Evaluation of Seed Traits and Determining Genetic Diversity and Population Structure in Medicinal Legume *Mucuna pruriens* (L.) DC.

A Thesis Submitted to

S I K K I M U N I V E R S I T Y

In partial fulfillment of the requirement for the

Degree of Doctor of Philosophy

By

Patrush Lepcha

Department of Botany
School of Life Sciences
Gangtok – 737102
Sikkim, India

सामदुर, तादोंग - 737102
सिक्किम, भारत
92-251212, 251415, 251656
- 251067
- www.cus.ac.in

6th Mile, Samdur, Tadong-737102
Gangtok, Sikkim, India
Ph. 03592-251212, 251415, 251656
Telefax : 251067
Website : www.cus.ac.in

सिक्किम विश्वविद्यालय
SIKKIM UNIVERSITY

(भारत के संसद के अधिनियम द्वारा वर्ष 2007 में स्थापित और नैक (एनएएसी) द्वारा वर्ष 2015 में प्रत्यायित केंद्रीय विश्वविद्यालय)
(A central university established by an Act of Parliament of India in 2007 and accredited by NAAC in 2015)

CERTIFICATE

This is to certify that the Ph.D. thesis entitled **"Evaluation of Seed Traits and Determining Genetic Diversity and Population Structure in Medicinal Legume *Mucuna pruriens* (L.) DC."** submitted to Sikkim University in partial fulfillment for the requirement of the degree of Doctor of Philosophy in Botany embodies the research work carried out by **Mr. Patrush Lepcha** at the Department of Botany, School of Life Sciences, Sikkim University. It is a record of a bonafide investigation carried out and completed by him under my supervision. He has followed the rules and regulations prescribed by the University. The results are original and have not been submitted anywhere else for any other degree or diploma.

It is recommended that this Ph.D. thesis be placed before the examiners for evaluation.

Place : Gangtok
Date : 21/09/2021

Prof. N. Sathyanarayana
Research Supervisor
Department of Botany
School of Life Sciences
Sikkim University
Gangtok

ामदुर, तादोंग - 737102
विकम, भारत
2-251212, 251415, 251656
251067
www.cus.ac.in

6th Mile, Samdur, Tadong-737102
Gangtok, Sikkim, India
Ph. 03592-251212, 251415, 251656
Telefax : 251067
Website : www.cus.ac.in

सिक्किम विश्वविद्यालय
SIKKIM UNIVERSITY
(भारत के संसद के अधिनियम द्वारा वर्ष 2007 में स्थापित और नैक (एनएएसी) द्वारा वर्ष 2015 में प्रत्यायित केंद्रीय विश्वविद्यालय)
(A central university established by an Act of Parliament of India in 2007 and accredited by NAAC in 2015)

CERTIFICATE

This is to certify that the Ph.D. thesis entitled **"Evaluation of Seed Traits and Determining Genetic Diversity and Population Structure in Medicinal Legume *Mucuna pruriens* (L.) DC."** submitted to Sikkim University in partial fulfillment for the requirement of the degree of Doctor of Philosophy in Botany embodies the research work carried out by **Mr. Patrush Lepcha** at the Department of Botany, School of Life Sciences, Sikkim University under the supervision of **Prof. N. Sathyanarayana**. The results are original and have not been submitted anywhere else for any other degree or diploma.

It is recommended that this Ph.D. thesis be placed before the examiners for evaluation.

Place : Gangtok
Date : 21/09/2021

Prof. Shanti S. Sharma
Head
Department of Botany
School of Life Sciences
Sikkim University
Gangtok

ामदुर, तादोंग - 737102
क्किम, भारत
2-251212, 251415, 251656
251067
www.cus.ac.in

6th Mile, Samdur, Tadong-737102
Gangtok, Sikkim, India
Ph. 03592-251212, 251415, 251656
Telefax : 251067
Website : www.cus.ac.in

सिक्किम विश्वविद्यालय
SIKKIM UNIVERSITY

(भारत के संसद के अधिनियम द्वारा वर्ष 2007 में स्थापित और नैक (एनएएसी) द्वारा वर्ष 2015 में प्रत्यायित केंद्रीय विश्वविद्यालय)
(A central university established by an Act of Parliament of India in 2007 and accredited by NAAC in 2015)

CERTIFICATE OF PLAGIARISM TEST

This is to certify that the plagiarism check has been carried out for the following Ph.D. thesis entitled **"Evaluation of Seed Traits and Determining Genetic Diversity and Population Structure in Medicinal Legume *Mucuna pruriens* (L.) DC."** submitted by **Mr. Patrush Lepcha** under the supervision of **Prof. N. Sathyanarayana,** Department of Botany, School of Life Science, Sikkim University with help of **URKUND SOFTWARE.** The result is 5% which is within the permissible limit (below 10% tolerance rate) as per the norms of Sikkim University.

Patrush Lepcha.
Patrush Lepcha
Signature of the candidate

N.S.
Prof. N. Sathyanarayana
Research Supervisor
Department of Botany
School of Life Sciences
Sikkim University
Gangtok

15/11/21
पुस्तकालयाध्यक्ष
Librarian
केन्द्रीय पुस्तकालय Central Library
सिक्किम विश्वविद्यालय
Sikkim University
Dr. Shri Ram
Librarian
Central Library
Sikkim University
Gangtok

DECLARATION

I do hereby declare that the present Ph.D. thesis entitled **"Evaluation of Seed Traits and Determining Genetic Diversity and Population Structure in Medicinal Legume *Mucuna pruriens* (L.) DC."** submitted by me for the award of the degree of Doctor of Philosophy (Botany) is a bonafide research work carried out by me at the Department of Botany, Sikkim University under the supervision of **Prof. N. Sathyanarayana**. The thesis contains no material which has been accepted for a degree or diploma of any other University or Institution, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgment is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Patrush Lepcha

Place : Gangtok
Date : 21/09/2021

Patrush Lepcha
Registration No.: 15/Ph.D/BOT/02
Department of Botany
School of Life Sciences
Sikkim University
Gangtok

ACKNOWLEDGEMENTS

"Listen to advice and accept instruction, that you may gain wisdom in the future."

–Proverbs 19:20

The journey of my Ph.D. program at Sikkim University was incredible and challenging as well. After years of labor in the laboratory and field, finally, the time has arrived to accomplish this goal. I always felt the warmth of God's grace and the blessing of my beloved Mother in all the difficulties throughout this journey. The completion of my Ph.D. course would have been impossible without the guidance, support, and encouragement bestowed upon me by several concerned individuals. At the end of my Ph.D. journey, I take the opportunity to convey my sincere gratitude to all people who have contributed in many ways to make this accomplishment possible.

First of all, I was fortunate to get enrolled in the Ph.D. program under the supervision of ***Prof. N. Sathyanarayana*** *(Ph.D.), Department of Botany, Dean of School of Life Sciences, Sikkim University. This is how he provided me with research opportunities and became the turning point of my academic career. I express my deep sense of gratitude and indebtedness to him for his selfless, proficient, and diligent guidance, and valuable suggestions and encouragement during this research. He has always been an excellent teacher, supportive mentor, and wonderful human being. Despite many other commitments, he has always stepped forward with solutions and suggestions whenever needed and helped made this thesis successful. It was a wonderful privilege and honor to be his Ph.D. student and work under his supervision. His systematic and meticulous way of working will always be my inspiration and it is the trait that I wish to inherit and nourish for a lifetime.*

I am extremely thankful to Prof. Shanti S. Sharma, Head, Department of Botany, Sikkim University and acknowledge all the teachers from the Department of Botany- Dr. Dhani Raj Chhetri, Dr. N. Bijayalaxami Devi, Dr. Santosh K. Rai, Dr. Arun Chettri, and Dr. Arun K. Rai for their invaluable suggestions, moral support, and motivations during the advancement of my research. I also extend my reverence and gratitude to all the non-teaching staff; Mr. Deepak Chettri, Mr. N.B. Subba, Mr. Roshan Rai, and other supporting staff from the Department of Botany for their encouragement, cooperation, and technical support during my research work.

Without the availability of resources, necessary laboratory, and field facilities in the University, my Ph.D. work would have been impossible. Thus, I extend my profound gratitude to Prof. Avinash Khare, the Hon'ble Vice-Chancellor of Sikkim University, and Prof. T.B. Subba, (former Vice-Chancellor) for allowing me to pursue my Ph.D. in this institution and permitting the use of resources and infrastructure of the University.

I acknowledge the Ministry of Tribal Affairs (MoTA), Government of India (GoI) for providing me with the National Fellowship for Higher Education of Scheduled Tribe Students (Award No: NFST-2015-17-ST-SIK-1633, dated 01/04/2016).

I am extremely thankful to Dr. Vikas Belamkar, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, the USA for guiding me to carry out the association mapping analysis and Dr. Ratan Chopra, Department of Plant and Microbial Biology, University of Minnesota, the USA for helping me in doing gene annotation. These two sections are the heart and soul of my Ph.D. thesis.

I am indebted to my two senior research scholars and brothers; Dr. Nima Tshering Lepcha and Dr. Dawa Lhendup Lepcha, for showing me the path to the Ph.D. course. It was their endorsement that made me pursue Ph.D. and they were always there to help me in the ups and downs of my life. I feel so lucky to come across such wondrous, compassionate, and pure-hearted human beings in the journey of my life.

I am thankful to my all senior and fellow research scholars from the Department of Botany - Dr. Pankaj Kr. Tripathi, Dr. Pittala Ranjith Kumar, Dr. Aditya Pradhan, Mr. Sai Sundeep, Mrs. Karuna Gurung, Mr. Bupen Roka, Ms. Aita Rani Subba, Mr. Shubhakar Gurung, Mr. Ashish Rai, Mr. Hemant Sharma, Mrs. Alice Rachel Subba, Ms. Raksha Mukhia, Ms. Puja Shashankar, Mr. Abhijeet Chhetri, Mrs. Punam Chetry, and Ms. Binu Gogoi. It was indeed a wonderful family where we learned and grew together.

I would also thank Dr. Sailendra Dewan, Mr. Aaron Lepcha, Mr. Saurav Chettri, Mr. Jiwan Gurung, Mr. Dipendra Chamlagain, Ms. Prerna Dahal, Ms. Ananta Rai, Dr. Nisal Rai, Dr. Sayak Das, Mr. Abijith Tiwari, Ms. Rachna Bista, Mr. Anirudh Rai, Mr. Milan Subba, Mr. Tushar Gogoii, Dr. Ugen Bhutia, Dr. Deep Gogoi, Mr. Dawa Sherpa, Dr. Roshan Gurung, Mr. Purna Hang Subba, and Mr. Vishal Tamang from the other departments and offices of Sikkim University. Their encouragement, support, and indelible friendship have made this Ph.D. course memorable.

My Ph.D. thesis would be incomplete if I don't acknowledge and mention my friends from all the states of Seven Sisters (Northeast India) who have selflessly helped me with a sample collection from the entire region. Their hospitality, welcoming nature and unconditional support to unknown person like me was astonishing and

overwhelming. Therefore, I acknowledge and express my sincere thanks to Mr. Mrinal Devberman and group (Tripura), Mr. Anas Khan, Ms. Bandana Pukhram, Mr. Santosh, Mr. Ojit (Manipur), Mr. Ben, Mrs. Kiki, Mrs. Nunu Jamir (Annu), and family, Mr. Grason Makma, Ms. Lochumbini Sitiri (Nagaland), Ms. Ari (Meghalaya), Mr. Mama, (Mizoram University), Mr. Rakesh Basneth (NERIST, Arunachal Pradesh), Ms. Janeshweri Basumateri, Mr. Denish Rajbongshi (Assam) for accompanying me during my field trips in their respective states.

Without the support of my family, my dream of pursuing a Ph.D. would have remained a dream. But, because of their persuasion and support, today my dream has turned into reality. Thus, I feel so privileged to be part of this wonderful family and I am extremely thankful to my Late Father Mr. James Lepcha, elder brother (Mayal), sisters (Dezy, Rupa, Eli, Hanna, Jacienth), uncles (Mr. Lukesh Lepcha, Mr. G.D. Subba, Rev. Shem Subba), aunts (Mrs. Kala Gurung, Mrs. Lediya Lepcha, Mrs. Kumari Subba), nephews (Semon, Jonah, Nathan), nieces (Carnelian, Joana), brother in law (Mr. Peter Limboo), and companion (Ms. Sadhana Rai). It was their love, prayers, support, and faith in me that has always kept me motivated in all the hardship and challenges of my life.

*Alas, my words are insufficient to express my gratefulness to my beloved Mother. It was her love, benevolence, and consistent prayer that had sustained me this far. Although she isn't here in this world, I believe she might be smiling and watching me from somewhere above with pride and joy. Thank you "***Ama***", for showing me this world and trusting me. This entire Ph.D. thesis is dedicated to you "Ama".*

Patrush Lepcha
September 2021

In the loving memory of
my beloved Ama

Mrs. Purna Maya Lepcha

CONTENTS

Page No.

ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AM	Association Mapping
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
APG	Angiosperm Phylogeny Group
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
CTAB	Cetyl Trimethyl Ammonium Bromide
DArT	Diversity Array Technology
DH	Double Haploids
DMT	N, N-Dimethyltryptamine
EH	Eastern Himalayas
EST	Expressed Sequence Tags
FAO	Food and Agriculture Organization
FDR	False Discovery Rate
GLM	General Linear Model
GMCC	Green Manure Cover Crop
GWAS	Genome-Wide Association Studies
HPLC	High Performance Liquid Chromatography
HWE	Hardy-Weinberg Equilibrium
IB	Indo-Burma
ISSR	Inter Simple Sequence Repeat
LD	Linkage Disequilibrium
L-DOPA	L-3,4-Dihydroxyphenylalanine
MAF	Major Allele Frequency

MCMC	Markov Chain Monte Carlo
MLM	Mixed Linear Model
MTA	Marker-Trait Association
NGS	Next-Generation Sequencing
NILs	Near Isogenic Lines
NJ	Neighbor Joining
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PCs	Principal Components
PEM	Protein Energy Malnutrition
PIC	Polymorphic Information Content
PVE	Phenotypic Variance Explained
QTL	Quantitative Trait Loci
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RILs	Recombinant Inbred Lines
SNP	Single Nucleotide Polymorphism
SPSS	Statistical Package for the Social Sciences Software
SSR	Simple Sequence Repeats
var.	Variety

SYMBOLS AND UNITS

%	Percentage
%CV	Coefficient of Variation
%P	Percentage of Polymorphic Loci
cM	Centi Morgan
cm	Centimeter
g	Gram
Gy	Gray
h	Hour
min	Minutes
mm	Millimeter
mM	Millimolar
ng	Nano gram
s	Seconds
SD	Standard Deviation
SE	Standard Error
Tm	Temperature
γ	Gamma-ray
µl	Microlitre
µM	Micromolar

LIST OF TABLES **Page No.**

LIST OF FIGURES **Page No.**

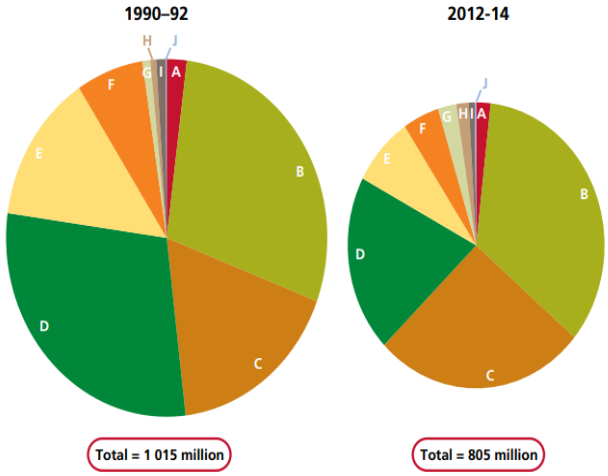
CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1. Food security and tropical agriculture

Tropical areas of the world are home to nearly 40% of the human population on this planet. Access to adequate safe and nutritional food for them has been one of the major challenges of the twenty-first century. Almost 70% of these people, living in the arid and semi-arid regions of Asia, Africa, and Latin America, depend on agriculture as their chief source of livelihood (Rockstrom 2003; Graeub et al. 2016). Ironically, a greater dearth of nutritional food and clean drinking water is also felt in these regions (Chibarabada et al. 2017). As of date, South Asia and Sub-Saharan Africa are reporting the highest number of undernourished populations, particularly women and children. This is likely to grow further in time due to rapid expansion in the population (Figure 1) (FAO et al. 2015). As a consequence, there is rising pressure on the global agriculture sector in general, and the tropical countries in particular, to produce more with the limited availability of land and water (Godfray et al. 2010).



	Number (millions)		Regional share (%)	
	1990–92	2012–14	1990–92	2012–14
A Developed regions	20	15	2.0	1.8
B Southern Asia	292	276	28.8	34.3
C Sub-Saharan Africa	176	214	17.3	26.6
D Eastern Asia	295	161	29.1	20.0
E South-Eastern Asia	138	64	13.6	7.9
F Latin America and the Caribbean	69	37	6.8	4.6
G Western Asia	8	19	0.8	2.3
H Northern Africa	6	13	0.6	1.6
I Caucasus and Central Asia	10	6	0.9	0.7
J Oceania	1	1	0.1	0.2
Total	1015	805	100	100

Figure 1. The numbers and the shares of undernourished people by different regions of the world. Note: The areas of the pie charts are proportional to the total number of undernourished in each period (Source: FAO et al. 2015).

In most of the tropics, major crops are staples such as rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), and maize (*Zea mays* L.) in addition to root and tuber crops such as Irish potato (*Solanum tuberosum* L.), sweet potato [*Ipomea batatas* (L.) Poir.], and cassava (*Manihot esculenta* Crantz.). But their production and nutritional quality are inadequate to address the food security challenges of these regions. Many of them are poor in protein and micronutrient content (Sarwar 2013; Bohra et al. 2015). This has caused a greater prevalence of protein-energy malnutrition (PEM) and micronutrient deficiency in the region (Bhat and Karim 2009). It is, therefore, crucial to develop an alternative source of protein and micronutrient diet for these regions. Introduction of soybean [*Glycine max* (L.) Merr.] was endeavored with much enthusiasm earlier as it is the most popular source of plant edible protein. But its production could not match with that realized in the temperate regions due to climatic factors (Saryoko et al. 2017). Soybean is a short-day and photoperiod-sensitive species best suited in temperature between 20-30 °C (de Avila et al. 2013); when introduced in the tropical areas, it flowered early and produced low-yield (Lyu 2017). Thus, soybean cultivation has not been rewarding.

Tapping the potential of underutilized tropical legume species offers a viable alternative. Many of these are rich in protein, low saturated fats, and micronutrient contents (zinc, calcium, folate, and tocopherols) (Seena and Sridhar 2005; Boschin and Arnoldi 2011; Akinyele and Shokunbi 2015). Further, improving their productivity is also easier to achieve, as they are adapted to local conditions, and in turn, can augment the nitrogen-deficient tropical soils for fertility (Chibarabada et al. 2017; Chivenge et al. 2015; Mabhaudhi et al. 2016). Therefore, there is a need to refocus our efforts to mainstream the lesser-known legume crops through the deployment of modern breeding tools.

1.2. Legumes and tropical agriculture

Tropical agriculture is characterized by prolonged dry seasons, low or erratic rainfall, and less fertile soils (Varshney et al. 2013). In recent years, increased episodes of mid-season drought are being witnessed possibly due to climate change (Serdeczny et al. 2017). In addition, only 5% of the total smallholder farms have irrigation (Rosegrant et al. 2009), seriously hampering crop and livestock production (Muoni 2019). As most of the smallholder farmers still follow age-old agricultural practices involving hand hoes and ox-drawn mouldboard plows (Zingore et al. 2008), soil fertility is depleting due to the loss of organic matter and erosion (Amini et al. 2015). Intensified use of legumes can help ameliorate many of these challenges (Garcia-Estringana et al. 2013).

Legumes offer a distinct advantage over other crops due to their multiple uses (Graham and Vance 2003; Mousavi-Derazmahalleh et al. 2019). They are the crucial component of the low-input agriculture system (Massawe et al. 2016) and an important source of protein, particularly for low-income families and/or individuals under the vegan diet (Young et al. 2003). Their protein content ranges from 5% to 39% with white lupin (*Lupinus* spp.) and soybean among the richest sources (Messina 1999; Vecerek et al. 2008). Grain legumes are also rich in vitamin A, minerals such as iron (Fe), zinc (Zn), calcium (Ca), etc. (Chibarabada et al. 2017). Thus, they not only enhance the dietary diversity but also complement the protein and micronutrient deficiency in tropical diets (Iqbal et al. 2006; Chibarabada et al. 2017). Besides, grain, leaves, and husks of several legumes can be used as animal feed (Crepon et al. 2010; Jezierny et al. 2010). Some medicinal legumes are the source of antibiotics, alkaloids, flavonoids, glycosides, and phytochemicals (Tyler et al. 1976; Morris 2003). The overview of the benefits offered by the legumes is given in Figure 2.

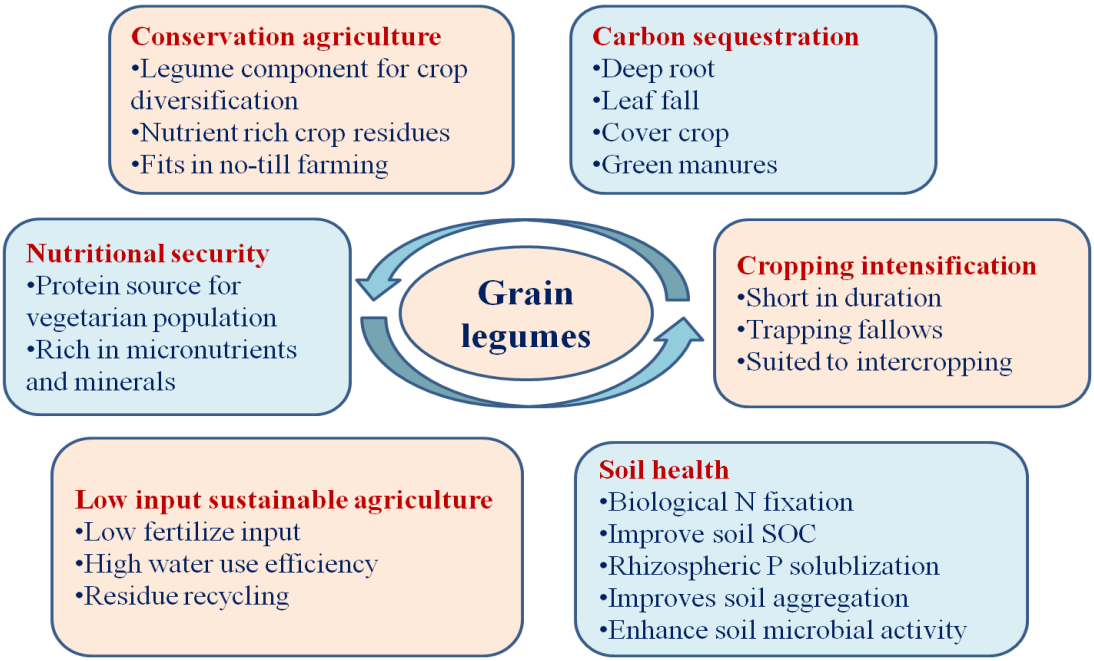


Figure 2. The benefits of legumes in soil-plant systems and nutrition (Source: Kumar et al. 2018).

Another significant characteristic of legume species is the symbiotic nitrogen-fixing ability, which increases soil fertility (Zahran 1999). They can also be used as a cover crop to conserve soil and water (Muoni 2019). Their ability to provide dense soil cover prevents the soil-moisture evaporation by reducing the effect of the direct heat (Farzi et al. 2017). Most importantly, they contribute to income generation through the sale of legume-based products including grain, livestock feed, etc. (Muoni 2019). But the success of the latter critically depends upon the market availability and access to value chains in different stakeholder countries (Muoni 2019).

About 30 legume species are cultivated across the tropical regions (Chibarabada et al. 2017). Soybean, peanut (*Arachis hypogea* L.), chickpea (*Cicer arietinum* L.), pigeon pea [*Cajanus cajan* (L.) Millsp.], cowpea [*Vigna unguiculata* (L.) Walp.], and common bean (*Phaseolus vulgaris* L.) contributes about 90% of the total production;

the remaining 10% or less are accounted for by faba bean (*Vicia faba* L.), common pea (*Pisum sativum* L.), lablab bean [*Lablab purpureus* (L.) Sweet], Bambara groundnut [*Vigna subterranean* (L.) Verdc.], lentil (*Lens culinaris* Medik.), etc. (Abate et al. 2012). The comparative production trends of legumes across the world, arid and semi-arid tropics are given in Table 1.

Table 1. Production trends (2010-2012) of some grain legumes in the world and arid and semi-arid tropics (Sub-Saharan Africa, and South Asia) (Source: Abate et al. 2012, Nedumaran et al. 2015, Chibarabada et al. 2017).

Crop	Area (1000 ha)	Yield (kg/ha)	Production (1000 Metric Ton)	% of World Production
World				
Chickpea	10,914	818	8929	-
Common bean	27,232	723	19,705	-
Cowpea	14,500	454	6155	-
Peanut	22,633	1607	36,379	-
Pigeon Pea	4655	885	3463	-
Soybean	92,622	2348	217,397	-
Lentil	3571	1904	2900	-
Sub-Saharan Africa				
Chickpea	398	769	315	3.5
Common bean	5190	596	3045	16
Cowpea	11440	450	5145	84
Peanut	9057	1007	8942	40
Pigeon Pea	499	729	363	10
Soybean	1228	1060	1279	1.3
Lentil	100	1094	90	2
South Asia				
Chickpea	8334	855	6792	76
Common bean	11,532	985	5908	30
Cowpea	159	975	154	3
Peanut	7038	1122	8457	31
Pigeon Pea	4118	840	3068	88
Soybean	8490	1275	5735	9.2
Lentil	1700	633	1088	33

1.3. Underutilized tropical legumes

Underutilized or orphan crops are defined as the minor crops of regional importance, but with minimum economic relevance at the global level (Naylor et al. 2004; Cullis and Kunert 2017). They are used for food, feed, and fodder (Tadele 2009). Due to the lack of economic significance, they are neglected by the scientific community and industries (Foyer et al. 2016). But in recent years, unlocking the nutritional and genetic potential of underutilized crops has been pursued as the key strategy to address global food security challenges (Cheng 2018).

There are several tropical underutilized legume crops whose potential is untapped and underexploited (Sathyanarayana et al. 2016). This includes Bambara groundnut, faba bean, lablab bean, winged bean [*Psophocarpus tetragonolobus* (L.) DC.], azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi], rice bean [*Vigna umbellate* (Thunb.) Ohwi & Ohashi], moth bean (*Vigna aconitifolia* L.), Lathyrus pea (*Lathyrus sativus* L.), horse gram [*Macrotyloma uniflorum* (Lam.) Verdc.], and velvet bean [*Mucuna pruriens* (L.) DC.]. Studies from across the disciplines over the past two decades have reaffirmed their prospects for nutritional security (Massawe et al. 2016; Mabhaudhi et al. 2019). Recently, they have gained unprecedented visibility as the demand for alternative sources of protein is raising (Pugalenthi et al. 2005). They possess excellent nutritional value and offer a vital source of protein, dietary fiber, essential amino acids, polyunsaturated fatty acids (PUFAs), vitamins, and essential minerals along with beneficial bioactive compounds (Palai et al. 2019). A broad overview of the uses and nutritional potential of underutilized legumes are given in Figure 3 and Table 2.

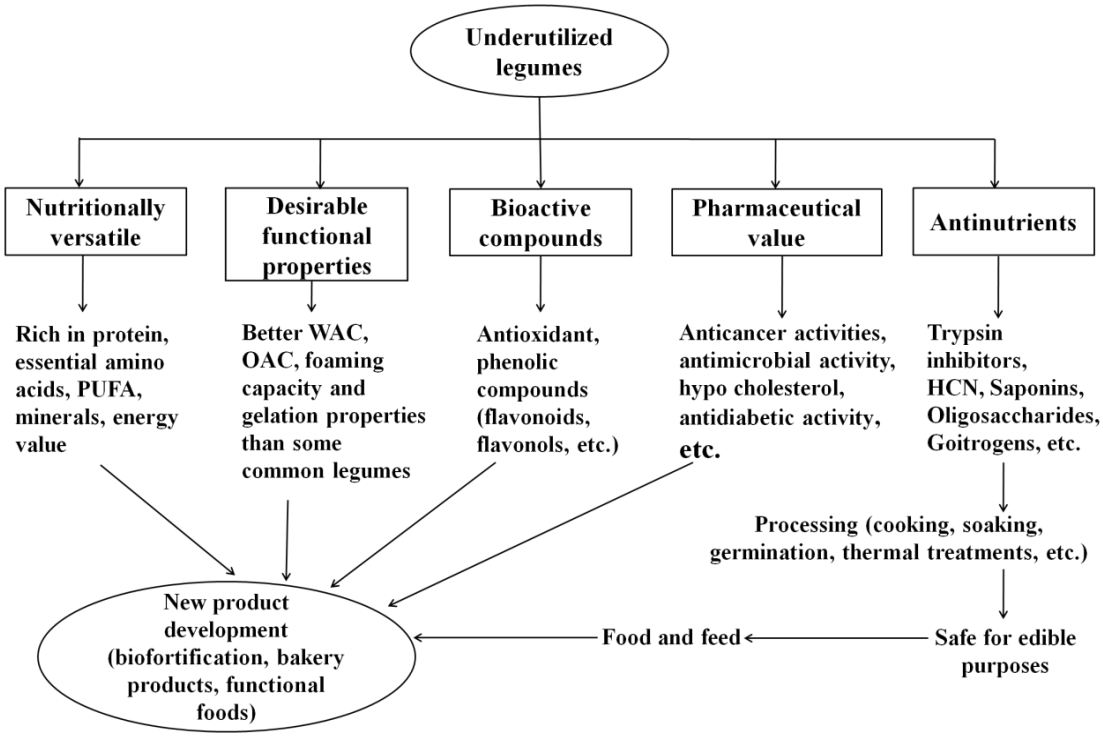


Figure 3. The potential value and the possible utility of the underutilized legumes for developing new products (Source: Bhat and Karim 2009).

Table 2. Nutritional properties of some underutilized legumes (Source: Palai et al. 2019).

Legumes	Protein (g)	Fat (g)	Carbohydrate (g)	Fiber (g)
Lathyrus pea	28	1	47	2
Winged bean	30-40	15-20	35-45	6-7
Jack bean	29-30	-	50.77-54.28	7.34-9.9
Bambara groundnut	20	6-8	60	3-6
Kidney bean	24	0.8	60.0	25.0
Faba bean	26.12	1.53	58.59	25.0
Lima bean	19-25	1-2	70-75	4-6
Cowpea	8.0	0.5	21.0	7.0
Horse gram	22	0.5	57.2	5.3
Rice bean	20.9	0.9	60.7	4.0
Moth bean	23	1.6	62	5.0
Velvet bean	20.2-29.3	6.3-7.4	49.9-61.2	8.7-10.5
Adzuki beans	20.0	0.5	6.0	13.0
Chickpea	19.0	6.0	61.0	17.0

Further, the underutilized legumes are capable of adapting to adverse environmental conditions including extreme temperatures, soil salinity, and pH (Padulosi et al. 2011; Ebert 2014; Dhillon and Tanwar 2018). For example, the Bambara groundnut is adapted to acidic soils and has been well-integrated into cereal-based cropping systems (Musa et al. 2016). It is more resistant to biotic (pests, diseases) and abiotic (drought) stresses compared to many other commonly grown legumes like peanut and cowpea (Heller 1997). The crop has become the cheap source of protein for poor farmers residing in several semi-arid areas (Yao et al. 2015).

Furthermore, as the consumption of legumes at the global level is poised to increase in both developed and underdeveloped countries due to increasing preference for the vegetarian diet (Cheng et al. 2019), the global market for underutilized legumes may attain new niches. Nonetheless, the 'omics' level characterization of these species is still in the nascent stage and a large number of them have remained unexplored at the proteomic, transcriptomic, and genomic levels (Cullis and Kunert 2017). Since these resources are crucial for the molecular breeding approach, initiatives like the African Orphan Crops Initiative (http://africanorphancrops.org), etc., have been undertaken to make up the resource gap. But research investments hitherto are far from satisfactory to realize their full potential for the global community (Cullis and Kunert 2017).

1.4. Genomic resources in underutilized legumes

In recent years, next-generation sequencing (NGS) has advanced drastically and becoming a cost-effective tool, opening up newer opportunities for the improvement of underutilized legumes (Kulski 2016). During the past decade, the genome sequence of soybean (Schmutz et al. 2010) was the only reference genome available for work on underutilized legume species. Now the information is available for many other

legumes such as barrel medic (*Medicago tranculata* Gaertn.) (Young et al. 2011), common bean (Schmutz et al. 2014), wild peanuts (*Arachis duranensis* and *A. ipaensis* Krapov. and W.C. Greg.) (Bertioli et al. 2016), and cultivated peanut (*A. hypogaea* L.) (Zhuang et al. 2019). Similar efforts have also been initiated in the nutrient-dense underutilized legumes like pigeon pea (Singh et al. 2012), lablab bean, Bambara groundnut (Chang et al. 2019), and cowpea (Xia et al. 2019). The increasing availability of such genomic resources has added valuable information about agriculturally important genes in these legume species (Dhaliwal et al. 2020). They are proving to be useful to understand the crucial gene families, re-arrangements of chromosomal structure, and evolution in the related legume species (Dhaliwal et al. 2020). Genome sequencing is also playing an important role in comparative genetics as it facilitates researchers to compare genomes, identify orthologous and paralogous genes, transfer traits and markers, and provide a comprehensive insight into domestication and evolution (Huang et al. 2018). The progress of genomic resource development in some underutilized legumes is outlined in Table 3.

Table 3. Genomic resources development in some underutilized legumes (Source: Dhaliwal et al. 2020).

Crop	Objective	Description	Reference
Pigeon pea	Transcriptome seq	50,566 SSRs, 12,000 SNPs, 0.12 million unique sequences and 150.8 million sequence reads	Dubey et al. 2011
	RNA-seq	1.696 million reads, 3771 SSRs	Dutta et al. 2011
	Comparative transcriptome	*Cajanus cajan* (L.) and *Cajanus platycarpus* (Benth.) sequence revealed 0.11 million transcripts, 82%annotated	Rathinam et al. 2019
Cluster bean	RNA-Seq	5773 SSR, 3594 SNPs, 62,146 unigenes with mean 679 bp length, and 11,000 genes annotated for biochemical pathways	Tanwar et al. 2017
	RNA-Seq	127,706 transcripts, 48,007 non-redundant unigenes, 79% annotations,8687 SSRs	Rawal et al. 2017
	Whole-genome sequencing	1859 SSRs from 1091 scaffolds constituting 60% genome of the cluster bean	Tribhuvan et al. 2019
Winged bean	CPP34 (PI 491423) andCPP37 (PI 639033) accessions	16,115 total contigs, 12,956 SSRs and 5190 SNPs developed	Vatanparast et al. 2016
	Tissue-specific (leaf, pod root, and reproductive tissues)	198,554 contigs, 24,598 SSR motifs detected	Wong et al. 2017
Cowpea	SNP chip-Cowpea iSelect Consortium Array	51 128 SNPs obtained by WGS sequencing of 37 different cowpea accessions	Muñoz-Amatriaín et al. 2017
Lablab bean	ORCAE-AOCC	Genomic portal for orphan crops such as Dolichos bean	Yssel et al. 2019

1.5. Molecular markers in underutilized legumes

Advancements in genomics platforms and techniques have accelerated the plant breeding efforts in several lesser-known legume species (Varshney et al. 2005), particularly for large-scale molecular markers and genome sequence development (Mousavi-Derazmahalleh et al. 2019). Earlier, the molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSR), and amplified fragment length polymorphism (AFLP) were deployed largely to assess the genetic diversity, etc., in underutilized legumes like cowpea (Fang et al. 2007; Zannouou et al. 2008), pigeon pea (Malviya and Yadav 2010), lablab bean (Kinmani et al. 2012), winged bean (Chen et al. 2015), cluster bean (*Cyamopsis tetragonoloba* L.) (Gresta et al. 2016), etc. But, their utility was constrained due to dominant nature (except RFLP) and reproducibility issues. Therefore, the development of species-specific co-dominant markers, which deliver better reproducibility, polymorphism, and trait-mapping abilities, was on the priority agenda for these species (Mondini et al. 2009).

In this context, one of the major contributions of genome and transcriptomic sequencing has been the rolling-out of a large number of species-specific microsatellite or simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) markers for some of the underutilized legume species. These sequence-based molecular markers are providing better reliability for trait mapping, linkage analysis, and fine-mapping (Dhaliwal et al. 2020). SNPs are ubiquitous and abundant in the genome due to which they are the marker of choice. In legumes, they have been developed in cowpea (Muchero et al. 2009), lablab bean (Venkatesha et al. 2013), pigeon pea (Saxena et al. 2014), winged bean (Vatanparast et al. 2016), cluster bean (Thakur and Randhawa 2018), etc. In addition, the application of hybridization-

based diversity array technology (DArT) has also increased recently. Despite these developments, SSRs are still a preferred marker system for breeding programs (Varshney et al. 2013), as they are cost-effective and even transferable among the related species (Mason 2015). In the case of the wild species, they can be employed for wide-ranging applications such as estimating the genetic diversity, gene flow, crossing over rates, and evolutionary studies; and in the case of the cultivated species, they find extensive utility for genotyping of the cultivars, estimating kinships, QTL mapping and marker-assisted selections (Jonah et al. 2011; Kalia et al. 2011). Hundreds of species-specific SSR markers are now available for the underutilized legumes such as cowpea (Ogunkanm et al. 2008), pigeon pea (Varshney et al. 2012), winged bean (Yang et al. 2018), cluster bean (Kumar et al. 2020), velvet bean (Sathyanarayana et al. 2017), etc. Besides, transferability of the inter-species SSRs have been successfully tested for lablab bean (Rai et al. 2016; Shivakumar et al. 2017), grass pea (*Lathyrus sativus* L.) (Shiferaw 2013), velvet bean (Shetty et al. 2015), African yam bean [*Sphenostylis stenocarpa* (Hochst Ex A. Rich.) Harms] (Shitta et al. 2015), etc.

1.6. QTL mapping

Most of the agronomically important traits such as yield, quality, disease resistance, etc., are governed by many genes, and such traits are known as quantitative, complex, or polygenic traits (Collard et al. 2005). They are influenced by the environment as well as the gene-environment interactions (Stich and Melchinger 2010). After the inception of the molecular marker technology in crop breeding, QTL (quantitative trait loci) mapping has been extensively used to identify the chromosomal location of the several genes controlling complex traits in many crop species (Mohan et al. 1997). The QTL mapping is based on the idea that the segregation of DNA markers in the

offspring of genetically divergent parents can help determine their relative distance on the chromosome (Paterson 1996). Thus, it requires a bi-parental mapping population. Although, a wide-varieties of them are available such as F_2, test cross, and first back cross, the most effective, commonly used, and reproducible are recombinant inbreed lines (RILs), near-isogenic lines (NILs), and double haploids (DH) which achieves homozygosity up to 99% (Taylor 2019). Except for the DH population which can be developed quickly (but only possible in a few species due to the requirement of anther/pollen culture protocol), the genetic stability for other mapping populations requires several generations of backcrossing or self-pollination (Taylor 2019). Once produced, the bi-parental mapping population is genotyped using the high-throughput genomic technologies followed by the development of high-resolution linkage/QTL maps. Recently, the QTL mapping has been successfully carried out in many underutilized legumes to dissect the important agronomic traits (Table 4).

However, the conventional linkage analysis and/or QTL mapping suffers from several limitations. Firstly, the development of bi-parental mapping populations is time-consuming (Ambreen et al. 2018). Secondly, the possible allelic variation would be limited as allelic diversity is restricted to what is present in the parents of specific F_2 cross or within the RIL population (Borevitz and Nordborg 2003). Thirdly, it might yield low-resolution maps as only fewer recombination events may occur during the development of the F_2 or RIL population (Flint-Garcia et al. 2005).

Table 4. QTL mapping in some underutilized legumes (Source: Dhaliwal et al. 2020).

Crop	Molecular marker/QTL	Trait/Objective	Reference
Pigeon pea	*qSMD4* major QTL and minor QTLs	Sterility mosaic resistance	Gnanesh et al. 2011
	13 QTLs for six traits	Earliness, plant type, high-density linkage map	Kumawat et al. 2012
	339 SSR,4 QTLs	Linkage map, fertility restoration	Bohra et al. 2012
	*C.cajan*_01839 for sterility mosaic,*C. cajan*_03203 for Fusarium wilt	Fusarium wilt, sterility mosaic disease	Singh et al. 2016
	3 major QTLs (CcLG11)	Sterility mosaic resistance	Saxena et al. 2017a
	547 SNP (bead-array), 319 SNP (RAD), 65 SSR	Molecular linkage map	Arora et al. 2017
	CcLG08 carry major QTL	Fertility restoration	Saxena et al. 2017b
	CcLG07 (8 QTLs), SNP S7_14185076 (linked to 4 traits)	Yield related traits	Saxena et al. 2020
Cluster bean	*L19, D1, AB7* and *QLTY 3* (Bacterial blight) and *OPQ 20,OPD10,OPD14,OPQ 12, OPAC 8* and *OPF 9* (drought tolerance) QTLs	Mapping Bacterial blight resistance and drought tolerance	Bajaj 2011
Lablab	127 RFLP, 91 RAPD	F_2 population for genetic linkage map	Konduri et al. 2000
	41 main effect QTLs (22 for growth phenological traits and 19 for fruit traits)	Growth phenological and fruit traits	Yuan et al. 2009
	40 QTLs (8.1 to 55.0% variation)	Inflorescence length traits	Yuan et al. 2011
	PvTFLy1 locus	Photoperiod responsive flowering	Ramtekey et al. 2019
Cowpea	191 SNP and 184 SSR loci 3 QTLs for pod tenderness, 2 QTLs for total soluble solid	Molecular linkage map Pod tenderness and total soluble solid	Xu et al. 2011 Kongjaimun 2013
	Major QTLs on LG 11	Days to first flowering, leaf senescence, nodes to first flower, and pod number per plant	Xu et al. 2013
	QTLs on LG 1,4,7	Pod fiber content and pod shattering	Suanum et al. 2016
	Ruv2 locus *qCel7.1*, *qHem7.1*, and *qLig7.1*	Rust resistance Pod fiber content	Wu et al. 2018 Watcharatpong et al. 2020

1.7. Association mapping

Association mapping (AM) overcomes many of the above limitations of conventional linkage mapping/QTL mapping and offers a faster and efficient means to dissect the complex traits (Yu and Buckler 2006; Oraguzie and Wilcox 2007; Abdurakhmonov and Abdukarimov 2008). The AM was first initiated in human genetics to identify the loci influencing the diseases (Burton et al. 2007). In the case of the plants, it was first reported for maize flowering time (Thornsberry et al. 2001). Since then, AM has been widely utilized in several important crop plants with greater advances in statistical methods, high-throughput genotyping technologies, and increased interest in the identification of novel alleles (Zhu et al. 2008). The AM has several advantages over traditional QTL mapping as it is (i) time and cost-effective as natural population or germplasm collection is used instead of bi-parental mapping population, (ii) generates higher map resolution as it allows the exploitation of all the recombination events that occurred during the evolutionary history of a plant (Figure 4), and (iii) facilitates to test a larger number of alleles as compared to traditional QTL mapping (Mir et al. 2012).

The AM analysis determines the relationship between the molecular marker and the trait based on linkage disequilibrium (Flint-Garcia et al. 2003) which is defined as a non-random association between the alleles at different loci in a particular population (Hill and Robertson 1968). However, it can also face serious setbacks if not executed properly. For instance, false-positive associations resulting from a strong population structure have long been considered as a hindrance to detect reliable marker-trait associations (MTAs) (Zhu et al. 2008). To overcome this and other such experimental limitations, various statistical models including the mixed linear model (MLM) have been suggested. The MLM incorporates both population structure (Q matrix), and

family relatedness or kinship (K matrix) which reduces the rate of false positives (Yu et al. 2006; Zhang et al. 2010). Defining the correct *P*-value threshold is another critical step to discern the true positives (Kaler and Purcell 2019). False discovery rate (FDR), Sidak correction, Bonferroni correction, permutation test, and Bayesian approaches are some of the methods used to determine statistically significant MTAs. In the case of crop plants, FDR and Bonferroni corrections have been widely used for such rectifications (Kaler and Purcell 2019).

To date, the AM analysis has yielded beneficial results in several underutilized legume species including root architecture traits in cowpea (Burridge et al. 2017), fusarium wilt resistance in pigeon pea (Patil et al. 2017), seed-related traits in peanut (Zhao et al. 2017), iron and zinc concentration in lentil (Singh et al. 2017), flowering, pod yield per plant and fresh pod per pant in lablab bean (Vaijayanthi et al. 2018), etc.

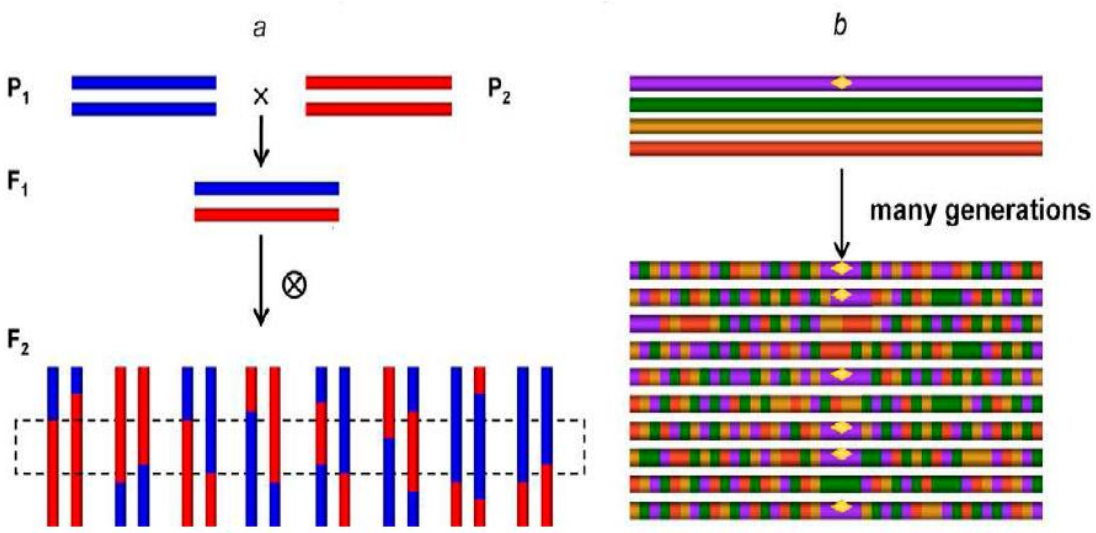


Figure 4. Schematic comparison of conventional linkage mapping and association mapping **(a)** linkage analysis using F_2 design as an example; only a few possibilities of recombination to occur within family and pedigree with known ancestry resulting in low mapping resolution **(b)** association mapping (showing only in haplotype); historical recombination and natural genetic diversity are accounted for high mapping resolution (Source: Zhu et al. 2008).

1.8. The significance of *Mucuna pruriens* (L.) DC.

Mucuna pruriens (L.) DC. (Figure 5) is a promising legume for tropical agriculture as it offers a multitude of uses such as food, feed, fodder, and cover crop (Siddhuraju et al. 2000; Siddhuraju and Becker 2001). It is a well-known source of L-3,4-dihydroxyphenylalanine (Levodopa or L-DOPA) - a choice drug in the treatment of Parkinson's disease (Soares et al. 2014) and other neurologically active compounds such as bufotenine, serotonin, nicotine, 5-HTP, N, N-DMT, etc. (Chikagwa-Malunga et al. 2009). The Ayurvedic system of medicine has long been using it for the treatment of neuronal disorders, diabetes, gout, tuberculosis (Sathiyanarayanan and Arulmozhi 2007). Additionally, it exhibits anti-diabetic, anti-neoplastic, anti-microbial, aphrodisiac, learning, and memory-enhancing properties (Poornachandra et al. 2005).

From an agronomic standpoint, it grows well in the less fertile and dry soils (Siddhuraju et al. 2000), produces a seed yield of about 1.3-2.4 t/ha (Kumwenda and Gilbert 1998; Gurumoorthi et al. 2003), and possesses disease-resistant (Eilitta et al. 2002), nematicidal (Carsky and Ndikawa 1998) as well as allelopathic properties (Fujii et al. 1991). Its effectiveness as a green manure cover crop (GMCC) due to its high N_2 fixing ability has earned it the name 'magic bean' (Eilitta et al. 2003). Besides, cultivated *M. pruriens* (var. *utilis*, velvet bean) contain high seed protein (20-30%; Buckles 1995) akin to other popular legumes such as soybean, pigeon pea, chickpea (Kumar et al. 1991; Hira and Chopra 1995; Mang et al. 2016) offering a cheap source of edible protein. Seed powder is a source of high-value industrial starch (Lawal and Adebowale 2004) and can be supplemented with livestock feed (Burgess et al. 2003).

Figure 5. *Mucuna pruriens* (L.) DC. **(a)** habit **(b)** flower **(c)** different stages of fruit.

Identifying/tagging loci influencing the key agronomic traits is lacking in many underutilized crop species, *M. pruriens* being no exception. In the realm of molecular breeding, so far, only two linkage maps based on AFLP markers for the floral, pod, and seed traits (Capo-chichi et al. 2004; Mahesh et al. 2016) and one based on SSR marker (Kumar 2019) have been developed. There aren't any efforts on the association mapping so far. Therefore, the present research work was attempted to identify and validate the *M. pruriens* association panel in addition to determine significant marker-trait association for key agronomic traits related to seed, pods, and inflorescence.

CHAPTER 2

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1. Botany

The genus *Mucuna* is classified under the phaseloid clade of Leguminosae. It comprises about 150 species of annual and perennial legumes of pantropical distribution (Buckles 1995). About nine of them are recorded from the Indian subcontinent with the majority being perennial species except for *M. pruriens* (Wilmot-Dear 1987). The plant is known by several common names such as velvet bean, Bengal bean, Mauritius bean, itchy bean, buffalo bean, cowhage, cow-itch, etc. It is a self-pollinated species with a diploid genome (2n = 2x = 22) of approximately 1281 to 1361 Mbp (Sastrapradja et al. 1974; Sathyanarayana et al. 2017), and occurs both in wild (var. *pruriens* and var. *hirsuta*) and cultivated (var. *utilis*) forms (Figure 6) (Wilmot-Dear 1987; Sasidharan 2004). The wild varieties are known as "cow-itch" or "itching bean" as they possess highly itching dense trichomes on the pod, while the cultivated variety is called "velvet bean" as they contain non-itching trichomes (Pugalenthi and Vadivel 2007a, 2007b). The systematic position and general characteristics of *M. pruriens* are presented in Tables 5 and 6, respectively.

Table 5. Systematic position of *M. pruriens* (L.) DC. based on APG-II.

Division	Tracheophyta
Class	Magnoliopsida
Sub-class	Rosidae
Order	Fabales
Family	Fabaceae
Sub-Family	Papilionoideae
Tribe	Phaseoleae
Genus	*Mucuna* Adans.
Species	*pruriens* (L.) DC
Variety	*pruriens* (L.) DC., *hirsuta* (Wight and Arn.) Wilmot-Dear, *utilis* (Wall. ex Wight) Bak. ex Burck

Figure 6. Varieties of *M. pruriens* (**a)** var. *pruriens* (**b)** var. *hirsuta* (**c)** var. *utilis*.

Table 6. Characteristics of *Mucuna pruriens* (L.) DC. (Source: Sastry and Kavathekar 1990; Agharkar 1991; Rastogi and Mehrotra 1994).

Habit	Annual herbaceous climber with long thin branches; grows up to 3 to 18m in height
Leaves	Trifoliolate; lanceolate; opposite; up to 15 to 30 cm in length
Leaflets	Ovate; rhomboid ovate, or elliptic; unequal at the base
Flower	Yellowish-white to dark-purple in color
Inflorescence	Long clusters or pendulous racemes
Fruit	Pod: leathery, thick, dark green or silvery grey; curved, longitudinally ribbed, turgid. In wild varieties (var. *pruriens* and var. *hirsuta*), it is covered with reddish-orange colored trichomes that are easily dislodged and cause an itching sensation while collecting the plant in the field. While itching trichomes are absent in cultivated variety (var. *utilis*).
Seed	Black, brown, dark brown, white, or grey, and sometimes black /brown mottled seed coat patterns are also found. The number of seeds ranges from 4-6 in a pod, which may be oval, cuboid, or round which are 6-12 mm long.

2.2. Distribution

M. pruriens flourishes well under the acidic soil (pH < 5.8), elevation below 1,600 m, annual rainfall > 400 mm, warm (19-27°C), and moist tropical climatic conditions (Pugalenthi et al. 2005; Kumar and Saha 2013). It is native to Eastern India and Southern China (Duke 1981; Wilmot-Dear 1987); but is now distributed across Asia, the Americas, West Indies, Africa, and the Pacific Islands (Figure 7) (Fung et al. 2011). In Southeast Asia, it is mainly found in India, Nepal, Bangladesh, Myanmar, Sri Lanka, and Malaysia (Fung et al. 2011; Kumar and Saha 2013). Within the Indian sub-continent, its distribution ranges from the lower Himalayan range to the entire tropical plains of India (Muralia and Pathak 2003). In this region, it is found in dry evergreen low forests, and throughout the plain mostly inhabiting hedges and bushes.

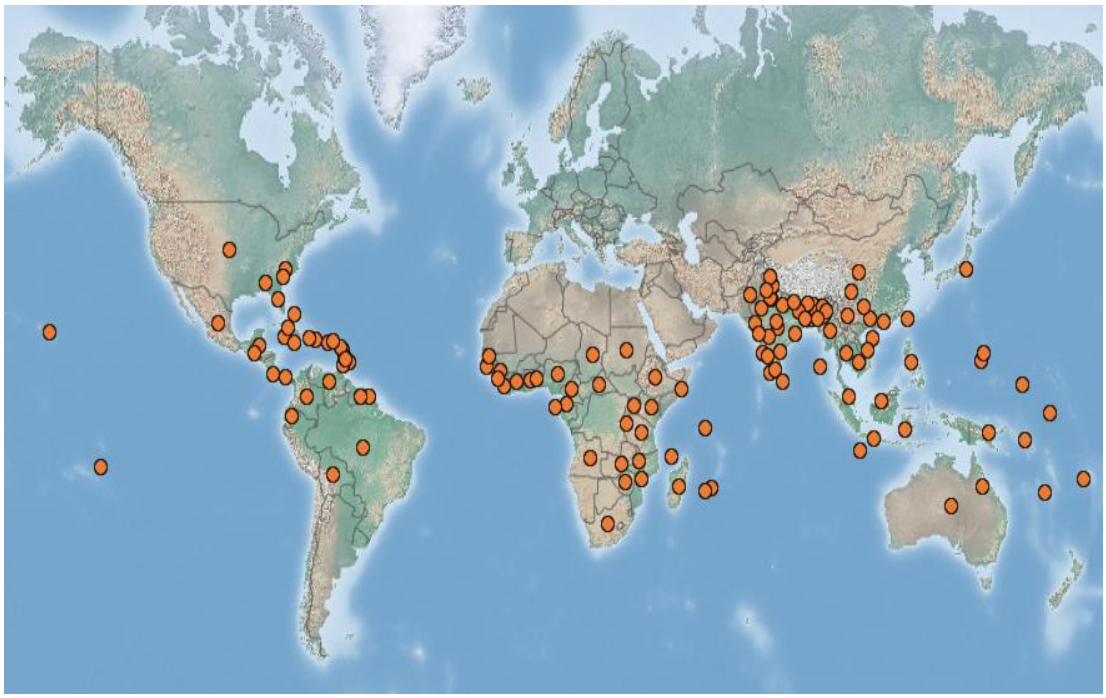


Figure 7. Distribution map of *M. pruriens*
(Source: https://www.cabi.org/isc/datasheet/35134).

2.3. Agronomic potential

M. pruriens is the only economically important species of the genus *Mucuna*. It is regarded as one of the most productive legumes (Fujii et al. 1991) due to its multiple agronomic potentials. The plant produces a seed yield of 1.3-2.4 t/ha (Kumwenda and Gilbert 1998; Gurumoorthi et al. 2003), the total biomass of 20-30 t/ha, and dry matter yield 7-9 t/ha. It is tolerant to drought, high soil acidity, and low soil fertility (Pugalenthi et al. 2005); but grows poorly in the cold environment and wet soil (Duke 1981). It is resistant to a wide range of pests and diseases due to the presence of N, N-Dimethyltryptamine (DMT), and L-DOPA which provides a chemical defense (Ortiz-Ceballos et al. 2012). It also exhibits allelopathic properties (Carsky et al. 1998; Fuji et al. 1991) and is effective in lowering the nematode population (Carsky et al. 1998; Queneherve et al. 1998). But much needs to be done in terms of the breeding efforts to augment its production potential for large-scale cultivation. Developing improved cultivars with low/high L-DOPA content, biotic and abiotic stress-resistance, and identifying markers and genes for important agronomic traits are some of the challenges and opportunities in the velvet bean genetic improvement program.

2.4. Nutritional potential

Seed and immature pods of velvet bean have been used as food by several indigenous communities in India, the Philippines, Nigeria, Ghana, Brazil, and Malawi (Eilitta et al. 2002; Janardhan et al. 2003). Cultivated variety (var. *utilis*) particularly possesses high crude protein content in seeds ranging from 26.26 - 31.24% (Kalidas and Mohan 2011; Kalidas and Mahapatra 2014) which is akin to prominent legumes such as pigeon pea (Kumar et al. 1991), chickpea (Hira and Chopra 1995), and soybean (Mang et al. 2016). The total dietary fiber is similar to that of jack bean

[*Canavalia ensiformis* (L.) DC.] (Doss et al. 2011) and the lipid content is higher than that of other pulses such *Vigna* spp. and *Rhychosia* spp. (Kalidass and Mohan 2012a, 2012b). It is also a good source of minerals and amino acids (Kalidas and Mohan 2011; Kalidas and Mahapatra 2014) as well as oils such as oleic, linoleic, palmitic, and stearic acids (Pathania et al. 2020). The comparative nutritional attributes of *M. pruriens* varieties vis-à-vis *Phaseolus vulgaris* L. are provided in Table 7.

Table 7. Comparative nutritional composition of *M. pruriens* and *Phaseolus vulgaris* seeds (Sources: Audu and Aremu 2011; Kalidass and Mohan 2011; Kalidass and Mahapatra 2014).

Component	***M. pruriens* var. *pruriens***	***M. pruriens* var. *utilis***	***Phaseolus vulgaris***
Proximate composition			
Crude Protein (%)	27.51-31.24	26.26-28.82	22.50-27.96
Crude Lipid %	6.57-7.84	6.45-7.19	0.92-4.4.3
Total Dietary Fiber (TDF) %	6.52-9.71	8.46-9.73	2.3-5.7
Calorific Value (KJ 100 g-1)	1576.18-1620.43	1567.93-1602.14	1549.5-1840.9
Total sulfur amino acids (%)	1.16-1.32		2.0-2.9
Mineral composition (mg 100 g-1)			
Sodium	34.53-98.51	98.39-128.12	40-210
Potassium	1395.41-1600.93	1628.36-1846.23	1320-1780
Calcium	584.15-740.72	689.45-746.15	70-210
Magnesium	440.78-594.60	298.41-341.44	160-230
Phosphorus	410.37-583.07	327.47-456.35	380-570
Manganese	5.55-8.36	3.78-4.49	1.0-2.0
Iron	4.45-7.56	12.41-14.74	3.34-8.00
Zinc	1.42-2.63	6.12-8.25	1.40-6.50
Essential amino acids (g 100 g-1)			
Threonine	4.08-5.28	3.40-3.52	3.83-4.01
Valine	3.33-6.30	3.64-4.60	4.82-5.41
Cystine	1.12-2.20	0.38-0.78	0.06-0.09
Methionine	0.54-1.10	0.84-0.96	1.04-1.61
Isoleucine	2.54-5.30	5.26-6.01	4.06-4.27
Leucine	5.17-6.50	5.36-7.01	7.10-7.57
Tyrosine	2.43-5.11	4.36-5.12	2.30-2.94
Phenylalanine	3.13-4.80	3.42-3.78	4.96-5.55
Lysine	4.85-5.40	5.36-5.94	5.87-6.72
Histidine	2.23-4.31	2.20-3.44	2.28-2.57
Tryptophan	0.92-1.21	0.96-1.10	1.20-1.28

While the high protein content of *M. pruriens* seeds offers potential alternatives for mitigating the protein deficiency gaps in the tropics (Muoni et al. 2019), its utilization as a mainstream food/feed is constrained due to high levels of anti-nutritional factors such as polyphenols, phytate, alkaloids, saponins, lectins, oligosaccharides, cyanogenic glycosides, and trypsin inhibitors in its seeds (Pugalenthi et al. 2005). Thus, it is necessary to integrate the traditional processing methods used by the indigenous communities with the modern processing technologies to exploit its full potential as a food and feed.

2.5. Medicinal properties

M. pruriens is well known medicinal plant in India. In the Indian Ayurvedic system of medicine, it is used for the treatment of various diseases (Sathiyanarayanan and Arulmozhi 2007) and supplemented as the key ingredient in more than 200 drug formulations (Chinapolaiah et al. 2019a). Due to the high content of L-DOPA, it is used in the treatment of Parkinson's disease and serves as its commercial source (Lampariello et al. 2012). It also reportedly increases the sperm count and its motility and augments the transformation of spermatocytes to sperm (Pulikkalpura et al. 2015). The paste of *M. pruriens* seeds is used to treat scorpion stings and the seed powder is used to avert the toxic effects of snake bites (Fung et al. 2010; Kumar et al. 2016; Pathania et al. 2020). It also exhibits analgesic, anti-inflammatory (Hishika et al. 1981), anti-neoplastic, anti-epileptic (Gupta et al. 1997), neuroprotective (Misra and Wagner 2004), antioxidant (Tripathi and Upadhyay 2001; Rajeshwar et al. 2005; Kumar et al. 2010), anti-diabetic (Majekodunmi et al. 2011), and antimicrobial activities (Rajeshwar et al. 2005; Verma et al. 2014). Thus, a detailed investigation of the safety and efficacy of these extracts/drugs could help design and develop potential therapeutic targets from this medicinal plant (Pathania et al. 2020).

2.6. Industrial potential

The seeds of the *M. pruriens* plant are the best natural source of anti-Parkinson's drug L-DOPA (Chinapolaiah et al. 2019b) although the synthetic alternative is also available (Katzenschlager et al. 2004). The high seed-yielding plants are in great demand from the herbal industries (Chinapolaiah et al. 2019b). It is also the source of other bioactive alkaloids like serotonin, nicotine (Duke 1981), N, N-dimethyltryptamine, choline, bufotenine, indole-3-alkylamine, 5-oxindole-3-alkylamines, and β-carboline, which are of significant pharmaceutical interest (Ghosal et al. 1971; Kumar et al. 2012). Thus, the plant could effectively substitute for other natural/synthetic sources of these drugs for industrial production (Shanmugavel and Krishnamoorthy 2018).

The demand for starch is rising, and the food industry needs an alternative source of this polysaccharide. The most commonly used starches are derived from grains of maize, rice, wheat, and *Sorghum* spp. and tuber plants such as Irish potato, sweet potato, sago palm (*Cycas revoluta* Thunb.), and Adam's needle (*Yucca* spp.) (Carcea et al. 1992; Wang et al. 1993; Wiesenborn et al. 1994). The *M. pruriens* can offer a potential alternative to these popular sources as it contains about 52% of carbohydrates in seeds (Betancur-Ancona et al. 2002). Reportedly, *Mucuna* starches are beneficial as an ingredient to the production of cookies, custards, sausages, porridges, puddings, jellies, jams, emulsifiers, and canned products (Betancur-Ancona et al. 2002; Lawal and Adebowale 2004). Likewise, oxidized starch of *M. pruriens* is useful to produce mayonnaise, salad cream, and lemon curd that requires low viscosity (Adebowale and Lawal 2003). Oil extracts are useful to prepare skin cream, liquid soap, polish, paint, wood varnish, and resin (Ajiwe et al. 1997).

2.7. Genetic resources

Germplasm is a source of information of all genes present in a specific plant that can be preserved for future use (Bhatia 2015). It conserves a wide range of genetic diversity, generating a pool of diverse genes to act as a reservoir of genetic resources (Mathur 2013). In the context of *M. pruriens*, very few research institutes /organizations across the world are maintaining the germplasm collection (Sathyanarayana et al. 2016). They include the International Institute of Tropical Agriculture (IITA), Nigeria; Centro Internacional de Agricultura Tropical (CIAT), Colombia; United States Department of Agriculture (USDA 1994), USA; National Biological Institute (NBI), Indonesia; AVRDC - The World Vegetable Centre, Taiwan (Jorge et al. 2007), etc.

In India, institutes like the National Bureau of Plant Genetic Resources (NBPGR), New Delhi (Jorge et al, 2007; Raina et al. 2012), Jawaharlal Nehru Tropical Botanical Garden and Research Institute (JNTBGRI), Thiruvananthapuram (Padmesh et al. 2006), Indian Institute of Horticultural Research (IIHR), Bengaluru (Mamatha et al. 2010), Arya Vaidya Sala, Kottakkal, Bharathiar University, Coimbatore (Siddhuraju and Becker 2005), Zandu Foundation for Health Care, Valsad, Gujarat (Krishnamurthy et al. 2005), etc. have been preserving the *M. pruriens* germplasm. However, the data on the exact number of accessions maintained by them is not available (Sathyanarayana et al. 2016).

2.8. Morphological diversity

The natural population of *M. pruriens* shows significant variation in flower, pod, and seed traits (Figure 8) (Sathyanarayana et al. 2016). It produces two flower colors- white and purple. Pod hairs also show two phenotypes, the long rough ones present in the wild genotypes, and the short smooth ones found in the cultivated ones (Aminah et al. 1974). In seeds, a wide range of phenotypic variability has been reported with colors ranging from extreme black to white with varying shades of brown (Lubis et al. 1980). Wide variability is also reported for other seed traits as well, such as seed length, width, and thickness in both the wild and the cultivated varieties (Figure 8) (Leelambika et al. 2010; Sathyanarayana et al. 2012).

Figure 8. Morphological diversity in *M. pruriens* **(a)** leaves **(b)** inflorescence **(c)** pods **(d)** seeds.

In addition, significant variability has been observed for the agronomic traits such as fertility index, days to flowering, harvest index, seed recovery percentage, etc. from Indian accessions (Gurumoorthi et al. 2003). Early maturing genotypes have been identified from the velvet bean accessions of Ghana (Bennet-Lartey 1998). Such diversity provides *Mucuna* breeders with an elaborate opportunity to breed novel and improved cultivars with desirable traits (Bhandari et al. 2017).

2.9. Conventional breeding

The first conventional breeding in *M. pruriens* targeted seed traits including L-DOPA content (Krishnamurthy et al. 2002). The researchers examined a total of 69 F_1 hybrids and their back-crosses (BC_1) and recovered accessions showing high seed yield and L-DOPA content together. The heterosis analysis of selected 37 F_1 lines over the mid parent, better parent and best parent on the seed yield traits generated a high magnitude of heterosis (40-96%) over the best parent in four F_1 lines. A more recent study on the heterosis and combining ability analysis for yield and yield contributing traits have recommended different sets of parents for specific combiner and heterosis for economically important traits such as pod length, pod weight, dry pod yield per plant, seed yield per plant, etc., (Chinapolaiah et al. 2017).

Recently, the genetic analysis of the L-DOPA trait by hybridizing the contrasting parents with low (1.55%) and high (9.03%) L-DOPA content found a significant reduction in L-DOPA (3.82%) among the F_1 hybrids suggesting low L-DOPA may be dominant over high L-DOPA content (Tripathi 2018). High heritability (71.3%) and moderate genetic advance (GAM = 196.23) in this study suggested that the L-DOPA trait is highly influenced by the additive genes, confirming the selection based on this metabolic trait would be effective and reliable.

2.10. Mutation breeding

Mutation breeding is yet another prominent approach to produce superior plant varieties within a short period (Singh et al. 2016). In this method, chemical treatments or irradiations are used to produce heritable mutations (Stadler 1928), which sometimes results in improved traits. In *M. pruriens*, mutation breeding work has been initiated recently using γ-radiation (Singh et al. 2016). The early results show that the seeds treated with lower doses (50-200 Gy) produce higher germination and growth rates vis-à-vis non-treated one. Treatment with 500 Gy produced the highest (5.7%) and 200 Gy produced the lowest (3.1%) L-DOPA compared to the parents (3.18%). Mutants were stable up to the M4 generation. This work has demonstrated the potential of mutation breeding for developing improved varieties.

2.11. Genetic diversity and molecular markers

The genetic diversity in the germplasm collection and the natural population is a prerequisite for any crop improvement program. In *M. pruriens*, earlier studies using the dominant markers such as RAPD (Padmesh et al. 2006; Leelambika et al. 2010; Patil et al. 2016), ISSR (Chinapolaiah et al. 2018; Patil et al. 2016), and AFLP (Capo-chichi et al. 2001; Sathyanarayana et al. 2011; Tripathi et al. 2018) have revealed moderate genetic diversity. But, the need for co-dominant markers has been felt increasingly for more reliable estimates due to their advantage over the dominant markers (Mondini et al. 2009).

The first such attempt to develop cross-species SSR markers was initiated by screening 2,86,488 EST-sequences from four closely related legume species *viz*., soybean, common bean, cowpea, and chickpea (Shetty et al. 2015) identifying 22,457 SSR markers. For validation, 522 primer combinations were designed and 50

were randomly picked and validated using a diverse panel of 25 *M. pruriens* genotypes which revealed the high polymorphism information content value (PIC = 0.65). However, the first species-specific SSR marker development based on the transcriptome sequence was reported by Sathyanarayana et al (2017), who identified 7,493 SSR motifs and validated 132 SSR of them using 23 *M. pruriens* accessions (Sathyanarayana et al. 2017). A chosen set of 50 markers from this study were further deployed to assess the genetic diversity among 18 accessions of the cultivated variety (var. *utilis*) (Kumar et al. 2019). All these studies not only reported the usefulness of these markers but also revealed the presence of high genetic diversity in the *M. pruriens* germplasm of India (Shannon's information index, I > 0.6). However, so far, high-throughput sequence-based SNP markers are not available in this legume.

2.12. Parent line assessments and trait-specific breeding

In *M. pruriens*, rigorous breeding efforts are needed to develop varieties with enhanced nutritional content, safe levels of L-DOPA, and resistance to biotic and abiotic stresses (Mahesh 2015). A glance at this effort indicates that Mahesh and Sathyanarayana (2011) screened 15 *M. pruriens* accessions in an attempt to develop resistant varieties for Fusarium wilt and found four of them (IC471870, 500101KA, 500108KA, 500155AP) to be useful as parental lines (Mahesh and Sathyanarayana 2011). Similarly, the same authors tested 35 accessions for salinity tolerance and identified several putative parental combinations for breeding/mapping (Mahesh and Sathyanarayana 2015). L-DOPA is an important constituent of *M. pruriens*. Combining both metabolic and genetic diversity data from 40 *M. pruriens* accessions, seven putative parental combinations have been suggested for genetic improvement and mapping of this trait (Leelambika et al. 2016).

2.13. Genomic and transcriptome resources

The transcriptome information has served as a key resource to identify and characterize the biosynthesis pathway of important secondary metabolites in several plant species (Rama Reddy et al. 2015; Hagel et al. 2015; Xu et al. 2015). It has also facilitated gene expression studies in resource-scarce species (Grabherr et al. 2011; Garg and Jain 2013) besides offering easy and economic methods to develop species-specific markers. In the case of *M. pruriens*, the first transcriptome analysis reported *de novo* assembly, functional annotation, and the differential gene expression among pod, leaf, and root tissues (Sathyanarayana et al. 2017). The *de novo* assembly was generated using 67,561 transcripts with an N50 length of 987 bp and a mean transcript length of 641 bp (Table 8). About 80% of the transcripts were successfully annotated using the BLASTx tool. In the case of maker development, a total of 67,561 sequences were examined to identify 7,943 SSR motifs validating 134 of them (Lepcha et al. 2019). In another study that followed, detailed transcriptome analysis has been carried out along with biochemical characterization to track the presumptive biosynthetic pathway of L-DOPA and the genes associated with it (Singh et al. 2018). However, whole-genome sequencing efforts are still awaited in this species.

Table 8. Data on *M. pruriens* transcripts from the Trinity assembly and SSR identification (Source: Sathyanarayana et al. 2017).

Transcripts obtained from the Trinity assembly		**SSRs mining**	
Total assembled bases	46,525,999	Number of sequences examined	67,561
Number of transcripts	72,561	Total examined sequences (bp)	42,340,968
The total number of transcripts after clustering	67,561	Total number of identified SSRs	7,943
The mean sequence length	626	Number of SSR containing sequences	6,284 (9.3%)
Average % of N	0.00	Number of sequences containing more than one SSR	1,174
Average % of GC content	44.58	Number of SSRs present in compound formation	963
N50	987	Frequency of SSRs	One per 5.3 kb
Maximum transcript length	17,978	**Distribution of SSRs in different repeat types**	
Average transcript length	641	Mono-nucleotide	3,638 (45.80%)
Number of putative non coding sequences	1,493	Di-nucleotide	1,674 (21.07%)
Length of the longest ORF (bp)	2,362	Tri-nucleotide	2,240 (28.20%)
Number of ORFs ≥ 100 bp	36,228	Tetra-nucleotide	146 (1.83%)
Number of ORFs on plus (+) strand	36,421	Penta-nucleotide	64 (0.80%)
Number of ORFs on minus (−) strand	31,140	Hexa-nucleotide	100 (1.25%)

2.14. Linkage mapping

In the realm of molecular breeding, AFLP based genetic linkage maps have been developed in *M. pruriens* reporting the QTL positions for the floral, pod, and seed traits (Capo-chichi et al. 2004; Mahesh et al. 2016). The first linkage map was based on the segregation analysis of traits like pod pubescence and pod color in 82 F_2 individuals produced from divergent parents (Capo-chichi et al. 2004). A total of 166 AFLP markers were used which generated 20 linkage groups covering over 687.9 cM with an average distance of 34.4 cM between the markers. The morphological traits such as pod pubescence and pod color were found to co-segregate at a distance of 4.2 cM. The second genetic linkage map, developed from the Indian *M. pruriens* accessions, identified QTL positions for the floral, pod, and seed traits using 200 F_2 progenies from an inter-varietal cross involving wild and cultivated genotypes (Figure 9) (Mahesh et al. 2016). This map comprised 129 AFLP markers distributed over 13 linkage groups with an average marker interval of 4.79 cM and spanning a total distance of 618.88 cM.

However, the use of dominant DNA marker (AFLP) and low genome coverage has been the critical limitations in the above two maps. Also, the influence of the environment on F_2 progeny is not accounted for in these studies. Although SSR-based genetic linkage map has been attempted (Kumar 2019), due to its low resolution, it is considered as a preliminary result. Therefore, testing mapping populations in varying environmental conditions, and the use of co-dominant molecular markers such as SSR and SNP to develop dense and saturated linkage maps need to be undertaken in this plant.

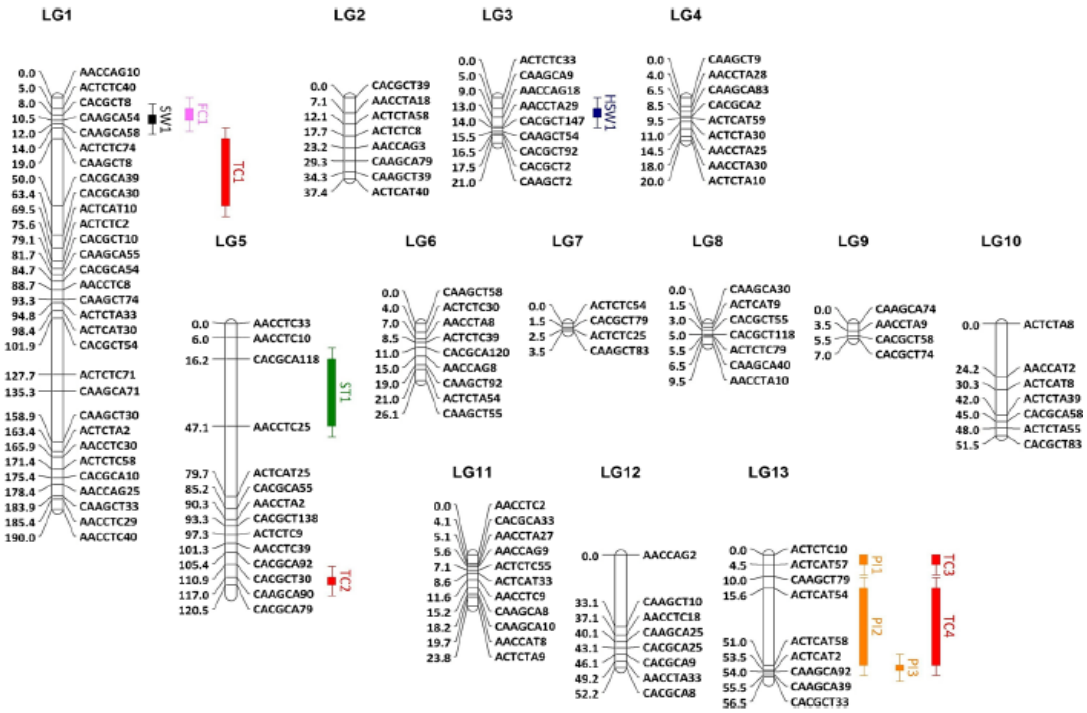


Figure 9. Linkage map of *M. pruriens* indicating the QTL positions (Source: Mahesh et al. 2016).

2.15. Association mapping in *M. pruriens*

Association mapping (AM) identifies the relationship between phenotypic variations with genetic polymorphism among the unrelated individuals by utilizing naturally occurring or historical recombination and thus serves as a quick and efficient alternative to conventional linkage mapping (Ambreen et al. 2018). The success of association mapping critically depends on the choice of the germplasm which ought to encompass a wide range of diversity capturing the maximum number of historical recombination events (Flint-Garcia et al. 2005; Yan et al. 2011). Conventionally, core collection, which is available only for the selected/mainstream crop species has done well as association mapping panels (Liu et al. 2017; Zhao et al. 2017; Ambreem et al. 2018). Alternatively, for crops that lack these resources, a diverse panel of accessions from the germplasm collection can be used, provided they demonstrate the presence

of high genetic variance and possess weak population structure and low kinship association among the individuals (Mahajan et al. 2017; Hu et al. 2014). This approach has been successfully implemented for identifying marker-trait associations for several agronomic traits in various legume species including seed-related traits in groundnut (Zhao et al. 2017), iron and zinc concentration in lentil (Singh et al. 2017), frost tolerance in common pea (Lui et al. 2017), flowering, pod yield per plant and fresh pod per pant in lablab or Dolichos bean (Vaijayanthi et al. 2018), crude protein and mineral concentration in *Medicago sativa* (Jia et al. 2017), etc. However, to date, not even a single study has explored this opportunity in *M. pruriens*.

Conventionally, AM analysis is carried out on a large number of populations with more than 300 accessions. But in the cases where the numbers are limited, it is still possible to explore marker-trait associations (Soumya et al. 2021). Recently, several genome-wide association analysis (GWAS) studies have been successfully carried out in population sizes between 60 and 150 (Ma et al. 2018; Lehnert et al. 2018; Rohila et al. 2020). Taking cues from these studies, in the present study (a) we evaluated a panel of sixty-one diverse *M. pruriens* accessions for suitability as an AM panel; (b) and performed association analysis on sixteen agronomically important traits using microsatellite markers.

CHAPTER 3

OBJECTIVES

3. OBJECTIVES

The thesis covered the following four objectives:

1. To analyze the variability for different seed traits in the natural population of *M. pruriens*.
2. To estimate genetic diversity in wild and cultivated genotypes using molecular markers.
3. To determine population structure and linkage disequilibrium in *M. pruriens* population.
4. To carry out association analysis for selected seed traits.

CHAPTER 4

MATERIALS & METHODS

4. MATERIALS AND METHODS

Two different sets of *M. pruriens* accessions have been used in this study. For evaluation of seed traits and estimation of genetic diversity (objectives 1 and 2), 60 accessions of *M. pruriens* collected from across Northeast India were used. For the linkage disequilibrium and association analysis (objectives 3 and 4), a different set of 61 accessions from the germplasm collection at Sir Mokshagundam Visvesvaraya Institute of Technology (Sir MVIT), Bangalore was used. This is because the association mapping panel requires precise phenotyping of the traits for the reliable marker-trait association. *M. pruriens* germplasm of Sir MVIT is well characterized and offered a much reliable descriptor for different traits. Therefore, this set of germplasm was borrowed from Prof. N Sathyanarayana's previous students and used for the association analysis.

4.1. Variability for seed traits in *M. pruriens* natural population

4.1.1. Plant material

Seeds of *M. pruriens* were collected from all the eight states of Northeast India (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura) during the year 2016-17 (Figure 10). The collection of 60 accessions comprised two botanical varieties of which 16 were of var. *utilis* (cultivated) and the remaining 44 were of the var. *pruriens* (wild). The accessions of cultivated variety were collected from the households and kitchen gardens and, those of the wild variety was collected directly from the wild. The details of the accessions along with their collection locations are provided in Table 9. For evaluation, the seeds were raised in the Sikkim University botanical garden using standard practices.

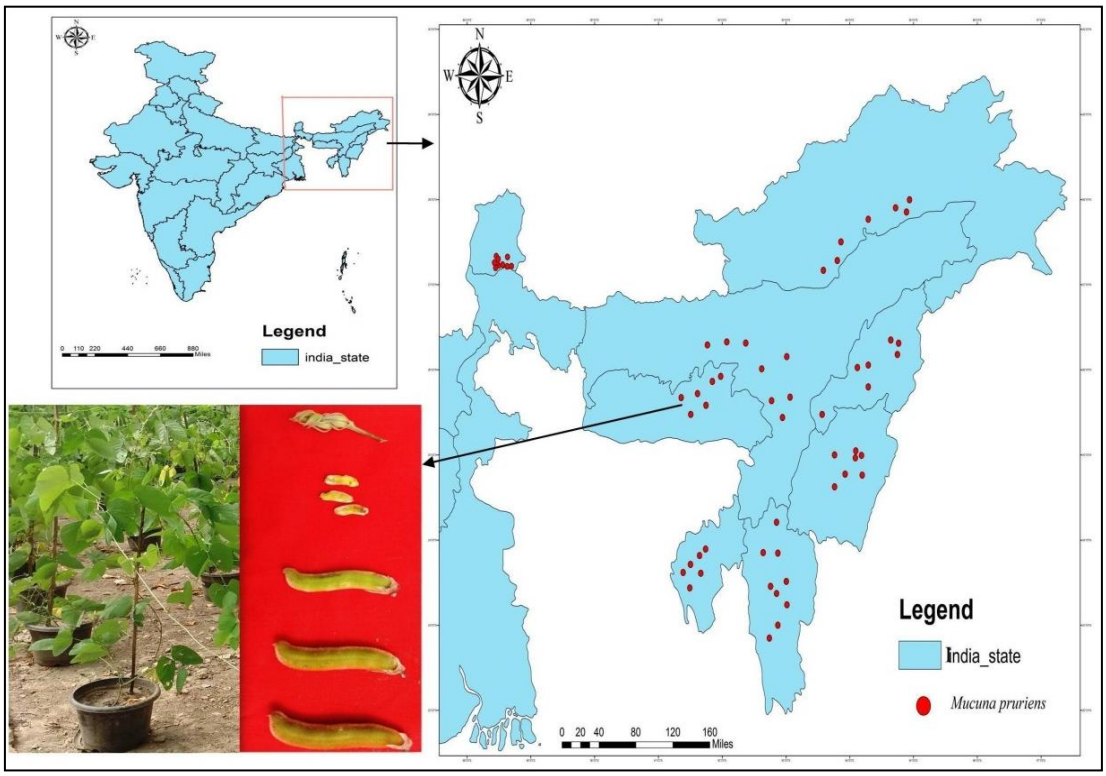


Figure 10. *M. pruriens* germplasm collection sites in Northeast India.

Table 9. *M. pruriens* accessions collected from Northeast India.

Sl. No.	Accessions	Botanical variety	Area of collection	Geographic region
1	500270-AR	var. *utilis*	Upper Siang (AR)	EH
2	500317-AR	var. *utilis*	Upper Siang (AR)	EH
3	500314-AR	var. *utilis*	Upper Siang (AR)	EH
4	500315-AR	var. *pruriens*	Upper Siang (AR)	EH
5	500316-AR	var. *pruriens*	Papum Parey (AR)	EH
6	500318-AR	var. *pruriens*	Papum Parey (AR)	EH
7	500302-AR	var. *pruriens*	Papum Parey (AR)	EH
8	500286-SK	var. *pruriens*	West Sikkim (SK)	EH
9	500287-SK	var. *pruriens*	West Sikkim (SK)	EH
10	500288-SK	var. *pruriens*	South Sikkim (SK)	EH
11	500289-SK	var. *pruriens*	South Sikkim (SK)	EH
12	500293-SK	var. *pruriens*	South Sikkim (SK)	EH
13	500299-SK	var. *pruriens*	South Sikkim (SK)	EH
14	500304-SK	var. *pruriens*	East Sikkim (SK)	EH
15	500325-SK	var. *pruriens*	East Sikkim (SK)	EH
16	500326-SK	var. *pruriens*	East Sikkim(SK)	EH
17	500327-SK	var. *pruriens*	East Sikkim(SK)	EH
18	500275-SK	var. *pruriens*	East Sikkim(SK)	EH
19	500285-AS	var. *pruriens*	Kamrup Metropolitan (AS)	EH

20	500290-AS	var. *pruriens*	Kamrup Metropolitan (AS)	EH
21	500292-AS	var. *pruriens*	Kamrup Metropolitan (AS)	EH
22	500305-AS	var. *pruriens*	Kamrup Metropolitan (AS)	EH
23	500322-AS	var. *pruriens*	Jorhat (AS)	EH
24	500323-AS	var. *pruriens*	Jorhat (AS)	EH
25	500324-AS	var. *pruriens*	Jorhat (AS)	EH
26	500274-AS	var. *pruriens*	Jorhat (AS)	EH
27	500268-NL	var. *utilis*	Kohima (NL)	IB
28	500271-NL	var. *utilis*	Kohima (NL)	IB
29	500273-NL	var. *utilis*	Dimapur (NL)	IB
30	500276-NL	var. *utilis*	Dimapur (NL)	IB
31	500278-NL	var. *utilis*	Dimapur (NL)	IB
32	500279-NL	var. *utilis*	Mokokchung (NL)	IB
33	500280-NL	var. *utilis*	Mokokchung (NL)	IB
34	500281-MZ	var. *pruriens*	Mamit (MZ)	IB
35	500282-MZ	var. *pruriens*	Mamit (MZ)	IB
36	500283-MZ	var. *pruriens*	Mamit (MZ)	IB
37	500269-MZ	var. *utilis*	Aizawl (MZ)	IB
38	500295-MZ	var. *pruriens*	Aizawl (MZ)	IB
39	500296-MZ	var. *pruriens*	Aizawl (MZ)	IB
40	500297-MZ	var. *pruriens*	Kolasib (MZ)	IB
41	500303-MZ	var. *pruriens*	Kolasib (MZ)	IB
42	500272-MN	var. *utilis*	Imphal East (MN)	IB
43	500320-MN	var. *utilis*	Imphal East (MN)	IB
44	500321-MN	var. *utilis*	Imphal East (MN)	IB
45	500284-MN	var. *utilis*	Imphal East (MN)	IB
46	500294-MN	var. *pruriens*	Bishnupur (MN)	IB
47	500300-MN	var. *pruriens*	Bishnupur (MN)	IB
48	500319-MN	var. *pruriens*	Bishnupur (MN)	IB
49	500309-TR	var. *utilis*	West Tripura (TR)	IB
50	500301-TR	var. *pruriens*	West Tripura (TR)	IB
51	500306-TR	var. *pruriens*	West Tripura (TR)	IB
52	500307-TR	var. *pruriens*	West Tripura (TR)	IB
53	500308-TR	var. *pruriens*	West Tripura (TR)	IB
54	500310-TR	var. *pruriens*	West Tripura (TR)	IB
55	500311-ML	var. *pruriens*	East Khasi Hills (ML)	IB
56	500312-ML	var. *pruriens*	East Khasi Hills (ML)	IB
57	500313-ML	var. *pruriens*	East Khasi Hills (ML)	IB
58	500291-ML	var. *pruriens*	East Khasi Hills (ML)	IB
59	500277-ML	var. *pruriens*	East Khasi Hills (ML)	IB
60	500298-ML	var. *pruriens*	East Khasi Hills (ML)	IB

EH, Eastern Himalaya; IB, Indo-Burma; AR, Arunachal Pradesh; SK, Sikkim; AS, Assam; NL, Nagaland; MZ, Mizoram; MN, Manipur; TR, Tripura; ML, Meghalaya.

4.1.2. Evaluation of seed traits

Data on seed length, seed width, and seed thickness were obtained by manual measurements with the Vernier caliper. Hundred (100) seed weight was obtained using weighing balance, while the total protein and carbohydrate contents were estimated using Bradford (Bradford 1976) and Anthrone (Hedge and Hofreiter 1962) methods. All statistical analyses such as correlation, principal component analysis (PCA), one-way analysis of variance (ANOVA) were performed using the R program (R Core Team 2014) and the SPSS v. 17 (SPSS Inc. 2008).

4.2. Genetic diversity among the Northeast Indian accessions

4.2.1. DNA Isolation and SSR data generation

Leaf tissue from 15-days old seedlings was used for the DNA isolations as described previously (Doyle and Doyle 1990). Genotyping was carried out using a sub-set of 25 species-specific genic SSR primer pairs (Table 10) chosen from Sathyanarayana et al. (2017). Polymerase chain reaction (PCR) mixture included: template DNA (50ng/µl), primers (1µM each), dNTPs (2.5mM), Taq polymerase (1U), PCR buffer (1X), and $MgCl_2$ (1.5mM), maintaining the final volume of 25µl. DNA amplification condition comprised of initial denaturation at 94°C for 3min followed by 35 cycles of 94°C for 30s, appropriate annealing temperature (*Tm*) for 30s, and an extension at 72°C for 20s with a final extension of 7min at 72°C. PCR was performed on a Bio-Rad thermal cycler (C1000 Touch™). The amplified products were visualized on a UV illumination gel documentation system (ChemiDoc XRS+, Bio-Rad).

Table 10. Details of the SSR markers used for genetic diversity analysis.

Markers	Forward	Reverse
MPU_01	TTCTCACCTGTGTTGCTTATC	GCAGGCATATGATAGTGTGAT
MPU_02	CTGAAACACCACCTTTCATAG	GGATTAGTGGCTCTATTCAGG
MPU_04	CGATTATACAGCACTCCATGT	ACTTAATTCTGGTGGTTTTCC
MPU_05	AAGCTTGTCACTGTCAAAAAG	ATGCAACACATGTCAACACT
MPU_06	GCATTTCAAATAAACGTCAAC	TTGACACAATTTGACAGAGGT
MPU_07	AATGGATCCCTTTTCTCTATG	TATTGGAATAGATCCCCTTGT
MPU_08	CTGCAGGAGGAACTGTTG	TGCTGTGTATAATTTGCAATG
MPU_09	ACACCCAGTATCTTCCTCTTC	TTGTGCTTTTGATTCTTTAGG
MPU_10	CCTTCCCTTTAGATGTGAAAT	ACATTGATAGCAGTGGAGAAA
MPU_11	CCTTCCCTTTAGATGTGAAAT	ACATTGATAGCAGTGGAGAAA
MPU_12	AGGAGAGAGAGTGAAATTGGA	ACAACGTGAACAGAGAGAGAA
MPU_13	AGGAGAGAGAGTGAAATTGGA	ACAACGTGAACAGAGAGAGAA
MPU_14	GACTCCAACTCCTCCTTCA	ACTGTTGTTGTTGCTGATGTT
MPU_15	GAAGAGGAGGTTCAGAACAGT	GCATCAGATACAACAAAGGAG
MPU_16	ATTCTACAATCACGCATCATC	TTTGATGCATAGAAAAGGAGA
MPU_17	TAAAACTCCTTTTCCTTCTCC	AGTTCCTTCAAAATACGCTTC
MPU_18	GCATTCTCTGGACATACAAAC	GGGAAAATCTGAGAAGAAAAA
MPU_19	CAAAGCTAGCATTAGAAGCAG	AAAAATACGTAAAAGGGATGG
MPU_20	TCGTTTCGTTTTCGTTTACTA	GATTTGAACTGGGGAAAAA
MPU_22	CACGTTTGTAGGGTAAGATCA	TTTTTGGAACAAAAGAGTGTG
MPU_23	CTAGCATTCTTCTTTGGATCA	CAAAGCCACTTAAGAGAGAGA
MPU_25	ACAGTTTTGATCCATTTTCCT	ACAAAATTGATGCAGCTTTAG
MPU_61	AGAAATTTGTCCCAGGTAGAG	ACAAACACAAACACCACAAAC
MPU_124	CCTTCCCTTTAGATGTGAAAT	ACATTGATAGCAGTGGAGAAA
MPU_125	AGGAGAGAGAGTGAAATTGGA	ACAACGTGAACAGAGAGAGAA

4.2.2. Marker attributes and genetic diversity

Marker attributes such as polymorphism information content (PIC) and major allele frequency (MAF) were calculated by PowerMarker v 3.25 (Liu and Muse 2005). The total number of alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), percentage of polymorphic loci (%P), gene flow (Nm), and Hardy-Weinberg equilibrium (HWE) were assessed by GenAlEx v 6.5

(Peakall and Smouse 2012). HP-Rare v 1.0 (Kalinowski 2005) was used to determine private/rare allelic richness per locus (R_p) based on the rarefaction approach. The genetic diversity indices such as - Shannon's information index (I), Nei's gene diversity (h), total genetic diversity (H_T), genetic diversity within the population (H_S), and genetic differentiation (G_{ST}) were determined by using POPGENE v 1.32 (Yeh et al. 1999).

4.2.3. Genetic relationship

Population structure was analyzed using the Bayesian clustering method implemented in STRUCTURE v 2.3.4 (Pritchard et al. 2000). For each accession, the proportion of ancestral contribution was estimated using the admixture model and correlated allele frequencies. K-values ranging from 1-10 were tested with 10 independent replications, 100,000 lengths of the burn-in period, and 200,000 Markov Chain Monte Carlo (MCMC) repetitions for each K. The optimal K-value was obtained by using STRUCTURE analysis results with STRUCTURE HARVESTER (Earl 2012). Accessions were assigned to a subgroup depending on the Q-value (membership proportion), and if the Q-value is < 80%, they were termed admixtures. Analysis of molecular variance (AMOVA) and pairwise F_{ST} of geographical population and subpopulations were carried out using GenAlEx v 6.5 (Peakall and Smouse 2012) with 1,000 permutations. Genetic relationship based on the distance was arrived through the construction of unrooted neighbor-joining (NJ) dendrogram and principal coordinate analysis (PCoA) using Darwin v 6.0 (Perrier and Jacquemoud-Collet 2006). The reliability of the NJ-dendrogram was tested with the bootstrap value of 1000 replicates.

4.3. Feasibility analysis of association mapping panel

4.3.1. Plant material

A subset of 61 *M. pruriens* accessions from the germplasm collection of Sir MVIT, Bangalore was used as the association mapping (AM) panel. The panel comprised eight accessions of var. *utilis*, 28 of var. *pruriens*, and 25 of var. *hirsuta* (Figures 11-13). They represented three large geographic regions in India - Eastern, Southern, and West-Central India (Table 11).

Figure 11. *M. pruriens* germplasm procured from Sir MVIT for association analysis (8 accessions of var. *utilis*; cultivated variety).

Figure 12. *M. pruriens* germplasm procured from Sir MVIT for association analysis (25 accessions of var. *hirsuta*; wild variety).

Figure 13. *M. pruriens* germplasm procured from Sir MVIT for association analysis (28 accessions of var. *hirsuta*; wild variety).

Table 11. Details of *M. pruriens* accessions procured from Sir MVIT for association analysis.

Sl. No.	Botanical Variety	Accession No.	Location of collection	Geographical region
1	*M. pruriens var. pruriens*	IC-265577	NBPGR, Delhi	Eastern India
2	*M. pruriens var. utilis*	IC-369144	Jharkhand	Eastern India
3	*M. pruriens var. pruriens*	IC-391885	Orissa	Eastern India
4	*M. pruriens var. utilis*	IC-392241	Jharkhand	Eastern India
5	*M. pruriens var. utilis*	IC-392850	Orissa	Eastern India
6	*M. pruriens var. utilis*	IC-471870	NBPGR, Delhi	Eastern India
7	*M. pruriens var. utilis*	IC-471876	NBPGR, Delhi	Eastern India
8	*M. pruriens var. pruriens*	500191-OR	Orissa	Eastern India
9	*M. pruriens var. pruriens*	500192-OR	Orissa	Eastern India
10	*M. pruriens var. pruriens*	500193-OR	Orissa	Eastern India
11	*M. pruriens var. pruriens*	500194-OR	Orissa	Eastern India
12	*M. pruriens var. pruriens*	500195-OR	Orissa	Eastern India
13	*M. pruriens var. pruriens*	500196-OR	Orissa	Eastern India
14	*M. pruriens var. pruriens*	500197-WB	West Bengal	Eastern India
15	*M. pruriens var. pruriens*	500199-WB	West Bengal	Eastern India
16	*M. pruriens var. pruriens*	500113-MH	Maharashtra	West-Central India
17	*M. pruriens var. pruriens*	500172-MH	Maharashtra	West-Central India
18	*M. pruriens var. pruriens*	500173-MH	Maharashtra	West-Central India
19	*M. pruriens var. pruriens*	500174-MH	Maharashtra	West-Central India
20	*M. pruriens var. pruriens*	500175-MH	Maharashtra	West-Central India
21	*M. pruriens var. pruriens*	500176-MH	Maharashtra	West-Central India
22	*M. pruriens var. pruriens*	500177-MH	Maharashtra	West-Central India
23	*M. pruriens var. pruriens*	500178-MH	Maharashtra	West-Central India
24	*M. pruriens var. hirsuta*	500183-MH	Maharashtra	West-Central India
25	*M. pruriens var. pruriens*	500184-MH	Maharashtra	West-Central India
26	*M. pruriens var. pruriens*	500190-MH	Maharashtra	West-Central India
27	*M. pruriens var. utilis*	500108-KA	Karnataka	Southern India
28	*M. pruriens var. hirsuta*	500109-KA	Karnataka	Southern India
29	*M. pruriens var. pruriens*	500110-KA	Karnataka	Southern India
30	*M. pruriens var. hirsuta*	500111-KA	Karnataka	Southern India
31	*M. pruriens var. pruriens*	500112-KA	Karnataka	Southern India
32	*M. pruriens var. hirsuta*	500126-KA	Karnataka	Southern India
33	*M. pruriens var. pruriens*	500130-KA	Karnataka	Southern India
34	*M. pruriens var. hirsuta*	500131-KA	Karnataka	Southern India
35	*M. pruriens var. hirsuta*	500132-KA	Karnataka	Southern India
36	*M. pruriens var. hirsuta*	500133-KA	Karnataka	Southern India
37	*M. pruriens var. hirsuta*	500134-KA	Karnataka	Southern India
38	*M. pruriens var. hirsuta*	500135-KA	Karnataka	Southern India
39	*M. pruriens var. utilis*	500188-KA	Karnataka	Southern India
40	*M. pruriens var. hirsuta*	500189-KA	Karnataka	Southern India
41	*M. pruriens var. hirsuta*	500142-KA	Karnataka	Southern India

42	*M. pruriens var. hirsuta*	500144-AP	Andhra Pradesh	Southern India
43	*M. pruriens var. hirsuta*	500145-AP	Andhra Pradesh	Southern India
44	*M. pruriens var. hirsuta*	500146-AP	Andhra Pradesh	Southern India
45	*M. pruriens var. hirsuta*	500148-AP	Andhra Pradesh	Southern India
46	*M. pruriens var. hirsuta*	500149-AP	Andhra Pradesh	Southern India
47	*M. pruriens var. hirsuta*	500151-AP	Andhra Pradesh	Southern India
48	*M. pruriens var. hirsuta*	500152-AP	Andhra Pradesh	Southern India
49	*M. pruriens var. hirsuta*	500153-AP	Andhra Pradesh	Southern India
50	*M. pruriens var. hirsuta*	500165-AP	Andhra Pradesh	Southern India
51	*M. pruriens var. hirsuta*	500166-AP	Andhra Pradesh	Southern India
52	*M. pruriens var. utilis*	500159-PY	Pondicherry	Southern India
53	*M. pruriens var. pruriens*	500162-PY	Pondicherry	Southern India
54	*M. pruriens var. pruriens*	500163-PY	Pondicherry	Southern India
55	*M. pruriens var. pruriens*	500164-PY	Pondicherry	Southern India
56	*M. pruriens var. hirsuta*	500115-TN	Tamil Nadu	Southern India
57	*M. pruriens var. hirsuta*	500120-TN	Tamil Nadu	Southern India
58	*M. pruriens var. hirsuta*	500121-TN	Tamil Nadu	Southern India
59	*M. pruriens var. hirsuta*	500122-TN	Tamil Nadu	Southern India
60	*M. pruriens var. pruriens*	500137-TN	Tamil Nadu	Southern India
61	*M. pruriens var. pruriens*	500138-TN	Tamil Nadu	Southern India

4.3.2. Genotyping

Accessions of the AM panel were genotyped using a sub-set of 66 species-specific genic SSR primer pairs (Table 12; Sathyanarayana et al. 2017) as per the protocol described in the previous section (4.2.1).

Table 12. Details of SSR primer pairs used for association analysis.

Marker ID	Forward Primer	Reverse Primer
MPU_06	GCATTTCAAATAAACGTCAAC	TTGACACAATTTGACAGAGGT
MPU_14	GACTCCAACTCCTCCTTCA	ACTGTTGTTGTTGCTGATGTT
MPU_15	GAAGAGGAGGTTCAGAACAGT	GCATCAGATACAACAAAGGAG
MPU_16	ATTCTACAATCACGCATCATC	TTTGATGCATAGAAAAGGAGA
MPU_18	GCATTCTCTGGACATACAAAC	GGGAAAATCTGAGAAGAAAAA
MPU_19	CAAAGCTAGCATTAGAAGCAG	AAAAATACGTAAAAGGGATGG
MPU_20	TCGTTTCGTTTTCGTTTACTA	GATTTGAACTGGGGAAAAA
MPU_22	CACGTTTGTAGGGTAAGATCA	TTTTTGGAACAAAAGAGTGTG
MPU_23	CTAGCATTCTTCTTTGGATCA	CAAAGCCACTTAAGAGAGAGA
MPU_24	GGTGTTATTTGTGTGAATTGG	ACCAATATTTTGCTCTTTTCC

MPU_27	ACAGTTTTGATCCATTTTCCT	ACAAAATTGATGCAGCTTTAG
MPU_30	AGTTAAAACTCACCCACCCTA	TGGAGGTGAGTGAATAGATGA
MPU_31	AGTACATGGGGATAGAGCAAT	GGTGTCTTTTCTCTCCTCACT
MPU_37	TGGTCTCTCTAAACGACAGAG	AATAAGCATGCAGAGAGAGTG
MPU_39	AGTGGCAACTATATGACGGTA	GAGTCACAACCAAATAGCTCA
MPU_40	GCTACATATGGAATTGATTGG	AGATAAGGGACGAAAAATCAG
MPU_42	CAAAGATGACGAAGATGGATA	TCTTTCAACTCAATCTTCCAA
MPU_43	GTCTAGGGTGGCTTTGTAGAC	CAACACCCACCAAATCTAGTA
MPU_45	ATCTCTGTCTCTGTCCCTTTC	GCGATCTCTCTCTTTCTCTCT
MPU_46	TAGGAGACCCAATTAGAGGAG	CCGTAACAACACATCCTTAAA
MPU_47	GTTCACCACATCCATTATGAC	ACCACACACACACATACACAC
MPU_48	GGTGGAGACCTTACACAATATG	CACCTCCAGTTTATTCACCTC
MPU_49	GGAGGATAAAAGTGCTTCATT	ACCGTCCTCTTAATTGTTGAT
MPU_51	ATCGTTTGTCAACCATTAGAG	ATAATTATCACCCCACAAGGT
MPU_52	AAGGAAAGCATATGAAAGGAG	ATTCCTCCAATTCTATCTCGT
MPU_54	TCTCTCAACCAGTATCTCACC	CCACTGAACAACTCATGGTAT
MPU_57	TATCGATCTCAATTCCCATAA	TCTCTCTCTAAAGCCCAAAAC
MPU_58	CACTCTCAGTGAGTCTTCTGG	GAATGCTTGAGGCATAGAAG
MPU_59	GAGGTGTTGAAGAAGTGTGAG	ATTCCCACCATTCTCATTAGT
MPU_60	CTCTGTCCAACTCCTGAAGAT	CTACATCGCAACCGACAC
MPU_62	ATGAATGACTTGTTCCATGAT	GCAAGATGAAGTCATAAATCG
MPU_63	GGATCTTCTAAACCCGTTATC	AGCAAAAACAAAGTGGTAGTG
MPU_64	GCTTTCTATGAATGACTGCAT	AATTTCCAACCATGGATTAAG
MPU_65	CCAGTTTAGCTTGCACATACT	CCAATAGGCGAATTATATGAC
MPU_68	GGGTTAGGGTTACGGAAA	GCGATACTTCCTTTTCAAGTC
MPU_72	AAGAGCGATAGAGGAAGACAT	TGAAGTAGATTGGCAGGTTAG
MPU_73	CGATATAGCACCAGAAATAGC	AGTAACGCTTTCTCTCTCCTC
MPU_78	TGAAGAGCTTTCAAAGTGAAG	GAACGAATTTTTAGGGTGATT
MPU_80	ACCAGAAAGGTAAAGCTCTCA	TTTTTGTTACTCTCCAAGCAG
MPU_81	TTCTATCGGAGAAGAAGTCG	CACACCTCTCTCTCTCTCTCTC
MPU_83	CACTCATGGAAGGTTCTAGGT	CTCATTTTGCTTTTCAAGAGA
MPU_85	GAGTGCTAACAAGCTCAAGAA	AGGGTTCTTCTTCTTCATCAC
MPU_87	CTAAACCTGTTGAGACAAGGA	CAGACACATTTCAACAAGATG
MPU_88	GGAACTGGGTATGAGGTATTT	AGATGAGACCCCATTTGTTAC
MPU_89	ATAGCTCCGATTACCCAAAT	CACCCATTTCTCTCTTTCTCT
MPU_90	TTCCAAACAGTGGTCATAATC	TGTATTGCTGTTGGTAAGGAT
MPU_92	CGGTCGTAGTAATCGCTTC	GGTTTTGCTCTGTTTTCTCTT
MPU_94	AGTATTTGAACGAGAGGCTTT	ATCGTCGTCTTCTTCTTCTTC
MPU_96	CTCCTCTGAATACACAGCATC	TAACAGGTCCAAGAGTTCTCA
MPU_98	TGCCTTATGAGGAATCTTACA	GTATTTCCCCTTGTTCATCTT

MPU_99	ACTGCAGAATGGAACTTTGTA	TAAAAGCCCTACTCAAAATCC
MPU_100	GGGTACGGATAAAGAGATGAT	CACAACATTCATTTGATTGAC
MPU_101	TGGAAATAGGAATGAGTTTGA	TATAAGAAAATCCCGTTGTCA
MPU_102	GAGAGAACTTCAATTCCCCTA	GCTTGTTCATGTTTCTTTGTC
MPU_106	GGAGAAATCAAATAGGTCGTC	AGTTCAAACCTAACCGAACTC
MPU_107	CGTATTTGTTGAACAGGATGT	GTCTCTCCTACCTCGATCAGT
MPU_108	GGCATAGATACTGAACGAGAA	GCAGAGAAAATACAAAACCTG
MPU_111	CATCAACCTTATTCGACTCAT	AGTTCCTCCGTTACGTCAT
MPU_114	GTCTCTTTCACTCTTCCGTTT	GCAAGGTCTCTTTCACTCTTT
MPU_115	AAAGGCAATTGAGATCTAAGAA	CCTTTGTCTCCTCTTTCTTTC
MPU_121	TCATCACCATCACCATCAC	GATATCTGCCAGGTCCATC
MPU_122	TTCCTCGTGGAGATACAGATA	TGATGCATGCTATGATTAGAA
MPU_123	TTGTGCTGTTGTTATGATTGA	CTTTGTTTAAAACTGAACCA
MPU_124	CCTTCCCTTTAGATGTGAAAT	ACATTGATAGCAGTGGAGAAA
MPU_125	AGGAGAGAGAGTGAAATTGGA	ACAACGTGAACAGAGAGAGAA
MPU_130	TCTATGGGTTTGTCCTCATC	CATTTTTCCACAATCACTTTC

4.3.3. Assignment of putative chromosome location

As reference genome and microsatellite-based high-resolution genetic linkage map are not available for *M. pruriens*, the putative or pseudo chromosome locations of all 66 SSR markers were allocated using chromosome sequences of *Phaseolus vulgaris* with the assumption that chromosomes of both the species are syntenic and most species within the Milletieae tribe of Fabaceae have 11 pairs of chromosomes (Doyle 2012; Cannon et al. 2015). The putative chromosome number was allocated based on the homology mapping in EnsemblPlants - *Phaseolus vulgaris* database (https://plants.ensembl.org). The sequences of SSR markers of *M. pruriens* validated by Sathyanarayana et al. (2017) were used as a search query in EnsemblPlants BLAST tool (https://plants.ensembl.org/Multi/Tools/Blast) against the genome sequence of *Phaseolus vulgaris* where an *e-value* $\leq$ 1e-05 was used as a threshold. The chromosome number with maximum hit was allocated as the putative location of the particular SSR marker in the genome.

4.3.4. Population structure

Population structure and genetic relationship of the AM panel accessions were determined according to the method described in section 4.2.3 of the materials and methods. In addition, the kinship coefficients (F_{ij}; individual level) were estimated following Loiselle et al. (1995) to determine the relatedness between the individuals. TASSEL v 5.0 (Bradbury et al. 2007) was used to generate the kinship coefficient matrix among all pairs of accessions. The kinship heat map was obtained using GAPIT, R package (Wang et al. 2014).

4.3.5. Linkage disequilibrium

Linkage disequilibrium (LD) was estimated using the square value of the correlation coefficient (R^2). The values of R^2 between all pairs of SSR markers were obtained using software package TASSEL v 5.0 (Bradbury et al. 2007) and R^2 values with $P < 0.001$ were considered significant. To determine the extent of LD in the mapping population, the LD was estimated separately for each pseudo-*M. pruriens* chromosome as well as across the genome.

4.4. Association analysis

4.4.1. Phenotyping

Sixteen traits related to seed (seed length, seed width, seed thickness, seed yield per plant, hundred seed weight), pod (pod length, pod width, number of pods per cluster, number of pods per plant), inflorescence (inflorescence length, flower buds per inflorescence, flower length, pedicel length), and biochemical attributes (L-DOPA, total protein, total carbohydrate) were evaluated for two consecutive years (2015 and 2016) as per the descriptor (Sathyanarayana et al. 2012). For data scoring, we used manual measurements using Vernier caliper, weighing balance, and counting as

relevant for the trait. The L-DOPA content was estimated as per Daxenbichler et al. (1972). The total protein and total carbohydrate contents were estimated using Bradford's (Bradford 1976) and Anthrone (Hedge and Hofreiter 1962) methods, respectively. All statistical analyses were performed using the R program (R Core Team, 2014).

4.4.2. Estimation of variance components and broad-sense heritability

To estimate the variance, genotypic and phenotypic coefficients of variation were evaluated using Syukur et al. (2012) as follows:

$\sigma^2G = (MSG - MSE) / r$

$\sigma^2P = \sigma^2G + \sigma^2E/r$

Where, σ^2G = genotypic variance, σ^2P = phenotypic variance, σ^2E = environmental variance (error mean square from the analysis of variance); MSG = mean square of genotypes; MSE = error mean square; r = number of replications.

The broad-sense heritability for each trait was estimated according to Allard (1960) as follows:

$H^2 = (\sigma^2G/ \sigma^2 P) \times 100$

Where, H^2= heritability in broad-sense; σ^2G = genotypic variance; σ^2P = phenotypic variance.

4.4.3. Genetic diversity

Genetic diversity analysis of the AM panel was carried out using the SSR data generated from the 66 species-specific SSR markers (Table 12) as per the method described in section 4.2.2.

4.4.4. Association mapping

TASSEL v 5.0 (Bradbury et al. 2007) was used for the association analysis. A mixed linear model (MLM) was employed to determine the marker-trait association (MTA). The MLM is considered superior over the general linear model (GLM), as GLM incorporates only population structure (Q matrix) which often results in false positives; instead, MLM incorporates both Q and K matrix (kinship) which overcomes this limitation (Yu et al. 2006). Normally in such studies, Bonferroni multiple test correction ($P = 0.05/n$; n is the number of markers used in the study) is applied to obtain the threshold *P*-value as it provides a stringent cut-off to avoid false positives. However, it assumes that the loci are independent, which is not always true, given that certain loci may be in linkage disequilibrium. Therefore, to avoid loss of beneficial MTAs, in addition to Bonferroni correction, we also used a less stringent criterion, $P = 1/n$ as an additional cut-off. This approach has been beneficially used in some earlier studies (Li et al. 2012; Yang et al. 2014; Xu et al. 2018). The Manhattan plots and quantile-quantile (Q-Q) plots for MTAs were generated using CMplot, R package (Yin et al. 2021).

4.4.5. Annotation of marker function

Further, to test the reliability of our MTAs, the putative function of each trait-associated marker was determined using the *Arabidopsis thaliana* (L.) Heynh. genome database (https://www.arabidopsis.org). For this, the transcripts of SSR markers were used as a query sequence and matched against the reference genome sequences of *A. thaliana* with the BLAST tool, and then the top hits were selected as the putative gene/function of the respective marker. An e-value ≤ 1e-05 was used as a threshold.

CHAPTER 5

RESULTS

5. RESULTS

5.1. Variability for seed traits in *M. pruriens* population

5.1.1. Germplasm establishment

A total of 60 *M. pruriens* accessions obtained from different locations of Northeast India were established in the garden of Sikkim University (Figure 14). The collection comprised of two botanical varieties, of which 16 were of *M. pruriens* var. *utilis* (cultivated) and the remaining 44 were of the var. *pruriens* (wild). They belonged to two geographic regions *viz.* Eastern Himalaya (25 accessions) and Indo-Burma (35 accessions) (Table 9).

Figure 14. *M. pruriens* germplasm established at Sikkim University.

5.1.2. Variability for the seed traits

The one-way ANOVA revealed significant variability ($P < 0.05$) for the six seed-based economic traits such as seed length, seed width, seed thickness, hundred seed weight, total protein, and total carbohydrate contents in 60 accessions (Table 13). The statistics of six seed-based traits for var. *utilis* and var. *pruriens* and entire accessions of Northeast India are depicted in Tables 14 and 15. In var. *utilis*, the total protein and carbohydrate contents ranged from 27.35 to 33.16% and 27.50 to 60.50%. The seed length, seed width, seed thickness, and hundred seed weight varied from 15.40 to 21.20 mm, 11.20 to 15.20 mm, 7.20 to 11.40 mm, and 108.86 to 263.99 gm, respectively. The highest and the lowest coefficient of variation (CV) were observed for hundred seed weight (CV = 26.84%) and total protein content (CV = 6.08%). In var. *pruriens*, the total protein and carbohydrate contents ranged from 19.21 to 31.75% and 23.03% to 56.33%. The seed length, seed width, seed thickness, and hundred seed weight varied from 7.80 to 15.80 mm, 5.40 to 10.80 mm, 3.20 to 7.60 mm, and 11.51 to 80.92 gm, respectively. The highest and the lowest CV were observed in hundred seed weight (CV = 51.30%) and total protein content (CV = 6.24%).

Table 13. One-way ANOVA for six seed-based traits in *M. pruriens* of Northeast India.

Trait	F	F crit	*P*-value	Sig
Seed length	210.881	4.007	5.6404E-21	***
Seed width	207.298	4.007	8.33688E-21	***
Seed thickness	151.962	4.007	7.64141E-18	***
Hundred seed weight	270.026	4.007	1.72642E-23	***
Total protein	8.357	4.007	0.0054	**
Total carbohydrate	17.087	4.007	0.0001	***

*F, calculated F value; F crit, F critical value; Sig, significance; ***significant at P < 0.001; **significant at P < 0.01.*

Table 14. Variability for six seed-based traits in *M. pruriens* of Northeast India.

Mucuna pruriens* var. *utilis					
Traits	**Minimum**	**Maximum**	**Mean**	**SE**	**CV (%)**
Seed length (mm)	15.40	21.20	17.58	0.34	7.63
Seed width (mm)	11.20	15.20	12.69	0.24	7.59
Seed thickness (mm)	7.20	11.40	9.11	0.33	14.29
Hundred seed weight (gm)	108.86	263.99	155.22	10.41	26.84
Protein (%)	27.35	33.16	29.97	0.45	6.08
Carbohydrate (%)	27.50	60.50	43.87	2.60	23.73
Mucuna pruriens* var. *pruriens					
Traits	**Minimum**	**Maximum**	**Mean**	**SE**	**CV (%)**
Seed length (mm)	7.80	15.80	10.16	0.28	18.42
Seed width (mm)	5.40	10.80	7.55	0.19	17.23
Seed thickness (mm)	3.20	7.60	5.13	0.15	20.10
Hundred seed weight (gm)	11.51	80.92	32.48	2.51	51.30
Protein (%)	19.21	31.75	28.46	0.26	6.24
Carbohydrate (%)	23.03	56.33	34.78	0.94	17.91

SE, standard error; CV, coefficient of variation.

Table 15. Variability for six seed-based phenotypic traits in *M. pruriens* from Northeast India.

Accession No.	**SL**	**SW**	**ST**	**SWt**	**TPr**	**TCr**
500268-NL	21.20	15.20	11.20	263.99	30.85	41.05
500269-MZ	16.80	12.80	7.40	137.90	32.04	28.11
500270-AR	17.80	12.60	9.20	152.95	27.63	47.83
500271-NL	18.80	14.20	11.40	226.16	33.16	27.50
500272-MN	17.40	12.40	7.80	122.02	29.73	62.50
500273-NL	15.80	11.60	9.40	134.38	29.44	48.33
500276-NL	18.20	12.60	9.80	167.13	29.15	39.66
500278-NL	16.60	12.60	7.20	108.86	27.92	45.33
500279-NL	18.20	12.60	9.20	157.68	29.13	43.83
500280-NL	17.80	12.60	9.80	147.35	32.33	39.66
500284-MN	15.40	11.20	7.20	115.65	29.94	51.66
500309-TR	17.80	12.60	8.40	194.48	30.61	43.02
500314-AR	18.40	11.60	9.20	115.77	32.46	33.83
500317-AR	17.20	12.60	9.20	143.14	29.77	65.50
500320-MN	17.20	12.40	8.80	152.89	27.35	37.33
500321-MN	16.60	13.40	10.60	143.13	27.96	46.83
500274-AS	11.40	8.40	5.40	44.92	30.27	30.33

500275-SK	9.80	7.80	5.80	44.41	27.62	28.33
500277-ML	14.20	10.60	7.60	80.92	30.13	29.16
500281-MZ	8.20	6.40	3.60	22.55	27.58	41.33
500282-MZ	7.80	5.80	4.20	18.50	27.25	33.04
500283-MZ	7.80	5.60	3.60	16.72	29.18	25.50
500285-AS	9.80	6.20	5.20	29.60	29.79	33.66
500286-SK	11.20	8.40	6.60	46.93	27.77	30.16
500287-SK	9.40	6.60	4.40	41.11	28.62	38.33
500288-SK	8.20	7.20	4.40	21.05	28.98	36.50
500289-MZ	8.60	6.40	3.80	16.92	30.32	39.83
500290-AS	10.60	8.20	3.60	27.30	29.21	36.16
500291-ML	9.40	6.60	4.60	27.94	29.50	39.50
500292-AS	9.40	6.60	3.80	20.72	27.04	23.03
500293-SK	10.40	7.80	5.60	32.43	27.56	30.66
500294-MN	8.40	5.80	4.20	16.51	28.25	34.33
500295-MZ	9.40	7.60	5.20	31.14	19.21	56.33
500296-MZ	8.60	6.80	4.20	29.02	27.63	37.33
500297-MZ	7.80	5.40	3.20	15.07	27.06	42.16
500298-ML	8.20	6.40	4.60	25.12	29.42	35.66
500299-SK	10.40	7.80	5.20	39.90	28.89	31.33
500300-MN	10.20	7.20	5.40	29.96	29.36	31.83
500301-TR	8.20	6.20	3.60	15.61	31.75	41.33
500302-AR	12.40	10.40	6.40	72.51	29.29	39.83
500303-MZ	11.80	8.20	5.20	42.63	28.65	35.05
500304-SK	9.80	7.60	5.20	33.68	28.37	30.16
500305-AS	11.60	8.60	6.20	23.76	27.63	34.16
500306-TR	9.40	6.80	4.40	12.21	27.75	26.33
500307-TR	9.40	6.80	5.60	29.00	28.11	32.16
500308-TR	10.40	7.80	5.40	11.51	28.71	38.66
500310-TR	10.20	7.80	6.60	28.01	30.23	33.16
500311-ML	10.40	7.60	4.60	24.55	29.02	39.16
500312-ML	8.40	8.20	5.60	23.73	26.61	33.33
500313-ML	8.40	7.60	5.60	27.70	28.25	49.33
500315-AR	15.80	10.80	5.80	78.10	28.46	34.01
500316-AR	11.20	8.40	6.20	49.50	28.73	37.03
500318-AR	10.60	8.40	5.60	23.20	28.96	24.50
500319-MN	10.20	7.20	5.40	69.41	27.58	26.83
500322-AS	14.60	10.20	7.40	23.52	28.75	40.03

500323-AS	12.20	8.20	5.20	42.90	30.06	36.33
500324-AS	13.60	9.20	6.40	24.24	27.63	32.02
500325-SK	10.40	7.20	5.20	34.51	29.08	30.83
500326-SK	8.60	6.20	4.40	27.86	28.77	37.33
500327-SK	10.20	7.20	5.40	32.31	29.16	34.16
Minimum	7.80	5.40	3.20	11.51	19.21	23.03
Maximum	15.80	10.80	7.60	80.92	31.75	56.33
Average	10.16	7.55	5.13	32.48	28.46	34.78
SE	0.14	0.09	0.08	1.21	0.13	0.45
CV	18.42	17.23	20.10	51.30	6.24	17.91

SL, seed length (mm); SW, seed width (mm); ST, seed thickness (mm), HSW, hundred seed weight (gm); TPr, total protein (%); TCr, total carbohydrate (%); SE, standard error; CV, coefficient of variation.

The relationships between the six seed-based traits were further analyzed by correlation coefficients coupled with a two-tailed t-test for determining the strength of the correlation. A significant positive correlation ($P < 0.05$) was observed among all the studied seed traits except between the protein and carbohydrate content. The correlation matrix of six seed-based traits is depicted in Table 16.

Table 16. Correlations among the six seed-based traits.

	SL	**SW**	**ST**	**HSW**	**TPr**	**TCr**
SL	1					
SW	0.973^{**}	1				
ST	0.922^{**}	0.935^{**}	1			
HSW	0.928^{**}	0.922^{**}	0.903^{**}	1		
TPr	0.390^{**}	0.344^{**}	0.341^{**}	0.386^{**}	1	
TCr	0.347^{**}	0.377^{**}	0.321^{*}	0.339^{**}	-0.136^{ns}	1

*SL, seed length; SW, seed width; ST, seed thickness; HSW, hundred seed weight; TPr, total protein; TCr, total carbohydrate, **significant at correlation at 0.01 level (2-tailed), *significant correlation at 0.05 level (2-tailed), ^{ns}not significant.*

To understand the pattern of diversity among the seed traits, principal component analysis (PCA) was performed. The first two components (PCs) explained 87.29% of the total variance. The PC1 contributed 68.34% of the total phenotypic variation and revealed the highest positive correlation with seed length (0.98), seed width (0.98), seed thickness (0.96), and hundred seed weight (0.96). The PC2 contributed 18.95% of the phenotypic diversity, which was positively correlated with total carbohydrate content (0.76) and negatively correlated with the total protein content (-0.74). Wild and cultivated varieties of *M. pruriens* formed a separate cluster in the PCA scatter plot. The PCA scatter plot and Eigen values of five PCs for six phenotypic traits are provided in Figure 15 and Table 17.

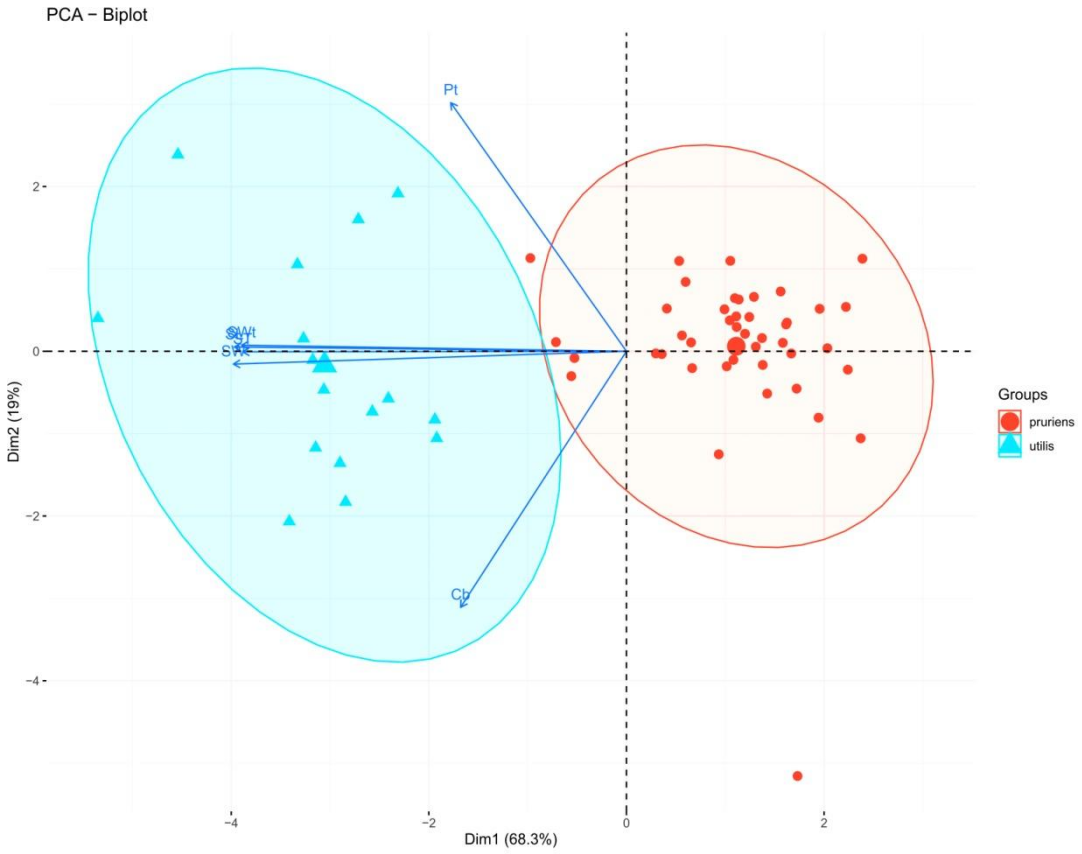


Figure 15. Scatter plot of *M. pruriens* accessions from Northeast India based on six seed-based traits from the principal component analysis (PCA). Note: red circles and sky-blue triangles represent var. *pruriens*, and var. *utilis*, respectively.

Table 17. Eigen values of the five principal components (PCs) for six seed-based traits.

Traits	PC1	PC2	PC3	PC4	PC5
Seed width	0.98	0.04	-0.11	0.08	-0.10
Seed length	0.98	-0.01	-0.09	0.09	-0.15
Hundred seed weight	0.96	-0.02	-0.08	-0.27	0.01
Seed thickness	0.96	0.00	-0.16	0.10	0.23
Total carbohydrate	0.41	0.76	0.49	0.00	0.01
Total protein	0.44	-0.74	0.51	0.01	0.01
Eigen values	4.10	1.14	0.55	0.10	0.08
% of Variance	68.34	18.95	9.24	1.64	1.42
Cumulative %	68.34	87.29	96.53	98.18	99.59

5.2. Genetic diversity among the Northeast Indian accessions

5.2.1. Marker attributes and genetic diversity

Of the total 25 genic-microsatellite markers used in the study, 22 produced polymorphic and 3 markers (MPU_07, MPU_09, and MPU_61) produced monomorphic profiles after three repeated amplifications. Therefore, further analyses were carried out using the data from the 22 polymorphic markers. The details of the marker attributes are provided in Table 18. A total of 61 alleles were generated with an average of 2.77 alleles per marker. The number of alleles ranged from 2 (in many) to 4 (in MPU_04, MPU_05, MPU_13, and MPU_23) per marker. The mean value of major allele frequency (MAF = 0.692) was high and ranged from 0.51 in MPU_124 to 0.88 in MPU_05. The expected heterozygosity (He) per marker varied from 0.219 (MPU_14) to 0.646 (MPU_13) with a mean value of 0.477. A low average value of observed heterozygosity (Ho = 0.020) was recorded, which ranged from 0 in many microsatellite markers to 0.401 in marker MPU_13. The average value of polymorphism information content (PIC = 0.312) was moderate (PIC between 0.25 - 0.5), which varied from the lowest of 0.172 in MPU_05 to the highest of 0.375 in MPU_124. A major proportion of microsatellite markers (19) were moderately

polymorphic with PIC > 0.25. Only three markers *viz*. MPU_05, MPU_06, and MPU_14 produced low information with PIC < 0.25. Based on the PIC values, markers such as MPU_10, MPU_19, MPU_20, MPU_124, and MPU_125 were determined as informative with all of them showing PIC > 0.35. All the microsatellite markers were found to have deviated from the Hardy-Weinberg equilibrium (HWE) at a significance level of P < 0.001.

Table 18. Marker attributes of microsatellite markers used for genetic diversity analysis.

Markers	No. of Alleles	MAF	Gene Diversity	He	PIC	HWE
MPU_01	3	0.706	0.364	0.566	0.283	***
MPU_02	3	0.720	0.380	0.515	0.304	***
MPU_04	4	0.667	0.372	0.627	0.333	***
MPU_05	4	0.882	0.197	0.367	0.172	***
MPU_06	3	0.816	0.291	0.392	0.245	***
MPU_08	2	0.733	0.391	0.367	0.315	***
MPU_10	2	0.569	0.490	0.469	0.370	***
MPU_11	3	0.756	0.354	0.511	0.288	***
MPU_12	3	0.651	0.454	0.359	0.351	***
MPU_13	4	0.663	0.425	0.646	0.331	***
MPU_14	2	0.873	0.222	0.219	0.197	***
MPU_15	3	0.745	0.363	0.537	0.294	***
MPU_16	2	0.714	0.408	0.401	0.325	***
MPU_17	3	0.692	0.401	0.556	0.318	***
MPU_18	2	0.695	0.424	0.381	0.334	***
MPU_19	2	0.634	0.464	0.447	0.356	***
MPU_20	2	0.591	0.483	0.481	0.367	***
MPU_22	3	0.667	0.390	0.580	0.303	***
MPU_23	4	0.667	0.401	0.598	0.313	***
MPU_25	3	0.667	0.420	0.600	0.328	***
MPU_124	2	0.517	0.499	0.407	0.375	***
MPU_125	2	0.593	0.483	0.479	0.366	***
Mean	2.773	0.692	0.394	0.477	0.312	-
SE	0.160	0.019	0.017	0.023	0.011	-

*MAF, major allele frequency; He, expected heterozygosity; PIC, polymorphism information content; HWE, Hardy-Weinberg equilibrium; SE, standard error; ***significant at P-value < 0.001.*

The effective number of alleles (Ne) ranged from 1.29 to 3.03, with an average of 2.03. The mean value of Ne was lower than that of the total number of alleles (Na = 2.773), indicating the contribution of fewer alleles for the variation. The average values of Shannon's information index (I = 0.496) and Nei's gene diversity (h = 0.326) were suggestive of moderate to high genetic diversity which was further supported by total gene diversity (H_T) value of 0.330. The average values of the genetic differentiation indices such as G_{ST} (0.038) and F_{ST} (0.061) were high indicating low genetic differentiation among the population groups. The high value of gene flow (Nm = 7.48) indicated widespread genetic exchange among the population groups. The genetic diversity indicators obtained across different microsatellite alleles are depicted in Table 19.

To ascertain the genetic diversity between and within the populations, the accessions were assumed into different population groups based on geographical locations of collection and varietal affiliation. Based on geographical location, accessions were assigned into two groups *viz.* Indo-Burma (IB) region, and Eastern Himalayan (EH) region. Likewise, the accessions were grouped based on varietal affiliation into var. *utilis* (cultivated) and var. *pruriens* (wild). Among the two population groups based on the geographical location, I varied from 0.468 to 0.479 with an average of 0.474, and h varied from 0.307 to 0.312 with an average of 0.309 (Table 20). The percentage of polymorphic SSR loci (%P) was found to be 95.45% in the IB subgroup, followed by 95.21% in the EH subgroup. Between the two botanical varieties, I varied from 0.444 to 0.476 with an average of 0.460 and h from 0.290 to 0.311, with an average of 0.311 (Table 20). The percentage of polymorphic SSR loci was found to be 95.45 and 90.91% in var. *pruriens* and var. *utilis*.

Table 19. Genetic diversity indicators obtained using microsatellite markers across all populations of Northeast India.

Markers	Na	Ne	h	I	H_T	Hs	G_{ST}	F_{ST}
MPU_01	3	2.304	0.259	0.410	0.264	0.263	0.005	0.004
MPU_02	3	2.122	0.283	0.447	0.291	0.271	0.057	0.114
MPU_04	4	2.699	0.295	0.468	0.295	0.292	0.009	0.015
MPU_05	4	1.595	0.186	0.301	0.184	0.181	0.019	0.026
MPU_06	3	1.868	0.255	0.401	0.260	0.232	0.076	0.124
MPU_08	2	1.610	0.373	0.552	0.370	0.359	0.027	0.039
MPU_10	2	1.893	0.411	0.600	0.411	0.402	0.021	0.033
MPU_11	3	2.066	0.275	0.435	0.274	0.271	0.011	0.017
MPU_12	3	1.580	0.398	0.584	0.421	0.353	0.164	0.292
MPU_13	4	3.036	0.313	0.485	0.318	0.287	0.084	0.101
MPU_14	2	1.293	0.291	0.447	0.291	0.284	0.019	0.025
MPU_15	3	2.164	0.278	0.439	0.278	0.277	0.004	0.007
MPU_16	2	1.699	0.380	0.561	0.387	0.375	0.030	0.046
MPU_17	3	2.311	0.289	0.457	0.297	0.280	0.052	0.095
MPU_18	2	1.679	0.386	0.569	0.394	0.362	0.076	0.113
MPU_19	2	1.823	0.401	0.588	0.401	0.394	0.017	0.027
MPU_20	2	1.930	0.408	0.597	0.410	0.407	0.005	0.009
MPU_22	3	2.386	0.288	0.448	0.288	0.287	0.004	0.008
MPU_23	4	2.505	0.290	0.455	0.289	0.287	0.009	0.020
MPU_25	3	2.543	0.294	0.465	0.294	0.289	0.018	0.031
MPU_124	2	1.704	0.414	0.605	0.426	0.378	0.112	0.178
MPU_125	2	1.922	0.408	0.597	0.410	0.407	0.009	0.014
Mean	2.773	2.033	0.326	0.496	0.330	0.315	0.038	0.061
SE	0.160	0.091	0.014	0.018	0.015	0.013	0.009	0.015

Na, number of different alleles; Ne, effective no. of alleles; h, Nei's gene diversity; I, Shannon's information index; H_T, total gene diversity; Hs, subdivided population; G_{ST}, genetic differentiation; F_{ST}, genetic differentiation among populations, SE, standard error.

Table 20. Genetic diversity estimates in populations based on the collection location and varietal affiliation.

Geographical location	**Na**	**Ne**	**Ho**	**He**	**h**	**I**	**%P**
Indo-Burma Hotspot	2.686	2.110	0.021	0.486	0.312	0.479	95.45%
Eastern Himalaya	2.591	2.002	0.022	0.460	0.307	0.468	95.21%
Mean	2.638	2.056	0.022	0.473	0.309	0.474	95.33%
SE	0.034	0.038	0.001	0.009	0.002	0.003	0.001
Botanical varieties	**Na**	**Ne**	**Ho**	**He**	**h**	**I**	**%P**
var. *utilis*	2.500	1.935	0.006	0.442	0.29	0.444	90.91%
var. *pruriens*	2.682	2.062	0.027	0.460	0.311	0.476	95.45%
Mean	2.591	1.998	0.017	0.451	0.301	0.460	93.23%
SE	0.091	0.064	0.011	0.009	0.011	0.016	0.023

Na, number of different alleles; Ne, effective no. of alleles; Ho, observed heterozygosity; He, expected heterozygosity; h, Nei's gene diversity; I, Shannon's information index; %P, percentage of polymorphic loci; SE, standard error.

5.2.2. Population structure

The population structure based on Bayesian statistics revealed the presence of two major groups as ΔK was maximum at K = 2 (Figure 16). The 60 *M. pruriens* accessions from the Northeast were assigned into two major groups: STR-I and STR-II. STR-I comprised of 35 accessions (11 var. *utilis* and 24 var. *pruriens*) of which 17 were derived from EH and 18 were from the IB region. Similarly, STR-II was comprised of 25 accessions (5 var. *utilis* and 20 var. *pruriens*) of which 9 were affiliated to EH and 16 were to IB. Out of the total accessions evaluated, 26 showed >80% of shared ancestry and these accessions were considered as admixtures. The STRUCTURE analysis was heterogeneous, i.e., no clear-cut grouping was observed either based on geographical or varietal affiliation.

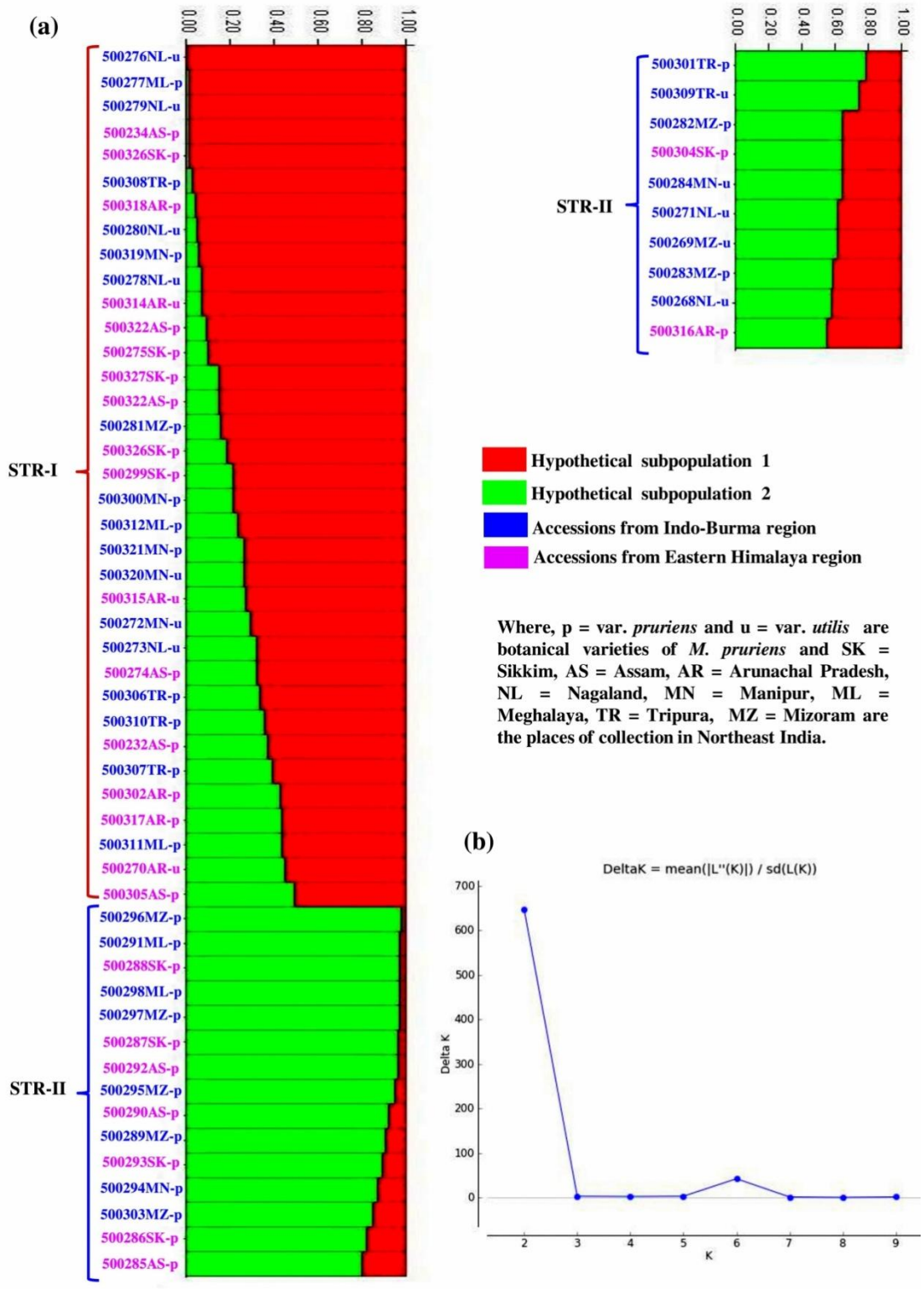


Figure 16. Population structure of *M. pruriens* population from Northeast India inferred using STRUCTURE software **(a)** population structure based on inferred ancestry (Q-matrix); STR- I and STR-II indicate the names of two subpopulations **(b)** hypothetical subpopulation estimation of *M. pruriens* from Northeast India using ΔK-values revealing the highest peak at K = 2.

The minor peak at K = 6 indicated the existence of subgroups in two major groups. Therefore, independent STRUCTURE analysis for both STR-I and STR-II was performed in which the highest peak was observed at K = 3 for STR-I, and the resultant subgroups were named as STR-I (A), STR-I (B), and STR-I (C) (Figure 17). Subgroup STR-I (A) was comprised of var. *pruriens* majority of which were from the EH region, while subgroup STR-I (B) was comprised of var. *utilis*, all derived from the IB region; STR-I (C) had accessions from both the regions and varieties. For STR-II, the highest peak was observed at K = 2, and the resultant subgroups were named STR-II (A) and STR-II (B) (Figure 17). Subgroup STR-II (A) was entirely comprised of var. *pruriens* in which the greater number of the accessions represented the EH region, while a similar trend for the IB derived accessions was noticed in subgroup STR-II (B).

Three independent analysis of molecular variance (AMOVA) was performed for the populations based on geographical location, botanical varieties, and subpopulations derived from the STRUCTURE analysis. Results of AMOVA are depicted in Table 21. In the populations based on geographical location, 96% of the variance was contributed by within-population variance and only 4% contributed to variation among the populations. In the case of botanical varieties, 6% variance was explained by among variety variance, and 94% was within the variety. In STRUCTURE-derived five subpopulations, the molecular variance among the populations was 27% and within subpopulations was 73%, respectively. The overall F_{ST} values for the populations based on the geographical locations, botanical varieties, and STRUCTURE-derived subpopulations were 0.032, 0.058, and 0.269, respectively. These data suggested maximum genetic differentiation in STRUCTURE-derived subpopulations followed by populations based on botanical variety and geographical

origin. To find out the genetic divergence among the STRUCTURE-derived five subpopulations, pairwise F_{ST} was estimated, which ranged from 0.189 between STR-II (A) and STR-II (B) to 0.410 between STR-I (B) and STR-I (C) at $P < 0.001$ (Table 22). The expected heterozygosity (He) was also estimated for each subpopulation which varied from 0.191 in STR-I (C) to 0.371 in STR-II (B) with an average of 0.317 (Table 23).

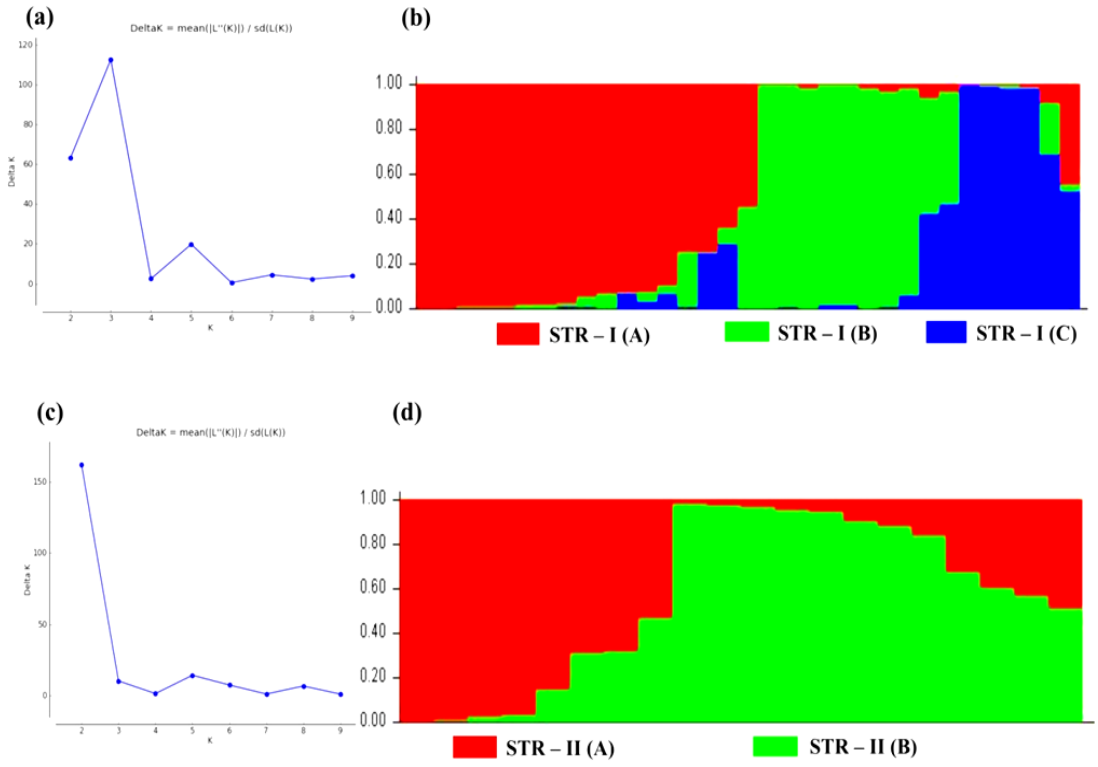


Figure 17. Population structure of STRUCTURE-derived subpopulations STR-I and STR-II **(a)** hypothetical subpopulation estimation of STR-I using ΔK values, revealing the highest peak at K = 3 **(b)** population structure analysis of STR-I based on inferred ancestry (Q-matrix); STR-I (A), STR-I (B), and STR-I (C) indicate the names of three subpopulations **(c)** hypothetical subpopulation estimation of STR-II using ΔK-values, revealing the highest peak at K = 2 **(d)** population structure analysis of STR-II based on inferred ancestry (Q-matrix); STR-II (A) and STR-II (B) indicate the names of two subpopulations.

Table 21. Analysis of molecular variance (AMOVA) among and within the populations based on geographical distribution, varietal affiliation, and STRUCTURE-derived subpopulations.

Geographical distribution				
Sources of variation	**Est Var**	**%Var**	**F_{ST}**	***P*-value***
Among population	0.217	4%		0.001
Within population	6.503	96%	0.032	0.001
Total	6.72	100%		0.001
Varietal affiliation				
Sources of variation	**Est Var**	**%Var**	**F_{ST}**	***P*-value***
Among population	0.385	6%		0.001
Within population	6.496	94%	0.058	0.001
Total	6.881	100%		0.001
Structure subpopulations				
Sources of variation	**Est Var**	**%Var**	**F_{ST}**	***P*-value***
Among populations	1.920	27%		0.001
Within Population	5.444	73%	0.269	0.001
Total	7.364	100%		0.001

*Est Var, estimated variance; %Var, percentage of variance; *With 999 data permutations.*

Table 22. Matrix depicting pairwise F_{ST} of subpopulations from STRUCTURE analysis.

	STR-I (A)	**STR-I (B)**	**STR-I (C)**	**STR-II (A)**	**STR-II (B)**
STR-I (A)	*				
STR-I (B)	0.218	*			
STR-I (C)	0.365	0.410	*		
STR-II (A)	0.228	0.336	0.347	*	
STR-II (B)	0.243	0.268	0.328	0.189	*

**F_{ST} values between the same subpopulation.*

Table 23. Expected heterozygosity (He) of STRUCTURE derived subpopulations.

STRUCTURE derived subpopulations	**Expected heterozygosity (He)**
STR-I (A)	0.370
STR-I (B)	0.324
STR-I (C)	0.191
STR-II (A)	0.328
STR-II (B)	0.371
Mean	0.317
Standard error	0.033

5.2.3. Genetic relationship

To determine the genetic relationship, we constructed the neighbor-joining (NJ) dendrogram based on a simple matching coefficient and performed principal coordinate analysis (PCoA). The accessions in the NJ tree and PcoA plots (PC1 vs. PC2) are color-coded from the information of two subgroups identified using the STRUCTURE analysis (STR-I and STR-II). The NJ algorithm grouped 60 accessions (25 from EH and 35 from IB regions) into two main clusters: Cluster-I and Cluster-II (Figure 18). Cluster-I was comprised of 26 accessions (19 var. *pruriens* and 7 var. *utilis*) derived from both the geographical regions (10 from EH and 16 from IB regions). The other cluster, Cluster-II was comprised a total of 32 accessions (24 var. *pruriens* and 8 var. *utilis*), among which 13 were from EH and 19 were from IB regions. Two accessions from EH emerged as out-group accessions. Similar to STRUCTURE, the clustering pattern in the NJ tree did not reveal any clear-cut differentiation either based on the geographical location or varietal affiliation. The PCoA scatter plot depicted a similar trend as that of the NJ tree (Figure 19). Axes I and II of PCoA explained 15.85 and 9.08% of the total variance.

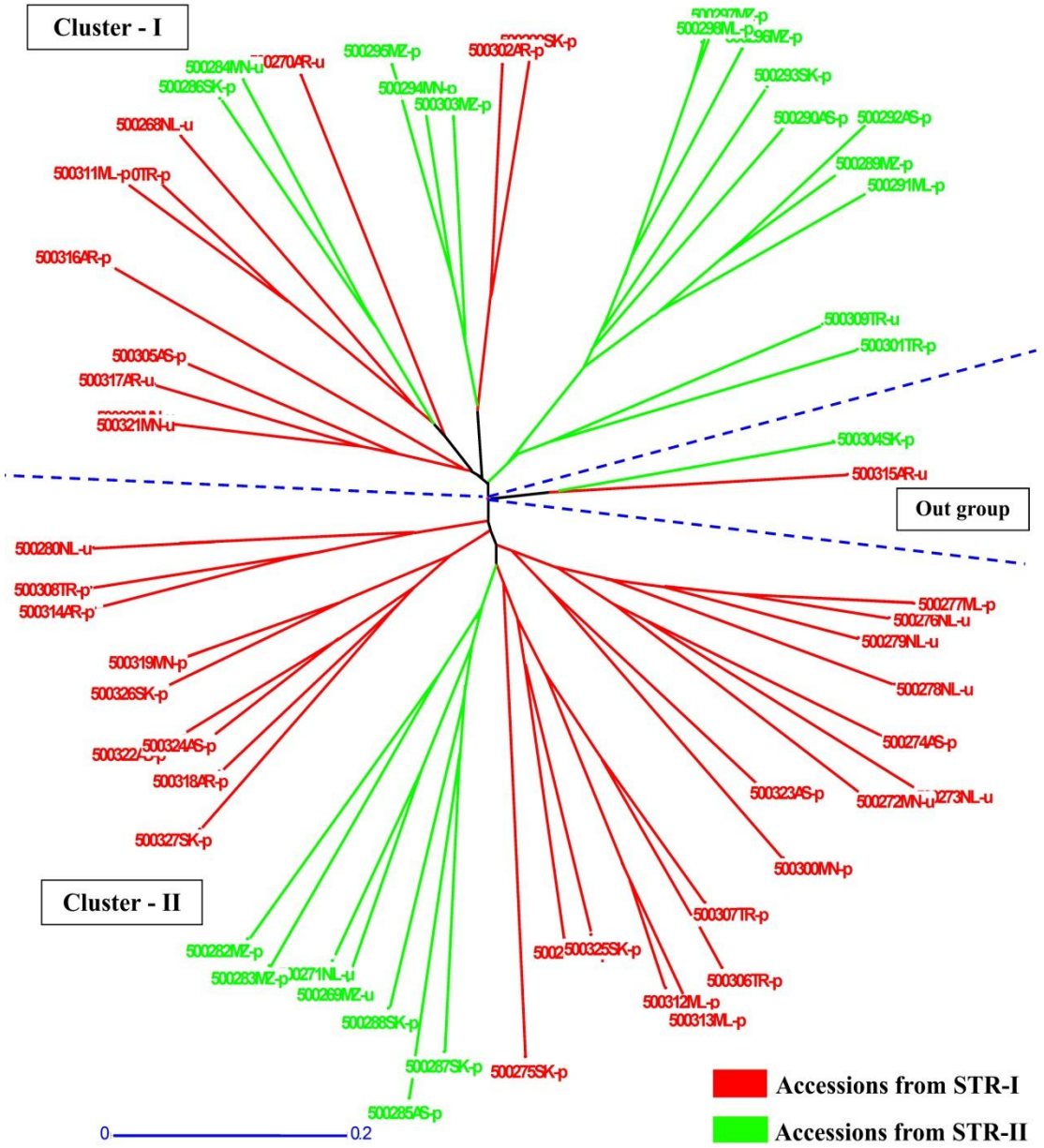


Figure 18. NJ tree depicting the genetic relationship among *M. pruriens* accessions from Northeast India. Note: Red and green colors depict the accessions belonging to two subgroups of STRUCTURE viz. STR-I and STR-II.

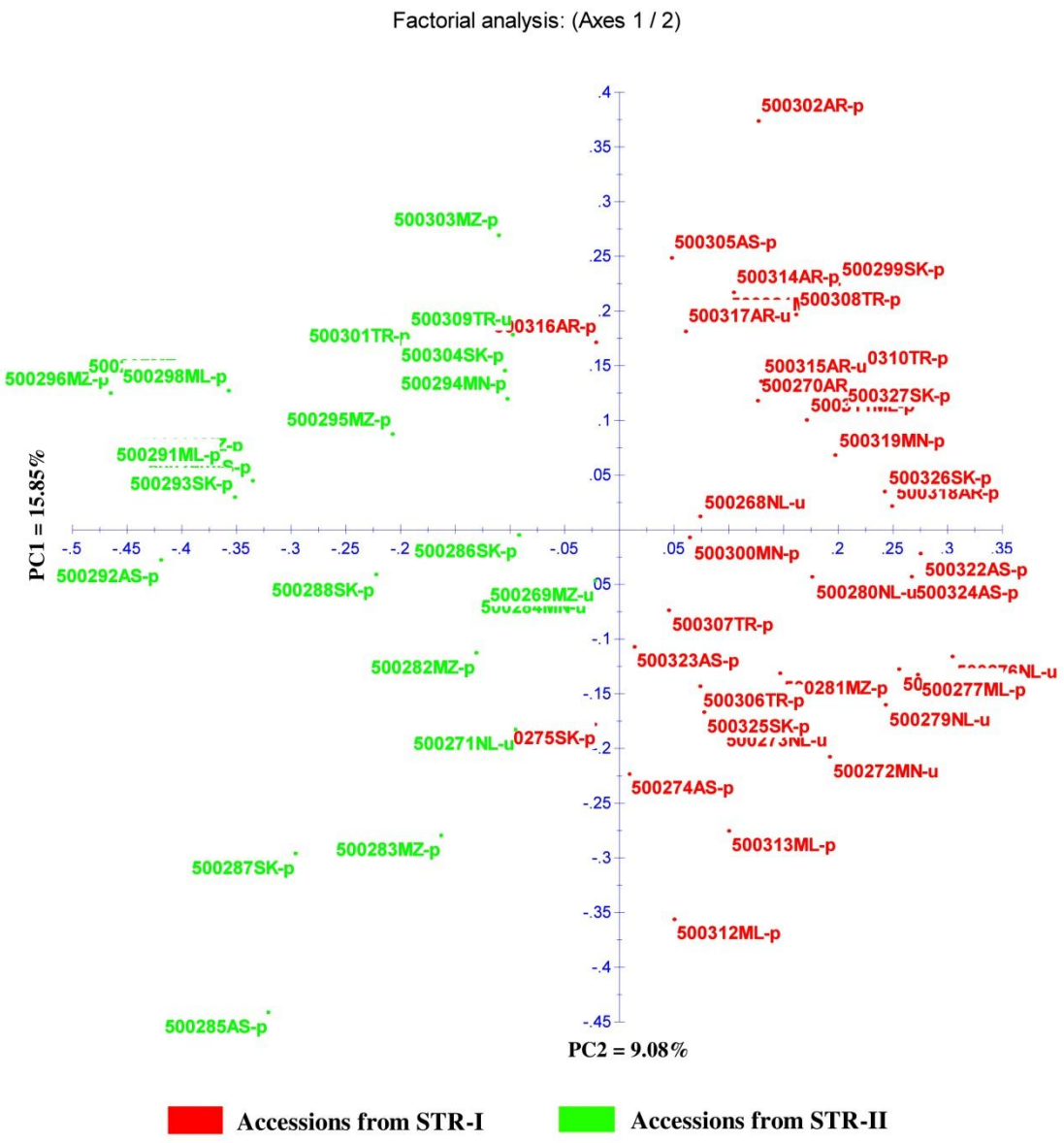


Figure 19. PCoA scatter plot based on SSR data depicting the distribution of *M. pruriens* accessions in two axes. Note: Red and green colors depict the accessions belonging to two subgroups of STRUCTURE viz. STR-I and STR-II.

5.3. Feasibility analysis of association mapping panel

5.3.1. Germplasm establishment

A total of 61 *M. pruriens* accessions obtained from the germplasm collection of Sir MVIT, Bangalore was used for linkage disequilibrium and association analysis. The panel comprised of 8 accessions of var. *utilis*, 28 of var. *pruriens*, and 25 of var. *hirsuta*. They were established in the Botanical garden of Sikkim University. These accessions represented three large geographic regions in India such as Eastern (15 accessions), West-Central (11 accessions), and Southern (35 accessions) India (Table 11). All the statistical estimates required for association mapping were performed on this panel.

5.3.2. Population structure

STRUCTURE analysis (ΔK vs. K plot) revealed a sharp peak at K = 2, indicating the presence of two genetic groups in our association mapping (AM) panel (Figure 20). Forty-eight accessions showed > 80% membership within respective subgroups, and 13 were admixtures with < 80% of shared ancestry (Figure 20). Majority of the 17 accessions in the subgroup-1 (designated as MpSTR-I), and 44 accessions in the subgroup-2 (MpSTR-II) aligned with Eastern (14) and Southern India (34), respectively. Accessions from West-Central India did not form an independent subgroup but merged with one of the two major clusters.

To ascertain the partitioning of variation, we performed AMOVA on three population groups assumed based on: (a) geographical distribution, (b) varietal affiliation, and (c) subgroups identified in STRUCTURE analysis (Table 24). All three revealed higher within gene pool variance vis-à-vis among gene pools. The ratios for within to between gene pools variance were 93:7 in case of geographical distribution, 97:3 for

varietal affiliations, and 89:11 for STRUCTURE-derived subpopulations. We estimated population differentiation (F_{ST}) within each group separately. The results revealed low to moderate differentiation in groups based on the varietal affiliation (F_{ST} = 0.032) and geographical origin (F_{ST} = 0.070) and high genetic differentiation (F_{ST} = 0.105) in groups based on STRUCTURE analysis (Table 24).

Table 24. Population differentiation and AMOVA based on geographical distribution, varietal affiliation, and STRUCTURE analysis within *M. pruriens* association panel.

Geographical distribution				
Sources of variation	Est Var	%Variance	*P*-value*	F_{ST}
Among populations	1.048	7%	0.001	
Within populations	13.841	93%	0.001	0.070
Total	14.889	100%	0.001	
Varietal affiliation				
Sources of variation	Est Var	%Variance	*P*-value*	F_{ST}
Among populations	0.463	3%	0.001	
Within populations	14.172	97%	0.001	0.032
Total	14.635	100%	0.001	
STRUCTURE				
Sources of variation	Est Var	%Variance	*P*-value*	F_{ST}
Among populations	1.619	11%	0.001	
Within populations	13.794	89%	0.001	0.105
Total	15.413		0.001	

*Est Var, estimated variance; %Var, percentage of variance; *with 999 data permutations.*

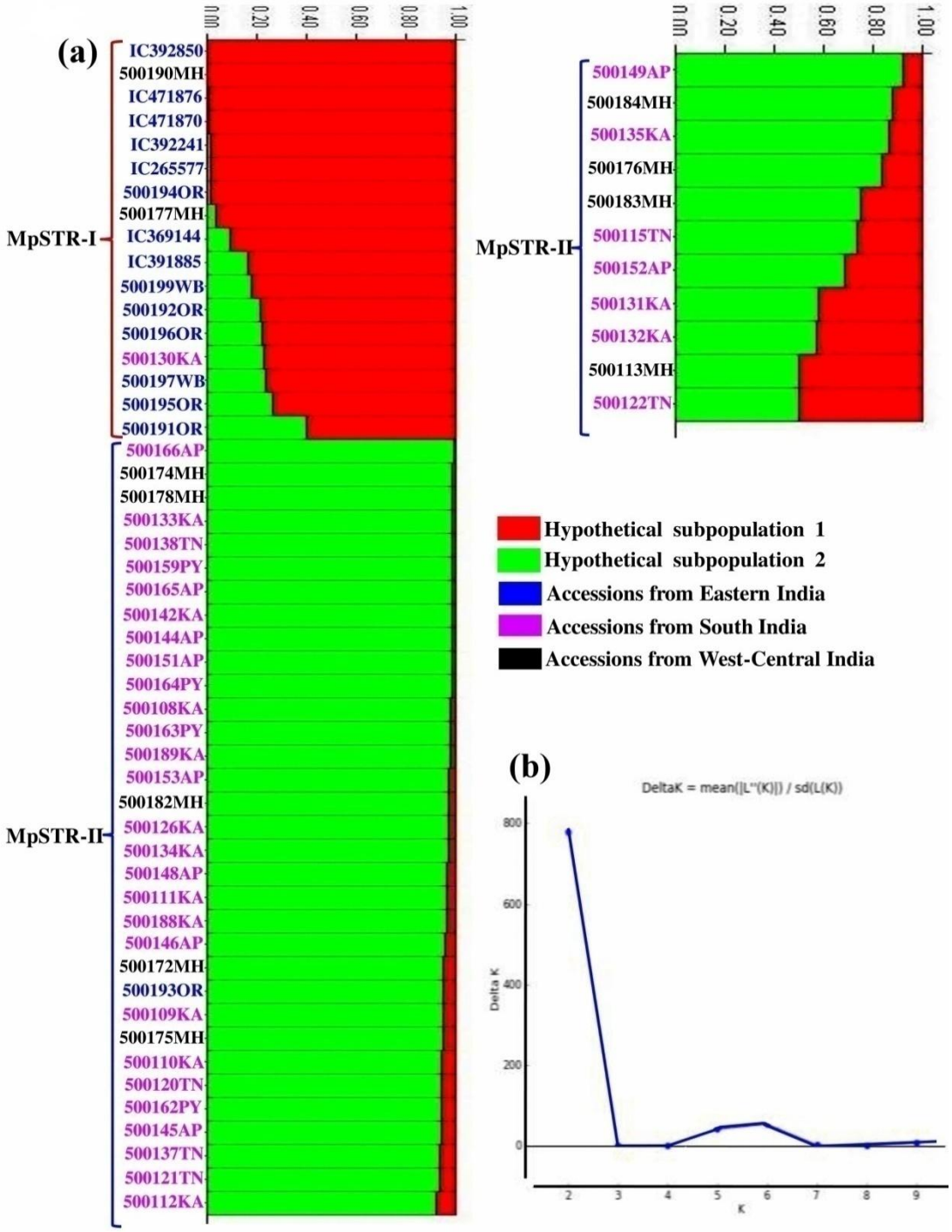


Figure 20. Population structure of *M. pruriens* association mapping panel inferred using STRUCTURE software **(a)** population structure at K = 2 based on inferred ancestry (Q-matrix) in which two sub-populations are indicated as MpSTR-I (red color) and MpSTR-II (green color) **(b)** hypothetical sub-population estimation using ΔK-values (K = 2) indicating two subpopulations.

To determine the genetic relationship, we constructed the NJ dendrogram based on a simple matching coefficient and performed PCoA. Two main clusters were revealed by the NJ algorithm, which was named MpNJ-I and MpNJ-II (Figure 21). Similar to STRUCTURE results, the majority of the accessions (25 out of 28) in MpNJ-I were derived from Southern India, three were from West-Central India. But considerable mixing was observed among the 33 accessions in MpNJ-II, as 15 of them belonged to Eastern India, eight were from West-Central India, and 10 were from Southern India. The accessions in the NJ tree and PCoA plots (PC1 vs. PC2) are color-coded from the information of two subgroups identified using the STRUCTURE analysis (MpSTR-I and MpSTR-II). We didn't find any indication of grouping based on geographical origin or varietal affiliation as accessions representing these groups were pretty much mixed up among both the clusters.

In PCoA, the principal axes 1 and 2 explained 28 and 26% of the total variance (Figure 22). The dispersion was relatively homogeneous in all four quadrants signifying the diverse nature of the accessions. The clustering pattern in PCoA largely corresponded with that of the NJ dendrogram. We also estimated the relative kinship between these accessions to measure their relatedness. About 55.29% of kinship coefficient (F_{ij}) values between any two accessions were within 0 to 0.05 and there was a subsequent reduction in frequency with an increase in kinship value (Figure 23). The heat map (Figure 24) revealed substantial differences among the accessions.

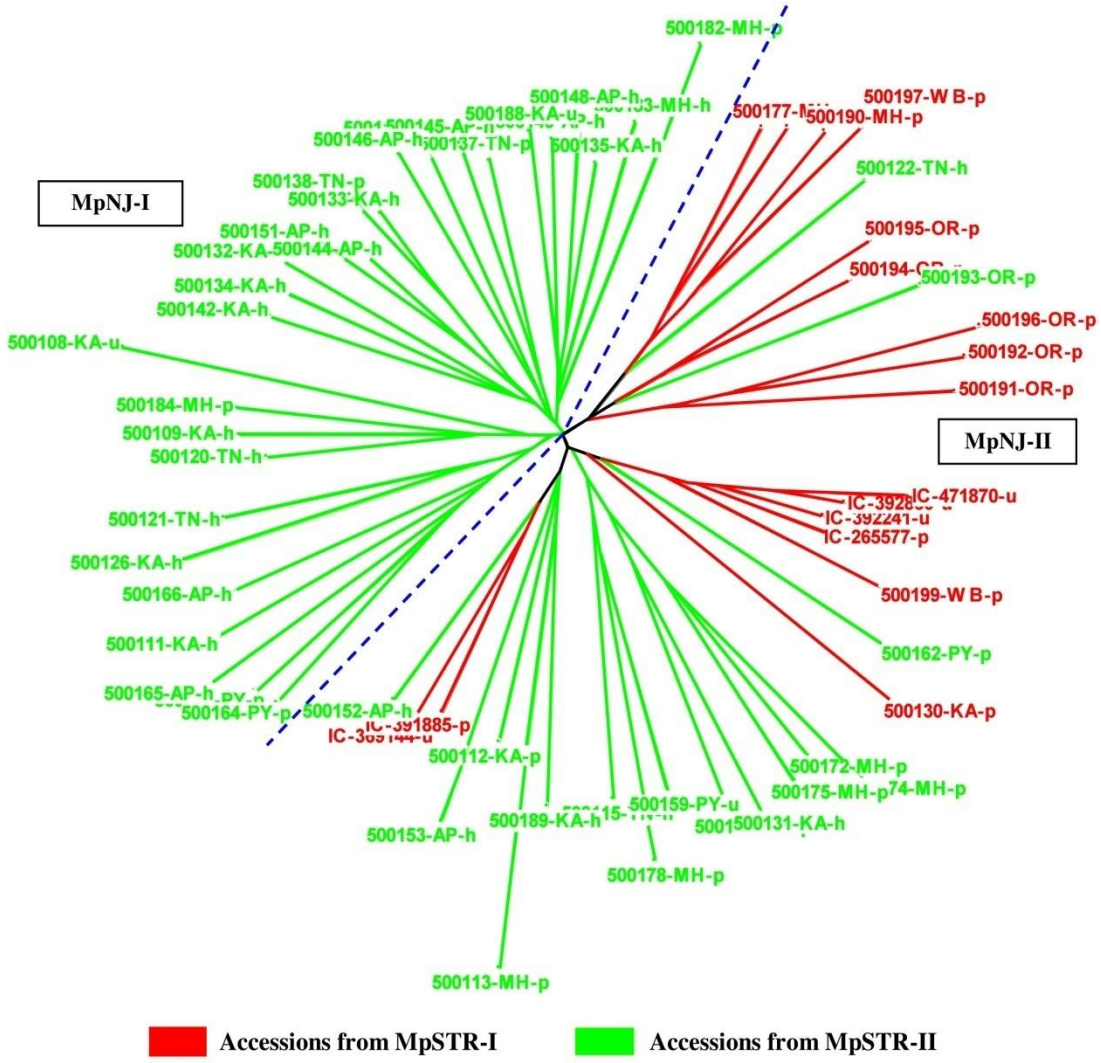


Figure 21. NJ dendrogram of AM panel based on the genetic distance. Red and green color indicates subgroup-1 (MpSTR-I) and subgroup-2 (MpSTR-II) from STRUCTURE analysis.

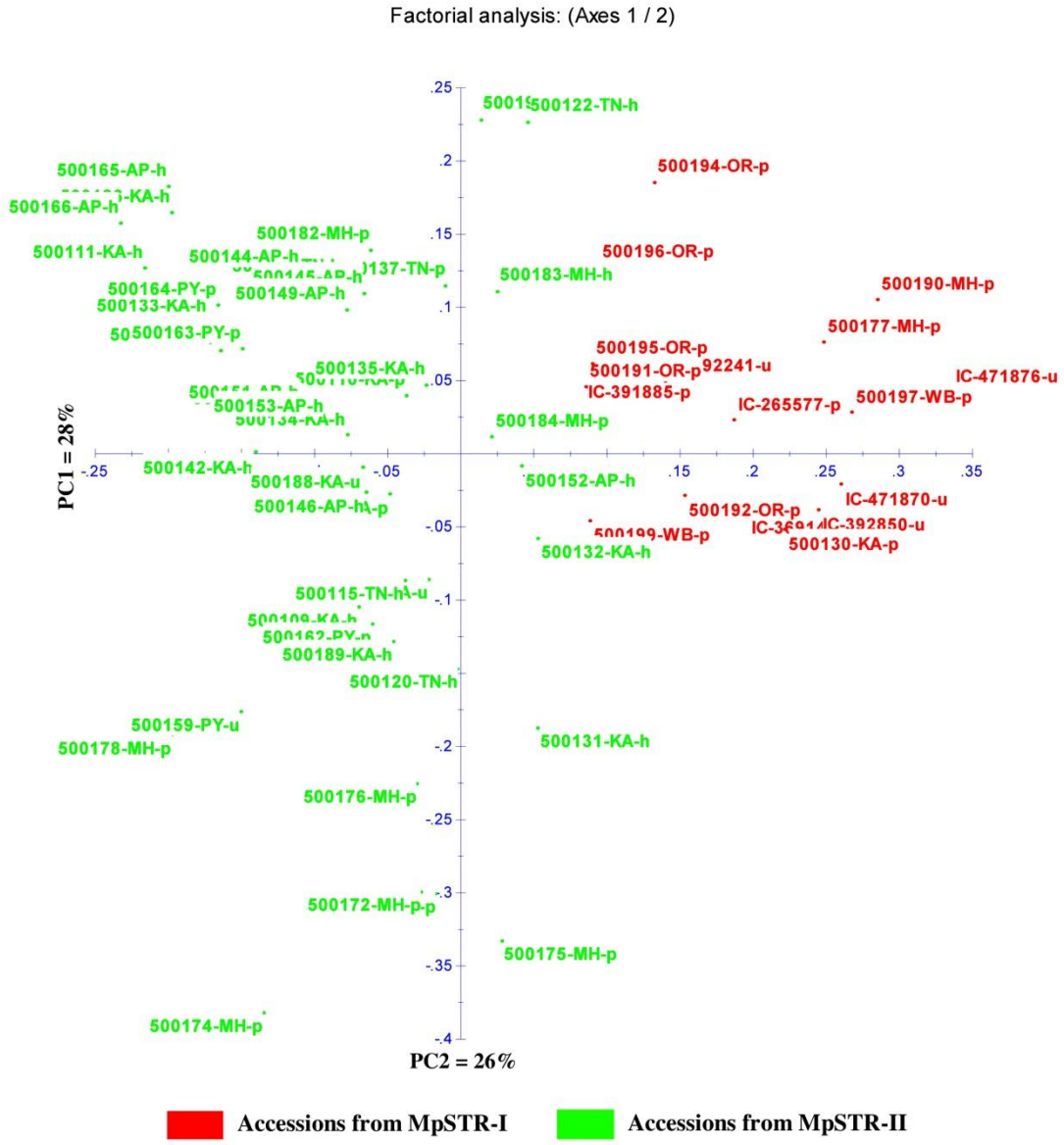


Figure 22. Scatter plot depicting dispersion of *M. pruriens* accessions (AM panel) based on principal component analysis. Red and green color indicates subgroup-1 (MpSTR-I) and subgroup-2 (MpSTR-II) from STRUCTURE analysis.

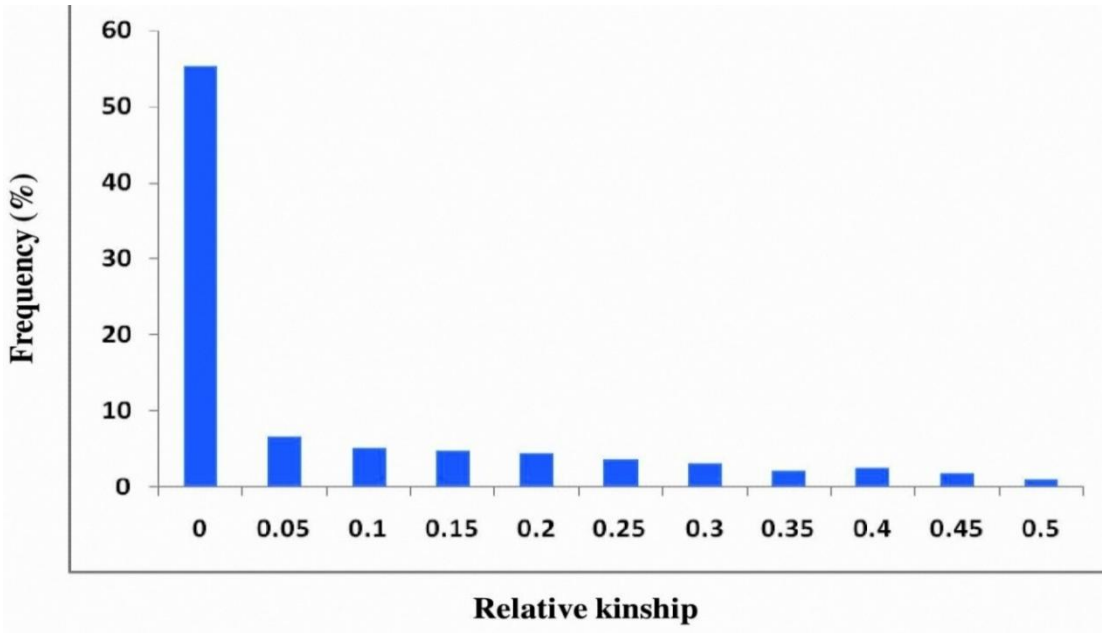


Figure 23. Distribution of global pairwise kinship coefficients (F_{ij}) of *M. pruriens* AM panel.

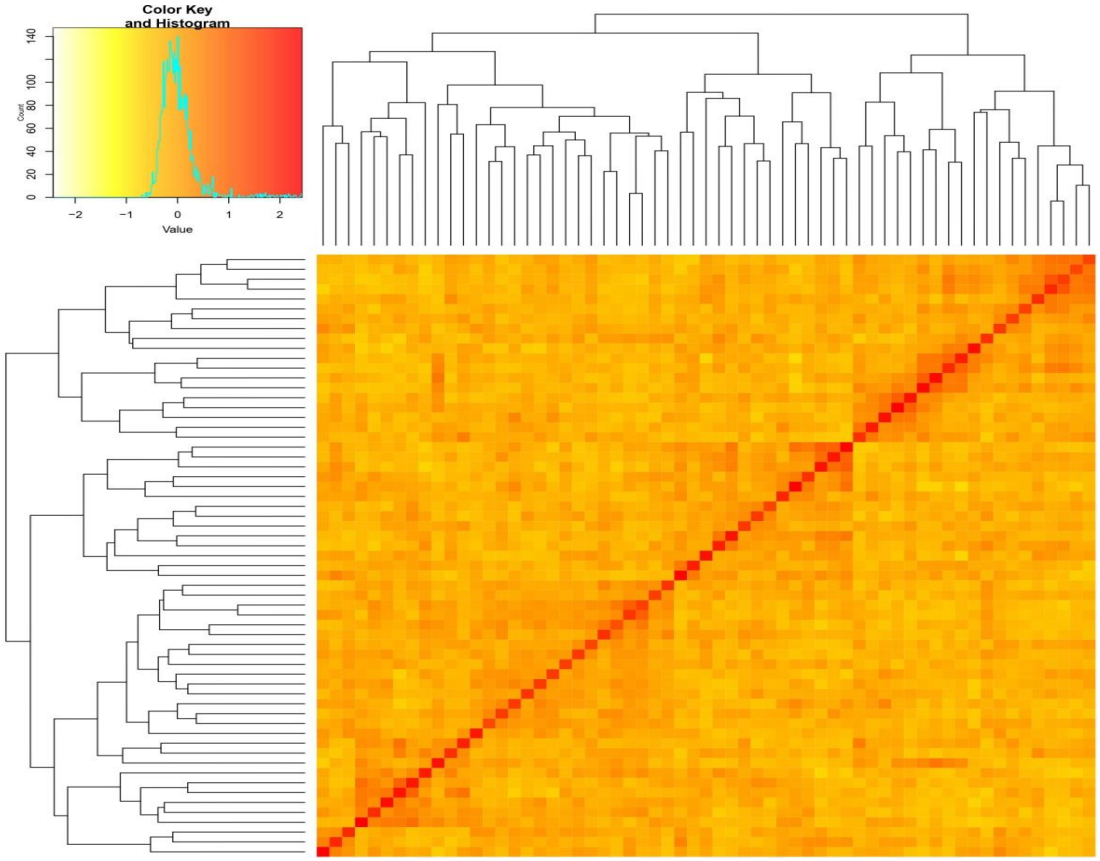


Figure 24. Heat map of kinship matrix generated for *M. pruriens* AM panel based on 180 filtered SSR markers data. The dendrogram is shown on the top and left.

5.3.3. Linkage disequilibrium

Linkage disequilibrium (LD) analysis revealed that 1.84% of SSR marker pairs were in LD ($R^2 \geq 0.1$) at $P \leq 0.001$, in which 1.76% of SSR marker pairs were in LD with $R^2 \geq 0.2$. This result was based on 7684 pair-wise comparisons generated by 180 SSR alleles. The average value of R^2 (0.032) revealed that there is no high LD among the 7684 SSR pairs in the tested AM panel. Further, LD between each pair of SSR markers located on the same chromosome was tested (Table 25). Among the pseudo-chromosomes, overall LD ranged from 0.037 to 0.091 with an average of 0.052. The $P \leq 0.001$ was used as a level of significance for testing LD between each SSR pair located on the same chromosome. Based on the average of significant LD within the chromosome, chromosomes 3, 4, 8, and 11 were found to possess the highest significant LD with $R^2 \geq 0.5$, followed by chromosomes 1 and 5 ($R^2 \geq 0.4$), chromosomes 2, 6, 7, and 9 ($R^2 \geq 0.3$) and chromosome 10 ($R^2 \geq 0.2$). Likewise, the highest average non-significant LD was found in chromosomes 4, 8, and 10 ($R^2 \geq 0.3$), followed by chromosomes 1, 2, 3, 5, 6, 7, 9, and 11 ($R^2 \geq 0.2$). The highest ratio between the number of significant LD and the number of non-significant LD was recorded in chromosome 3 (0.15), followed by chromosome 9 (0.07), chromosomes 1, 8, and 10 (0.06), chromosomes 2, 4, 6, and 11 (0.05) and chromosomes 5 and 7 (0.04). The LD decay plot is depicted in Figure 25 where R^2 between each pair of markers was plotted against genetic distance (kb).

Table 25. Linkage disequilibrium between SSR markers located on the same chromosome.

Chr. No.	Avg.R^2	No. of Sig.R^2	Avg. Sig. R^2	% sig. R^2	No. non sig. R^2	Avg. non sig.R^2	No. sig.R^2/ No. non sig.R^2
1	0.049	10	0.458	5.85	161	0.024	0.06
2	0.044	7	0.399	5.15	129	0.025	0.05
3	0.091	14	0.502	13.33	91	0.028	0.15
4	0.055	6	0.503	4.41	130	0.034	0.05
5	0.041	5	0.414	3.68	131	0.027	0.04
6	0.046	7	0.370	5.15	129	0.029	0.05
7	0.037	11	0.354	3.38	314	0.026	0.04
8	0.059	3	0.539	5.45	52	0.031	0.06
9	0.051	6	0.369	6.59	85	0.029	0.07
10	0.046	4	0.280	6.06	62	0.031	0.06
11	0.052	5	0.551	4.76	100	0.027	0.05
Average	0.052	7.09	0.431	5.80	125.82	0.028	0.06
Genome	0.032						

Chr, chromosome; Avg, average; R^2, estimated of linkage disequilibrium; Sig, significant.

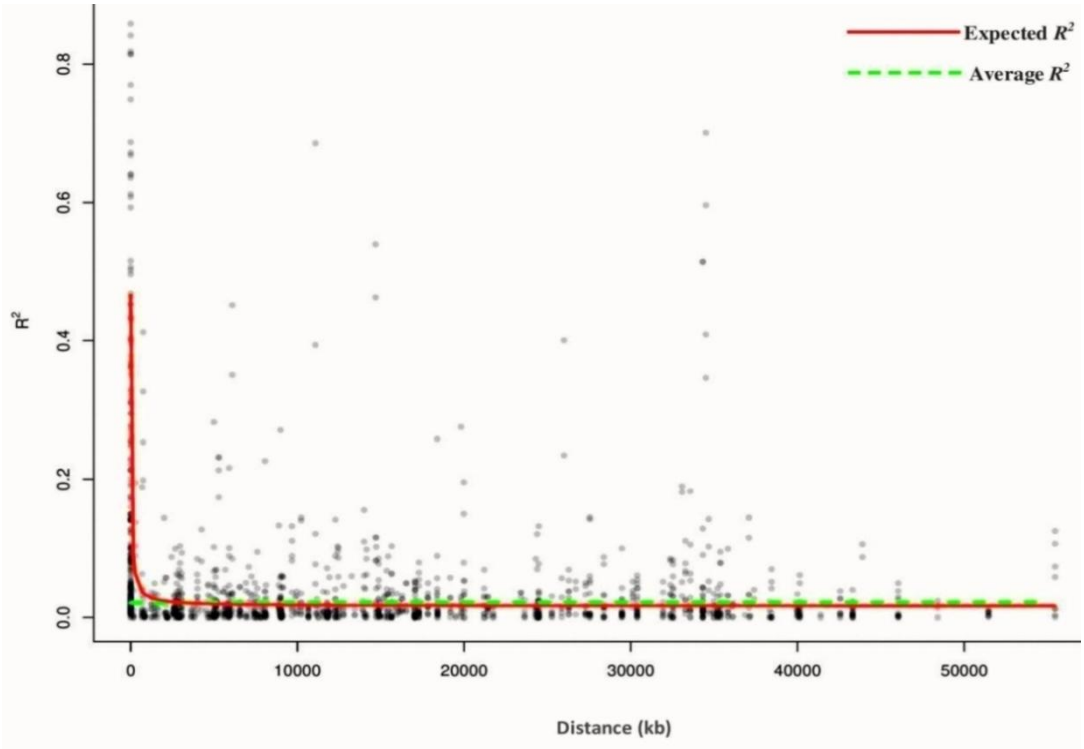


Figure 25. Linkage disequilibrium decay plot of *M. pruriens* accessions where R^2 between each pair of markers was plotted against genetic distance (kb = kilobases).

5.4. Association analysis

5.4.1. Phenotypic variability

The one-way ANOVA revealed significant variability ($P < 0.05$) for the majority of the phenotypic traits evaluated, except for pedicel length, and L-DOPA content (Table 26). Inflorescence length and flower length showed the highest (CV =84.97%) and the lowest variations (CV = 6.90%) (Table 27). Variability of the entire AM panel is provided in Table 28. We observed moderate to high estimates (42.86% to 99.93%) of broad-sense heritability (H^2) for the 16 phenotypic traits (Table 27). Traits such as number of pods per cluster, number of pods per plant, hundred seed weight, seed yield per plant, etc., showed very high heritability values ranging from 93.34 to 99.93%. Among the biochemical traits, total protein (98.58%) and total carbohydrate (95.71%) recorded high heritability, and L-DOPA content revealed a moderately high heritability (62.50%).

Table 26. One-way ANOVA for sixteen phenotypic traits in *M. pruriens* AM panel.

Trait	F	F crit	*P*-value	Sig
Inflorescence length	6.17923	3.15593	0.0036918	**
Flower buds per inflorescence	6.07893	3.15593	0.00401051	**
Pedicel length	0.1835	3.15593	0.83282943	ns
Flower length	1.44912	3.15593	0.24315107	ns
Pod length	55.7498	3.15593	3.1155E-14	***
Pod width	10.9331	3.15593	9.3518E-05	***
Number of pods per cluster	3.77094	3.15593	0.02886454	*
Number of pods per plant	5.95592	3.15593	0.0044406	**
Seed length	92.7291	3.15593	8.5725E-19	***
Seed width	85.0834	3.15593	5.6245E-18	***
Seed thickness	89.9942	3.15593	1.6569E-18	***
Seed yield per plant	8.34948	3.15593	0.00065056	***
Hundred seed weight	120.926	3.15593	2.04E-21	***
L-DOPA	0.78607	3.15593	0.46042512	ns
Total protein	3.403	3.15593	.039	*
Total carbohydrate	5.648	3.15593	.005	**

*F, calculated F value; F crit, F critical value; Sig, significance; ns, non significance at P < 0.05; ***significant at P < 0.001; **significant at P < 0.01; *significant at P < 0.05.*

Table 27. Variability and broad-sense heritability of phenotypic traits in *M. pruriens* AM panel.

Traits (unit of measurement)	Range	Average±SE	CV(%)	H^2(%)
Inflorescence length (cm)	1.60-93.00	18.85±2.05	84.97	96.93
No of flower buds per inflorescence	4.25-83.75	25.54±2.40	73.37	62.23
Flower length (cm)	4.05-5.18	4.61±0.04	6.90	52.94
Pedicel length (cm)	0.30-4.50	0.60±0.07	80.45	42.86
Pod length (cm)	6.43-14.03	8.80±0.21	18.71	99.93
Pod width (cm)	1.43-2.55	1.82±0.03	14.07	98.14
No of pods per cluster	2.50-28.00	10.83±0.77	55.71	98.83
No of pods per plant	27.25-405.75	147.43±11.66	61.77	96.77
Seed length (mm)	7.73-16.90	10.88±0.28	20.19	96.62
Seed width (mm)	6.15-12.77	8.18±0.20	19.17	93.92
Seed thickness (mm)	3.01-9.61	5.33±0.18	26.08	97.27
Seed yield per plant (gm)	41.13-1136.60	303.61±30.94	79.60	93.34
Hundred seed weight (gm)	12.83-146.36	41.71±3.75	70.31	95.93
L-DOPA (%)	0.95-3.24	2.07±0.06	22.81	62.50
Total protein (%)	15.07-29.19	22.94±0.42	14.14	98.58
Total carbohydrate (%)	6.54-28.25	14.60±0.68	36.44	95.71

*SE, standard error; CV, coefficient of variation; ANOVA, one-way analysis of variance; H^2, broad-sense heritability; ns, not significant; ***significant at P < 0.001; **significant at P < 0.01; *significant at P < 0.05.*

The values of correlation coefficients indicated a significantly high positive correlation ($R > 0.70$) between some important seed, pod, and inflorescence traits (Table 29). For instance, hundred seed weight was strongly correlated ($R > 0.70$) with seed thickness, seed width, seed length, and pod length, and moderately correlated ($0.30 < R < 0.70$) with pod width and flower length. A similar trend was observed for seed yield as well. In PCA, the first five principal components (PCs) explained 81.39% of the total phenotypic variance of which PC1 accounted for 34.15% and PC2 accounted for 24.68% (Figure 26; Table 30). The resultant scatter plot distinctly separated the cultivated and wild accessions. However, we couldn't find clear-cut separation for the two wild varieties - var. *pruriens* and var. *hirsuta* (Figure 26).

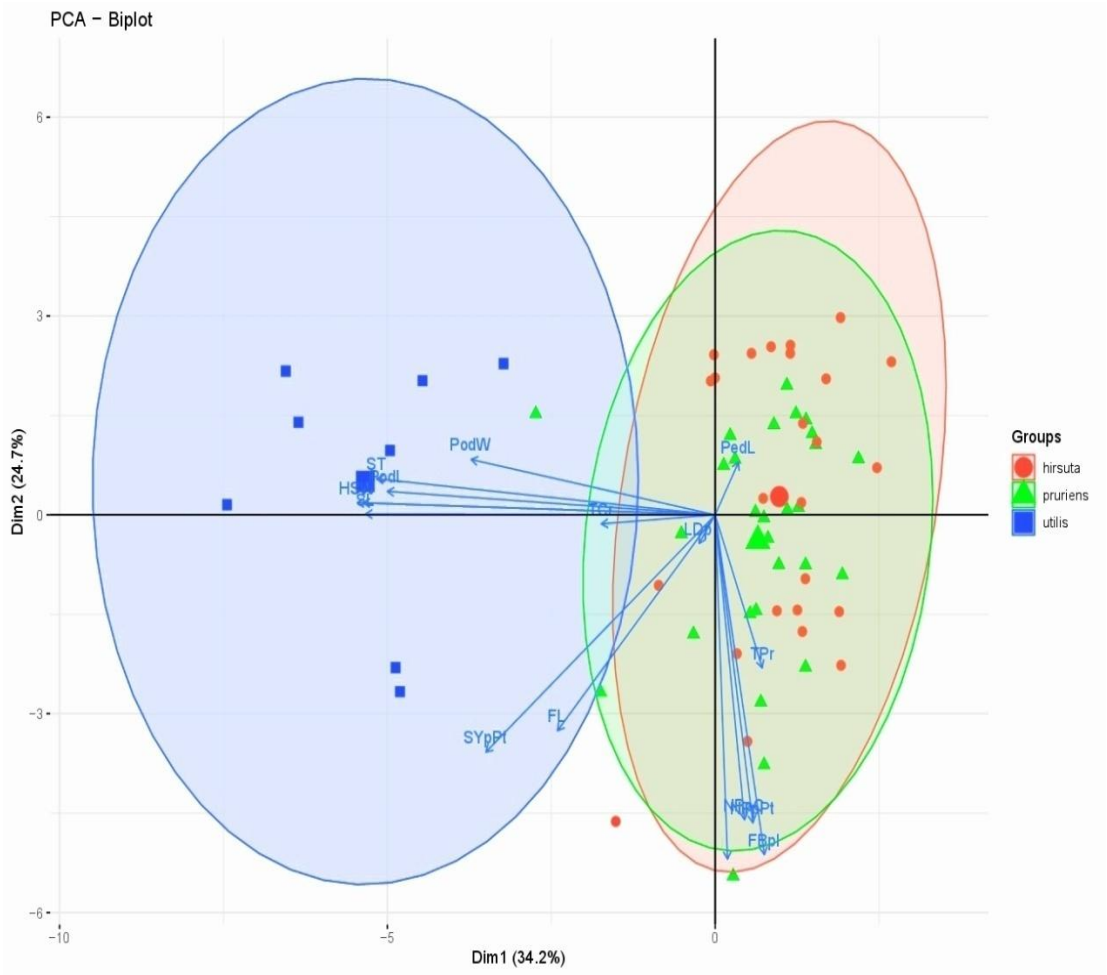


Figure 26. Scatter plot of *M. pruriens* accessions (AM panel) based on sixteen phenotypic traits from the principal component analysis (PCA). Note: red circles, green triangles, and blue squares represent var. *hirsuta*, var. *pruriens*, and var. *utilis*, respectively.

Table 28. Variability for phenotypic traits in *M. pruriens* association mapping panel.

Accessions	IL	FBpI	FL	PedL	PodL	PodW	NPpC	NPpPt	SL	SW	ST	SYpPt	HSW	LDp	TPr	TCr
IC-265577	13.00	14.00	4.67	0.48	12.40	2.35	5.75	64.25	13.92	9.25	5.64	200.77	54.30	2.40	21.72	21.04
IC-369144	4.88	8.25	4.93	0.65	13.10	2.45	6.50	130.50	15.14	12.77	9.61	1091.40	145.87	1.09	27.86	19.37
IC-391885	38.00	65.25	4.80	0.65	8.13	1.55	14.00	166.75	10.11	6.93	4.27	315.29	32.94	1.52	25.97	17.51
IC-392241	3.63	5.25	4.85	0.38	10.23	1.80	2.50	33.00	15.72	10.24	7.93	163.64	86.17	2.35	20.02	11.10
IC-392850	5.00	6.75	4.60	0.58	13.45	2.08	5.75	49.50	16.54	12.37	9.03	415.91	146.36	1.69	18.85	24.84
IC-471870	3.88	5.25	4.88	0.63	14.03	2.55	4.50	96.25	16.90	11.29	8.25	553.00	112.21	2.40	24.13	18.01
IC-471876	8.00	9.00	5.00	0.50	11.18	2.20	4.50	51.75	15.07	11.54	8.83	241.51	81.20	2.42	16.40	10.60
500108-KA	34.25	30.75	5.16	0.57	11.25	2.10	21.75	203.75	15.64	11.68	7.29	929.11	95.65	2.10	23.60	13.79
500109-KA	38.75	62.25	4.43	0.48	8.25	2.23	28.00	333.50	9.92	6.91	4.98	463.64	26.44	2.32	21.66	22.71
500110-KA	11.75	18.00	4.20	0.30	9.50	2.05	10.25	116.00	9.78	7.77	4.93	194.19	33.45	2.10	21.23	20.74
500111-KA	31.25	54.00	4.80	0.55	7.30	1.78	15.75	198.50	9.22	7.37	5.07	262.56	27.96	2.08	22.37	10.86
500112-KA	41.00	35.50	4.85	0.58	8.83	2.08	23.50	277.50	11.41	8.39	5.53	579.45	84.83	2.15	20.36	28.25
500113-MH	93.00	83.75	4.68	0.70	8.70	2.15	15.00	343.25	10.69	8.03	3.78	544.06	28.84	2.76	28.05	22.38
500115-TN	27.50	20.25	4.98	0.53	8.48	1.58	18.50	150.25	9.27	6.62	3.88	239.80	35.13	1.57	21.99	16.67
500120-TN	2.50	6.50	4.15	0.50	8.13	1.83	4.50	27.25	10.11	7.70	4.63	41.13	30.26	2.01	25.73	16.46
500121-TN	25.25	36.00	5.15	0.58	8.25	1.85	8.75	162.50	13.00	9.24	5.15	316.41	40.99	2.95	24.77	16.89
500122-TN	4.50	6.25	4.15	4.50	8.33	1.80	4.25	35.25	9.46	7.02	4.22	48.42	24.05	2.03	24.28	12.28
500126-KA	28.00	42.00	4.90	0.48	8.15	1.68	14.75	283.75	8.20	7.13	4.11	293.10	19.76	1.98	28.55	9.78
500130-KA	27.63	39.25	4.15	0.63	7.90	1.43	19.50	150.25	9.91	8.62	5.57	287.85	38.39	1.48	21.81	18.26
500131-KA	11.50	11.75	4.45	0.35	9.13	1.63	7.00	214.50	7.73	6.50	3.01	136.73	12.83	1.47	23.08	13.73
500132-KA	32.00	36.75	5.05	0.45	7.65	1.55	12.50	212.50	9.30	6.24	4.43	236.99	22.38	1.48	22.36	10.30
500133-KA	4.85	5.75	4.45	0.61	8.18	1.63	5.25	136.25	9.80	7.68	4.73	199.40	29.60	1.47	25.78	8.28

500134-KA	6.50	7.75	4.25	0.61	10.18	2.20	4.50	157.50	8.96	7.37	5.93	214.65	28.86	3.24	17.35	13.52
500135-KA	2.88	7.50	4.05	0.41	6.93	2.20	5.00	73.00	9.23	7.70	4.11	94.43	24.45	2.10	28.76	7.50
500137-TN	29.12	30.75	4.80	0.60	8.80	1.78	9.25	124.50	11.87	9.97	5.49	264.58	40.46	1.94	22.79	7.82
500138-TN	7.80	8.30	4.10	0.50	8.60	1.80	5.50	143.50	9.50	7.00	4.50	210.70	28.90	1.94	17.78	20.86
500142-KA	6.88	7.25	4.30	0.48	7.78	1.63	7.00	218.50	9.69	7.43	4.14	255.24	23.49	1.46	20.83	16.65
500144-AP	10.55	10.00	4.08	0.35	8.30	1.80	9.75	305.50	9.09	7.59	5.39	465.63	32.23	2.11	21.71	18.83
500145-AP	10.00	13.25	4.15	0.35	8.40	1.73	11.00	284.25	8.31	7.18	4.86	386.34	27.08	2.03	21.94	18.84
500146-AP	3.50	4.50	4.55	0.44	8.80	1.95	4.00	55.50	11.01	8.41	4.92	94.40	29.61	1.68	19.23	19.65
500148-AP	2.73	8.00	4.38	0.50	8.70	1.73	4.25	80.00	10.81	8.91	5.03	182.26	42.80	2.31	20.94	18.36
500149-AP	2.13	4.25	4.15	0.60	6.58	2.20	3.50	44.75	9.50	7.81	4.09	58.72	25.15	2.50	24.48	24.28
500151-AP	1.60	7.25	4.18	0.53	8.38	1.50	5.50	49.75	7.99	6.15	3.81	49.73	17.52	1.73	22.85	13.04
500152-AP	2.00	6.25	4.45	0.55	7.73	1.70	3.75	35.50	10.66	7.61	5.89	65.45	35.15	1.93	21.57	19.02
500153-AP	41.75	60.75	5.15	0.65	9.00	1.68	14.00	375.50	11.98	8.62	5.66	1136.60	48.33	3.02	25.70	21.39
500159-PY	38.25	40.25	5.18	0.64	11.98	2.08	14.25	165.75	14.91	11.73	7.04	884.89	96.44	2.15	23.71	18.44
500162-PY	9.75	16.50	4.70	0.70	9.13	1.80	11.00	82.50	10.72	7.51	4.98	158.48	32.00	2.24	22.98	19.37
500163-PY	10.00	13.50	4.45	0.60	8.63	1.78	7.50	117.25	8.89	6.73	4.38	178.18	29.00	2.10	18.54	12.76
500164-PY	30.00	42.00	5.10	0.48	7.45	1.68	7.25	77.00	10.21	7.46	4.20	146.22	26.63	2.25	26.41	9.04
500165-AP	2.25	7.25	4.15	0.60	6.43	2.18	5.50	33.50	9.69	7.84	5.14	64.16	31.88	2.28	23.10	20.87
500166-AP	29.13	46.50	4.60	0.45	8.50	1.63	21.25	133.75	9.95	7.64	5.57	264.20	34.35	1.80	23.10	9.22
500172-MH	16.50	29.25	4.78	0.63	7.75	1.75	11.50	140.00	10.82	7.41	4.36	150.04	22.78	2.36	20.31	14.21
500174-MH	18.12	33.75	4.25	0.58	7.88	1.62	6.25	74.00	10.88	7.90	4.79	133.04	31.29	2.52	15.07	19.19
500175-MH	31.75	51.00	4.75	0.43	8.98	1.73	10.50	178.00	10.75	7.75	5.13	267.77	26.14	2.22	25.88	12.00
500176-MH	37.00	61.50	4.73	0.65	9.13	1.65	26.50	195.75	10.77	7.47	4.79	327.65	29.11	2.99	20.79	15.23
500177-MH	23.00	32.25	4.60	0.50	8.38	1.78	15.25	281.50	11.72	9.21	5.13	606.52	39.25	2.30	25.73	12.93
500178-MH	29.00	33.75	4.70	0.50	7.88	1.63	15.75	150.75	10.39	8.12	5.47	338.25	39.00	2.67	29.11	16.15

500182-MH	12.38	17.25	4.50	0.58	8.40	1.75	13.75	143.25	10.16	7.39	4.69	203.77	25.95	1.54	27.90	10.14
500183-MH	36.00	39.75	4.93	0.53	8.00	1.58	16.00	214.50	11.71	8.42	5.16	410.29	33.28	1.62	24.71	11.68
500184-MH	38.25	47.75	4.88	0.70	7.63	1.53	14.75	405.75	10.65	7.44	5.70	727.43	32.65	1.62	29.19	7.94
500188-KA	12.05	17.00	4.80	0.58	12.35	1.95	9.25	75.00	15.59	11.35	9.30	476.88	105.97	1.72	20.07	10.30
500189-KA	37.25	36.00	4.88	0.60	8.70	1.65	18.75	165.75	9.20	6.52	5.08	279.50	29.35	1.42	19.77	9.74
500190-MH	18.38	30.00	4.78	0.48	9.13	1.75	15.00	163.75	8.47	6.87	3.72	165.64	17.59	1.98	29.12	7.60
500191-OR	18.38	32.75	4.55	0.58	8.13	1.78	10.50	124.50	10.46	7.84	5.66	234.79	32.87	2.63	23.17	9.56
500192-OR	6.25	11.00	4.45	0.48	7.70	1.60	8.00	71.75	10.83	7.69	4.48	135.70	32.98	2.82	21.45	9.04
500193-OR	12.25	24.75	4.28	0.45	8.00	1.68	7.25	76.50	10.21	6.86	5.17	120.91	27.59	2.27	22.68	9.28
500194-OR	9.35	17.00	4.70	0.65	8.13	1.50	7.00	84.00	9.49	7.08	4.92	130.39	27.22	2.15	22.75	9.56
500195-OR	20.25	30.75	4.80	0.58	7.60	1.55	14.00	116.75	10.69	8.33	5.37	218.72	32.50	0.95	18.75	11.28
500196-OR	10.88	23.00	4.55	0.55	8.18	1.68	12.25	74.25	11.08	8.76	6.05	236.84	41.00	1.67	22.00	8.46
500197-WB	9.20	19.50	4.78	0.45	6.55	1.53	6.75	98.50	9.36	6.63	4.71	120.56	21.32	2.12	22.68	6.54
500199-WB	16.50	25.50	4.55	0.60	7.30	1.75	15.00	164.75	10.62	8.10	5.62	306.48	32.35	2.24	24.12	6.64
Minimum	1.60	4.25	4.05	0.30	6.43	1.43	2.50	27.25	7.73	6.15	3.01	41.13	12.83	0.95	15.07	6.54
Maximum	93.00	83.75	5.18	4.50	14.03	2.55	28.00	405.75	16.90	12.77	9.61	1136.60	146.36	3.24	29.19	28.25
Average	18.85	25.54	4.61	0.60	8.79	1.82	10.83	147.43	10.88	8.18	5.33	303.61	41.71	2.07	22.94	14.60
SE	2.05	2.39	0.04	0.06	0.21	0.03	0.77	11.66	0.28	0.20	0.18	30.94	3.75	0.06	0.42	0.68
CV%	84.97	73.37	6.89	85.46	18.71	14.07	55.72	61.77	20.19	19.17	26.08	79.60	70.32	22.81	14.138	36.44

IL, inflorescence length (cm); FBpI, flower buds per inflorescence; FL, flower length (cm); PedL, pedicel length (cm); PodL, pod length (cm); PodW, pod width (cm); NPpC, number of pods per cluster; NPpPt, number of pods per plant; SL, seed length (mm); SW, seed width (mm); ST, seed thickness (mm); SYpPt, seed yield per plant (gm); HSW, hundred seed weight (gm); LDp, L-DOPA content (%); TPr, total protein (%); TCr, total carbohydrate; SE, standard error; CV, coefficient of variation.

Table 29. Correlation matrix of phenotypic traits evaluated on the *M. pruriens* AM panel.

	IL	FBpI	FL	PedL	PodL	PodW	NPpC	NPpPt	SL	SW	ST	SYpPt	HSW	LDp	TPr	TCr
IL		**	**	ns	ns	ns	**	**	ns	ns	ns	**	ns	ns	*	ns
FBpI	0.92		**	ns	ns	ns	**	**	ns	ns	ns	**	ns	ns	*	ns
FL	0.50	0.46		ns	**	ns	**	*	**	**	**	**	**	ns	ns	ns
PedL	-0.06	-0.08	-0.13		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
PodL	-0.08	-0.18	0.33	-0.01		**	ns	ns	**	**	**	**	**	ns	ns	*
PodW	-0.10	-0.21	0.00	0.00	0.60		ns	ns	**	**	**	*	**	*	ns	**
NPpC	0.70	0.73	0.35	-0.11	-0.10	-0.18		**	ns	ns	ns	**	ns	ns	ns	ns
NPpPt	0.67	0.63	0.25	-0.14	-0.11	-0.11	0.62		ns	ns	ns	**	ns	ns	*	ns
SL	-0.01	-0.08	0.49	-0.04	0.79	0.53	-0.10	-0.18		**	**	**	**	ns	ns	ns
SW	-0.05	-0.14	0.38	-0.06	0.77	0.57	-0.11	-0.16	0.93		**	**	**	ns	ns	ns
ST	-0.15	-0.19	0.34	-0.07	0.76	0.49	-0.10	-0.18	0.87	0.88		**	**	ns	ns	ns
SYpPt	0.46	0.39	0.51	-0.08	0.48	0.30	0.41	0.61	0.51	0.56	0.50		**	ns	*	ns
HSW	-0.07	-0.17	0.38	-0.04	0.84	0.56	-0.07	-0.14	0.89	0.91	0.90	0.59		ns	ns	*
LDp	0.13	0.17	-0.01	0.00	-0.01	0.27	-0.04	0.05	0.09	0.01	-0.03	0.02	-0.08		ns	ns
TPr	0.29	0.29	0.19	0.07	-0.16	-0.02	0.16	0.31	-0.12	-0.07	-0.21	0.26	-0.12	-0.05		ns
TCr	0.09	-0.01	-0.17	-0.04	0.26	0.45	0.03	0.11	0.19	0.20	0.09	0.25	0.30	0.17	-0.22	

*IL, inflorescence length; FBpI, flower buds per inflorescence; FL, flower length; PedL, pedicel length; PodL, pod length; PodW, pod width; NPpC, number of pods per cluster; NPpPt, number of pods per plant; SL, seed length; SW, seed width, ST, seed thickness; SYpPt, seed yield per plant; HSW, hundred seed weight; LDp, L-DOPA; TPr, total protein; TCr, total carbohydrate; ns, non significance; **significant at P < 0.01; *significant at P < 0.05. Note: Below diagonal elements represent the Pearson's correlation coefficients and the above diagonal represent the significance values.*

Table 30. Eigen vector and Eigen values of the first five principal components (PCs) for phenotypic traits.

Traits (Abbreviation)	PC1	PC2	PC3	PC4	PC5
Hundred seed weight (HSW)	0.963	-0.032	-0.076	-0.023	-0.110
Seed width (SW)	0.947	-0.032	-0.097	0.032	0.032
Seed length (SL)	0.941	-0.002	-0.103	0.017	0.178
Seed thickness (ST)	0.911	-0.096	-0.176	-0.080	0.036
Pod length (PodL)	0.883	-0.062	0.004	0.004	-0.057
Pod width (PodW)	0.657	-0.147	0.479	0.220	-0.033
Inflorescence length (IL)	-0.034	0.917	0.103	0.020	0.098
Flower buds per inflorescence (FBpI)	-0.133	0.905	0.051	-0.004	0.190
No. of pods per plant (NPpPt)	-0.102	0.820	0.179	-0.039	-0.260
No. of pods per cluster (NPpC)	-0.079	0.812	0.010	-0.187	-0.099
Seed yield per plant (SYpPt)	0.617	0.632	0.032	0.077	-0.243
Flower length (FL)	0.424	0.575	-0.420	-0.017	0.315
Total carbohydrate (TCr)	0.307	0.024	0.776	-0.135	-0.334
Pedicel length (PedL)	-0.063	-0.141	-0.003	0.752	-0.127
Total protein (TPr)	-0.127	0.408	-0.222	0.610	-0.163
L-DOPA (LDp)	0.043	0.076	0.596	0.206	0.714
Eigenvalue	5.464	3.950	1.515	1.098	0.996
% of Variance	34.152	24.686	9.468	6.861	6.228
Cumulative %	34.152	58.838	68.306	75.167	81.395

5.4.2. Genetic diversity

The summary of the marker attributes, as well as individual markers, are given in Tables 31 and Table 32. Initially, 66 markers produced 232 alleles with an average of 3.52 alleles per locus. Stringent filtering criterion was applied that retain 180 SSR alleles based on minor allele frequency > 0.05, along with sixty-one accessions which showed maximum missing site < 20% for further analysis. The major allele frequency (MAF) varied from 0.50 to 0.94, with an average of 0.73 indicating high genetic diversity and polymorphism at the observed loci. The heterozygosity (Ho) values ranged from 0.00 to 0.59 with a mean value of 0.12 and the gene diversity (He) varied from 0.03 (MPU_122) to 0.73 (MPU_42) with a mean value of 0.36.

Table 31. Marker attributes of SSR markers used for association analysis.

Major allele frequency (MAF)	
Average	0.73
SSR marker with lowest MAF	0.5 (MPU_52)
SSR marker with highest MAF	0.95 (MPU_20, MPU_122)
Observed heterozygosity (Ho)	
Average	0.12
SSR marker with lowest Ho	0.00 (MPU_39, MPU_47, MPU_80)
SSR marker with highest Ho	0.59 (MPU_90)
Expected heterozygosity (He)	
Average	0.36
SSR marker with lowest He	0.03 (MPU_122)
SSR marker with highest He	0.73 (MPU_42)
Polymorphism Information Content (PIC)	
Average	0.29
SSR marker with the lowest PIC	0.09 (MPU_20)
SSR marker with the highest PIC	0.38 (MPU_52)

The PIC, which provides the relative informativeness of each marker, ranged from 0.09 (MPU_22) to 0.38 (MPU_52) with an average of 0.29 (Table 31 and 32). The major proportion of SSR markers (43) were moderately polymorphic (PIC-value > 0.25) with PIC-values ranging from 0.26 to 0.38 (Table 32). Based on the marker statistics, the SSR marker MPU_52 was highly informative. The mean and private allelic richness per marker were 2.82 and 0.06 respectively. The majority of the SSR markers deviated from the Hardy-Weinberg equilibrium (HWE) at a significance level of $P < 0.001$ and very few markers deviated at $P < 0.01$ (MPU_46 and MPU_115) and $P < 0.05$ (MPU_92) (Table 32). The SSR markers MPU_122 and MPU_130 showed no significant deviation.

Table 32. Details of marker attributes and diversity of individual SSR markers used for the association mapping.

SSR Markers	Putative chromosome location of marker	Na	Ne	MAF	PIC	Ho	He	I	h	H_T	HWE
MPU_16	Chr01	4	1.83	0.61	0.36	0.16	0.39	0.74	0.43	0.45	***
MPU_42	Chr01	7	3.83	0.69	0.34	0.08	0.73	1.49	0.24	0.32	***
MPU_48	Chr01	6	2.01	0.57	0.37	0.22	0.50	0.85	0.48	0.50	***
MPU_114	Chr01	2	1.27	0.91	0.14	0.02	0.19	0.33	0.22	0.25	***
MPU_115	Chr01	2	1.23	0.81	0.26	0.13	0.19	0.33	0.30	0.30	**
MPU_130	Chr01	2	1.16	0.56	0.37	0.14	0.12	0.21	0.37	0.50	ns
MPU_23	Chr02	5	1.83	0.71	0.33	0.07	0.42	0.82	0.34	0.36	***
MPU_37	Chr02	3	1.42	0.79	0.28	0.06	0.29	0.48	0.30	0.30	***
MPU_51	Chr02	3	2.05	0.65	0.35	0.18	0.49	0.82	0.35	0.44	***
MPU_52	Chr02	4	1.76	0.50	0.38	0.35	0.43	0.77	0.44	0.49	***
MPU_88	Chr02	4	1.67	0.63	0.36	0.10	0.38	0.61	0.36	0.41	***
MPU_89	Chr02	3	1.77	0.71	0.32	0.07	0.42	0.63	0.43	0.45	***
MPU_46	Chr03	2	1.74	0.59	0.37	0.57	0.42	0.61	0.48	0.48	**
MPU_47	Chr03	2	1.39	0.90	0.17	0.00	0.25	0.40	0.25	0.27	***
MPU_57	Chr03	4	1.64	0.69	0.34	0.07	0.37	0.65	0.35	0.37	***
MPU_64	Chr03	3	2.69	0.68	0.33	0.03	0.62	1.03	0.42	0.42	***
MPU_96	Chr03	2	1.47	0.71	0.33	0.17	0.29	0.45	0.34	0.36	***
MPU_98	Chr03	3	1.37	0.90	0.17	0.02	0.23	0.39	0.26	0.30	***
MPU_06	Chr04	2	1.24	0.83	0.24	0.08	0.19	0.33	0.26	0.26	***
MPU_14	Chr04	2	1.97	0.56	0.37	0.07	0.49	0.68	0.49	0.49	***
MPU_18	Chr04	3	1.40	0.88	0.19	0.04	0.19	0.32	0.13	0.14	***
MPU_49	Chr04	4	1.78	0.54	0.37	0.03	0.42	0.64	0.31	0.43	***
MPU_102	Chr04	3	1.12	0.92	0.13	0.03	0.09	0.17	0.18	0.18	***

MPU_111	Chr04	5	2.32	0.51	0.37	0.22	0.56	1.05	0.48	0.49	***
MPU_125	Chr04	4	1.18	0.87	0.20	0.03	0.14	0.26	0.19	0.19	***
MPU_54	Chr05	6	2.95	0.64	0.35	0.29	0.66	1.24	0.44	0.44	***
MPU_62	Chr05	2	1.83	0.59	0.37	0.16	0.45	0.64	0.41	0.46	***
MPU_72	Chr05	3	1.62	0.70	0.33	0.08	0.32	0.48	0.23	0.31	***
MPU_80	Chr05	2	1.23	0.91	0.15	0.00	0.17	0.30	0.16	0.16	***
MPU_83	Chr05	2	1.10	0.93	0.13	0.01	0.09	0.17	0.09	0.09	***
MPU_92	Chr05	5	1.53	0.74	0.31	0.21	0.32	0.62	0.40	0.41	*
MPU_22	Chr06	5	2.65	0.69	0.33	0.04	0.61	1.05	0.40	0.42	***
MPU_68	Chr06	2	1.53	0.74	0.31	0.11	0.33	0.50	0.39	0.39	***
MPU_78	Chr06	3	2.04	0.58	0.37	0.05	0.51	0.78	0.47	0.48	***
MPU_81	Chr06	3	1.12	0.88	0.19	0.02	0.10	0.19	0.18	0.19	***
MPU_94	Chr06	3	2.39	0.68	0.34	0.25	0.57	0.96	0.34	0.42	***
MPU_123	Chr06	5	1.32	0.78	0.28	0.16	0.24	0.50	0.33	0.33	***
MPU_19	Chr07	5	2.25	0.53	0.37	0.22	0.51	0.92	0.43	0.49	***
MPU_45	Chr07	5	1.98	0.67	0.34	0.19	0.39	0.69	0.25	0.44	***
MPU_60	Chr07	4	3.04	0.73	0.31	0.26	0.66	1.19	0.32	0.36	***
MPU_87	Chr07	4	2.94	0.54	0.37	0.08	0.66	1.20	0.49	0.50	***
MPU_100	Chr07	3	1.36	0.81	0.26	0.03	0.26	0.45	0.34	0.34	***
MPU_101	Chr07	4	1.87	0.61	0.36	0.19	0.45	0.76	0.47	0.47	***
MPU_124	Chr07	7	2.79	0.74	0.30	0.19	0.64	1.20	0.29	0.31	***
MPU_20	Chr08	2	1.16	0.94	0.09	0.05	0.11	0.19	0.07	0.07	***
MPU_39	Chr08	4	1.35	0.88	0.18	0.00	0.24	0.44	0.26	0.27	***
MPU_40	Chr08	3	1.84	0.68	0.34	0.08	0.44	0.69	0.43	0.47	***
MPU_58	Chr08	2	1.42	0.74	0.31	0.03	0.28	0.44	0.32	0.34	***
MPU_65	Chr08	4	2.75	0.71	0.31	0.25	0.62	1.08	0.35	0.37	***

MPU_108	Chr08	3	1.04	0.93	0.12	0.02	0.04	0.09	0.08	0.08	***
MPU_27	Chr09	5	1.09	0.94	0.10	0.04	0.08	0.17	0.07	0.07	***
MPU_31	Chr09	6	2.13	0.62	0.35	0.17	0.52	0.87	0.36	0.40	***
MPU_43	Chr09	3	1.55	0.71	0.32	0.07	0.33	0.52	0.33	0.35	***
MPU_59	Chr09	2	1.37	0.74	0.31	0.10	0.26	0.42	0.28	0.31	***
MPU_73	Chr09	3	1.50	0.78	0.29	0.13	0.31	0.51	0.31	0.32	***
MPU_122	Chr09	2	1.03	0.94	0.10	0.03	0.03	0.06	0.07	0.07	ns
MPU_15	Chr10	4	2.13	0.64	0.33	0.07	0.51	0.91	0.44	0.47	***
MPU_30	Chr10	4	1.20	0.84	0.23	0.05	0.16	0.33	0.21	0.22	***
MPU_99	Chr10	3	1.32	0.82	0.25	0.11	0.23	0.44	0.33	0.34	***
MPU_107	Chr10	5	2.99	0.67	0.33	0.14	0.64	1.19	0.41	0.42	***
MPU_24	Chr11	6	1.98	0.73	0.31	0.05	0.48	0.84	0.40	0.40	***
MPU_63	Chr11	2	1.85	0.53	0.37	0.09	0.46	0.65	0.46	0.49	***
MPU_85	Chr11	3	1.43	0.83	0.24	0.11	0.24	0.43	0.23	0.24	***
MPU_90	Chr11	4	3.29	0.75	0.30	0.59	0.69	1.27	0.22	0.27	***
MPU_106	Chr11	2	1.32	0.87	0.20	0.08	0.17	0.25	0.14	0.16	***
MPU_121	Chr11	3	1.27	0.83	0.24	0.02	0.19	0.32	0.29	0.29	***
Minimum		2	1.03	0.50	0.09	0.00	0.03	0.06	0.07	0.07	-
Maximum		7	3.83	0.94	0.38	0.59	0.73	1.49	0.49	0.50	-
Average		3.52	1.78	0.73	0.29	0.12	0.36	0.62	0.32	0.34	-
SE		0.17	0.08	0.02	0.01	0.01	0.02	0.04	0.01	0.02	-

*Chr, chromosome; Na, total number of alleles; Ne, effective number of alleles; MAF, Major allele frequency; Ho, observed heterozygosity; He, expected heterozygosity; I, Shannon's information index; h, Nei's gene diversity; H_T, total gene diversity; HWE, Hardy-Weinberg Equilibrium; ns, non significance; ***significant at P < 0.001, **Significant at P < 0.01, *Significant at P < 0.05*

Shannon's information index (I = 0.62) and Nei's gene diversity (h = 0.32) suggest high genetic diversity within our AM panel (Table 32). Total gene diversity value (H_T = 0.34) (Table 32) substantiated this finding. For more perceptive analyses, we subdivided the AM panel accessions into two groups: the first one based on three regional gene pools (Eastern, Central-West, and Southern India) and the second one on three botanical varieties (var. *utilis*, var. *pruriens*, and var. *hirsuta*). We estimated genetic diversity statistics independently for each one of these gene pools (Table 33). In the case of regional gene pools, I varied from 0.56 to 0.67, and h varied from 0.28 to 0.34. Accessions from West-Central India (I = 0.67, h = 0.34) and Southern India (I = 0.64, h = 0.34) were more diverse as compared to the ones from Eastern India (I = 0.0.56, h = 0.31). The percentage of polymorphic SSR loci was 96.97% for South Indian accessions followed by 89.39% and 87.88% for West-Central and Eastern Indian accessions, respectively. Among the three varietal gene pools, I varied from 0.44 to 0.72, and h varied from 0.28 to 0.36 (Table 33). Wild varieties, var. *pruriens* (I = 0.72, h = 0.36) and var. *hirsuta* (I = 0.62, h = 0.34) were genetically more diverse than the cultivated var. *utilis* (I = 0.44, h = 0.28). The percentage of polymorphic SSR loci were recorded as: 98.48%, 96.97%, and 72.73% in var. *pruriens*, var. *hirsuta*, and var. *utilis*.

Table 33. Genetic diversity estimates for the regional gene pools and botanical varieties within *M. pruriens* association mapping panel.

Regional gene pool	**Na**	**Ne**	**I**	**h**	**Ho**	**He**	**%P**
Eastern India	2.48	1.65	0.56	0.31	0.09	0.33	87.88
West-Central India	2.56	1.95	0.67	0.34	0.13	0.40	89.39
Southern India	3.15	1.75	0.64	0.34	0.13	0.35	96.97
Varietal gene pool	**Na**	**Ne**	**I**	**h**	**Ho**	**He**	**%P**
M. pruriens var. *utilis*	1.98	1.53	0.44	0.28	0.06	0.27	72.73
M. pruriens var. *pruriens*	3.11	1.92	0.72	0.36	0.12	0.41	98.48
M. pruriens var. *hirsuta*	3.00	1.72	0.62	0.34	0.14	0.35	96.97

Na, total no. of alleles; Ne, effective no. of alleles; I, Shannon's information index; h, Nei's gene diversity; Ho, observed heterozygosity; He, expected heterozygosity; %P, percentage of polymorphic loci

5.4.3. Association mapping

For association analysis, we employed the mixed linear model (MLM) which integrates both population structure (Q matrix) and kinship (K matrix) to avoid false positive marker-trait associations (MTAs). Only two MTAs involving two SSR markers (MPU_83 and MPU_122) passed the stringent Bonferroni adjusted threshold $P < 2.77 \times 10^{-4}$ ($P = 0.05/180$, where 180 is the number of markers used in the analysis). These two associations were for the seed yield per plant (phenotypic variance explained, PVE = 31.12%) and the pedicel length (PVE = 25.07%) (Figure 27).

We also used a less stringent criterion of $P < 5.55 \times 10^{-3}$ (1/180) as a cut-off for the reasons explained in the materials and method (section 4.4.4). By this method, a total of 15 additional MTAs related to inflorescence length, flower buds per inflorescence, flower length, pedicel length, seed length, seed width, seed thickness, hundred seed weight, and seed yield per plant were identified (Table 34). The PVE for these MTAs ranged from 14.72 to 31.12%. These MTAs were contributed mainly by 10 SSR

markers: MPU_19, MPU_42, MPU_54, MPU_57, MPU_58, MPU_83, MPU_89, MPU_108, MPU_111, and MPU_122. We found four MTAs contributing for hundred seed weight, two each for seed length, seed width, and pedicel length, and at least one each for inflorescence length, flower buds per inflorescence, flower length, seed thickness, and seed yield per plant. The highest PVE (31.12%) was observed in MTA between SSR marker MPU_83 and important agronomic trait, seed yield per plant. This crucial MTA was also spotted at Bonferroni adjusted threshold $P < 2.77 \times 10^{-4}$ (Figure 27). The Manhattan and Q-Q plots of MTAs at $P < 5.55 \times 10^{-3}$ are given in Figures 28 and 29, and the MTAs significant at threshold $P < 0.05$ and PVE > 10% are provided in Table 35.

The association analysis also revealed four SSR markers which associated with multiple agronomic traits ($P < 5.55 \times 10^{-3}$; PVE > 10%). Marker MPU_83 was not only associated with the seed length and seed width, but also with the seed yield per plant. SSR loci MPU_54 was associated with seed thickness and hundred seed weight. MPU_42 was associated with seed length and seed width, and MPU_19 was associated with flower length and flower buds per inflorescence. These MTAs were justified by the high significant correlations ($P < 0.05$) values between these traits (Table 29).

Table 34. Marker-trait associations identified using mixed linear model (MLM) at $P < 5.55 \times 10^{-3}$ with phenotypic variance explained (PVE) >10%.

Trait	Marker	*P*-value	PVE (%)
Inflorescence length	MPU_58	4.50×10^{-3}	14.72
Flower buds per inflorescence	MPU_19	2.33×10^{-3}	19.43
Flower length	MPU_19	4.09×10^{-3}	15.48
Pedicel length	MPU_57	3.03×10^{-3}	15.81
	MPU_122	2.60×10^{-4}	25.07
Seed length	MPU_42	4.35×10^{-3}	15.43
	MPU_83	4.99×10^{-3}	14.94
Seed width	MPU_42	5.20×10^{-3}	15.61
	MPU_83	2.55×10^{-3}	18.48
Seed thickness	MPU_54	5.04×10^{-3}	15.15
Seed yield per plant	MPU_83	1.00×10^{-4}	31.12
Hundred seed weight	MPU_54	2.88×10^{-4}	26.7
	MPU_89	1.10×10^{-3}	20.27
	MPU_108	7.68×10^{-4}	22.08
	MPU_111	2.78×10^{-3}	16.85

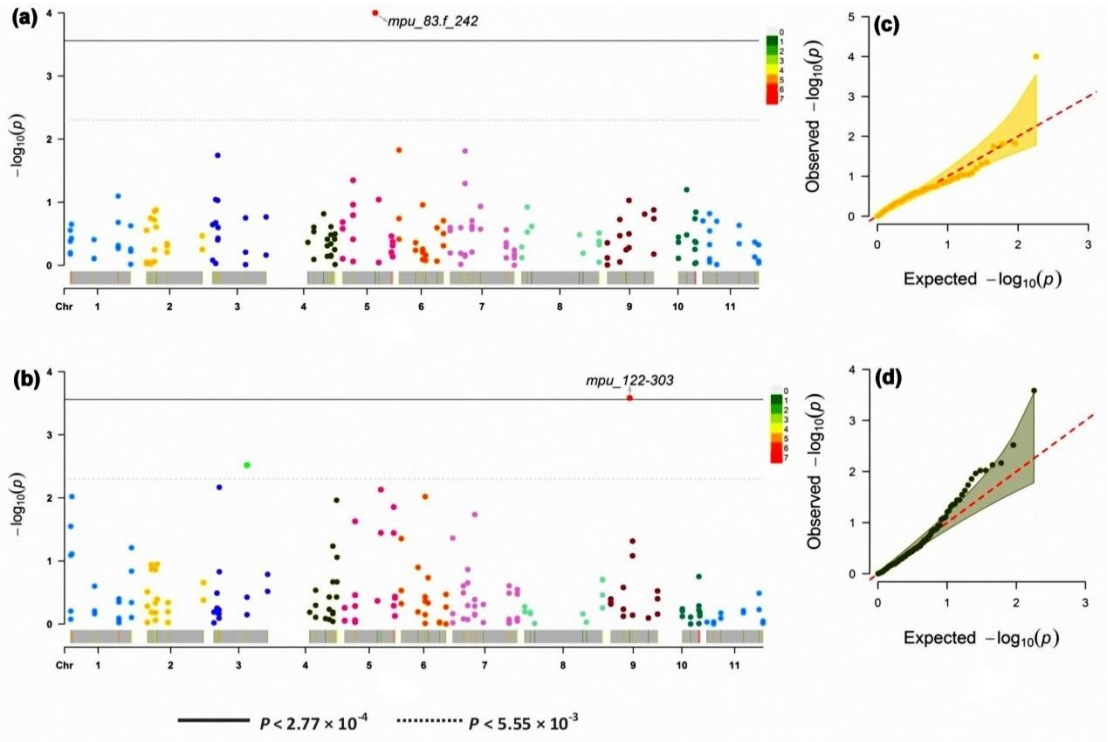


Figure 27. Manhattan plot (*P*-value) for MLM depicting significant marker-trait association at adjusted Bonferroni threshold $P < 2.77 \times 10^{-4}$ **(a)** seed yield per plant **(b)** pedicel length. Q-Q plot (quantile-quantile) of MLM **(c)** seed yield per plant **(d)** pedicel length.

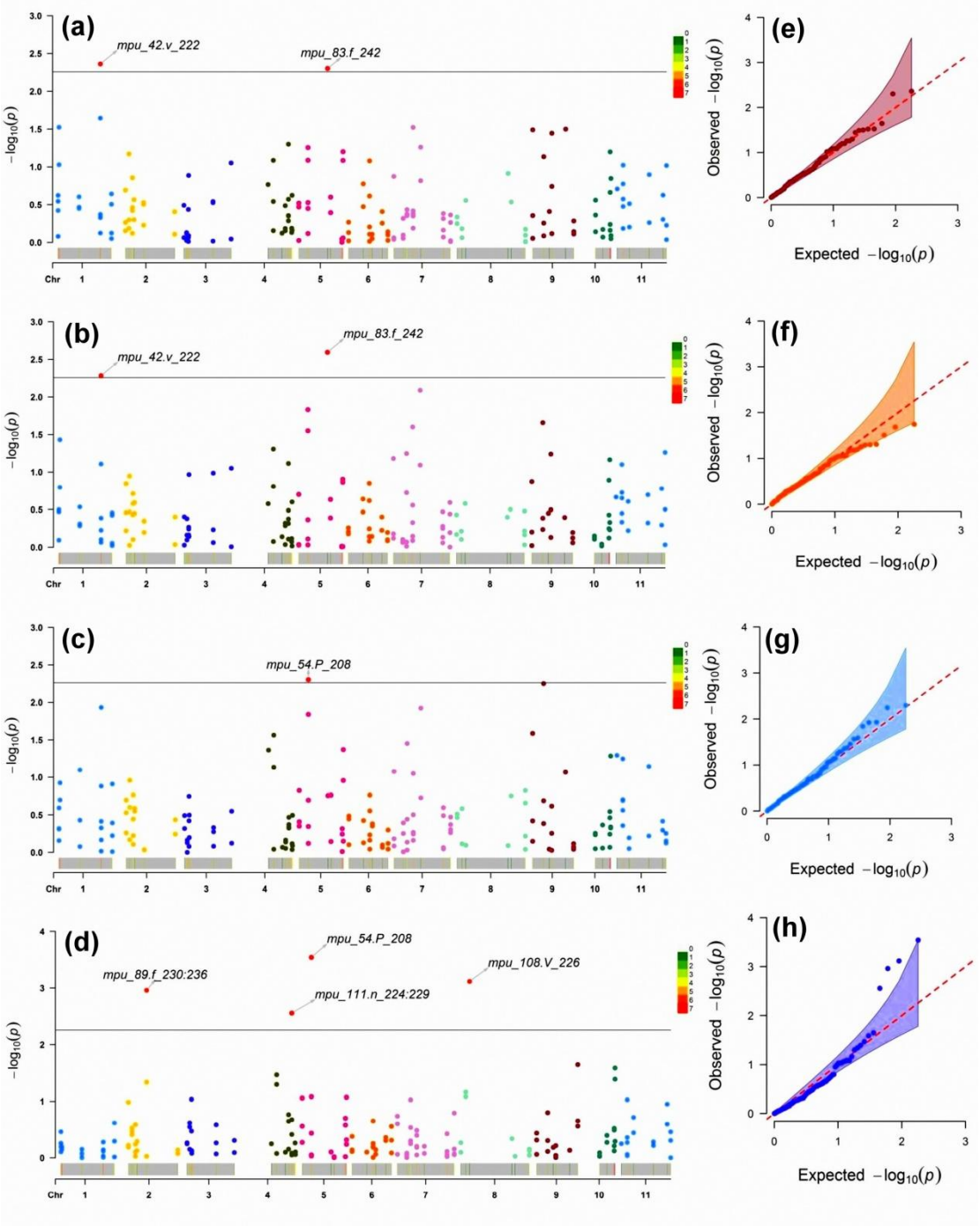


Figure 28. Manhattan plot (*P*-value) for MLM depicting significant marker-trait association at adjusted threshold $P < 5.55 \times 10^{-3}$ **(a)** seed length **(b)** seed width **(c)** seed thickness **(d)** hundred seed weight. Q-Q plot (quantile-quantile) of MLM **(e)** seed length **(f)** seed width **(g)** seed thickness **(h)** hundred seed weight.

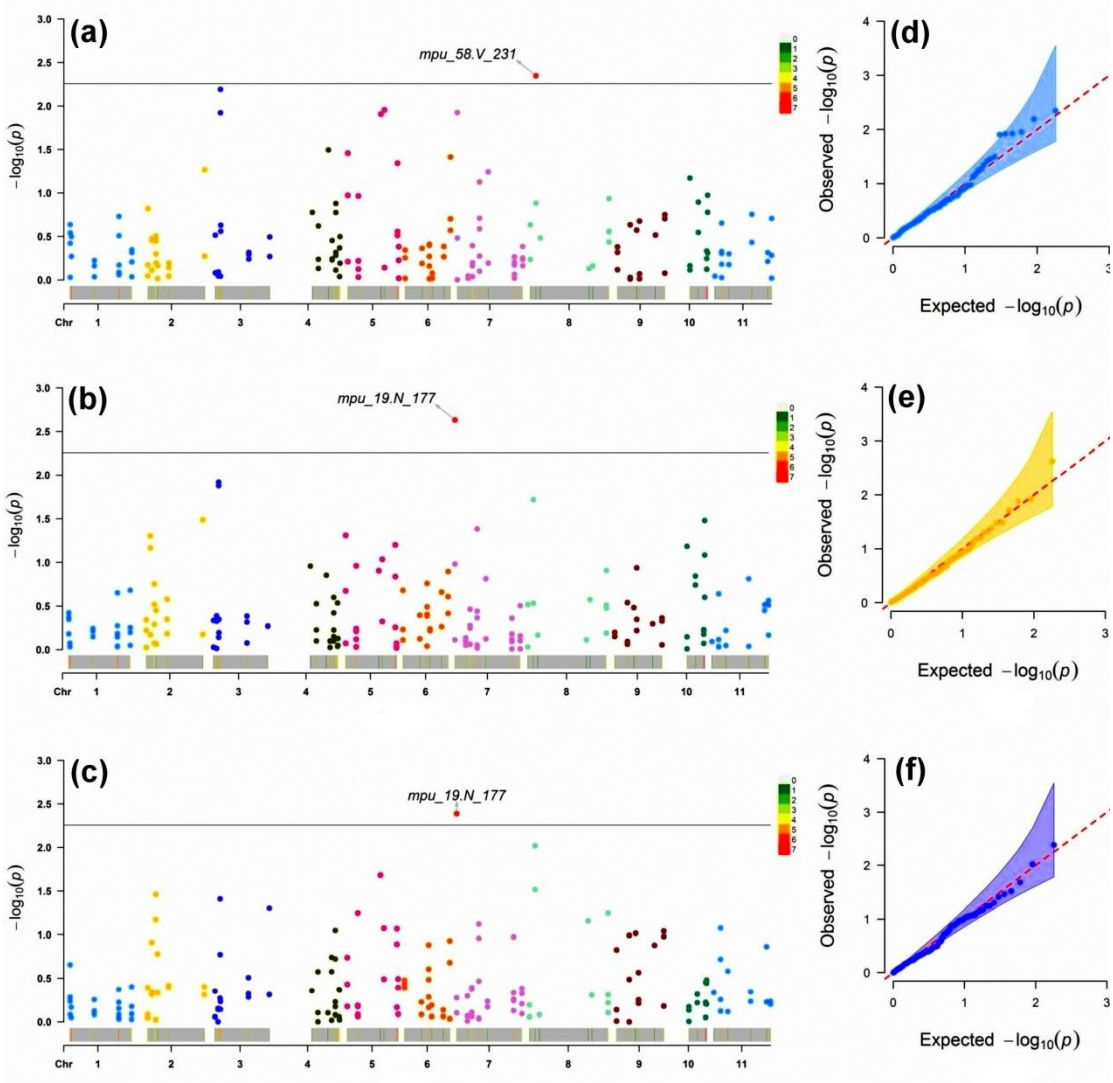


Figure 29. Manhattan plot (*P*-value) for MLM depicting significant marker-trait association at adjusted threshold $P < 5.55 \times 10^{-3}$ **(a)** inflorescence length **(b)** flower buds per inflorescence **(c)** flower length. Q-Q plot (quantile-quantile) of MLM **(d)** inflorescence length **(e)** flower buds per inflorescence **(f)** flower length**.**

Table 35. MTAs identified at a significance level of $P < 0.05$ with phenotypic variance explained (PVE) >10%.

Trait	Marker	*P*-value	PVE
Inflorescence length	MPU_19	0.01189	16.54
	MPU_58	0.0045	14.72
	MPU_62	0.0455	10.53
	MPU_64	0.00644	13.88
	MPU_80	0.01113	11.42
	MPU_83	0.01239	11.07
Flower buds per inflorescence	MPU_19	0.00233	19.43
	MPU_64	0.01205	11.80

Flower length	MPU_19	0.00409	15.48
	MPU_58	0.00959	11.95
Pedicel length	MPU_57	0.00303	15.81
	MPU_62	0.01397	10.61
	MPU_80	0.00742	12.79
	MPU_94	0.00958	11.89
	MPU_98	0.00681	13.04
	MPU_122	2.60E-04	25.07
	MPU_125	0.01089	11.40
	MPU_130	0.00953	11.81
Pod width	MPU_60	0.01803	11.60
	MPU_89	0.02074	10.17
No of pods per cluster	MPU_19	0.00658	14.75
	MPU_59	0.018	10.77
	MPU_62	0.01583	12.01
	MPU_64	0.008	13.87
	MPU_107	0.01259	11.98
No of pods per plant	MPU_30	0.01939	10.29
	MPU_62	0.02502	10.72
	MPU_64	0.03645	10.06
	MPU_92	0.00613	14.80
Seed length	MPU_42	0.00435	15.43
	MPU_83	0.00499	14.94
Seed width	MPU_42	0.0052	15.61
	MPU_54	0.01477	13.41
	MPU_60	0.02505	10.70
	MPU_83	0.00255	18.48
	MPU_100	0.00815	18.07
Seed thickness	MPU_14	0.02739	10.60
	MPU_42	0.01169	12.11
	MPU_43	0.00568	13.29
	MPU_54	0.00504	15.15
	MPU_73	0.02589	10.27
	MPU_100	0.01194	13.27
Seed yield per plant	MPU_22	0.01493	11.16
	MPU_64	0.01813	10.85
	MPU_83	1.00E-04	31.12
	MPU_87	0.01548	10.94
Hundred seed weight	MPU_54	2.88E-04	26.70
	MPU_89	0.0011	20.27
	MPU_108	7.68E-04	22.08
	MPU_111	0.00278	16.85
Total protein	MPU_80	0.0279	11.08
	MPU_83	0.01775	12.92

5.4.4. Annotation of marker function

To test the reliability of our MTAs, putative gene/function for each trait-associated marker ($P < 5.55 \times 10^{-3}$; PVE > 10%) was determined by annotating the related SSR primers using the *Arabidopsis thaliana* (L.) Heynh. genome database. Out of the 10 trait-associated markers, the BLAST analysis annotated seven markers *viz*., MPU_19, MPU_42, MPU_54, MPU_89, MPU_108, MPU_111, and MPU_122 (Table 36). Importantly, the SSR marker MPU_19, associated with flower length and flower buds per inflorescence, was annotated as 4-coumarate-CoA ligase 2, and MPU_42 associated with seed length and seed width was annotated for Zinc transporter 7 precursor. Likewise, the SSR marker MPU_54 associated with traits such as seed thickness and hundred seed weight was annotated for sucrose non-fermenting 1- related protein kinase 2 (SnRK2) gene, and MPU_89 associated with hundred seed weight was annotated as 3-dehydroquinate synthase. The SSR MPU_108, associated with a hundred seed weight matched with NAC domain-containing protein which is a plant-specific transcription factor (TF) associated with multiple aspects of stress and development. Further, MPU_111 (hundred seed weight) and MPU_122 (pedicel length) were annotated with Ubiquitin-specific protease and LEUNIG homolog (LUG and LUH), respectively.

Table 36. Putative Gene ID/description of the trait-associated markers obtained from *Arabidopsis* annotated genes using BLAST results.

Markers	***P*-value**	**Associated Traits**	**Gene ID/ Description**	**Role attributed**
MPU_19	0.00671	FL, FBpI	4-coumarate-CoA ligase 2 *(4CL; EC 6.2.1.12)*	Catalyses the activation of 4-coumarate and a few related substrates to the respective CoA esters and thus channels the common, phenylalanine-derived building block into the otherwise widely distinct branches of general phenylpropanoid metabolism (Hamburger and Hahlbrock 2004).
MPU_42	0.00326	SL, SW	Zinc transporter 7 precursor (*AT2G04032*)	Mediates zinc uptake from the rhizosphere (Grotz et al. 1998; Milner et al. 2013).
MPU_54	6.95E-04	ST, HSW	Sucrose nonfermentin g 1- related protein kinase 2 (SnRK2)	Involved in the abscisic acid (ABA) signalling and plays a central role in plant stress signal transduction (Feng et al. 2018)
MPU_89	0.00294	HSW	3-dehydroquinat e synthase	Catalyzes the transformation of the seven-carbon sugar 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAH7P) into the carbocycledehydroquinate (DHQ) (Negron et al. 2011).
MPU_108	0.00133	HSW	NAC domain-containing protein	Plant-specific transcription factors (TFs) associated with multiple aspects of the stress and development region (Mathew et al. 2016).
MPU_111	0.00311	HSW	Ubiquitin-specific protease 23 *(AT5G57990)*	The organ size in plants is reported to be regulated by two reversible processes called ubiquitination and deubiquitinating (Shi et al. 2019)

MPU_122	1.53E-04	PedL	LEUNIG homolog (LUG and LUH)	The two proteins may act cooperatively to coordinate inflorescence architecture through their influences on auxin biosynthesis, transport, and perception (Douglas et al. 2017)

FL, flower length; FBpI, flower buds per inflorescence; SL, seed length; SW, seed width; ST, seed thickness; HSW, hundred seed weight; PedL, pedicel length.

CHAPTER 6

DISCUSSION

6. DISCUSSION

6.1. *M. pruriens* in Northeast India

6.1.1. Regional significance of *M. pruriens*

India's Northeastern region represents a culturally and geographically unique landscape in the subcontinent as it embodies elements of two biodiversity hotspots, *viz.*, Eastern Himalayas and Indo-Burma regions (Mittermeier et al. 2004; Kamei et al. 2012). *M. pruriens* is thought to have originated in Eastern India (Burkill 1966; Duke 1981; Wilmot-Dear 1987), which encompasses Northeast India (Tripathi et al. 2018). Both cultivated (var. *utilis*) and wild (var. *pruriens*) varieties grow profusely in lower to middle elevations all along this region. Seed is the most important part as it serves as the chief source of edible protein (Janardhanan et al. 2003; Pugalenthi et al. 2005), animal feed (Eilitta and Carsky 2003; Muinga et al. 2003), as well as a source of pharmaceutical ingredients (Lampariello et al. 2012; Sathyanarayana et al. 2016). Anecdotal shreds of evidence suggest the presence of great variability in *M. pruriens* of Northeast India. In addition to the diversity contributed by the wild variety (var. *pruriens*), indigenous communities have moved forward with the artificial selection of velvet beans by growing them in their home garden for dry seeds and young pods (Arora 1991) resulting in rich germplasm. Nevertheless, despite their promising potential for genetic improvement programs (Sathyanarayana et al. 2017), *M. pruriens* germplasm from Northeast India is poorly explored for their offerings. Thus, the first two objectives (1 and 2) of this thesis were undertaken to address this gap.

6.1.2. Variability for seed-traits

The phenotypic evaluation revealed a good variability for seed-based traits. The data obtained for seed length, seed width, seed thickness, and hundred seed weight

conformed with the earlier reports from the South Indian accessions (Leelambika et al. 2010; Sathyanarayana et al. 2012) indicating their likely recent diffusion from the common genetic stock. The two varieties - var. *pruriens* and var. *utilis* showed a wide difference in the seed characters. The minimum ranges of *utilis* were higher than the maximum values obtained in the case of *pruriens*. Var. *utilis* was more than double the size of var. *pruriens* as far as seed characters are concerned. Such wide variation indicates the role of selection either by breeding tools or by farmers' selection. However, the availability of the limited data concerning the genetic background of these accessions, and the nature and specific outcomes from such efforts (very minimal) restricts us from expanding the relevance of this observation for larger inferences.

The seed protein and carbohydrate contents in Northeast Indian accessions were consistent with the previous reports (Gurumoorthi et al. 2003; Kalidass and Mohan 2011) and were comparable to soybean (Mang et al. 2016). A significant correlation was observed between different seed-based traits suggesting their utility in breeding programs. Further, detailed studies on lipid, fibers, moisture, ash, and minerals contents along with the reliable processing methods to overcome the anti-nutritional factors, as described in some earlier studies (Leelambika et al. 2010; Sathyanarayana et al. 2012), will reinforce the utility of this germplasm as food and feed source in this region.

6.2. Genetic diversity and relationship among the Northeast Indian accessions

6.2.1. Genetic diversity

In any genetic diversity studies, estimates of expected heterozygosity (He) and the polymorphism information content (PIC) provides insight into the evolutionary

pressure on the alleles and the rate of mutation that a locus might have undergone over a while (Botstein et al. 1980; Shete et al. 2000). Additionally, the PIC values reveal information on the utility of the markers for linkage analysis (Shete et al. 2000, Salem and Sallam 2016), while He indicates gene diversity and provides an estimate of genetic distance and average heterozygosity among the genotypes in a population (Nei 1990; Shete et al. 2000). Our study revealed that the mean PIC from SSR markers was moderate (PIC between 0.25-0.50) and was analogous to earlier report with ISSR markers (Chinapolaiah et al. 2018); but superior over AFLP (Sathyanarayana et al. 2011; Tripathi et al. 2018) and previous study with SSR markers (Kumar et al. 2019). The average He (0.477) along with Shannon's information index (I = 0.496) further supported this. The value of I in the present study was higher than that of AFLP markers (I = 0.340) reported by Tripathi et al. (2018) for *M. pruriens* from Northeast India. The superiority of the genic-microsatellites over the AFLPs might be due to their improved efficacy over the dominant AFLP for revealing the genetic differences. It is recognized that AFLPs detect multiple loci distributed throughout the genome as compared to SSR markers which detect multiple alleles at a given locus (Li et al. 2011).

Further, independent analysis of the two population groups representing Eastern Himalayas and Indo-Burma regions revealed similar He and I, suggesting the identical distribution of genetic diversity among these regions. This might be due to a close similarity in their environmental conditions. Even the two population groups that arrived based on the botanical varieties showed comparable genetic diversity, which is in deviation with the earlier reports on this aspect (Tripathi et al. 2018). The overall high genetic diversity observed in the Northeast Indian germplasm might be due to its status as a primary center of diversity of *M. pruriens*. This provides supplementary

evidence to its origin in Eastern India, as reported earlier (Sathyanarayana et al. 2011; Triathi et al. 2018). However, caution must be exercised while using this information as additional studies involving a large number of accessions from multiple geographical areas are needed before we conclusively settle this issue. On the application side, the superior allelic diversity in *M. pruriens* germplasm of the Northeast region might offer a storehouse of genes and novel genetic resources for future molecular breeding programs in this crop.

6.2.2. Genetic relationship

Knowledge of population structure is necessary for comprehending the genetic diversity information, particularly in the case of mapping applications (Eltaher et al. 2018). In genome-wide association studies (GWAS), determining the population structure of the association panels is the first step before specking the true marker-trait associations and underlying genes controlling the traits (Luo et al. 2017). In the present study, STRUCTURE analysis, in addition to NJ dendrogram and the PCoA results, identified two genetic subgroups among the *M. pruriens* accessions of Northeast India. However, none of the three analyses established a clear-cut relationship between them either based on the geographical location or varietal affiliation, evocating other factors to warrant due consideration. Further, the STRUCTURE plot revealed that 43.33% of these accessions display >80% of shared ancestry, suggesting the presence of widespread genetic admixtures. These observations are consistent with Tripathi et al. (2018). Perhaps the well-acknowledged fact that *M. pruriens* varieties can interbreed readily (Sathyanarayana et al. 2011) could explain the presence of a large number of genetic mosaics. Nonetheless, the weak population structure in *M. pruriens* of Northeast India points to their significance for the GWAS.

Population differentiation represented by F_{ST} is yet another important estimate of genetic structure, and its value > 0.15 is proof of significant population divergence (Frankham et al. 2002). By this account, even though considerable divergence was observed among the hypothetical STRUCTURE-based subgroups in our population, very low differentiation appears to have shaped the two other subgroups - geographical origin and botanical varieties. These observations coincide with the AMOVA results. The major part of the total variation was accounted for within subgroups, and only a minor proportion was contributed among the subgroups. In any such population, the gene flow value (Nm) < 1 suggests restricted gene flow (Wright 1965). But we observed a very high Nm value (7.483), indicating high genetic exchange between the population groups leading to low genetic differentiation (Eltaher et al. 2018). It is well known that Nm > 1 is strong enough to hinder genetic differentiation due to genetic drift (Slatkin and Barton 1989). Earlier Tripathi et al. (2018), based on the AMOVA results, recorded the absence of barriers for gene flow between the *M. pruriens* populations of Northeast India. Perhaps this further explains the presence of the high genetic admixtures. Population history and past anthropogenic activities, which are not considered for want of more data in this study, might be the other factors triggering low population divergence.

6.3. Feasibility analysis of association mapping panel

6.3.1. Population structure

As stated previously, 61 *M. pruriens* accessions representing different varietal and geographical populations obtained from the germplasm collection of Sir MVIT, Bangalore was used for LD analysis and association mapping. The low to moderate F_{ST} values (0.032, 0.070) in *M. pruriens* varietal and geographical populations used indicated less divergence among different populations in AM panel. The population

structure analysis from the distance-based NJ tree, PCoA, and Bayesian-based STRUCTURE largely conformed to each other in suggesting the minimal influence of the geographical origin and/or the varietal affiliation on the grouping of the accessions. High gene flow estimates (Nm = 3.03), presence of genetic admixtures, and higher within-population variance in AMOVA further concur with this. These observations are also chronicled in different germplasm collections in earlier studies (Sathyanarayana et al. 2016). Put together, these imageries suggest one or more of the (i) extensive pollen flow across long geographic distances, (ii) lack of inter-varietal barriers for hybridization, (iii) long seed dispersal, and (iv) high seed germination rates are contributing to low divergence of *M. pruriens* gene pools. The resultant genetic events appear to have shaped it as a highly adaptable species as evident from its wide and diverse distribution range. Alternatively, this might also suggest incomplete lineage sorting during diversification as observed in safflower (*Carthamus tinctorius* L.) (Ambreen et al. 2018). However, we warrant caution here as our experiments are based on the limited sample size. Future studies must examine more samples from a broad geographical range to confirm these results. Further, low pairwise kinship estimates between the majority of the accessions in the AM panel suggested a weak relationship. Low population divergence and kinship estimates further support the suitability of this panel for the association analysis.

6.3.2. Linkage disequilibrium

In association mapping studies, determining the magnitude of linkage disequilibrium (LD) and LD decay is crucial as they affect the mapping resolution and the marker deployment (Flint-Garcia et al. 2003). So far, there was no report on LD in *M. pruriens*; therefore, we attempted to determine this in the first place before undertaking the AM. Our results returned a low level of LD (average R^2 = 0.032) for

all the possible combinations of SSR marker pairs which was further confirmed by the fact that only about 1.84% of them were at significant LD ($P < 0.001$). There may be several reasons for this observation. Firstly, a large distance between the marker pairs on the same chromosome might have caused low LD. This inference needs further examination as our results are based on the pseudo-chromosomal locations assigned based on the BLAST analysis as the original locations were unknown from any of the previous studies. Secondly, the low LD may be also due to the high levels of out-crossing (Jin and Bao 2009). In our study, we recorded the presence of significant gene flow among the members of AM panel indicating this could be another factor contributing to the low LD. Nonetheless, low LD has been a common observation in other related legume species such as faba bean (Sallam and Martsch 2015), barrel medic (Branca et al. 2011), etc. On the implication side, low LD and sharp LD decay underscore the need for the deployment of more markers in association mapping studies.

6.4. Association analysis

6.4.1. Phenotypic variability

We observed significant variability for the majority of the flower, pod, seed, and biochemical characteristics in our AM panel. This is consistent with earlier reports in different *M. pruriens* germplasm (Gurumoorthi et al. 2003; Kalidas and Mohan 2011; Sathyanarayana et al. 2012). The majority of the seed and pod-based traits revealed a significant positive correlation possibly due to a key role played by the pod in the seed development as a protection and nutrient source (Bennet et al. 2011). Few accessions (IC-369144, 500113-MH, 500126-KA, 500135-KA, 500178-MH, 500184-MH, and 500190-MH) showed high seed protein content akin to soybean. These stocks will be useful in breeding for protein content. But, unlike previous

studies (Tripathi et al. 2018; Kumar et al. 2019), we found less variability for the L-DOPA content. The estimation method used (Daxenbichler et al. 1972) for the L-DOPA analysis is old and laborious; thus, it is plausible that poor recovery and associated shortcomings might have skewed this experiment. It is thus necessary to revisit this data using a more reliable L-DOPA estimation method available in recent literature (Pulikkalpura et al. 2015; Singh et al. 2018). For other traits, we recorded moderately-high to high heritability (H^2 > 60%). Two earlier studies found analogous observation in *M. pruriens* for traits such as pod length, pod width, pod weight, hundred seed weight, seed yield per plant, and inflorescence length (Hadapad et al. 2018; Chinapolaiah et al. 2019a). Present results reinforce them. High heritability means less influence of the environmental factors rendering phenotypic selection reliable besides contributing to high additive effect in the breeding programs (Tiwari et al. 2011; Rosmaina et al. 2016).

6.4.2. Genetic diversity

For genetic diversity analysis, we deployed 66 SSR markers on 61 *M. pruriens* accessions of AM panel which generated a moderate PIC-value (0.25 < PIC < 0.50). This corroborates with the earlier report in *M. pruriens* (Kumar et al. 2019) albeit slightly higher allelic range and allele per locus endorsing our choice of the SSR markers. Shannon's information index suggests high genetic diversity (I = 0.62) in our germplasm collection. The value is higher than that reported using AFLP (0.34) and SSR (0.47) markers in earlier studies (Tripathi et al. 2018; Kumar et al. 2019). In a predominantly self-pollinating species like *M. pruriens*, one will expect lower levels of genetic diversity. But higher values recorded here and in some other earlier studies (discussed in Sathyanarayana et al. 2016) categorically point to the presence of out-crossing in this species as observed by Padmesh et al. (2006). This is partly supported

by a high average gene flow among the population groups (Nm = 3.03). Thus, further studies on the pollination mechanism in *M. pruriens* can throw more light on the drivers of genetic diversity in this species. Among the three geographical regions, accessions from West-Central and Southern India revealed higher diversity than that of Eastern India which is consistent with Kumar et al. (2019). Of the three botanical varieties, wild varieties (var. *pruriens* and var. *hirsuta*) were more diverse than the cultivated variety (var. *utilis*) - the finding documented in many earlier studies (Leelambika and Sathyanarayana 2011; Sathyanarayana et al. 2012, 2016; Tripathi et al. 2018; Kumar et al. 2019). Nonetheless, wide phenotypic and genetic diversity in our AM panel signifies its utility for the association analysis.

6.4.3. Association mapping

Among the recent approaches used for the fine-scale mapping of the desirable traits, association mapping has produced fast and reliable results in several legume taxa. This includes MTAs for seed-related traits in peanut (Zhao et al. 2017), iron and zinc concentration in lentil (Singh et al. 2017), frost tolerance in pea (Lui et al. 2017), flowering, pod yield per plant, and fresh pod per pant in lablab (Vaijayanthi et al. 2018), crude protein and mineral concentration in alfalfa (*Medicago sativa* L.) (Jia et al. 2017), etc. However, in *M. pruriens* it has never been put to use (Lepcha et al. 2019), perhaps due to lack of resources common in neglected or underutilized species. Thus, the present study attempted AM of key agronomic traits in *M. pruriens*. Usually, a large number of accessions (>300) are used in the association studies. In comparison, the number of accessions used here is small (61). But several recent studies have demonstrated the feasibility of finding the right marker-trait associations even using a small number of accessions (Soumya et al. 2021; Ma et al. 2018; Lehnert et al. 2018; Rohilla et al. 2020).

The present AM analysis identified 15 significant MTAs ($P < 5.55 \times 10^{-3}$) using MLM with high PVE ranging from 14.72 to 31.12%. These associations are also benefitted by the moderate to high heritability observed for all the trait-associated markers (H^2 ranging from 42.86 to 97.27%). Four SSR markers were associated with multiple traits. They are mostly comprised of seed or inflorescence traits that are highly correlated. Such an account of a single marker associated with multiple traits can be attributed to closely linked QTLs affecting different traits (Rakshit et al. 2010). Alternatively, it may also be due to the pleiotropic effect of the linked QTLs on different traits (Miller and Rawlings 1967).

6.4.4. Annotation of marker function

To further confirm our AM results, we determined the putative functions of the trait-associated markers by annotating them against the corresponding genes/loci in the *Arabidopsis thaliana* (L.) Heynh. database (https://www.arabidopsis.org). The results produced some useful insights. For instance, the SSR marker MPU_111, associated with a hundred seed weight corresponded with the ubiquitin-specific protease 23 gene in *A. thaliana*. Another gene belonging to the same family in rice (*Oryza sativa* L.) *viz.,* ubiquitin-specific protease 23 is known to exert a positive regulatory influence on grain width and size (Shi et al. 2019). SSR marker MPU_54 associated with the traits such as seed thickness and hundred seed weight was annotated for sucrose non-fermenting 1-related protein kinase 2 (SnRK2) gene. This gene is involved in the abscisic acid (ABA) signaling during seed germination, dormancy, and seedling growth and has a central role in plant stress signal transduction (Feng et al. 2018). The SSR MPU_108, associated with a hundred seed weight was traced to NAC domain-containing protein which is a plant-specific transcription factor (TF) associated with the multiple aspects of the stress and

development. It is reported that three NAC TF encoding genes (*ONAC020, ONAC026,* and *ONAC023*) are expressed at very high levels during the seed development in rice (*Oryza sativa* L.) and exhibits a strong association with the seed size or weight with the sequence variations located in the upstream regulatory region (Mathew et al. 2016). SSR marker-MPU_122 associated with pedicel length was annotated for LEUNIG homolog (LUG and LUH). Likely, these two proteins act cooperatively to coordinate inflorescence architecture through their influence on auxin biosynthesis, transport, and perception (Douglas et al. 2017). The marker MPU_89 associated with a hundred seed weight was annotated to 3-dehydroquinate synthase gene which catalyzes the transformation of the seven-carbon sugar 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAH7P) into the carbocycledehydroquinate (DHQ) (Negron et al. 2011). This pathway controls the production of precursors of aromatic amino acids in several prokaryotes, fungi, and plants (Ganem 1978). Scrutinizing its relationship with L-DOPA biosynthesis in *M. pruriens* might throw open fascinating insights on the regulation of aromatic amino acids, particularly Tyrosine production during L-DOPA production. Thus, MTAs identified in the present study will be useful for improving the seed yield and related economic traits in *M. pruriens*. To the best of our understanding, this is the first report on marker-trait association based on AM in any *Mucuna* species from anywhere in the world.

CHAPTER 7

SUMMARY & CONCLUSIONS

7. SUMMARY AND CONCLUSIONS

The *M. pruriens* can meet three burgeoning needs of tropical agriculture: low-cost protein source, high-value medicinal plant, and an excellent green manure cover crop. But like other underutilized crops, it lacks modern breeding resources. Identifying marker-trait associations (MTAs) could prompt marker-assisted breeding paving a way for the early development of the improved varieties in this plant. Recent studies have demonstrated the feasibility of identifying marker-trait associations even using a small number of accessions (<100) triggering the hope for resource-poor crops. Taking a cue from there, we characterized a panel of 61 *M. pruriens* accessions across two consecutive years and performed association mapping using 66 genic-microsatellite markers. In addition, a detailed analysis of genetic diversity and seed trait variability among the 60 *M. pruriens* germplasm accessions from Northeast India has been carried out to supplement the existing AM panel in future GWAS studies.

7.1. Variability for seed traits in *M. pruriens* natural population of Northeast India

M. pruriens is grown by several indigenous communities of Northeast India for edible pods and seeds. But, the germplasm is poorly explored for its offerings. In this study, 60 accessions collected from different locations of Northeast India revealed a significant variability for the six seed-based traits ($P < 0.05$). The two varieties-var. *pruriens* and var. *utilis* showed a wide difference in their seed characters and formed distinct clusters in PCA scatter plot. Almost all the seed-based traits were positively correlated ($P < 0.05$) suggesting their utility in the breeding programs. The presence of good phenotypic and genetic diversity in the *M. pruriens* natural population of Northeast India signifies their utility as an important resource base for future GWAS studies in this plant.

7.2. Genetic diversity among the Northeast Indian accessions

We observed a moderate to high genetic diversity in *M. pruriens* population from Northeast India (I = 0.496). This provides credence to the hypothesis of its origin in Eastern India. A moderate PIC value was recorded for most of the SSR markers (0.25 < PIC < 0.50) suggesting their efficiency for the genetic diversity analysis. The distance-based NJ-tree, PCoA plot, and Bayesian-based STRUCTURE revealed the absence of a structured population. We also observed that high gene flow has triggered low genetic differentiation among the population group. The overall high genetic diversity among the *M. pruriens* population from the Northeast region of India signifies a possible storehouse of several novel genes. Besides, high genetic diversity and weak population structure observed imply their use in future GWAS or genomic selection programs.

7.3. Feasibility analysis of association mapping panel

For association analysis, we used 61 accessions of *M. pruriens* procured from Sir MVIT, Bangalore as an association mapping (AM) panel. The distance-based NJ tree, PCoA plot, and Bayesian-based STRUCTURE analysis revealed an analogous trend suggesting the minimal influence of the geographical origin and/or the varietal affiliation on the grouping of the accessions in our AM panel. Low pairwise kinship (*Fij*) suggested weak relatedness between the accessions. A low level of LD (average R^2 = 0.032) for all the possible combinations of SSR marker pairs and its sharp decay implied the need for more markers for identifying reliable MTAs. Based on the weak population structure and low relatedness between accessions, the panel of accessions was determined to be suitable for the association mapping.

7.4. Association analysis

A significant phenotypic variability and high broad-sense heritability were observed for many of the traits measured in AM panel accessions. Fifteen MTAs were identified in association analysis for agronomically important traits with PVE >10% from MLM. Their reliability tested through annotation against the *Arabidopsis* genome database lends further credence to this. This is the first report on association mapping study in *M. pruriens* and the results are expected to offer significant groundwork for future marker-assisted breeding and mining candidate genes for important agronomic traits in this promising underutilized legume for the tropics.

CHAPTER 8

FUTURE PROSPECTS

8. FUTURE PROSPECTS

Although *M. pruriens* has enormous nutritional, agricultural, and medicinal potential, it has remained largely underexploited. Research works focusing on developing early maturing, self-supporting varieties possessing low/high L-DOPA content and resistance to biotic and abiotic stresses are the key challenges and opportunities in this promising legume species. To further strengthen the molecular breeding efforts initiated through this work, we suggest the following future actions:

1. Fortification of AM panel to encompass a large number of accessions from diverse locations.
2. Determining the stability and utility of the identified MTAs under additional environments.
3. Development/validation of robust L-DOPA phenotyping method using LC-MS/MS.
4. Developing MTAs for other important traits relevant to this crop, particularly indeterminate growth habit - a major detriment in commercial cultivation of this crop.

CHAPTER 9

REFERENCES

9. REFERENCES

Abate T, Alene AD, Bergvinson D, Shiferaw B, Silim S, Orr A, Asfaw S (2012). Tropical grain legumes in Africa and South Asia: Knowledge and opportunities. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India.

Abdurakhmonov IY, Abdukarimov A (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics, 2, 574927.

Adebowale KO, Lawal OS (2003). Functional properties and retrogradation behaviour of native and chemically modified starch of Mucuna bean (*Mucuna pruriens*). Journal of the Science of Food and Agriculture, 83(15), 1541-1546.

Agharkar SP (1991). Medicinal plants of Bombay presidency. Scientific publishers, Jodhpur, India.

Ajiwe VIE, Okeke CA, Nnabuike B, Ogunleye GA, Elebo E (1997). Applications of oils extracted from African star apple (*Chrysophyllum africanum*), horse eye bean (*Mucuna sloanei*) and African pear (*Dacryodes edulis*) seeds. Bioresource Technology, 59, 259-261.

Akinyele IO, Shokunbi OS (2015). Concentrations of Mn, Fe, Cu, Zn, Cr, Cd, Pb, Ni in selected Nigerian tubers, legumes and cereals and estimates of the adult daily intakes. Food Chemistry, 173, 702-708.

Allard RW (1960). Principles of plant breeding. John Wiley and Sonc Inc. New York, USA.

Ambreen H, Kumar S, Kumar A, Agarwal M, Jagannath A, Goel S (2018). Association mapping for important agronomic traits in safflower (*Carthamus tinctorius* L.) core collection using microsatellite markers. Frontiers in Plant Sciences, 9, 402.

Aminah SH, Sastrapradja S, Lubis I, Sastrapradja D, Idris S (1974). Irritant hairs of *Mucuna* species. Annales Bogoriensis, 5(4), 179-186.

Amini S, Asoodar MA, Iran K (2015). The effect of conservation tillage on crop yield production (the review). New York Science Journal, 8(3), 25-29.

Arora KR (1991). Native food plants of Northeastern India. Scientific Publishers, Jodhpur, India, pp. 137-152.

Arora S, Mahato AK, Singh S, Mandal P, Bhutani S, Dutta S, Kumawat G, Singh BP, Chaudhary AK, Yadav R, Gaikwad K, Sevanthi AM, Datta S, et al. (2017). A high-density intraspecific SNP linkage map of pigeonpea (*Cajanas cajan* L. Millsp.). PLoS One, 12(6), e0179747.

Audu SS, Aremu MO (2011). Effect of processing on chemical composition of red kidney bean (*Phaseolus vulgaris* L.) flour. Pakistan Journal of Nutrition, 10(11), 1069-1075.

Bajaj A (2011). Molecular mapping of bacterial blight resistance gene, drought tolerant QTL(s) and genetic diversity analysis in clusterbean {*Cyamopsis tetragonoloba* (L) Taub}. Doctorial thesis, Chaudhary Charan Singh Haryana Agricultural University, India.

Bennett EJ, Roberts JA, Wagstaff C (2011). The role of the pod in seed development: strategies for manipulating yield. New Phytologist, 190(4), 838-853.

Bennett-Lartey SO (1998). Characterization and preliminary evaluation of some accessions of local germplasm of velvet bean (*M. pruriens* var. *utilis* Wall) of Ghana. Ghana Journal of Agricultural Science, 31(1), 131-135.

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature Genetics, 48(4), 438-446.

Betancur-Ancona DA, Chel-Guerrero LA, Bello-Pérez LA, Dávila-Ortiz G (2002). Isolation of velvet bean (*M. pruriens*) starch: Physicochemical and functional properties. Starch-Stärke, 54(7), 303-309.

Bhandari HR, Bhanu AN, Srivastava K, Singh MN, Shreya, Hemantaranjan A (2017). Assessment of genetic diversity in crop plants - an overview. Advances in Plants and Agriculture Research, 7(3), 279-286.

Bhat R, Karim AA (2009). Exploring the nutritional potential of wild and underutilized legumes. Comprehensive Reviews in Food Science and Food Safety, 8(4), 305-331.

Bhatia S (2015). Application of plant biotechnology. In: Modern applications of plant biotechnology in pharmaceutical sciences (Eds. Bhatia S, Bera T, Dahia R, Sharma K). Academic Press, New York, USA, pp. 157-207.

Bohra A, Jha R, Pandey G, Patilm PG, Saxena RK, Singh IP, Singh D, Mishra RK, Mishra A, Singh F, Varshney RK, Singh NP (2017). New hypervariable SSR markers for diversity analysis, hybrid purity testing and trait mapping in pigeon pea [*Cajanus cajan* (L.) Millspaugh]. Frontier in Plant Sciences, 8, 377.

Bohra A, Sahrawat KL, Kumar S, Joshi R, Parihar AK, Singh U, Singh D, Singh NP (2015). Genetics- and genomics-based interventions for nutritional enhancement of grain legume crops: Status and outlook. Journal of Applied Genetics, 56(2), 151-161.

Borevitz JO, Nordborg M (2003). The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiology, 132(2), 718-725.

Boschin G, Arnoldi A (2011). Legumes are valuable sources of tocopherols. Food Chemistry, 127(3), 1199-1203.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. The American Journal of Human Genetics, 32(3), 314-331.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633-2635.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Analytical Biochemistry, 72, 248-254

Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Tiffin P (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the

model legume *Medicago truncatula*. Proceedings of the National Academy of Sciences USA, 108(42), 864-870.

Buckles D (1995). *M. pruriens*: A "new" plant with a history. Economic Botany, 49(1), 13-25.

Burgess S, Hemmer A, Myhrman R (2003). Examination of raw and roasted *Mucuna pruriens* for tumorigenic substances. Tropical and Subtropical Agroecosystem, 1, 287-293.

Burkill IH (1966). A Dictionary of the economic products of the Malay Peninsula. Ministry of Agriculture and Cooperatives, Kuala Lumpur, Malaysia.

Burridge JD, Schneider HM, Huynh Roberts PA, Bucksch A, Lynch JP (2017). Genome-wide association mapping and agronomic impact of cowpea root architecture. Theoretical and Applied Genetics, 130(2), 419-431.

Burton PR, Clayton DG, Cardon L, Craddocj N, Deloukas, Duncanson A, et al. under The Welcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. Nature, 447, 661-678.

Cannon SB, McKainMR, Harkess A, NelsonMN, Dash S, DeyholosMK, Peng Y, Joyce B, Stewart CN Jr, Rolf M et al. (2015). Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Molecular Biology and Evolution, 32(1), 193-210.

Capo-chichi LJA, Morton CM, Weaver DB (2004). An intraspecific genetic map of velvet bean (*Mucuna* sp.) based on AFLP markers. Theoretical and Applied Genetics, 108(5), 814-821.

Capo-chichi LJA, Weaver DB, Morton CM (2001). AFLP assessment of genetic variability among velvet bean (*Mucuna* sp.) accessions. Theoretical and Applied Genetics, 103(8), 1180-1188.

Carcea M, Cubadda R, Acquistucci R (1992). Physicochemical and rheological characterization of sorghum starch. Journal of Food Science, 57(4), 1024-1028.

Carsky RJ, Ndikawa R (1998). Identification of cover crops for the semi-arid savannah zone of West Africa. In: Cover crops in West Africa - contributing to sustainable agriculture (Eds. Buckles D, Eteka A, Osiname M, Galiba M, Galiano G). International Development Research Centre (IDRC), Ottawa, Canada, pp. 179-187.

Carsky RJ, Tarawali SA, Becker M, Chikoye D, Tian G, Sanginga N (1998). *Mucuna* - herbaceous cover legume with potential for multiple uses. Resource and crop management research monograph 25. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria.

Chang Y, Liu H, Liu M, Liao X, Sahu SK, Fu Y, Song B, Cheng S, Kariba R, Muthemba S, Hendre PS, Mayes S, Ho WK, Yssel AEJ, Kendabie P, et al. (2019). The draft genomes of five agriculturally important African orphan crops. GigaScience, 8(3), giy152.

Chen D, Yi X, Yang H, Zhou H, Yu Y, Tian Y, Lu X (2015). Genetic diversity evaluation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) using inter-

simple sequence repeat (ISSR). Genetic Resources and Crop Evolution, 62(6), 823-828.

Cheng A (2018). Shaping a sustainable food future by rediscovering long-forgotten ancient grains. Plant Science, 269, 136-142.

Cheng A, Raai MN, Zain NAM, Massawe F, Singh A, Wan-Mohtar WAAQI (2019). In search of alternative proteins: Unlocking the potential of underutilized tropical legumes. Food Security, 11(6), 1205–1215.

Chibarabada TP, Modi AT, Mabhaudhi T (2017). Expounding the value of grain legumes in the semi- and arid tropics. Sustainability, 9(1), 60.

Chikagwa-Malunga SK, Adesogan AT, Sollenberger LE, Badinga LK, Szabo NJ, Littell RC (2009). Nutritional characterization of *Mucuna pruriens* 1. Effect of maturity on the nutritional quality of botanical fractions and the whole plant. Animal Feed Science and Technology, 148(1), 34-50.

Chinapolaiah A, Bindu HH, Khadke GN, Manjesh GN, Rao NH, Kumar SS, Suthar MK (2018). Genetic diversity analysis in underutilized medicinal climber *Mucuna pruriens* (L.) DC. germplasm revealed by inter simple sequence repeats markers. Legume Research, 43(1), 32-37.

Chinapolaiah A, Bindu KH, Manjesh GN, Thondaiman V, Shivakumara KT (2019a). Genetic variability, correlation and path analysis for yield and biochemical traits in velvet bean [*Mucuna pruriens* (L.)]. Journal of Pharmacognosy and Phytochemistry, 8(4), 2698-2704.

Chinapolaiah A, Bindu KH, Thondaiman V, Manjesh GN, Rao NH, Kumar SS, Reddy RN (2019b). Variability of L-DOPA and its association with morphological growth and yield traits of *Mucuna pruriens* (L.) germplasm. Legume Research, 44(1), 8-14.

Chinapolaiah KHBA, Manjesh GN, Rao NH, Kumar SS, Kumar TV (2017). Heterosis and combining ability analysis for yield and yield contributing traits in velvet bean *M. pruriens* (L.) DC. Legume Research, 42(1), 10-17.

Chivenge P, Mabhaudhi T, Modi AT, Mafongoya P (2015). The potential role of neglected and underutilized crop species as future crops under water-scarce conditions in Sub-Saharan Africa. International Journal of Environmental Research and Public Health, 12(6), 5685-5711.

Collard BYC, Jahufer MZZ, Brouwer JB, Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica, 142(1), 169-196.

Crépon K, Marget P, Peyronnet C, Carrouée B, Arese P, Duc G (2010). Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. Field Crops Research, 115(3), 329-339.

Cullis C, Kunert KJ (2017). Unlocking the potential of orphan legumes. Journal of Experimental Botany, 68(8), 1895-1903.

Daxenbichler, M.E., Van Etten C.H., Earle, F.R., Tallent, W.H., 1972. L-DOPA recovery from *Mucuna* seed. Journal of Agriculture and Food Chemistry, 20(5), 1046-1048.

deAvila AMH, Farias JRB, Pinto HS, Pilau FG (2013). Climatic restrictions for maximizing soybean yields. In: A comprehensive survey of international soybean research - genetics, physiology, agronomy and nitrogen relationships (Eds. James E. Board). IntechOpen, DOI: 10.5772/52177

Dhaliwal SK, Talukdar A, Gautam A, Sharma P , Sharma V, Kaushik P (2020). Developments and prospects in imperative underexploited vegetable legumes breeding: A review. International Journal of Molecular Sciences, 21(24), 9615.

Dhillon PK, Tanwar B (2018). Rice bean: A healthy and cost-effective alternative for crop and food diversity. Food Security, 10(3), 525-535.

Doss A, Pugalenthi M, Vadivel VG, Subhashini G, Subash AR (2011). Effects of processing technique on the nutritional composition and antinutritients content of under-utilized food legume *Canavalia ensiformis* L. DC. International Food Research Journal, 18(3), 928-933.

Douglas SJ, Li B, Kliebenstein DJ, Nambara E, Riggs CD (2017). A novel filamentous flower mutant suppresses brevipedicellus developmental defects and modulates glucosinolate and auxin levels. PLoS One, 12(5), e0177045.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus, 12, 13-15.

Doyle JJ (2012). Polyploidy in legumes. In: Polyploidy and genome evolution (Eds. Soltis PS, Soltis DE). Springer, Heidelberg, Germany, pp. 147-180.

Dubey AN, Farmer AN, Schlueter JE, Cannon SB, Abernathy BR, Tuteja RE, Woodward JI, Shah TR, Mulasmanovic BE, Kudapa HI, Raju NL, et al. (2011).

Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeon pea (*Cajanus cajan* L.). DNA Research, 18(3), 153-164.

Duke (1981). Handbook of legumes of world economic importance. Plenum Press, New York, USA, pp. 170-184.

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S, Singh MN, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeon pea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biology, 11, 17.

Earl DA (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetic Resources, 4(2), 359-361.

Ebert AW (2014). Potential of underutilized traditional vegetables and legume crops to contribute to food and nutritional security, income and more sustainable production systems. Sustainability, 6(1), 319-335.

Eilitta M, Bressani R, Carew LB, Carsky RJ, Flores M, Gilbert R, Huyck L, St. Laurent L, Szabo NJ (2002). *Mucuna* as a food and feed crop: An overview. In: Food and feed from *Mucuna*: Current Uses and the Way Forward, Workshop (Eds. Flores BM, Eilitta M, Myhrman R, Carew LB, Carsky RJ). CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, Honduras, pp. 18-47.

Eilitta M, Carsky RJ (2003). Efforts to improve the potential of *Mucuna* as a food and feed crop: background to the workshop. Tropical and Subtropical Agroecosystem, 1, 47-55.

Eilitta M, Sollenberger LE, Littell RC Harrington LW (2003). On-farm experiments with maize-*Mucuna* systems in the Los Tuxtlas region of Veracruz, S. Mexico: I. *Mucuna* biomass and maize grain yield. Experimental Agriculture, 39(1), 5-17.

Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, Poland J, Baenziger PS (2018). Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Frontiers in Genetics, 9, 76.

Fang J, Chao CC, Roberts PA, Ehlers JD (2007). Genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] in four West African and USA breeding programs as determined by AFLP analysis. Genetic Resources and Crop Evolution, 54(6), 1197-1209.

FAO, IFAD and WFP (2015). The State of Food Insecurity in the World 2015. Meeting the 2015 international hunger targets: taking stock of uneven progress. FAO, Rome.

Farzi R, Gholami M, Baninasab B, Gheysari M (2017). Evaluation of different mulch materials for reducing soil surface evaporation in semi-arid region. Soil Use and Management, 33(1), 120-128.

Feng J, Wang L, Wu Y, Luo Q, Zhang Y, Qiu D, Han J, Su P, Xiong Z, Chang J, Yang G, He G (2018). *TaSnRK2.9*, a sucrose non-fermenting 1-related protein kinase

gene, positively regulates plant response to drought and salt stress in transgenic tobacco. Frontiers in Plant Science, 9, 2003.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal, 44(6), 1054-1064.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003). Structure of linkage disequilibrium in plants. Annual Review of Plant Biology, 54(1), 357-374.

Foyer CH, Lam H-M, Nguyen HT, Siddique KHM, Varshney RK, Colmer TD, Cowling W, Bramley H, Mori TA, Hodgson JM, Cooper JW, et al. (2016). Neglecting legumes has compromised global food and nutritional security. Nature Plants, 2, 16112.

Frankham R, Ballou JD, Briscoe DA (2002). Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.

Fujii Y, Shibuya T, Usami Y (1991). Allelopathic effect of *M. pruriens* on the appearance of weeds. Journal of Weed Science and Technology, 36(1), 43-49.

Fung SY, Tan NH, Sim SM (2010). Protective effects of *Mucuna pruriens* seed extract pretreatment against cardiovascular and respiratory depressant effects of *Calloselasma rhodostoma* (Malayan pit viper) venom in rats. Tropical Biomedicine, 27(3), 366-372.

Fung SY, Tan NH, Sim SM, Marinello E, Guerranti Aguiyi JC (2011). *M. pruriens* Linn. seed extract pretreatment protects against cardiorespiratory and neuromuscular

depressant effects of *Naja sputatrix* (Javan spitting cobra) venom in rats. Indian Journal of Experimental Biology, 49(4), 254-259.

Ganem B (1978). From glucose to aromatics: recent developments in natural products of the shikimic acid pathway. Tetrahedron, 34(23), 3353-3383.

Garcia-Estringana P, Alonso-Blázquez N, Marques MJ, Bienes R, González-Andrés F, Alegre J (2013). Use of Mediterranean legume shrubs to control soil erosion and runoff in central Spain. A large-plot assessment under natural rainfall conducted during the stages of shrub establishment and subsequent colonization. Catena, 102, 3-12.

Garg R, Jain M (2013). Transcriptome analyses in legumes: A resource for functional genomics. The plant genome, DOI: 10.3835/plantgenome2013.04.0011

Ghosal S, Singh S, Bhattacharya SK (1971). Alkaloids of *Mucuna pruriens*: Chemistry and pharmacology. Planta Medica, 19(1), 279-284.

Gnanesh BN, Bohra A, Sharma M, Byregowda M, Pande S, Wesley V, Saxena RK, Saxena KB, Kishor PK, Varshney RK (2011). Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeon pea [*Cajanus cajan* (L.) Millsp.]. Field Crops Research, 123(2), 53-61.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010). Food security: The challenge of feeding 9 billion people. Science, 327(5967), 812-818.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Nir, et al. (2011). Full-length

transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7). 644-652.

Graeub B, Chappell M, Wittman H, Ledermann S, Kerr R, Gemmill-Herre B (2016). The state of family farms in the world. World Development, 87, 1-15.

Graham PH, Vance CP (2003). Legumes: Importance and constraints to greater use. Plant Physiology, 131(3), 872-877.

Gresta F, Mercati F, Santonoceto C, Abenavoli MR, Ceravolo G, Araniti F, Anastasi U, Sunseri F (2016). Morpho-agronomic and AFLP characterization to explore guar (*Cyamopsis tetragonoloba* L.) genotypes for the Mediterranean environment. Industrial Crops and Products, 86, 23-30.

Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D (1998). Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. Proceedings of the National Academy of Sciences USA, 95(12), 7220-7224.

Gupta M, Mazumder UK, Chakraborti S, Bhattacharya S, Rath N, Bhawal SR (1997). Antiepileptic and anticancer activity of some indigenous plants. Indian Journal of Physiology and Allied Science, 51, 53-56.

Gurumoorthi P, Senthil KS, Vadivel V, Janardhanan K (2003). Studies on agrobotanical characters of different accessions of velvet bean collected from Western Ghats, South India. Tropical and Subtropical Agroecosystems, 2(3), 105-115.

Hadapad B, Ravi CS, Shivaprasad M, Bindu H, Nadukeri S, Devaraju (2018). Genetic variability and correlation studies for quantitative and qualitative traits in velvet bean

(*Mucuna pruriens* L.) genotypes in rubber plantation under hill zone of Karnataka. Journal of Pharmacognosy and Phytochemistry, 7(3), 86-90.

Hagel JM, Morris JS, Lee EJ, Desgagné-Penix I, Bross CD, Chang L, Chen X, Farrow SC, Zhang Y, Soh J, Sensen CW, Facchini PJ (2015). Transcriptome analysis of 20 taxonomically related benzylisoquinolinealkaloid-producing plants. BMC Plant Biology, 15, 227.

Hamberger B, Hahlbrock K (2004). The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences USA, 101(7), 2209-2214.

Hedge JE, Hofreiter BT (1962). Determination of reducing sugars and carbohydrates. In: Methods in carbohydrate chemistry (Eds. Whistler RL, BeMiller JN). Academic Press, New York, USA, pp. 380-394.

Heller, J. (1997). Bambara groundnut: *Vigna subterranea* (L.) Verdc. Promoting the conservation and use of under-utilized and neglected crops. IPGRI, Zimbabwe.

Hill WG, Robertson A (1968). Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38(6), 226-231.

Hira CK, Chopra N (1995). Effects of roasting on protein quality of chickpea (*Cicer arietinum*) and peanut (*Arachis hypogaea*). Journal of Food Science and Technology, 32(6), 501-503.

Hishika R, Shastry S, Shinde S, Guptal SS (1981). Preliminary phytochemical and anti-inflammatory activity of seeds of *Mucuna pruriens*. Indian Journal of Pharmacology, 13(1), 97-98.

Hu Z, Zhang D, Zhang G, Kan G, Hong D, Yu D (2014). Association mapping of yield-traits and SSR markers in wild soybean *(Glycine soja* Sieb. and Zucc.). Breeding Science, 63(5), 441-449.

Huang H, Tan H, Xu D, Tang Y, Niu Y, Lai Y, Tie M, Li H (2018). High-density genetic map construction and comparative genome analysis in asparagus bean. Scientific Reports, 8, 4836.

Iqbal A, Khalil IA, Ateeq N, Khan MS (2006). Nutritional quality of important food legumes. Food Chemistry, 97(2), 331-335.

Janardhanan K, Gurumoorthi P, Pugalenthi M (2003). Nutritional potential of five accessions of a south Indian tribal pulse, *M. pruriens* var. *utilis.* 1. The effect of processing methods on the content of L-DOPA, phytic acid and oligosaccharides. Tropical and Subtropical Agroecosystem, 1(2), 141-152.

Jezierny D, Mosenthin R, Bauer E (2010). The use of grain legumes as a protein source in pig nutrition: A review. Animal Feed Science Technology, 157(3), 111-128.

Jia C, Wu X, Chen M, Wang Y, Liu X, Gong P, Xu O, Wang X, Gao H, Wang Z, (2017). Identification of genetic loci associated with crude protein and mineral concentrations in alfalfa (*Medicago sativa*) using association mapping. BMC Plant Biology, 17, 97.

Jin L, Bao JS (2009). Progress on the trait-marker association analysis in plants. Molecular Plant Breeding, 7(6), 1048-1063.

Jonah P, Bello L, Lucky O, Midau A, Moruppa S (2011). Review: The importance of molecular markers in plant breeding programs. Global Journal of Science Frontier Research, 11(5), 5-12.

Jorge M, Eilitta M, Proud F (2007). *Mucuna* species: Recent advances in Application of Biotechnology. Fruit, Vegetable and Cereal Science and Biotechnology, 1(2), 80-94.

Kaler AS, Purcell LC (2019). Estimation of a significance threshold for genome-wide association studies. BMC Genomics, 20, 618.

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011). Microsatellite markers: an overview of the recent progress in plants. Euphytica, 177(3), 309-334.

Kalidass C, Mahapatra AK (2014). Evaluation of the proximate and phytochemical compositions of an underexploited legume *Mucuna pruriens* var. *utilis* (Wall ex Wight) Baker ex Burck. International Food Research Journal, 21(1), 303-308.

Kalidass C, Mohan VR (2011). Nutritional and antinutritional composition of itching bean [*M. pruriens* (L.) DC. var. *pruriens*] an underutilized tribal pulse in Western Ghats Tamil Nadu. Tropical and Subtropical Agroecosystems, 14(1), 279-293.

Kalidass C, Mohan VR (2012a). Biochemical composition and nutritional assessment of selected under-utilized food legume of the genus *Rhynchosia*. International Food Research Journal, 19(3), 977-984.

Kalidass C, Mohan VR (2012b). Nutritional composition and anti-nutritional of factors of little known species *Vigna*. Tropical and Subtropical Agroecosystems, 15(3), 525-538.

Kalinowski ST (2005). HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes, 5(1), 187-189.

Kamei RG, Mauro DS, Gower DJ, vanBocxlaer I, Sherratt E, Thomas A, Babu S, Bossuyt F, Wilkinson Z, Biju SD (2012). Discovery of a new family of amphibians from northeast India with ancient links to Africa. Proceedings of the Royal Society: Biological Sciences, 279(1737), 2396-2401.

Katzenschlager R, Evans A, Manson A, Patsalos PN, Ratnaraj N, Watt H, Timmermann L, Van der Giessen R, Lees AJ (2004) *Mucuna pruriens* in Parkinson's disease: a double blind clinical and pharmacological study. Journal of Neurology, Neurosurgery and Psychiatry, 75(12), 1672-1677.

Kinmani EN, Wachira FN, Kinyua MG (2012). Molecular diversity of Kenyan lablab bean (*Lablab purpureus* L. Sweet) accessions using amplified fragment length polymorphism markers. American Journal of Plant Sciences, 3(3), 313-321.

Konduri V, Godwin ID, Liu CJ (2000). Genetic mapping of the *Lablab purpureus* genome suggests the presence of “cuckoo” gene(s) in this species. Theoretical and Applied Genetics, 100(6), 866-871.

Kongjaimun A, Somta P, Tomooka N, Kaga A, Vaughan DA, Srinives P (2013). QTL mapping of pod tenderness and total soluble solid in yardlong bean

[*Vigna unguiculata* (L.) Walp. subsp. *unguiculata* cv.-gr *sesquipedalis*]. Euphytica, 189(2), 217-223.

Krishnamurthy R, Chandorkar MS, Palsuledesai MR, Pathak JM, Gupta R (2002). Breeding in velvet bean (*M. pruriens*) for improvement in seed yield and quality traits. Indian Journal of Agricultural Sciences, 72(12), 709-715.

Krishnamurthy R, Chandrokar MS, Kalzunkar BG, Palsule DMR, Pathak JM, Gupta R (2005). Diversity evaluation in velvet bean (*M. pruriens*) germplasm for seed yield and associated agronomic traits. International Journal of Medicinal and Aromatic Plants, 27, 291-296.

Kulski JK (2016). Next-generation sequencing - An overview of the history, tools, and "omic" applications. In: Next generation sequencing - advances, applications and challenges (Eds. Kulski JK). IntechOpen, London, UK, pp. 1-3.

Kumar A, Gupta C, Nair DT, Salunke DM (2016). MP-4 contributes to snake venom neutralization by *Mucuna pruriens* seeds through an indirect antibody-mediated mechanism. Journal of Biological Chemistry, 291(21), 11373-11384.

Kumar DS, Muthu AK, Smith AA, Manavalan R (2010). Free radical scavenging activity of various extracts of whole plant of *Mucuna pruriens* (Linn): an in vitro evaluation. Journal of Pharmacy Research, 3(4), 718-721.

Kumar N, Hazra KK, Nath CP, Praharaj CS, Singh U (2018). Grain legumes for resource conservation and agricultural sustainability in South Asia. In: Legumes for soil health and sustainable management (Eds. Meena R, Das A, Yadav G, Lal R). Springer, Singapore, pp. 79-107.

Kumar P, Saha S (2013). An updated review on taxonomy, phytochemistry, pharmacology and toxicology of *Macuna pruriens*. Journal of Pharmacognosy and Phytochemistry, 2(1), 306-314.

Kumar PR (2019). Development and characterization of microsatellite markers in *mucuna pruriens* (L.) DC. Doctoral thesis, Sikkim University, India.

Kumar PR, Sundeep S, Sathyanarayana N (2019). Microsatellite analysis reveals low interpopulation differentiation in velvet bean (*Mucuna pruriens* var. *utilis*) of India. Nucleus, doi:10.1007/s13237-019-00276-1

Kumar S, Palve AS, Patel SK, Selvanayagam S, Sharma R, Rathore A (2020). Development of genomic microsatellite markers in clusterbean using next-generation DNA sequencing and their utility in diversity analysis. Current Plant Biology, 21, 100134.

Kumar S, Singh GK, Kumar R, Bhatia NK, Awasthi CP (1991). Variation in quality traits of pigeon pea (*Cajanus cajan* L. Millsp.) varieties. Journal of Food Science and Technology, 28(3), 173-174.

Kumar SB, Shamim A, Rahul S, Kumar VR, Nilesh K (2012). A review on *Mucuna pruriens*: Its phyto constituents and therapeutic uses. Novel Science, International Journal of Pharmaceutical Science, 1(6), 308-312.

Kumawat G, Raje RS, Bhutani S, Pal JK, Mithra AS, Gaikwad K, Sharma TR, Singh NK (2012). Molecular mapping of QTLs for plant type and earliness traits in pigeon pea (*Cajanus cajan* L. Millsp.). BMC Genetics, 13, 84.

Kumwenda JDT, Gilbert RA (1998). Biomass production by legume green manures on exhausted soils in Malawi: A soil fertility network trial. In: Soil fertility research for maize-based farming systems in Malawi and Zimbabwe (Eds. Waddington SR, Murwira HK, Kumwenda JDT, Hikwa D and Tagwira F). Proceedings of the soil fertility network results and planning workshop, Mutare, Zimbabwe, pp. 85-86.

Lampariello LR, Cortelazzo A, Guerranti R, Sticozzi C, Valacchi G (2012). The magic velvet bean of *Mucuna pruriens*. Journal of Traditional and Complementary Medicine, 2(4), 331-339.

Lawal OS and Adebowale KO (2004). Effect of acetylation and succinylation on solubility profile, water absorption capacity, oil absorption capacity and emulsifying properties of *Mucuna* bean (*Mucuna pruriens*) protein concentrate. Nahrung/Food, 48(2), 129-136.

Leelambika M, Mahesh S, Jaheer M, Sathyanarayana N (2010). Comparative evaluation of genetic diversity among Indian *Mucuna* species using morphometric biochemical and molecular approaches. World Journal of Agricultural Sciences, 6(5), 568-578.

Leelambika M, Mahesh S, Jaheer M, Tripathi PK, Pittala RK, Sathyanarayana N (2016). Targeted metabolic and genomic profiling reveals parents for L-DOPA breeding in *M. pruriens* (L.) DC. Tropical Plant Biology, 9(4), 239-251.

Leelambika M, Sathyanarayana N (2011). Genetic characterization of Indian *Mucuna* (Leguminoceae) species using morphometric and random amplification of polymorphic DNA (RAPD) approaches. Plant Biosystems, 145(4), 786-797.

Lehnert H, Serfling A, Friedt W, Ordon F (2018). Genome-wide association studies reveal genomic regions associated with the response of wheat (*Triticum aestivum* L.) to mycorrhizae under drought stress conditions. Frontier and Plant Sciences, 9, 1728.

Lepcha P, Kumar PR, Sathyanarayana N (2019). Exploring genomics research in the context of some underutilized legumes-a review. In: OMICS-based approaches in plant biotechnology (Eds. Banerjee R, Kumar GV, Kumar SPJ). Scrivener Publishing, Wiley, pp. 1-8.

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, et al. (2012). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics, 45(1), 43-50.

Li L, Chokchai W, Xianqun H, Tuan H, Qiyi L, Yi P, Guimin H (2011). Comparison of AFLP and SSR for genetic diversity analysis of *Brassica napus* hybrids. Journal of Agricultural Science, 3(3), 101-110.

Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics, 21(9), 2128-2129.

Liu R, Fang L, Yang T, Zhang X, Hu J, Zhang H, Han W, Hua Z, Hao J, Zong X, (2017). Marker-trait association analysis of frost tolerance of 672 worldwide pea (*Pisum sativum* L.) collections. Scientific Reports, 7, 5919.

Loiselle BA, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany, 82(11), 1420-1425.

Lubis SHA, Sastrapradja S, Lubis I, Sastrapradja D (1980). Genetic variation of *M. pruriens*. IV: Inheritance and genotypes of seed coat colors. Annals of Bogorienses, 8, 79-87.

Luo ZA, Iaffaldano BJ, Zhuang XF, Fresnedo-Ramirez J, Cornish K (2017) Analysis of the first *Taraxacum kok-saghyz* transcriptome reveals potential rubber yield related SNPs. Scientific Reports, 7, 9939.

Lyu J (2017). Soybean genetics: Adapting to the tropics. Nature Plants, 3, 17050.

Ma L, Liu M, Yan Y, Qing C, Zhang X, Zhang Y, Long Y, Wang L, Pan L, Zou C, Li Z, Wang Y, Peng H, Pan G, Jiang Z, Shen Y (2018). Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. Frontier in Plant Sciences, 9, 561.

Mabhaudhi T, Chibarabada TP, Modi AT (2016). Water-food-nutrition-health nexus: Linking water to improving food, nutrition and health in Sub-Saharan Africa. International Journal of Environmental Research and Public Health, 13(1), 107.

Mabhaudhi T, Chimonyo VGP, Hlahla S, Massawe F, Mayes S,Nhamo L, Modi AT (2019). Prospects of orphan crops in climate change. Planta, 250 (3), 695-708.

Mahajan R, Zargar SM, Salgotra RK, Singh R, Wani AA, Nazir M, Sofi, PA (2017). Linkage disequilibrium based association mapping of micronutrients in common bean (*Phaseolus vulgaris* L.): A collection of Jammu & Kashmir, India. 3 Biotech ,7(5), 295.

Mahesh S (2015). Identification of elite germplasm, DNA fingerprinting and development of framework linkage map in Indian velvet bean (*M. pruriens*). Doctoral thesis, Visvesvaraya Technological University, India.

Mahesh S, Leelambika M, Anithakumari AM, Sathyanarayana N (2016). Genetic mapping and QTL analysis of agronomic traits in Indian *M. pruriens* using an intraspecific F2 population. Journal of Genetics, 9 (1), 35-44.

Mahesh S, Sathyanarayana N (2011). Identification of contrasting genotypes for Fusarium wilt disease in *M. pruriens* germplasm through combined *in vitro* screening and AFLP analysis. Electronic Journal of Plant Breeding, 2(4), 510-519.

Mahesh S, Sathyanarayana N (2015). Intra-specific variability for salinity tolerance in Indian *M. pruriens* L. (DC.) germplasm. Journal of Crop Science and Biotechnology, 18(3), 181-194.

Majekodunmi SO, Oyagbemi AA, Umukoro S, Odeku OA (2011). Evaluation of the anti-diabetic properties of *Mucuna pruriens* seed extract. Asian Pacific Journal of Tropical Medicine, 4(8), 632-636.

Malviya N, Yadav D (2010). RAPD analysis among pigeon pea [*Cajanus cajan* (L.) Mill sp.] cultivars for their genetic diversity. Genetic Engineering and Biotechnology Journal, 2010, GEBJ-1.

Mamatha B, Siddaramaa R, Shivananda TN (2010). Evaluation of *M. utilis* germplasm for higher biomass production, active principle and seed yield. Journal of Medicinal Plants Research, 4(13), 1297-1300.

Mang YD, Njintang YN, Abdou BA, Scher J, Bernard C, Mbofung MC (2016). Dehulling reduces toxicity and improves in vivo biological value of proteins in vegetal milk derived from two mucuna (*Mucuna pruriens* L.) seeds varieties. Journal of Food Science and Technology, 53(6), 2548-2557.

Mason AS (2015) SSR Genotyping. In: Plant genotyping (Eds. Batley J). Springer, New York, USA, pp. 77-89.

Massawe F, Mayes S, Cheng A (2016). Crop diversity: An unexploited treasure trove for food security. Trends in Plant Science, 21(5), 365-368.

Mathew IE, Das S, Mahto A, Agarwal P (2016). Three rice NAC transcription factors heteromerize and are associated with seed size. Frontiers in Plant Science, 7, 1638.

Mathur S (2013). Conservation of biodiversity through tissue culture. Research and Reviews: Journal of Microbiology and Biotechnology, 2(3), 1-6.

Messina MJ (1999). Legumes and soybeans: overview of their nutritional profiles and health effects. The American Journal of Clinical Nutrition, 70(3), 439-450.

Miller PA, Rawlings JO (1967). Breakup of initial linkage blocks through intermating in a cotton breeding population. Crop Science 7(3), 199-204.

Milner MJ, Seamon J, Craft E, Kochian LVJ (2013). Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. Journal of Experimental Botany, 64(1), 369-381.

Mir RR, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney RK (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theoretical and Applied Genetics, 125(4), 625-645.

Misra L, Wagner H (2004) Alkaloidal constituents of *Mucuna pruriens* seeds. Phytochemistry, 65(18), 2565-2567.

Mittermeier RA, Gils PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreaux J, da Fonseca GAB (2004). Hotspots revisited: Earth's biologically richest and most endangered terrestrial eco-regions. CEMEX, Mexico.

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding, 3(2), 87-103.

Mondini L, Noorani A, Pagnotta MA (2009). Assessing plant genetic diversity by molecular tools. Diversity, 1(1), 19-35.

Morris JB (2003). Bio-functional legumes with nutraceutical, pharmaceutical, and industrial use. Economic Botany, 57(2), 254-261.

Mousavi-Derazmahalleh M, Bayer PE, Hane JK, Valliyodan B, Nguyen HT, Nelson MN, Erskine W, Varshney RK, Papa R, Edwards D (2019). Adapting legume crops to climate change using genomic approaches. Plant, Cell and Environment, 42(1), 6-19.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottor M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009). A consensus genetic map of cowpea [*Vigna unguiculata* (L)Walp.] and synteny based on EST-derived

SNPs. Proceedings of the National Academy of Sciences USA, 106(43), 18159-18164.

Muinga RW, Saha HM, Mureithi JG (2003). The effect of *Mucuna* (*Mucuna pruriens*) forage on the performance of lactating cows. Tropical and Subtropical Agroecosystem, 1, 87-91.

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, et al. (2017). Genome resources for climate-resilient cowpea, an essential crop for food security. The Plant Journal 89(5), 1042-1054.

Muoni T (2019). Integrating legumes in mixed crop-livestock systems in East Africa: Farmers' perceptions, ecosystem services and support for decision making. Doctoral thesis, Swedish University of Agricultural Sciences, Sweden.

Muoni T, Öborn I, Mhlanga B, Okeyo I, Mutemi M, Duncan A (2019). The role of Mucuna pruriens in smallholder farming systems of Eastern and Southern Africa: A review. In: Agronomic crops (Eds. Hasanuzzaman M). Springer, Singapore, DOI: 10.1007/978-981-32-9783-8_23

Muralia S, Pathak AK (2003). Database of medicinal plant used in Ayurveda: Medicinal and aromatic plants cultivation and uses, pp. 185–187.

Musa M, Massawe F, Mayes S, Alshareef I, Singh A (2016). Nitrogen fixation and N-balance studies on Bambara groundnut (*Vigna subterranea* L. Verdc) landraces grown on tropical acidic soils of Malaysia. Communications in Soil Science and Plant Analysis, 47(4), 533-542.

Naylor RL, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H, Nelson RJ (2004). Biotechnology in the developing world: A case for increased investments in orphan crops. Food Policy, 29(1), 15-44.

Nedumaran S, Abinaya P, Jyosthnaa P, Shraavya B, Parthasarathy Rao, Cynthia Bantilan (2015) Grain Legumes Production, consumption and trade trends in developing countries. Working paper series No 60. ICRISAT research program, markets, institutions and policies. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India.

Negron L, Patchett ML, Parker EJ (2011). Expression, purification, and characterization of dehydroquinate synthase from *Pyrococcus furiosus*. Enzyme Research, 2011(22), 134893.

Nei M (1990). Heterozygosity and genetic-distance - a citation classic commentary on estimation of average heterozygosity and genetic-distance from a small number of individuals. Genetics, 89, 583-590.

Ogunkanmi LA, Ogundipe OT, Ng NQ, Fatokun CA (2008). Genetic diversity in wild relatives of cowpea (*Vigna unguiculata*) as revealed by simple sequence repeats (SSR) markers. Journal of Food, Agriculture and Environment, 6, 253-268.

Oraguzie NC, Wilcox PL (2007). An overview of association mapping. In: Association mapping in plants (Eds. Oraguzie NC, Rikkerink EHA, Gardiner SE). Springer, New York, USA, pp. 1-9.

Ortiz-Ceballos AI, Aguirre-Rivera JR, Osorio-Arce MM, Peña-Valdivia C (2012). Velvet bean (*Mucuna pruriens* var. *utilis*) a cover crop as Bioherbicide to preserve

the environmental services of soil. In: Herbicides-environmental impact studies and management approaches (Eds. Alvarez-Fernandez R). University of Cambridge, Cambridge, UK, pp. 167-184.

Padmesh P, Reji JV, Dhar MJ, Seeni S (2006). Estimation of genetic diversity in varieties of *M. pruriens* using RAPD. Biologia Plantarum, 50(3), 367-372.

Padulosi S, Heywood V, Hunter D, Jarvis A (2011). Underutilized species and climate change: Current status and outlook. In: Crop adaptation to climate change (Eds. Yadav SS, Redden RJ, Hatfield JL, Lotze-Campen H, Hall AE). Blackwell Publishing Ltd., pp. 507-521.

Palai JB, Jena J, Maitra S (2019). Prospects of underutilized food legumes in sustaining pulse needs in India-A review. Crop Research, 54, 82-88.

Paterson AH (1996). Making genetic maps. In: Genome mapping in plants (Eds. Paterson AH). Academic Press, New York, USA, pp. 23-39.

Pathania R, Chawla P, Khan H, Kaushik R, Khan MA (2020). An assessment of potential nutritive and medicinal properties of *Mucuna pruriens*: A natural food legume. 3 Biotech, 10(6), 261.

Patil PG, Dubey J, Bohra A, Mishra RK, Saabale PR, Das A, Rathore M, Singh NP (2017). Association mapping to discover significant marker-trait associations for resistance against fusarium wilt variant 2 in pigeon pea [*Cajanus cajan* (L.) Millspaugh] using SSR markers. Journal of Applied Genetics, 58(3), 307-319.

Patil RR, Pawar KD, Rane MR, Yadav SR (2016). Assessment of genetic diversity in *Mucuna* species of India using Randomly Amplified Polymorphic DNA and Inter

Simple Sequence Repeat markers. Physiology and Molecular Biology of Plants, 22(2), 207-217.

Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28(19), 2537-2539.

Perrier X, Jacquemoud-collet JP (2006). DARwin Software, version 5.0.158. Montpellier: Department Systems Biologiques (BIOS), CIRAD.

Poornachandra MN, Khanam S, Shivananda BGTN, Shivananda TN, Dris R (2005). *Mucuna pruriens* (L.) DC. - A novel drug for learning and memory retrieval. International Journal of Food, Agriculture and Environment, 3, 13-15.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Pugalenthi M, Vadivel V (2007a). Agro biodiversity of eleven accessions of *M. pruriens* (L.) DC. var. *utilis* (Wall ex Wight) Baker ex Burck (velvet bean) collected from four districts of south India. Genetics Resources and Crop Evolution, 54(5), 1117-1124.

Pugalenthi M, Vadivel V (2007b). L-Dopa (L-3, 4-Dihydroxyphenylalanine): A nonprotein toxic amino acid in *M. pruriens* seeds. Food, 1(2), 322-343.

Pugalenthi M, Vadivel V, Siddhuraju P (2005). Alternative food/feed perspectives of an underutilized legume *M. pruriens* var. *utilis* - a review. Plant Foods for Human Nutrition, 60(4), 201-218.

Pulikkalpura H, Kurup R, Mathew PJ, Baby S (2015). Levodopa in *Mucuna pruriens* and its degradation. Scientific Reports, 5(1), 11078.

Queneherve P, Topart P and Martiny B (1998). *Mucuna pruriens* and other rotational crops for control of *Meloidogyne incognita* and *Rotylenchulus reniformis* in vegetables in polytunnels in Martinique. Nematropica, 28(1), 19-30.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rai N, Kumar S, Singh RK, Rai KK, Tiwari G, Kashyap SP, Singh M, Rai AB (2016). Genetic diversity in Indian bean (*Lablab purpureus*) accessions as revealed by quantitative traits and cross-species transferable SSR markers. Indian Journal of Agricultural Sciences, 86(9), 1193-1200.

Raina AP, Tomar JB, Dutta M (2012). Variability in *M. pruriens* L. germplasm for L-Dopa - An anti parkinsonian agent. Genetic Resources and Crop Evolution, 59(6), 1207-1212.

Rajeshwar Y, Kumar SGP, Gupta M, Mazumder KU (2005). Studies on in vitro antioxidant activities of methanol extract of *Mucuna pruriens* (Fabaceae) seeds. European Bull of Drug Research, 13(1), 31-39.

Rakshit A, Rakshit S, Singh J, Chopra SK, Balyan HS, Gupta PK, Bhat S (2010). Association of AFLP and SSR markers with agronomic and fiber quality traits in *Gossypium hirsutum* L. Journal of Genetics, 89(2), 155-162.

Rama Reddy NR, Mehta RH, Soni PH, Makasana J, Gajbhiye NA, Ponnuchamy M, Kumar J (2015). Next generation sequencing and transcriptome analysis predicts

biosynthetic pathway of sennosides from senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties. PLoS One, 10(6), e0129422.

Ramtekey V, Bhuriya A, Ayer D, Parekh V, Modha K, Kale B, Vadodariya G, Patel R (2019). Molecular tagging of photoperiod responsive flowering in Indian bean [*Lablab purpureus* (L.) Sweet]. Indian Journal of Genetics and Plant Breeding, 79(1), 264-269.

Rastogi RP, Mehrotra BN (1994). Compendium of Indian medicinal plants. CDRI, Lucknow, India, pp. 554.

Rathinam M, Mishra P, Vasudevan M, Budhwar R, Mahato A, Prabha AL, Singh NK, Rao U, Sreevathsa R (2019). Comparative transcriptome analysis of pigeon pea, *Cajanus cajan* (L.) and one of its wild relatives *Cajanus platycarpus* (Benth.) Maesen. PLoS One, 14(7), e0218731.

Rawal HC, Kumar S, Mithra SV, Solanke AU, Nigam D, Saxena S, Tyagi A, Yadav NR, Kalia P, Singh NP, Sharma TR, Gaikwad K (2017). High quality unigenes and microsatellite markers from tissue specific transcriptome and development of a database in clusterbean (*Cyamopsis tetragonoloba*, L. Taub). Genes (Basel), 8(11), 313.

Rockström J (2003). Water for food and nature in drought–prone tropics: Vapour shift in rain-fed agriculture. Philosophical Transactions of the Royal Society B: Biological Sciences, 358(1440), 1997-2009.

Rohilla M, Singh N, Mazumder A, Sen P, Roy P, Chowdhury D, Singh NK, Mondal TK (2018). Genome-wide association studies using 50 K rice genic SNP chip unveil

genetic architecture for anaerobic germination of deep‐water rice population of Assam, India. Molecular Genetics and Genomics, 295(2), 1211-1226.

Rosegrant MW, Ringler C, De Jong IJ (2009). Irrigation: Tapping potential. International Bank for Reconstruction and Development/The World Bank.

Rosmaina, Syafrudin, Hasrol, Yanti F, Juliyanti and Zulfahmi (2016). Estimation of variability, heritability and genetic advance among local chilli pepper genotypes cultivated in peat lands. Bulgarian Journal of Agricultural Science, 22(3), 431-436.

Salem KFM, Sallam A (2016). Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. Comptes Rendus Biologies, 339(1), 1-9.

Sallam A, Martsch R (2015). Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). Genetica, 143(4), 501-514.

Sarwar H (2013). The importance of cereals (Poaceae: Gramineae) nutrition in human health: A review. Journal of Cereals and Oilseeds, 4(3), 32-35.

Saryoko A, Homma K, Lubis I, Shiraiwa T (2017). Plant development and yield components under a tropical environment in soybean cultivars with temperate and tropical origins. Plant Production Science, 20(4), 375-383.

Sasidharan N (2004). Biodiversity documentation for Kerala Part 6: Flowering plants. Kerala Forest Research Institute (KFRI), Peechi, India.

Sastrapradja S, Sastrapradja D, Aminah SH, Lubis I and Idris S (1974). Morphological and cytological investigation on some species of *Mucuna* (Papilionaceae). Annals of Bogorienses, 5(4), 173-178.

Sastry CST, Kavathekar YY (1990). Plants for reclamation of wastelands. Publications and Information Directorate, New Delhi, India, pp. 317-318.

Sathiyanarayanan L, Arulmozhi S (2007). *M. pruriens* Linn. - A comprehensive review. Pharmaconosy Review, 1(1), 157-162.

Sathyanarayana N, Leelambika M, Mahesh S, Jaheer M (2011). AFLP assessment of genetic diversity among Indian *Mucuna* accessions. Physiology and Molecular Biology of Plants, 17(2), 171-180.

Sathyanarayana N, Mahesh S, Jaheer M, Leelambika M (2012). Genetic diversity of wild and cultivated *M. pruriens* (L.) DC. accessions analyzed using 30 morphoagronomical characters. Tropical and Subtropical Agroecosystems, 15(2), 249-259.

Sathyanarayana N, Mahesh S, Leelambika M, Jaheer M, Chopra R (2016). Role of genetic resources and molecular markers in *M. pruriens* (L.) DC. improvement. Plant Genetic Resources Characterization and Utilization, 14(4), 270-282.

Sathyanarayana N, Pittala RK, Tripathi PK, Chopra R, Singh HR, Belamkar V, Bhardwaj PK, Doyle JJ, Egan AN (2017). Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *De novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers. BMC Genomics, 18, 409.

Saxena RK, Kale S, Mir RR, Mallikarjuna N, Yadav P, Das RR, Molla J, Sonnappa M, Ghanta A, Narasimhan Y, Ghanta A, Narasimhan Y, et al. (2020). Genotyping-by-sequencing and multi location evaluation of two interspecific backcross populations identify QTLs for yield-related traits in pigeon pea. Theoretical and Applied Genetics, 133(3), 737-749.

Saxena RK, Kale SM, Kumar V, Parupali S, Joshi S, Singh V, Garg V, Das RR, Sharma M, Yamini KN, Ghanta A, Rathore A, Sameerkumar CV, et al. (2017a). Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeon pea. Scientific Reports 7, 1813.

Saxena RK, Obala J, Sinjushin A, Kumar CS, Saxena KB, Varshney RK (2017b). Characterization and mapping of Dt1 locus which co-segregates with CcTFL1 for growth habit in pigeon pea. Theoretical and Applied Genetics, 130(9), 1773-1784.

Saxena RK, Von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena K, Kimurto P, Varshney RK (2014). Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. PLoS One, 9(2), e88568.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Jianlin C, Dong X; Hellsten U, May GD, Yu Y, et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature, 463(7278), 178-183.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics, 46(7), 707-713.

Seena S, Sridhar KR (2005). Physicochemical, functional and cooking properties of under explored legumes, Canavalia of the southwest coast of India. Food Research International, 38(7), 803-814.

Serdeczny O, Adams S, Baarsch F, Coumou D, Robinson A, Hare W, Schaeffer M, Perrette M, Reinhardt J (2017). Climate change impacts in Sub-Saharan Africa: From physical changes to their social repercussions. Regional Environmental Change, 17(6), 1585-1600.

Shanmugavel G, Krishnamoorthy G (2018). Nutraceutical and phytochemical investigation of *Mucuna pruriens* seed. The Pharma Innovation Journal, 7(11), 273-278.

Shete S, Tiwari H, Elston RC (2000). On estimating the heterozygosity and polymorphism information content value. Theoretical Population Biology, 57(3), 265-271.

Shetty P, Sharma S, Sathyanarayana N (2015). Exploiting legume EST databases for the development of gene-derived SSR-markers in medicinal legume *Mucuna pruriens* L. (DC.). Electronic Journal of Plant Breeding, 6(4), 1041-1051.

Shi C, Ren Y, Liu L, Wang F, Zhang H, Tian P, Tian P, Wang Y, Jing R, Liu T, Wu F, Lin Q, Lei C, Zhang X, Zhu S, Guo X, Wang, J, et al. (2019). Ubiquitin specific protease 15 has an important role in regulating grain width and size in rice. Plant Physiology, 180(1), 381-391.

Shiferaw E (2013). Development and cross-species amplification of grass pea EST-derived markers. African Crop Science Journal, 21(2), 165-172.

Shitta NS, Abberton MT, Adesoye AI, Adewale DB, Oyatomi O (2015). Analysis of genetic diversity of African yam bean using SSR markers derived from cowpea. Plant Genetic Resources: Characterization and Utilization, 14(1), 1-7.

Shivakumar MS, Ramesh S, Rao AM, Udaykumar HR, Keerthi CM (2017). Cross legume species/genera transferability of SSR markers and their utility in assessing polymorphism among advanced breeding lines in Dolichos bean (*Lablab purpureus* L.). International Journal of Current Microbiology and Applied Sciences, 6(8), 656-668.

Siddhuraju P, Becker K (2001). Effect of various indigenous processing methods on the ∝- galactoside and mono- and disaccharide content of an Indian tribal pulse, *Mucuna pruriens* var. *utilis*. Journal of the Science of Food and Agriculture, 81(8), 718-725.

Siddhuraju P, Becker K (2005). Nutritional and antinutritional composition *in vitro* amino acid availability, starch digestibility and predicted glycemic index of differentially processed mucuna beans (*M. pruriens* var. *utilis*): An under-utilized legume. Food Chemistry, 91(2), 275-286.

Siddhuraju P, Becker K, Makkar HP (2000). Studies on the nutritional composition and antinutritional factors of three different germplasm seed materials of an under-utilized tropical legume, *Mucuna pruriens* var. *utilis*. Journal of Agriculture and Food Chemistry, 48(12), 6048-6060.

Singh A, Sharma V, Dikshit K, Aski M, Kumar H, Thirunavukkarasu N, Patil BS, Kumar S, Sarkar A (2017). Association mapping unveils favorable alleles for grain

iron and zinc concentrations in lentils (*Lens culinaris* subsp. *culinaris*). PLoS One, 12(11), e0188296.

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, Pal JK, Pandit A, Singh A, et al. (2012). The first draft of the pigeon pea genome sequence. Journal of Plant Biochemistry and Biotechnology, 21(1), 98-112.

Singh SK, Yadav D, Lal RK, Gupta MM, Dhawan SS (2016). Inducing mutations through γ-irradiation in seeds of *Mucuna pruriens* for developing high L-DOPA-yielding genotypes. International Journal Radiation Biology, 93(4), 426-432.

Singh SK, Dhawan SS, Lal RK, Shanker K, Singh M (2018). Biochemical characterization and Spatio-temporal analysis of the putative L-DOPA pathway in *Mucuna pruriens*. Planta, 248(5). 1277-1287.

Singh VK, Khan AW, Saxena RK, Kumar V, Kale SM, Sinha P, Chitikineni A, Pazhamala LT, Garg V, Sharma M, Kumar CVS, Parupalli S, et al. (2016). Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeon pea (*Cajanus cajan*). Plant Biotechnology Journal, 14(5), 1183-1194.

Slatkin M, Barton NH (1989). A comparison of three indirect methods for estimating average levels of gene flow. Evolution, 43(7), 1349-1368.

Soares AR, Marchiosi R, Soares RCS, Lima RB, Santos WD, Ferrarese-Filho O (2014). The role of L-DOPA in plants. Plant Signaling and Behavior, 9(3), e28275.

Soumy PR, Burridge AJ, Singh N, Batra R, Pandey R, Kalia S, Rai V, Edwards KJ (2021). Population structure and genome-wide association studies in bread wheat for phosphorus efficiency traits using 35 K Wheat Breeder's Affymetrix array. Scientific Reports, 11, 7601.

SPSS Inc. Released 2008. SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.

Stadler LJ (1928). Mutations in barley induced by x-rays and radium. Science, 68(1756), 186-187.

Stich B, Melchinger (2010). An introduction to association mapping in plants. CAB Reviews Perspectives in Agriculture Veterinary Science Nutrition and Natural Resources, 5(039), 1-9.

Suanum W, Somta P, Kongjaimun A, Yimram T, Kaga A, Tomooka N, Takahashi Y, Srinives P (2016). Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yard long bean and wild cowpea. Molecular Breeding, 36(6), 80.

Syukur M, Sujiprihati S, Yunianti R (2012). Teknik PemuliaanTanaman, Penebar Swadaya, Jakarta, Indonesia.

Tadele Z (2009). Role of orphan crops in enhancing and diversifying food production in Africa. African Technology Development Forum Journal, 6(3), 9-15.

Tanwar UK, Pruthi V, Randhawa GS (2017). RNA-seq of guar (*Cyamopsis tetragonoloba*, L. Taub.) leaves: *De novo* transcriptome assembly,

functional annotation and development of genomic resources. Frontier in Plant Sciences, 8, 91.

Taylor CM (2019). Dissecting the genetic control of flowering time for improved phenological adaptation in narrow-leafed lupin (*Lupinus angustifolius* L.). Doctoral thesis, The University of Western Australia, Australia.

Thakur O, Randhawa GS (2018). Identification and characterization of SSR, SNP and InDel molecular markers from RNA-Seq data of guar (*Cyamopsis tetragonoloba*, L. Taub.) roots. BMC Genomics, 19, 951.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics, 28, 286-289.

Tiwari DK, Pandey P, Tripathi S, Giri SP, Dwivedi JL (2011). Studies on genetic variability for yield components in rice (*Oryza sativa* L.). Advances in Agriculture and Botanics-International Journal of the Bioflux Society, 3(1), 76-81.

Tribhuvan KU, Mithra SVA, Sharma P, Das A, Kumar K, Tyagi A, Solanke AU, Sharma R, Jadhav PV, Raveendran M, Fakrudin B, Sharma TR, Singh NK, Gaikwad K (2019). Identification of genomic SSRs in cluster bean (*Cyamopsis tetragonoloba*) and demonstration of their utility in genetic diversity analysis. Industrial Crops and Products, 133, 221-231.

Tripathi PK (2018). Genetic analysis of L - DOPA trait in *Mucuna pruriens* (L.) DC. Doctoral Thesis, Sikkim University, India.

Tripathi PK, Jena SN, Rana TS, Sathyanarayana N (2018). High levels of gene flow constraints population structure in *Mucuna pruriens* L. (DC.) of Northeast India. Plant Gene, 15, 6-14.

Tripathi YB, Updhyay AK (2001). Antioxidant property of *Mucuna pruriens* Linn. Current Science, 80(11), 1377-1378.

Tyler VE, Brady LR, Robbers JE (1976). Glycosides. In: Pharmacognosy. Lea and Febiger, Philadelphia, USA, pp. 76-103.

USDA (1994). PGRCU GA. http://www.ars.usda.gov/

Vaijayanthi PV, Ramesh S, Gowda MB, Rao AM, Keerthi CM (2018). Genome-wide marker-trait association analysis in a core set of Dolichos bean germplasm. Plant Genetic Resources, 17(1), 1-11.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, et al. (2012). Draft genome sequence of pigeon pea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology, 30(1), 83-89.

Varshney RK, Graner A, Sorrells ME (2005). Genomics - assisted breeding for crop improvement. Trends in Plant Science, 10(12), 621-630.

Varshney RK, Mohan SM, Gaur PM, Gangarao NV, Pandey MK, Bohra A, Sawargaonkar SL, Chitikineni A, Kimurto PK, Janila P, Saxena KB, et al. (2013). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Advances, 31(8), 1120-1134.

Vatanparast M, Shetty P, Chopra R, Doyle JJ, Sathyanarayana N, Egan AN (2016). Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Scientific Reports, 6, 29070.

Vecerek V, Suchý P, Straková, E, Machácek M, Palta JA, Berger JD (2008). Nutritive composition of seeds of the lupin varieties registered in the Czech Republic. In: Lupins for Health and Wealth, Proceedings of the 12th International Lupin Conference, Fremantle, Australia. International Lupin Association, Canterberry, New Zealand, pp. 14-18.

Venkatesha SC, Ganapathy KN, Gowda MB, Gowda PR, Mahadevu P, Girish G, Ajay BC (2013). Variability and genetic structure among lablab bean collections of India and their relationship with exotic accessions. Vegetos, 26(2s), 121-130.

Verma SC, Vashishth E, Singh R, Pant P, Padhi MM (2014). A review on phytochemistry and pharmacological activity of parts of *Mucuna Pruriens* used as an ayurvedic medicine. World Journal of Pharmacy and Pharmaceutical Sciences, 3(5), 138-158.

Wang Q, Tian F, Pan Y, Buckler ES, Zhang ZA (2014). SUPER powerful method for genome wide association study. PLoS One, 9(9), e107684.

Wang YJ, White PJ, Pollack L (1993). Physicochemical properties of starches from mutant genotypes of the Oh43 inbread line. Cereal Chemistry, 70(2), 199-203.

Watcharatpong P, Kaga A, Chen X, Somta P (2020). Narrowing down a major QTL region conferring pod fiber contents in yardlong bean (*Vigna unguiculata*), a vegetable cowpea. Genes (Basel), 11(4), 363.

Wiesenborn DP, Orr PH, Casper HH, Tacke BK (1994). Potato starch paste behavior as related to some physical/chemical properties. Journal of Food Science, 59(3), 644-648.

Wilmot-Dear CM (1987). A revision of *Mucuna* (Leguminosae Phaseoleae) in the Indian sub-continent and Burma. Kew Bulletin, 42(1), 23-46.

Wong QN, Tanzi AS, Ho WK, Malla S, Blythe M, Karunaratne A, Massawe F, Mayes S (2017). Development of gene-based SSR markers in winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for diversity assessment. Genes (Basel), 8(3), 100.

Wright S (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, 19(3), 395-420.

Wu X, Li G, Wang B, Hu Y, Wu X, Wang Y, Lu Z, Xu P (2018). Fine mapping Ruv2, a new rust resistance gene in cowpea (*Vigna unguiculata*), to a 193-kb region enriched with NBS-type genes. Theoretical and Applied Genetics, 131(12), 2709-2718.

Xia Q, Pan L, Zhang R, Ni X, Wang Y, Dong X, Gao Y, Zhang Z, Kui L, Li Y, Wang W, Yang H, Chen C, Miao J, Chen W, Dong Y (2019). The genome assembly of asparagus bean, *Vigna unguiculata* ssp. s*esquipedialis*. Scientific Data, 6(1), 124.

Xu P, Wu X, Wang B, Hu T, Lu Z, Liu Y, Qin D, Wang S, Li G (2013). QTL mapping and epistatic interaction analysis in asparagus bean for several characterized and novel horticulturally important traits. BMC Genetics 14, 4.

Xu P, Wu X, Wang B, Liu Y, Ehlers JD, Close TJ, Roberts PA, Diop NN, Qin D, Hu T, Lu Z, Li G (2011). A SNP and SSR based genetic map of asparagus bean (*Vigna. unguiculata* ssp. *sesquipedialis*) and comparison with the broader species. PLoS One, 6(1), e15952.

Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, Xu S, Yang Z, Xu C (2018). Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. Frontiers in Plant Science, 9, 1311.

Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF, Song J, Chen S (2015). Full- length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. The Plant Journal, 82(6), 951-961.

Yan J, Warburton M, Crouch J (2011). Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Science, 51(2), 433-449.

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J (2014). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genetics, 10(9), e1004573.

Yang S, Grall A, Chapman MA (2018). Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.), a multipurpose underutilized legume. American Journal of Botany, 105(5), 888-897.

Yao DN, Kouassi KN, Erba D, Scazzina F, Pellegrini N, Casiraghi MC (2015). Nutritive evaluation of the bambara groundnut Ci12 landrace [*Vigna subterranea* (L.) Verdc. (Fabaceae)] produced in Côte d'Ivoire. International Journal of Molecular Sciences, 16(9), 21428-21441.

Yeh FC, Yang RC, Boyle T, Ye ZH, Mao JX (1999). POPGENE, Version 1.32: The User-Friendly Software for Population Genetic Analysis. Edmonton, AB: Molecular Biology and Biotechnology Centre, University of Alberta.

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan Z, Zhu M, Zhao S, Li X, Liu X (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics, Proteomics and Bioinformatics, doi: 10.1016/j.gpb.2020.10.007

Young ND, Debelle F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, de Peer YV, et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature, 480, 520-524.

Young ND, Mudge J, Ellis TH (2003). Legume genomes: More than peas in a pod. Current Opinion in Plant Biology, 6(2), 199-204.

Yssel EJ, Kao S, Peer YV, Sterck L (2019). ORCAE-AOCC: A centralized portal for the annotation of African orphan crop genomes. Genes (Basel), 10(12), 950.

Yu J, Buckler ES (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17(2), 155-160.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38(2), 203-208.

Yuan J, Wang B, Wu TL (2011). Quantitative trait loci (QTL) mapping for inflorescence length traits in *Lablab purpureus* (L.) Sweet. African Journal of Biotechnology, 10(18), 3558-3566.

Yuan J, Yang R, Wu T (2009). Bayesian mapping QTL for fruit and growth phenological traits in *Lablab purpureus* (L.) Sweet. African Journal Biotechnology, 8(2), 167-175.

Zahran HH (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiology and Molecular Biology Reviews, 63(4), 968-989.

Zannouou A, Kossou DK, Ahanchede A, Zoundjihékpon J, Agbicodo E, Struik PC, Sanni A (2008). Genetic variability of cultivated cowpea in Benin assessed by random amplified polymorphic DNA. African Journal of Biotechnology, 7(24), 4407-4414.

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics, 42(4), 355-360.

Zhao J, Huang L, Ren X, Pandey MK, Wu B, Chen Y, Zhou X, Chen W, Xia Y, Li Z, Luo H, Lei Y, Varshney RK, Liao B, Jiang H (2017). Genetic variation and association mapping of seed-related traits in cultivated peanut (*Arachis hypogaea* L.)

using single-locus simple sequence repeat markers. Frontier in Plant Sciences, 8, 2105.

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

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W, Zhang L, Zhang X, Tang R, Garg V, Wang X, Tang H, Chow C, Wang J, et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics, 51 (5), 865-876.

Zingore S, Delve RJ, Nyamangara J, Giller KE (2008). Multiple benefits of manure: The key to maintenance of soil fertility and restoration of depleted sandy soils on African smallholder farms. Nutrient Cycling in Agroecosystems, 80(3), 267-282.

Agric Res
https://doi.org/10.1007/s40003-021-00568-6

FULL-LENGTH RESEARCH ARTICLE

Variability for Seed-based Economic Traits and Genetic Diversity Analysis in *Mucuna pruriens* Population of Northeast India

Patrush Lepcha[1] · **N. Sathyanarayana**[1]

Received: 8 June 2020 / Accepted: 29 May 2021

Abstract The tropical legume *Mucuna pruriens* (L.) DC. is one of the protein-rich crops well suited for the arid regions of the world, which suffers from low soil fertility and protein-energy malnutrition. Though thought to have originated in Southern China or Eastern India, which includes parts of Northeast (NE) India, the genetic diversity of *M. pruriens* from this region is poorly documented. In this study, we used 25 species-specific genic-microsatellite markers to investigate the diversity and population structure of sixty (60) *M. pruriens* accessions from Northeast India alongside assessing variability for the six seed-based economic traits. The study revealed high genetic diversity (I = 0.496), poor population differentiation (GST = 0.038 and FST = 0.061), extensive gene flow (Nm = 7.48), and admixture genotypes in addition to good variability for the seed-based economic traits. These findings will provide a strong basis for future studies on association mapping in addition to broadening the germplasm base for breeding programs in this lesser-known legume crop.

Keywords Genetic diversity · Microsatellite · *Mucuna pruriens* · Population structure · Seed traits

Introduction

The *Mucuna pruriens* ($2n = 2x = 22$) is a self-pollinated tropical legume classified within the phaseoloid clade of Leguminosae. It is botanically represented by two varieties (sometimes also referred to as subspecies in literature), namely var. *utilis* (cultivated) and var. *pruriens* (wild)—while the presence of a third entity viz. var. *hirsuta* (wild) is also documented [61]. The species is native of southern China or eastern India [5, 8, 60, 61]. It grows opulently in Northeast (NE) India, which houses two critical biological hot spots of the world, namely Eastern Himalaya (Sikkim, Arunachal Pradesh, and some parts of Assam) and Indo-Burma (Meghalaya, Manipur, Mizoram, Nagaland, and Tripura) [4]. All along this region, the wild variety (var. *pruriens*) occurs ubiquitously while the cultivated ones (var. *utilis*; velvet bean) are found mainly near the settlements of the indigenous people like Naga, Kuki, Mizo, Chakma, Khasi, Jaintia, etc. They grow it as a home garden/backyard/hedge plant for edible seeds and unripe pods [1]. Seed is the economically vital part of the plant and is copiously produced in all three varieties [4]. They are the source of valuable phytochemicals, as well as nutritional substances [36, 56]. Most importantly, seeds contain (up to 9%) non-protein amino acid- L-3, 4 dihydroxy-L-phenylalanine (L-DOPA)—a choice drug in the treatment of Parkinson's disease [55]. Besides, *M. pruriens* extracts are known to possess anti-microbial, anti-diabetic, aphrodisiac, anti-neoplastic, learning, and memory-enhancing properties [39, 46]. The protein content of the cultivated velvet bean (27.5–28.8%) is comparable to other popular food legumes such as *Cajanus cajan* [25], *Cicer arietinum* [18], and *Glycine max* [32]. It also promises high energy content (1568–1620 kJ 100–1) akin to pulse crops like *Vigna unguiculata*, *V. radiata*, *V. aconitifolia*, *Macrotyloma uniflorum*, and *Pisum sativum* [43, 59]. It has been a preferred green manure cover crop [11, 20] and generates lofty

✉ N. Sathyanarayana
nsathyanarayana@cus.ac.in

[1] Department of Botany, Sikkim University, 6th Mile, Samdur, Tadong-737102, Gangtok, East Sikkim 737102, Sikkim, India

Published online: 25 June 2021

1

Exploring Genomics Research in the Context of Some Underutilized Legumes—A Review

Patrush Lepcha, Pittala Ranjith Kumar and N. Sathyanarayana*

Department of Botany, Sikkim University, Gangtok, East Sikkim, India

Abstract

Broadening legume resource base is imperative to meet the ever-increasing demand for protein-rich diet in the developing world. Many legumes species considered to be minor on a global scale have now been investigated and found to possess excellent nutritional value. Some of them are even a storehouse of rare drug molecules. Till date, their large-scale adoption for cultivation has remained unmet owing to poor research investments in these crops. Many of them have skipped genomics revolution and lack targeted genetic improvement programs. Recently, there has been renewed interest in these crops, and progress in genetic and genomics resources development is catching up, fueling greater promise toward molecular breeding and gene discovery programs in the near future. This review focuses on providing nutritional potential and prospects of genomic research in four lesser-known legume species: velvet bean, winged bean, rice bean, and lablab bean, which are grown as minor crops across the Indian subcontinent.

Keywords: Genomics, legumes, genomic resources, transcriptome, nutritional potential, segregant population, genetic map

1.1 Introduction

Trends in human population growth and pattern of consumption imply that the global demand for food will continue to grow for the next 40 years. This, along with depleting land and water resources in addition to climate change,

**Corresponding author*: nsathyanarayana@cus.ac.in

Rintu Banerjee, Garlapati Vijay Kumar, and S.P. Jeevan Kumar (eds.) OMICS-Based Approaches in Plant Biotechnology, (3–18) © 2019 Scrivener Publishing LLC