which was used for screening thirty turmeric accessions. It was found that each locus has anywhere between two and eight alleles, with an average of 4.7 alleles per locus. These markers had an average discrimination power of 0.6 and a range of 0.25 to 0.67. These markers can be used to determine the polymorphism rate at specific loci in other turmeric genotypes and were recommended as highly informative for any future genetic investigative studies. In another study on the same line, genetic relatedness of some thirty turmeric genotypes was investigated using nine SSR markers (Singh et al., 2018). Out of these nine markers, six markers have been found very helpful as they produced higher levels of polymorphism and may be utilised to distinguish between the distinct turmeric genotypes. Here, the genotypes under study, were divided into two clusters with a significant amount of genetic diversity based on SSR based molecular analysis. Among the examined genotypes, NVST-80 and Pratibha, NVST-55 and GNT-2, and NVST-53 demonstrated 100% similarity, which implies identical genetic background for all the loci studied most possibly due to shared origin. Genotypes NVST-85 and NVST-70 had the lowest similarity level hinting to a greater genetic variation between them. Thus, all these studies conclude that for the genetic investigation of turmeric accessions, the simple sequence repeat (SSR) markers can supplement the molecular markers that are already in use. SSR utilized in majority of these studies fall into either genomic SSR (Sigrist et al., 2011; Siju et al., 2010a) or EST-SSR (Siju et al., 2010b). Genomic SSR isolated and developed based on whole genome library preparation can have origin in coding and non-coding regions of genome in contrast to EST-SSR that have origin only in coding region of genome (Siju et al., 2010b). But over the years, development and characterization of EST-SSRs have presented additional benefits over genomic SSRs (Joshi et al., 2010). EST-SSR development is much simplified due to the

accessibility to the expressed sequence data, which is easily available from public sources (Joshi et al., 2010). Moreover, EST-SSRs have the potential to increase the usefulness of DNA markers by capturing the variability within transcribed genes (Sahoo et al., 2017). In addition to development and characterization of EST-SSR for diversity screening studies (Siju et al., 2010b), studies have also developed and utilized distinct EST-SSR markers as a quick and effective means for establishment of cultivar identity (Sahoo et al., 2017). Due to their considerable agro-morphological similarities, turmeric cultivar identification is often difficult, which hinders commercialization. A previous study has attempted to address this problem by the molecular identification of turmeric genotypes (that included eight elite cultivars and 88 accessions) using EST-SSR markers (Sahoo et al., 2017). In this study, Eleven EST-SSR primers showed polymorphic banding patterns when evaluated against the eight cultivars (Suroma, Roma, Lakadong, Megha, IISR Alleppey Supreme, IISR Kedaram, Pratibha, and Suvarna). The polymorphic information content value (PIC) of these EST-SSR primers varied between 0.13 and 0.48. Notably, the 'Lakadong' and 'Suvarna' cultivars were successfully identified from others by three SSR loci (CSSR 14, CSSR 15, and CSSR 18), which consistently displayed distinctive banding patterns. These three SSR markers were effective in distinguishing the 'Lakadong' cultivars from the 88 accessions that were gathered from various agro-climatic regions. Moreover, the distinct phenotype allows for exact distinction of these two identified cultivars namely 'Lakadong' and 'Suvarna', through study and comparison with a phylogenetic tree encompassing 94 additional turmeric genotypes. Thus, this study highlights the general efficacy of EST-SSR markers in varietal identification based on the ability of these markers to distinguish between and authenticate the two economically important turmeric cultivars, 'Lakadong' and 'Suvarna' (Sahoo et al., 2017).

In general, it is evident that majority of the molecular studies in turmeric that focus on the development and characterization of molecular markers in turmeric genotypes deals with two prominent aspects – 1) validation of marker efficacy based on level of polymorphism and 2) elucidating the pattern of genetic variation among genotypes by clustering them. This grouping is done based on UPGMA cluster

analysis that relies on genetic distance values (Rohlf et al., 2009). These studies help to resolve genetic identity of genotype as marker-based characterization have resulted in determination of identical genetic backgrounds of morphologically diverse turmeric genotypes often with different vernacular identities in germplasm (Aswathi et al., 2023). Thus, molecular marker-based methods can aid in germplasm management and preservation of genetic resources. Another common finding from many studies is that the clustering pattern of turmeric genotypes was independent of their geographic origin in most cases (Verma et al., 2015).

Karyomorphological attributes were also studied with molecular marker analysis to find genetic distinctiveness among turmeric cultivars (Bhadra et al., 2018). Although this study, reports karyotypic symmetry with a uniform somatic chromosome number ($2n = 63$) across the studied cultivars, the molecular marker analysis done (RAPD and ISSR marker analyses) revealed significant genetic variability among the cultivars. Pattern of genetic variation revealed by Unweighted pair group method with arithmetic average (UPGMA) dendrogram as well as clustering using principal coordinate analysis on the basis of Jaccard's coefficient showed no geographical bias in this study. Thus, this study reveals the practical limitation of karyomorphological characters in cultivar identification alongside emphasizing the importance of molecular markers in determining the authenticity of cultivars (Bhadra et al., 2018).

2.6.1.1 Transcriptome analysis in turmeric

Molecular marker studies in turmeric received an impetus with many works on transcriptome sequencing and assembly (Annadurai et al., 2013; Sahoo et al., 2016a,b; Sheeja et al., 2015). Transcriptome sequencing have resulted in identification of potential genes involved in curcuminoid biosynthesis pathway (Annadurai et al., 2013; Sahoo et al., 2016a,b; Sheeja et al., 2015). The prominent approach followed for this is of comparative transcriptome profiling (Sheeja et al., 2015) in addition to annotation of *de novo* transcriptome assembly (Annadurai et al., 2013).The earliest of transcriptome profiling was done by a study (Sheeja et al., 2015) that led to discovery of novel polyketide synthase genes and transcription

factors involved in phenylpropanoid pathway using comparative transcriptome profiling of *C. longa* and *C. aromatica* which had a highly contrasting curcuminoid content. This study has also found many single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) within these potential genes. These results offer an invaluable resource for functional investigations, marker development, and expanded knowledge of turmeric's genomic structure and metabolic pathways. Transcriptome data was utilized to develop cultivar specific markers as well. In another such study, the transcriptome analysis of turmeric cultivars namely, Suroma, Roma, Lakadong, and Megha was done that resulted in identification of high-quality cultivar specific SSRs, SNPs, and pigment production gene transcripts through annotation and functional categorization (Sahoo et al., 2019). It also shed light on the cultivars' various genetic backgrounds by revealing the varying expression of pigment production genes among them. The results obtained in this study not only allow for the precise identification of elite cultivars that have a comparable appearance utilising cultivar specific SSRs and SNPs, but also help us to understand the molecular processes that underlie differential gene expression in turmeric.

2.6.1.2 Turmeric reference genome

The construction of a draft whole genome genome sequence of *C. longa* using advanced sequencing techniques has now opened a gateway to genetic information and molecular mechanisms of the crop (Chakraborty et al., 2021). This draft genome assembly was approximately 1.02 Gbp in size, with a significant portion consisting of repetitive sequences. It contained around 50,401 coding gene sequences. Through phylogenetic analysis involving 16 other plant species, the evolutionary position of *C. longa* was also determined by this study. Also, a comparative evolutionary analysis across 17 species, including *C. longa,* revealed significant evolutionary changes in genes related to secondary metabolism, plant phytohormone signalling, and stress response mechanisms (Chakraborty et al., 2021). Comparative genomic analysis revealed a recent whole-genome duplication event shared between turmeric and ginger These findings shed light on the molecular mechanisms responsible for the wide range of medicinal properties associated with *C. longa*, including its

defense mechanisms and ability to tolerate environmental stresses. Another research group has recently published a chromosomal-level genome assembly for turmeric with the aim of investigating the regulatory mechanisms of curcumin biosynthesis and tuber formation (Yin et al., 2022). They have utilized Pacbio long reads and Hi-C technologies to assemble the turmeric genome into 21 pseudochromosomes. The completed genome is 1.11 Gb in length, with a scaffold N50 of 50.12 Mb. It contains 49,612 protein-coding genes. Here, the analysis of gene families indicated the potential involvement of transcription factors, phytohormone signalling, and genes associated with plant-pathogen interactions in adaptation to challenging environments. In this study, candidate genes involved in phytohormone signalling and glucose metabolism were discovered through transcriptomic study of tubers at various developmental stages, suggesting that they may play a role in promoting tuber formation. The variation in curcumin levels between rhizomes and tubers is a result of secondary metabolites' adaptation to environmental stress, demonstrating that these compounds are involved in plant defence against abiotic stresses (Yin et al., 2022). Thus, the availability of the turmeric genome sequence improves our knowledge of how turmeric rhizome develop and how curcumin is made, as well as making it easier to study other Curcuma species.

2.6.2 Molecular marker-based trait improvement studies

Molecular marker-based trait enhancement studies are very limited in turmeric. Unavailability of reference genome for a long time may be one of the possible reasons. But still few studies can be found in the literature. One such study utilizes the chloroplast genome sequence to determine high curcumin containing turmeric genotypes (Hayakawa et al., 2011). In this study, they analysed the chloroplast DNA sequences to precisely identify *C. longa* lines with a high curcumin concentration. This was done by firstly creating a molecular phylogenetic tree of *C. longa* and allied species in order to choose appropriate outgroup taxa for infraspecific analysis. The results indicated that *C. aromatica* and *C. zedoaria* are the most closely related to *C. longa*. A molecular marker for identifying *C. longa* lines with high curcumin content was developed following a network analysis of chloroplast microsatellite

regions. A distinct haplotype within *C. longa* that corresponds to the line with high curcumin concentration was discovered through research. Thus, this study was able to recognise *C. longa* lines with a high curcumin concentration by using chloroplast microsatellite regions. Another study has developed molecular marker for disease resistance (Kar et al., 2013). The current cultivated turmeric germplasm is very susceptible to rhizome rot disease caused by the oomycete *Pythium aphanidermatum*. Traditional breeding methods for host resistance have not been effective due to limited genetic diversity, asexual reproduction, and stigmatic incompatibility. This study aimed to develop molecular markers for identifying new turmeric germplasm with resistance to *P. aphanidermatum* (Kar et al., 2013). By conducting bulk segregant analyses using 40 inter simple sequence markers, a potential resistance-specific marker named ClRSM was isolated in this study. This marker was then converted into a sequence tagged marker (ClSTS) that successfully amplified a specific fragment in resistant turmeric genotypes but not in susceptible plants. Southern blotting confirmed the presence of this marker as a single copy locus associated with resistant genotypes. The ClSTS marker correctly identified ten resistant and five susceptible genotypes from 15 turmeric germplasm samples with uncertain disease response from various areas. Subsequent inoculation tests with a virulent strain of *P. aphanidermatum* supported the findings of the STS marker. Thus, the identified STS marker of this study is a valuable tool for early and efficient identification of new turmeric cultivars resistant to rhizome rot using marker-assisted selection.

The emergence of high throughput Next generation sequencing (NGS) technologies has made the DNA based marker studies simpler and cost effective (Shen, 2019). Among the various NGS methods double digest restriction-site associated sequencing (ddRAD-seq) represents a versatile and economically efficient approach for gaining comprehensive understanding of the genetic makeup of any germplasm collection (Esposito et al., 2020). The dd-RAD sequencing technique coupled with high throughput NGS technologies (e.g., Illumina HiSeq 2000) and genotyping tools (e.g., GATK, samtools), allows for efficient simultaneous detection and genotyping of sequence variations, regardless of the availability of a reference genome (Peterson

et al., 2012) . In contrast to existing RADseq methods, dd-RAD sequencing offers increased adaptability and reliability in recovering target regions, along with a substantial decrease in costs, sample genomic material requirements and time (Peterson et al., 2012). DdRAD sequencing have been employed in many plant species to study genomic diversity, population structure, phylogenetic and demographic history of plant population (Aballay et al., 2021; Esposito et al., 2020; Liu et al., 2020; Yang et al., 2016). For instance, in a study carried out in *Brassica juncea* genotypes, a modified ddRAD sequencing method was adopted to partially sequence six *B. juncea* genotypes for SNP identification and genotyping (Sudan et al., 2019). In this study a tailored bioinformatics pipeline was used for SNP detection which were then utilized to assess molecular diversity, population structure, and conduct association mapping for morphological traits of interest.

Among the various molecular markers available, single nucleotide polymorphisms (SNPs) represent the most prevalent form of genomic variation, characterized by alternative alleles occurring at single base positions (Brookes, 1999). They are ubiquitous, primarily biallelic, and highly amenable to high-throughput automation (Brookes, 1999). These attributes had helped SNPs to emerge as key genetic markers, supplanting microsatellites, and have been instrumental in genetic marker-assisted breeding across economically significant species (Rafalski, 2002). Their adoption has facilitated significant time and cost savings in breeding programs (Rafalski, 2002). Notably, SNPs offer advantages in kinship analysis and pedigree reconstruction, owing to their suitability for high-throughput genotyping methodologies (Moragues et al., 2010). Sequencing data obtained from high-throughput genotyping technologies are aligned to a reference genome if its available or to a *de novo* reference in the absence of reference genome for discovery of SNP variants (Yao et al., 2020). Among the various alignment tools available, previous studies revealed that the BWA-mem mapping tool is a good choice as an alignment tool as it exhibited superior performance compared to other tools like Bowtie2, demonstrating both a higher mapping rate and accuracy rate (Yao et al., 2020). The Genome Analysis Toolkit (GATK) is a widely used tool for variant calling of single nucleotide polymorphisms (SNPs) and small insertions and

deletions (indels) from short-read sequencing data that is aligned against a reference genome (Warden et al., 2014). While Samtools mpileup (now Bcftools mpileup) was previously widely used, it has been surpassed by GATK HaplotypeCaller, primarily due to the latter's faster performance, particularly for large sample sizes (Lefouili & Nam, 2022). GATK HaplotypeCaller is widely recognized as the preferred option for variant calling, noted for its high accuracy in SNP and indel identification (Lefouili & Nam, 2022). Numerous benchmark studies have consistently found GATK HaplotypeCaller to outperform Bcftools solidifying its status as the current gold standard for variant-calling pipelines (Lefouili & Nam, 2022). These SNP variants identified by variant calling pipelines is the basis for genetic diversity, population structure analysis marker trait association studies etc (Rafalski, 2002).

Marker trait association studies is a pre-requisite for marker assisted breeding programs in any crop. Association analysis, or association mapping, has become a potent tool in plant genomics, aided by advancements in statistical software and high-throughput molecular marker technologies (Zhu et al., 2008). Genome-wide association studies (GWAS) leveraging genome-wide markers have proven highly effective in unravelling the genetic architecture of complex traits across plant species (Padmashree et al., 2023; Peng et al., 2023). Studies combining sequencing data from high-throughput methods like ddRAD sequencing with agro-morphological trait information have provided valuable insights into genetic diversity and trait-associated markers (Esposito et al., 2020; Han et al., 2023). The GAPIT (Genome Association and Prediction Integrated Tool) R package is a useful package for conducting Genome-Wide Association Studies (GWAS) and genome prediction (Ksouri et al., 2023; Tandayu et al., 2022). Among the various models that can be utilized by GAPIT, BLINK emerges as the preferred model, recognized for its robust calibration, effectively managing false positive and false negative effects (Ksouri et al., 2023). Consequently, BLINK is deemed as one of the most suitable models, exhibiting optimal adjustment with phenotypic data (Ksouri et al., 2023). Henceforth, the GWAS results derived from BLINK provides a reliable basis for further analysis. The BLINK model, employed in GWAS studies, harnesses linkage disequilibrium to enhance statistical power, and the maximum likelihood is

estimated using Bayesian Information Content (BIC) (Han et al., 2023). There have been studies utilizing genotypic data from ddRAD sequencing for marker trait studies. For instance, a study in *Ceiba* plants utilized ddRAD sequencing for GWAS studies to identify SNPs linked to morphological traits of interest and the candidate genes involved were found thereby providing a basis for marker assisted selection and molecular breeding in *Ceiba* plants (Han et al., 2023). Similarly, in another study, ddRAD sequencing was utilized to identify genomic regions associated with traits of interest, including grain zinc (Zn) content which can be further exploited for Zn biofortification breeding programs in rice (Babu et al., 2020). Through this research, the scientists uncovered genomic regions responsive to traits of interest and identified candidate genes associated with these key physiological traits, thereby laying the groundwork for marker-assisted selection and the development of fortified rice varieties (Babu et al., 2020). In summary, ddRAD sequencing have emerged as a promising tool for advancing our comprehension of plant genetics and streamlining crop improvement endeavours through targeted genomic analyses. Though one of the earlier studies (Liang et al., 2021) have recommended the utilization of reduced representation sequencing methods like ddRAD to understand phylogeny and evolution of curcuma species but there has been no such study done till date. Thus, this doctoral study is aimed to fill in this gap of genetic diversity and marker trait association assessment in turmeric.

Chapter 3

MATERIALS AND METHODS

3.1 Field experiment: Experimental site

The field experiments pertaining to this study was carried out at the experimental farm of ICAR-Indian Institute of Spices Research (ICAR-IISR) in Peruvannamuzhi, Kozhikode, Kerala (North latitude 11°36'34"; East longitude 75°49'12"; 60 m MSL). Genotypes utilized in this study were obtained from National Active Germplasm Site (NAGS) of ICAR-IISR which has India's biggest and most significant turmeric germplasm collections.

3.1.1 Climate

The experimental site falls in the tropical region and hence under the tropical monsoon climate. Tropical monsoon climate is characterized by seven months of rain (mean 4374.0 mm) from May to December and monthly mean temperatures above 18 °C (64 °F) in every month of the year and a dry season. The climate is characterized by a relative humidity range of 75.0-90.0% and high temperatures in dry seasons that can go up to 35 °C which usually lasts from December to April. Weather data is given in Appendix 1.

3.1.2 Soil

Soil here is a clay loam Ustic Humitropept and the soil profile is given in Table 3.1.

Table 3.1 Soil profile of the experimental site

Soil Parameters	Values
pH	5.18
Electrical conductivity	0.13 dS m^{-1}
Organic carbon	13.9 g kg^{-1}
Mineral nitrogen	125 mg kg^{-1}
Bray phosphorous	11.6 mg kg^{-1}
Exchangeable potassium	176 mg kg^{-1}
Bulk density	1.48 mg m^{-3}
Particle density	2.6 mg m^{-3}
Water holding capacity	0.74 mm cm^{-1}

3.2 Experiment details

The field experiment began in May 2020 and was repeated in the next year; 2021. The field experiments were conducted at different plots in the same site to avoid nutrient depletion and incidence of soil borne diseases.

3.2.1 Land preparation, experiment design and maintenance

Typically, turmeric is cultivated in rainfed settings on raised beds. Here, weeds were removed from the plot to be planted, and the soil was fine-tilled with a tractor-mounted disc harrow before being levelled. Due to the soil's inherent acidity, lime ($CaCO_3$) was added at a rate of 500 kg per hectare, and the soil was thoroughly mixed before being levelled. Then, using a garden spade, raised beds of 3 x 1 x 0.30 m (l x b x h) were created. The beds were kept 40 cm apart from one another. On the beds, little shallow trenches with a spacing of 30-20 cm were dug, and in these pits, 20–30 g of turmeric seed-rhizome with at least two sprouted buds was planted at a depth of 3.5–5.0 cm from soil surface. Genotypes were planted in a randomised blocks design with two replicates and field was maintained with recommended package of practices (Prasath et al., 2019). This includes mulching, weeding, fertilizer application, etc. Mulching was done using leaves of *Gliricidia sepium* (Jacq.) Kunth ex Walp at 15 t ha^{-1} immediately after planting to prevent exposure of planted rhizomes and shield the raised beds from heavy rains thereby preventing soil erosion. Following weeding, fertiliser application, and earthing up, green leaf mulching was applied again at 7.5 t ha^{-1} at 45 and 90 DAP.

Fig. 3.1. Land preparation, planting, weeding, and mulching.

Fig. 3.2. Turmeric plants after 4 months of planting

Fig. 3.3. Fully grown turmeric plants after six months of planting.

3.3 Plant materials

For this study, a diversity panel comprising 93 turmeric genotypes from germplasm were used. The phenotypic evaluation of 93 turmeric genotypes was conducted for two years (2020 and 2021). The list of genotypes is given in Table 3.2.

Fig. 3.4. The characteristics of core collection of turmeric utilized in this study.

Table 3.2 List of turmeric genotypes used in this study

S. No	Genotype name	Category
1	Varna	
2	Megha Turmeric	Released varieties
3	Suranjana	
4	Suguna	
5	Kedaram	
6	Rajendra Sonia	
7	Suvarna	
8	IISR Pragati	
9	IISR Prabha	
10	Sudharsana	
11	IISR Alleppey Supreme	
12	IISR Pratibha	
13	Punjab Haldi 1	
14	BSR2	

15	CO 1	Released varieties
16	CO 2	
17	Narendra Haldi 1	
18	Narendra Haldi 3	
19	Rajapuri	
20	Roma	
21	Suroma	
22	Rasmi	
23	Ranga	
24	Duggirala red	
25	Palam Pitambar	
26	Kanthi	
27	Sobha	
28	Sona	
29	Pant Peetabh	
30	Punjab Haldi 2	
31	CIM Pitambar	
32	Rajendra Sonali	
33	NDH 98	
34	Kadappa local	
35	Erode turmeric	Local varieties
36	Waigon turmeric	
37	Salem local	
38	Santra	
39	Kandaila Haldi	
40	Dehati Haldi	
41	Futi Halood	Farmers varieties
42	Surkha	
43	Hardi	
44	Koirana	
45	Hasgova	
46	Acc 2	
47	Acc 23	
48	Acc 224	
49	Acc 246	
50	Acc 780	Promising accessions
51	Acc 799	
52	Acc 821	
53	Acc 880	
54	Acc 849	
55	Acc 884	
56	Acc 887	
57	Acc 449	
58	Acc 300	
59	Acc 415	

60	IISR RRN1	Breeding materials
61	IISR RRN2	
62	IISR RRN3	
63	IISR RRN4	
64	SLP 389	
65	SL 3	
66	KTS 611	
67	SC 61	
68	Acc 19	
69	Acc 138	
70	Acc 1046- Cambodia	Exotic collections
71	Acc 1053 - Uganda	
72	Acc 1054 - Uganda	
73	Nepal Turmeric	
74	Mydukkur White	Distinct types
76	Acc 37	
77	Acc 286	
78	Acc 297	
79	Acc 52	
80	Acc 161	
81	Acc 84	
82	Acc 313	
83	Acc 180	
84	Acc 79	Nematode tolerant genotypes
85	Acc 142	
86	Acc 200	
87	Acc 376	
88	Acc 1	High curcumin genotypes
89	Acc 146	
90	Acc 8	Low curcumin genotypes
91	Acc 130	
92	Acc 134	
93	Acc 902	

3.4 Phenotypic evaluation

Phenotyping of 93 turmeric genotypes was conducted based on general plant morphology, rhizome characters, and biochemical character i.e curcuminoid content. General quantitative morphological characters were recorded 150 days after planting based on DUS guidelines (PPV & FRA, 2009; Aarthi et al., 2018a). The general plant morphological characters observed are plant height (PH), the number of shoots

(NS), the number of leaves on the main shoot (NL), the total number of leaves per plant (TL), the length of the petiole (LP), the length of the leaf lamina (LL) and the width of the leaf lamina (LW). Rhizome characterization was recorded for the following characters after harvesting: number, length, thickness, weight per plant of mother rhizome (MN, ML, MG, MWT), primary rhizome (PN, PL, PG, PWT), secondary rhizome (SN, SL, SG, SWT) respectively, total rhizome weight per plant (TWT), internodal length (IL), inner core width/diameter (ID) and outer core width (OW) of primary rhizome, dry rhizome yield per plant (DY).

3.4.1 General plant morphological characterization

3.4.1.1 Plant height (cm)

The measurement of plant height was determined by measuring from the soil level up to the tip of the leaf on the primary shoot. A total of five clumps were selected from each replication, and an average height was calculated.

3.4.1.2 Number of shoots per plant

To determine the number of shoots per plant, five clumps were randomly chosen from each replication, and the average number of shoots was counted.

3.4.1.3 Number of leaves on main shoot

The number of leaves on the main shoot was assessed by selecting five clumps from each replication and calculating the average number of leaves present on the main shoot.

3.4.1.4 Total number of leaves per plant

The total number of leaves was assessed by selecting five clumps from each replication and calculating the average number of total leaves present on all shoots.

3.4.1.5 Length of petiole (cm)

The measurement of leaf petiole length was taken from the pseudostem to the base of the leaf blade, specifically from the middle three leaves of the main shoot. Five

clumps were randomly selected from each replication, and the average petiole length was calculated based on these measurements.

3.4.1.6 Leaf length (cm)

The measurement of leaf lamina length was taken from the tip of the petiole to the tip of the leaf blade, specifically from the middle three leaves of the main shoot. A total of five clumps were randomly selected from each replication, and the average lamina length was calculated.

3.4.1.7 Leaf width (cm)

To determine leaf lamina width, measurements were taken at the widest point of the middle three leaves on the main shoot. Five clumps were selected from each replication, and the average lamina width was recorded.

3.4.2 Rhizome characterization

3.4.2.1 Number of mother rhizomes per plant

The number of mother rhizomes per clump was counted. A total of five clumps were assessed in each replication, and the average number of mother rhizomes was recorded.

3.4.2.2 Length of mother rhizomes per plant (cm)

To measure the length of a mother rhizome, a measuring tape or ruler was used to determine the distance from the point of attachment to the main plant up to the tip of the mother rhizome. This measurement represented the overall length of the mother rhizome.

3.4.2.3 Girth of mother rhizomes per plant (cm)

To measure the girth of a mother rhizome, a calliper or flexible measuring tape was employed. The measuring device was wrapped around the widest section of the mother rhizome, ensuring a snug fit without excessive pressure. The measurement was recorded as the girth of the mother rhizome.

3.4.2.4 Number of primary rhizomes per plant

To measure the number of primary rhizomes, individual rhizomes present in each clump were counted. This process was repeated for multiple clumps to obtain an average number of primary rhizomes per clump. The total number of primary rhizomes per clump found in each replication was recorded.

3.4.2.5 Length of primary rhizomes per plant (cm)

The length of the primary rhizomes was measured from the point of attachment to the mother rhizome up to the tip. This measurement was taken for each individual rhizome, and the recorded values represented the length of the primary rhizomes.

3.4.2.6 Girth of primary rhizomes per plant (cm)

To measure the girth of a primary rhizome, a calliper or a flexible measuring tape was used. The measuring device was wrapped around the widest part of the primary rhizome, ensuring it was snug but not overly tight. The measurement was taken, and the girth of the primary rhizome was recorded. This process was repeated for each primary rhizome to ensure accurate and consistent measurements.

3.4.2.7 Number of secondary rhizomes per plant

The number of secondary rhizomes per clump was counted. Multiple clumps were assessed, and the average number of secondary rhizomes per clump was recorded.

3.4.2.8 Length of secondary rhizomes per plant (cm)

The length of the secondary rhizomes was measured from the point of attachment to the primary rhizome up to the tip. Each individual rhizome was measured, and the recorded values represented the length of the secondary rhizomes.

3.4.2.9 Girth of secondary rhizomes per plant (cm)

To measure the girth of a secondary rhizome, a flexible measuring tape was utilized. The measuring device was wrapped around the widest part of the secondary

rhizome, ensuring a snug fit without excessive pressure. The measurement was taken, and the girth of the secondary rhizome was recorded. This process was repeated for each secondary rhizome to ensure accurate and consistent measurements.

3.4.2.10 Internodal length (cm)

The internodal length was measured between two consecutive nodes on the primary rhizome. Measurements were taken for multiple internodes, and the average internodal length was recorded.

3.4.2.11 Inner core diameter of primary rhizome (cm)

The inner core width or diameter of the primary rhizome was measured using a calliper or a measuring device (scale). Primary rhizome was cut, and width or diameter of core measured using a scale and recorded. This process was repeated for multiple primary rhizomes, and the average inner core width or diameter was calculated.

3.4.2.12 Outer core width of primary rhizome (cm)

The outer core width of the primary rhizome was measured using a calliper or a measuring device. Primary rhizome was cut, and width of outer core measured using a scale and recorded This process was repeated for multiple primary rhizomes, and the average outer core width was calculated.

3.4.3 Yield related characters

3.4.3.1 Total weight of primary rhizome per plant (g)

The total weight of the primary rhizome per plant was determined by collecting all primary rhizomes from a clump and measuring their combined weight. This was repeated for five clumps each for two replications.

3.4.3.2 Total weight of mother rhizome per plant (g)

The total weight of the mother rhizome per plant was determined by collecting all mother rhizomes from a clump and measuring their combined weight. This was repeated for five clumps each for two replications.

3.4.3.3 Total weight of secondary rhizome per plant (g)

The total weight of the secondary rhizome per plant was determined by collecting all secondary rhizomes from a clump and measuring their combined weight. This was repeated for five clumps each for two replications.

3.4.3.4 Total weight of rhizome per plant (g)

The total weight of the rhizome per plant was determined by weighing the total clump that belongs to a plant. This was repeated for five clumps each for two replications.

3.4.3.5 Dry weight of rhizome per plant (g)

The dry weight of the rhizome per plant was determined by collecting whole rhizomes (clump) from each plant, boiled and sun dried to remove moisture, and then recording their individual weight. This was repeated for five clumps each for two replications.

3.4.4 Biochemical characterization

The percentage of curcuminoid content (CUR) was examined in all 93 varieties using ASTA method (ASTA, 2004):

- 1 g of dried turmeric powder was weighed and put into a 125 ml Erlenmeyer flask, added approximately 75 ml of acetone was added and the mixture was stirred.
- The mixture was gently refluxed on a heated stirring plate with a West Condenser for one hour. Afterward, it was cooled to room temperature and filtered volumetrically using a Whatman filter paper into a 200 ml capacity

flask. It was thoroughly washed with acetone and adjusted to a final volume of 200 ml.

- 1 ml of this solution was then transferred into a 100 ml volumetric flask, and then the volume was adjusted to 100 ml with acetone, and thoroughly mixed.
- Absorbance reading of above solution was determined using spectrophotometer. For that, the spectrophotometer was first blanked with acetone in the cuvette at 425 nm, and then the absorbance of the above solution at 425 nm was determined. The percentage of curcuminoids determined by the below formula:

% of curcuminoids = absorbance at 425 nm x 11.11

3.5 Determination of ploidy level using flow cytometry

Samples were analyzed for ploidy levels using flow cytometry (Prasath et al., 2022). Here, turmeric leaf tissue was used for ploidy estimation. FCM based ploidy estimation of a sample from a species requires a reference standard of known ploidy level from the same species. IISR Prabha and IISR Kedaram with known ploidy level i.e. triploids were used as reference standards. The protocol followed is given below:

- A small amount of tender leaf tissue was taken from the reference standard plant as well as from genotype of unknown ploidy level.
- The leaves were finely chopped with scissors into to 1 ml hypotonic propidium iodide lysis buffer (Krishan, 1975) with minor modification.

Composition of nuclei isolation buffer used:

Constituent	Concentration
Sodium citrate tribasic dehydrate	0.1 % (w/v)
RNase A	2 mg/ml
PI	0.05 mg/ml
Tween-20	0.3 % (v/v).
β -mercaptoethanol	1%

- After the chopping of leaf tissue into buffer, it was filtered through a 10-µm cell strainer to obtain nuclear suspension which was collected in 1.5 ml eppendorf tubes. This suspension of isolated nuclei was incubated for 15 min.
- Nuclear suspension from reference standards and unknown samples were run on a Cytoflex flow cytometer (Beckman Coulter) with a 488 nm laser and a 585/42 Band Pass filter. For each sample, around 500-2000 nuclei were gathered and analyzed.
- Reference standard was used to set the detector gain, and the other unknown samples were analyzed using the same gain settings.
- The analysis results are given as one-parameter histograms with G_0/G_1 peaks plotted on a linear scale. The median value of G_0/G_1 gives the fluorescence intensity and it was obtained for both reference standard and unknown sample.
-
- The data obtained was then analyzed in FCS Express Software (DENOVO software, USA) and the ploidy level of unknown sample was calculated based on following formula:

$$\text{Ploidy of unknown sample} = \frac{\text{Ploidy of reference standard} \times \text{median value } G_0/G_1 \text{ peak of unknown sample}}{\text{median value of } G_0/G_1 \text{peak of reference standard}}.$$

3.6 Statistical analysis

3.6.1 Analysis of variance (ANOVA)

Statistical calculations were performed on the pooled morphological and biochemical (curcuminoid content) data of two years. A one-way analysis of variance (ANOVA) to test for significant differences among the 93 varieties for the characters under study evaluated using at $p \leq 0.05$ and 0.01. Analysis of variance (ANOVA) for RBD was done as per previous study (Panse et al., 1954).

3.6.2 Duncan multiple range test (DMRT)

When ANOVA was significant, Duncan's mean range test (DMRT) was employed to specific varieties with different mean values (Duncan, 1955). It is a post hoc test, a commonly used to compare treatment means.

3.6.3 T test

T test is widely employed statistical test to compare the means of two groups (Student, 1908). It is frequently employed in hypothesis testing to establish whether a procedure or treatment truly affects the population of interest or whether two groups differ from one another. T test was used to determine whether there is any difference between the observed ploidy levels, triploids, and tetraploids, for the 25 characters under examination.

3.6.4 Principal component analysis (PCA)

Principal component analysis is a dimensionality reduction method used for analysis of large complex datasets (Hotelling, 1933). It improves interpretability of the dataset as well as minimise the information loss. It accomplishes this by the means of creation of new and uncorrelated variables called principal co-ordinates that maximise variance one after the other. Here, we have utilized PCA analysis to determine the pattern of genetic variation between the ploidy levels based on their phenotypic characters. PCA analysis was performed in the R programming language.

3.6.5 Estimation of genetic parameters

The variability package in R software was used to analyse genetic variability, which included determining GCV, PCV, heritability, and genetic advance.

3.6.6 Phenotypic and genotypic variance

The genotypic and phenotypic variances were estimated according to following equation (Johnson et al., 1955) :

$$\text{Genotypic variance } (\sigma^2_g) = \frac{\text{Mean sum of squares due to genotypes} - \text{Error mean sum of squares}}{\text{Replication}}$$

Phenotypic variance (σ^2_p) = σ^2_g + σ^2_e

Where Environmental variance (σ^2_e) = Error mean sum of squares / Replication

3.6.7 Phenotypic and genotypic coefficients of variation (%)

Phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) is expressed in percentage and is given by the following equation (Burton & DeVane, 1953) :

a. PCV = ($\sqrt{V_P}$ / Mean) * 100

b. GCV = ($\sqrt{V_G}$/Mean) * 100

Where V_P is phenotypic variance and V_G is genotypic variance.

The PCV and GCV values were categorized as per (Sivasubramanian & Menon, 1973) as follows:

PCV and GCV	Category
< 10 per cent	Low
10 – 20 per cent	Moderate
> 20 per cent	High

3.6.8 Heritability (h^2)

Heritability (h^2) estimate in broad sense is expressed in percentage was calculated based on the methods of (Lush & others, 1949) and (Allard, 1960).

Heritability (h^2) = (V_G/V_P) x 100

Heritability (%) was categorized as suggested by (Johnson et al., 1955).

Heritability in per cent	Category
< 30	Low
31 – 60	Medium
> 60	High

3.6.9 Genetic advance (GA)

Genetic advance was estimated based on the formula given by (Johnson et al., 1955) and is as follows:

Genetic advance = h^2 x σ_P x k

where,

h^2 = Heritability in broad sense

σ_P = Phenotypic standard deviation

k = 2.06, Selection differential (at 5% selection intensity) (Falconer, 1960)

3.6.10 Genetic advance as per cent of mean (GAM)

Genetic advance as per cent of mean was calculated using the following equation:

Genetic advance (as per cent of mean) = (Genetic advance / Grand mean) * 100

Categorization of GAM was done based on range specified by (Johnson et al., 1955) and is as follows:

GAM (%) value	Category
< 10 per cent	Low
10 - 20 per cent	Moderate
> 20 per cent	High

3.6.11 Heritability in broad sense

Heritability in broad sense (h^2 (b)) was calculated using the formula given by (Allard, 1960).

$$\text{Heritability in broad sense } \left(h^2(b)\right) = \frac{\text{Genotypic variance}}{\text{Phenotypic variance}} \times 100$$

The classification of heritability in broad sense was done using ranges given by (Johnson et al., 1955)

h^2 (b) value	Category
0-30%	Low
31 – 60%	Medium
> 61%	High

3.7 Genotypic analysis

3.7.1. Genomic DNA isolation

All the 93 genotypes were genotyped using SSR markers. For this, genomic DNA was isolated from leaf tissue of the 93 turmeric genotypes using CTAB method (Doyle & Doyle, 1987) with minor modification as follows:

- 100 mg of fresh, clean, and young leaf of each genotype was weighed and each sample was grinded well in 1.2 ml CTAB buffer using pestle and mortar.
- Each of this grinded sample solution was transferred to a 2ml eppendorf tube and incubated in water bath at 65°C for 60 min. After incubation, sample were allowed to cool and bought to room temperature.
- 800 µl of chloroform-iso amyl alcohol solution prepared in 24: 1 composition was added to each of these tubes. It was mixed gently by swinging the eppendorf tubes and then centrifuged at 12,000 rpm for 15 min.
- The clear supernatant (aqueous layer) obtained after centrifugation was carefully transferred to a 1.5 ml eppendorf tube and chloroform isoamyl extraction step and centrifugation was repeated once again.
- The supernatant obtained after the second centrifugation was again transferred to 1.5 ml eppendorf tube and 500 µl of isopropanol was added to it. This was kept for overnight at 4°C.
- After overnight incubation, the solution was centrifuged at 12000 rpm for 10 min at 4°C to pellet down the precipitated DNA.
- After centrifugation, supernatant was removed and 100 µl of 70% ethanol was added to the tube to wash the precipitate.

- This was again centrifuged at 10,000 rpm for 5 min. supernatant obtained was carefully removed and DNA pellet was allowed to air dry in laminar air flow.
- The DNA obtained thus was dissolved in 30 µl of nuclease-free water. DNA concentration and quality were determined using the DeNovix DS-11 spectrophotometer (DeNovix, Wilmington, DE, USA) followed by agarose gel electrophoresis on 0.8% agarose gel.
- The genomic DNA was then diluted to 50 ng/µl and stored at − 20°C for downstream PCR reactions.

The released turmeric varieties were screened, first, with 56 molecular markers including both SSR and ISSR to check for the efficiency of already developed markers in turmeric. (Table 3.3a and 3.3b). Here, we have utilized SSR markers derived from genomic regions of turmeric as well as the EST derived markers (Siju et al., 2010a,b). This molecular characterization has confirmed the reliability of these markers to find genetic similarity. The most reliable and polymorphic markers obtained based on this were further used to screen 93 turmeric genotypes. A UPGMA based dendrogram was constructed based on these screening on 93 genotypes.

Table 3.3a. Characteristics of ISSR Primers used in the study

Primer	Sequence FP -Forward primer (5'- 3') RP -Reverse primer (5'-3')	Annealing temperature T_a(°C)
UBC 815	CTCTCTCTCTCTCTCTG	43.7
UBC 818	CACACACACACACACAG	45.3
UBC 826	ACACACACACACACACC	45.3
UBC 827	ACACACACACACACACG	46.5
UBC 834	$(CA)_8$AG	45.3
UBC 842	GAGAGAGAGAGAGAGAYG	50.4
UBC 845	CTCTCTCTCTCTCTCTRG	49.1
UBC 850	GTGTGTGTGTGTGTGTCA	45.1
UBC 856	ACACACACACACACACCA	52.7
UBC 857	ACACACACACACACACYG	54.8
UBC 860	TGTGTGTGTGTGTGTGRA	66.7
UBC 866	CTCCTCCTCCTCCTCCTC	50.4
UBC 884	HBHAGAGAGAGAGAGAG	50.4
UBC 889	ACACACACACACACACDBD	42.3
UBC 896	AGGTCGCGGCCGCNNNNNNATG	50.4
UBC 897	CCGACTCGAGNNNNNNATGTGG	53.8
ISSR-02	AGTGAGTGAGTG	55.0
ISSR-03	$(GACA)_4$	43.7
ISSR-06	GAGAGAGAGAGAGAGAC	51.7
ISSR-07	GAGAGAGAGAGAGAGAG	52.7
ISSR 13	AGTGAGTGAGTGGG	43.7
ISSR 14	AGCAGCAGCAGCGT	44.2
ISSR 15	TCCTCCTCCTCCTCC	52.1
ISSR 17	CACACACACACACACAG	51.7

Table 3.3 b Characteristics of SSR Primers used in the study

Primer	Sequence FP -Forward primer (5'- 3') RP -Reverse primer (5'-3')	Annealing temperature T_a(°C)
Cumisat 01	FP-AAACCGCAAGAAAACTGAAG RP-CTCTTCCCTGAACGATTCC	55.0
Cumisat 02	FP-TATGTGATGGTTGGGACG RP-GTAGTGGAGGAAGACGCC	55.8
Cumisat 03	FP-GCACTACTTCCTTCTCGTTCAA RP-CGTCGTAAAGATTAGCGTGTG	60.4
Cumisat 04	FP-TCAGGTTTCAGGGTGTAGAAG RP-CCCAGCAAGATTTTACCAAG	55.8
Cumisat 05	FP-AGCAGTGCGTCTTTCATC RP-CTCTTGTCACGGAACCTC	57.8
Cumisat 08	FP-CATTGCGTGCCCACTTCC RP-CCTCCCTGTCGCTCTCCTC	60.4
Cumisat 13	FP-CCCGAAGCCATTTCTCAG RP-TCGTCTCTCCTCTGCCAAC	55.0
Cumisat19	FP-CATGCAAATGGAAATTGACAC RP-TGATAAATTGACACATGGCAGTC	55.0
Cumisat 20	FP-CGATACGAGTCCATCTCTTCG RP-CCTTGCTTTGGTGGCTAGAG	62.9
Cumisat 22	FP-AATTTATTAGCCCGGACCA RP-AAGAAAGTGAGTAGAAACCAAAGC	57.8
Cumisat 23	FP-CGTGGAAGGTGAGTTTGAC RP-CAGAAGGGAACTGAGATGG	57.8
Cumisat 25	FP-TACATGAGAAACAACAAAGCCC RP-AGTTAGCCAAGTCCCAATTTAGC	55.8
Cumisat 28	FP-TTCAACTTCTCCTCGCTCAG RP-GCAAGGTCTGCATCTATTTCTC	55.0
Cumisat 29	FP-GTGGTATCCCCATGAAGAGC RP-ATGACCAAGCCCTTTCACC	57.8
Cumisat 31	FP-GGAGGAGGAGAAGCAGAAG RP-GACAGGCGAAGGAAGAAAC	62.9
Cumisat 32	FP-TGTTGTAGGTAGAAGCAAATGAC RP-TTGGTGTCCTAATTCTTTCAAC	55.0
Cumisat 33	FP-ATGGATGGATACAACAACAAC RP-TATAAACACACTCCCTCTTGG	57.8
Cumisat 36	FP-TGGGCTCAATGGTTGATACG	55.0

	RP-CTCCTCATCGCTATCCGAGG	
Cumisat 37	FP-CCATTGGCGAGGATGAAGC RP-CCTGCCAAGCAAAGCCAAG	55.8
Clest 02	FP-ACCGTAGCAAAGAAATAGGAC RP-AAGGTGGAAGGAAACTCG	55.0
Clest 03	FP-AGGGAAAATAGAGTAGGCAAG RP-TGAAGGATTACAGTCAGCAAA	57.8
Clest 04	FP-ACACAACATTCAGTTTAGCAC RP-TCCCTATTCTTTCCTCTCG	55.0
Clest 06	FP-TCATCGTCTGCTTTAGTTTTC RP-ACGCTCTGCTCCTTCAAC	55.0
Clest 08	FP-GATGCACACATTGCCCGTG RP-GGGTGCAATTCTTGGTCCG	55.0
Clest 09	FP-TCGGTTCTACTGAATCTTTACTCG RP-AGACTGTTTTCCCATTGTTGC	60.4
Clest10	FP-GTGGTGGAGGAGGAAGAGAAG RP-TTGAGGGAACAAAAGGAAGAC	60.4
Clest 11	FP-TTCATTCGACGCAAACAGC RP-CGACGCAATAGTCGAAGGC	55.0
Clest 12	FP-GGGATTGAGGTGGAGGTAGG RP-GCTGGCGAAGTAGAAGAAGAAG	57.8
Clest 13	FP-TGTACAAGCTCCAAATAAGTCAAG RP-CAGGAGTGTTCTAATGTTGCCC	64.6
Clest 15	FP-GCCAAAGAAAGAACTGACATCC RP-TTACAACCCTCCTCCCATTAGA	55.0
Clest 16	FP-AAGCAGTCCGTGGGAGAAG RP-CTTCCTCAATCGAATGGCCG	64.6
Clest 17	FP-GTGCCTGTGGACCTATCCG RP-GAAGCATGCGAATTCATCTAAAC	55.0

3.7.2 PCR analysis

Both SSR and ISSR molecular markers were utilized for PCR based molecular characterization.

PCR reaction components and reaction set up

PCR component	:	Volume (µl)
Nuclease free water	:	8
Master mix	:	10
Primer	:	1
Template DNA	:	1

Total volume of PCR product made up to 20 µl.

3.8 ddRAD based genotyping

Since the available set of markers are limited, the ddRAD sequenced data developed in turmeric from ICAR-IISR was used to find SNP markers. Double digest restriction site associated DNA sequencing (ddRAD) is a next-generation sequencing-based method that permits high throughput simultaneous discovery and genotyping of SNPs, either with or without an existing reference genome (Peterson et al., 2012). Single Nucleotide Polymorphism (SNPs) so obtained are utilized as genetic markers and are amenable to high-throughput assays (Trebbi et al., 2011).

3.8.1 Genotypic data

Genomic DNA was extracted from the leaves of 51 turmeric accessions using DNeasy plant mini kit (Qiagen) and were sent to Scigenome, Kochi, Kerala to perform double digest restriction site associated DNA sequencing. The SNP genotyping was performed using Illumina HiSeq 2000 platform.

3.8.1.1 ddRAD library preparation and sequencing

Double digestion of genomic DNA (1 µg) was done using the restriction enzymes *SphI* and *MlucI*. After double digestion the cleanup of the digested product was done using Ampure beads. Ligation with adaptors P1 (Barcoded) and P2 was conducted using T4 DNA ligase. Thereafter the cleanup and pooling of the ligated products were done. Size selection of the product was done after 2% agarose gel electrophoresis. The next step was PCR amplification to enrich and add the Illumina specific adapters and flow cell annealing sequences. QC check was done on bioanalyzer, and the final pooling and sequencing was done. After sequencing,

demultiplexing of the reads was done using the Axe-demultiplexer software version 0.3.3. The result of ddRAD based sequencing are paired end sequences per sample named as R1 and R2 respectively. The workflow is given in Fig. 3.5.

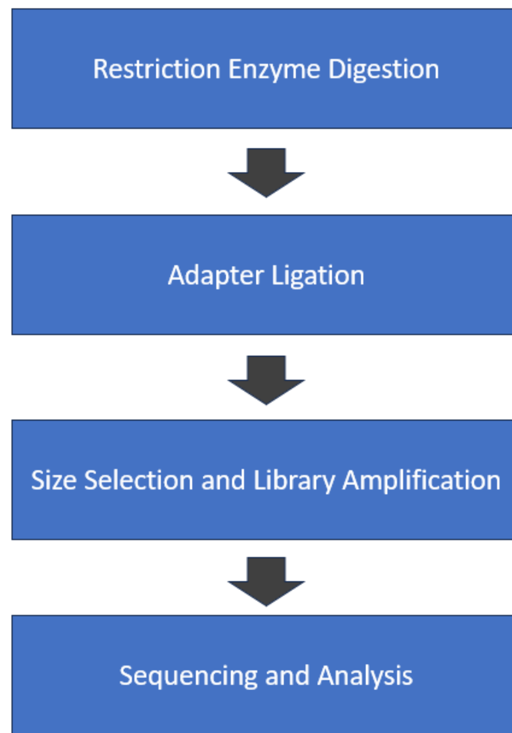


Fig. 3.5. ddRAD library preparation and sequencing workflow

3.9 NGS data analysis: available tools and programs

After acquiring the initial sequence data, there are various options for processing the data. If there is no existing reference sequence for a particular species, it becomes necessary to generate a *de novo* assembly, which is a complex and computationally demanding process (Zhang et al., 2011).

In the presence of a reference genome, sequence reads can be aligned (mapped) back to the reference to facilitate SNP calling across multiple samples. Short-read alignment tools such as BWA and Bowtie, which employ the Burrows-Wheeler Transform (BWT) method have gained prominence due to their efficiency and lower memory requirements compared to earlier hash-based methods like MAQ

and Novoalign. Recent versions, such as BWA-mem and Bowtie 2, have demonstrated improved accuracy and performance for longer reads and better handling of indels, respectively (Li & Durbin, 2009).

Alignments are typically stored in the sequence alignment/map (SAM) format (Li et al., 2009) and can be converted to binary (BAM) files. SNP and indel detection were carried out on SAM/BAM files using programs like SAMtools, Freebayes, and GATK (Garrison & Marth, 2012). SAMtools and GATK identify variants based on read alignment to the reference, while Freebayes determines SNPs based on the most likely combination of genotypes at each reference position. Bayesian methods are employed by Freebayes and GATK to model sequencing errors. Freebayes additionally utilizes Bayesian framework for multi-allelic haplotype detection. SAMtools incorporates a hidden Markov mapping and assembly quality (MAQ) model to estimate error and assumes a prior probability of observing a heterozygote to be 0.001, resulting in less frequent heterozygous genotype calls. SAMtools and Freebayes are characterized by their simplicity and flexibility, while GATK has stricter data formatting requirements, necessitating additional steps in the SNP calling pipeline (Lefouili & Nam, 2022). GATK offers a local realignment step to enhance alignment around indels, eliminating frameshifts that might generate false-positive SNP calls. Furthermore, GATK employs variant quality score recalibration (VQSR) on large datasets to differentiate true SNPs from artifacts during data processing (Lefouili & Nam, 2022). Both GATK and Freebayes allow the user to select ploidy levels beyond haploid or diploid. The output format for all three tools is Variant Call Format (VCF), which lists the identified variants along with their positions, scores, and other.

3.9.1 Bioinformatics pipeline

Here, the goal of bioinformatics pipeline was the analysis of NGS data for variant discovery (SNP genotyping) using genotypic data from 51 turmeric genotypes. A basic overview of the bioinformatics pipeline followed for NGS data analysis is given in Fig.3.6. Like other NGS applications, the variant calling workflow starts with quality control to find the suitability of our data for downstream analysis. This

is followed by alignment to a reference genome as reference genome was available in turmeric (Chakraborty et al., 2021). Alignment is generally followed by alignment clean-up to get the data ready for variant calling (Li & Durbin, 2009). This involves the pre-processing of BAM files. After that, variant calling was done. Further, variant calls were filtered and their functional significance was examined. Here, the annotation of variant was not done as annotated form of genome was not available. Variant calling in turmeric is challenging as the crop is polyploid and most of bioinformatics tools available are suited to diploids (Yao et al., 2020). To begin with we have tried with various tools for alignment and variant calling to finally optimize a pipeline for variant calling in our crop of interest (Begali, 2018; Gubaev et al., 2020). Detailed workflow shown in Fig. 3.7.

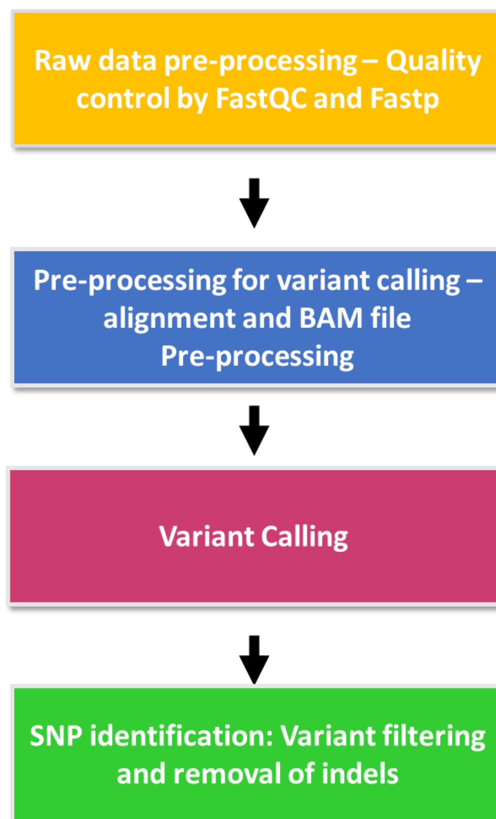


Fig. 3.6. NGS data analysis pipeline followed for SNP genotyping

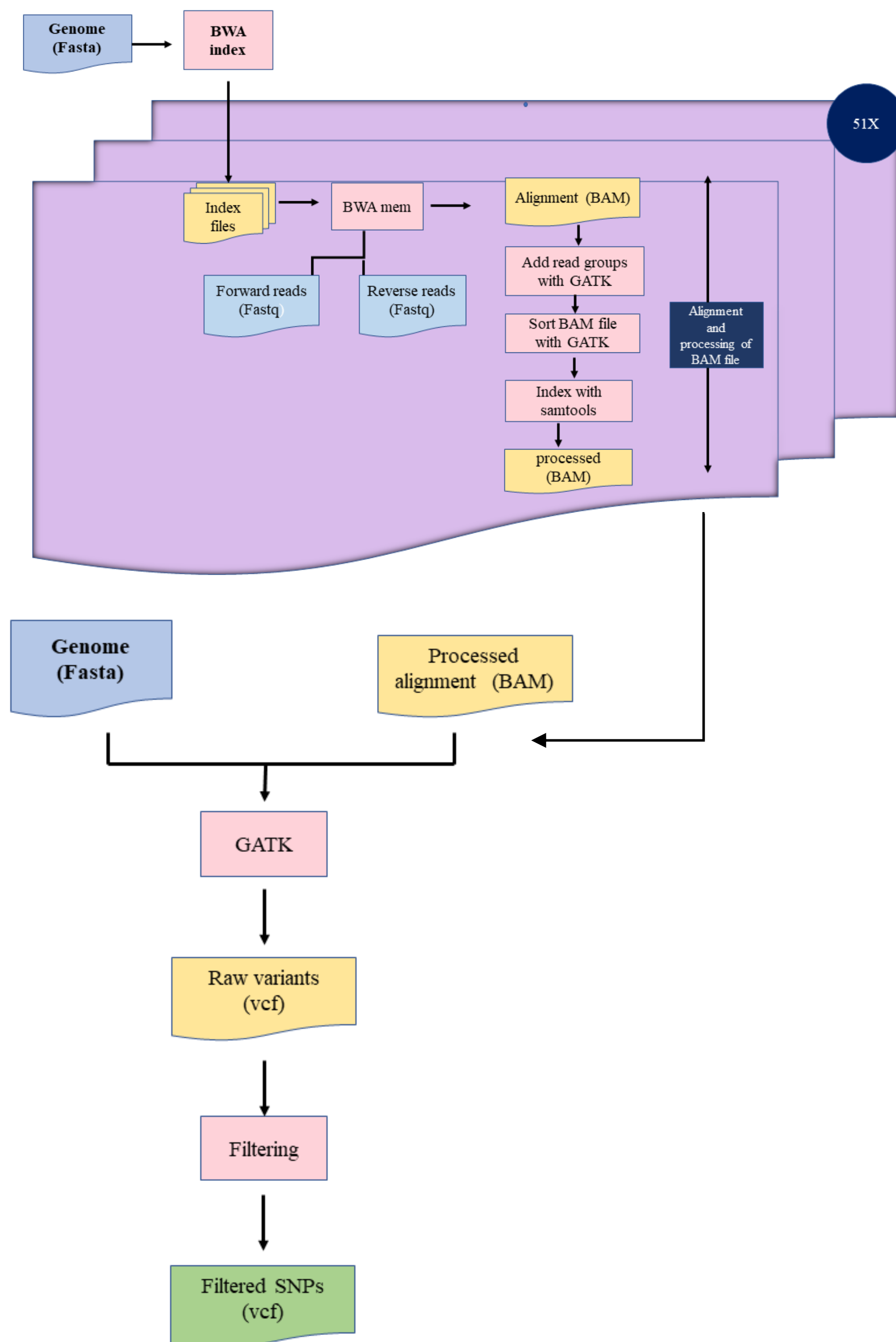


Fig. 3.7. Detailed workflow of bioinformatic pipeline for SNP discovery

3.9.1.1 Step 1: Pre-processing of raw data

3.9.1.1.1 Quality control

NGS sequencing result in tens of millions of reads in a single run. A quick quality check of this raw sequenced data helps to detect issues or biases that would adversely affect downstream data analysis and biological inferences (Sudan et al., 2019) . Here, QC check was done using FastQC for the individual samples. It was done in a miniconda environment in Linux virtual machine created in personal computer using virtual box (https://www.virtualbox.org/wiki/Downloads).

Input files: Paired end ddRAD sequencing data from 51 genotypes.

Tool: FastQC, multiQC and fastp.

The basic command for FastQC run was:

fastqc -t 2 -o QC*.gz.

A multiQC report was generated based on all FastQC files (multiqc -p).

Based on QC reports, trimming of low quality reads as well as removal of Illumina sequencing adapters were performed using fastp:

Basic FastP Command:

fastp -w 2 -a AGATCGGAAGAGC -i input filename 1.fastq.gz -I Input filename 2.fastq.gz -o Trimmed_R1.fq.gz -O Trimmed_R2.fq.gz

where input file 1 and 2 are paired end reads R1 and R2 of a sample respectively and Trimmed_R1.fq.gz and Trimmed_R2.fq.gz are the corresponding trimmed files. (AGATCGGAAGAGC is the illumina adapter sequence).

3.9.1.2. Step 2 Pre-processing for variant calling

3.9.1.2.1 Alignment to reference genome

For alignment, the turmeric reference genome (CL_assembled _genome.fasta) available in FASTA format was utilized. FASTA format is a widely utilized text-

based format to store nucleotides or amino acids sequence information (Yang et al., 2016). The software programme called BWA to map trimmed paired end ddRAD sequencing reads against the turmeric reference genome was sued (Li & Durbin, 2009). BWA consist of three algorithms: BWA-backtrack, BWA-SW, and BWA-MEM. However, BWA-MEM which is the most recent version, is typically advised for high-quality queries due to its speed and accuracy (Li & Durbin, 2009).

The following steps were followed in alignment:

1) Running BWA mem: to index reference genome

- Directory containing reference genome is created.
- Path for dictionary containing summary information also created.
- Reference genome is indexed using the command.

```
bwa index $ref

samtools faidx $ref
```

- This led to creation of few BWA index files like: genome.fa.ann, genome.fa.bwt, genome.fa.fai, genome.fa.pac, genome.fa.sa
- A dictionary is created using GATK:

```
Java -jar $gatk_dir/gatk-package-4.2.6.1-local.jar CreateSequenceDictionary R=CL_assembled_genome.fasta O=CL_assembled_genome.dict
```

2) Running BWA mem: align reads to reference genome

- Set the reads (trimmed paired end ddRAD reads) directory, output directory and reference genome directory.
- Collect the sample names in variable using command: sample_names=$(awk '{print $2}' $reads/names.txt | sort -u)
- Run alignment and sorting over set of fastq files using command:

```
for name in $sample_names
do
```

bwa mem -M -t 24 $ref/CL_assembled_genome.fasta $reads/$name_R1.fq.gz $reads/$name_R2.fq.gz | samtools view -Sb - > $output_dir/$name.bam

echo Alignment completed for $name

done

where, -M: if a read is divided (i.e different parts or components of a given read is mapped to various locations of reference genome), then a "secondary alignment" label is created for all components other than the primary or main component (a technicality, but crucial for GATK, which ignores secondary alignments).

-R: description to add Read Group. The Read Group ID given to each alignment record helps to let other programmes that perform downstream analysis know where the read originated.

-t 4: means run comprising 4 CPU cores. If more CPUs available, bwa mem can go up to 12 CPU cores.

Output (i.e., alignments) will be written to the file $name.bam (BAM format). BAM (Binary Alignment/Map) is a compact binary file format. Typically, it is obtained by sorting the alignments over genomic coordinate and indexing them using samtools (Li et al., 2009). The output of alignment is in SAM (Sequence Alignment/Map) format. But as it generates a lot of text file and consume large space, it is converted to more compact BAM files.

3.9.1.3 Step 3: Variant calling in GATK

Genome Analysis Toolkit is referred to as GATK. It is a set of command-line tools with a main goal of variant finding for analysing high-throughput sequencing data (Lefouili & Nam, 2022). The tools can be utilised singly or in full workflows when linked together. It offers complete workflows, referred to as GATK Best Practises, that are designed for certain use situations. A Unix-style OS and Java 1.8 are the two software prerequisites for the majority of GATK4 utilities. However, a small number of tools also require R or Python. The main steps followed are:

- Set the input directory containing the aligned files, output directory, reads directory and reference genome.
- Set the path to GATK.
- Collect the sample names in variable.
- The variant calling was done using GATK HaplotypeCaller and ploidy = 3 was used while variant calling for triploids and ploidy = 4 for tetraploids.

The workflow followed for variant discovery is given in Fig. 3.8.

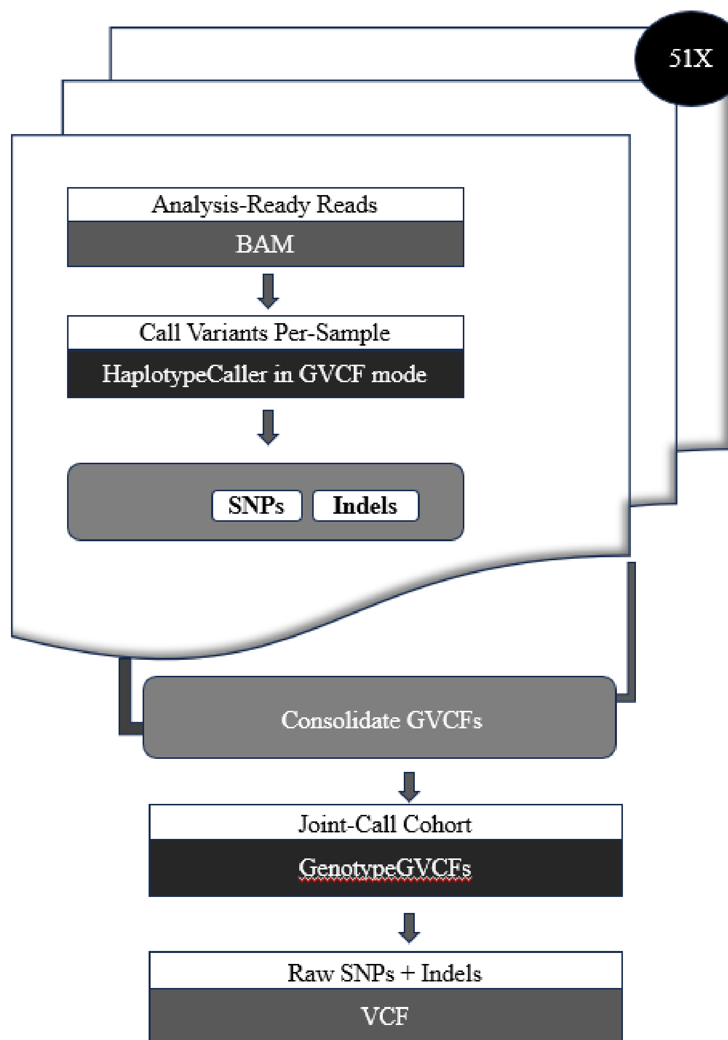


Fig. 3.8. Variant discovery pipeline using GATK

Potential variant sites for each sample are called using HaplotypeCaller, and the output is saved in GVCF format. During the calling procedure, GVCF delivers variant sites and divides non-variant sites into blocks based on genotype quality.

Command: gatk --java-options "-Xmx4g" HaplotypeCaller -R -I input.bam -O output.g.vcf.gz -ERC GVCF

- Combine GVCFs

Created a multi-sample GVCF file by combining the per sample GVCF files generated by HaplotypeCaller.

Command: gatk CombineGVCFs R --variant sample1.g.vcf.gz --variant sample2.g.vcf.gz –O cohort.g.vcf.gz

- Genotype GVCFs

Combined GVCF files were genotyped and major command is:

gatk --java-options "-Xmx4g" GenotypeGVCFs –R -V cohort.g.vcf.gz -O output.vcf.gz

- Output is a VCF file

3.9.1.4 Variant refining-SNP identification

Variant calling tools are sensitive and prone to errors. The aim of variant filtering is to identify potential false positive variants and apply filters to remove those that are less likely to be genuine variants (Yao et al., 2020). Strategies include:

1. Recalibrating the variant quality score (using known sites)
2. Applying hard filters based on quality criteria

Filtering is applied to flag low-quality variant calls. The parameters used for filtering here are MQ<40, QUAL<20, and DP<10.

3.10. Population structure and genetic diversity assessment

The SNP genotypic data obtained after variant calling was used for detection of population structure using Structure 2.3.4 software (Porras-Hurtado et al., 2013). An admixture model (Liu et al., 2020; Rosenberg et al., 2002) was used with following parmeters; length of Burnin period: (100000), and number of Markov chain Monte Carlo (MCMC) Reps (simulations) after Burnin: (100000). K-value was set from 1 to 10 with number of replications = 5. The optimum population structure (K-value) as well as the inferred ancestry (Q matrix) of genotypes were determined based on structure analysis.

3.10.1 Admixture model

The admixture model is a statistical method commonly used in genetics and population genetics research to examine and understand population structure (Liu et al., 2020). Utilizing genetic markers, such as single nucleotide polymorphisms (SNPs), the admixture model can estimate the ancestral proportions of individuals and identify subgroups or clusters within a population. This analysis aids in the understanding of genetic diversity, population history, and potential associations with diseases or traits (Werner et al., 2020). Admixture model relies on certain assumptions, such as marker independence, equilibrium between populations, and the absence of substructure within populations (Pfaffelhuber et al., 2022).

3.10.2 K-value

In the context of population structure analysis using the Structure software, the "K" value represents the number of genetic clusters or subpopulations assumed in the analysis (Porras-Hurtado et al., 2013). It is a user-specified parameter that denotes the hypothesized number of distinct populations within the dataset. Researchers typically set the value of "K" before running the population structure analysis.

3.10.3 Q matrix

Q matrix, also known as the ancestry or membership matrix, is an output generated by the Structure software for each individual in the dataset (Porras-Hurtado et al., 2013). It provides information about the estimated proportion of genetic ancestry for

each individual from each assumed genetic cluster or subpopulation. The Q matrix is presented as a table or matrix, where each row corresponds to an individual and each column corresponds to a genetic cluster. The values within the Q matrix indicate the estimated proportion of ancestry for each individual from each genetic cluster. The sum of the values in each row of the Q matrix is equal to 1, indicating that the proportions across clusters for each individual add up to 100%.

3.10.4 Population structure analysis

The general protocol followed for population structure analysis is given below:

1. Data preparation

The genotype data from SNP genotyping was utilized for population structure analysis. The genotype data was ensured to be in the appropriate format for analysis with the Structure software (HapMap format).

2. Installing and preparing Structure software

Structure software 2.3.4 (Falush et al., 2003; Kaeuffer et al., 2007) was downloaded and installed from the official source. The software's documentation and user manual were reviewed to understand its usage and parameter settings.

3. Setting up the analysis

A new project was created in Structure software. The genotype data was imported into the project. The parameters for the analysis, such as the number of populations (K), burn-in length, and number of Markov Chain Monte Carlo (MCMC) iterations, were defined. Multiple analyses were run with different parameter settings to assess the robustness of the results.

4. Running the analysis

The Structure analysis was initiated by executing the MCMC algorithm. The progress of the analysis was monitored, ensuring convergence by examining the log likelihood values, trace plots, and other diagnostic outputs. The burn-in length and MCMC iterations were adjusted as necessary to achieve convergence and adequate sampling of the parameter space.

5. Analyzing the results

The output files generated by Structure, including the log file, summary statistics, and individual assignment files, were analysed. Appropriate statistical methods were used to interpret the results.

6. Interpreting and reporting the findings

The results were interpreted within the context of the research objectives and hypotheses. The observed population structure, genetic differentiation, and any significant patterns or trends were discussed. Comparisons were made with previous studies, if available, and possible explanations were provided for any observed differences. Relevant figures, tables, and statistical analyses were included in this thesis to support the conclusions.

3.11 Association analysis

Association analysis in crop genetics is a valuable approach employed to investigate the genetic factors responsible for complex traits in plants (Zhu et al., 2008). It entails studying the association between genetic markers and specific phenotypic traits within a given population. By scrutinizing the patterns of genetic variation across individuals, researchers can identify connections between particular markers and traits, shedding light on the genes that influence significant agronomic characteristics. This knowledge proves vital for crop enhancement initiatives, as it allows for the selection of favourable alleles through marker-assisted breeding or genomic selection. Association analysis aids in the discovery of genetic variants underlying traits, facilitating targeted breeding strategies and the development of improved crop varieties endowed with enhanced yield, disease resistance, and other desirable attributes (Hazzouri et al., 2019; Khan & Korban, 2012).

Here, the association between SNP genotypic data with phenotypes of 51 turmeric genotypes was performed using GAPIT software (Garcia et al., 2019; Lipka et al., 2012). The phenotypic traits considered for analysis are curcuminoid content, length of primary rhizome or rhizome length, girth of primary rhizome girth or rhizome girth.

3.11.1 Data preparation: Input files

3.11.1.1 Genotypic data

The SNP data for turmeric obtained after variant calling and refinement of ddRAD sequencing was used for association analysis. Before conducting association analysis in crop genetics, it is necessary to perform SNP data imputation to the HapMap (HMP) format. This was carried out using the Tassel software (Bradbury et al., 2007). Imputation involves predicting and filling in missing genotypic data by leveraging patterns observed in the available SNP markers (Sun & Kardia, 2008). Converting the SNP data to the standardized HMP format ensures compatibility and comparability across different datasets, enhancing the quality and completeness of the data for association analysis. This imputation step facilitates accurate and comprehensive investigations into the relationships between genetic markers and phenotypic traits, providing valuable insights for crop improvement and breeding programs.

3.11.1.2 Phenotypic data

Phenotypic data for 3 traits rhizome length, rhizome girth and curcuminoid content were utilized for association analysis. The data of two-year field trial was utilized here. Firstly, a BLUP (Best Linear Unbiased Prediction) analysis was performed on the phenotypic data using the meta-R package. BLUP analysis estimates breeding values for individuals by considering genetic relationships and environmental effects (Sood et al., 2020). This approach enhances the accuracy of phenotype predictions by incorporating information from relatives and accounting for confounding factors. Using the meta-R package, researchers can conduct meta-analyses by combining data from multiple studies or populations, which increases statistical power and improves result reliability (Ahn & Kang, 2018). This step enables the extraction of high-quality phenotype predictions used in association analysis to identify genetic variants associated with important traits. These findings aid in the development of targeted crop improvement strategies. The text file containing the BLUP values obtained for the three phenotypic traits were utilized as input file in GAPIT analysis.

3.11.1.3 Kinship data

Kinship data obtained from structure analysis called as Q matrix or Q file was also utilized as input in GAPIT analysis. By including the Q file in the analysis, GAPIT can account for population stratification and correct for any confounding effects that might arise due to genetic substructure and reduce false positive associations (Lipka et al., 2012). The Q file thus helps to improve the accuracy of association results by controlling for population structure and enhancing the identification of true genetic associations with phenotypic traits.

3.11.2 Association analysis

3.11.2.1 GAPIT installation and preparation in R software:

a. Installed and loaded the required R packages: The "GAPIT" package and its dependencies (Lipka et al., 2012; Wang & Zhang, 2021) were installed and loaded in R software.
b. The necessary input files for GAPIT (genotype data in HMP format, phenotype file, and kinship data file (Q file)). The genotype data was read using the appropriate function provided by GAPIT. The phenotype data was also read into the R environment. The kinship data was loaded from the Q file.

3.11.2.2 Association analysis in GAPIT

a. Association analysis was performed using BLINK (Huang et al., 2019), a statistical model by incorporating all the input data.
b. The results of the association analysis were assessed, with significant SNPs or genomic regions identified based on statistical significance thresholds. In association analysis, P values are calculated to assess the significance of genetic variant-trait associations (Gubaev et al., 2020). It quantifies the probability of observing the data or more extreme results under the assumption that there is no true association (null hypothesis) (Gubaev et al., 2020). A smaller P value indicates stronger evidence against the null hypothesis and supports the presence of an association. Researchers typically

set a significance threshold (e.g., 0.05) and consider associations with P values below this threshold as statistically significant. When performing these multiple hypothesis tests in association analysis, the probability of encountering false positive associations also increases. Bonferroni correction (Sedgwick, 2012) is a method to adjust the significance threshold (alpha level) to maintain the overall type I error rate. It divides the desired significance level (e.g., 0.05) by the number of tests conducted. This adjustment yields a stricter threshold for declaring statistical significance. It thus provides a conservative approach to control the family-wise error rate and ensure more reliable results.

3.12 Interpreting and reporting the findings

a. The association analysis results were interpreted within the context of the research objectives and hypotheses.
b. The biological relevance of the identified marker trait association (MTA) and their implications for turmeric traits were examined. This was done by extracting 1 Kb sequence containing MTAs (Marker-Trait Associations) and subjecting it to a BLAST analysis (Camacho et al., 2009) against the reference genome, which was assembled at the chromosome level. This analysis successfully determined the chromosomal location of the MTAs. Additionally, a functional significance of the MTAs was examined using BLASTN (Altschul et al., 1990) to identify any potential candidate genes or nucleotide sequences that have a considerable level of similarity with the discovered MTAs.
c. The findings were presented in a comprehensive manner, including appropriate figures, tables, and statistical analyses.
d. Limitations and potential confounding factors were acknowledged and addressed in the discussion.

Chapter 4

RESULTS AND DISCUSSION

4.1 Overview

This chapter of the Ph.D. thesis presents the outcomes and interpretations of our research study focused on investigating the genetic diversity and marker-trait associations in turmeric (*C. longa* L.). The study involved the assessment of turmeric core collection consisting of 93 genotypes at both phenotypic and genotypic levels. The evaluation of phenotypic characteristics in turmeric genotypes revealed significant variations in morphological traits and the key biochemical parameter, namely the curcuminoid content. Through flow cytometry-based ploidy level assessment, two distinct ploidy levels were identified (triploids and tetraploids) with triploids being the predominant group within the diversity panel. Comparative analysis of traits between triploids and tetraploids revealed variations across ploidy levels and highlighted important traits suitable for selection. Furthermore, molecular characterization using microsatellite-based DNA markers provided insights into genetic variation at the DNA level among the genotypes of the diversity panel. Additionally, marker-trait association analysis was carried out using high-throughput SNP markers obtained from ddRAD sequencing, combined with phenotypic data. This analysis facilitated the identification of putative SNP markers associated with traits of interest. This comprehensive investigation into genetic diversity and marker-trait associations offers valuable insights into the genetic makeup of turmeric, paving the way for potential advancements in breeding and crop improvement strategies. The three objectives of the study have been completed and the findings are given in following subsections.

4.2 Objective 1. To study the level of phenotypic variability among core collection

Here, we aimed to assess the phenotypic variation within a diverse core collection of turmeric genotypes. A comprehensive phenotypic evaluation was conducted, encompassing various agronomic and morphological traits such as general plant morphology, leaf characteristics, rhizome characters and curcuminoid content. The results revealed substantial phenotypic variation among the genotypes, indicating the presence of diverse genotypes with distinct agronomic and chemical properties. These findings highlight the potential for exploiting this variation to develop improved turmeric varieties with desirable traits.

Phenotypic evaluation of the turmeric genotypes has revealed the morphological and biochemical variation present in the turmeric diversity panel. Fig 4.1 shows the variation of rhizome morphology among turmeric genotypes.

Fig 4.1. Morphological variation present in the rhizome morphology of turmeric genotypes; a. Punjab Haldi 1; b. Sudharsana; c. IISR Prabha; d. Suguna; e. Varna; f. Suroma; g. Suvarna; h. Megha Turmeric; i. Rajendra Sonia

4.2.1 Morphological characterization (pooled data over two years)

4.2.1.1 Plant height (cm)

The average plant height of 93 genotypes was measured at 112.24 cm, with a standard deviation of 18.70 cm, indicating a considerable range of heights within the collection. The observed plant heights ranged from 71.00 cm to 159.00 cm, emphasizing the presence of notable diversity among the cultivars. Highest plant height was recorded in Uganda turmeric (159.00 cm) followed by Megha turmeric (157.83 cm), Sona (156.17 cm), and Varna (145.33 cm). The genotypes Megha turmeric (157.83 cm) and Sona (156.17 cm) were statistically on par with the genotype of highest plant height, namely Uganda turmeric (159.00 cm) based on DMRT analysis. Smallest height was recorded for IISR RRN-4 (71.00 cm) followed by Acc. 884 (75.67 cm), and SLP 389 (77.00 cm). 31 genotypes with plant height shown in red colour were significantly different from the general mean value (112.24 cm). The results are shown in Table 4.1.

4.2.1.2 Number of shoots per plant

For the number of shoots per plant, significant genetic variation was observed among turmeric genotypes. The average number of shoots per plant was 2.76, with a standard deviation of 0.65. The range observed was from 1.17 to 4.16 shoots. Maximum number of shoots were observed in Mydukkur white (4.16), Suranjana (4.00), Rajendra Sonia (4.00), Acc. 286 (4.00). The genotypes Sudarsana (3.83), Suguna (3.83), Acc. 146 (3.50), Acc. 269 (3.83), Acc. 286 (4.00), IISR Alleppey Supreme (3.50), Suranjana (4.00), Rajendra Sonia (4.00), BSR-2 (3.50), Duggirala red (3.50), NDH-3 (3.50), Punjab Haldi 1 (3.50), Rajapuri (3.83), Acc. 180 (3.50), CO-1 (3.83) and Acc. 297 (3.66) were statistically on par with the genotype of highest number of shoots namely, Mydukkur White (4.16). Minimum shoot of 1.17 was observed in SL 3. 83 genotypes with number of shoots per plant (shown in red colour) were significantly different from the general mean value (2.76). The results are shown in Table 4.1.

4.2.1.3 Number of leaves on the main shoot per plant

The average number of leaves on main shoot per plant was recorded as 7.24, with a standard deviation of 1.04. The leaf number ranged from 5.33 to 9.67. Maximum leaves on main shoot (9.67) were observed in KTS 6 11. Sona (9.33), Pant Peetabh (9.16), Acc. 134 (8.83), NDH-98 (8.66), SLP 389 (9.00), SL 3 (9.17), Acc. 23 (8.67), Acc. 2 (8.67), SC 61 (8.50), Pant Peetabh (9.16), Sona (9.33), Uganda turmeric (8.50), Nepal turmeric (8.50), NDH-3 (8.50), Acc. 19 (8.50) and Acc. 37 (9.17) are statistically on par with the genotype of highest number of leaves on main shoot namely, KTS 611. Minimum number of leaves on main shoot (5.33) was observed in Acc. 130. 16 genotypes with number of leaves on main shoot per plant shown in red colour were significantly different from the general mean value (7.24). The results are shown in Table 4.1.

4.2.1.4 Total number of leaves per plant

The average number of total leaves per plant was recorded as 15.61, with a standard deviation of 2.96. The leaf number ranged from 8.17 to 24.00. Maximum total leaves were observed in Acc 849 (24.00), Mydukkur white (23.66), whose differences are not statistically significant based on DMRT analysis. The minimum was observed in Acc. 780 (8.17) and Acc. 884 (8.33). 18 genotypes with total number of leaves per plant shown in red colour were significantly different from general mean value (15.61) as shown in Table 4.1.

4.2.1.5 Length of petiole (cm)

The average length of leaf petiole was recorded as 14.95 cm, with a standard deviation of 3.35. The petiole length ranged from 5.50 cm to 28.33 cm. Maximum petiole length was observed in Mydukkur white (28.33 cm), CO-1 (25.78 cm). There was no genotype with petiole length on par with the genotype of highest petiole length namely, Mydukkur White. The minimum was observed in Waigon turmeric (5.50 cm). Based on DMRT test, it was found that 24 genotypes had petiole length that was significantly different from the mean petiole length of all genotypes (indicated in red colour in Table 4.1).

4.2.1.6 Leaf length (cm)

The average leaf length was recorded as 46.96 cm, with a standard deviation of 8.10. The leaf length ranged from 21.16 cm to 65.16 cm. Maximum leaf length was observed in Megha turmeric (65.16 cm) followed by Uganda turmeric (64.33 cm), Sona (63.55 cm), which were statistically on par. The minimum was observed in KTS 611 (21.16 cm). Leaf length of 71 genotypes showed in red colour were significantly different from the general mean value (Table 4.1).

4.2.1.7 Leaf width (cm)

The average width of leaf was recorded as 13.34 cm, with a standard deviation of 2.03 cm. The width of leaf ranged from 6.78 cm to 17.85 cm. Maximum leaf width was observed in IISR Kedaram (17.85 cm) followed by SC 61 (17.44 cm) whose differences are not statistically significant based on DMRT test. The minimum leaf width was observed in KTS 611 (6.78 cm). Leaf width of 70 genotypes showed in red colour were significantly different from the general mean value (Table 4.1).

Table 4.1 General plant morphological data of turmeric genotypes

S. No	Genotype	PH (cm)	NS	NL	TL	LP (cm)	LL (cm)	LW (cm)
1	Varna	145.33	2.83	7.33	18.33	16.16	60.12	13.50
2	Megha Turmeric	157.83	2.66	7.00	16.33	15.11	65.16	15.33
3	Suranjana	113.67	4.00	6.83	17.66	12.83	44.11	13.72
4	Suguna	104.00	3.83	6.50	17.00	14.50	44.89	13.56
5	IISR Kedaram	140.17	3.00	6.83	13.33	19.44	58.11	17.85
6	Rajendra Sonia	103.50	4.00	6.83	18.83	13.06	43.78	10.55
7	Suvarna	123.67	3.33	5.83	13.50	15.61	55.39	14.72
8	IISR Pragati	94.83	3.33	6.83	16.00	13.00	34.50	12.50
9	IISR Prabha	112.83	2.67	7.00	15.66	17.33	55.94	14.50
10	Sudarsana	100.33	3.83	6.66	18.50	13.50	39.39	12.67
11	IISR Alleppey Supreme	113.17	3.50	6.83	19.33	12.00	49.50	15.33
12	IISR Pratibha	109.33	3.17	6.16	17.00	11.78	47.44	12.55
13	Punjab Haldi 1	106.17	3.50	5.66	15.33	21.50	46.11	11.83
14	BSR-2	113.67	3.50	5.83	19.00	16.39	50.67	10.83
15	CO-1	116.00	3.83	6.00	16.89	25.78	35.00	9.83

16	CO-2	113.17	2.33	7.00	14.83	23.11	46.83	12.83
17	NDH-1	112.83	3.33	7.00	17.33	15.94	44.28	10.67
18	NDH-3	105.83	3.50	8.50	18.00	14.33	38.66	10.50
19	Rajapuri	104.83	3.83	7.00	15.55	12.50	43.50	16.50
20	Roma	123.67	2.66	9.33	16.33	18.50	48.83	12.17
21	Suroma	135.50	2.00	8.00	15.72	17.00	52.50	14.33
22	Reshmi	139.17	2.66	7.67	17.33	18.22	56.44	11.83
23	Ranga	112.83	2.50	7.83	15.17	17.27	46.50	11.16
24	Duggirala red	136.67	3.50	8.33	20.67	18.83	52.00	11.72
25	Palam Pitambar	95.50	2.67	6.33	13.00	14.83	42.44	12.39
26	Kanthi	141.00	2.83	8.33	20.50	17.16	56.05	15.39
27	Sobha	142.33	2.83	7.83	16.00	15.27	59.66	14.16
28	Sona	156.17	2.50	9.33	19.00	17.11	63.55	15.44
29	Pant Peetabh	121.33	3.00	9.16	21.50	19.78	46.44	12.56
30	Punjab Haldi 2	106.00	2.17	8.00	12.17	17.28	45.28	12.83
31	CIM Pitamber	113.00	2.83	7.83	15.83	22.17	50.28	12.50
32	Rajendra Sonali	99.83	2.33	7.50	15.33	16.56	47.89	12.72
33	NDH-98	134.00	1.33	8.66	11.33	12.83	46.94	13.14
34	Kadappa local	100.67	1.33	8.33	9.17	14.11	47.83	13.39
35	Erode Turmeric	90.67	2.00	7.83	12.33	12.50	45.39	12.11
36	Waigon Turmeric	96.00	2.33	6.50	13.00	5.50	29.50	7.83
37	Salem local	106.17	2.83	7.67	15.50	16.39	47.61	12.72
38	Santra	112.00	2.50	6.33	11.83	14.11	47.00	14.66
39	Kandaila Haldi	107.83	2.50	6.66	13.50	14.11	49.28	14.55
40	Dehati Haldi	135.50	2.83	6.33	16.67	14.22	54.44	15.11
41	Futi Halood	111.00	3.00	6.67	13.33	14.17	50.72	12.88
42	Surkha	113.33	2.50	6.17	13.33	14.22	51.94	13.17
43	Hardi	96.50	2.50	6.50	13.83	13.61	44.55	13.89
44	Koirana	105.83	2.66	7.33	18.67	14.00	46.61	13.44
45	Hasgova	102.17	2.67	6.50	13.33	13.94	45.05	12.27
46	Acc. 2	114.17	2.67	8.67	17.00	16.22	47.89	15.77
47	Acc. 23	137.50	2.17	8.67	16.17	16.00	52.56	16.33
48	Acc. 224	122.33	2.33	7.83	16.00	16.00	53.11	13.06
49	Acc. 246	134.33	3.00	8.16	18.00	15.94	56.89	14.72
50	Acc. 269	96.00	3.83	5.83	16.00	13.11	44.33	13.27
51	Acc. 780	112.83	1.67	6.33	8.17	15.78	49.78	15.33
52	Acc. 821	109.67	2.50	6.33	14.33	15.72	52.67	13.33
53	Acc. 880	101.33	2.83	5.67	12.83	14.50	41.10	12.50

54	Acc. 849	129.33	3.33	7.83	24.00	11.28	52.33	15.61
55	Acc. 884	75.67	1.50	7.00	8.33	13.17	33.11	10.33
56	Acc. 887	122.33	2.17	6.17	11.99	16.61	55.94	14.11
57	Acc. 449	97.50	3.00	6.00	17.17	10.22	46.78	12.94
58	Acc. 300	106.83	3.00	6.17	15.50	15.80	42.33	14.50
59	Acc. 415	107.67	3.17	6.66	17.33	12.39	46.33	14.78
60	IISR RRN-1	89.17	2.17	5.83	11.17	13.66	36.17	13.17
61	IISR RRN-2	99.67	2.33	6.66	15.50	12.27	44.00	11.17
62	IISR RRN3	81.17	2.17	6.50	14.00	11.89	39.33	10.89
63	IISR RRN-4	71.00	2.50	7.16	13.00	11.66	37.17	10.94
64	SLP 389	77.00	1.33	9.00	9.83	13.89	47.22	12.67
65	SL 3	108.50	1.17	9.17	9.50	11.16	35.22	11.16
66	KTS 611	85.50	1.33	9.67	11.67	6.39	21.16	6.78
67	SC 61	134.33	3.33	8.50	19.50	18.05	52.28	17.44
68	Acc. 19	112.33	2.00	8.50	17.67	14.50	47.16	15.50
69	Acc. 138	110.50	2.67	6.83	14.67	13.61	46.55	15.02
70	Acc. 1046	84.33	2.83	6.33	14.50	12.05	33.94	11.50
71	Uganda	159.00	2.50	8.50	17.33	18.50	64.33	14.83
72	Acc. 1054	102.17	2.83	7.17	17.50	12.33	46.78	13.11
73	Nepal Turmeric	124.00	2.83	8.50	18.66	17.83	51.33	13.50
74	Mydukkur White	127.33	4.16	6.67	23.66	28.33	52.83	9.50
75	Mydukkur Orange	123.67	2.33	7.00	19.16	18.94	57.78	13.33
76	Acc. 37	140.00	2.33	9.17	16.50	14.83	47.00	15.50
77	Acc. 286	93.00	4.00	5.67	17.17	10.61	38.83	13.61
78	Acc. 297	88.17	3.66	5.66	15.83	12.00	32.33	11.33
79	Acc. 52	127.00	3.00	7.83	13.50	14.00	44.11	15.67
80	Acc. 161	84.33	2.66	8.17	15.00	13.00	34.11	11.83
81	Acc. 84	106.83	2.17	7.00	15.00	17.05	46.55	16.55
82	Acc. 313	104.00	3.00	6.16	13.50	15.17	47.16	14.66
83	Acc. 180	82.17	3.50	7.83	17.00	12.50	33.05	10.55
84	Acc. 79	91.17	3.17	7.66	17.00	13.00	37.00	12.00
85	Acc. 142	130.17	2.66	6.33	13.50	17.79	61.05	14.95
86	Acc. 200	90.45	2.66	6.50	16.00	11.55	38.61	11.22
87	Acc. 376	98.83	2.50	7.33	14.00	14.00	45.61	14.55
88	Acc. 1	124.33	2.33	7.83	15.00	13.33	45.02	14.60
89	Acc. 146	125.33	3.50	6.50	16.17	12.61	58.17	16.11
90	Acc. 8	107.50	2.17	7.66	13.72	14.00	37.33	15.66

91	Acc. 130	127.00	3.00	5.33	14.16	14.05	54.39	15.00
92	Acc. 134	110.67	2.83	8.83	17.83	14.90	47.44	16.55
93	Acc. 902	124.67	3.33	7.83	18.16	10.94	52.83	15.67
General Mean		112.24	2.76	7.24	15.61	14.95	46.96	13.3 4
CV (%)		2.36	12.67	8.25	7.56	4.85	3.53	4.88
LSD at 5%		5.25	0.69	1.18	2.34	1.44	3.29	1.29

PH – Plant height (cm), NS – Number of shoots per plant, NL – Number of leaves on the main shoot per plant, TL – Total number of leaves per plant, LP – Length of petiole (cm), LL – Leaf length (cm), LW - Leaf width (cm)

4.2.2 Rhizome characters

4.2.2.1 Number of mother rhizomes per plant

The average number of mother rhizomes per plant was 3.75 (standard deviation of 1.04), with a range between 1.50 and 6.50. The maximum number of mother rhizomes was observed in Acc. 297 (6.50), while the minimum of 1.50 was observed in Suroma. This extent of variation was statistically significant as per DMRT analysis. The genotypes IISR Kedaram (6.00), Acc. 200 (5.61), Hasgova (5.50), IISR RRN-3 (5.83), CIM Pitamber (5.50) and Acc. 79 (5.60) were statistically on par with genotype of highest number of mother rhizomes namely, Acc 297. Number of mother rhizomes of nine genotypes were significantly different from the general mean value (showed in red colour in Table 4.2).

4.2.2.2 Length of mother rhizomes per plant (cm)

The average length of mother rhizomes was 5.78 cm (standard deviation of 1.43), ranging from 3.73 cm to 11.95 cm. The smallest recorded length was 3.73 cm by genotype Acc. 161, while the longest was 11.95 cm by genotype Acc. 902, demonstrating the variation in length of mother rhizomes, which was statistically significant as per DMRT analysis. The length of mother rhizome of genotype Suroma (11.78 cm) was statistically on par with the genotype of highest length of mother rhizome namely, Acc. 902. Nine genotypes (length of whose mother rhizomes indicated in red colour) were significantly different from the average length of mother rhizomes in the population (Table 4.2).

4.2.2.3 Girth of mother rhizomes per plant (cm)

For girth of mother rhizome, the average was 10.73 cm (standard deviation = 1.77), with a range of 7.55 cm to 16.03 cm. The smallest girth was observed in Suguna (7.55 cm) followed by Koirana (7.96 cm) while the largest was observed in Suroma (16.03 cm) followed by Acc. 902 (15.89 cm), Acc. 37 (15.25 cm), highlighting the diversity in rhizome thickness within the population. The genotypes Acc. 902 (15.89 cm), Acc. 37 (15.25 cm) and Acc. 849 (14.93 cm) was statistically on par with genotype of highest girth of mother rhizome namely, Suroma. 19 genotypes (girth of whose mother rhizomes indicated in red colour) were significantly different from the average girth of mother rhizomes in the population (Table 4.2).

4.2.2.4 Number of primary rhizomes per plant

The average number of primary rhizomes per plant was 12.53 (standard deviation of 3.10), with a range between 4.65 and 19.42. The maximum number of primary rhizomes was observed in Acc. 297 (19.42) while the minimum of 4.65 was observed in Waigon turmeric followed by Kadappa local (4.75). Palam Pitambar (17.00), IISR RRN-2 (17.00), Suvarna (17.39), Acc 286 (18.71), Acc 246 (18.67) and Kanthi (17.50) were statistically on par with genotype of highest number of primary rhizomes namely, Acc 297. Eighteen genotypes (number of whose primary rhizomes indicated in red colour) were significantly different from the average number of primary rhizomes in the population (Table 4.2).

4.2.2.5 Length of primary rhizomes per plant (cm)

The average primary rhizome length was found to be 7.95 cm (standard deviation of 1.12), ranging from 4.00 cm to 10.10 cm. The smallest recorded length was observed in genotype KTS611 (4.00 cm), followed by Kadappa local (4.90 cm) while the longest was 10.10 cm by genotype Reshmi followed by IISR RRN-1 (9.99 cm), demonstrating the variation in length of primary rhizomes. The genotype Suvarna (9.65 cm), IISR Alleppey Supreme (9.60 cm), Acc. 23 (9.35 cm), Acc. 2 (9.50 cm), Pant Peetabh (9.39 cm), Sobha (9.35 cm), Uganda (9.55 cm) and IISR RRN-1 (9.99 cm) were statistically on par with genotype of highest length of

primary rhizomes namely, Reshmi. The length of primary rhizomes of 22 genotypes (indicated in red colour) were significantly different from the average length of primary rhizomes in the population (Table 4.2).

4.2.2.6 Girth of primary rhizome per plant (cm)

For primary rhizome girth, the average was 7.34 cm (standard deviation = 1.21), with a range of 4.50 cm to 13.50 cm. The smallest girth was observed in KTS611 (4.50 cm) followed by Koirana (4.57 cm) while the largest was observed in Acc. 180 (13.50 cm), highlighting the diversity in primary rhizome thickness within the population. There exists no genotype with girth of primary rhizomes that is on par with genotype of highest girth of primary rhizome namely, Acc. 180. The girth of primary rhizomes of 20 genotypes (indicated in red colour) were significantly different from the average girth of primary rhizomes in the population (Table 4.2).

4.2.2.7 Number of secondary rhizomes per plant

The average number of secondary rhizomes per plant was 13.24 (standard deviation of 4.85), with a range between 2.22 and 34.95. The maximum number of secondary rhizomes was observed in Koirana (34.95) while the minimum of 2.22 was observed in KTS 611 followed by Kadappa local (2.50) etc. There exists no genotype with number of secondary rhizomes that is on par with genotype of highest number of secondary rhizomes namely, Koirana. The number of secondary rhizomes of 25 genotypes (indicated in red colour) were significantly different from the average number of secondary rhizomes in the population (Table 4.2).

4.2.2.8 Length of secondary rhizomes per plant (cm)

The average secondary rhizome length measured 4.25 cm (standard deviation of 0.76), ranging from 2.37 cm to 6.10 cm. The smallest recorded length was observed in genotype KTS 611 (2.37 cm), followed by IISR Kedaram (2.45 cm) while the longest was observed in genotype Reshmi (6.10 cm) followed by Acc. 887 (5.87 cm), demonstrating the variation in length of secondary rhizomes. The genotypes Acc. 300 (5.65 cm), Acc. 23 (5.68 cm), Pant Peetabh (5.60 cm), Sobha (5.70 cm), Ranga (5.63 cm) and Acc 887 (5.87 cm) were statistically on par with genotype of

highest length of secondary rhizomes namely, Reshmi. The length of secondary rhizomes of 21 genotypes (indicated in red colour) were significantly different from the average length of secondary rhizomes in the population (Table 4.2).

4.2.2.9 Girth of secondary rhizomes per plant (cm)

For secondary rhizome girth, the average was 5.82 cm (standard deviation = 1.06), with a range of 3.46 cm to 8.45 cm. The smallest girth of 3.46 cm was observed in CO-2 and Koirana, while the largest was observed in Ranga (8.45 cm), followed by NDH-3 (8.15 cm) highlighting the diversity in secondary rhizome thickness within the population. There exists a significant statistical difference between genotypes of highest girth of secondary rhizomes and those with lowest girth of secondary rhizomes. The genotypes Punjab Haldi 2 (8.09 cm), Sudarsana (7.90 cm), NDH-1 (7.90 cm), NDH-3 (8.15 cm) and IISR Pragati (8.05 cm) were statistically on par with genotype of highest girth of secondary rhizomes namely, Ranga. The girth of secondary rhizomes of 24 genotypes (indicated in red colour) were significantly different from the average girth of secondary rhizomes in the population (Table 4.2).

4.2.2.10 Internodal length (cm)

The average internodal length of primary rhizomes was 1.07 cm (standard deviation of 0.21), with a range between 0.53 cm and 1.46 cm. KTS 611 has the closest internodes (0.53 cm) and wider internodes observed in Pant Peetabh (1.46 cm) followed by Ranga (1.44 cm). The genotypes NDH-1 (1.39 cm) and Ranga (1.44 cm) were statistically on par with genotype of highest Internodal length namely, Pant Peetabh. The internodal length of primary rhizomes of 54 genotypes (indicated in red colour) were significantly different from the average internodal length of primary rhizomes in the population (Table 4.2).

4.2.2.11 Inner core width (cm)

The average inner core width of primary rhizomes was 1.16 cm (standard deviation of 0.21), with a range between 0.65 cm and 1.70 cm. KTS611 and Acc. 286 has the smallest inner core width (0.65 cm) and largest inner core width observed in IISR Pragati (1.70 cm) followed by Ranga (1.60 cm) and Suroma (1.60 cm). The genotypes Ranga (1.60 cm), Suroma (1.60 cm) and Nepal turmeric (1.55 cm) were

statistically on par with genotype of highest inner core width of primary rhizomes namely, IISR Pragati. The inner core width of primary rhizomes of 18 genotypes (indicated in red colour) were significantly different from the average inner core width of primary rhizomes in the population (Table 4.2).

4.2.2.12 Outer core width (cm)

The average outer core width of primary rhizomes was 0.44 cm (standard deviation of 0.08), with a range between 0.20 cm and 0.65 cm. Smallest outer core width observed in Kanthi (0.20 cm) followed by Koirana (0.25 cm) and the largest outer core width observed in Punjab Haldi-2 (0.65 cm) and NDH-1 (0.65 cm). Acc 313 (0.55 cm), Mydukkur Orange (0.55 cm), IISR RRN-4 (0.58 cm), SLP 389 (0.55 cm), Rajendra Sonali (0.55 cm), Acc 269 (0.60 cm), Suranjana (0.60 cm), NDH-1 (0.65 cm), Mydukkur White (0.55 cm), CIM Pitamber (0.55 cm), Pant Peetabh (0.55 cm), Duggirala red (0.55 cm), Nepal turmeric (0.55 cm), IISR RRN-1 (0.55 cm) and IISR Pragati (0.55 cm) were statistically on par with genotype of highest outer core width of primary rhizomes namely, Punjab Haldi-2. The outer core width of primary rhizomes of six genotypes (indicated in red colour) were significantly different from the average outer core width of primary rhizomes in the population (Table 4.2).

Table 4.2 Rhizome characterization data of turmeric genotypes

S.No	Genotype	MN	ML (cm)	MG (cm)	PN	PL (cm)	PG (cm)	SN	SL (cm)	SG (cm)	IL (cm)	ID (cm)	OW (cm)
1	Varna	4.50	5.55	10.50	15.50	8.55	6.70	15.77	5.35	5.50	1.10	1.20	0.35
2	Megha Turmeric	4.50	5.45	10.90	15.50	8.10	6.20	18.50	4.05	4.85	0.83	1.25	0.40
3	Suranjana	4.67	6.05	10.30	13.37	8.10	8.20	13.50	5.45	7.35	1.25	1.45	0.60
4	Suguna	4.83	5.52	7.55	15.83	7.37	7.93	15.83	3.88	6.66	1.26	1.33	0.46
5	IISR Kedaram	6.00	6.00	13.30	14.00	6.40	6.48	5.50	2.45	5.15	0.77	1.10	0.45
6	Rajendra Sonia	5.00	5.25	10.15	14.00	8.50	8.78	16.67	3.95	7.58	1.23	1.45	0.50
7	Suvarna	3.66	5.75	11.60	17.39	9.65	7.10	15.40	4.55	5.10	1.05	0.95	0.40
8	IISR Pragati	4.50	5.05	10.35	11.00	9.25	10.00	13.50	4.70	8.05	1.45	1.70	0.55
9	IISR Prabha	2.00	6.75	12.50	14.50	7.40	7.05	13.00	3.40	5.45	0.95	1.10	0.45
10	Sudarsana	4.00	5.45	11.40	14.00	8.35	9.25	14.50	4.75	7.90	1.21	1.25	0.45
11	IISR Alleppey Supreme	4.28	5.50	10.75	14.21	9.60	6.75	17.30	4.10	5.13	1.20	0.90	0.40
12	IISR Pratibha	4.67	8.75	13.80	12.08	7.65	7.60	15.00	4.18	5.70	0.72	1.35	0.45
13	Punjab Haldi 1	3.33	4.13	8.59	8.22	5.91	7.10	8.17	3.20	5.64	1.15	0.96	0.50
14	BSR-2	4.94	5.56	10.57	10.00	6.60	7.70	9.69	3.47	5.65	0.97	1.25	0.60
15	CO-1	5.30	5.25	10.65	12.50	5.60	6.55	10.00	2.60	5.05	0.92	1.10	0.45
16	CO-2	3.50	5.00	10.00	8.00	6.00	7.20	7.50	3.79	3.46	0.83	1.30	0.35

17	NDH-1	3.50	5.13	9.82	14.50	7.15	8.35	16.78	5.00	7.90	1.39	1.05	0.65
18	NDH-3	4.83	5.61	10.17	15.67	7.46	9.28	13.00	3.85	8.15	1.23	1.25	0.45
19	Rajapuri	4.50	4.95	10.50	12.00	8.94	6.44	11.50	4.90	5.30	1.25	1.15	0.35
20	Roma	3.17	6.68	14.25	8.75	7.85	6.75	13.71	3.65	6.20	0.89	1.15	0.45
21	Suroma	1.50	11.78	16.03	8.75	8.65	8.20	16.83	4.20	6.05	0.75	1.60	0.45
22	Reshmi	2.50	5.50	11.80	15.00	10.10	7.15	17.30	6.10	6.00	1.22	1.25	0.40
23	Ranga	4.28	5.75	10.80	13.50	9.05	10.10	16.00	5.63	8.45	1.44	1.60	0.45
24	Duggirala red	3.83	5.83	13.30	13.00	7.50	7.70	8.80	3.78	6.16	0.67	1.00	0.55
25	Palam Pitambar	4.00	4.95	9.16	17.00	7.14	7.60	17.50	5.36	6.52	1.14	1.25	0.50
26	Kanthi	3.00	5.45	11.15	17.50	8.77	6.45	16.94	5.02	5.26	1.11	1.25	0.20
27	Sobha	3.50	5.39	11.89	15.94	9.35	6.70	17.89	5.70	5.45	1.15	1.25	0.45
28	Sona	2.50	6.00	11.80	16.78	8.91	7.40	18.00	4.50	5.23	1.24	1.05	0.45
29	Pant Peetabh	4.44	4.55	9.08	10.72	9.39	9.04	14.83	5.60	7.39	1.46	1.35	0.55
30	Punjab Haldi 2	4.33	5.50	9.70	10.33	8.03	8.91	9.00	4.24	8.09	1.25	1.40	0.65
31	CIM Pitamber	5.50	5.75	9.65	14.00	8.50	8.35	15.67	4.10	7.35	1.37	1.45	0.55
32	Rajendra Sonali	4.16	6.25	9.30	10.50	7.75	8.50	13.50	3.70	6.60	1.28	1.40	0.55
33	NDH-98	1.83	10.33	13.23	9.50	7.71	7.87	15.67	4.45	6.26	0.73	1.25	0.45
34	Kadappa local	3.00	8.00	14.00	4.75	4.90	6.80	2.50	3.95	5.75	0.72	1.05	0.45
35	Erode Turmeric	2.33	5.53	11.05	9.10	5.58	5.93	13.72	4.11	5.07	0.76	1.10	0.40

36	Waigon Turmeric	2.67	5.68	10.93	4.65	8.58	8.31	3.00	3.47	4.37	0.74	1.35	0.45
37	Salem local	3.00	7.33	13.03	15.50	7.59	7.46	10.16	4.00	5.87	0.86	1.10	0.45
38	Santra	3.50	6.10	12.00	14.05	7.90	7.30	14.83	3.90	5.25	0.95	1.10	0.45
39	Kandaila Haldi	3.86	5.44	9.30	13.50	8.80	6.80	11.00	3.40	4.45	1.00	1.05	0.35
40	Dehati Haldi	3.50	5.75	10.25	14.50	9.20	7.55	14.50	3.35	5.40	1.15	1.05	0.40
41	Futi Halood	4.33	5.50	8.57	11.33	7.55	8.23	16.67	4.06	7.09	1.35	1.45	0.45
42	Surkha	4.65	5.05	9.16	13.38	7.89	8.27	15.05	4.89	7.16	1.20	1.50	0.50
43	Hardi	3.50	5.45	10.45	15.15	8.10	6.60	13.42	4.65	5.30	1.13	1.25	0.40
44	Koirana	3.83	4.46	7.96	13.50	5.64	4.57	34.95	3.11	3.46	0.75	0.75	0.25
45	Hasgova	5.50	5.75	8.73	11.16	7.35	8.39	11.83	4.05	6.81	1.19	1.25	0.35
46	Acc. 2	3.50	5.65	11.70	14.50	9.50	7.13	16.33	5.00	5.50	1.05	1.15	0.45
47	Acc. 23	1.83	6.00	11.65	13.50	9.35	6.50	15.82	5.68	5.75	1.14	1.30	0.40
48	Acc. 224	2.00	6.50	12.15	9.50	8.50	7.45	20.00	4.40	5.15	1.15	1.15	0.30
49	Acc. 246	2.17	5.50	11.18	18.67	7.70	6.70	10.50	4.25	5.35	1.11	1.00	0.40
50	Acc. 269	3.50	5.00	9.21	12.33	7.28	7.67	20.83	4.16	6.80	1.33	1.40	0.60
51	Acc. 780	3.17	5.15	10.16	13.50	8.08	6.94	15.00	4.10	5.39	0.88	1.05	0.40
52	Acc. 821	3.00	5.02	9.67	13.00	8.58	6.67	17.05	4.65	4.75	0.86	1.10	0.35
53	Acc. 880	3.67	4.75	9.10	10.83	7.29	6.55	11.00	3.65	4.65	0.98	0.75	0.35
54	Acc. 849	2.50	10.93	14.93	15.79	8.40	7.84	18.79	3.50	5.15	0.66	1.30	0.40
55	Acc. 884	3.50	4.65	8.30	10.50	7.60	8.20	12.50	4.55	6.65	1.15	1.20	0.45
56	Acc. 887	2.50	5.74	11.08	10.66	8.81	6.84	16.83	5.87	5.84	1.13	1.10	0.35
57	Acc. 449	3.93	5.42	10.50	10.45	8.60	5.60	15.78	3.70	4.30	0.98	0.75	0.35

58	Acc. 300	4.50	5.65	11.15	13.50	8.80	6.75	15.50	5.65	5.13	1.00	1.06	0.40
59	Acc. 415	4.37	6.58	11.10	15.16	7.55	6.60	7.50	3.55	4.80	0.95	1.10	0.35
60	IISR RRN-1	3.17	4.83	9.67	9.33	9.99	8.16	5.50	4.07	6.24	1.08	1.20	0.55
61	IISR RRN-2	4.50	4.63	8.53	17.00	9.10	5.49	13.00	4.18	4.80	1.19	0.75	0.40
62	IISR RRN3	5.83	4.38	9.02	12.27	7.23	6.85	7.00	2.96	4.77	0.76	0.95	0.50
63	IISR RRN-4	3.78	4.13	8.10	10.50	7.05	6.70	5.50	3.46	5.47	0.98	1.20	0.58
64	SLP 389	2.50	5.65	10.00	7.83	8.60	7.65	3.00	3.40	6.29	1.30	1.20	0.55
65	SL3	1.83	7.58	13.10	7.00	7.68	8.09	10.33	4.51	6.63	0.74	1.25	0.45
66	KTS 611	2.50	5.88	9.53	5.83	4.00	4.50	2.22	2.37	4.80	0.53	0.65	0.35
67	SC 61	3.50	6.50	11.25	13.50	8.55	6.38	8.00	5.35	6.27	1.32	1.05	0.50
68	Acc. 19	4.16	5.00	10.40	13.00	7.75	6.80	16.00	3.90	5.00	0.98	1.00	0.30
69	Acc. 138	4.00	5.40	10.75	11.00	8.70	7.55	11.00	4.16	5.16	1.27	0.85	0.45
70	Acc. 1046	4.33	5.45	8.35	15.00	7.42	7.65	17.89	4.90	6.64	1.13	1.05	0.45
71	Uganda	3.00	5.85	11.95	8.15	9.55	7.50	12.92	5.06	6.10	1.13	1.00	0.30
72	Acc. 1054	3.00	4.95	10.00	11.00	8.65	7.00	13.00	4.55	5.35	1.35	0.95	0.35
73	Nepal Turmeric	4.00	4.55	9.60	10.00	7.05	8.55	11.00	4.15	7.25	1.32	1.55	0.55
74	Mydukkur White	3.83	5.00	10.10	8.00	5.52	7.45	4.60	3.18	5.68	0.96	1.10	0.55
75	Mydukkur Orange	2.33	4.83	10.28	6.17	6.31	6.49	7.16	4.04	5.47	0.92	1.05	0.55
76	Acc. 37	2.83	8.16	15.25	11.50	8.04	7.23	14.50	3.99	5.73	0.85	1.40	0.50
77	Acc. 286	4.83	4.85	8.45	18.71	6.83	5.22	16.00	4.45	4.25	0.84	0.65	0.30
78	Acc. 297	6.50	4.35	9.06	19.42	8.13	5.50	22.44	3.69	4.17	0.95	0.80	0.45
79	Acc. 52	4.50	5.30	10.50	15.00	8.00	6.45	5.50	4.30	4.70	1.15	0.95	0.35

80	Acc. 161	4.83	3.73	8.45	14.78	7.10	7.05	10.50	4.24	6.70	1.13	1.25	0.53
81	Acc. 84	4.69	5.85	10.70	11.00	7.60	6.05	10.00	4.25	5.85	1.17	1.00	0.40
82	Acc. 313	4.00	5.20	9.75	15.50	7.70	7.75	14.50	3.75	6.25	1.07	1.40	0.55
83	Acc. 180	4.63	5.70	9.70	15.00	9.05	13.50	13.50	4.75	7.55	1.30	1.20	0.40
84	Acc. 79	5.60	3.94	8.68	14.50	8.06	8.94	13.00	4.20	6.75	1.23	1.25	0.45
85	Acc. 142	2.67	5.66	12.58	13.33	8.07	6.50	14.00	4.20	4.95	1.20	1.15	0.35
86	Acc. 200	5.61	5.14	9.35	11.43	8.00	7.85	11.00	4.68	7.25	1.20	1.39	0.50
87	Acc. 376	3.50	5.65	10.15	14.94	7.65	6.25	14.96	3.65	4.60	1.17	0.90	0.45
88	Acc. 1	4.00	5.67	11.20	10.80	8.45	6.48	11.67	4.20	5.19	1.13	1.15	0.35
89	Acc. 146	2.67	6.45	12.32	14.17	9.15	7.55	18.02	5.39	5.99	1.03	1.05	0.35
90	Acc. 8	3.67	5.59	10.00	10.60	8.54	6.34	10.83	3.77	4.98	1.11	1.00	0.40
91	Acc. 130	2.50	5.56	12.14	10.27	8.00	7.25	16.94	5.10	5.23	1.12	1.20	0.40
92	Acc. 134	3.55	5.61	10.90	10.61	8.55	7.05	12.40	4.49	5.40	1.34	1.25	0.45
93	Acc. 902	3.17	11.95	15.89	12.17	8.80	7.93	15.50	4.90	6.53	0.88	1.05	0.40
General Mean		3.75	5.78	10.73	12.53	7.95	7.34	13.24	4.25	5.82	1.07	1.16	0.44
CV (%)		15.25	8.50	5.59	10.08	5.05	5.74	10.31	7.43	6.23	4.26	7.20	13.99
LSD at 5%		1.14	0.98	1.19	2.50	0.80	0.84	2.71	0.63	0.72	0.09	0.17	0.12

MN – number of mother rhizomes per plant, ML- length of mother rhizomes per plant (cm), MG – girth of mother rhizomes per plant (cm), PN – number of primary rhizomes per plant, PL- length of primary rhizomes per plant (cm), PG – girth of primary rhizomes per plant (cm), SN – number of secondary rhizomes per plant, SL- length of secondary rhizomes per plant (cm), SG – girth of secondary rhizomes per plant (cm), IL- internodal length (cm), ID – inner core width of primary rhizomes (cm), OW – outer core width of primary rhizomes (cm).

4.2.3 Yield related characters

4.2.3.1 Weight of mother rhizomes per plant (g)

The average weight of the mother rhizomes in the study was found to be 87.54 g (standard deviation of 40.22). The range of rhizome weights varied from 24.50 g to 230.00 g. The minimum recorded weight was observed in IISR RRN-2 (24.50 g), while the maximum weight was observed in Acc. 849 (230 g) and Acc. 902 (225 g). There was no genotype with weight of mother rhizomes that was on par with genotype of highest weight of mother rhizomes namely, Acc 849. It was found that the weight of mother rhizomes of 38 genotypes (indicated in red colour) were significantly different from the average weight of mother rhizomes in the population (Table 4.3).

4.2.3.2 Weight of primary rhizomes per plant (g)

The average weight of the primary rhizomes in the study was found to be 197.37 g (standard deviation of 90.73). The range of primary rhizome weights varied from 10.00 g to 425.00 g . The minimum recorded weight was observed in KTS 611 (10.00 g), while the maximum weight of 425.00 g was observed in Rajendra Sonia followed by Ranga (420.00 g) and IISR Pragati (420.00 g). There was no genotype with weight of primary rhizomes that was on par with genotype of highest weight of primary rhizomes namely, Rajendra Sonia. Based on DMRT test, it was found that weight of primary rhizome per plant of 43 genotypes were significantly different from the general mean value of the same in the population (Table 4.3).

4.2.3.3 Weight of secondary rhizomes per plant (g)

The average weight of the secondary rhizomes in the study was found to be 82.83 g (standard deviation of 75.10). The range of secondary rhizome weights varied from 3.00 g to 205.00 g. The minimum recorded weight was observed in KTS 611 (3.00 g), while the maximum weight of 205.00 g was observed in Ranga followed by Suranjana (200.00 g). There was no genotype with weight of secondary rhizomes that was on par with genotype of highest weight of secondary rhizomes namely, Ranga. Based on DMRT test, it was found that the weight of secondary rhizomes of

40 genotypes were significantly different from the general mean value of the same in the population (Table 4.3).

4.2.3.4 Total rhizome weight per plant (g)

The average total weight of the rhizomes per plant in the study was found to be 364.06 g (standard deviation of 149.91). The range of total rhizome weights varied from 51.00 g to 775.00 g. The minimum recorded weight was observed in KTS 611 (51.00 g), while the maximum weight of 775.00 g was observed in Ranga followed by Rajendra Sonia (730.00 g) and Suranjana (730.00 g). There was no genotype with total rhizome weight that was on par with genotype of highest rhizome weight namely, Ranga. Based on DMRT test, it was found that total rhizome weight of 42 genotypes (indicated in red colour) were significantly different from general mean value for the same in the population (Table 4.3).

4.2.3.5 Dry rhizome weight per plant (g)

The average dry weight of rhizomes per plant in the study was found to be 68.01g (standard deviation of 22.79). The range of dry rhizome weights varied from 19.17 g to 133.17 g. The minimum recorded weight was observed in Mydukkur white (19.17 g), while the maximum weight was observed in Acc. 902 (133.17 g) followed by Acc. 849 (128.50 g), Sona (112.00 g). There was no genotype with dry rhizome weight that was on par with genotype of highest dry rhizome weight namely, Acc. 902. Based on DMRT test, it was found that dry rhizome weight of 40 genotypes (indicated in red colour) were significantly different from the general mean value for the character in the population (Table 4.3).

4.2.3.6 Biochemical characterization

4.2.3.6.1 Curcuminoids content (%)

The average curcuminoids content in the rhizomes was measured at 3.26% (standard deviation of 1.75%). The range of curcuminoids content varied from 0.51% to 7.17%. The lowest recorded curcuminoids content was recorded in Santra (0.51%) followed by Koirana (0.75%) while the highest was recorded in Sudarsana (7.17%),

followed by NDH-3 (6.85%), and Acc. 269 (6.64%). There was no genotype with curcuminoid content that was on par with genotype of highest curcuminoid content namely, Sudarsana. Based on DMRT test, it was found that curcuminoid content of 39 genotypes (indicated in red colour) were significantly different from the general mean value for the same character in the population (Table 4.3).

Table 4.3 below shows the mean recorded data for yield related characters and curcuminoid content.

Table 4.3 Yield characterization and curcuminoid content of turmeric genotypes

S.No	Genotype	MWT (g)	PWT (g)	SWT (g)	TWT (g)	DY (g)	CUR (%)
1	Varna	90.00	246.50	85.00	460.00	90.50	1.65
2	Megha Turmeric	80.00	185.00	70.00	335.00	90.50	1.70
3	Suranjana	125.00	400.00	200.00	730.00	99.33	5.04
4	Suguna	130.00	300.00	100.00	540.00	64.33	5.95
5	IISR Kedaram	87.00	92.50	35.50	225.00	54.50	4.64
6	Rajendra Sonia	109.00	425.00	175.00	730.00	79.33	5.43
7	Suvarna	106.50	260.00	101.00	450.00	98.50	3.22
8	IISR Pragati	100.00	420.00	187.60	685.00	98.83	6.14
9	IISR Prabha	60.00	157.00	45.00	315.00	59.50	2.09
10	Sudarsana	120.00	355.00	120.00	610.00	64.50	7.17
11	IISR Alleppey Supreme	95.00	220.00	70.00	383.00	91.00	2.02
12	IISR Pratibha	175.00	206.00	100.00	455.00	91.83	4.10
13	Punjab Haldi 1	70.00	115.00	35.00	230.00	21.50	4.00
14	BSR-2	115.50	114.50	42.50	270.00	32.17	3.54
15	CO-1	102.50	151.00	35.33	233.00	29.50	2.32
16	CO-2	60.00	70.00	79.50	155.00	23.33	3.84
17	NDH-1	94.50	295.00	135.00	525.00	55.33	4.43
18	NDH-3	119.50	285.00	114.00	505.00	60.67	6.85

19	Rajapuri	75.00	155.00	45.00	275.00	67.33	1.44
20	Roma	88.00	103.00	54.50	243.00	49.33	4.65
21	Suroma	165.00	210.00	147.33	500.00	105.33	1.43
22	Reshmi	58.50	190.00	125.00	395.00	95.33	2.03
23	Ranga	135.66	420.00	205.00	775.00	104.17	5.35
24	Duggirala red	123.00	175.00	55.50	335.00	87.00	4.75
25	Palam Pitambar	50.00	190.00	115.00	360.00	66.17	5.56
26	Kanthi	75.00	235.00	80.00	395.00	90.83	2.45
27	Sobha	68.50	245.00	112.50	410.00	91.17	1.67
28	Sona	80.00	260.00	144.00	520.00	112.00	2.12
29	Pant Peetabh	168.33	270.00	160.00	520.00	73.83	4.75
30	Punjab Haldi 2	80.00	225.00	70.00	390.00	56.17	6.21
31	CIM Pitamber	107.00	405.00	185.00	700.00	90.67	3.98
32	Rajendra Sonali	95.00	240.00	61.66	455.00	53.17	5.39
33	NDH-98	121.00	141.50	65.00	340.00	77.67	1.70
34	Kadappa local	145.00	60.00	15.00	205.00	44.83	5.64
35	Erode Turmeric	40.00	70.00	30.00	95.00	36.67	2.62
36	Waigon Turmeric	50.00	60.00	35.00	115.00	67.17	2.88
37	Salem local	130.00	155.00	45.00	335.00	90.33	5.04
38	Santra	85.00	190.00	125.00	355.00	81.83	0.51
39	Kandaila Haldi	60.00	160.00	45.00	275.00	60.83	1.59
40	Dehati Haldi	60.00	200.00	66.66	345.00	82.67	1.24
41	Futi Halood	80.00	265.00	135.00	495.00	48.50	5.97
42	Surkha	90.00	277.00	117.50	492.50	74.67	4.26
43	Hardi	75.50	182.50	53.00	315.00	65.33	2.11
44	Koirana	45.00	75.00	40.00	180.00	48.00	0.75
45	Hasgova	115.00	275.00	80.00	480.00	56.33	6.06
46	Acc. 2	120.00	235.00	85.00	415.00	92.67	2.25
47	Acc. 23	48.00	255.00	135.00	430.00	90.33	1.26

48	Acc. 224	46.00	140.00	112.50	284.50	89.00	1.69
49	Acc. 246	42.00	185.00	155.00	284.50	56.83	1.62
50	Acc. 269	70.00	270.00	165.00	510.00	65.50	6.64
51	Acc. 780	43.50	150.00	180.00	280.00	58.83	0.99
52	Acc. 821	75.00	140.00	80.00	285.00	60.00	1.52
53	Acc. 880	49.50	90.50	24.00	165.00	41.17	2.04
54	Acc. 849	230.00	325.00	95.00	575.00	128.50	1.79
55	Acc. 884	39.50	210.00	115.00	405.00	47.00	4.41
56	Acc. 887	50.00	156.00	90.00	305.00	74.00	1.86
57	Acc. 449	80.50	106.00	33.50	200.00	62.67	1.87
58	Acc. 300	73.50	117.50	45.50	280.50	60.67	2.29
59	Acc. 415	135.00	180.00	35.00	340.00	67.67	2.47
60	IISR RRN-1	65.00	230.00	35.00	345.00	78.50	5.38
61	IISR RRN-2	24.50	135.00	25.66	190.00	52.17	1.61
62	IISR RRN3	61.50	140.00	20.00	255.00	47.50	2.02
63	IISR RRN-4	81.66	110.00	27.33	200.00	42.17	2.58
64	SLP 389	61.00	155.00	70.00	235.00	44.67	5.10
65	SL3	80.50	121.00	65.00	280.00	60.33	3.51
66	KTS 611	25.00	10.00	3.00	51.00	29.83	3.69
67	SC 61	40.50	150.50	43.00	322.50	65.67	1.84
68	Acc. 19	70.00	105.00	30.00	213.00	58.67	1.55
69	Acc. 138	93.50	204.00	44.00	335.50	72.33	1.60
70	Acc. 1046	81.00	260.00	135.00	475.00	61.00	5.25
71	Uganda	46.50	185.00	114.50	356.50	69.83	1.75
72	Acc. 1054	67.50	116.00	40.50	263.50	59.17	1.39
73	Nepal Turmeric	110.00	340.00	90.33	476.50	71.33	6.56
74	Mydukkur White	55.00	75.00	40.00	155.00	19.17	1.95
75	Mydukkur Orange	46.50	55.00	23.50	130.00	21.83	3.25
76	Acc. 37	180.00	260.00	85.00	455.00	91.00	1.70

77	Acc. 286	52.00	110.50	50.00	215.00	62.17	3.20
78	Acc. 297	59.00	110.00	53.00	215.00	55.50	4.18
79	Acc. 52	53.50	156.50	18.00	200.00	57.83	1.54
80	Acc. 161	65.00	245.00	195.00	465.00	49.17	5.73
81	Acc. 84	119.5	170.00	41.00	313.50	71.83	1.76
82	Acc. 313	65.00	255.00	80.00	380.00	47.67	5.62
83	Acc. 180	115.00	360.00	110.00	625.00	75.67	4.41
84	Acc. 79	65.00	265.00	90.00	380.00	66.17	5.79
85	Acc. 142	109.00	235.00	109.50	415.00	100.83	2.51
86	Acc. 200	138.50	335.00	70.00	410.00	73.67	4.92
87	Acc. 376	45.00	125.00	38.50	415.00	50.17	1.83
88	Acc. 1	85.00	110.00	40.50	260.00	70.33	2.28
89	Acc. 146	75.00	230.00	95.00	420.00	101.17	2.09
90	Acc. 8	135.00	141.00	45.00	255.00	58.83	1.93
91	Acc. 130	40.00	140.00	70.00	260.00	58.33	2.72
92	Acc. 134	98.00	214.50	62.50	405.00	70.83	2.08
93	Acc. 902	225.00	210.00	150.00	590.00	133.17	1.67
General Mean		87.54	197.37	82.83	364.06	68.01	3.26
CV (%)		19.78	20.00	28.46	17.29	10.52	1.91
LSD at 5%		0.03	0.08	0.06	0.125	0.01	0.124

MWT - Weight of mother rhizomes per plant, PWT - Weight of primary rhizomes per plant, SWT - Weight of secondary rhizomes per plant, TWT - Total rhizome weight per plant, DY - Dry rhizome weight per plant, CUR – Curcuminoid content.

4.3 Statistical analysis

4.3.1 ANOVA results

Phenotypic variability was examined by phenotypic evaluation of general plant morphology, rhizome characters and curcuminoid content. Analysis of variance (ANOVA) was performed to find out the phenotypic traits of significant difference

among the genotypes. Based on the F statistics obtained from the ANOVA test (Table 4.4 a-c), it can be concluded that there was a significant difference (with a P value of 0.01) among the turmeric genotypes for all the 25 phenotypic traits. The DMRT test conducted subsequently identified specific differences between the means of individual genotypes, providing further evidence of the morphological and biochemical variations present among the genotypes (results already summarised in Table 4.1, 4.2 and 4.3). The coefficient of variation (CV) values ranged from 1.91% to 28.56% for percentage of curcumin content and weight of secondary rhizome, respectively.

Table 4.4 a. ANOVA of general plant morphological characters of turmeric genotypes

	df	**Mean square**						
		PH (cm)	**NS**	**NL**	**TL**	**LP (cm)**	**LL (cm)**	**LW (cm)**
Genotypes	92	706.1**	0.86**	2.16**	17.5**	23.23**	131.08**	8.36**
Error	93	7.055	0.124	0.36	1.40	0.53	2.63	0.424

PH – Plant height (cm), NS – Number of shoots per plant, NL – Number of leaves on the main shoot per plant, TL – Total number of leaves per plant, LP – Length of petiole (cm), LL – Leaf length (cm), LW - Leaf width (cm)

Table 4.4b ANOVA of rhizome and yield characters of turmeric genotypes

	df	Mean square											
		MN	ML (cm)	MG (cm)	PN	PL (cm)	PG (cm)	SN	SL (cm)	SG (cm)	IL (cm)	ID (cm)	OW (cm)
Genotypes	92	2.187**	4.16**	6.34**	19.38**	2.54**	2.99**	47.44**	1.11**	2.30**	0.08**	0.08**	0.01**
Error	93	0.33	0.24	0.36	1.60	0.161	0.18	1.86	0.09	0.31	0.002	0.006	0.003

MN – number of mother rhizomes per plant, ML- length of mother rhizomes per plant (cm), MG – girth of mother rhizomes per plant (cm), PN – number of primary rhizomes per plant, PL- length of primary rhizomes per plant (cm), PG – girth of primary rhizomes per plant (cm), SN – number of secondary rhizomes per plant, SL- length of secondary rhizomes per plant (cm), SG – girth of secondary rhizomes per plant (cm), IL- internodal length (cm), ID – inner core width of primary rhizomes (cm), OW – outer core width of primary rhizomes (cm).

Table 4.4c. ANOVA of yield characters and curcuminoid content of turmeric genotypes

	df	Mean square					
		MWT (g)	PWT (g)	SWT (g)	TWT(g)	DY (g)	CUR (%)
Genotypes	92	3067.59**	16467**	19178.8**	44946**	1039.26**	6.16**
Error	93	0.0003	0.001	0.0001	0.004	0.0004	0.003

MWT - Weight of mother rhizomes per plant, PWT - Weight of primary rhizomes per plant, SWT - Weight of secondary rhizomes per plant, TWT - Total rhizome weight per plant, DY - Dry rhizome weight per plant, CUR – Curcuminoid content

4.4 Flow cytometric analysis

Flow cytometry (FCM) analysis conducted on intact leaf nuclei, resulted in distinct histograms that facilitated accurate identification of ploidy levels (Fig. 4.2 a-c). Table 4.5 presents the inferred ploidy levels for each turmeric genotype. Our findings revealed that out of all the turmeric genotypes examined, 84 were triploids, while nine exhibited tetraploidy (Table 4.6). The coefficients of variation (CV) values of the G_0/G_1 peaks fell within the acceptable range of 3.5% (Table 4.6).

In summary, FCM analysis proved to be a reliable method for precise determination of ploidy levels in turmeric genotypes. The clear histograms obtained through FCM analysis provided valuable insights into the genetic composition and diversity of turmeric populations. The identification of triploids and tetraploids among the genotypes studied enhances our understanding of turmeric's reproductive biology and genetic variability. Overall, FCM analysis of intact leaf nuclei contributes to our knowledge of ploidy variation in turmeric, a significant plant species.

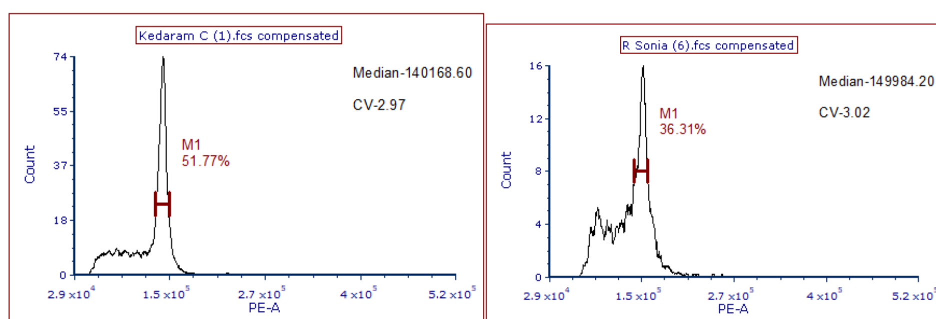


Fig 4.2a **Fig 4.2b**

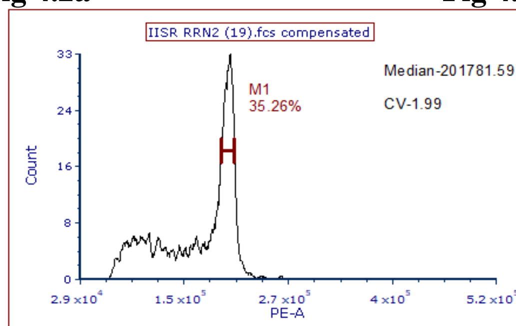


Fig 4.2c

Fig 4.2a. A triploid, IISR Kedaram with known ploidy (2n =63) was used as an external reference standard and its G_0/G_1 peak was positioned at channel with median of 140168.60; **Fig 4.2b.** G_0/G_1 peak of unknown sample (Rajendra Sonia) appeared at median value of 149984.20 and its ploidy was calculated as triploid; **Fig 4.2c**. G_0/G_1 peak of unknown sample (IISR RRN2) appeared with median value of 201781.59 and its ploidy was calculated as tetraploid.

Table 4.5. List of triploid and tetraploids turmeric genotypes based on flow cytometric analysis.

Triploid		Tetraploid
IISR Kedaram (Control)	Santra	IISR RRN1
Varna	Kandaila Haldi	IISR RRN2
Megha Turmeric	Dehati Haldi	IISR RRN3
Suranjana	Futi Halood	IISR RRN4
Suguna	Surkha	SLP 389
Rajendra Sonia	Hardi	Acc 19
Suvarna	Koirana	Acc. 138
IISR Pragati	Hasgova	Acc. 1054 (exotic)
IISR Prabha	Acc. 2	Acc. 449
Sudarsana	Acc. 23	
IISR Alleppey Supreme	Acc. 224	
IISR Pratibha	Acc. 246	
Punjab Haldi 1	Acc. 780	
BSR-2	Acc. 821	
CO-1	Acc. 880	
CO-2	Acc. 849	
NDH-1	Acc. 884	
NDH-3	Acc. 887	
Rajapuri	Acc. 300	
Roma	Acc. 415	
Suroma	SL-3	
Reshmi	KTS 6 11	
Ranga	SC 61	
Duggirala Red	Acc. 1046	
	Acc 1053	
Kanthi	Nepal Turmeric	
Sobha	Acc. 37	
Sona	Acc. 286	
Pant Peetab	Acc. 297	
Punjab Haldi 2	Acc. 52	
CIM Pitambar	Acc. 161	
Rajendra Sonali	Acc. 84	
NDH 98	Acc. 313	

Kadappa Local Erode Turmeric Waigon Turmeric Salem Local Acc. 1 Acc. 146 Acc. 8 Acc. 130 Acc. 134 Acc. 902 Acc. 1504	Acc. 180 Acc. 79 Acc. 142 Acc. 200 Acc. 376 Mydukkur Orange Mydukkur White Acc. 269	

Table 4.6. Summary of ploidy level analysis

Ploidy level	No of individuals	CV (%)	CV range	
			Minimum	Maximum
Triploid	84 (90.3%)	2.77	1.82	4.87
Tetraploid	09 (9.7%)	2.34	1.19	3.16

4.5 Comparison between triploids and tetraploids

A comparison between triploids and tetraploids was conducted using a two-sample t-test to assess various traits. The results indicated significant differences in several characteristics, including plant height, number of shoots, petiole length, leaf length, length of mother rhizome, length of primary rhizomes, length of secondary rhizomes, girth of secondary rhizomes, inner core diameter of primary rhizome, weight of mother rhizomes, weight of primary rhizomes, total fresh yield of rhizome per plant, and dry yield of rhizome per plant. Table 4.7 provides the mean, standard deviation (SD), t-value (t), and p-value (p) for each trait in both triploid and tetraploid groups. The density plot (Fig.4.3 a-n) illustrates the variation of these traits in triploids (group A) and tetraploids (group B). Furthermore, additional genetic variability analysis was performed to determine the genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability (h-square), and genetic advance (GA), which are presented in Table 4.8. Table 4.9 summarizes the key characteristics identified based on the variability analysis. Principal Component Analysis was also employed to further illustrate the grouping of triploids and tetraploids.

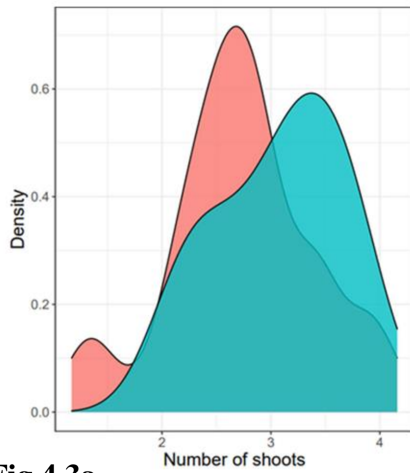


Fig.4.3a

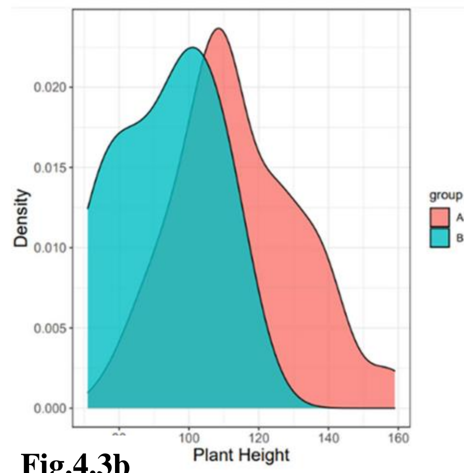


Fig.4.3b

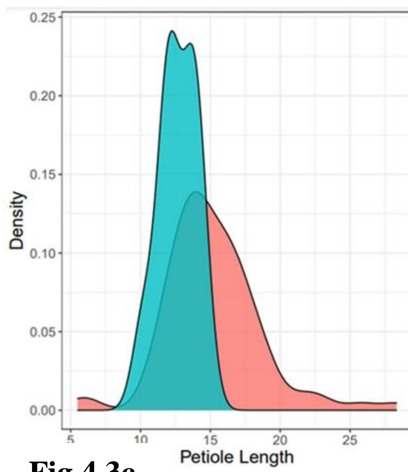


Fig.4.3c

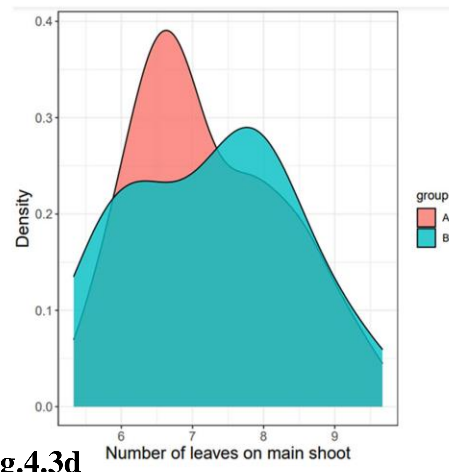


Fig.4.3d

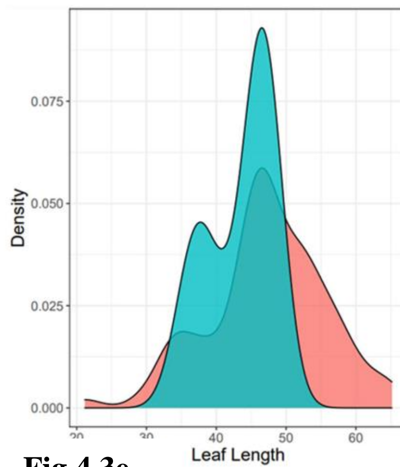


Fig.4.3e

Fig.4.3 a-e . Density plot of phenotypic characters with significant difference between triploids and tetraploids; Fig.4.3a. number of shoots per plant, Fig.4.3b. plant height, Fig.4.3c. petiole length, Fig.4.3d. number of leaves on main shoot, Fig.4.3e. leaf length

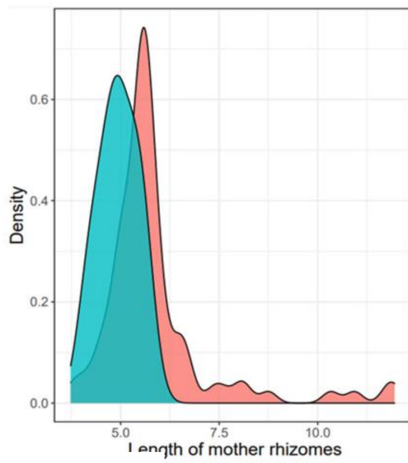


Fig 4.3f

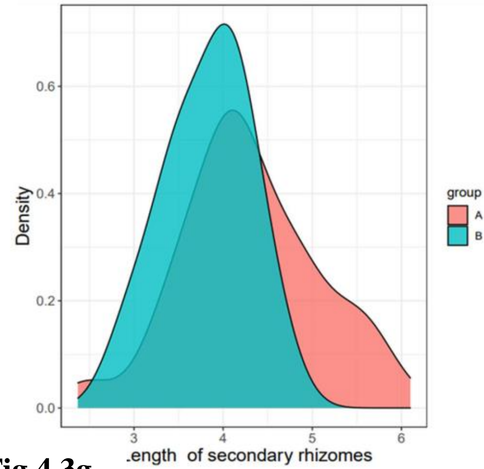


Fig 4.3g

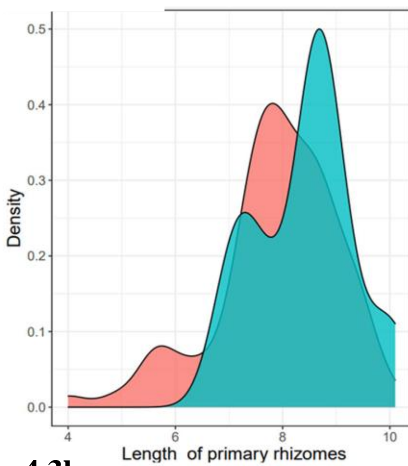


Fig.4.3h

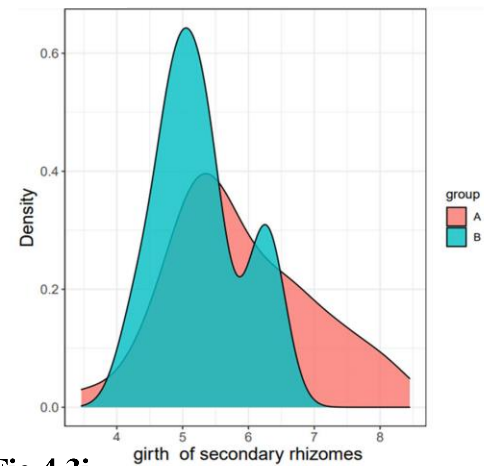


Fig.4.3i

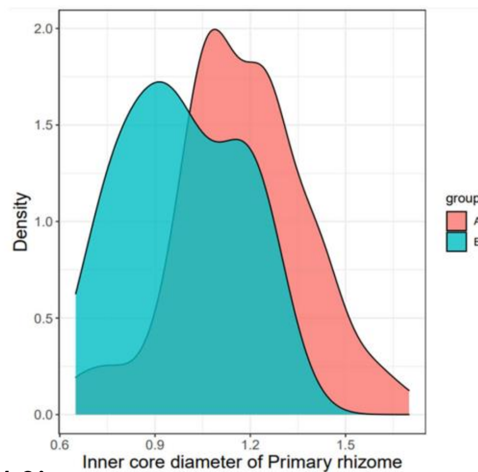


Fig.4.3j

Fig.4.3 f-j. Density plot of phenotypic characters with significant difference between triploids and tetraploids; Fig.4.3f. length of mother rhizomes, Fig.4.3g. length of secondary rhizomes, Fig.4.3h. length of primary rhizomes, Fig.4.3i. girth of secondary rhizomes, Fig.4.3j. inner core diameter of primary rhizome.

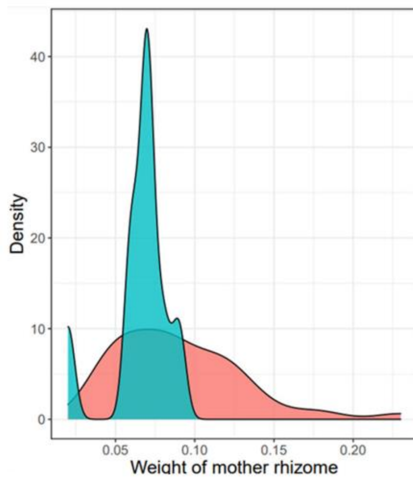


Fig.4.3k

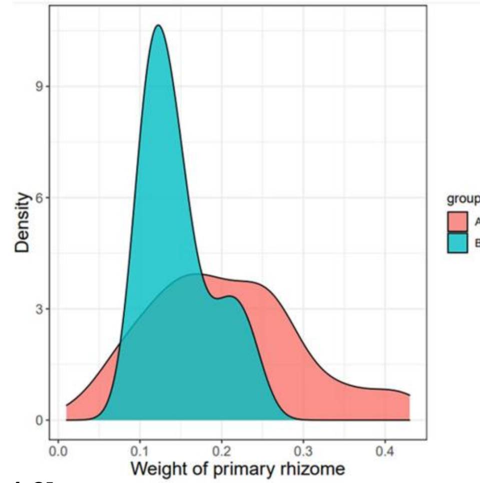


Fig.4.3l

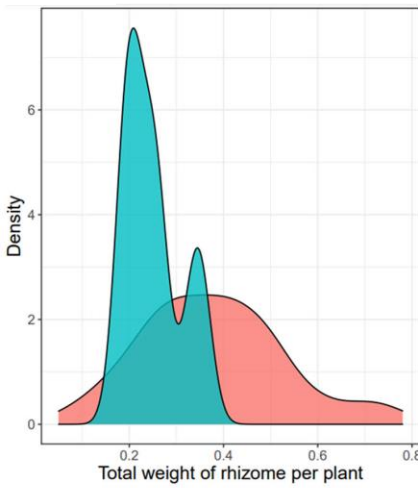


Fig.4.3m

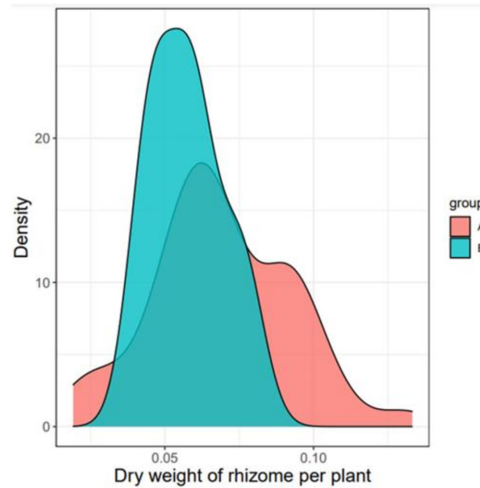


Fig.4.3n

Fig.4.3 k-n . Density plot of phenotypic characters with significant difference between triploids and tetraploids; Fig.4.3k. weight of mother rhizomes per plant, Fig.4.3l. weight of primary rhizomes per plant, Fig.4.3m. total weight of rhizomes per plant, Fig.4.3n. dry weight of rhizomes per plant.

Table 4.7 Phenotypic characters between triploid and tetraploid genotypes

Characters	Group/N	Mean	SD	*t*	*P*
Plant height (cm)	Triploid/84	114.17	18.11	3.93	0.0024
	Tetraploid/9	93.39	14.67		
No. of shoots/plant	Triploid/84	2.82	0.65	2.69	0.0204
	Tetraploid/9	2.33	0.50		
No. of leaves on main shoot	Triploid/84	15.18	3.47	4.25	0.0003
	Tetraploid/9	12.67	1.35		
Petiole length (cm)	Triploid/84	47.22	8.32	2.10	0.0264
	Tetraploid/9	43.46	4.60		
Leaf length (cm)	Triploid/84	5.87	1.48	4.04	0.004
	Tetraploid/9	4.93	0.50		
Length of mother rhizome (cm)	Triploid/84	10.83	1.81	3.20	0.005
	Tetraploid/9	9.66	0.92		
Length of primary rhizome (cm)	Triploid/84	13.54	4.77	2.11	0.03
	Tetraploid/9	9.97	4.83		
Length of secondary rhizome (cm)	Triploid/84	4.28	0.76	2.52	0.013
	Tetraploid/9	3.82	0.49		
Girth of secondary rhizome (cm)	Triploid/84	5.88	1.09	2.47	0.027
	Tetraploid/9	5.26	0.66		
Inner core width of primary rhizome (cm)	Triploid/84	1.17	0.20	2.96	0.013
	Tetraploid/9	0.98	0.18		
Weight of mother rhizome (g)	Triploid/84	0.09	0.04	3.20	0.0053
	Tetraploid/9	0.07	0.02		
Weight of primary rhizome (g)	Triploid/84	0.20	0.09	3.32	0.0038
	Tetraploid/9	0.15	0.04		
Total rhizome weight g/plant	Triploid/84	0.37	0.15	4.97	<0.0001
	Tetraploid/9	0.25	0.06		
Dry weight of rhizome g/plant	Triploid/84	0.20	0.07	2.18	0.045
	Tetraploid/9	0.17	0.04		

Table 4.8 GCV, PCV, h^2 and GA estimates among genotypes

Character	GCV	PCV	Heritability	Genetic advance
Plant height	11.52	26.49	0.18	10.32
No. of shoots/plant	20.10	22.9	0.76	36.25
No. of leaves on main shoot	10.60	14.13	0.57	16.60
Petiole length	17.94	23.69	0.57	28.00
Leaf length	16.71	17.16	0.94	33.54
Length of mother rhizome	21.35	22.51	0.89	41.71
Length of primary rhizome	14.14	14.21	0.99	29.01
Length of secondary rhizome	16.64	17.08	0.94	33.40
Girth of secondary rhizome	17.89	18.24	0.96	36.13
Inner core width of primary rhizome	17.88	18.50	0.93	35.59
Weight of mother rhizome	49.78	52.33	0.90	97.53
Weight of primary rhizome	45.12	46.51	0.94	90.20
Total rhizome weight/plant	38.86	39.72	0.95	78.31
Dry weight of rhizome/plant	33.47	35.38	0.89	65.24

Table 4.9 Important characters that can be considered during selection process

Phenotype	High GCV and PCV	Heritability	Genetic advance
Plant height	-	-	-
No. of shoots/plant	√	√	√
No. of leaves on main shoot	-	-	-
Petiole length	-	-	√
Leaf Length	-	√	√
Length of mother rhizome	√	√	√
Length of primary rhizome	-	√	√
Length of secondary	-	√	√
Girth of secondary rhizome	-	√	√
Inner core width of primary rhizome	-	√	√
Weight of mother rhizome/plant	√	√	√
Weight of primary rhizome/plant	√	√	√
Total rhizome weight/plant	√	√	√
Dry weight of rhizome/plant	√	√	√

4.5.1. Principal component analysis and genetic variability analysis

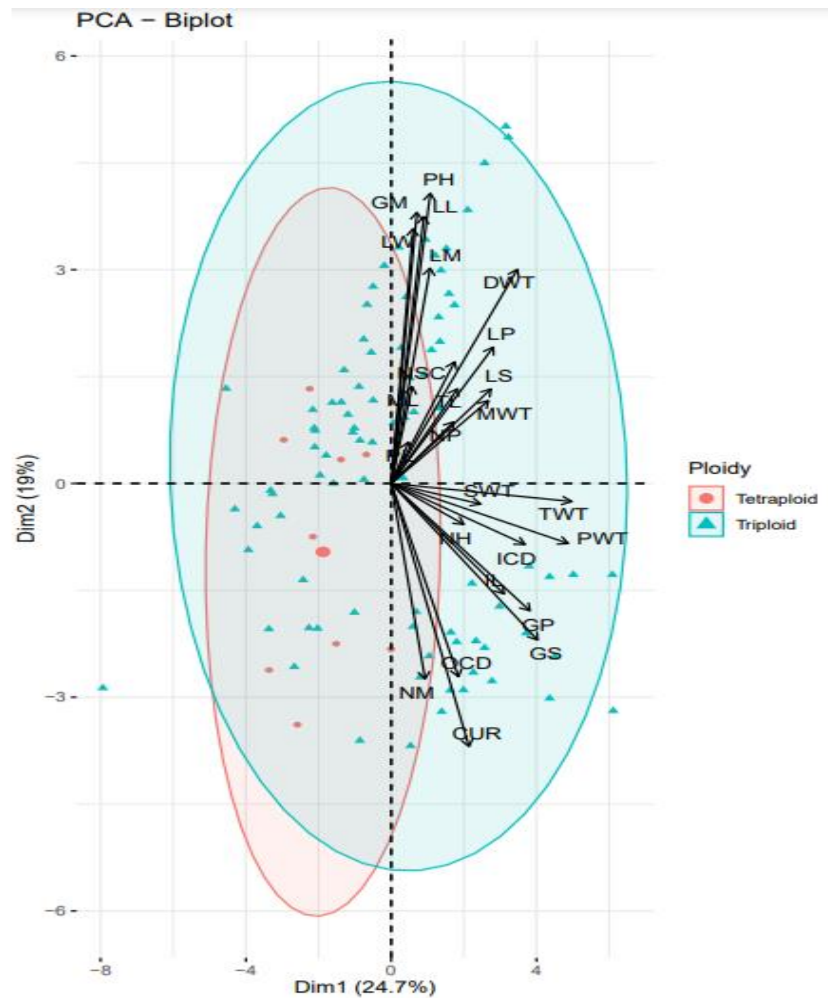


Fig.4.4. PCA Biplot based on phenotypic evaluation of triploids and tetraploids

In the analysis of genetic variability, the calculated GCV values ranged from 10.6% to 49.78%, while PCV values ranged from 14.13% to 52.33% for the characters number of leaves on main shoot and weight of mother rhizome respectively. In general, most of the characters displayed high PCV and GCV values, except for eight specific characters as indicated in Table 4.9.

Furthermore, most of the characters exhibited high heritability values, except for plant height, number of leaves on main shoot, and petiole length, which showed relatively lower heritability values. Similarly, most of the characters demonstrated a

high genetic advance as a percentage of the mean, except for plant height and number of leaves on main shoot, which displayed moderate genetic advance.

Based on Table 4.9, certain characters stand out as important for the selection process, including number of shoots/plants, length of mother rhizome, weight of mother rhizome/plant, weight of primary rhizomes/plant, total rhizome weight/plant, and dry weight of rhizome/plant. These characters possess high GCV, PCV heritability and genetic advance.

Additionally, the elliptical biplot generated from the Principal Component Analysis reveals the pattern of genetic variation and the clustering of genotypes based on phenotypic characters across the two ploidy levels (Fig.4.4).

4.6 Discusssion on results of objective 1

4.6.1 Morphological, yield and curcuminoid content characterization in turmeric

Morphological characterization is the primary step in the description and classification of any crop species (Smith & Smith, 1989). It helps in identification and selection of desirable traits in crop plants (Malek et al., 2014). Previous studies in assessment of morphological diversity in turmeric have reported considerable level of diversity in the crop (Roy et al. 2011; Ullah Jan et al. 2012; Bahadur et al. 2016; Anindita et al. 2020). Our study also reveals a significant phenotypic variation in the crop based on the phenotypic characterization of turmeric core collection. According to an earlier study, experimental precision is categorized as high when the coefficient of variation (CV) falls below 10%, moderate when it ranges between 10% and 20%, and low when it lies between 20% and 30% (Pimentel-Gomes, 2009). Values exceeding 30% are deemed to have very low precision. Based on the variation observed in general plant morphology in turmeric across the genotypes, we can find that all the characters had lower CV values except for the trait of number of shoots per plants that had a comparatively higher CV value (12.67). Earlier studies have reported a moderate heritability for the trait of number of shoots per plant (Mamatha et al., 2020; Paw et al., 2020; Singh et al., 2020). Overall, these general

morphological characters are important as some of these like plant height, number of leaves, size of primary fingers and number of suckers had showed positive and significant association with rhizome yield in earlier studies (Bahadur et al., 2016; Roy et al., 2011). The association of these characters with traits of economic interest, makes them putative morphological markers for traits of our interest and variation in them provide us with ample genetic resources for trait improvement (Lavudya et al., 2024).

Most of the rhizome characters that we studied here are of importance and had been reported as associated with yield, for instance in an earlier study, it was noted that the primary rhizome's weight exhibited the most significant, positive as well as direct impact on yield, succeeded by the length of the primary rhizome, the core diameter of the mother rhizome, as well as the length and weight of the secondary rhizome (Dev & Sharma, 2020, 2022).

Another study found that rhizome yield correlated significantly with the weight and number of secondary rhizomes, weight of mother rhizomes, primary rhizome length, number of tillers, mother rhizome length, weight of primary rhizomes, and leaf area (Jagadeeshkanth, 2017). In this study they also found that selecting for increased mother rhizome length showed the greatest positive impact on traits such as number of leaves, number of mother rhizomes, length of secondary rhizomes, curcumin and essential oil content, number of secondary rhizomes, plant height, and weight of mother rhizomes, which are key for enhancing turmeric rhizome yield (Jagadeeshkanth, 2017). In our characterization, a moderate to low CV was observed for most of the rhizome characters except for number of mother rhizomes, number of primary rhizomes, number of secondary rhizomes and outer core width of primary rhizome that showed moderate CV values. For yield related characters, most of them had high to moderate CV. Curcuminoid content had the lowest CV (1.91%) of all the characters studied. An earlier study has also reported a lower CV (5.87) for curcuminoid content analysed in 15 genotypes (Aarthi et al., 2018).

4.6.2 Ploidy level variation in turmeric, implications in breeding

Ploidy level is an important cytogenetic marker that governs the morphology, physiology, cellular and biochemical characteristics in an organism (Gauthier et al., 1998; Yao Xinzhuan et al., 2023). Polyploidy is a common natural phenomenon, especially in plants. It has been a major driving force in plant evolution and diversification (Heslop-Harrison et al., 2023). As in most other plants, polyploidy may have a significant role in the diversification of the genus *Curcuma* evident from the ploidy levels reported by previous studies of the genus (Bonna et al., 2021; Chen et al., 2013; Nair & Sasikumar, 2009). In plants, it has been hypothesized that there is a strong association between polyploidy and vegetative reproduction wherein the rate of polyploidy speciation is augmented by the existence of vegetative reproduction (Herben et al. 2017). Turmeric, a member of genus Curcuma is mostly a vegetatively propagated crop and is considered as a triploid with less than 60% pollen fertility (Nair & Sasikumar, 2009). There have been reports of chromosome number variation in turmeric based on chromosome counting (Nair et al., 2010; Nair & Sasikumar, 2009; Ramachandran, 1961). Based on these earlier studies, triploid and tetraploid are considered as the major ploidy levels in turmeric (Nair & Sasikumar, 2009). These earlier studies report a wide variation in chromosome number in seedling progenies and few germplasm accessions (which has origin as seedling progenies) from the common 2n = 63 (triploid) and 2n = 84 (tetraploid) (Nair et al., 2010; Nair & Sasikumar, 2009). This study revealed an almost stable chromosome number corresponding to the triploid status for the vegetatively propagated germplasm lines and a wide variation in chromosome number for the seedling progenies possibly due to the sexual mode of propagation (Nair & Sasikumar, 2009). But this information is limited to few genotypes from germplasm and does not capture nor represent the diverse germplasm.

In comparison to tedious chromosome counting, flow cytometry has now emerged as a powerful alternative for high throughput ploidy analyses (Doležel & Bartoš, 2005). In this study we have analysed the ploidy level of the most diverse turmeric varieties which involves genotypes from released varieties, germplasm collections,

farmers varieties, seedling progenies etc. Our findings are also in alignment with previous study as most of the genotypes from germplasm collection are triploid while seedling progenies are tetraploids. Thus, we have examined the diversity in ploidy level among the most diverse turmeric genotypes in the germplasm. Here, we have utilized IISR Kedaram and IISR Prabha with a known ploidy of 2n = 63 as reference standard (triploid control). Earlier studies have recorded the production of seed set and its germination in turmeric leading to the development of high yielding varieties from open pollinated progenies (Nair & Sasikumar, 2009). The ploidy estimates from this study supplemented by actual chromosome counting in the desired genotypes can help in designing breeding experiments in turmeric and to enhance the existing variability in the crop.

4.6.3 Ploidy level and agro-morphological traits

The morphological characterization and estimation of curcuminoids content of these genotypes have revealed the morphological and biochemical variability present among these genotypes. Triploids exhibit a wide variation in morphological features compared to tetraploids in all the traits. Morphological evaluation shows that tetraploids studied are significantly shorter than triploids. Earlier study done on morphological analysis of seedling progenies (most of which were tetraploid) have recorded higher leaf length, leaf breadth and internode length in seedling progenies compared to triploid mother plant (Nair et al., 2010). The characters like plant height, number of shoots, petiole length and dry yield showed a mixed observation as some of the progeny's had significantly higher values, but others had significantly lower values for these traits. One of the seedlings progenies SLP 389 which is also used in our study had a red pigmentation in its emerging shoots which was observed as a heritable morphological marker. Many studies have reported this trend of an increased leaf size in response to increased ploidy level (Dixit & Chaudhary, 2014; Sugiyama, 2005) attributed to a changed cellular parameter like increased cell size and elongation rate. It was observed that polyploids have shoots with thicker and shorter internodes (Yildiz 2013). Leaf length and internodal observations of our study are in alignment with these common observations of polyploids. Regarding

one of the most important qualitative parameters of turmeric rhizomes, i.e., curcuminoid content, the earlier study has reported wide variation in among seedling progenies and curcumin content of all seedling progenies were less than mother plant (Nair et al., 2010). Our observations also support this wherein curcuminoids content in tetraploids were not significantly higher than triploids as expected. This contrasts with the most likely and expected outcome wherein production of metabolites increases because of increased ploidy level due to an enhanced genetic activity (Lavania, 2005). This Enhanced genetic activity can be attributed to an increased copy number of genes or due to a decrease in the ratio of the nuclear membrane to chromatin enabling more chromatins to come in contact with the nuclear membrane. But this decreased level of secondary metabolite accompanying polyploidisation have also been reported in some studies (Caruso et al., 2013; Madani et al., 2021) which speculate this as a probable gene suppression consequence due to suppression of certain genes.

Phenotypic variation is the most likely consequence of genetic variation due to changes in ploidy level as reported in many plants (Gauthier et al., 1998; Yao Xinzhuan et al., 2023) A comparative study of this phenotypic variation between the ploidy levels helps to improve the efficacy of breeding programmes (Huang et al., 2022). Similarly in our study, the comparative evaluation of phenotypic characters between triploid and tetraploid turmeric was studied using a t test. Among the 14 characters in which triploids and tetraploids differ significantly, a genetic variability analysis was conducted to find the traits that are valuable to selection. Number of shoots per plant, length of mother rhizome, total weight of mother rhizomes per plant, total weight of rhizome per plant and dry weight of rhizome per plant were identified as the most important traits. Earlier studies had reported a high GCV, PCV, heritability (>50) and genetic gain for these characters i.e. weight of mother rhizomes, weight of primary rhizomes and weight of secondary rhizomes (Dev & Sharma, 2022; Vinodhini et al., 2018). It can be inferred that selection may not be effective for characters with low heritability and genetic advance i.e plant height and number of leaves per main shoot as they are highly influenced by environment. Earlier works report low GCV, PCV for number of leaves per main

shoot (Vinodhini et al., 2018) but high heritability for plant height (Aarthi et al., 2018; Luiram et al., 2018; Prajapati et al., 2014). The character petiole length that shows a moderate heritability but high genetic advance indicate that the character may be governed by additive gene effects and the heritability is lowered due to environmental influence and hence selection may be effective in such cases as well (Luiram et al., 2018). The rhizome characters studied that include dimensions of rhizomes (length of primary rhizome, length of secondary rhizome, girth of secondary rhizome and inner core width of primary rhizome) except length of mother rhizome showed a moderate GCV and PCV values but high heritability and genetic advance which is likely due to additive gene effects and hence can be putative selective traits (Luiram et al., 2018). Our findings support previous studies wherein they report a comparatively low genotypic and phenotypic variation but high heritability and moderate genetic gain for the dimensions of rhizomes (viz., length and girth of mother, primary and secondary rhizomes) in relation to the weight and number of primary, secondary, and mother rhizomes (Dev & Sharma, 2022; Vinodhini et al., 2018) among the rhizome characters studied. In the light of these previous studies, the highly heritable characters viz. length, core diameter and weight of mother, primary and secondary rhizomes, girth of primary and secondary rhizomes, yield per plant and dry matter recovery (dry yield) were recommended as selective traits in turmeric (Aarthi et al., 2018; Dev & Sharma, 2022; Luiram et al., 2018; Prajapati et al., 2014; Salimath, 2017; Vinodhini et al., 2018). Present study is mostly in alignment with the earlier results and thus affirm these findings.

It is important to acknowledge that the phenotypic evaluation, ploidy analysis, and biochemical assessments represent specific observations under particular environmental conditions and growth stages. Further investigations are necessary to validate and expand these findings across multiple environments and seasons. Additionally, the incorporation of molecular approaches, such as genotyping and transcriptomic analyses, can provide deeper insights into the genetic mechanisms underlying the observed phenotypic variations and assist in identifying key genes or pathways associated with the evaluated traits.

In conclusion, the comprehensive evaluation of phenotypic traits, ploidy analysis, and biochemical assessments in this study yielded valuable insights into the performance, genetic diversity, and biochemical profiles of the investigated genotypes. These outcomes contribute to our understanding of the genotypic potential for crop improvement and facilitate the selection of superior genotypes with desirable agronomic traits. Future research endeavors should focus on leveraging these findings through molecular investigations and multi-environmental evaluations to enhance our understanding of crop genetics and facilitate targeted breeding strategies.

4.7 Objective 2: To genetically characterize the genotypes using available molecular markers and develop novel markers

4.7.1 Molecular characterization with available set of markers: preliminary screening of released varieties

Here, we have screened the released or improved turmeric varieties with 56 microsatellite markers. Out of the 56 microsatellite markers screened, it was found that 55 of them were polymorphic. The alleles per genotype ranged between 1.00 and 3.44 on average for each marker. CuMiSat 08 and UBC 889 had the highest average number of alleles per genotype (3.44 alleles), whereas UBC 896, CuMisat-01, CuMisat-25, CuMisat-31, CuMisat-33, Clest-04, Clest-06, Clest-08, Clest-09, Clest-11, Clest-13, and Clest-17 had the lowest average number of alleles per genotype (1 allele).

The Polymorphic Information Content (PIC) that measure the discriminatory power of a given marker, ranged between 0 and 0.5 for CuMiSat 36 and CuMisat-33 respectively. The mean PIC value obtained is 0.38. The PIC (Polymorphic Information Content) value serves as an indicator of the usefulness of a molecular marker in distinguishing and selecting markers that effectively differentiate between individuals (Gogoi et al., 2023). Hence, the markers employed in our study demonstrate high informativeness and are recommended for molecular investigations in turmeric. Fig.4.5 shows the PCR amplification profiles of 18

improved varieties using various SSR/ISSR primers. Table 4.10 summarises our findings.

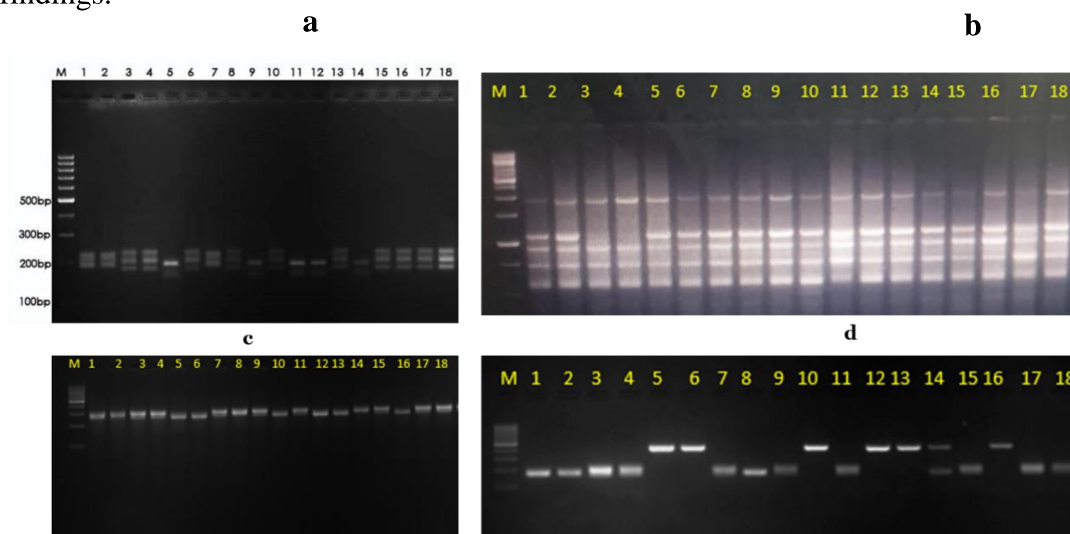


Fig.4.5. PCR amplification profile of 18 improved varieties using CuMisat 19 (a), ISSR-02 (b), CuMisat 13 (c), CuMisat 03 (d). Loading pattern: M-100 bp DNA ladder; 1-Varna; 2-Megha Turmeric; 3-Suranjana; 4-Suguna; 5-Kedaram; 6-Rajendra Sonia; 7-Suvarna; 8-IISR Pragati; 9- IISR Prabha; 10-Sudarsana; 11-IISR Alleppey Supreme; 12-IISR Pratibha; 13-Punjab Haldi 1; 14-BSR-2; 15-CO-1; 16-CO-2; 17-NDH-1; 18-NDH-3.

Table 4.10 Characteristics of SSR and ISSR markers obtained after the PCR screening of varieties

Primer	Allele size range	Total number of alleles	Avg no of alleles per genotype	monomorphic alleles	Polymorphic alleles	PIC
UBC 815	850-1650	5	1.05	0	5	0.32
UBC 818	875-1350	7	1.27	0	7	0.29
UBC 825	750-1600	4	2.61	0	4	0.40
UBC 826	400-1500	10	1.27	0	10	0.38
UBC 834	600-1500	5	1.55	0	5	0.40
UBC 842	500-1900	9	2.05	0	9	0.35
UBC 845	600-2250	4	1.27	0	4	0.40
UBC 850	500-1600	5	2.11	0	5	0.45
UBC 856	750-850	6	1.72	0	6	0.39
UBC 857	490-2000	2	1.11	0	2	0.44
UBC 860	500-1400	8	2.16	0	8	0.37
UBC 866	730-1450	7	1.27	0	7	0.29
UBC 884	300-1200	7	2.38	0	7	0.43
UBC 889	400-1200	8	3.44	0	8	0.46
UBC 896	600-1700	3	1.00	0	3	0.44

UBC 897	1200-1450	5	1.55	0	5	0.40
ISSR-02	600-1100	7	2.00	0	7	0.40
ISSR-03	400-1200	7	2.44	0	7	0.42
ISSR-06	500-1350	5	1.05	0	5	0.32
ISSR-07	600-1400	4	1.22	0	4	0.40
ISSR 13	350-1200	7	2.38	0	7	0.39
ISSR 14	500-1400	7	2.44	0	7	0.42
ISSR 15	600-1450	10	1.72	0	10	0.27
ISSR 17	875-1350	7	1.27	0	7	0.29
Cumisat 01	238–218	4	1.00	0	4	0.38
Cumisat 02	146–118	7	2.38	0	7	0.42
Cumisat 03	200–140	5	1.55	0	5	0.40
Cumisat 04	204–188	7	1.27	0	7	0.29
Cumisat 05	186–160	8	2.00	0	8	0.38
Cumisat 08	178–132	8	3.44	0	8	0.46
Cumisat 13	268–238	7	2.00	0	7	0.40
Cumisat19	204–154	7	2.44	0	7	0.42
Cumisat 20	158–148	7	1.27	0	7	0.29
Cumisat 22	158–122	5	1.55	0	5	0.40
Cumisat 23	165–132	2	1.05	1	1	0.47
Cumisat 25	146–140	4	1.00	0	4	0.38
Cumisat 28	160–139	5	2.11	0	5	0.45
Cumisat 29	177–150	6	1.72	0	6	0.39
Cumisat 31	169–148	3	1.00	0	3	0.44
Cumisat 32	135–120	6	2.11	0	6	0.40
Cumisat 33	170–140	2	1.00	0	2	0.5
Cumisat 36	220–208	1	1.00	1	0	0.00
Cumisat 37	218–192	4	1.27	0	4	0.40
Clest 02	204–152	6	2.00	0	6	0.44
Clest 03	173–133	5	1.44	0	5	0.39
Clest 04	188–172	4	1.00	0	4	0.38
Clest 06	199–191	4	1.00	0	4	0.38
Clest 08	181–154	4	1.00	0	4	0.38
Clest 09	218–184	4	1.00	0	4	0.38
Clest 10	200–188	3	1.50	0	3	0.44
Clest 11	305–292	5	1.00	0	5	0.32
Clest 12	162–140	6	1.55	0	6	0.34
Clest 13	158–136	3	1.00	0	3	0.44
Clest 15	198–162	4	1.27	0	4	0.41
Clest 16	174–164	6	1.44	0	6	0.35
Clest 17	192–174	3	1.00	0	3	0.44

*Avg no – Average number

4.7.2 Genetic similarity analysis

In addition to the allelic variation at the marker level, genetic similarity is an important and useful information which help us to understand genetic relatedness, diversity and to find duplicates if any. In this study, similarity matrix generated based on Jaccard index using SIMQUAL module in NTSYSpc was used for sequential, agglomerative, hierarchical clustering as defined by Sneath and Sokal in SAHN module of NTSYSpc. The Jaccard's similarity coefficients ranged between 0.04 and 1 (Fig. 4.6). The highest similarity coefficient of 1 was observed between Varna and Megha turmeric; Sudarsana and Punjab Haldi. The dendrogram constructed grouped the genotypes into three clusters. Cluster I was the smallest cluster and subdivided into two sub-clusters at similarity of ~0.28. cluster IA comprised of identical genotypes (Varna and Megha Turmeric). Cluster IB contained the variety Suvarna. Cluster II was further split and comprised of IISR Kedaram, IISR Alleppey Supreme, IISR Pratibha, IISR Prabha, BSR 2. Cluster III included majority of the variety under study (10 varieties) at ~0.91 similarity. Cluster III is divided into subclusters III A (Suranjana, Suguna, Rajendra Sonia, IISR Pragati, Sudarsana, Punjab Haldi 1, CO 1, CO2 and NDH 1) and III B (NDH 3).

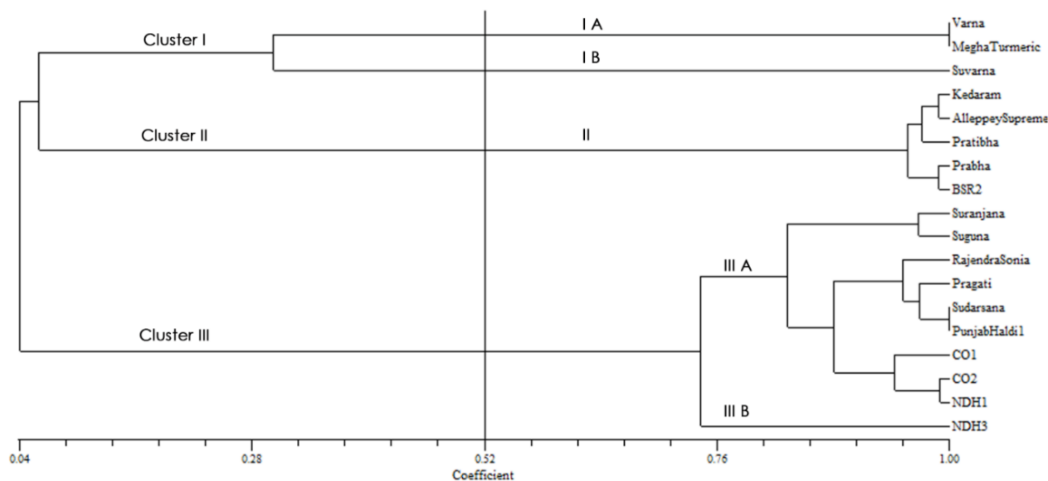


Fig.4.6. Dendrogram based on hierarchical clustering

4.7.3 Identification of most unique or divergent genotypes

As per Jamshidi et al., (2011), the scored data from PCR based molecular screening of genotypes was further subjected to cluster analysis and dendrogram construction for each of individual markers to identify individuals with least relativeness. Here, number of individuals a given individual variety is common in the same clade in each primer's dendrogram was totalled across all 56 primers. Genotypes are arranged from least relativeness to high relativeness as we move from top to bottom of the table. The same is depicted graphically in Fig.4.7. The variety Suvarna shares the least number of clusters with rest of the varieties and hence is the most divergent and unique variety among the varieties under study. This is followed by identical Varna and Megha Turmeric, IISR Kedaram, BSR-2, IISR Pratibha, IISR Alleppey Supreme and IISR Prabha which shares an intermediate number of clusters with rest of the varieties. NDH-3, Suguna, Suranjana, NDH-1, CO-2, Rajendra Sonia, IISR Pragati, CO-1, Sudarsana and Punjab Haldi 1 shares a high number of clusters with the rest of the genotypes and hence have lower divergence. Especially, Sudarsana and Punjab Haldi 1 have the maximum number of shared clusters and hence least unique or divergent of all varieties under study.

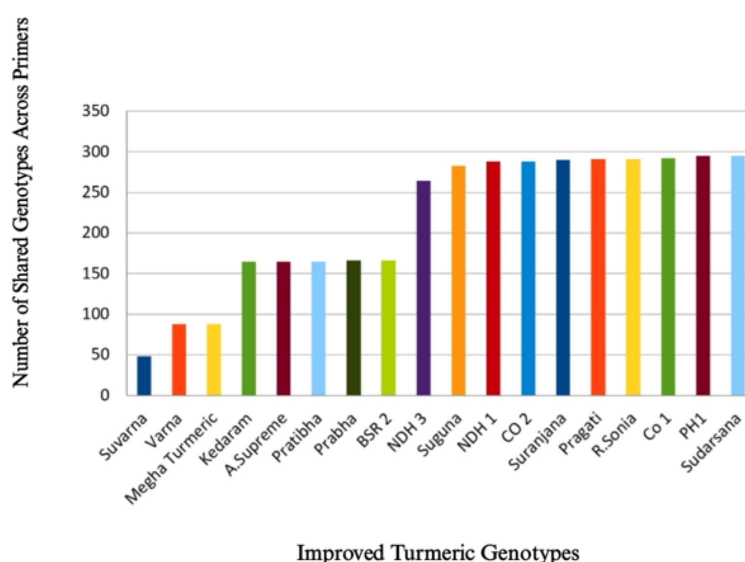


Fig.4.7. Improved turmeric genotypes are shown in X-axis and the total number of individual genotypes shared in the same cluster across all primer's dendrograms by the respective genotype in X-axis is given in y-axis.

4.7.4 Screening of turmeric core collection with SSR markers

The core collection of 93 turmeric genotypes were genotyped using five SSR markers (selected based on initial screening studies in released turmeric varieties) to understand their molecular genetic variation. The markers used were CuMisat 13, CuMisat 19, CuMisat 23, Clest 02 and clest 17.

4.7.4.1 Dendrogram analysis

Dendrogram analysis using NTSYSpc software was conducted, leveraging genetic data from a molecular scored dataset comprising 93 turmeric genotypes. The primary objective of this analysis was to unveil genetic similarities among these species and to organize them into clusters based on these similarities. The resulting dendrogram as depicted in Fig.4.8, visually portrays the hierarchical clustering of the 93 turmeric genotypes, displaying their genetic relationships. In this representation, shorter branch lengths indicate closer genetic ties among species, while longer branches suggest greater genetic divergence.

Fig.4.8. Dendrogram of genetic similarity between 93 turmeric genotypes.

The dendrogram analysis revealed several prominent clusters of plant genotypes. Notably, three significant clusters can be discerned:

Cluster I: This cluster comprises the majority of genotypes and many sub clustering indicating a diverse group of genotypes. It comprises the tetraploids, majority of accessions among other genotypes. All the tetraploids exhibited identical molecular patterns among the markers studied. This cluster also had many genotypes with more or less distinct molecular patterns among the markers studied. Megha Turmeric and Varna (which also show identical banding pattern) along with Acc. 52 (comparatively distinct molecular pattern) are genotypes with high genetic relatedness with tetraploids among all genotypes studied. Acc. 849 (with distinct molecular pattern in most markers) is the nearest subcluster of the above. Suvarna, Kanthi, identical Sona, and Sobha (showed identical banding patterns) along with distinct pattern showing Acc. 286 and Acc. 297 forms the nearby cluster that is genetically similar to the above discussed genotypes. Acc. 821, Acc. 224, Acc. 887 (identical molecular pattern) shares genetic similarity with the variety Reshmi and clusters with identical Dehati Haldi and Acc. 142, Acc. 130 and Acc. 146 (identical banding patterns) shares genetic similarity with Acc. 1 and shares genetic similarity with exotic varieties Nepal turmeric and Uganda turmeric (identical banding patterns). It can be noted that Acc. 161 and Acc. 884 have identical molecular banding patterns indicating high genetic relatedness. It is also observed that released varieties IISR Pratibha, IISR Prabha, IISR Kedaram, IISR Alleppey Supreme, BSR-2 share identical molecular banding pattern with SL 3 and KTS611. It is also observed that Kadappa Local, Duggirala red and Erode turmeric with identical banding pattern share high genetic similarity with Salem Local. Acc. 37 and Acc 902 which are genotypes with distinct molecular pattern share genetic similarity with each other. Acc. 134, Acc. 8, SC 61 and Acc 84 shows identical molecular banding patterns indicating a high genetic relatedness here. These genotypes are genetically similar to genotypes with distinct patterns i.e Rajapuri and Acc. 2, respectively. The identical molecular banding patterns of Acc. 300, Acc. 880, Acc. 23, Acc. 376, Acc. 415 and Kandaila Haldi indicate a high genetic relatedness

among these genotypes. It can be observed that Acc 246, Santra and Waigon turmeric have genetic similarities indicating possible genetic relatedness.

Cluster II: Cluster II comprises two subclusters which is again divided into two clusters each. Most of the genotypes in subclusters show identical molecular pattern here except one or two distinct genotypes. Here, the identical banding patterns of most of the farmers varieties (Hasgova, Hardi, Punjab Haldi-2, Surkha, Futi Halood) with released varieties (Roma, Ranga, Pant Peetabh, CIM Pitambar and Rajendra Sonali) and accessions (Acc 79, Acc 1046 and Acc 200) indicate a high genetic relatedness among them. Palam Pitambar (Acc 1504) with distinct molecular patterns is genetically similar to the above group. It can be noted that both Mydukkur white and Mydukkur Orange showed identical banding patterns and were identical to Acc 269, Acc 313 and Acc 180. Acc 780 showed genetic genetic similarity to the above group containing Mydukkur Orange.

Cluster III: This cluster has the least number of genotypes of all the major clusters. All the genotypes in this cluster share identical molecular patterns for the markers studied. Cluster III with 10 identical genotypes (NDH-1, Punjab Haldi 1, Suranjana, Suguna, Rajendra Sonia, NDH-3, IISR Pragati, CO-1, Sudarsana, and CO-2) reveal a high genetic similarity and relatedness among these genotypes.

4.8 Discussion on results of objective 2

4.8.1 Molecular characterization of released varieties

Molecular characterization of these released varieties has revealed the presence of molecular genetic variation. Even though we have used codominant SSR markers along with dominant ISSR markers for genotyping, there is no information about the location of targeted loci. SSR Primers may target a locus or in some cases multiple loci owing to the polyploid nature of crop leading to multiple alleles per genotype. Hence, characterization of microsatellite loci is complicated in polyploids as allele dosage of SSRs cannot be determined (Pfeiffer et al., 2011). Here as well, the possibility of multilocus amplification and polyploid nature of turmeric (Sigrist et al., 2011) presents difficulties in scoring and analysis of polyploid marker data if we

treat the marker as codominant. So, instead of scoring per locus/primer, we have scored considering each band like a locus and hence scored for the presence (as '1') and absence (as 0) of these bands in a binary format across varieties just as in the case of a dominant marker. Hence, polymorphism information content value of SSR markers were also calculated by treating them as dominant markers like ISSR markers (Rawat et al., 2014). From the dendrogram constructed based on this scored molecular data, we can infer genetic similarity of varieties. Earlier molecular studies involving some of these improved varieties have found that IISR Pratibha, IISR Kedaram, IISR Alleppey Supreme, and Megha turmeric were grouped nearby to each other because of the high similarity among them and the varieties Lakadong and Suvarna gave a unique banding profile (Sahoo et al., 2017). Our results also confirm these findings wherein Varna and Megha turmeric which are identical in all loci assessed here share a similarity coefficient of ~0.28 with Suvarna which showed a unique banding pattern in gel profile for most of the loci tested. These varieties which form Cluster I are the most divergent genotypes with lowest intracluster coefficient of similarity (0.28) and hence broad genetic base compared to genotypes of cluster II (IISR Kedaram, IISR Alleppey Supreme, IISR Pratibha, IISR Prabha, and BSR-2) and cluster III (Suranjana, Suguna, Rajendra Sonia, IISR Pragati, Sudarsana, Punjab Haldi 1, CO-1, CO-2, NDH-1 and NDH-3) which have similarity coefficient of ~0.95 and ~0.76 respectively. This indicates that the genotypes in Cluster II share high relativeness within the cluster (similarity coefficient above 0.9) with highly shared genetic background and hence least divergence followed by genotypes in Cluster II and Cluster I, respectively.

Also, it is found that of all the primers evaluated, varieties Varna and Megha turmeric, Sudarsana and Punjab Haldi 1 respectively had identical genetic background indicating a close genetic relatedness. Suvarna followed by Varna and Megha turmeric form the most divergent group of genotypes which was confirmed by the analysis to find the unique genotypes (Jamshidi & Jamshidi, 2011). In general, we can understand that of all the loci assessed, the genetic similarity of all the improved varieties under study is at least 0.04 (Jaccards similarity coefficient of 0.04) which hints to a broad genetic base collectively. This can be possibly since

these improved genotypes are from diverse locations from India. It can be inferred that in general, there was no relation between the origin of the improved variety and its clustering as most of the varieties from the same place of origin are dispersed among separate groups. This can be explained by the probable exchange of planting material or rhizome between farmers or stakeholders.

4.8.2 Genotyping of core collection

Here, we have characterized the turmeric core collection (93 genotypes) with five dependable and easy to score markers based on our preliminary screening of released varieties. The observed clustering patterns of this turmeric core collection (93 genotypes) based on molecular characterization appear to correspond closely to known morphological and ploidy level relationships, validating the efficacy of our dendrogram analysis. The tight clustering of genotypes for instance, reaffirms their genetic proximity (Jamshidi & Jamshidi, 2011). Conversely, the distinctiveness of genotypes from other genotypes underscores their unique genetic attributes (Rohlf et al., 2009). Cluster I with the largest and highly branched subclusters encompasses the majority of turmeric genotypes. It can be noted that all tetraploids showed identical banding patterns in the markers screened here indicating a highly similar genetic environment among them. Earlier molecular studies that reaffirms our findings involved some of these improved varieties have found that IISR Pratibha, IISR Kedaram, IISR Alleppey Supreme, and Megha turmeric were grouped nearby to each other because of the high similarity among them and the varieties Lakadong and Suvarna gave a unique banding profile (Sahoo, 2017). Cluster II is the second largest cluster. Cluster III with the least number of genotypes of all the major clusters had all genotypes showing identical molecular patterns for the markers studied. With regard to large scale molecular marker based genetic diversity studies in turmeric, only limited reports are available so far (Sahoo et al., 2017; Singh et al., 2012). The clustering pattern observed in these studies revealed the genetic relatedness of the genotypes (Singh et al., 2018; Verma et al., 2015). Although most of these studies had reported geographically independent clustering pattern,

geographic dependent clustering have also been reported in turmeric (Gogoi et al., 2023).

The comprehension of genetic relationships among the diverse set of turmeric genotypes assumes vital importance in germplasm management and conservation endeavours (Islam, 2004). The delineation of these clusters aids in identifying genetically distinct groups, potentially help in tailored conservation measures (Sahoo et al., 2017). Plant breeding initiatives can benefit significantly from these genetic clusters. Breeders can focus their efforts on genotypes within the same genetic group, thereby maximizing genetic compatibility in hybridization programs (Lavudya et al., 2024; Singh et al., 2018).

It is necessary to acknowledge certain limitations associated with this analysis. Our results are reliant solely on molecular genetic data with limited molecular markers, and they may not encapsulate all facets of relatedness among genotypes (Verma et al., 2015). Future investigations may consider augmenting genetic markers which we have attempted here as well by ddRAD sequencing or incorporating ecological data to provide a more comprehensive understanding of genotypic relationships. Moreover, delving into the functional consequences of genetic clustering presents a promising avenue for further research.

In summary, our molecular characterization has unveiled genetic clusters among turmeric diversity panel which cannot be distinguished solely on phenotypic evaluation. These findings hold substantial implications for conservation strategies and plant breeding programs, underscoring the valuable insights this genetic analysis can offer in the realm of turmeric crop improvement. As the accuracy of clustering depends on the quality and quantity of genetic markers employed, we have undertaken an NGS based sequencing analysis which incorporates a large molecular data for the purpose of marker trait study.

4.9 Objective 3: To determine the association of rhizome characters and quality traits with molecular markers

We have carried out association analysis or association mapping to find molecular markers associated with traits of interest. Based on our phenotypic characterization as well as based on earlier field trial observations, we have narrowed our traits of interest to three characters, i.e. rhizome characters (length of primary rhizome per plant and girth of primary rhizome per plant) and quality trait (curcuminoid content). As association analysis requires millions of markers, we have utilized NGS based genotyping strategy of ddRAD sequencing. For genotyping purpose, we have narrowed our sample size into 51 genotypes based on their phenotypic and molecular characterization.

4.9.1 Genotyping

4.9.1.1 ddRAD sequencing and bioinformatic analysis

Genomic DNA was isolated from each of 51 turmeric genotypes and utilized for ddRAD library preparation and sequencing (Fig.4.9a and Fig.4.9b). Sequencing data from ddRAD libraries was further utilized for a comprehensive investigation of the genetic diversity and population structure of the turmeric genotypes. A total of 51 individuals were included in the study out of 93 genotypes, and after demultiplexing and quality filtering, we obtained 164,380,624 raw sequencing reads. The reads had an average length of 100-150 base pairs (bp), resulting in a total of 1.5 billion bp of high-quality data for downstream analysis.

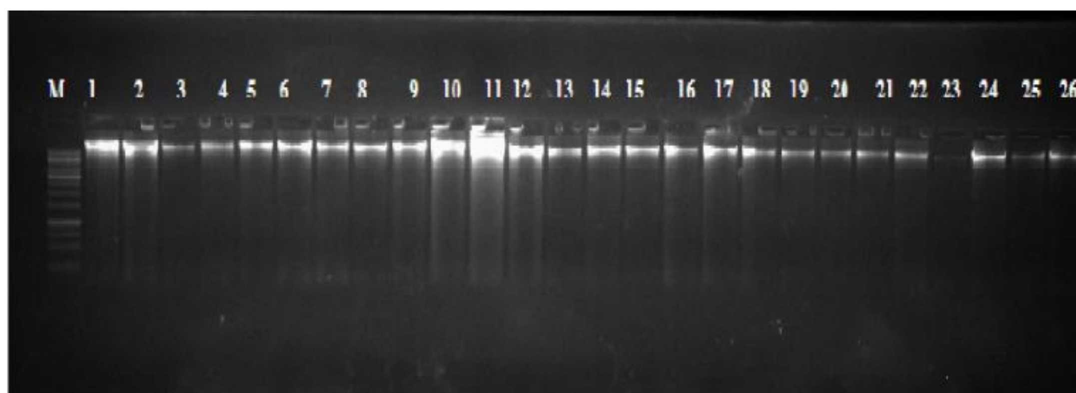


Fig. 4.9a Agarose gel electrophoresis showing QC of samples

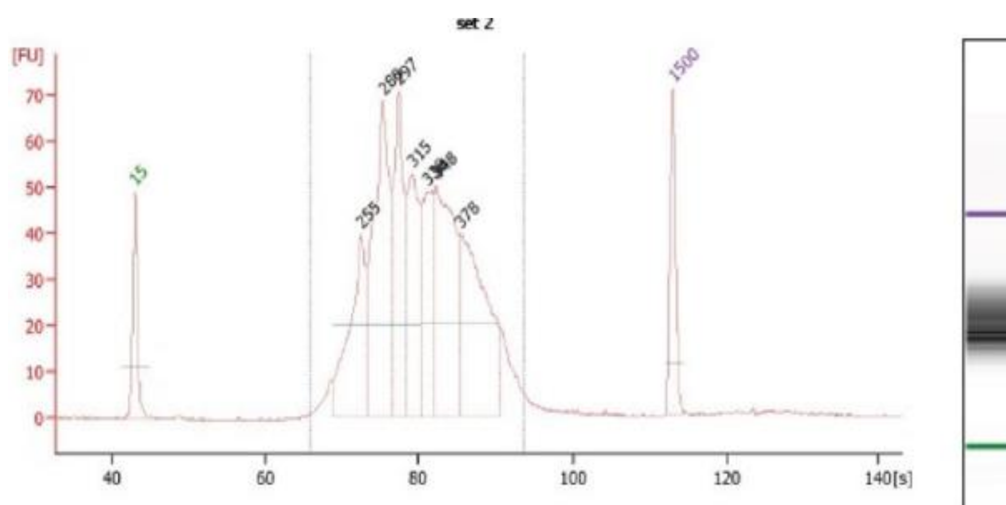


Fig. 4.9b. Bioanalyzer image of sample

4.9.1.2 FastQC Check for NGS Data Quality

The raw sequencing data from 51 ddRAD samples obtained via the Illumina HiSeq platform underwent meticulous quality control assessment using FastQC. This analysis provided invaluable insights into the overall quality and integrity of the NGS data. The results obtained are shown in Fig.4.10 to Fig.4.15.

The average Phred score of above 30 for over 90% of bases indicated robust base calling accuracy (Fig.4.10). A minor decline in quality towards the 3' end of the reads was detected, but it fell within acceptable parameters, posing no substantial impact on downstream analysis.

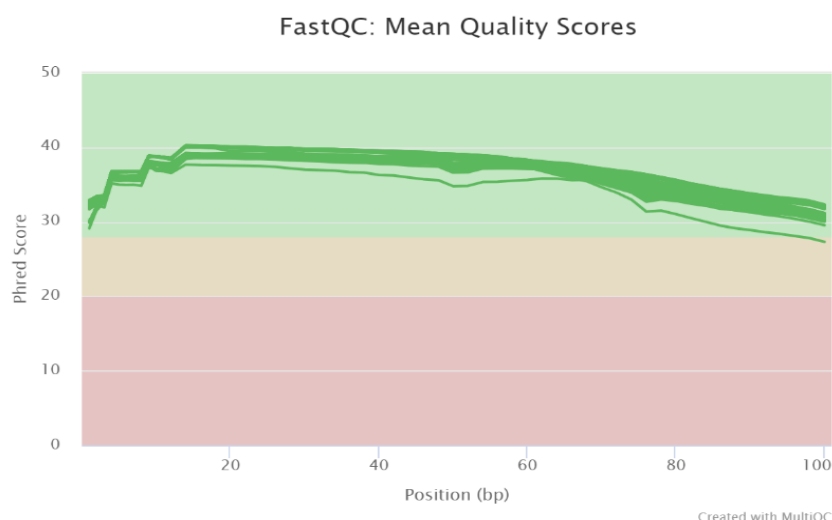


Fig. 4.10 Per base sequence quality analysis, demonstrating consistently high-quality scores across the reads.

Most reads exhibited high-quality scores above 30, ensuring reliable base calling accuracy (Fig.4.11). Although a small fraction of bases had slightly lower quality scores, they were evenly dispersed throughout the reads and did not pose significant concerns.

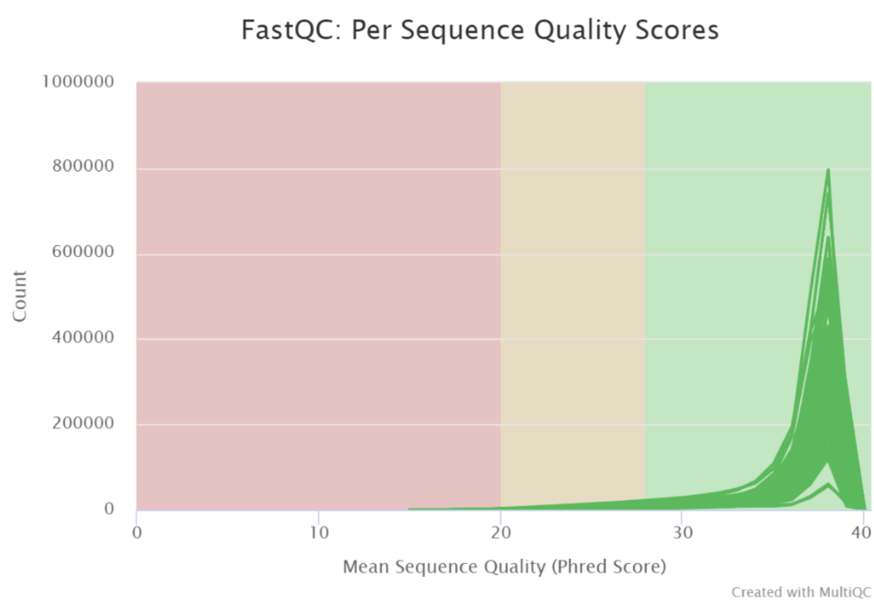


Fig. 4.11 Distribution of quality scores

The read length ranged from 75 to 150 base pairs (Fig 4.12) , with an average length of approximately 100 base pairs. No anomalies or biases were detected, affirming the absence of technical issues during sequencing.

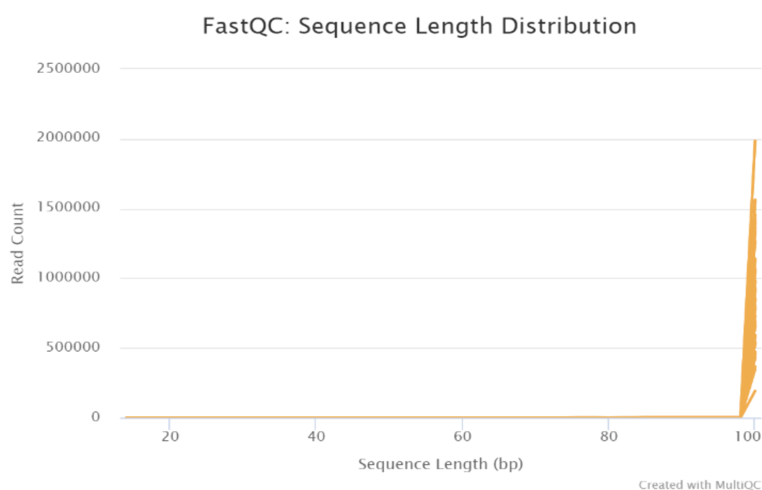


Fig. 4.12 Sequence length distribution

Based on the GC content distribution (Fig 4.13), the average GC content was found approximately as 40%, reflective of the organism's genomic composition.

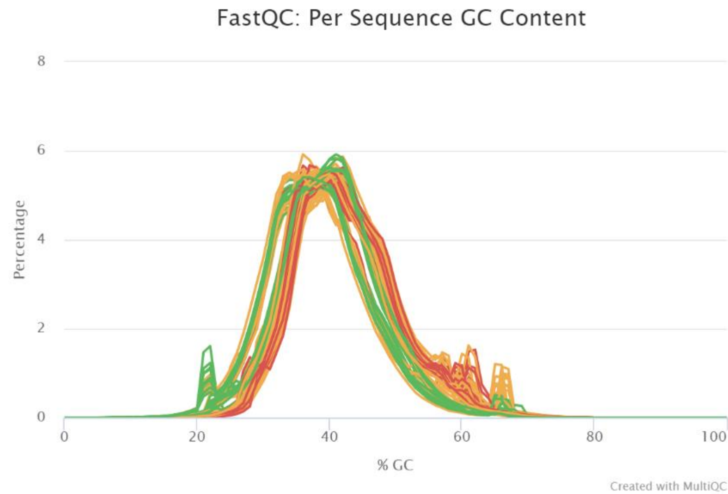


Fig. 4.13. GC content distribution

The stacked bar plot (Fig 4.14) demonstrated negligible levels of adapter contamination, with adapter-containing reads accounting for less than 1% of the dataset. This outcome highlighted the successful removal of adapter-related artifacts during the data preprocessing phase.

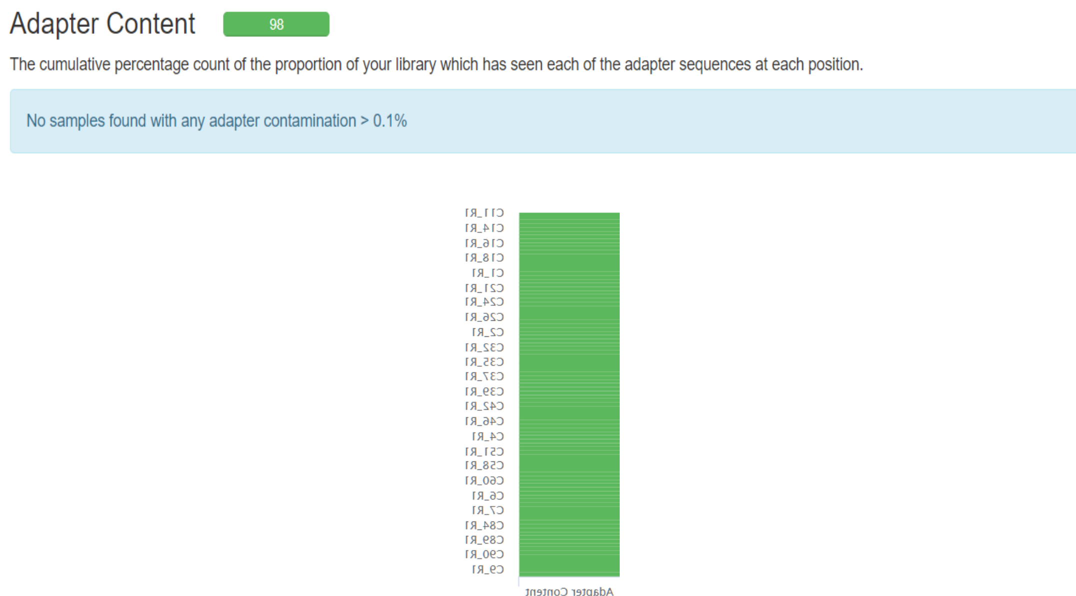


Fig. 4.14. Adapter contamination analysis

Detecting and addressing overrepresented sequences is another crucial step for ensuring the accuracy of downstream analyses such as alignment, assembly, and variant calling (Fig.4.15).

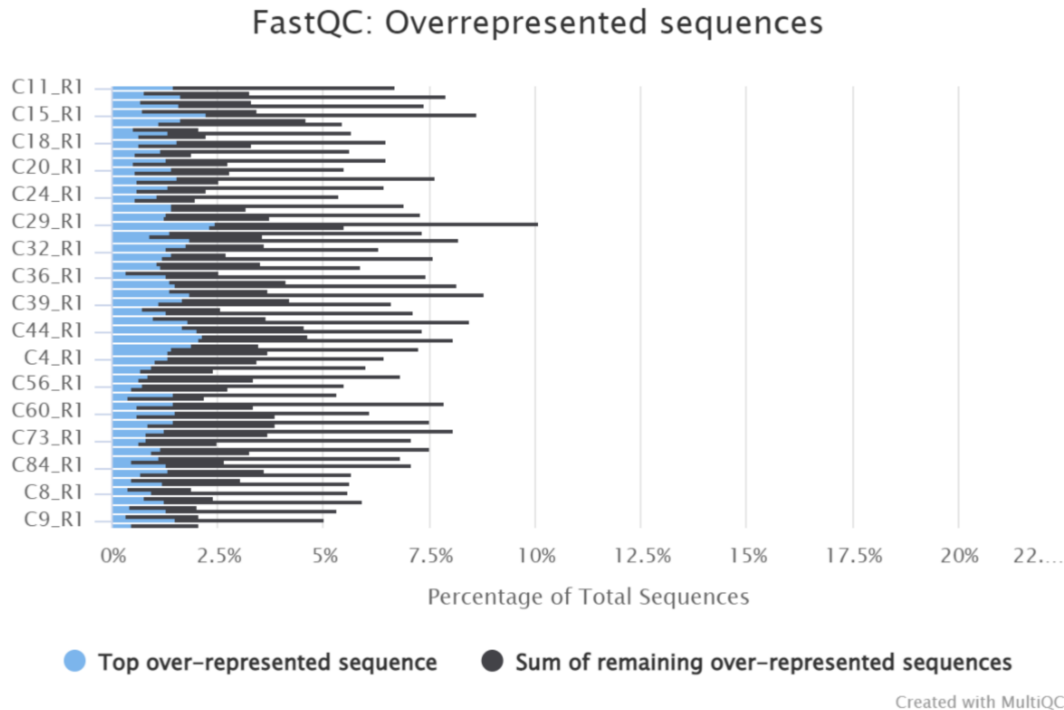


Fig. 4.15 Overrepresented sequences

In summary, quality assessment by FastQC of the sequencing data revealed promising results. The average Phred quality score was 30, indicating a high level of base call accuracy (99.9%). The GC content of the reads was determined to be 40%, which closely matched the expected value for the species under investigation. Duplicate reads were identified but not removed and resulting dataset was utilized for subsequent analysis.

4.9.1.3 BWA Alignment to reference genome

The NGS data obtained from the ddRAD samples were aligned to the reference turmeric genome (PRJNA660606) using the BWA-MEM alignment algorithm (Chakraborty et al., 2021). The alignment process provided valuable insights into the mapping efficiency, uniqueness, and coverage of the aligned reads. Results are given below in Fig. 4.16:

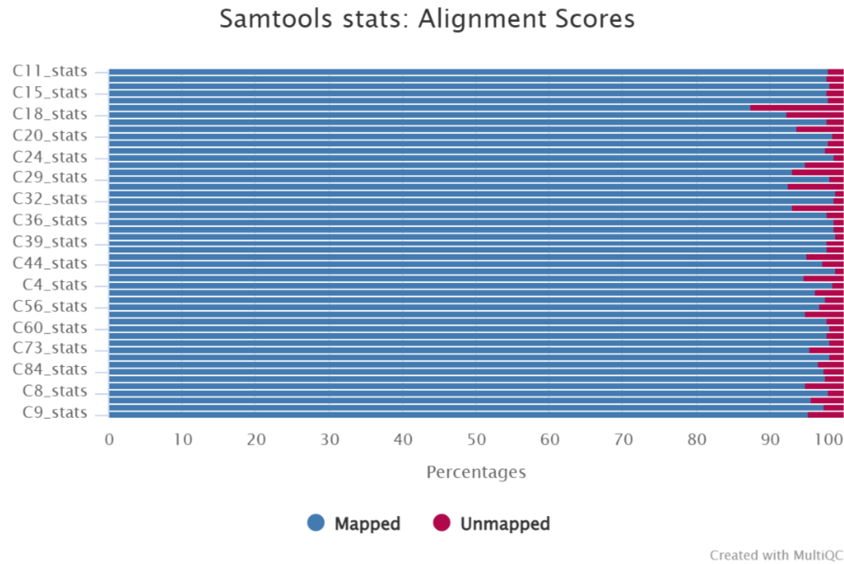


Fig. 4.16: BWA alignment statistics

The BWA alignment resulted in the mapping of a total of 10 million reads to the reference genome. Of the mapped reads, 80% were uniquely aligned to a specific location, ensuring high-confidence mapping. The remaining 20% of the mapped reads had multiple alignments, indicating potential repetitive regions or homologous genes. Additionally, 5% of the reads were either unmapped or had low mapping qualities. The overall mapping rate of 95% demonstrates the efficiency of the alignment process and the successful mapping of the majority of reads to the reference genome.

4.9.1.4 Variant calling

The NGS data obtained from the 51 genotypes underwent variant calling using the GATK tool. This analysis aimed to identify genetic variations within the polyploid genome and provide insights into its genomic landscape.

The GATK variant calling analysis identified variants across the polyploid genome. These variants encompassed both single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). Most of the detected variants (70%) were SNPs, while the remaining 30% were indels. These findings offer an overview of the genetic variations present within the polyploid genome. Filtering of variants resulted in removal of indels. The resultant VCF file is shown below in Fig. 4.17.

##source=CombineGVCFs
##source=SelectVariants

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	C1	C11	C13	C14	C15	C16	C17	C18	C19	C2	C20	C21	C23	C24
scaffold1_	102007	.	C	T	738.63	PASS	AC=18;AF=	GT:AD:DP:	1/1/1:0,1:	1/1/1:0,3:	././.:1,0:1:	0/0/0:7,0:	././.:0,0:0:	0/0/0:7,0:	0/0/0:1,0:	0/0/0:6,0:	0/1/1:1,1:	1/1/1:0,4:	0/0/0:14,0	0/0/0:18,0	././.:0,0:0:	0/0/0:3,0:
scaffold1_	102129	.	T	G	5216.59	PASS	AC=55;AF=	GT:AD:DP:	1/1/1:0,2:	1/1/1:0,3:	././.:1,0:1:	0/1/1:3,4:	././.:0,0:0:	0/1/1:3,4:	0/0/0:2,0:	0/1/1:2,4:	1/1/1:0,2:	1/1/1:0,4:	0/1/1:6,8:	././.:19,0:1	././.:0,0:0:	0/0/0:3,0:
scaffold1_	102137	.	T	C	5204.48	PASS	AC=55;AF=	GT:AD:DP:	1/1/1:0,2:	1/1/1:0,3:	././.:1,0:1:	0/1/1:3,4:	././.:0,0:0:	0/1/1:3,4:	././.:2,0:2:	0/1/1:2,4:	1/1/1:0,2:	1/1/1:0,4:	0/1/1:6,8:	././.:19,0:1	././.:0,0:0:	0/0/0:3,0:
scaffold1_	105644	.	G	T	1301.76	PASS	AC=52;AF=	GT:AD:DP:	0/0/1:1,1:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,1:	././.:0,0:0:	0/1/1:1,2:	1/1/1:0,2:	0/0/0:2,0:	0/0/1:2,2:	1/1/1:0,1:	1/1/1:0,3:	././.:4,0:4:	1/1/1:0,1:	1/1/1:0,2:
scaffold1_	105658	.	T	C	871.75	PASS	AC=39;AF=	GT:AD:DP:	0/0/1:1,1:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,1:	././.:0,0:0:	0/1/1:1,2:	1/1/1:0,2:	0/0/0:2,0:	0/0/1:2,2:	1/1/1:0,1:	0/0/0:3,0:	0/0/0:5,0:	1/1/1:0,1:	././.:0,0:0:
scaffold1_	105678	.	T	A	1108.39	PASS	AC=49;AF=	GT:AD:DP:	0/0/1:1,1:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,1:	././.:0,0:0:	0/1/1:1,2:	1/1/1:0,3:	0/0/0:2,0:	0/0/1:2,2:	1/1/1:0,1:	1/1/1:0,2:	././.:6,0:6:	1/1/1:0,1:	././.:0,0:0:
scaffold1_	105705	.	G	A	920.14	PASS	AC=41;AF=	GT:AD:DP:	0/0/1:1,1:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,1:	././.:0,0:0:	0/1/1:1,2:	1/1/1:0,3:	0/0/0:2,0:	0/0/0:3,0:	1/1/1:0,1:	././.:0,0:0:	0/0/0:6,0:	1/1/1:0,1:	././.:0,0:0:
scaffold1_	115054	.	A	T	558.44	PASS	AC=7;AF=(	GT:AD:DP:	0/0/0:4,0:	././.:0,0:0:	././.:2,0:2:	0/0/0:2,0:	././.:2,0:2:	0/0/0:2,0:	././.:0,0:0:	0/0/0:1,0:	0/0/0:1,0:	0/0/0:1,0:	0/0/0:3,0:	0/0/0:8,0:	0/0/0:2,0:	0/0/0:3,0:
scaffold1_	122908	.	A	T	534.78	PASS	AC=30;AF=	GT:AD:DP:	0/0/0:8,0:	0/0/0:2,0:	0/0/0:2,0:	././.:1,0:1:	0/0/0:1,0:	1/1/1:0,1:	0/0/0:1,0:	1/1/1:0,2:	0/0/0:3,0:	0/0/0:2,0:	1/1/1:0,3:	1/1/1:0,2:	././.:0,0:0:	././.:0,0:0:
scaffold1_	148684	.	C	G	104.98	PASS	AC=9;AF=(	GT:AD:DP:	0/0/0:1,0:	0/0/0:3,0:	././.:1,0:1:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	././.:0,0:0:	0/0/0:2,0:	0/0/0:4,0:	0/0/0:5,0:	0/0/0:5,0:	././.:0,0:0:	././.:0,0:0:
scaffold1_	148700	.	C	T	173.74	PASS	AC=9;AF=(	GT:AD:DP:	0/0/0:2,0:	0/0/0:2,0:	././.:1,0:1:	0/0/0:1,0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	././.:0,0:0:	0/0/0:2,0:	0/0/0:4,0:	0/0/0:4,0:	0/0/0:5,0:	././.:0,0:0:	././.:0,0:0:
scaffold1_	148701	.	A	G	175.02	PASS	AC=9;AF=(	GT:AD:DP:	0/0/0:2,0:	0/0/0:2,0:	././.:1,0:1:	0/0/0:1,0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	././.:0,0:0:	0/0/0:2,0:	0/0/0:4,0:	0/0/0:4,0:	0/0/0:5,0:	././.:0,0:0:	././.:0,0:0:
scaffold1_	191358	.	T	A	1858.83	PASS	AC=26;AF=	GT:AD:DP:	././.:2,0:2:	././.:1,0:1:	././.:0,0:0:	././.:0,0:.:	././.:0,0:0:	0/0/0:3,0:	0/0/1:4,1:	0/0/0:3,0:	0/0/0:1,0:	././.:0,0:0:	0/0/1:7,3:	././.:0,0:0:	0/0/1:4,2:	././.:0,0:0:
scaffold1_	205399	.	T	C	2354.17	PASS	AC=49;AF=	GT:AD:DP:	1/1/1:0,2:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,2:	././.:1,0:1:	0/0/1:10,2	././.:0,0:0:	1/1/1:0,1:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	0/0/0:6,0:
scaffold1_	205547	.	C	T	1271.24	PASS	AC=25;AF=	GT:AD:DP:	0/0/0:1,0:	././.:2,0:2:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	././.:0,0:0:	././.:1,0:1:	././.:0,0:0:	1/1/1:0,1:	1/1/1:0,3:	0/0/0:1,0:	0/0/0:3,0:
scaffold1_	205588	.	G	C	1431.55	PASS	AC=22;AF=	GT:AD:DP:	././.:0,0:0:	././.:0,0:0:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	././.:1,0:1:	1/1/1:0,3:	0/0/0:1,0:	0/0/0:3,0:
scaffold1_	216522	.	A	G	240.32	PASS	AC=7;AF=(	GT:AD:DP:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	1/1/1:0,3:	././.:0,0:0:	0/0/0:1,0:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	././.:1,0:1:	././.:0,0:.:	0/0/0:1,0:	0/0/0:2,0:
scaffold1_	216535	.	C	A	809.34	PASS	AC=18;AF=	GT:AD:DP:	././.:0,0:0:	././.:0,0:0:	././.:1,0:1:	1/1/1:0,5:	././.:0,0:0:	0/0/0:1,0:	././.:2,0:2:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	././.:1,0:1:	1/1/1:0,2:	0/0/0:1,0:	0/0/0:2,0:
scaffold1_	216553	.	C	A	897.78	PASS	AC=23;AF=	GT:AD:DP:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	1/1/1:0,5:	././.:0,0:0:	0/0/0:1,0:	././.:2,0:2:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	1/1/1:0,2:	0/0/0:1,0:	0/0/0:2,0:
scaffold1_	216565	.	A	G	907.8	PASS	AC=23;AF=	GT:AD:DP:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	1/1/1:0,5:	././.:0,0:0:	0/0/0:1,0:	././.:2,0:2:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	././.:0,0:0:	1/1/1:0,2:	0/0/0:1,0:	0/0/0:2,0:
scaffold1_	220966	.	A	G	3339.24	PASS	AC=32;AF=	GT:AD:DP:	0/0/1:5,5:	0/0/1:5,3:	0/1/1:1,2:	0/0/0:2,0:	././.:1,0:1:	0/0/0:2,0:	1/1/1:0,5:	0/0/0:5,0:	././.:0,0:.:	0/1/1:1,4:	0/0/0:8,0:	0/0/0:11,0	0/1/1:2,8:	0/0/0:6,0:
scaffold1_	279544	.	G	A	339.06	PASS	AC=5;AF=(	GT:AD:DP:	0/0/0:9,0:	0/0/0:3,0:	././.:8,0:8:	././.:6,0:6:	././.:0,0:0:	././.:4,0:4:	././.:0,0:.:	0/0/0:7,0:	././.:0,0:0:	0/0/0:3,0:	././.:0,0:.:	././.:9,0:9:	././.:0,0:.:	0/0/0:2,0:
scaffold1_	298405	.	C	G	635.04	PASS	AC=17;AF=	GT:AD:DP:	0/0/0:2,0:	././.:1,0:1:	0/0/0:4,0:	1/1/1:0,2:	0/0/0:1,0:	././.:2,0:2:	0/0/0:3,0:	0/0/0:2,0:	././.:0,0:0:	0/0/0:1,0:	0/0/0:4,0:	././.:8,0:8:	0/0/0:4,0:	1/1/1:0,2:
scaffold1_	298525	.	A	G	2897.45	PASS	AC=44;AF=	GT:AD:DP:	1/1/1:0,2:	././.:1,0:1:	././.:0,0:.:	0/0/0:3,0:	1/1/1:0,1:	././.:2,0:2:	1/1/1:0,3:	././.:2,0:2:	././.:0,0:0:	././.:0,0:.:	0/0/1:5,4:	././.:0,0:.:	././.:0,0:.:	0/0/0:3,0:
scaffold1_	311508	.	T	C	79.78	PASS	AC=6;AF=(	GT:AD:DP:	0/0/0:10,0	././.:0,0:0:	././.:0,0:0:	0/0/0:2,0:	././.:0,0:0:	././.:0,0:0:	0/0/0:1,0:	0/0/0:2,0:	././.:0,0:0:	0/0/0:1,0:	0/0/0:2,0:	././.:0,0:0:	0/0/0:2,0:	0/0/0:1,0:
scaffold1_	311806	.	T	C	2296.46	PASS	AC=35;AF=	GT:AD:DP:	0/0/1:5,3:	0/1/1:3,4:	1/1/1:0,4:	0/0/0:3,0:	1/1/1:0,2:	0/0/0:1,0:	1/1/1:0,2:	0/0/0:6,0:	0/0/0:3,0:	1/1/1:0,1:	0/0/0:7,0:	0/0/0:3,0:	1/1/1:0,7:	0/0/0:1,0:

Fig. 4.17. VCF file obtained after GATK variant calling

In conclusion, the GATK variant calling analysis performed on the polyploid NGS data successfully identified 30438 SNP variants. The variant allele frequency spectrum demonstrated a broad range of allele frequencies, highlighting the genetic diversity within the polyploid genome. The variant annotation analysis can further provide insights into the potential functional impact of the identified variants. These results contribute to our understanding of the genomic landscape and genetic variations within the polyploid species. VCF file was then sorted in HapMap (HMP) format in TASSEL software for further downstream bioinformatics analysis (Fig.4.18).

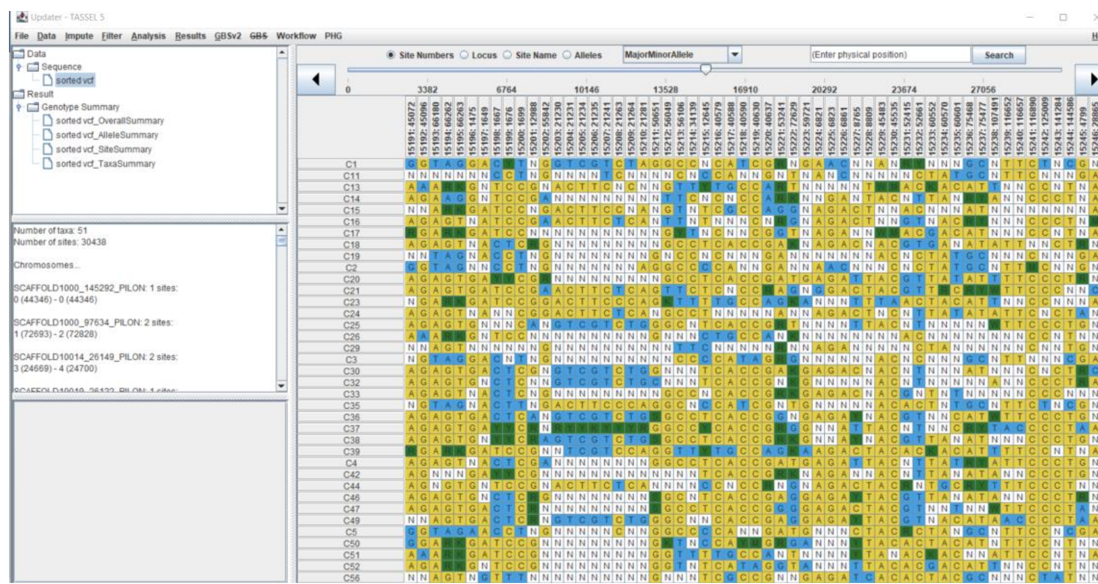


Fig. 4.18 VCF file sorted in HapMap format using TASSEL

4.10 Population structure analysis using STRUCTURE software

The population structure of the 51 genotypes was investigated using the STRUCTURE software. The dataset consisted of 51 turmeric genotypes and 30,438 SNP markers. The analysis was performed using the STRUCTURE software version.2.3.4. The results obtained are shown in Fig.4.19.

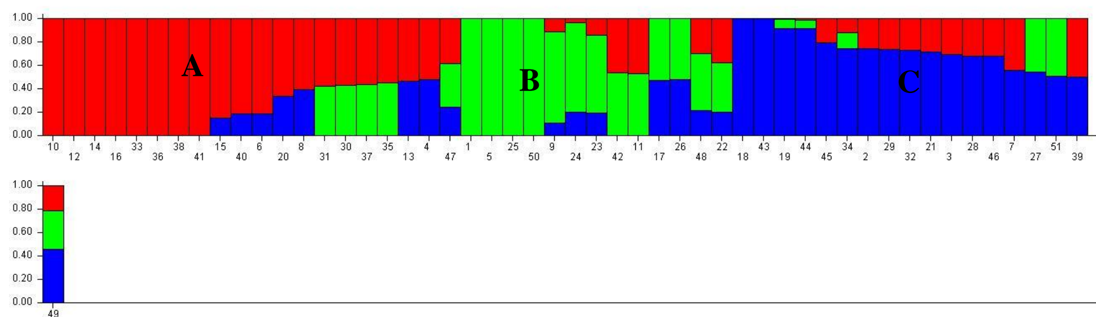


Fig. 4.19: Bar plot of population structure

The results of the STRUCTURE analysis unveiled three distinct population clusters (A; red, B; green and C; blue) and varying levels of genetic admixture within the studied population. The bar plot demonstrated the inferred population membership proportions, revealing genetic differentiation and admixture patterns. Individuals sharing similar colour patterns indicated shared ancestry, while those with mixed colours indicated genetic admixture.

Table 4.11 below shows the population membership proportions (Q values) for each individual genotype. Individuals were given in rows, while assumed populations were shown in columns. The Q values denote the estimated ancestry proportions for each individual in the inferred populations. Thus, the population membership matrix provides the detailed information on the estimated ancestry proportions of individuals in the assumed populations. This matrix shed light on the genetic contribution from different populations for each individual.

Table 4.11 Population membership matrix

Genotypes	Q1	Q2	Q3
Reshmi	0.00	1.00	0.00
Suranjana	0.00	0.45	0.55
Ranga	0.25	0.00	0.75
Varna	0.32	0.00	0.68
Rajendra Sonia	0.30	0.00	0.70
Suvarna	0.26	0.00	0.74
Sudarsana	0.52	0.00	0.48
Suguna	0.57	0.43	0.00
Sona	0.00	1.00	0.00
Suroma	0.58	0.42	0.00
Sobha	0.81	0.00	0.19
NDH-1	0.27	0.00	0.73
IISR Prabha	0.44	0.00	0.56
IISR Pratibha	1.00	0.00	0.00
IISR Alleppey Supreme	0.61	0.00	0.39
IISR Kedaram	0.12	0.14	0.75
Megha Turmeric	0.11	0.78	0.11
Kanthi	0.54	0.46	0.00
Acc 79	1.00	0.00	0.00
Acc 146	1.00	0.00	0.00
Acc 130	0.47	0.53	0.00
Acc 200	0.56	0.44	0.00
Acc 376	1.00	0.00	0.00
Acc 142	1.00	0.00	0.00
Punjab Haldi 1	0.53	0.00	0.47

Punjab Haldi 2	0.50	0.00	0.50
CO-1	1.00	0.00	0.00
Acc 23	0.81	0.00	0.19
Acc 52	0.85	0.00	0.15
Acc 134	1.00	0.00	0.00
Acc 449	1.00	0.00	0.00
Acc 1	0.46	0.54	0.00
BSR-2	0.00	0.53	0.47
Acc 2	0.00	0.00	1.00
Acc 8	0.00	0.00	1.00
Acc 19	0.01	0.07	0.92
Acc 224	0.00	0.08	0.92
Acc 821	0.20	0.00	0.80
Rajapuri	0.66	0.00	0.34
Acc 887	0.32	0.00	0.68
Acc 902	0.28	0.00	0.72
Pant Peetabh	0.38	0.37	0.25
Acc 1504	0.38	0.42	0.20
Roma	0.30	0.48	0.22
Acc 84	0.14	0.66	0.20
Acc 161	0.21	0.33	0.46
Acc 269	0.04	0.76	0.21
Acc 286	0.00	1.00	0.00
Duggirala Red	0.00	1.00	0.00
Acc 297	0.00	0.49	0.51
Acc 313	0.00	0.52	0.48

The ΔK plot obtained from structure analysis (Fig.4.20) showed the rate of change in the log probability of data for successive K values. This plot helps to find the optimal number of populations that best describes the genetic structure. The peak at a specific K value gives the most likely number of populations found.

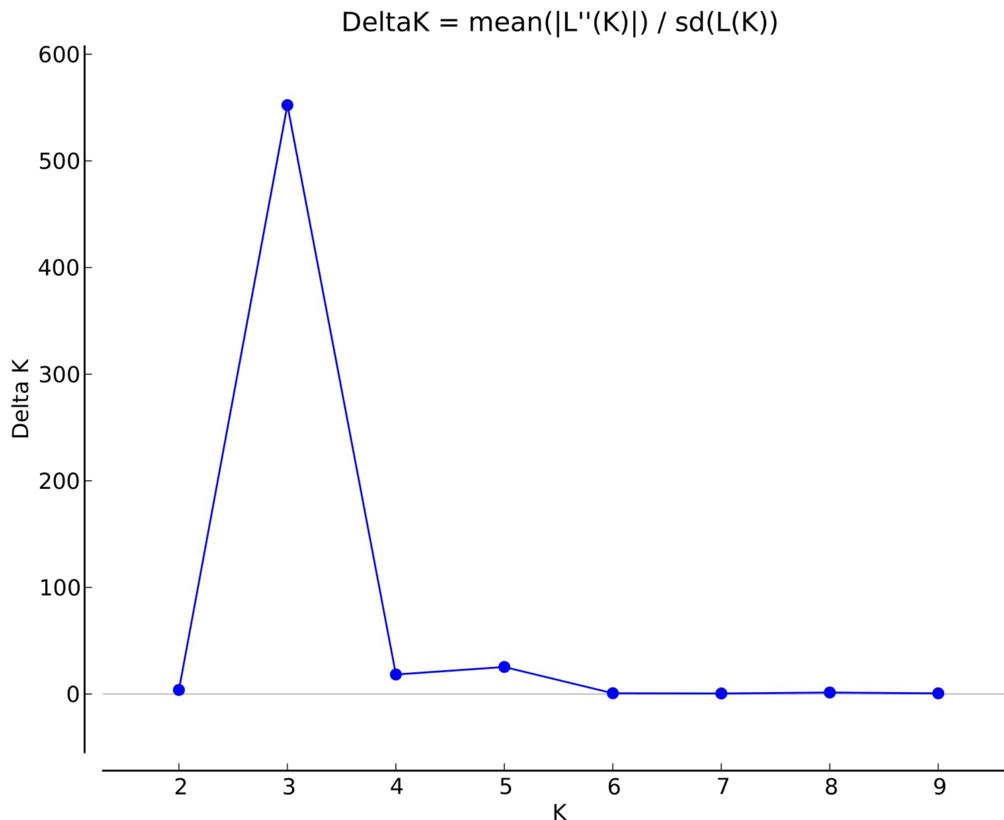


Fig.4.20. ΔK Plot showing the optimum number of subpopulation present.

The ΔK plot indicated the optimal number of populations that captured the underlying genetic structure. The peak at a specific K value (K =3) indicated the most suitable number of assumed populations.

In Conclusion, these findings from population structure analysis suggested the presence of genetic substructure and admixture within the population. Further exploration and interpretation of these population structure patterns can be conducted in relation to known population history, geographical factors, or other relevant biological aspects.

4.11 Phenotypic data for association analysis

4.11.1 BLUP value generation through Meta-analysis using R

For the investigated phenotypic traits, a meta-analysis was done to get the Best Linear Unbiased Prediction (BLUP) values. Data of relevant traits from two years field trial (2020 and 2021) was utilized in the meta-analysis. The BLUP values generated are given in Table 4.12.

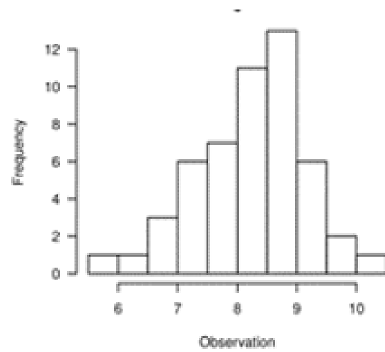
Table 4.12. Meta-analysis results - BLUP values

Environment	Statistic	BLUP values for curcuminoid content	BLUP values for length of Primary rhizomes	BLUP values for girth of Primary rhizomes
Overall	Heritability	1.000	0.983	0.989
Overall	Genotype variance	2.951	0.877	0.988
Overall	GenxLoc variance	0.000	0.000	0.000
Overall	Residual variance	0.003	0.061	0.045
Overall	Grand mean	3.295	8.159	7.236
Overall	LSD	0.078	0.404	0.370
Overall	CV	1.661	3.026	2.924
Overall	n Replicates	2.000	2.000	2.000
Overall	n Environments	2.000	2.000	2.000
Overall	Genotype significance	0.000	5.40418E-3 9	0.000
Overall	GenxEnv significance	1.000	1.000	1.000

The computed BLUP values from the meta-analysis for each analysed characteristic are shown in the table. Each column represents a study that contributed to the meta-analysis, and each row represents a particular feature. The BLUP values given correspond to the predicted phenotypic values for the relevant trait of interest.

The produced BLUP values provide valuable information on how the examined traits perform across various genotypes and environment. These values provide

more precise estimates of the genuine phenotypic values since they consider random effects and potential confounders. Fig.4.21, 4.22 and 4.23 shows the phenotypic distribution of the 51 turmeric genotypes for the traits under study. It can be seen that the genotypes follow a normal distribution for all the trait except for the trait of curcuminoid content. So, the phenotypic data was transformed and normalised prior to the marker trait association study.



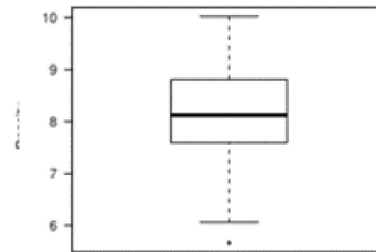
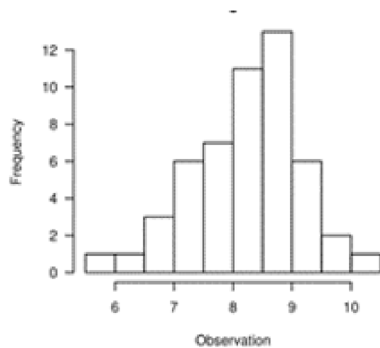


Fig.4.21 a **Fig.4.21 b**

Fig.4.21. Distribution of length of primary rhizomes in histogram and box plot



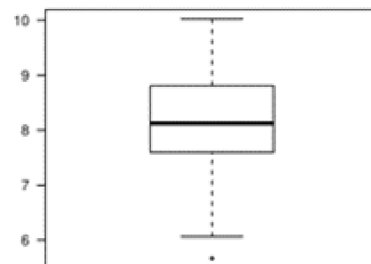


Fig.4.22 a **Fig.4.22 b**

Fig.4.22. Distribution of girth of primary rhizomes in histogram and box plot

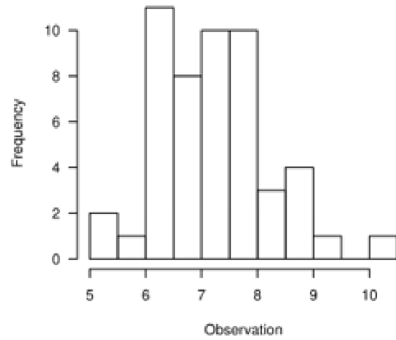


Fig.4.23a

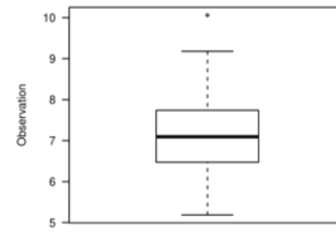


Fig.4.23b

Fig. 4.23. Distribution of curcuminoid content depicted in histogram and box plot

A final dataset of 30438 high-quality SNPs was produced after the implementation of preliminary quality control steps to ensure data integrity. Based on the structure analysis, it was discovered that there were three distinct subpopulations. Kinship was included as a fixed effect in the association analysis model (BLINK model) to account for this population structure.

4.12 Marker-trait association analysis using GAPIT and the BLINK model

In this work, we investigated the genetic basis of trait variation in our population using an SNP marker trait association analysis with the GAPIT software. The dataset focused on three main traits: length of primary rhizomes per plant, girth of primary rhizomes per plant and curcuminoid content. It included phenotypic data from 51 distinct genotypes of turmeric and 30438 SNP markers. Here, genome-wide association analysis was conducted using the GAPIT R package and the BLINK (Bayesian Logistic Interaction Kernel) model to find genetic relationships with the phenotypic traits under study. The research sought to identify potential genetic variations linked to the desired phenotypic features. To identify SNP markers significantly associated with traits of our interest, BLUP values derived from transformed trait measurements over 2 years were utilized as input phenotypic data for GWAS analysis. BLINK algorithm utilized here is a model selection technique inspired by FarmCPU. BLINK employs a bin approach to prevent marker selection from the same genomic regions and does not assume an equal distribution of causal

genes across the genome. Instead, it utilizes the linkage disequilibrium (LD) method to enhance statistical power (Huang et al., 2019). Markers were ranked based on their association significance, and redundant markers were eliminated if they were in LD with the most associated marker, through an iterative process. BLINK adopted a fixed-effect model, approximating maximum likelihood using Bayesian Information Content (BIC), thereby reducing computational complexity. In GWAS conducted *via* GAPIT software, BLINK analysis incorporated structure and kinship information as covariates. The default P-value threshold employed by BLINK was a Bonferroni-corrected threshold of 0.01, with a -log10P value threshold set at ≥ 5.00 for this analysis. The marker-trait associations identified by BLINK models for each trait in turmeric are summarized in Table 4.13. These associations indicate potential genetic variants influencing the phenotypic variation.

Table 4.13. Significant marker trait associations

Character	Chromosome	Position	P value-overall	Alleles	Bonferonni correction value
Curcumin content	SCAFFOLD89_247814_PILON	195259	0.000150202	G/T	3.823324284
Rhizome length	SCAFFOLD7_640062_PILON	36714	3.89E-05	C/G	4.410050399
Rhizome length	SCAFFOLD7_640062_PILON	36731	4.56E-05	G/T	4.341035157
Rhizome length	SCAFFOLD7_640062_PILON	36732	4.56E-05	C/A	4.341035157
Rhizome length	SCAFFOLD7_640062_PILON	36736	4.56E-05	T/A	4.341035157
Rhizome length	SCAFFOLD1660_108323_PILON	33359	8.30E-05	G/A	4.080921908
Rhizome girth	SCAFFOLD1658_73987_PILON	153029	7.43E-05	C/T	4.129011186
Rhizome girth	SCAFFOLD20303_6028_PILON	5594	9.93E-05	C/T	4.003050752
Rhizome girth	SCAFFOLD716_115384_PILON	4272	0.000135505	C/T	3.868044679

Thus, the association analysis using the BLINK model in GAPIT identified several significant SNP-trait associations. The Manhattan plot (Fig.4.24, 4.25 and 4.26) illustrates the genomic distribution of associations, with distinct peaks indicating potential genomic regions of interest. The Q-Q plot (Fig.4.27a to 4.27c) demonstrates that the observed p-values deviate from the null distribution, supporting the presence of true associations. The Q-Q plot compares the observed p-values with the expected p-values under the null hypothesis of no association. Deviation from the expected line indicates the presence of genetic associations. The closer the observed p-values align with the expected line; the fewer false positives are present in the analysis.

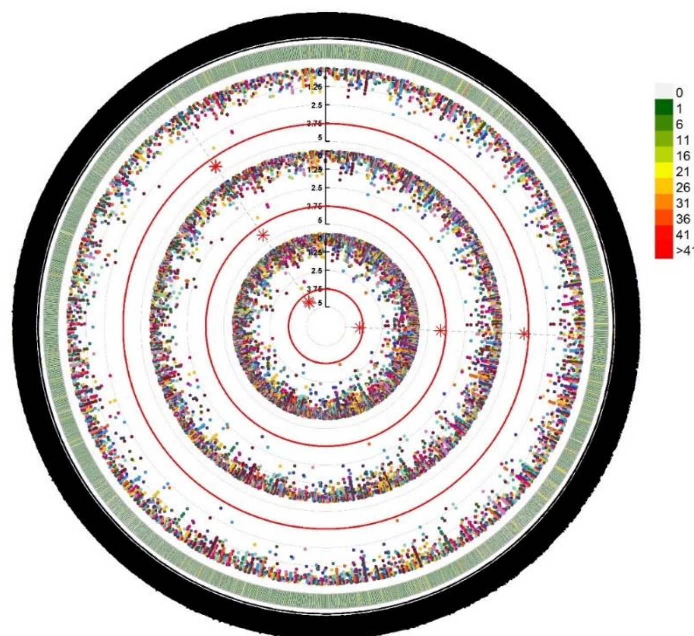


Fig 4.24: Manhattan plot of SNP association for length of primary rhizomes. The Manhattan plot above displays the genome-wide associations between single nucleotide polymorphisms (SNPs) and the trait, primary length of rhizome per plant. Each point represents a SNP, with the x-axis indicating its genomic position, and the y-axis showing the -log10 (p-value) of the association. Significant associations are represented by points above the significance threshold line.

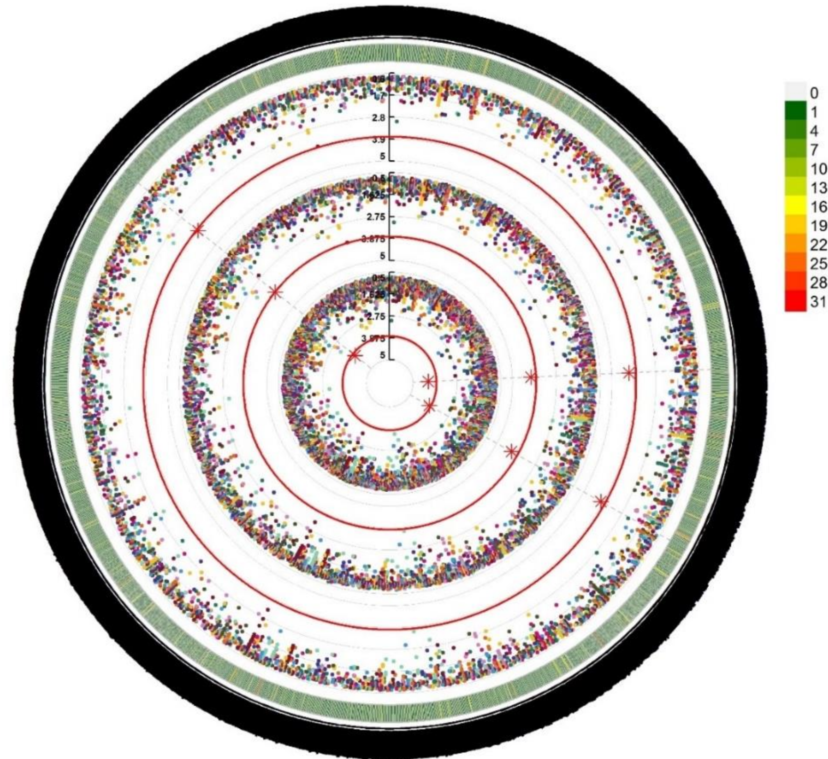


Fig 4.25: Manhattan plot of SNP associations for girth of primary rhizomes. The Manhattan plot displays the genome-wide associations between single nucleotide polymorphisms (SNPs) and the trait, girth of primary rhizomes per plant. Each point represents a SNP, with the x-axis indicating its genomic position, and the y-axis showing the -log10(p-value) of the association. Significant associations are represented by points above the significance threshold line.

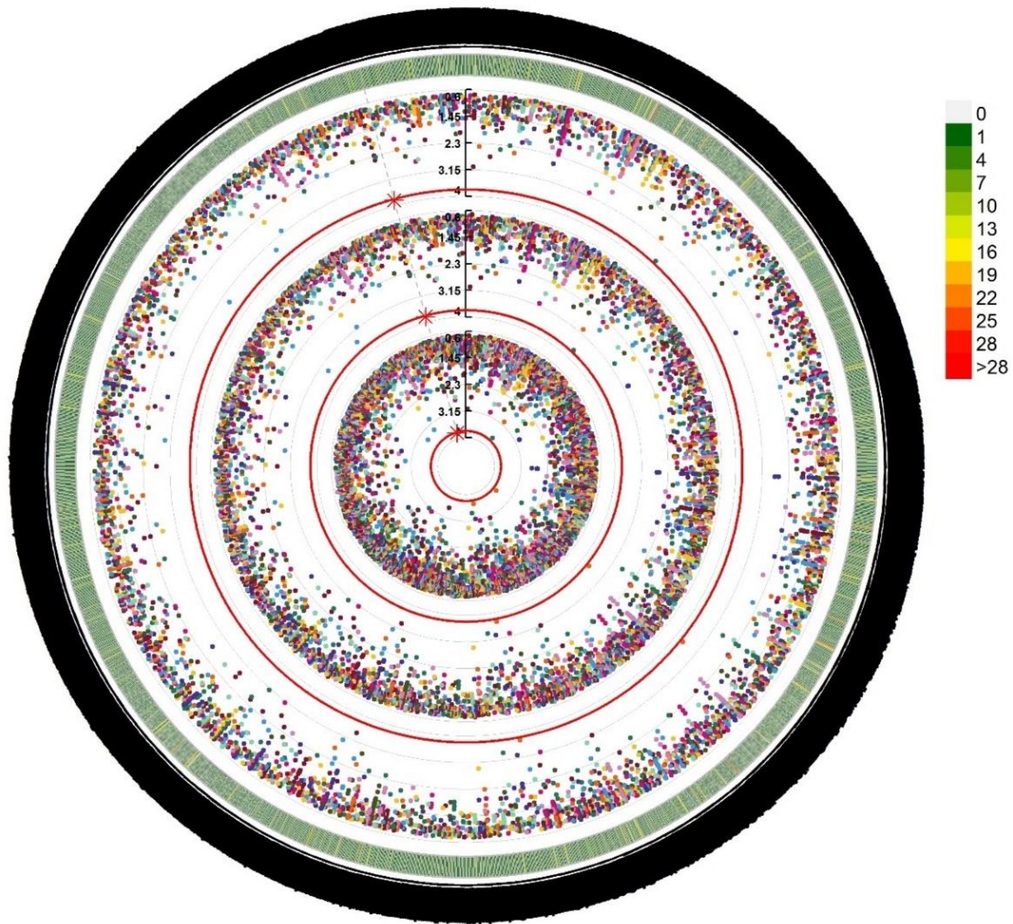


Fig 4.26: Manhattan plot of SNP associations for curcuminoid content. The Manhattan plot displays the genome-wide associations between single nucleotide polymorphisms (SNPs) and the trait, curcuminoid content. Each point represents a SNP, with the x-axis indicating its genomic position, and the y-axis showing the -log10 (p-value) of the association. Significant associations are represented by points above the significance threshold line.

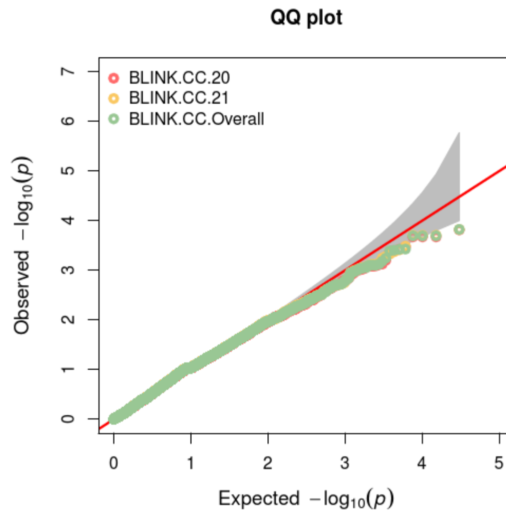


Fig 4.27a

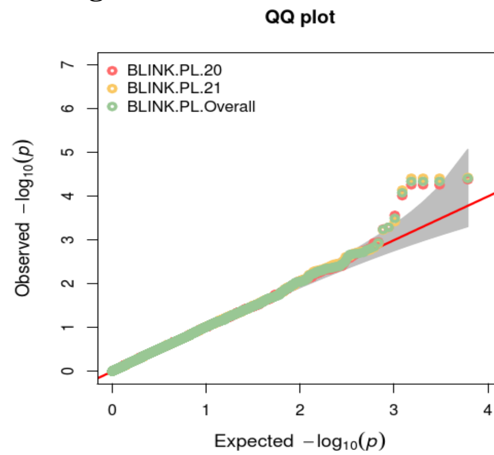


Fig 4.27b

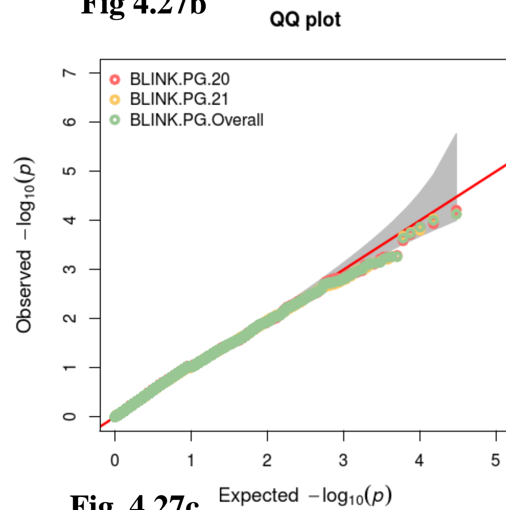


Fig 4.27c

Fig. 4.27a. Q-Q plot for the trait curcuminoid content; **Fig. 4.27b.** Q-Q plot for the trait of length of primary rhizomes per plant; **Fig. 4.27c.** Q-Q plot for the trait of girth of primary rhizomes per plant.

Thus, association analysis has resulted in identification of nine significant MTAs (marker-trait associations) for the three phenotypic traits (Table 4.13). The number of significant MTAs per trait are as follows: curcumin content (1), rhizome length (5), rhizome girth (3). These associations provide insights into the genetic architecture underlying the phenotypic variation and can serve as valuable markers for future studies or breeding programs.

4.13 BLAST analysis of GWAS-identified MTAs

To gain insights into the genomic location of the Marker-Trait Associations (MTAs) identified through Genome-Wide Association Study (GWAS), a BLAST analysis was conducted. For this, a 1 Kb sequence comprising MTAs was retrieved, and a BLAST comparison was made with the reference genome that had been assembled at the chromosomal level (Yin et al., 2022). This allowed us to infer the chromosomal location of MTAs. Further, to find whether the MTAs fall in any potential candidate genes, a similarity search using BLASTN was carried out. This analysis aimed to find the functional significance of MTAs by comparing them to known nucleotide sequences available in public databases.

Table 4.14. Summarizes the top hits obtained from the BLASTN analysis. Each hit is accompanied by relevant information such as the hit ID, description, alignment score, and other pertinent details. Lower e-values suggest stronger similarities to known nucleotide sequences. These results provide initial clues about the potential functions and roles of the MTAs in various biological processes or pathways.

Table 4.14. BLASTN similarity analysis results

Trait	SNP	Chromosomal location	Putative candidate genes if any	Scientific name	Max score	Total score	Query cover	E value	Percent identity (%)	Alignment length (bp)
Curcumin content	MTA 1	Chr 11	*Zingiber officinale* probable ethanolamine kinase (LOC121974481), mRNA	*Z. officinale*	165	247	13%	2E-35	100 %	1415
Rhizome length	MTA 2	Chr 4	*Z. officinale* MADS-box transcription factor 34-like (LOC122034397), mRNA	*Z. officinale*	159	159	11%	1e-33	91.45%	984
	MTA 3	Chr 4	*Z. officinale* MADS-box transcription factor 34-like (LOC122034397), mRNA	*Z. officinale*	159	159	11%	1e-33	91.45%	984
	MTA 4	Chr 4	*Z. officinale* MADS-box transcription factor 34-like (LOC122034397), mRNA	*Z. officinale*	159	159	11%	1e-33	91.45%	984
	MTA 5	Chr 4	*Z. officinale*	*Z. officinale*	159	159	11%	1e-33	91.45%	984

			MADS-box transcription factor 34-like (LOC122034397), mRNA							
	MTA 6	Chr 21	*Chenopodium album* genome assembly; chromosome 18	*C. album*	97.1	97.1	16%	9e-15	77.38%	50815545
Rhizome girth	MTA 7	Chr 19	No significant similarity found							
	MTA 8	Chr 20	No significant similarity found							
	MTA 9	Chr 13	*Myrica gale* genome assembly, chromosome: 14	*M. gale*	137	137	15%	5e-27	82.80%	26419627

The association analysis identified several significant SNP markers associated with the target traits. For the trait of curcuminoid content, SNP located on chromosome 11 showed a strong association ($p < 0.001$) and this genomic locus showed similarity to *Z. officinale* probable ethanolamine kinase (LOC121974481), mRNA. For length of primary rhizome, all the SNPs detected fall on chromosome 4 and were highly similar to *Z. officinale* MADS-box transcription factor 34-like (LOC122034397), mRNA except for one SNP (MTA 6) that fall on chromosome 21 and was similar to chromosome 18 of *C. Album* genome assembly. Meanwhile, for the three SNPs associated with rhizome girth that falls on chromosome 19, chromosome 20 and chromosome 13 respectively, we did not find any similarity match except for MTA 9 which most likely hints their location in non-coding regions. MTA 9 that falls on chromosome 13 of turmeric chromosomal assembly finds similarity with chromosome 14 of *M. gale* genome assembly.

In association analysis, the likelihood of getting false positive results (Type I errors) rises when running several statistical tests at once. Bonferroni correction was used to address this problem. By changing the p-value cut-off for each individual test, the Bonferroni correction helps regulate the overall significance level or family-wise error rate (Gubaev et al., 2020). Thus, results obtained indicate that these SNP markers may play a crucial role in determining the observed variation in length of primary rhizome, girth of primary rhizome and curcuminoid content.

Further validation and replication of these associations in independent datasets will be necessary to confirm their robustness. Nevertheless, our findings provide valuable insights into the genetic architecture underlying these traits, which can inform future breeding efforts aimed at improving plant performance and disease resistance.

Overall, our SNP marker trait association analysis using GAPIT has uncovered significant associations between specific genetic variants and important agronomic traits, shedding light on the genetic factors contributing to phenotypic variation in our population.

4.14 Discussion of objective 3

Here, the turmeric genotypes were characterised using the double-digest restriction-site associated DNA sequencing (ddRAD) method, and the population structure of the genotypes were then elucidated through variant calling using the Genome Analysis Toolkit (GATK). One of the earlier studies (Liang et al., 2021) have recommended the utilization of reduced representation sequencing methods like ddRAD to understand phylogeny and evolution of *Curcuma* species but there has been no such study done till date. The identification and characterization of genetic variations through the processing of ddRAD data will advance our knowledge of turmeric crop genomics and will aid efforts in subsequent study and breeding. The choice of restriction enzymes and size selection range are the most crucial factors for any ddRAD experiment (Peterson et al., 2012; Yang et al., 2016). The restriction enzyme pair utilized here, *SphI* and *MlucI* produced sufficient number of fragments well within our size selection range of 100-200 bp and hence can be utilized for ddRAD experiments in similar crops. The recovered fragments after digestion is 2-3% for say 1 Gb genome and the depth of a read is atleast 5X and reads below 5X is filtered out and the size selection range was of 100- 200 bp. The recommended size selection in ddRAD sequencing for plants varies depending on factors such as the species under study, research objectives, and the sequencing platform employed (Peterson et al., 2012). However, a commonly advised size range for plants typically falls between 250 to 500 base pairs (bp) (Peterson et al., 2012) . This range is often preferred as it strikes a balance between capturing sufficient genomic information and minimizing biases associated with extremely short or long fragments (Aballay et al., 2021). Opting for smaller fragments as in here (100-200 bp) offers potential advantages such as heightened resolution and sensitivity, along with the capability to detect finer genetic variations for complex trait analysis as in GWAS (Miller et al., 2007). But this approach also poses potential drawbacks as well, including heightened bias and complexities in computational analysis (Andrews et al., 2016).

After the ddRAD sequencing, quality control of sequencing data is an important step as it ensures the suitability of the data for downstream analysis (Davey et al., 2011) .

In our study here, the absence of significant concerns pertaining to sequence quality, read length, GC content, and adapter contamination validated the data's suitability for subsequent analysis (Catchen et al., 2013). These results instil confidence in the reliability of the acquired sequencing data. Unlike many NGS data quality control analysis, here we have not removed the duplicate reads in the overrepresented reads (Li et al., 2009). This is because these sequences may correspond to known target regions of interest, such as highly expressed genes or repetitive elements or can arise due to polyploid nature of the crop. Hence, it was determined that these sequences were biologically relevant rather than indicative of contamination or technical artifacts (Davey et al., 2011).

After quality check, alignment to a reference genome or *de novo* genome is necessary to discover genomic variants present (Clevenger et al., 2017) . Here, the alignment statistics demonstrated high-confidence mapping, with the majority of reads uniquely aligned to specific genomic locations. These results validate the effectiveness of the BWA alignment algorithm for aligning NGS data to the reference genome, enabling reliable downstream analysis and interpretation as reported in other studies as well (Chakraborty et al., 2021; Hazzouri et al., 2019; Li & Durbin, 2009).

The variant calling process using GATK has successfully identified a comprehensive set of single nucleotide polymorphisms (SNPs), i.e. 30438 SNP variants in the crop genotypes. These variants were utilized as genetic markers for population genetic analysis and association studies. Thus, the high-throughput nature of ddRAD sequencing combined with the accuracy of GATK variant calling has enabled the efficient and reliable identification of genetic variations across the crop genome (Clevenger et al., 2015).

The genotypic data containing the identified SNP variants were then utilized for the study of genetic population structure present among the genotypes. Identification of intraspecific population structure is one of the pre-requisites before embarking on an in-depth genetic diversity or evolution study (Boukteb et al., 2021). Here, we have conducted population structure for marker-trait association studies. The population

structure analysis based on the identified variants has revealed three distinct genetic clusters within the studied crop population. Presence of three subgroups in structure analysis indicate population stratification because of genetic variation and capacity for adaptability within the crop species (Szczecińska et al., 2016). Because it enables the selection of a wide and representative sample of individuals for hybridization and trait improvement, understanding the population structure is essential for the design and implementation of efficient breeding techniques (Werner et al., 2020).

The study of population structure is important as it can shed light on the patterns of crop domestication and evolutionary history (Bird et al., 2017). For example, the discovery of numerous subpopulations within a species of say a cereal crop can reveal the influence of several breeding programmes and the geographic origins of the genotypes. Several studies underline how crucial it is to take population structure into account when conducting association mapping research to prevent false-positive associations and to enhance the precision of genotype-phenotype connections (Kaler et al., 2020; Naret et al., 2018; Szczecińska et al., 2016).

The genetic variants discovered by ddRAD genotyping and GATK variant calling can be used for a variety of downstream investigations in addition to population structure. These include marker-assisted selection (MAS) in breeding programmes, candidate gene identification, and genome-wide association studies (GWAS) (Scheben et al., 2017). The accessibility of high-quality genomic variants enables researchers and breeders to relate certain genetic loci to critical agronomic qualities in less time. At the same time, we need to acknowledge the drawbacks and limitations of ddRAD genotyping and GATK variant calling. Here, the representation of genomic region may be impacted by the choice of appropriate restriction enzymes and library preparation techniques, which may also cause biases in variant calling (Herry et al., 2023). Additionally, the presence of repeated areas and structural variations may make it difficult to detect variants effectively (Peterson et al., 2012). To guarantee accurate and reliable results, careful consideration and validation of the detected variants are required.

Hence, the ddRAD genotyping approach combined with GATK variant calling has provided valuable insights into the genetic diversity, population structure, and genetic variation of crop genotypes. The identified variants serve as valuable resources for population genetics, association studies, and breeding applications. Future research should focus on expanding the sample size, incorporating additional genomic data, and integrating functional genomics approaches to further enhance our understanding of crop genetics and enable more precise crop improvement strategies.

Although a lot of genetic diversity studies have been done in turmeric based on agro-morphological characterization (Anindita et al., 2020; Sasikumar, 2005) and molecular marker based characterization (Siju et al., 2010a; Siju et al., 2010b; Singh et al., 2018; Verma et al., 2015), there have not been any marker-trait association work in the crop. The genomic ambiguity due to the absence of a reference genome was a hindrance to effective marker-trait studies as it can result in bias and misidentification of variants (Formenti et al., 2022; Valiente-Mullor et al., 2021). The advent of reference genome paved way to association studies in turmeric (Chakraborty et al., 2021). In annual crops like turmeric marker trait studies can expedite the identification and selection of plants with desired traits (Kumawat et al., 2020). The traditional methods identification of marker trait association (QTL mapping) relies on crossing over studies and linkage analysis which are challenging in crops like turmeric that is mostly vegetatively propagated with only occasional seed setting (Khan & Korban, 2012). Association mapping that exploits the inherent genetic diversity in the population provides an alternative strategy to traditional QTL mapping (Gupta et al., 2014) .

The association analysis or association mapping has now emerged as a powerful tool in plant genomics with the advent of new statistical softwares and high throughput molecular marker technologies (Zhu et al., 2008). Approach of genome wide association studies or GWAS that leverages on genome wide markers has become highly useful in the elucidation of the genetic architecture of complex traits across plant species (Padmashree et al., 2023; Peng et al., 2023). There have been studies

combining the sequencing data from high throughput method of ddRAD sequencing with agro-morphological trait information to gain insights into the genetic diversity as well as markers associated with traits of interest (Esposito et al., 2020; Han et al., 2023). In our study, we have used the BLINK model in the GAPIT package to find the marker-trait association in turmeric. BLINK model that has been used in GWAS studies utilize linkage disequilibrium to increase statistical power and the maximum likelihood is estimated through the application of Bayesian Information Content (BIC) (Huang et al., 2019).

Here, we have identified nine significant Marker-Trait Associations (MTAs) that can contribute to our understanding of trait variation in this crop. This is the first study of its kind in turmeric. The variants discovered here may most likely contribute to the complex traits of interest. For instance, MTA1 that is likely to be in an ethanolamine kinase had showed a significant association with curcuminoid content, hinting to its role in this complex trait expression. Ethanolamine kinase are enzymes that participate in phospholipid metabolism that has wide ranging impacts on plant growth and metabolism (Lin et al., 2020; Rabeler et al., 2022). This highlights the reliability of the BLINK model in capturing meaningful associations. Four MTAs pertaining to rhizome length, namely MTA 2, MTA 3, MTA 4 and MTA 5 are likely to be in MADS-box transcription factor had showed a significant association with rhizome length. The role of MADS-box transcription factors in development of vegetative tissues like rhizomes and tuber have been reported in some plants (Fatima et al., 2020; Sun et al., 2018, 2021). In a microarray study specifically in ginger and turmeric, it was found that there was an upregulated rhizome specific expression of few MADS box genes which were homologous to MADS box genes which are found near quantitative trait loci (QTLs) linked to rhizomatousness in Oryza and Sorghum plant species (Koo et al., 2013). MTA 6 associated with rhizome length was found to be similar with chromosome 18 of *Chenopodium album* (*C. Album*) genome assembly. *C. album*, commonly known as lamb's quarters or pigweed, is an annual weed prevalent in various environments, including agricultural fields as well as urban areas and rich in polyphenolic phytochemicals of medicinal value (Saini et al., 2019; Poonia & Upadhayay, 2015).

Interestingly, *C. album* have been reported as common weed in turmeric fields (Tahira et al., 2010) Genomic similarity of host plant and weed hints to molecular dynamics of weed host interaction and likely incidence of horizontal gene transfer (Ashapkin et al., 2023). Among the MTAs associated with rhizome girth, MTA 7 and MTA 8 did not show any similarity with known nucleotide sequences except for MTA 9 that showed similarity with regions of chromosome 14 of *Myrica gale* genome assembly (*M. gale)*, often referred to as sweet gale or bog myrtle is a perennial shrub native to wetland areas of Europe, Asia, and North America and is known for polyphenolic content, which supports its traditional culinary and medicinal use (Ložienė et al., 2023). Absence of similarity matches for MTA 7 and MTA 8 hints to their likely location at non-coding region and species specific genomic regions. Although many GWAS studies across species have found MTAs in non-coding regions, the interpretation of these variants is still challenging due to the uncertainty in the functional consequences and regulatory genes involved (Smits et al., 2004; Wu et al., 2020).

It is important to acknowledge the limitations and challenges encountered during the association analysis. These limitations include potential biases arising from available data and the impact of sample size on detecting associations, particularly for rare variants (Korte & Farlow, 2013). Conducting future studies with larger sample sizes and more comprehensive genotyping approaches can offer a more comprehensive understanding of the genetics of the traits.

The utilization of the BLINK model in the GAPIT package provides advantages by effectively accounting for population structure and kinship, thus improving the accuracy of association mapping (Huang et al., 2019). However, the model assumes no interactions between markers, which may result in overlooking important genetic effects. Future research can enhance the model's performance by incorporating epistatic interactions.

The implications of the findings extend beyond the scope of this study. Gaining knowledge about the genetic basis of traits can inform breeding programs aimed at developing improved turmeric varieties with improved characters. The identified

MTAs serve as targets for further functional studies, enabling the exploration of the molecular mechanisms underlying these traits.

In short, the association analysis employing the BLINK model in the GAPIT package has provided valuable insights into the genetic architecture of the examined traits in turmeric. The identified MTAs, contribute to our understanding of trait variation and establish a foundation for future investigations. Future research directions should prioritize increasing sample sizes, investigating epistatic interactions, and conducting functional studies to unravel the molecular mechanisms driving these significant traits.

Chapter 5

CONCLUSION

Phenotypic characterization revealed significant genetic variation among 93 turmeric genotypes. Flow cytometric analysis revealed that among these 93 genotypes, majority were triploids (84) and nine were tetraploids. Most of the tetraploids have origins as seedling progenies while few are germplasm collections. It was found that triploids and tetraploids differ significantly in their plant height, number of shoots per plant, number of leaves on main shoot, length of petiole, leaf length, length of mother rhizomes per plant, length of primary rhizomes per plant, length of secondary rhizomes per plant, girth of secondary rhizomes per plant, inner core diameter of primary rhizomes, weight of mother rhizomes per plant, weight of primary rhizomes per plant, total rhizome weight per plant and dry rhizome weight per plant. Genetic variability analysis revealed that characters; number of shoots per plant, length of mother rhizome, total weight of mother rhizomes per plant, total weight of rhizome per plant and dry weight of rhizome per plant were the most important traits for selection across ploidy levels. Here, we have utilized molecular marker based genotyping as well as NGS based genotyping to study the genetic relationship of the genotypes under study. Although, reliable and reproducible, the inherent limitations of molecular marker to achieve the requisite resolution for association studies led us to proceed with NGS based genotypic data analysis. Quality check and control of NGS ddRAD data enhanced the quality of genotypic data for a reliable downstream analysis result. Also, we have optimized a bioinformatics pipeline or workflow for variant calling which have resulted in discovery of 30438 SNP variants. Further, we have analysed the population structure of genotypes under study and have found that there are 3 subpopulations. This is important as not properly corrected, these factors (population structure and relatedness) can cause false positives and false negatives in downstream analysis. These 30438 SNP markers were utilized for association mapping in turmeric using the GAPIT package in R. 9 significant marker trait association were observed:

curcumin content (1), rhizome length (5), and rhizome girth (3). Most marker trait associations were from chromosome 4. *Z. officinale* MADS-box transcription factor 34-like (LOC122034397) can be a putative locus involved in rhizome length morphology. These associations can help in the reveal of the molecular mechanisms involved in generation of complex traits and result in the identification of candidate genes involved. Thus, results obtained here will add onto the understanding of genetic base of these complex traits in turmeric and to further crop improvement efforts in turmeric.

In conclusion, this study on marker-trait association in turmeric has revealed significant findings that contribute to the understanding of the genetic basis of key traits in turmeric. Our work not only advances the current knowledge in the field but also holds practical implications for improving turmeric varieties through marker-assisted selection. Despite acknowledged limitations (biological complexity, incomplete genomic information), the study provides a foundation for future research directions (validation and replication), highlighting opportunities to explore the impact of environmental factors and suggesting avenues for refining marker-assisted breeding strategies. Overall, this research contributes valuable insights to the scientific community, paving the way for enhanced turmeric crop management and genetic improvement.

Chapter 6

RECOMMENDATIONS

The overall aim of this Ph.D. work was to study genetic diversity and marker trait association in turmeric (*Curcuma longa* L.). The major objectives were to study the level of phenotypic variability among core turmeric collection, to genetically characterize these genotypes using molecular markers and finally to determine the association of rhizome characters and quality traits with molecular markers in turmeric. The first component of the study that involves the phenotypic characterization of turmeric genotypes for two years revealed significant genetic diversity among these 93 genotypes that form the core collection. Flow cytometric analysis revealed that among these 93 genotypes, majority were triploids (84) and nine were tetraploids. Most of the tetraploids have origins as seedling progenies while few are germplasm collections. It was found that triploids and tetraploids differ significantly in their plant height, number of shoots, number of leaves on main shoot, petiole length, leaf length, length of mother rhizome, length of primary rhizomes, length of secondary rhizomes, girth of secondary rhizomes, inner core diameter of primary rhizome, weight of mother rhizomes per plant, weight of primary rhizomes per plant, total rhizome weight per plant and dry rhizome weight per plant. Genetic variability analysis revealed that characters; number of shoots per plant, length of mother rhizome, total weight of mother rhizomes per plant, total weight of rhizome per plant and dry weight of rhizome per plant were the most important traits for selection across ploidy levels. As most of these phenotypic traits in turmeric is highly influenced by environment, there is a need for replication of these experiment in various environment conditions before arriving at generalisation and conclusive findings.

As part of second objective of molecular characterization, we have carried out the genotypic characterization of turmeric using both microsatellite (SSR and ISSR) and next-generation sequencing (NGS) markers. The molecular marker based characterization have revealed three major clusters that groups the genotypes of core

collection based on their molecular genetic similarity. Although, reliable and reproducible, the inherent limitations of molecular marker to achieve the requisite resolution for association studies led us to proceed with NGS based genotypic data analysis. Quality check and control of ddRAD data enhanced the quality of genotypic data for a reliable downstream analysis result. Also, we have optimized a bioinformatics pipeline or workflow for variant calling which have resulted in discovery of 30438 SNP variants. Even though earlier studies have recommended the utilization of reduced representation sequencing methods like ddRAD to understand phylogeny and evolution of *Curcuma* species but there has been no such study done till date and this is the first study of its kind. The genomic study of turmeric is still in its infant stage and hence there is a need for continued and sustained efforts to expand turmeric genomic resources by exploiting the emerging utilities of high-throughput genotyping platforms and enhancing bioinformatics pipelines and analytical tools for efficient analysis of NGS data for turmeric breeding and genetic studies.

The final objective of identification of marker trait association was accomplished with the discovery of nine significant marker-trait associations, notably linked to curcumin content and rhizome attributes. Most marker trait associations were from chromosome 4. *Zingiber officinale* MADS-box transcription factor 34-like (LOC122034397) was found out as a putative candidate gene locus involved in rhizome length morphology. These associations can help in the reveal of the molecular mechanisms involved in generation of complex traits and result in the identification of candidate genes involved. Thus, results obtained here will add onto the understanding of genetic base of these complex traits in turmeric and to further crop improvement efforts in turmeric.

As a future perspective, this work needs to be further validated by conducting validation studies to confirm the marker-trait associations observed in this research using independent datasets or populations. The stability and generalizability of these marker-trait associations can be examined by replicating the study in diverse environments and genetic backgrounds. Also, more functional studies are required to

validate the role of candidate genes identified through marker-trait associations in regulating key traits such as rhizome length and curcumin content. Future studies involving gene expression analysis, gene editing techniques (e.g., CRISPR/Cas9), or transgenic approaches can help to elucidate the molecular mechanisms underlying trait variation and will enhance our understanding of the genetic architecture of these traits. This can be supplemented with an integrative approach by combining the multiple omics datasets (e.g., transcriptomics, metabolomics) to gain a comprehensive understanding of these complex traits in turmeric by investigating gene regulatory networks and metabolic pathways associated with turmeric rhizome development and secondary metabolite biosynthesis. As most of these traits are highly influenced by environment, future studies in the direction of environmental effects and GxE interactions will add on to our understanding about the trait expression. Exploring the influence of environmental factors (e.g., climate, soil conditions) on turmeric trait expression and marker-trait associations by investigating the genotype-by-environment interactions (GxE) to identify genomic regions or markers associated with trait stability across diverse environmental conditions. Findings from all these studies can be collated and used to develop and implement marker-assisted breeding strategies in turmeric targeting key traits identified in this study, such as rhizome length and curcumin content. Incorporating this marker information into breeding programs also helps to accelerate the development of improved turmeric varieties with desirable agronomic and medicinal traits.

REFERENCES

Aarthi, S., Suresh, J., Leela, N. K., & Prasath, D. (2020). Multi environment testing reveals genotype-environment interaction for curcuminoids in turmeric *(Curcuma longa* L.). *Industrial Crops and Products*, *145*, 112090.

Aarthi, S., Suresh, J., & Prasath, D. (2018a). Morphological characterization of Indian turmeric (*Curcuma longa* L.) genotypes using DUS descriptor. *Journal of Plantation Crops*, *46*(3), 173-179.

Aarthi, S., Suresh, J., & Prasath, D. (2018b). Variability and association analysis of curcumin content with yield components in turmeric (*Curcuma longa* L.). *Electronic Journal of Plant Breeding*, *9*(1), 295-303.

Aballay, M. M., Aguirre, N. C., Filippi, C. V., Valentini, G. H., & Sánchez, G. (2021). Fine-tuning the performance of ddRAD-seq in the Peach genome. *Scientific Reports*, *11*(1), 6298.

Ahmad, R. S., Hussain, M. B., Sultan, M. T., Arshad, M. S., Waheed, M., Shariati, M. A., Plygun, S., & Hashempur, M. H. (2020). Biochemistry, safety, pharmacological activities, and clinical applications of turmeric: a mechanistic review. *Evidence-Based Complementary and Alternative Medicine*, Article ID 7656919. https://doi.org/10.1155/2020/7656919.

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, *71*(2), 103-112.

Allard, R. W. (1960). Principles of plant breeding (Ed.2), John Willey and Sons. *Inc. New York*, *485*.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403- 410.

Ammon, H. P. T., & Wahl, M. A. (1991). Pharmacology of *Curcuma longa* L. *Planta Medica*, *57 1*, 1-7.

Amom, T., Tikendra, L., Apana, N., Goutam, M., Sonia, P., Koijam, A. S., Potshangbam, A. M., Rahaman, H., & Nongdam, P. (2020). Efficiency of RAPD, ISSR, iPBS, SCoT and phytochemical markers in the genetic relationship study of five native and economical important bamboos of North-East India. *Phytochemistry*, *174*, 112330.

Anandaraj, M., Prasath, D., Kandiannan, K., Zachariah, T. J., Srinivasan, V., Jha, A. K., Singh, B. K., Singh, A. K., Pandey, V. P., Singh, S. P., Shoba, N., Jana, J. C., Kumar, K. R., & Maheswari, K. U. (2014). Genotype by environment interaction effects on yield and curcumin in turmeric (*Curcuma longa* L.). *Industrial Crops and Products*, *53*, 358-364.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81-92.

Anindita, P. A., Putri, T. K., Ustari, D., Maulana, H., Rachmadi, M., Concibido, V., Suganda, T., & Karuniawan, A. (2020). Dataset of agromorphological traits in early population of turmeric (*Curcuma longa* L.) local accessions from Indonesia. *Data in Brief*, *33*, 106552.

Annadurai, R. S., Neethiraj, R., Jayakumar, V., Damodaran, A. C., Rao, S. N., Katta, M. A. V. S. K., Gopinathan, S., Sarma, S. P., Senthilkumar, V., Niranjan, V., Gopinath, A., & Mugasimangalam, R. C. (2013). De Novo Transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLoS ONE*, *8*(2), e56217.

Anne Plotto. (2004). *Turmeric: Post Production Management*. Food and Agriculture Organization of the United Nations (FAO), AGST, Rome, Italy,1-21.

Anonymous, *ITIS - Report: Curcuma longa*. https://www.itis.gov

Arif, I. A., Bakir, M. A., Khan, H. A., Al Farhan, A. H., Al Homaidan, A. A., Bahkali, A. H., Al Sadoon, M., & Shobrak, M. (2010). A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences, 11*(5), 2079-2096.

Ashapkin, V. V., Kutueva, L. I., Aleksandrushkina, N. I., Vanyushin, B. F., Teofanova, D. R., & Zagorchev, L. I. (2023). Genomic and epigenomic mechanisms of the interaction between parasitic and host plants. *International Journal of Molecular Sciences*, *24*(3), 2647.

ASTA. (2004). Curcumin content of turmeric spice and oleoresin. *Official Analytical Methods* (second edition).

Aswathi, A. P., Raghav, S. B., & Prasath, D. (2023). Assessment of genetic variation in turmeric (*Curcuma longa* L.) varieties based on morphological and molecular characterization. *Genetic Resources and Crop Evolution*, *70*(1), 147-158.

Ayer, D. K. (2017). Breeding for quality improvement in Turmeric (*Curcuma longa* L.): a review. *Advances in Plants & Agriculture Research*, *6*(6), 1- 4.

Babu, P. M., Neeraja, C. N., Rathod, S., Suman, K., Uttam, G. A., Chakravartty, N., Lachagari, V. B. R., Chaitanya, U., Rao, L. V. S., & Voleti, S. R. (2020). Stable SNP allele associations with high grain zinc content in polished rice (*Oryza sativa* L.) identified based on ddRAD sequencing. *Frontiers in Genetics*, 11, 763. https://doi.org/10.3389/fgene.2020.00763.

Bahadur, V., Yeshudas, V., & Meena, O. P. (2016). Nature and magnitude of genetic variability and diversity analysis of Indian turmeric accessions using agro-morphological descriptors. *Canadian Journal of Plant Science*, *96*(3), 371-381.

Begali, H. (2018). A pipeline for markers selection using restriction site associated DNA sequencing (Radseq). *Journal of Applied Bioinformatics & Computational Biology*, *7*(1), doi: 10.4172/2329-9533.1000147.

Bhadra, S., & Bandyopadhyay, M. (2015). Karyomorphological investigations on some economically important members of Zingiberaceae from Eastern India. *Caryologia*, *68*(3), 184-192.

Bhadra, S., Maity, D., & Bandyopadhyay, M. (2018). Correlating karyomorphology and molecular marker analyses in turmeric: a case study. *Journal of Crop Improvement*, *32*(5), 657-680.

Bhanu, A. N. (2017). Assessment of genetic diversity in crop plants - an overview. *Advances in Plants & Agriculture Research*, *7*(3), 279-286.

Bird, K. A., An, H., Gazave, E., Gore, M. A., Pires, J. C., Robertson, L. D., & Labate, J. A. (2017). Population structure and phylogenetic relationships in a diverse panel of *Brassica rapa* L. *Frontiers in Plant Science*, 8, 243265.

Bonna, I. J., Akter, S., & Sultana, S. S. (2021). Somatic chromosome number and ploidy level in three Curcuma spp. from Bangladesh. *Dhaka University Journal of Biological Sciences*, *30*(2), 133-140.

Boukteb, A., Sakaguchi, S., Ichihashi, Y., Kharrat, M., Nagano, A. J., Shirasu, K., & Bouhadida, M. (2021). Analysis of genetic diversity and population structure of *Orobanche foetida* populations from Tunisia using RADseq. *Frontiers in Plant Science*, 12, 618245.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633-2635.

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177-186.

Burton, G. W., & DeVane, E. H. (1953). estimating heritability in Tall Fescue (*Festuca Arundinacea*) from replicated clonal material. *Agronomy Journal*, *45*(10), 478-481.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Caruso, I., Dal Piaz, F., Malafronte, N., De Tommasi, N., Aversano, R., Wulff Zottele, C., Scarano, M.-T., & Carputo, D. (2013). Impact of ploidy change on secondary metabolites and photochemical efficiency in Solanum bulbocastanum. *Natural Product Communications*, *8*(10), 1387-1392.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140.

Chakraborty, A., Mahajan, S., Jaiswal, S. K., & Sharma, V. K. (2021). Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Communications Biology*, *4*(1), 1193.

Chakravorthy, A. K. (1948). Multiplication of chromosome number in relation to speciation of Zingiberaceae. *Science and Culture*, *14*, 137-140.

Chen, J., Xia, N., Zhao, J., Chen, J., & Henny, R. J. (2013). Chromosome numbers and ploidy levels of Chinese Curcuma Species. *Hortscience*, 48(5), 525-530.

Chintakovid, N., Tisarum, R., Samphumphuang, T., Sotesaritkul, T., & Cha-Um, S. (2022). Evaluation of curcuminoids, physiological adaptation, and growth of *Curcuma*

longa under water deficit and controlled temperature. *Protoplasma*, *259*(2), 301-315.

Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., & Jackson, S. A. (2015). Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. *Molecular plant 8*(6), 831-846.

Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C. M., Pandey, M. K., Vaughn, J., Abernathy, B., Barkley, N. A., Hovav, R., Burow, M., Nayak, S. N., Chitikineni, A., Isleib, T. G., Holbrook, C. C., Jackson, S. A., Varshney, R. K., & Ozias-Akins, P. (2017). Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in Peanut. *Molecular plant*, *10*(2), 309-322.

Cooper, H. D., Spillane, C., & Hodgkin, T. (2001). Broadening the genetic base of crop production. International Plant Genetic Resources Institute (IPGRI) Food and Agriculture Organization of the United Nations (FAO), CABI.

Crespel, L., & Meynet, J. (2003). Biotechnologies for breeding | Manipulation of Ploidy Level. *Encyclopedia of Rose Science*, Elsevier, 5-11.

Das, Rai S, & das P. (1999). Karyotype analysis and cytophotometric estimation of nuclear DNA content in some members of the Zingiberaceae. *Cytobios*, *97*(384), 23-33.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499-510.

Davoodvandi, A., Farshadi, M., Zare, N., Akhlagh, S. A., Alipour Nosrani, E., Mahjoubin-Tehran, M., Kangari, P., Sharafi, S. M., Khan, H., Aschner, M., Baniebrahimi, G., & Mirzaei, H. (2021). Antimetastatic effects of curcumin in oral and gastrointestinal cancers. *Frontiers in Pharmacology 12*, 668567.

Deng, H., Wan, M., Li, H., Chen, Q., Li, R., Liang, B., & Zhu, H. (2021). Curcumin protection against ultraviolet-induced photo-damage in Hacat cells by regulating nuclear factor erythroid 2-related factor 2. *Bioengineered*, *12*(2), 9993-10006.

Dev, H., & Sharma, V. (2020). Correlation studies in turmeric (*Curcuma longa* L.). *International Journal of Current Microbiology and Applied Sciences*, *9*(8), 3063-3071.

Dev, H., & Sharma, V. (2022). Genetic Variability in Turmeric (*Curcuma longa* L.). *International Journal of Bio-Resource and Stress Management*, *13*(6), 595-604.

Dixit, V., & Chaudhary, B. R. (2014). Colchicine-induced tetraploidy in garlic (*Allium sativum* L.) and its effect on allicin concentration. *Journal of Horticultural Science and Biotechnology*, *89*(5), 585-591.

Doležel, J., & Bartoš, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany*, *95*(1), 99-110.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11-15.

Duke, J. (2007). Turmeric -The Genus Curcuma. Medicinal and Aromatic Plants - Industrial Profiles Volume 45. *Economic Botany*, *61*(4), 397-398.

Duncan, D. B. (1955). Multiple range and multiple F tests. *Biometrics*, *11*(1), 1-42.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422-433.

Esposito, S., Cardi, T., Campanelli, G., Sestili, S., Díez, M. J., Soler, S., Prohens, J., & Tripodi, P. (2020). ddRAD sequencing-based genotyping for population structure analysis in cultivated tomato provides new insights into the genomic diversity of Mediterranean 'da serbo' type long shelf-life germplasm. *Horticulture Research*, *7*(1), 1-14.

Essam, A. M., & Ashraf, A. E. (2013). Effect of curcumin on hematological, biochemical and antioxidants parameters in *Schistosoma mansoni* infected mice. *International Journal of Sciences*, 2, 1-14.

Falconer, D. S. (1960). Introduction to quantitative genetics (Ed.3), Pearson Education, India.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567-1587.

Fatima, M., Zhang, X., Lin, J., Zhou, P., Zhou, D., & Ming, R. (2020). Expression profiling of MADS-box gene family revealed its role in vegetative development and stem ripening in S. spontaneum. *Scientific Reports*, *10*(1), 20536.

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology and Evolution*, *37*(3), 197-202.

Garcia, M., Eckermann, P., Haefele, S., Satija, S., Sznajder, B., Timmins, A., Baumann, U., Wolters, P., Mather, D. E., & Fleury, D. (2019). Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLOS ONE*, *14*(2), e0211730.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. https://doi.org/10.48550/arXiv.1207.3907

Gascon, C., Lougheed, S. C., & Bogart, J. P. (1996). Genetic and morphological variation in Vanzolinius discodactylus: a test of the river hypothesis of speciation. *Biotropica*, *28*(3), 376.

Gauthier, P., Lumaret, R., & Bédécarrats, A. (1998). Genetic variation and gene flow in Alpine diploid and tetraploid populations of Lotus (*L. alpinus* (D.C.) Schleicher/*L. corniculatus* L.). I. Insights from morphological and allozyme markers. *Heredity*, *80*(6), 683-693.

Gogoi, A., Munda, S., Paw, M., Begum, T., Siddiqui, M. H., Gaafar, A. R. Z., Kesawat, M. S., & Lal, M. (2023). Molecular genetic divergence analysis amongst high curcumin lines of golden Crop (*Curcuma longa* L.) using SSR marker and use in trait-specific breeding. *Scientific Reports*, *13*(1), 19690.

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics Research International*, Article ID 431487, 1-14.

Gubaev, R., Gorlova, L., Boldyrev, S., Goryunova, S., Goryunov, D., Mazin, P., Chernova, A., Martynova, E., Demurin, Y., & Khaitovich, P. (2020). Genetic characterization of Russian rapeseed collection and association mapping of novel loci affecting glucosinolate content. *Genes*, *11*(8), 926.

Guerra, M. (2008). Chromosome numbers in plant cytotaxonomy: concepts and implications. *Cytogenetic and Genome Research*, *120*(3-4), 339-350.

Gul, P., & Bakht, J. (2015). Antimicrobial activity of turmeric extract and its potential use in food industry. *Journal of Food Science and Technology*, *52*(4), 2272-2279.

Gupta, A. K., Mishra, R., & Lal, R. K. (2016). Genetic variability and character interrelationship among indigenous germplasm of turmeric (*Curcuma longa* L.). *Journal of Herbs, Spices and Medicinal Plants*, *22*(2), 190-201.

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Chapter Two - Association mapping in crop plants: opportunities and challenges. In T. Friedmann, J. C. Dunlap, & S. F. Goodwin (Eds.), *Advances in Genetics*, 85, 109-147.

Han, B., Huang, S., Huang, G., Wu, X., Jin, H., Liu, Y., Xiao, Y., & Zhou, R. (2023). Genetic relatedness and association mapping of horticulturally valuable traits for the Ceiba plants using ddRAD sequencing. *Horticultural Plant Journal*, *9*(4), 826-836.

Hayakawa, H., Kobayashi, T., Minaniya, Y., Ito, K., Miyazaki, A., Fukuda, T., & Yamamoto, Y. (2011). Development of a molecular marker to identify a candidate line of turmeric (*Curcuma longa* L.) with a high curcumin content. *American Journal of Plant Sciences*, *02*(01), 15-26.

Hazzouri, K. M., Gros-Balthazard, M., Flowers, J. M., Copetti, D., Lemansour, A., Lebrun, M., Masmoudi, K., Ferrand, S., Dhar, M. I., Fresquez, Z. A., Rosas, U., Zhang, J., Talag, J., Lee, S., Kudrna, D., Powell, R. F., Leitch, I. J., Krueger, R. R., Wing, R. A., Purugganan, M. D. (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications*, *10*(1), 4680.

Herben, T., Suda, J., & Klimešová, J. (2017). Polyploid species rely on vegetative reproduction more than diploids: A re-examination of the old hypothesis. *Annals of Botany*, *120*(2), 341-349.

Herry, F., Hérault, F., Lecerf, F., Lagoutte, L., Doublet, M., Picard-Druet, D., Bardou, P., Varenne, A., Burlot, T., Le Roy, P., & Allais, S. (2023). Restriction site-associated DNA sequencing technologies as an alternative to low-density SNP chips for genomic selection: a simulation study in layer chickens. *BMC Genomics*, *24*(1), 271.

Heslop-Harrison, J. S. (Pat), Schwarzacher, T., & Liu, Q. (2023). Polyploidy: its consequences and enabling role in plant diversification and evolution. *Annals of Botany*, *131*(1), 1-10.

Hewlings, S. J., & Kalman, D. S. (2017). Curcumin: A review of its effects on human health. *Foods*, *6*(10), 92.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, *8*(2), 1-12.

Huang, X., Ouyang, K., Luo, Y., Xie, G., Yang, Y., & Zhang, J. (2022). A comparative study of characteristics in diploid and tetraploid *Anoectochilus roxburghii*. *Frontiers in Nutrition*, 9, 1034751.

Islam, M. A. (2004). Genetic diversity of the genus Curcuma in Bangladesh and further biotechnological approaches for in vitro regeneration and long-term conservation of *C. longa* germplasm (PhD Thesis, University of Hannover), 1-137.

Jagadeeshkanth, R. P., Paramaguru, P., & Rameshkumar, D. (2017). Character association and path coefficient analysis in turmeric (*Curcuma longa* L.). *Journal of Pharmacognosy and Phytochemistry*, *6*(6), 979-983.

Jakubczyk, K., Drużga, A., Katarzyna, J., & Skonieczna-żydecka, K. (2020). Antioxidant potential of curcumin - a meta-analysis of randomized clinical trials. *Antioxidants*, *9*(11), 1-13.

Jamshidi, S., & Jamshidi, S. (2011, September). NTSYSpc 2.02e implementation in molecular biodata analysis (clustering, screening, and individual selection). *Proceedings of 4th International Conference on Environmental and Computer Science. Singapore* (pp. 165-169).

Jenfaoui, H., Uras, M. E., Bahri, B. A., Ozyigit, I. I., & Souissi, T. (2021). Morphological variation, genetic diversity, and phylogenetic relationships of Hypericum triquetrifolium Turra populations from Tunisia. *Biotechnology and Biotechnological Equipment*, *35*(1), 1505-1519.

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955). Estimates of genetic and environmental variability in Soybeans. *Agronomy Journal*, *47*(7), 314-318.

Joshi, R. K., Kuanar, A., Mohanty, S., Subudhi, E., & Nayak, S. (2010). Bioinformation mining and characterization of EST derived microsatellites in *Curcuma longa* L. *Bioinformation*, *5*(3), 128.

Kaeuffer, R., Réale, D., Coltman, D. W., & Pontier, D. (2007). Detecting population structure using STRUCTURE software: Effect of background linkage disequilibrium. *Heredity*, *99*(4), 374-380.

Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in Soybean and Maize. *Frontiers in Plant Science*, *10*, 486047.

Kar, B., Nayak, S., & Joshi, R. K. (2013). Development and evaluation of STS diagnostic marker to track turmeric (*Curcuma longa* L.) resistance against rhizome rot caused by *Pythium aphanidermatum*. *Australasian Plant Pathology*, *43*, 167-175.

Khajehdehi, P., Zanjaninejad, B., Aflaki, E., Nazarinia, M., Azad, F., Malekmakan, L., & Dehghanzadeh, G.R. (2012). Oral supplementation of turmeric decreases proteinuria, hematuria, and systolic blood pressure in patients suffering from relapsing or refractory Lupus Nephritis: a randomized and placebo-controlled Study. *Journal of Renal Nutrition*, *22*(1), 50-57.

Khan, M. A., & Korban, S. S. (2012). Association mapping in forest trees and fruit crops. *Journal of Experimental Botany*, *63*(11), 4045-4060.

Koo, H. J., McDowell, E. T., Ma, X., Greer, K. A., Kapteyn, J., Xie, Z., Descour, A., Kim, H. R., Yu, Y., Kudrna, D., Wing, R. A., Soderlund, C. A., & Gang, D. R. (2013). Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. *BMC Plant Biology*, *13*(1), 1-17.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, *9*, 1-9.

Ksouri, N., Sánchez, G., Font i Forcada, C., Contreras-Moreira, B., & Gogorcena, Y. (2023). ddRAD-seq-derived SNPs reveal novel association signatures for fruit-related traits in Peach. *BioRxiv*, 2023-07. https://doi.org/10.1101/2023.07.31.551252.

Kumar, A., Singh, A. K., Kaushik, M. S., Mishra, S. K., Raj, P., Singh, P. K., & Pandey, K. D. (2017). Interaction of turmeric (*Curcuma longa* L.) with beneficial microbes: A review. *3 Biotech*, *7*, 1-8.

Kumari, A., Prasad, C., & Kumar, R. (2022). Biochemical studies in different varieties of turmeric (*Curcuma longa* L.). *The Pharma Innovation Journal*, *11*(1), 1639-1645.

Kumawat, G., Kumawat, C. K., Chandra, K., Pandey, S., Chand, S., Mishra, U. N., Lenka, D., & Sharma, R. (2020). Insights into marker assisted selection and its applications in plant breeding. *Plant breeding-current and future views.* (p. Ch. 11). IntechOpen.

Lad, S. K. (1993). A case of seed-setting in cultivated turmeric types (*Curcuma longa* Linn.). *Journal of Soils and Crops*, *3*, 78-79.

Lahkar, L., & Tanti, B. (2017). Study of morphological diversity of traditional aromatic rice landraces (*Oryza sativa* L.) collected from Assam, India. *Annals of Plant Sciences*, *6*(12), 1855.

Lavania, U. C. (2005). Genomic and ploidy manipulation for enhanced production of phyto-pharmaceuticals. *Plant Genetic Resources*, *3*(2), 170-177.

Lavudya, S., Patel, M. P., Kumar, S., Singh, A., Sakure, A. A., & Pandya, M. (2024). Identification of elite genotypes through morpho-biochemical and molecular characterization of turmeric (*Curcuma longa* L.). *South African Journal of Botany*, *166*, 161-168.

Lefouili, M., & Nam, K. (2022). The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Scientific Reports*, *12*(1), 11331.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, S., Yuan, W., Deng Ping Wang, G., & Yang, P. (2011). Chemical composition and product quality control of turmeric (*Curcuma longa* L.). *Pharmaceutical Crops*, *2*, 28-54.

Liang, H., Deng, J., Gao, G., Ding, C., Zhang, L., Xu, K., Wang, H., & Yang, R. (2021). Inferring the phylogeny and divergence of Chinese Curcuma (Zingiberaceae) in the Hengduan mountains of the Qinghai–Tibet plateau by reduced representation sequencing. *Forests*, *12*(5), 520.

Lin, Y. C., Araguirang, G. E., Ngo, A. H., Lin, K. T., Angkawijaya, A. E., & Nakamura, Y. (2020). The four Arabidopsis choline/ethanolamine kinase isozymes play distinct roles in metabolism and development. *Plant Physiology*, *183*(5), 152-166.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397-2399.

Liu, C.C., Shringarpure, S., Lange, K., & Novembre, J. (2020). *Methods in Molecular Biology*, *2090*, 67-86.

Liu, J.X., Zhou, M.Y., Yang, G.Q., Zhang, Y.X., Ma, P.F., Guo, C., Vorontsova, M. S., & Li, D.Z. (2020). ddRAD analyses reveal a credible phylogenetic relationship of the four main genera of Bambusa-Dendrocalamus-Gigantochloa complex (Poaceae: Bambusoideae). *Molecular Phylogenetics and Evolution*, *146*, 106758.

Ložienė, K., Maskolaitytė, V., Labokas, J., Būdienė, J., & Vaičiulytė, V. (2023). Chemical composition of essential oils and local knowledge of Myrica gale in Lithuania. *Plants*, *12*(5), 1050.

Luiram, S., Barua, P. C., Saikia, L., Talukdar, M. C., Luikham, S., Verma, H., & Sarmah, P. (2018). Genetic variability studies of turmeric (*Curcuma longa* L.) genotypes of north eastern region of India. *International Journal of Current Microbiology and Applied Sciences*, *7*(07), 3891-3896.

Lush, J. L., & others. (1949). Heritability of quantitative characters in farm animals. Proceedings in 8 th international Congress in Genetics, Lowa state university. *Genetica Heridita* (Suppl). 356-357.

Madani, H., Escrich, A., Hosseini, B., Sanchez-Muñoz, R., Khojasteh, A., & Palazon, J. (2021). Effect of polyploidy induction on natural metabolite production in medicinal Plants. *Biomolecules*, *11*(6), 899.

Malek, M. A., Rafii, M. Y., Afroz, S. S., Nath, U. K., & Mondal, M. (2014). Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *The Scientific World Journal*, article ID 968796. 10.1155/2014/968796.

Maluszynska, J. (2003). Cytogenetic tests for ploidy level analyses—chromosome counting. In *Doubled haploid production in crop plants: a manual* (pp. 391-395). Dordrecht: Springer Netherlands.

Mamatha, K., Rao, M. B. N., Bhagavan, BVK., Umajyothi, K., Anuradha, M., & Sivaraju, K. (2020). Studies on genetic variability and heritability in Turmeric (*Curcuma longa* L.). *International Journal of Current Microbiology and Applied Sciences*, *9*(9), 329-335.

Mansouri, K., Rasoulpoor, S., Daneshkhah, A., Abolfathi, S., Salari, N., Mohammadi, M., Rasoulpoor, S., & Shabani, S. (2020). Clinical effects of curcumin in enhancing cancer therapy: a systematic review. *BMC Cancer*, *20*, 1-11.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*(2), 240-248.

Mishra, S., & Palanivelu, K. (2008). The effect of curcumin (turmeric) on Alzheimer's disease: an overview. *Annals of Indian Academy of Neurology*, *11*(1), 13-19.

Mohanta, S. (2015). Morphological and molecular screening of Turmeric (*Curcuma longa* L.) cultivars for resistance against parasitic nematode, Meloidogyne incognita. *Journal of Plant Pathology and Microbiology*, 6(5), 270-275.

Moragues, M., Comadran, J., Waugh, R., Milne, I., Flavell, A. J., & Russell, J. R. (2010). Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theoretical and Applied Genetics*, *120*(8), 1525-1534.

Nair, R. R., Dhamayanti, K. P. M., Sasikumar, B., Padmini, K., & Saji, K. V. (2004). Cytogenetics and reproductive biology of major spices. Annual Report -Indian Institute of Spices Research, 38-39.

Nair, R. R., & Sasikumar, B. (2009). Chromosome number variation among germplasm collections and seedling progenies in Turmeric, *Curcuma longa* L. *Cytologia*, *74*(2), 153-157.

Nair, R. R., Shiva, K. N., Anchu, S., & Zachariah, T. J. (2010). Characterization of open-pollinated seedling progenies of turmeric (*Curcuma longa* L.) based on chromosome number, plant morphology, rhizome yield and rhizome quality. *Cytologia*, *75*(4), 443-449.

Nambiar, M. C. (1979). Morphological and cytological investigations in the genus *Curcuma longa* L. (PhD Thesis, University of Bombay).

Naret, O., Chaturvedi, N., Bartha, I., Hammer, C., Fellay, J., & The Swiss HIV Cohort Study (SHCS). (2018). Correcting for population stratification reduces false positive and false negative results in joint analyses of host and pathogen genomes. *Frontiers in Genetics*, *9*, 266.

Nayak, S., Kumar Naik, P., Acharya, L. K., & Pattnaik, A. K. (2006). Detection and evaluation of genetic variation in 17 promising cultivars of Turmeric (*Curcuma longa* L.) using 4C nuclear DNA content and RAPD Markers. *Cytologia*, *71*(1), 49-55.

Nirmal Babu, K., Divakaran, M., Raj, R. P., Anupama, K., Peter, K. V., & Sarma, Y. R. (2015). Biotechnological approaches in improvement of spices: A review. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*, 487-516.

Ouaja, M., Bahri, B. A., Aouini, L., Ferjaoui, S., Medini, M., Marcel, T. C., & Hamza, S. (2021). Morphological characterization and genetic diversity analysis of Tunisian durum wheat (*Triticum turgidum* var. *durum*) accessions. *BMC Genomic Data*, *22*(1), 1-17.

Oza, K., Khumji, J. B., & Maitreya, B. (2021). Antifungal activity of turmeric (*Curcuma longa* L.) rhizome against different fungi mycology. *Journal of Natural Sciences and Mathematics*, *11*(64), 29014-29017.

Padmashree, R., Barbadikar, K. M., Honnappa, Magar, N. D., Balakrishnan, D., Lokesha, R., Gireesh, C., Siddaiah, A. M., Madhav, M. S., Ramesha, Y. M., Bharamappanavara, M., Phule, A. S., Senguttuvel, P., Diwan, J. R., Subrahmanyam, D., & Sundaram, R. M. (2023). Genome-wide association studies in rice germplasm reveal significant genomic regions for root and yield-related traits under aerobic and irrigated conditions. *Frontiers in Plant Science*, *14*, 1143853.

Panse, V. G., Sukhatme, P. V., & others. (1954). *Statistical Methods for Agricultural Workers,* (Ed. 3), Indian Council of Agricultural Research, New Delhi, India.

Paw, M., Munda, S., Borah, A., Pandey, S. K., & Lal, M. (2020). Estimation of variability, genetic divergence, correlation studies of *Curcuma caesia Roxb*. *Journal of Applied Research on Medicinal and Aromatic Plants*, *17*, 100251.

Peng, S., Liu, Y., Xu, Y., Zhao, J., Gao, P., Liu, Q., Yan, S., Xiao, Y., Zuo, S. M., & Kang, H. (2023). Genome-wide association study identifies a plant-height - associated gene OsPG3 in a population of commercial rice varieties. *International Journal of Molecular Sciences*, *24*(14). 11454.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: an inexpensive method for de novo SNP Discovery and genotyping in model and non-model species. *PLOS ONE*, *7*(5), e37135.

Petrova, G., Dzhambazova, T., Moyankova, D., Georgieva, D., Michova, A., Djilianov, D., & Möller, M. (2014). Morphological variation, genetic diversity, and genome size of critically endangered *Haberlea* (Gesneriaceae) populations in Bulgaria do not support the recognition of two different species. *Plant Systematics and Evolution*, *300*(1), 29-41.

Pfaffelhuber, P., Sester-Huss, E., Baumdicker, F., Naue, J., Lutz-Bonengel, S., & Staubach, F. (2022). Inference of recent admixture using genotype data. *Forensic Science International: Genetics*, *56*, 102593.

Pfeiffer, T., Roschanski, A. M., Pannell, J. R., Korbecka, G., & Schnittler, M. (2011). Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). *Journal of Heredity*, *102*(4), 479-488.

Pimentel-Gomes. (2009). *Curso de estatística experimental* (15th ed.). FEALQ: Piracicaba, Brazil, 451.

Poonia, A., & Upadhayay, A. (2015). Chenopodium album Linn: review of nutritive value and biological properties. *Journal of Food Science and Technology*, 52(7), 3977-3985.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, 4, 98.

PPV & FRA. (2009). Guidelines for the Conduct of Test for Distinctiveness, Uniformity and Stability on turmeric *(Curcuma longa L.)* India. http://plantauthority.gov.in/pdf/Turmeric.pdf.

Prajapati, K. N., Patel, M. A., Pandya, M. M., & Patel, J. R. (2014). Studies on the performance of Turmeric (*Curcuma longa* L.) accessions for rhizome yield and yield attributing parameters for Gujarat Condition. *Trends in Biosciences*, *7*(3), 204–206.

Prasad, S., & Aggarwal, B. B. (2011). *Turmeric, the Golden Spice: from traditional medicine to modern medicine* (2nd ed.).

Prasath D, K Kandiannan, N K Leela, S Aarthi, B Sasikumar, & K Nirmal Babu. (2019). Turmeric: Botany and production practices. *Horticultural Reviews,* 46, 99-184.

Prasath, D., Nair, R. R., & Babu, P. A. (2022). Effect of colchicine induced tetraploids of ginger (*Zingiber officinale* Roscoe) on cytology, rhizome morphology, and essential oil content. *Journal of Applied Research on Medicinal and Aromatic Plants*, *31*, 100422.

Rabeler, C., Chen, M., & Kaplinsky, N. (2022). BUMPY STEM is an Arabidopsis Choline/Ethanolamine Kinase required for normal development and chilling Responses. *Frontiers in Plant Science*, *13*, 851960.

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, *5*(2), 94-100.

Raghavan, T. S., & Venkatasubban, K. R. (1943). Cytological studies in the family zingiberaceae with special reference to chromosome number and Cyto-Taxonomy. *Proceedings of Indian Academy of Sciences*, *17*(4), 118-132.

Rajasekaran, S. A. (2011). Therapeutic potential of curcumin in gastrointestinal diseases. *World Journal of Gastrointestinal Pathophysiology*, *2*(1), 1.

Ramachandran, K. (1961). Chromosome numbers in genus Curcuma Linn. *Current Science* , *30*, 194-196.

Ramachandran, K. (1969). Chromosome Numbers in Zingiberaceae. *Cytologia*, 34(2), 213-221.

Ramsewak, R. S., DeWitt, D. L., & Nair, M. G. (2000). Cytotoxicity, antioxidant, and anti-inflammatory activities of Curcumins I–III from *Curcuma longa* L. *Phytomedicine*, *7*(4), 303-308.

Rawat, A., Barthwal, S., & Ginwal, H. S. (2014). Comparative assessment of SSR, ISSR and AFLP markers for characterization of selected genotypes of Himalayan Chir pine (*Pinus roxburghii* Sarg.) based on resin yield. *Silvae Genetica*, *63*(3), 94-109.

Renjith, D., Valsala, P. A., & Nybe, E. v. (2001). Response of turmeric (*Curcuma domestica* Val.) to in vivo and in vitro pollination. *Journal of Spices and Aromatic Crops*, 10(2), 135-139.

Rohlf, F. J., Applied Biostatistics, Inc., & Exeter Software (Firm). (2009). *NTSYS-pc : numerical taxonomy and multivariate analysis system*. Applied Biostatistics, Inc.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, *298*(5602), 2381-2385.

Roy, S., Kumar Verma, S., & Misra, A. K. (2011). Agro-morphological diversity in turmeric (*Curcuma longa* L.) accessions collected from north-eastern India. *Indian Journal of Agricultural Sciences*. *81*(10), 898.

Sahoo, A., Jena, S., Kar, B., Sahoo, S., Ray, A., Singh, S., Joshi, R. K., Acharya, L., & Nayak, S. (2017). EST-SSR marker revealed effective over biochemical and morphological scepticism towards identification of specific turmeric (Curcuma longa L.) cultivars. *3 Biotech*, *7*(1), 1-12.

Sahoo, A., Jena, S., Sahoo, S., Nayak, S., & Kar, B. (2016a). Resequencing of *Curcuma longa* L. cv. Kedaram through transcriptome profiling reveals various novel transcripts. *Genomics Data*, *9*, 160-161.

Sahoo, A., Kar, B., Sahoo, S., Jena, S., Kuanar, A., Parameswaran, S., Patnaik, J., & Nayak, S. (2019). De Novo transcriptome sequencing explored cultivar specific sequence variation and differential expression of pigment synthesis genes in turmeric (*Curcuma longa* L.). *Industrial Crops and Products*, *134*, 388-402.

Sahoo, A., Kar, B., Sahoo, S., Ray, A., & Nayak, S. (2016b). Transcriptome profiling of *Curcuma longa* L. cv. Suvarna. *Genomics Data*, *10*, 33-34.

Saini, R., Kumar, D., & Mittal, A. (2019). Antimicrobial and phytochemical potential of Chenopodium album Linn. *International Journal of Scientific and Technology Research*, *8*(7), 877-880.

Salimath, S. (2017). Genetic variability studies in Turmeric (*Curcuma longa* L.) under Southern dry zone of Karnataka. *International Journal of Pure & Applied Bioscience*, *5*(4), 719-722.

Sammour. (2014). Morphological, cytological, and biochemical characterization of soybean germplasm. *Research & Reviews in BioSciences*, *8*, 277-284.

Sandeep, I. S., Kuanar, A., Akbar, A., Kar, B., Das, S., Mishra, A., Sial, P., Naik, P. K., Nayak, S., & Mohanty, S. (2016). Agroclimatic zone based metabolic profiling of

turmeric (*Curcuma longa* L.) for phytochemical yield optimization. *Industrial Crops and Products*, *85*, 229-240.

Sasikumar, B. (2005). Genetic resources of Curcuma: diversity, characterization, and utilization. *Plant Genetic Resources*, *3*(2), 230-251.

Sasikumar, B., George, J. K., Zachariah, J., Ratnambav, M. J., Babu, N., & Ravindran, P. N. (1996). I1SR Prabha and IISR Prathibha-two new high yielding and high quality turmeric (*Curcuma longa* L.) varieties. *Journal of Spices and Aromatic Crops*, 5, 41-48.

Sato, D. (1948). The karyotype and phylogeny of Zingiberaceae. *Japanese Journal of Genetics*, *23*, 44.

Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. In *Plant Biotechnology Journal*, *15*(2), 149-161.

Sedgwick, P. (2012). Multiple significance tests: The Bonferroni correction. *BMJ (Online)*, *344*(7841). https://doi.org/10.1136/bmj.e509.

Selvam, R., Subramanian, L., Gayathri, R., & Angayarkanni, N. (1995). The anti-oxidant activity of turmeric (Curcuma longa). *Journal of Ethnopharmacology*, *47*(2), 59-67.

Senan, S., Kizhakayil, D., Sheeja, T. E., Sasikumar, B., Bhat, A. I., & Parthasarathy, V. A. (2013). Novel polymorphic microsatellite markers from turmeric, *Curcuma longa* L. (Zingiberaceae). *Acta Botanica Croatica*, *72*(2), 407-412.

Shamina, A., Zachariah, T. J., Sasikumar, B., & George, J. K. (1998). Biochemical variation in turmeric (*Curcuma longa* L.) accessions based on isozyme polymorphism. *Journal of Horticultural Science and Biotechnology*, *73*(4), 479-483.

Sharifi-Rad, J., Rayess, Y. El, Rizk, A. A., Sadaka, C., Zgheib, R., Zam, W., Sestito, S., Rapposelli, S., Neffe-Skocińska, K., Zielińska, D., Salehi, B., Setzer, W. N., Dosoky, N. S., Taheri, Y., El Beyrouthy, M., Martorell, M., Ostrander, E. A., Suleria, H. A. R., Cho, W. C. and Martins, N. (2020). Turmeric and Its major compound Curcumin on health: bioactive effects and safety profiles for food, pharmaceutical, biotechnological, and medicinal applications. *Frontiers in Pharmacology*, *11*, 01021.

Sharma, A. K., & Bhattacharya, N. K. (1959). Cytology of several members of Zingiberaceae. *Cellule*, *59*, 297-346.

Sheeja, T. E., Deepa, K., Santhi, R., & Sasikumar, B. (2015). Comparative transcriptome analysis of two species of Curcuma contrasting in a high-value compound curcumin: insights into genetic basis and regulation of biosynthesis. *Plant Molecular Biology Reporter*, *33*, 1825-1836.

Shen, C.H. (2019). Chapter 11 - Techniques in Sequencing. In C.H. Shen (Ed.), *Diagnostic Molecular Biology* (pp. 277-302). Academic Press.

Sigrist, M. S., Pinheiro, J. B., Azevedo-Filho, J. A., Colombo, C. A., Bajay, M. M., Lima, P. F., Camilo, F. R., Sandhu, S., Souza, A. P., & Zucchi, M. I. (2010). Development

and characterization of microsatellite markers for turmeric (*Curcuma longa* L). *Plant Breeding*, *129*(5), 570-573.

Sigrist, M. S., Pinheiro, J. B., Filho, J. A. A., & Zucchi, M. I. (2011). Genetic diversity of turmeric germplasm (*Curcuma longa* L; Zingiberaceae) identified by microsatellite markers. *Genetics and Molecular Research*, *10*(1), 419-428.

Siju, S., Dhanya, K., Syamkumar, S., Sasikumar, B., Sheeja, T. E., Bhat, A. I., & Parthasarathy, V. A. (2010b). Development, characterization, and cross species amplification of polymorphic microsatellite markers from expressed sequence tags of Turmeric (*Curcuma longa* L.). *Molecular Biotechnology*, *44*(2), 140-147.

Siju, S., Dhanya, K., Syamkumar, S., Sheeja, T. E., Sasikumar, B., Bhat, A. I., & Parthasarathy, V. A. (2010a). Development, characterization, and utilization of genomic microsatellite markers in turmeric (*Curcuma longa* L.). *Biochemical Systematics and Ecology*, *38*(4), 641-646.

Singh, D., Mishra, D. P., Dwibedi, D. K., Kumar, S., & Kumar, M. (2020). Genetic variability, and genetic advance as percent of mean in turmeric (*Curcuma longa* L.). *The Pharma Innovation Journal*, *9*(9), 402-404.

Singh, S., Joshi, R. K., & Nayak, S. (2013). Identification of elite genotypes of turmeric through agroclimatic zone based evaluation of important drug yielding traits. *Industrial Crops and Products*, *43*, 165-171.

Singh, S., Panda, M. K., & Nayak, S. (2012). Evaluation of genetic diversity in turmeric (*Curcuma longa* L.) using RAPD and ISSR markers. *Industrial Crops and Products*, *37*, 284-291.

Singh, T. J., Patel, R. K., kumar, S., Patel, N., & Patel, P. A. (2018). Molecular diversity analysis in Turmeric (*Curcuma longa* L.) using SSR markers. *International Journal of Current Microbiology and Applied Sciences*, *7*(11), 552-560.

Sivakumar, V., Umajyothi, K., Dorajeerao, A. V. D., & Umakrishna, K. (2019). Genetic variability, heritability, and genetic advance as per cent mean in turmeric (*Curcuma longa* L.) genotypes. *Journal of Pharmacognosy and Phytochemistry*, *8*(1), 1799-1801.

Sivasubramanian, S., & Menon, M. (1973). Heterosis and inbreeding depression in rice. *Madras Agricultural Journal, 60*(7), 1139-1140.

Smith, J. S. C., & Smith, O. S. (1989). The description and assessment of distances between inbred lines of maize. II: The utility of morphological biochemical, and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. *Maydica*, 34(2), 151-161.

Smits, B. M. G., van Zutphen, B. F. M., Plasterk, R. H. A., & Cuppen, E. (2004). Genetic variation in coding regions between and within commonly used inbred rat strains. *Genome Research*, *14*(7), 1285-1290.

Sood, S., Bhardwaj, V., Kaushik, S. K., & Sharma, S. (2020). Prediction based on estimated breeding values using genealogy for tuber yield and late blight resistance in auto-tetraploid potato (*Solanum tuberosum* L.). *Heliyon*, *6*(11), e05624.

Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.

Sudan, J., Singh, R., Sharma, S., Salgotra, R. K., Sharma, V., Singh, G., Sharma, I., Sharma, S., Gupta, S. K., & Zargar, S. M. (2019). ddRAD sequencing-based identification of inter-genepool SNPs and association analysis in Brassica juncea. *BMC Plant Biology*, *19*(1), 1-15.

Sugiura, T. (1931). A list of chromosome numbers in angiospermous plants. *Shokubutsugaku Zasshi*, *45*(535), 353-355.

Sugiyama, S. I. (2005). Polyploidy and cellular mechanisms changing leaf size: comparison of diploid and autotetraploid populations in two species of Lolium. *Annals of Botany*, *96*(5), 931-938.

Sun, C.H., Wang, J.H., Gu, K.D., Zhang, P., Zhang, X.Y., Zheng, C.S., Hu, D.G., & Ma, F. (2021). New insights into the role of MADS-box transcription factor gene CmANR1 on root and shoot development in chrysanthemum (*Chrysanthemum morifolium*). *BMC Plant Biology*, *21*(1), 1-16.

Sun, C.H., Yu, J.Q., Duan, X., Wang, J.H., Zhang, Q.Y., Gu, K.D., Hu, D.G., & Zheng, C.S. (2018). The MADS transcription factor CmANR1 positively modulates root system development by directly regulating CmPIN2 in chrysanthemum. *Horticulture Research*, *5*(1), 52.

Sun, Y. V., & Kardia, S. L. R. (2008). Imputing missing genotypic data of single-nucleotide polymorphisms using neural networks. *European Journal of Human Genetics*, *16*(4), 487-495.

Syamkumar S. (2008). Molecular Biochemical and morphological characterization of selected Curcuma accessions (Ph.D. Thesis, University of Calicut). https://shodhganga.inflibnet.ac.in /handle/10603/27442.

Szczecińska, M., Sramko, G., Wołosz, K., & Sawicki, J. (2016). Genetic diversity and population structure of the rare and endangered plant species *Pulsatilla patens* (*L.)* Mill in east Central Europe. *PLoS ONE*, *11*(3): e0151730.

Tahira, J. J., Khan, S. N., Suliman, R., & Anwar, W. (2010). Weed flora of *Curcuma longa* fields of district Kasur, Pakistan. *Pakistan Journal of Weed Science Research*, *16*(2), 241-246.

Tandayu, E., Borpatragohain, P., Mauleon, R., & Kretzschmar, T. (2022). Genome-wide association reveals trait loci for seed Glucosinolate accumulation in Indian Mustard (*Brassica juncea* L.). *Plants*, *11*(3), 364.

Tang, X., Xiong, K., Wassie, T., & Wu, X. (2022). Curcumin and Intestinal Oxidative Stress of Pigs With Intrauterine Growth Retardation: A Review. In *Frontiers in Nutrition*, *9*, 847673.

Tanvir, E. M., Hossen, Md. S., Hossain, Md. F., Afroz, R., Gan, S. H., Khalil, Md. I., & Karim, N. (2017). Antioxidant properties of popular Turmeric *(Curcuma longa* L.*)* varieties from Bangladesh. *Journal of Food Quality*, *2017*, 8471785. https://doi.org/10.1155/2017/8471785.

Thaikert, R., & Paisooksantivatana, Y. (2009). Variation of total curcuminoids content, antioxidant activity and genetic diversity in Turmeric (*Curcuma longa* L.) collections. *Agriculture and Natural Resources*, *43*(3), 507-518.

Tomeh, M. A., Hadianamrei, R., & Zhao, X. (2019). A review of curcumin and its derivatives as anticancer agents. *International Journal of Molecular Sciences*, *20*(5), 1033.

Trebbi, D., Maccaferri, M., de Heer, P., Sørensen, A., Giuliani, S., Salvi, S., Sanguineti, M. C., Massi, A., van der Vossen, E. A. G., & Tuberosa, R. (2011). High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theoretical and Applied Genetics*, *123*(4), 555-569.

Trujillo, J., Chirino, Y. I., Molina-Jijón, E., Andérica-Romero, A. C., Tapia, E., & Pedraza-Chaverrí, J. (2013). Renoprotective effect of the antioxidant curcumin: recent findings. *Redox Biology*, *1*(1), 448-456.

Ullah Jan, H., Ashiq Rabbani, M., & Khan Shinwari, Z. (2012). Estimation of genetic variability in turmeric (*Curcuma longa* L.) germplasm using agro-morphological traits. *Pakistan Journal of Botany*, *44*, 231-238.

Usha, H., Devi, N., & Chezhiyan, N. (2006). Impact of gamma rays on turmeric crop (*Curcuma longa* L.). *Journal of Horticultural Sciences*, *1*(2), 124-128.

Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P., & González-Candelas, F. (2021). One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLOS Computational Biology*, *17*(1), e1008678.

Valuyskikh, O. E., Shadrin, D. M., & Pylina, Ya. I. (2019). Morphological variation and genetic diversity of *Gymnadenia conopsea* (L.) R. Br. (Orchidaceae) populations in the northeast of European Russia (Komi Republic). *Russian Journal of Genetics*, *55*(2), 180-196.

Verma, S., Singh, S., Sharma, S., Tewari, S. K., Roy, R. K., Goel, A. K., & Rana, T. S. (2015). Assessment of genetic diversity in indigenous turmeric (*Curcuma longa* L.) germplasm from India using molecular markers. *Physiology and Molecular Biology of Plants*, *21*(2), 233-242.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. de F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*, 312-328.

Vijayalatha, K. R., & Chezhiyan, N. (2008). Multivariate based marker analysis in turmeric (*Curcuma longa* L.). *Journal of Horticultural Sciences*, *3*, 107-111.

Vinodhini, V., Senthamizh Selvi, B., Balakrishnan, S., & Muthuragavan, R. (2018). Studies on variability and genetic components of yield and quality traits in turmeric (*Curcuma longa* L.). *Electronic Journal of Plant Breeding*, *9*(3), 1060–1066.

Vo, T. S., Vo, T. T. B. C., Vo, T. T. T. N., & Lai, T. N. H. (2021). Turmeric (*Curcuma longa* L.): Chemical components and their effective clinical applications. *Journal of the Turkish Chemical Society Section A: Chemistry*, *8*(3), 883-898.

Wang, J., & Zhang, Z. (2021). GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, *19*(4), 629-640.

Warden, C. D., Adamson, A. W., Neuhausen, S. L., & Wu, X. (2014). Detailed comparison of two popular variant calling packages for exome and targeted exon studies. *PeerJ*, *2*, e600.

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., & Stahl, A. (2020). How Population Structure Impacts Genomic Selection Accuracy in Cross-Validation: Implications for Practical Breeding. *Frontiers in Plant Science*, *11,* 592977.

Wilken, R., Veena, M. S., Wang, M. B., & Srivatsan, E. S. (2011). Curcumin: A review of anti-cancer properties and therapeutic activity in head and neck squamous cell carcinoma. *Molecular Cancer*, *10*, 1-19.

Wu, Z., Ioannidis, N. M., & Zou, J. (2020). Predicting target genes of non-coding regulatory variants with IRT. *Bioinformatics*, *36*(16), 4440-4448.

Yang, G. Q., Chen, Y. M., Wang, J. P., Guo, C., Zhao, L., Wang, X. Y., Guo, Y., Li, L., Li, D. Z., & Guo, Z. H. (2016). Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods*, *12*(1), 1-17.

Yao Xinzhuan, Qi Yong, Chen Hufang, Zhang Baohui, Chen Zhengwu, & Lu Litang. (2023). Study of Camellia sinensis diploid and triploid leaf development mechanism based on transcriptome and leaf characteristics. *PLOS ONE*, *18*(2), 1-24.

Yao, Z., You, F. M., N'Diaye, A., Knox, R. E., McCartney, C., Hiebert, C. W., Pozniak, C., & Xu, W. (2020). Evaluation of variant calling tools for large plant genome re-sequencing. *BMC Bioinformatics*, *21*(1), 1-16.

Yildiz, M. (2013). Plant responses at different ploidy levels.

Current Progress in Biological Research (p. Ch. 16). *InTech, Rijeka*, 363-385.

Yin, Y., Xie, X., Zhou, L., Yin, X., Guo, S., Zhou, X., Li, Q., Shi, X., Peng, C., & Gao, J. (2022). A chromosome-scale genome assembly of turmeric provides insights into curcumin biosynthesis and tuber formation mechanism. *Frontiers in Plant Science*, *13,* 1003835.

Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., & Shen, B. (2011). A practical comparison of De Novo genome assembly software tools for next-generation sequencing technologies. *PLOS ONE*, *6*(3), e17915.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, *1*(1), 5-20.

APPENDIX I

Meteorological data during 2020 and 2021 at ICAR-IISR Experimental Farm, Peruvannamuzhi

Season	Month	Temperature (°C)		Rainfall	
		Maximum	Minimum	Total Rainfall (mm)	Rainy days
2020	January	34.3	21.6	0.0	0
	February	35.8	21.9	0.0	0
	March	37.1	23.7	0.0	0
	April	34.3	21.6	75.8	8
	May	34.5	24.8	511.4	20
	June	31.0	23.5	942.6	27
	July	34.5	24.8	902.6	26
	August	29.9	23.5	945.4	21
	September	30.5	23.6	933.0	22
	October	22.2	13.0	275.2	12
	November	34.4	22.5	80.4	8
	December	33.8	21.9	99.0	8
2021	January	33.5	21.5	119.0	6
	February	35.4	21.2	0.0	0
	March	35.5	23.2	23.9	4
	April	34.9	24.2	141.8	14
	May	32.9	23.3	799.3	21
	June	32.7	24.1	508.0	22
	July	29.9	23.4	868.3	27
	August	29.9	23.6	649.9	27
	September	31.0	23.8	438.3	18
	October	32.0	23.7	583.0	17
	November	32.5	23.8	455.0	18
	December	34.0	21.1	42.0	4

APPENDIX II

List of improved varieties with their institution of release and characteristic features

Improved variety	Institution of release	Characteristic features
Varna	Kerala Agricultural University (KAU), Thrissur, Kerala	Bright orange yellow rhizome, medium bold with closer internodes, medium duration crop.
Megha Turmeric	ICAR Research Complex for NEH Region, Shillong, Meghalaya	Selection from Lakadong turmeric, bold rhizomes with high curcumin content.
Suranjana	Uttar Bengal Krishi Vishwa Vidyalaya, West Bengal	Tolerant to leaf blotch and rhizome rot
Suguna	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	A high yielding and good quality variety with reddish yellow coloured rhizome.
Sudarsana	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	A high yielding high quality short duration turmeric with thick plumpy rhizomes. Field tolerant to rhizome rot.
Suvarna	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	A high yielding, short duration turmeric with deep orange coloured rhizome.
IISR Kedaram	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	Consistency in curcumin content and tolerant to leaf blotch.
IISR Pragati	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	Short duration, stable high curcumin, moderately resistant to root-knot nematode.
IISR Prabha	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	Selection from a seedling progeny, long duration, high yielding variety with high curcumin.
IISR Alleppey Supreme	ICAR-Indian Institute of Spices Research, Kozhikode, Kerala	Selection form AFT, long duration, high curcumin, tolerance to leaf blotch disease.
IISR Pratibha	ICAR-Indian Institute of Spices Research,	Selection from a seedling progeny, long duration,

	Kozhikode, Kerala	high curcumin, plumpy and bold rhizomes
Punjab Haldi 1	Punjab Agricultural University, Ludhiana, Punjab	Plants medium in height, long medium thick rhizomes. Rhizomes dark yellow in colour, harvest in 215 days.
BSR 2	Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu	It is a mutant from Erode local.The crop yields 32 t/ha in 240-250 days. The plants are medium statured, high yielding
CO 1	Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu	It is a vegetative mutant from Erode local. Rhizomes are bigger sized and bright orange coloured. Plants are tall, vigorous with more leaves and tillers. Suitable for drought and saline conditions.
CO 2	Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu	High curcumin content (4.02%), highly resistant to rhizome rot, field tolerant to thrips, shoot borer, leaf folder and scale insects, duration 250-260 days. Yield - 42 t fresh rhizome /ha.
Rajendra Sonia	Tirhut College of Agriculture, RAU, Dholi, Bihar	Bold and plumby rhizomes, high curcumin, medium dry recovery
NDH 1	Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Uttar Pradesh	High yield potential, good size and colour of rhizomes.
NDH 3	Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Uttar Pradesh	High yielding with good size fingers.

PUBLICATIONS AND PRESENTATIONS

Publications

- Aswathi, A.P., Raghav, S.B. & Prasath, D. Assessment of genetic variation in turmeric (*Curcuma longa* L.) varieties based on morphological and molecular characterization. *Genetic Resources and Crop Evolution*, 70, 147–158 (2023). https://doi.org/10.1007/s10722-022-01417-3
- Aswathi, A.P., Prasath, D. Ploidy level variation and phenotypic evaluation of turmeric (*Curcuma longa* L.) diversity panel. *Genetic Resources and Crop Evolution* (2024). https://doi.org/10.1007/s10722-023-01844-w.
- Neenu, M. G., Aswathi, P., & Prasath, D. (2023). Synthetic polyploidy in spice crops; A review. Crop Science, 1-22.

Presentations

- Oral presentation on "Characterization of turmeric genotypes based on ploidy levels by flow cytometry" in the National Conference entitled '*Enhancing Competitiveness of Horticulture through Technology Innovations*', to be organized during November 17-18, 2022, at ICAR-CPCRI, Kasaragod, Kerala.
- Poster presentation on "Genome wide association analysis for important agronomic traits in turmeric (*Curcuma longa* L.) based on ddRAD sequencing" at International Seminar on Gingers, KSCSTE-Malabar Botanical Garden and Institute for Plant Sciences, Kozhikode on March 01-03,2023.