GENETIC ANALYSIS OF L - DOPA TRAIT IN *MUCUNA PRURIENS* (L.) DC.

Submitted

To

Sikkim University

In Partial Fulfilment of the Requirements for the

Degree of Doctor of Philosophy

By

PANKAJ KUMAR TRIPATHI

Department of Botany
School of Life Sciences

August 2018
Gangtok – 737102, Sikkim, India

CERTIFICATE

This is to certify that the Ph.D thesis entitled **"GENETIC ANALYSIS OF L-DOPA TRAIT IN *MUCUNA PRURIENS* (L.) DC."** submitted to **Sikkim University** in partial fulfilment for the requirements of the degree of **DOCTOR OF PHILOSOPHY** in Botany embodies the research work carried out by **Mr. PANKAJ KUMAR TRIPATHI** at the Department of Botany, School of Life Sciences, Sikkim University. It is a record of *bona fide* 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 the evaluation.

PROF. N. SATHYANARAYANA
Ph.D Supervisor
Department of Botany
School of Life Sciences
Sikkim University, Gangtok

Place: Gangtok

Date: 27/08/2018

CERTIFICATE

This is to certify that the Ph.D thesis entitled **"GENETIC ANALYSIS OF L-DOPA TRAIT IN *MUCUNA PRURIENS* (L.) DC."** submitted to **Sikkim University** in partial fulfilment for the requirements of the degree of **DOCTOR OF PHILOSOPHY** in Botany embodies the research work carried out by **Mr. PANKAJ KUMAR TRIPATHI** 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 the evaluation.

DR. DHANI RAJ CHHETRI
Head, Dept. of Botany
School of Life Sciences
Sikkim University, Gangtok

Place: Gangtok
Date: 27/08/2018

DECLARATION

I declare that the present Ph.D thesis entitled **"GENETIC ANALYSIS OF L-DOPA TRAIT IN *MUCUNA PRURIENS* (L.) DC."** submitted by me for the award of the degree of **DOCTOR OF PHILOSOPHY (BOTANY)** is a *bona fide* research work carried out be 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, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

PANKAJ KUMAR TRIPATHI
Registration No.: 13SU12337
Ph.D Scholar
Department of Botany
School of Life Sciences
Sikkim University, Gangtok

Place: Gangtok

Date: 27/08/2018

Date: 27/08/2018

PLAGIARISM CHECK CERTIFICATE

This is to certify that the plagiarism check has been carried out for the following Ph.D thesis with help of **URKUND SOFTWARE**.

"GENETIC ANALYSIS OF L - DOPA TRAIT IN *MUCUNA PRURIENS* (L.) DC."

The result is within the permissible limit decided by the Sikkim University.

PANKAJ KUMAR TRIPATHI
Ph.D student

PROF. N. SATHYANARAYANA
Research Supervisor

ACKNOWLEDGEMENTS

*First and foremost, **I** owe my whole-hearted thanks to **Lord Almighty** who has provided so many helping hands to complete this piece of work successfully.*

I would like to convey my deep sense of gratitude and sincere thanks to my guide **Prof N. Sathyanarayana**, Department of Botany, Sikkim University for his expert guidance, sustained thought, encouragement, timely suggestion and invaluable help during the course of this study and in the preparation of the thesis. In spite of his many other commitments, he used to review my thesis progress, gave his valuable suggestions and made corrections.

I wish to express my sincere thanks to **Dr. Dhani Raj Chhetri,** Head, Department of Botany, Sikkim University for support and facilities provided in completing the present investigation. I also thank all the faculty members of the Department of Botany for their timely help during the course of my research work.

My sincere thanks to **Hon. Vice-Chancellor**, Sikkim University for providing lab facilities and botanical garden for my research work. I extend my profound gratitude to the authorities of Sikkim University for the award of UGC Non-NET Fellowship.

I express my sincere thanks to **Science and Engineering Research Board (SERB),** Govt. of India, New Delhi for providing financial support in the form of Senior Research Fellowship under the Project **"***DNA fingerprinting of important germplasm, assessment of genetic diversity and identification of elite parental lines for trait specific breeding in velvet bean (M. pruriens)"* undertaken by **Prof. N. Sathyanarayana.**

I express my whole hearted thanks and gratefulness to **Prof. T. S. Rana**, Senior Principal Scientist, CSIR-National Botanical Research Institute (NBRI), Lucknow for his invaluable help in AFLP analysis. I sincerely thank **Dr. S. N. Jena**, Senior Scientist and **Mr. Ravi Prakash Shukla**, Ph. D Scholar, CSIR-NBRI, Lucknow for their help in AFLP analysis.

I also thank **Dr. Biswajeet Gopal Roy**, Assistant Professor, Department of Chemistry, Sikkim University for his help in L-Dopa analysis. I express my cheerful thanks to Mr. Deepak Chettri and Mr. Roshan Rai, Dept. of Botany, Sikkim University for their help all throughout my study.

Mr. Pittala Ranjith Kumar who was my colleague, friend and roommate shared precious time with me and helped greatly in experiment designing, field work and survey. I sincerely thank him for his support both in professional and personal matters. I also thank **Mr. Sai Sundeep Y** who helped me in survey and collection.

I consider it as a pleasure to express my deep sense of gratitude to my friends - Dawa Lhendup Lepcha, Nima Tshering Lepcha, Patrush Lepcha, Aditya Pradhan, Shubhakar Gurung, Ashish Rai, Hemant Sharma, Aita Rani Subba, Alice Rachel Subba, Raksha Mukhia, Puja Shashankar and Abhijeet Chhetri. My sincere thanks are also due to my friends Mr. Bhusan Gurung, Mr. Lalit Chaurasia and Mr. Ishfaq Nabi for their helping hand.

I thank **Hostel warden**, Sikkim University who made my stay in the hostel most comfortable. I also extend my thanks to the **authors** of the research articles referred during this research work.

I thank all my teachers and friends who have come across in this journey of my research career directly or indirectly.

I find no words to express my sincere thanks to Ms. **Poonam Mishra** for her moral support and encouragement throughout the course of my research work.

Last but not the least, my beloved **parents** and **sisters**; honestly I am lucky to have you in all my odds as well as happy experiences throughout my life. Your moral support, love and care turned out to be my blessings and I owe a great deal of gratitude in this journey to all of you.

Once again I whole heartedly thank **one and all** who are responsible for completion of my Ph. D work directly or indirectly.

(PANKAJ KUMAR TRIPATHI)

CONTENTS

List of Tables

List of Figures

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AHC	Agglomerative Hierarchical Clustering
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
AP	Assam Plains
ATP	Adenosine Tri-Phosphate
BA	6-Benzyladenine
BAP	6-Benzylaminopurine
BSI	Botanical Survey of India
CAS	Chemical Abstract Service
CDA	Czapek Dox Agar
CIEPCA-IITA	Centre d'Information et d'echanges sur les Plantes de Couverture en Afrique - International Institute of Tropical Agriculture
CIAT	Centro Internacional de Agricultura Tropical
Cr	Crore
CSIR	Council of Scientific and Industrial Research
CTAB	Cetyl Trimethyl Ammonium Bromide
DDC	Dopa decorboxylase
DM	Dry Mass
EC	Electrical Conductivity
EDTA	Ethylene Diamine Tetra acetic Acid
EH	Eastern Himalayas
EST	Expressed Sequence Tags
FAO	Food and Agriculture Organization
FM	Fresh Mass
GA_3	Gibberellic Acid
GMCC	Green Manure Cover Crop

GEI	Genotype Environment Interaction
HPLC	High Performance Liquid Chromatography
IAA	Indole-3-Acetic Acid
IBA	Indole-3-Butyric Acid
IIHR	Indian Institute of Horticultural Research
Ib	Band Informativeness
ISSR	Inter Simple Sequence Repeat
KGMU	King George Medical University
LIFDCs	Low-Income Food Deficit Countries
LD	L-Dopa (L-3,4-Dihydroxyphenylalanine)
LOD	Limit of Detection
LOQ	Limit of Quantification
MAS	Marker Assisted Selection
MS	Murashige and Skoog
NAA	1-Naphthaleneacetic Acid
NBPGR	National Bureau of Plant Genetic Resources
NE	Northeastern
PAGE	Polyacrylamide Gel Electrophoresis
PCA	Principal Component Analysis
PCoA	Principal Co-ordinate Analysis
PCR	Polymerase Chain Reaction
PCs	Primer Combinations
PD	Parkinson's disease
PIC	Polymorphic Information Content
PVP	Polyvinyl Pyrrolidone
QTL	Quantitative Trait Loci
RAPD	Randomly Amplified Polymorphic DNA
RCBD	Randomized Complete Block Design
RFLP	Restriction Fragment Length Polymorphism
RILs	Recombinant Inbred Lines

SSR	Simple Sequence Repeats
TAE	Tris Acetic acid EDTA
TBE	Tris Boric acid EDTA
TE	Tris EDTA
TEMED	Tetra Methyl Ethylene Diamine
UPGA	Unweighted Pair-Group Average
UPGMA	Unweighted Pair-Group Method with Arithmetic mean
USDA	United States Department of Agriculture
UV	Ultra Violet
var.	Variety
WHO	World Health Organization
2,4-D	2,4-Dichlorophenoxyacetic Acid

Symbols and units

%	Percentage
cm	Centimeter
cM	centi-Morgan
%CV	Coefficient of Variation
ft	feet
g	Gram
h	Hour
h^2	Heritability
kg	Kilogram
m	Meter
MT	Metric ton
mAU	Milli-Absorbance Unit
MQ	MiliQ Water
mg	Milligram
min	Minutes
ml	Millilitre
mM	Millimolar
ppm	Parts Per Million
RT	Retention Time
SD	Standard Deviation
s	second
SE	Standard Error
µG	Microgram
µM	Micromolar
µl	Microlitre
nm	Nanometer
v/v	Volume by volume
w/v	Weight by volume
”	Inches

1. General Introduction

1.1 Medicinal plants in healthcare

According to World Health Organization (WHO), around 5.2 billion people around the world live in less developed countries, mostly in Asia and Africa (Rao and Rajput 2010). About 80% of them (~3.3 billion) depend on traditional medicines for their primary healthcare needs (Chakraborty 2018). Almost 2000 indigenous groups around the world practice their own healthcare systems conserved through oral traditions (Liu et al. 2009; Kebriaee-zadeh 2003).

Medicinal plants are the "backbone" of traditional healthcare system (Davidson-Hunt 2000). The use of plants as medicine dates back to antiquity and to the dawn of civilization. There are abundant evidences in favor of plants being used for the treatment of diseases and fortifying human body systems in ancient medical text such as Ayurveda and alternative medical systems such as Unani, Sidda and Chinese traditional medicine (Mahima et al. 2012; Sen and Chakraborty 2016). Plant based healing systems, typical of indigenous groups, also evolved in parallel over the centuries in different cultures and communities (Kirmayer et al. 2003). The major benefit of using plant based products for healthcare needs is their holistic action on the body and lack of side effects (Gautam et al. 2012). They also command better cultural acceptability (Williams and Ahmad 1999). Recently, bio-prospection of plants for therapeutically important and biologically active molecules is gaining impetus. Large scale screening of plant extracts have led to discovery of new pharmaceutical leads for a number of life threatening diseases (Dewick 2002; Gullo et al. 2006). Many successful plant based drugs are already in the market for the treatment of diseases ranging from fever to cancer to neurological problems (Table 1).

Table 1 High value drugs derived from the medicinal plants (Chakraborty 2018)

Drugs	Medicinal plant	Action against
Vinblastine	*Catharanthus roseus*	Cancer
Vincristine	*Catharanthus roseus*	
Camptothecin	*Camptotheca acuminate*	
Taxol	*Taxus baccata*	
Podophyllotoxin	*Podophyllum peltatum*	
Artemisinin	*Artemisia annua*	Malaria
Quinine	*Cinchona ledgeriana*	
Quinoline	*Cinchona ledgeriana*	
Digoxin	*Digitalis purpurea*	Cardiac disorder
Digitoxin	*Digitalis purpurea*	
Quinidine	*Cinchona ledgeriana*	
Codeine	*Papaver somniferum*	Analgesic
Morphine	*Papaver somniferum*	
Allicin	*Allium sativum*	Diabetes
S-Methyl Cysteine Sulfoxide	*Allium cepa*	
Nicotine	*Nicotiana tabaccum*	Brain disorder
Caffeine	*Coffea canephora*	
Diosgenin	*Dioscorea Mexicana*	Sexual problems
Stigmasterol	*Glycine max*	
Cannabinoids	*Cannabis sativa*	Psycho-disorder
Tubocurarine	*Chondodendron tomentosum*	Muscle disorder
Atropine	*Atropa belladonna*	Nervous system
Hyoscyamine	*Hysocyamus niger*	

India contributes to about 7% of the world biodiversity and is one of the 12 mega diverse countries in the world with 16 agro-climatic zones, 15 biotic provinces and 426 biomes (Ghose 2017). Here, plants are distributed from Himalayan to marine and desert to rain forest ecosystems with over 8000 of them reported to have medicinal uses (Kamble et al. 2009). The domestic demand for medicinal plants was estimated at 1, 95,000 MT with an export demand of 1, 34,500 MT during the year 2014-2015. The consumption of crude drug was 5,12,000 MT in the year of 2014-15 which accounted for net worth of ₹ 5,500 Cr (Rathore and Mathur 2018). A remarkable elevation in the export of herbal medicines has been witnessed from ₹ 345.80 Cr in 2005-2006 to ₹ 3,211 Cr in 2014-2015. This shows the market potential for herbal medicines have increased about nine folds in the last 10 years (Mungikar 2018),

suggesting India has excellent potential to harness the economic power of this unique natural resource.

1.2 Genetic improvement of medicinal plants

Genetic improvement of medicinal plants, as compared to traditional food crops is in the nascent stage with much desired in terms of breeding efforts. Increased yield of valuable compounds, elimination of unwanted compounds, tolerance to abiotic and biotic stresses and better homogeneity of the cultivars are some of the challenges that need attention (Carlen 2011). Encouragingly, medicinal plants show abundant diversity in wild habitats, germplasm collections and/or already developed cultivars which can be tapped for novel genes and metabolic pathways. It is therefore important to conserve and diversity germplasm base to prevent loss of genetic diversity, including the diversity within species (Kasso and Balakrishnan 2013). Such efforts have been initiated in *Acorus calamus* (Chauhan et al. 2006), *Aloe vera* (Lal et al. 2015), *Andrographis paniculata* (Mishra et al. 2003), *Artemisia annua* (Kumar et al. 1999), *Asparagus racemosus* (Singh et al. 2007), *Bacopa monnieri* (Darokar et al. 2007), *Casia senna* (Singh et al. 1997), *Catharanthus roseus* (Kulkarni et al. 1997), *Foeniculum vulgare* (Lal et al. 2006), *Lippia alba* (Kumar et al. 1999), *Mentha citrate* (Tyagi et al. 1988), *Ocimum aficanum* (Lal et al. 2014), *Papaver somniferum* (Lal et al. 2010), *Withania somnifera* (Lal et al. 2011) and *Hyoscyamus niger* (Sharma et al. 1987). Improved varieties/cultivars have been developed in *Artemisia annua* (Alin 1997), *Hypericum perforatum* (Debrunner et al. 2000), *Ginko biloba* (Canter et al. 2005), *Thymus vulgaris* (Mewes et al. 2008), *Catharanthus rose*us (Kulkarni et al. 2016), *Vetiveria zizaniodes* (Lavania et al. 2016), *Medicago truncatula* (Bell et al. 2001), *Salix aegyptiaca* (Asgarpanah 2012), *Origanum majorana*, *Foeniculum*

vulgare, Carum carvi and *Hypericum perforatum* (Pank, 2010) etc. both at National and International levels.

High natural variability present within medicinal plant species is one of the reasons that classical breeding approaches were mainly used till now (Pank 2010). The other reasons are that these methods are relatively cheap and allow a return on investments (Canter et al. 2005). On the other hand, breeding new cultivar by conventional methods generally takes 5 to 15 years according to species and the selection criteria (Pank 2010). Thus, to react more quickly to the requirements of the stakeholders, methods to accelerate the breeding procedures must be factored in, while planning the improvement programs.

Molecular breeding can be a promising approach to this effect. In recent years, targeted use of DNA markers has made it possible to identify genomic regions that control variations in secondary metabolites, including quantitative traits loci (QTLs) (Collard et al. 2005). For example, in *Allium cepa* (McCallum et al. 2007) QTL mapping allowed identification of a genomic regions controlling high-fructan content (McCallum et al. 2006). A QTL for pungency, mapped in *Capsicum annuum* (Blum et al. 2003), has been cloned and sequenced revealing it was different from structural genes involved in biosynthetic pathways of capsaicinoids (Blum et al. 2003). Other QTL analyses for plant secondary metabolites includes: Limonene content in *Eucalyptus grandis* (Shepherd et al. 1999); linalool, nerol and geraniol contents in *Vitis vinifera* (Ronning et al. 1999), fruit metabolites in *Solanum lycopersicum* (Schauer et al. 2006); *S. tuberosum* and *S. berthaultii* (Yencho et al. 1998) and swerchinin, swertanone, and swertianin contents in *Swertia chirayita* (Misra et al. 2010). Analyses have been also carried out in a number of plant species for several targeted and non-targeted metabolites in order to discover the nature of inheritance

and location of genetic controls - which are elegantly discussed in a review (Kumar and Gupta 2008). More importantly, the studies in *Arabiodopsis thaliana (*Kliebenstein et al. 2002) have shown that the content of plant secondary metabolites exhibits transgressive segregation which implies unique opportunity to recover extreme phenotypes for the desired traits (Keurentjes et al. 2006; Schauer et al. 2006).

In addition, EST databases can be used directly for developing functional markers for prediction of phentoypes (Kumar and Gupta 2008). EST databases for active ingredients have been developed in several medicinal plant species which include *Plantago major* (Aucubin), *Gloriosa superba* (Colchicine, sitosterol and luteolin), *Withania somnifera* (Withaferin A and other alkaloids), *Glycyrrhiza uralensis* (Flavanoids), *Papaver somniferum* (Morphine and codeine), *Mentha piperita* (Terpenoids), *Allium species* (Alliin), *Capsicum annuum* (Capsaicin, capsaicin and solanine), *Artemisia annua* (Artemisinin), *Nicotiana tabacum* (Nicotine). The details of these resources are available in http://www.ncbi.nlm.nih.gov; Kumar and Gupta (2008). Additionally, transcriptomes of hundreds of medicinal plants have been sequenced, eg, Caryophyllales (Yang et al. 2015), Fabaceae (Cannon et al. 2015), *Oenothera* (Onagraceae) (Hollister et al. 2015), *Rhodiola algida* (Zhang et al. 2014), *Salvia sclarea* (Hao et al. 2015), *Polygonum cuspidatum* (Hao et al. 2012) and *Taxus mairei* (Hao et al. 2011). The information uncovered from these studies serves as an important resource base for the characterization of traits related to secondary metabolite formation and for probing the relevant molecular mechanisms (Zhang et al. 2014). Further, several SNPs and single-copy orthologous gene sequences have been extracted from the unigene datasets of multiple medicinal plants (Hao et al. 2012). They can be effectively employed for large-scale phylogenetic reconstruction and evolutionary analyses (Yang et al. 2015) in these species.

Recently, whole genome sequencing has been implemented in several representative medicinal plant species including *Ganoderma lucidum* (Chen et al. 2012), *Ziziphus jujuba* (Li et al. 2014c), *Catharanthus roseus (*Kellner et al. 2015), *Panax notoginseng* (Chen et al. 2017), *Camptotheca acuminate (*Zhao et al. 2017), *Calotropis gigantea* (Genevieve et al. 2018), *Azadirachta indica* (Chakraborty 2018) and many others. They have emerged as valuable model for studying the genetics and metabolic activities of these herbs (Kellner et al. 2015). These efforts are providing greater impetus to breeding programs. In cases where conventional breeding is not successful, new approaches like RNAi technology is providing remarkable alternative for trait regulation (Allen et al. 2004). This technique, as applied in *Papaver somniferum*, resulted in accumulation of precursor (s)-retculine at the expense of morphine, codeine, oripavine and thebaine (Kumar and Gupta 2008). Besides, a systems biology approach along with metabolomics is providing promising results to identify elite cultivars that possess desirable levels of drug molecule/pharmaceuticals as reported in *A. cepa* and *S. lycopersicum* (Saxena and Cramer 2013). Thus, scientists are gradually recognizing the importance of integrated approach to study the organism at a cellular and organ levels and describe biological regulation of secondary metabolites in a most inclusive way. Among others, plant breeders will be the major beneficiaries of this approach to develop improved cultivars with better productivity.

1.3 Medicinal legumes

Fabaceae is the third-largest family of angiosperms comprising ca. 730 genera and ca. 19,400 species (Lewis et al. 2005). It includes many valuable plants such as crops, vegetables, timbers, ornamental and medicinal plants (Maesen and Somaatmadja 1992; Gepts et al. 2005; Brink & Belay 2006; Saslis-Lagoudakis 2011). More than 50

legume species belonging to 38 genera are used in local health traditions (Arora and Chandel 1972). Many of them possess unique secondary metabolites with important pharmacological activities (Southon 1994; Dixon and Summer 2003). This includes alkaloids, polyphenols, phytoestrogens, phenolic compounds, phytic acids, saponins, lupin, quinolizidine alkaloids and isoflavones (Lopez et al. 2004; Gonzalez-Castejon and Rodriguez-Casado 2011). Investigations have revealed high structural diversity among legume secondary metabolites comprising more than 21,000 alkaloids, 700 non-protein amino acids (NPAA), 200 cyanogenic glucosides and glucosinolates, 20,000 terpenoids, 10,000 polyphenols, 1500 polyacetylenes and fatty acids, 750 polyketides, and 200 carbohydrates (Wink 2013). They exhibit wide ranging properties such as antioxidant, anti-inflammatory and prevention of obesity, aging, diabetes and heart related ailments (Sharma et al. 2011; Prakash et al. 2011). Isoflavonoids are particularly widespread in the Papilonoideae subfamily (Wink 2013). They reveal estrogenic, antiangiogenic, antioxidant and anticancer activities (Dixon and Ferreira 2002) and also useful in chemoprevention of osteoporosis and complications associated with postmenopausal and cardiovascular diseases (Alekel et al. 2000). Legumes (soybean and chickpea) containing certain prenylated isoflavanone such as daidzein, genistein and biochanin (Dixon et al. 2003) have been used by the Zulu's of South Africa for the treatment of impotency and erectile dysfunction (Drewes et al. 2002).

1.4 An overview of *Mucuna pruriens* (L.) DC.

1.4.1 Botanical description

Mucuna pruriens (L.) DC. is a self-pollinated tropical legume classified within the phaseoloid clade of Leguminosae (Sathyanarayana et al. 2016). It has a diploid genome with 22 pairs of chromosomes (2n=2x=22) (Sastrapradja et al. 1974) and an estimated genome size of 1361 Mbp (Sathyanarayana et al. 2017). Botanical description and general characters of the species are presented in the Tables 2 and 3. Unique feature of the plant is densely covered pod hairs with persistent pale-brown or grey trichomes that cause intense irritation to human skin. The chemical compounds responsible for itching property have been identified as two proteins *viz*., mucunain and serotonin (Agharkar et al. 1991).

Table 2 Botanical classification of *Mucuna 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	*Mucuna pruriens* (L.) DC.
Variety	*Mucuna pruriens* var. *pruriens* (L.) DC.
Variety	*Mucuna pruriens* var. *utilis* (Wall. ex Wight) Baker ex Burck
Variety	*Mucuna pruriens* var. *hirsuta* (Wight and Arn.)
Common name	Cowhage, Cow-itch, Velvet bean, Bengal bean, Itchy bean, Krame, Picapia, Chiporra, Buffalo bean, Devil bean

Table 3 General characters of *M. pruriens*

Vegetative Characters	
Roots	Fleshy, usually well nodulated and produced near soil surface.
Stems	The long trailing stems are rather slender and sparsely pubescent.
Growth pattern	Indeterminate
Growth habit	Indeterminate climber
Twining tendency	Pronounced.
Leaves	Numerous, Alternate, Trifoliate, lanceolate/ovate to lanceolate
Leaf texture	Coriaceous/ Membranous
Plant hairiness	Glabrous/ Short appressed hairs/ Pubescent/ Hirsute
Floral and Fruit characteristics	
Flower colour	Purple or Yellowish white
Inflorescence	Long cluster raceme
Pod colour	Dark Green or silvery grey
Pod curvature	Slightly curved/ Curved/ 'S' Shaped
Pod pubescence colour	Golden Orange/ Velvety Black/ Creamish white
Pod sensation	Itching or Non-Itching
Seed colour	Brown, Black, White, Dark Brown, Grey
Seed shape	Oval, Cuboid, Round
Seed coat patterns	Black /Brown mottled/ Absent
Seed yield	12.83-200.92 g/100seeds
Chromosome Number	
2n=2x=	22

1.4.2 Distribution and agronomic potential

M. pruriens came from eastern India or southern China (Burkill 1966; Duke 1981; Wilmot-Dear 1984). It is now widely distributed in tropical and subtropical regions of the world including Asia, America and Africa (Fung et al. 2011; Kumar and Saha 2013) with predominant presence in southeast Asia mainly in India, Bangladesh, Sri Lanka and Malaysia (Fung et al. 2011; Kumar and Saha 2013). In India, it is found in wild from Himalayas to cape camorin and up to 3000 feet elevation in the hills (Arora 1991). The plant occurs in bushes and damp places of scrap jungles. It grows well in

acidic soil (pH<5.8) with annual rainfall > 400 mm and annual temperature 19^0C-27^0C (Kumar and Saha 2013).

M. pruriens offers promising agronomic benefits. It generates seed yield of 2000 kg/hectare (Buckles 1995), thrives well under low fertility soil (Siddhuraju et al. 2000), possess resistance against several diseases (Eilitta et al. 2002), shows allelopathic properties (Fujii et al. 1991) and effective in reducing nematode population (Carsky and Ndikawa 1998; Gowen et al., 2005). Its impact as green manure cover crop (GMCC) is well documented in earlier reports (Eilitta et al. 2003; Jorge et al. 2007). Fast-growing habit of the plant allows groundcover in 60-90 days resulting in large biomass accumulation vis-à-vis other cover crops (Tarawali et al. 1999). Due to high N_2 fixing ability, it is regarded as "example of green manures contribution to the sustainable agricultural system" (Buckles 1995). Seeds are also traditionally used as minor food by many indigenous communities in Asia and Africa (Iyayi and Egharevba 1998, Diallo et al. 2002) and seed-powder can be beneficially supplemented with livestock feed (Burgess et al. 2003). It is also a source high value industrial starch (Betancur-Ancona et al. 2002; Lawal 2004). Oil from the seed is used in the preparation of paint, polish, resin, dye, wood varnish, skin cream and soap (Ajiwe et al. 1997).

1.4.3 Therapeutic properties

Mucuna species in general and *M. pruriens* in particular have attracted great demand from pharmaceutical industries (Siddhuraju et al. 2000; Siddhuraju and Becker 2001a; Bressani 2000). This is largely due to presence of L-Dopa – the drug widely used for the treatment of Parkinson's disease and stimulate aphrodisiac actions (Pulikkalpura et al. 2015). Various constituents that support anti-parkinsonian activity of levodopa

in *Mucuna* seed is presented in Table 4. The seed power also reveals anti-diabetic (Horbovitz 1998), anti-inflammatory, antimicrobial (Sofowora 1982) neuroprotective (Misra and Wagner 2007), anti-venom (Guerranti et al. 2008) and antioxidant (Bravo 1998) properties (Sathiyanarayanan and Arulmozhi 2007).

Table 4 Constituents of *M. pruriens* supporting anti-parkinsonian activity (Kasture et al. 2013)

Constituents	Pharmacological action/use	References
Genisten	Inhibits dopa decarboxylase	Umezawa et al. 1975
	Neuroprotective	Baluchnejadmojarad et al. 2009
	Tyrosine kinase inhibitor	Dunne et al. 1998
	Improves neuronal plasticity	Mandel et al. 2005
Phytic acid	Iron chelator	Graf and Eaton 1990
	Suppresses MPTP induced hydroxyl radical generation	Obata 2003
Glutathione	Free radical scavenging	Kidd 1997
Nicotine	Reduces levodopa induced dyskinesia in rat model of PD	Quik et al. 2009; Huang et al. 2011
Bufotenine, DMT, 5MeO-DMT	5-HT1A agonist, reduces dyskinesia	Muñoz et al. 2008; Pytliak et al. 2011; Riahi et al. 2011
β-carboline	Neuroprotective, antioxidant, MAO inhibitor facilitates activity of DMT	Maher and Davis 1996; Ruscher et al. 2007; Moura et al. 2007
Stearic acid, oleic acid, linolenic acid	Neuroprotective: Activates peroxisome proliferator-activated receptor-gamma (arrest PD progression)	Lauritzen et al. 2000; Wang et al. 2006; Bousquet et al. 2008; Schintu et al. 2009
Lecithin	Decreased confusion, hallucinations and nightmares	Barbeau 1980
Gallic acid	Neuroprotective	Lu et al. 2006; Sameri et al. 2011
Coenzyme- Q10	Slows functional decline in early stage of PD	Shults et al. 2002
Harmine	Glutamate receptor antagonist	Serrano-Dueñas et al. 2001; Kari et al. 2009
L-tyrosine, L-tryptophan, serotonin	Nutritional value	Karobath et al. 1971; Charlton and Crowell 1992; Zeevalk et al. 2007; Hinz 2009

1.5 L-Dopa - The key secondary metabolite

L-3,4-Dihydroxyphenylalanine, commonly known as L-Dopa or levodopa (Chemical abstract service number 59-92-7) is a naturally occurring non-protein amino acid produced in animals and a few plant species (Kuklin and Konger 1995). It has chemical formula $C_9H_{11}NO_4$, giving a molar mass of 197.19 g/mole. The structure is characterized by the presence of two OH, NH_2 and a single COOH as a functional group attached with main skeleton (Fig. 1).

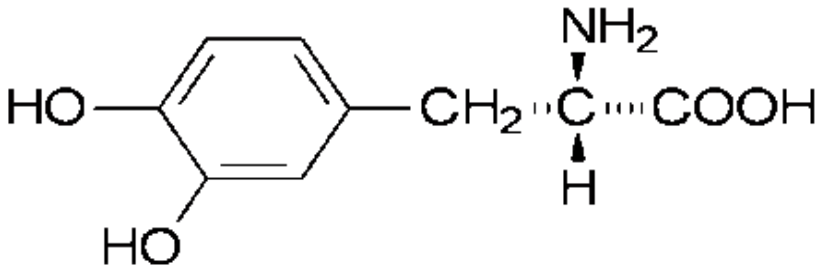


Fig. 1: Chemical structure of 3 - (3^I, 4^I – dihydroxyphenyl) - L - alanine (L - Dopa)

More importantly, L-Dopa is a precursor to dopamine and is considered as gold standard in the management of Parkinson's disease (PD) and dystonia (Calne and Sandler 1970; Khor and Hsu 2007; Abbott 2010). It is administered to patients with Parkinsonism to improve their motor functions. L-Dopa can overcome the blood brain barrier, and is converted into dopamine in the central nervous system by dopa carboxylase (Pulikkalpura et al. 2015). In PD patients, it is administered intravenously or orally and transferred through the blood circulatory system to the dopamine deficient cells in the brain, where it is converted to dopamine for neurotransmission across synapses with motor neurons (Swiedrych et al. 2004). Moreover, L-Dopa can recover spermatogenic loss which makes *M. pruriens* seeds the treatment of choice for infertility (Hornykiewicz 2002; Singh et al. 2013).

1.5.1 Distribution and biological role in plants

L-Dopa is reportedly found in at least 44 plant families (Daxenbichler et al. 1971; Table 5) where it acts mainly as an allelochemical to deter the competition with other plants and as an anti-herbivore (Awang et al. 1997; Nishihara et al. 2004).

Table 5 Sources of L-Dopa in plants

Sl. No.	Plants	Parts	L-Dopa (%)	References
1	*Mucuna pruriens*	Seeds	9.1%	Sathyanarayana et al. 2016
2	*Canavalia gladiata*	Seeds	4.22%	Gautam et al. 2012
3	*Bauhinia variegata*	Seeds	2.91%	Gautam et al. 2012
4	*Acacia leucophloea*	Seeds	2.39%	Gautam et al. 2012
5	*Entada scandens*	Seeds	1.67%	Gautam et al. 2012
6	*Sesbania bispinosa*	Seeds	4.25%	Gautam et al. 2012
7	*Tamarindus indica*	-	3.78%	Gautam et al. 2012
8	*Prosopis chilensis*		-	Ramya and Thaakur 2007
9	*Dalbergia retusa*	Seed	2.20	Hussain and Manyam 1997
10	*Glycine wightii*	Seed	0.20	Krishnan Marg 2003
11	*Alysicarpus rugosus*	Seed	0.65	Krishnan Marg 2003
12	*Cassia hirsuta*	Seed	2.37-2.82	Hussain and Manyam 1997

The herbicidal effects of L-Dopa has been tested in many plant species (Soares et al. 2014). Artificial treatment using seed extracts have shown changes in gene expression involved in amino acid metabolism, oxidative stress, melanin synthesis and lignification (Golisz et al. 2011). These changes appear to alter amino acid metabolism and iron homeostasis, important for many fundamental biological processes such as photosynthesis (Golisz et al. 2011). It is also possible that the action may be related to molecular mimicking of L-Dopa as an amino acid analogue with its subsequent incorporation into proteins (First 2011; Moor et al. 2011).

1.5.2 Chemistry and biosynthesis

Much of our understanding on L-Dopa biosynthesis has come from the study of human system where the main path to L-Dopa formation is via hydroxylation of tyrosine residues by the copper-containing enzyme tyrosine hydroxylase in the

company of molecular O_2 (Pulikkalpura et al. 2015). The shikimic acid pathway feeds this progression by converting simple carbohydrate precursors to aromatic amino acids, particularly tyrosine (Soares et al. 2014). L-Dopa further serves as a precursor to other neurologically important molecules such as dopamine, adrenaline, noradrenalin (George 2006) and melanin, which is present in both animals and plants (Cerenius and Söderhäll 2004; Muller et al 2007; Soares et al. 2014). The compound can also suffer decarboxylation by tyrosine decarboxylase, resulting in tyramine synthesis (Kulma and Szopa 2007) (Fig. 2a and b). The L-Dopa biosynthesis pathway in plants is reported to be analogous to that in mammals (Kwang-Hoon 1998).

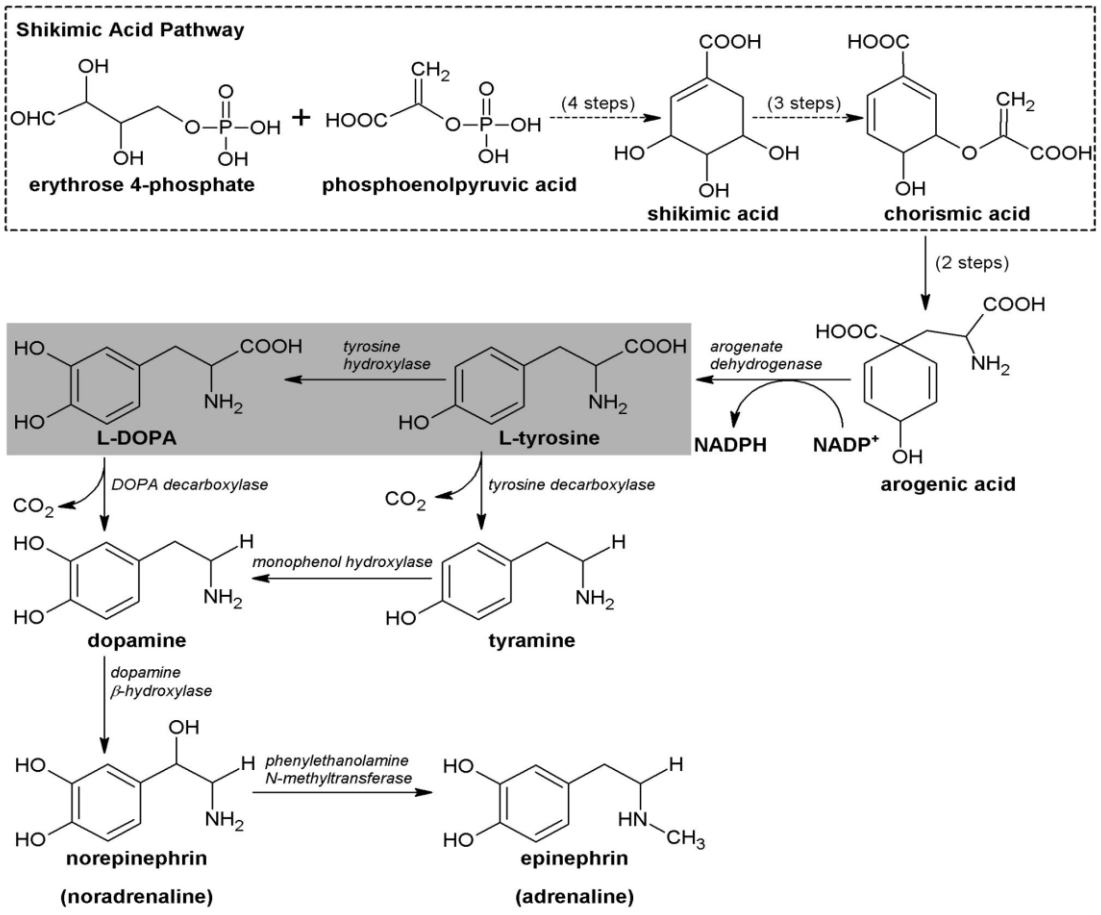


Fig. 2a: L-Dopa metabolic pathway (Soares et al. 2014)

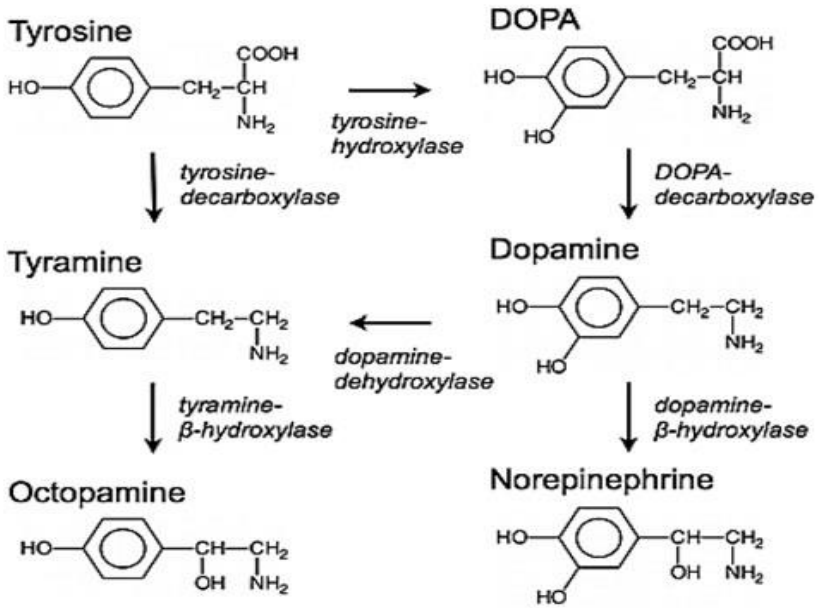


Fig. 2b: Biosynthesis of dopamine via L-Dopa (Barron et al. 2010)

1.5.3 Extraction and commercial production

Several attempts were made in the past to produce L-Dopa on a commercial scale. This includes production from Fungi (Sahi et al.1969; Haneda et al. 1971), yeast species (Ali et al. 2007; Doaa et al. 2010) and extraction from different plant sources such as *Vicia faba* (Guggenheim 1913), *Musa* spp. (Nagatsu et al. (1972), *Ginkgo biloba, Portulaca oleracea* (Simopoulos 2004), *Acacia nilotica* (Vadivel and Biesalski 2012), *Sesbania bispinosa* and *Entada scandens* (Gautam et al. 2012) - which were all met with limited success.

Among the plants, higher amount of L-Dopa was noticed only in *Mucuna s*pecies (Gautam et al. 2012; Ingle 2003). After this discovery, several groups attempted L-Dopa production through cultured cells of *M. pruriens* with varying degrees of success (Brain 1976; Wichers et al. 1985; Huizing et al. 1985; Huang et al. 1995; Huang and Hu 2001). Presently, large scale production of L-Dopa is being carried out by chemical synthesis. However, *M. pruriens* extract reportedly works faster and

better than synthetic L-Dopa in the treatment of Parkisonism (Hussain and Manyam; 1997; Manyam et al. 2004a 2004b). This makes the plant very important source of L-Dopa for commercial purification.

1.6 Genetic studies

Little is known on the genetics governing L-Dopa production in *Mucuna* spp. Three studies earlier attempted to investigate GXE interaction (GEI) effects on L-Dopa production. Lorenzetti et al. (1998), in their study, where latitude was used for environmental factor, found both environmental and genotypic factors responsible for L-Dopa production. This was supported by St-Laurent et al. (2002) who found marginal impact of latitude, but concluded that other factors were influential too. On the contrary, Capo-chichi et al. (2003b) and Mahesh and Sathyanarayana (2011b) found genotype had greater influence on L-Dopa production, whereas, GEI effect was minimal when compared to genotype/accession main effect.

Beyond this, so far no work has been carried out to analyze the genetics of L-Dopa trait in *M. pruriens* and generate resources for future breeding program. Thus, the present study was carried out with a primary objective to analyze the genetic and diversity aspects of L-Dopa trait in *M. pruriens* including heritability analysis.

1.7 Micropropagation of elite genotypes

Clonal propagation through tissue culture is a mass propagation method that can be achieved in short time and space. It is an alternative method of propagation (George and Sherrington 1984) and is being widely used for the commercial propagation of a large number of plant species including many medicinal plants (Rout et al. 2000a; Nalawade et al. 2003). The main advantage of this method is, its ability to regenerate disease free, true to type plants at high frequency from meristems, shoot tips, axillary

buds or nodal segments (Mallick et al. 2012). This technology could be an affordable and cost effective means of high-end production of elite planting materials throughout the year without any seasonal constraints. In view of this, mass multiplication protocol using axillary bud explants have been standardized for one of the elite parental lines *viz*., 500240SK (*M. pruriens* var. *pruriens*) as part of this study.

1.8 Amplified Fragment Length Polymorphism (AFLP) Markers

Molecular markers are powerful tool to asses and characterize the genetic resources (Caetamo-Anollés and Gresshoff 1998; Joshi et al. 2011). Of the several molecular markers available, ALFP has the capacity to discriminate closely related populations (Lall et al. 2010) and ability to detect large number of polymorphic loci in a single analysis (Garcia et al. 2004). AFLPs have also been applied earlier in different plant species for genome mapping (Zimnoch-GuZowska et al. 2000), DNA fingerprinting (Powell et al. 1996), genetic diversity studies (Russell et al. 1997) and parentage analysis (Gerber et al. 2000; Lima et al. 2002). In case of diversity analysis, AFLPs are particularly chosen as it covers large area of the genome (Karp and Edwards 1995) and are extremely proficient in revealing diversity at the infra-specific level (Karp and Edwards 1995; Capo-chichi et al. 2001). Therefore AFLP markers were specifically employed for genetic diversity analysis in the present study.

2. Review of Literature

2.1 Medicinal legume *Mucuna pruriens* (L.) DC.

Many legume species contain organic chemicals in sufficient quantity to be useful as raw materials for breakthrough scientific, technological and commercial applications. *M. pruriens* has fair share of reference in Ayurveda, traditional Chinese and Persian medicine (Oudhia 2002). The species is best known natural source of L-Dopa - widely used in the treatment of the Parkinson's disease and infertility (Haq 1983; Abbott 2010; Hornykiewicz 2002; Singh et al. 2013). On the other hand, L-Dopa is an anti-nutritional factor (ANF) in edible varieties (Pugalenthi and Vadivel 2007b). Therefore development of low/high L-Dopa containing varieties will render *Mucuna* breeders and farmers with an invaluable genetic resource for future cultivation.

The northeastern (NE) region of India comprising eight states of Sikkim, Arunachal Pradesh, Assam, Meghalaya, Manipur, Tripura, Mizoram and Nagaland is topographically categorized into eastern Himalayas, northeast hills and Brahmaputra and Barak valley plains (Chatterjee 2008). *M. pruriens* is thought to be native of eastern India (Burkill 1966, Wilmot-Dear 1984) which includes parts of northeast India. The region display abundant distribution of wild *M. pruriens* represented by var. *pruriens* and there is a long history of cultivation of velvet bean (*M.pruriens* var. *utilis*) in the region by several indigenous groups such as Khasi, Naga, Kuki, Jaintia, Chakma and Mizo etc. who have used it as a minor food for centuries (Arora 1991). Consequently, considerable work has been done on evaluating nutritional and anti-nutritional properties of this tribal pulse - the details of which are available in a review (Pugalenthi et al. 2005). On the other hand, development of superior varieties

either through conventional breeding or biotechnological methods has received little attention (Sathyanarayana et al. 2016). In view of this, the present thesis work attempted to address this gap through two connected objectives. Firstly, to collect and characterize *M. pruriens* germplasm from northeast India – the centre of origin of this crop to understand the diversity represented in the region and secondly, to conduct the genetic analysis of L-Dopa trait using F2 population to generate baseline data on this economic trait. Additionally, we demonstrate micropropagation protocol using a representative genotype to aid mass multiplication of the elite genotypes.

2.1.1 Germplasm collection and characterization

Gene pool provides basic genetic resource for selection and improvement of a crop species through targeted breeding program. Germplasm of *M. pruriens* is available in a few research institutes/organizations across the world (Sathyanarayana et al. 2016). This includes: United States Department of Agriculture (USDA 1994), International Institute of Tropical Agriculture (IITA), Nigeria, Centro Internacional de Agricultura Tropical (CIAT), Colombia, AVRDC - The World Vegetable Centre, Taiwan; National Biological Institute (NBI), Indonesia (Jorge et al. 2007) etc.

In India, National Bureau of Plant Genetic Resources (NBPGR), New Delhi (Jorge et al, 2007; Raina et al. 2012), Indian Institute of Horticultural Research (IIHR), Bengaluru (Mamatha et al. 2010), Jawaharlal Nehru Tropical Botanical Garden and Research Institute (JNTBGRI), Thiruvanantpuram (Padmesh et al. 2006), Zandu Foundation for Health Care, Valsad, Gujarat (Krishnamurthy et al. 2005), Arya Vaidya Sala, Kottakkal; Bharathiar University, Coimbatore (Siddhuraju and Becker 2005) etc. are reportedly maintaining the *M. pruriens* germplasm although the number

of collections available with these institutes is not clearly known except 182 accessions reported in case of NBPGR, New Delhi (Sathyanarayana et al. 2016).

However, till date no effort has been made to collect germplasm from northeastern region of India which is part of larger centre of origin of this species. Besides, characterization details for many of these germplasm are not available except a few isolated reports available in literatures. The significant outcomes from these studies are summarized below;

M. pruriens exhibits considerable variability in pod, flower and seed characteristics. Earliest indications on the factors governing these traits have come from the pioneering works of Lubis and co-workers during 1970s and 80s. *M. pruriens* pod hairs produces two distinct phenotypes - long rough ones and short smooth ones. Their work revealed that the combination of size, cell wall surface and type of pod hairs serve as diagnostic marker for species differentiation. They also showed that, the trait is controlled by two genes *viz*., R and N, and abnormalities generated are due to recessive alleles in homozygous condition (Lubis et al. 1979). Likewise, studies on flower colour demonstrated that, of the two flower colours produced by the species, purple is dominant over the white and the genes controlling them are present on a single locus (Lubis et al. 1978). In case of seed coat colour, their work confirmed that the range of phenotypes - dark black to white including different shades of brown and mottled ones are due to multigenic factors (Lubis et al. 1980). These works provide earliest clues on the genetics of some of the key morphological traits in *M. pruriens*.

Some independent studies have reported germplasm evaluation in *M. pruriens* from different parts of India and World. Important observations from these investigations are highlighted below:

Bennet-Lartey (1998) screened 11 accessions of *M. pruriens* var. *utilis* from Ghana for 5 qualitative and 11 quantitative traits and reported good diversity for phonological and morphological traits. The study identified early maturing genotypes that can be used for seed production.

First such work in India was reported by Gurumoorthi et al. (2003), who carried out evaluation of seven *M. pruriens* var. *utilis* accessions from Western Ghats of India for important traits such as - fertility index, seed recovery percentage, days to flowering, harvest index, leg hemoglobin and chlorophyll contents, besides profiling macro and micro nutrients in soil samples. The study revealed good variability for these traits in Indian accessions and established beneficial impact of *M. pruriens* on soil fertility.

Krishnamurthy et al. (2005) reported variability in seed yield and other agronomic traits in 19 *M. pruriens* accessions and identified early and late flowering ones among both the itching and non-itching types. In addition, they also identified superior genotypes producing bold seeds, higher seed weight and favourable physico-chemical properties. It was revealed that the itching trichome lines possessed higher L-Dopa (4.36-6.12%) content over non-itching lines (2.30-4.18%) possibly due to selection.

Mamatha et al. (2006) evaluated 13 *M. pruriens* accessions and identified genotypes with higher nitrogen fixing ability. The results showed positive correlation between nitrogen fixing ability with traits such as nodule number, nodule biomass and L-Dopa content in seeds.

Pugalenthi and Vadivel (2007a) studied 11 *M. pruriens* var. *utilis* accessions from South India for yield and vegetative characters and reported significant variability and its implications on the breeding programs.

Mamatha et al. (2010) evaluated several *M. pruriens* var. *utilis* accessions for production of biomass, active principle and seed yield. The study revealed that growth parameters did not vary significantly across the genotypes, but yield parameters did. Accession IIHR MP 09 recorded highest number of pods per bunch, number of seeds per pod and branches per plant. Other accessions: IIHR MP 03 recorded highest pod length (cm) and pod weight (g) and IC 33243 recorded highest 100 seed weight (g).

Mahesh and Sathyanarayana (2011) carried out stability analysis of L-Dopa trait by planting selected *M. pruriens* accessions across different locations in Karnataka varying in environmental conditions. L-Dopa content from different locations ranged from 1.564% to 3.641%. Despite this variation, the ranking of the genotypes remained constant without any crossovers. The combined ANOVA results indicated that the effects of genotypes and environments were significant at 0.05 probability levels. Based on the results, it was suggested that *Mucuna* genotypes are relatively stable across the environments.

Raina and Dutta (2012) evaluated five seed characters in 38 *M. pruriens* accessions. Seed colour showed wide variability ranging from creamy white to dark mottled ones. Seed length varied from 0.80 to 1.70 cm, thickness 0.40 to 1.10 cm and width from 0.50 to 1.30 cm. 100 seed weight (HSW) ranged from 20.80 to 170.40 g. Correlation analysis of seed characters indicated significant positive correlation between HSW and seed length, thickness and width.

Sathyanarayana et al. (2012) evaluated 40 *M. pruriens* (31 wild and 9 cultivated) accessions using 30 mopho-agronomical characters. Results of PCA for 22 quantitative and eight qualitative characters revealed that the two PCs explained 65.18% and 67.50% of the total variation, which were drawn across the range of

traits. Cluster analysis demonstrated high morphological diversity in case of wild accessions (63.05%) as compared to cultivated types (36.95%).

Onyishi et al. (2015) evaluated 15 *M. pruriens* accessions in Federal University of Technology, Owerri (FUTO). The results showed significant positive correlation between pod and seed weight (0.874). The seed weight and pod weight also positively correlated with number of branches (0.715) and percentage emergence (0.631), but negatively correlated with number of days to plant emergence (-0.438), days to flowering (-0.435) and number of days to fruiting (-0.465). This indicated that the seed and pod weight has significant statistical linkage with these traits which suggests possibility of achieving simultaneous improvement of the traits.

Kumar (2016) carried out analysis of six *M. pruriens* accessions using nine quantitative traits. Variability in 100 seed weight was significant in two genotypes and the variability in hilum length remained almost insignificant. This study revealed adequate genetic variability within the germplasm evaluated.

Singh and Dhawan (2017) analyzed trichomes and itching protein - Mucunain characters in *M. pruriens*. Results showed wild varieties possessed largest trichome length (2015±29µM) compared to cultivated genotypes. They found Mucunain is expressed in every stage of plant growth, but its expression was high during maturity (about 170 days). The study also identified unigenes regulating trichome development such as GLABRA-1, GLABRA-2, and cpr-5.

Hadapad et al. (2018) carried out genetic variability and correlation analysis for quantitative and qualitative traits in *M. pruriens* in rubber plantation under hill zone of Karnataka. High heritability along with genetic advance was recorded for a number of

traits including leaves at harvest, days to 50 percent flowering, leaf area at harvest etc. Correlation analysis revealed seed yield per plant showed highly significant and positive correlation with number of trifoliate leaves per plant, leaf area, number of pod per bunch and pod yield per plant.

Patil et al. (2018) evaluated *M. pruriens* genotypes for growth and flowering characters in areca nut plantation under hill zone of Karnataka. Results showed significant differences among the genotypes with respect to growth and flowering attributes. The genotype Arka Dhanvantari recorded maximum plant height (282.03 cm) and number of trifoliate leaves (71.03) at harvest. The days taken for first flowering and 50 percent flowering were minimum (65.73 and 72.33, respectively) and maximum (132.20 and 140.67, respectively) in the genotypes Arka Aswini and Arka Dhanvantari.

2.1.3 Medicinal attributes

Mucuna species has rich history of use in Indian traditional medicine (Oudhia 2002). They have been a major ingredient of over 60 drug formulations sold in Indian market (Mustang, useful in male sexual impotence and Gertiforte, useful in senile 'pruritus' and fatigue). In addition to Parkinson's disease (Tharakan et al. 2007), the seed powder is also used to treat joint pain, irregular menstruation, leucorrhea, spermatorrhea, and erectile dysfunction (Nadkarni 1982; Ding et al. 1991).

Joshi and Pant (1970) reported hypoglycemic effect of *Mucuna* seeds and weak neuromuscular blocking effects of *Mucuna* pods in rats. Several therapeutically valued alkaloids like mucunine, mucunadine, prurienine, prurieninine nicotine, physostinginine, and serotonin have also been isolated in seed extracts (Duke 1981; Mehta and Majumdar 1994).

Abraham (2011) studied the efficacy of *M. pruriens* on the fertility aspects using rat models. The results revealed significant increase in sperm count, motility as well as decrease in the quantity of abnormal sperm in the test mice. The resultant histological changes revealed enhanced effects on target organs which showed changes in the architecture of testicles with densely packed spermatids in the somniferous tubules of the mice. Researchers of the KGMU, Lucknow found further evidence on role of *M. pruriens* in supporting normal functioning of the male reproductive system particularly fertility (Shukla et al. 2010).

More recently, Lampariello et al. (2012) reviewed the medicinal properties of *M. pruriens* and reported its beneficial effects as aphrodisiac, neuroprotective, Anti-diabetic, anti-epileptic, anti-neoplastic, anti-helminthic, analgesic and nervous disorders. A summary of different uses of the *M. pruriens* including medicinal properties is given in Fig. 3.

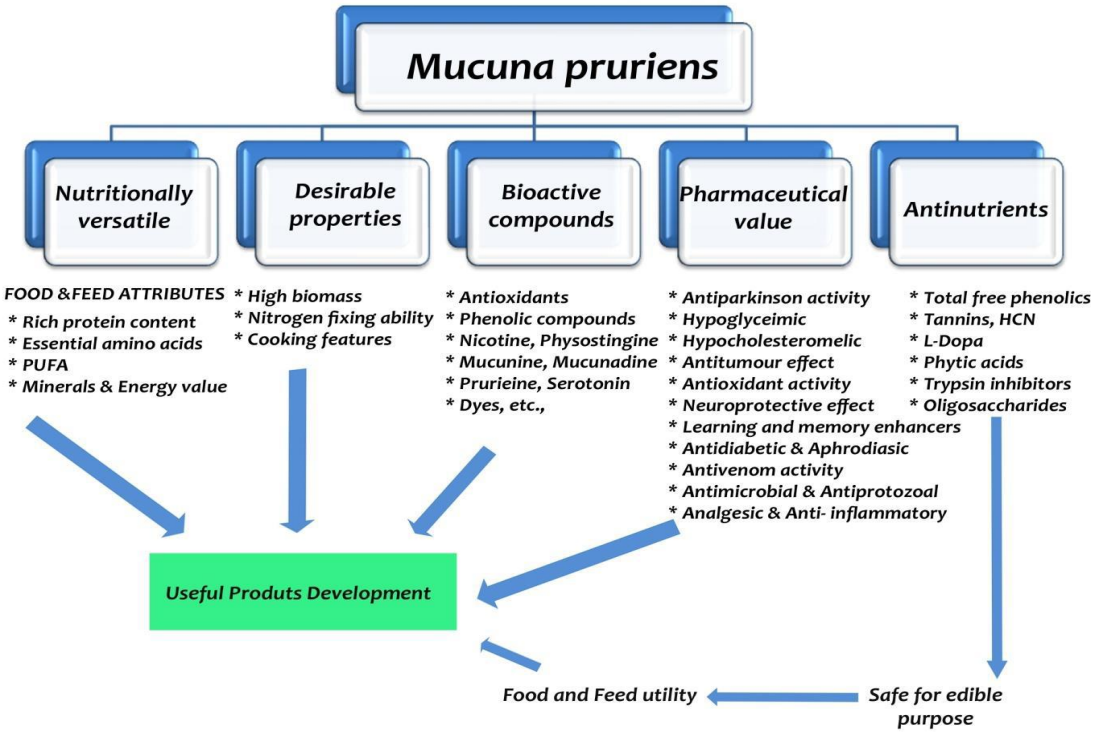


Fig. 3: Overview on uses of *M. pruriens* (Bhat and Karim 2009)

2.2 L-Dopa: Distribution and variability

L Dopa is distributed in a wide range of plant families. Earlier, Daxenbichler et al. (1971, 1972) surveyed 1062 species belonging to 160 plant families and determined that only *Mucuna* species produced L-Dopa in sufficient quantity (3.1% to 6.7%) for commercial extraction. Even though it can be produced more easily by synthetic methods, natural source has an advantage. Katzenschlager et al. (2004) assessed clinical effects and pharmacokinetics of natural L-Dopa vis-à-vis standard carbidopa and showed natural source has distinct edge over synthetic L-Dopa in terms of efficacy for the treatment of Parkinson's disease.

Several evaluation reports indicate great variability for L-Dopa content in *M. pruriens* accessions. Bell and Janzen (1971) showed L-Dopa varies from 5.9 to 9.0 % in *Mucuna* species. Rajaram and Janardhanan (1991) also reported higher L-Dopa content up to 9% in some *Mucuna* seeds. The variability for L-Dopa content in different *Mucuna* species reported by other groups is presented in Table 6. Further, Wichers et al. (1993) analyzed tissues such as root, stem and leaf of *M. pruriens* for L-Dopa content and showed that, besides L-Dopa, the leaves contains dopamine too.

Table 6 Distribution of L-Dopa in *Mucuna* species (Sathyanarayana et al. 2016)

Sl. no.	*Mucuna sp.*	L-Dopa (%)/ Yield	Reference
1	*Mucuna andreana* Micheli	6.3 - 8.9	Bell and Janzen 1971
2	*Mucuna aterrima* (Piper & Tracy)Holland	3.31 - 4.2	Daxenbichler et al. 1972; Pieris et al. 1980
3	*Mucuna birdwoodiana*	9.1	Chem Abstr 1990
4	*Mucuna cochinchinesis* (Lour.) A. Chev.	2.5 - 4.2	Prakash and Tewari 1999; Pieris et al. 1980; Chem Abstr. 1991
5	*Mucuna deeringiana* (Bort.) Merr.	2.7 - 3.13	Miller 1920; Daxenbichler et al. 1972; Pieris et al. 1980
6	*Mucuna gigantea* (Willd.) DC.	1.50 - 3.78	Daxenbichler et al. 1972; Rajaram and Janardhanan 1991
7	*Mucuna holtonii* (Kuntze) Mold.	6.13 - 7.5	Daxenbichler et al. 1972; Bell and Janzen 1971
8	*Mucuna monosperma* DC. Ex Wight	4.24 - 4.56	Arulmozhi and Janardhanan 1992; Mohan and Janardhanan 1995
9	*Mucuna mutisiana* (Kunth.) DC.	3.9 - 6.8	Daxenbichler et al. 1972; Bell and Janzen 1971; Bell et al. 1971
10	*Mucuna pruriens* (L.) DC.	3.8 - 9.16	Damodaran and Ramaswamy 1937; Daxenbichler et al. 1972; Mary et al. 1992; Chattopadhyay et al. 1995a; Mahajani et al. 1996; Prakash and Tewari 1999; Padmesh et al. 2006
		4.36 - 6.12	Krishnamurthy et al. 2005
		1.51 - 6.29	Singh et al. 2008
		5.6	Modi et al. 2008
		7.06 - 8.13	Fathima et al. 2010
		5.38 - 6.98	Kalidas and Mohan 2011
		1.41 - 6.63	Dhawan et al. 2011
		3.29 - 5.44	Raina et al. 2012
11	*Mucuna pruriens* f. *Hirsuta*	1.4 - 1.5	Chattopadhyay et al. 1995
12	*Mucuna pruriens* f.*utilis*	1.8	Chattopadhyay et al. 1995
13	*Mucuna pruriens*	0.945-3.585	Sathyanarayana et al. 2015
14	*Mucuna pruriens* var. *utilis*	3.6 - 8.05	Pieris et al. 1980; Janardhanan and Lakshmanan 1985; Mohan and Janardhanan 1995; Prakash and Tewari 1999; Siddhuraju and Becker 2001
		5.60 - 6.56	Janardhanan et al. 2003
		2.30 - 4.18	Krishnamurthy et al. 2005
		7.55 - 7.93	Kala and Mohan 2010
		4.11 - 6.61	Mamatha et al. 2010
		1.58 - 6.18	Dhawan et al. 2011
15	*Mucuna sloanei* Fawcett & Rendle	3.34 - 9.0	Bell and Janzen 1971; Daxenbichler et al. 1972; Rai and Saidu 1977.
16	*Mucuna urens* (Linn.) Medik.	4.92 - 7.4	Bell and Janzen 1971; Daxenbichler et al. 1971; Daxenbichler et al. 1972
17	*Mucuna atterrima*	4.5%	Pieris, N , 1980
18	*Mucuna pruriens*	24 g/DW	Chikagwa-Malunga, S.K 2009
19	*Mucuna monosperma*	5.48% DW	Inamdar 2012
20	*Mucuna pruriens*	3.54%DW	Teixeira 2003
21	*Mucuna pruriens* var. *utilis*	6.36%w/W	Egounlety (2003)

Conventionally L-Dopa estimation is carried out using spectrophotometric method described by Daxenbichler et al. (1971, 1972). However, recently more accurate methods have been developed using superior techniques such as HPLC and HPTLC.

Modi et al. (2008) developed HPTLC based method with good L-Dopa recovery (100.30%). This was confirmed by Raina and Khatri (2011) with mean recovery of 100.89%. Vachhani et al. (2011) further established efficacy of this method for precise and accurate molecule fingerprinting.

On the other hand, Rathod and Patel (2014) developed RP-HPLC method for L-Dopa estimation. Their results showed good linear correlation coefficient ($r^2 > 0.999$) for calibration plots in the range of 10-80 µg/ml. The report suggests HPLC is a preferred method over HPTLC as it is easy and fast. Also, in HPLC, the solvents used for mobile phase are not hazardous to the column as compared to other buffered mobile phase resulting in long life of the column.

However, Pulikkalpura et al. (2015) cautioned that most of the quantifications of L-Dopa content in *M. pruriens* seeds based on HPTLC/HPLC involves prolonged, multistep extraction procedure in acidic media and therefore the drug is subject to degradation. This needs careful consideration during the large-scale extraction processes.

Based on the above leads, we have used HPLC method for analysis of L-Dopa variability in *M. pruriens* germplasm of northeast India.

2.3 Genetic diversity estimation

Analyzing the genetic diversity in germplasm collections is an important component of the plant breeding program (Azhaguvel et al. 2006). Diversity estimation using morphological traits alone is not reliable owing to influence of environment on their expression (Tatikonda et al. 2009). Molecular markers, on the other hand, are independent of such influences and can be reliably employed for the purpose. Of the different markers used for the diversity analysis, PCR-based markers like RAPD (Williams et al. 1990), ISSR (Zietkiewicz 1994) and AFLP (Vos et al. 1995) have become handy as their use does not require sequence information. Conversely, simple sequence repeats (SSR) or microsatellite is the marker of choice for breeding applications- but its development is expensive (Thiel et al. 2003).

Of the different molecular markers used, AFLP provides an effective means for genetic diversity analysis as it covers large genome area in a single assay (Karp and Edwards 1995). This renders AFLP highly suitable for molecular characterization studies (Williams et al. 1990). Consequently, it has been widely applied in genetic diversity analysis in a several legume spp. such as soybean, dolichos bean and chickpea (Singh et al. 2010; Venkatesha et al. 2010; Sudupak et al. 2004) as well as in non-legume species such as mangroves (Jianab and Shib 2009), banana (El-Khishin et al. 2009) and mustard (Weerakoon et al. 2010).

Several earlier studies have established pre-eminence of molecular markers for diversity estimation in *Mucuna* species. A wide range of markers including RAPD, AFLP, ISSR and SSRs have been used (Capo-chichi et al. 2001, 2003a, 2003b, 2004; Krishnamurthy et al. 2005; Kumar et al. 2006; Padmesh et al. 2006; Gupta and Kak

2007, George et al. 2007 Sathyanarayana et al. 2008; 2017). The brief summary of the results of these studies are given below.

Capo-chichi et al. (2001) analyzed genetic diversity of 40 *M. pruriens* accessions using AFLP markers. The results revealed narrow genetic base (3-13%) among the US landraces. A more comprehensive analysis involving 64 accessions from a wider geographical region was implemented by the same authors (Capo-chichi et al. 2003a) which revealed enhanced genetic diversity (0-32%).

In India, Padmesh et al. (2006) carried out first genetic diversity analysis involving 13 *M. pruriens* (both var. *pruriens* and var. *utilis*) accessions collected from the Western Ghats of India. Their results showed good diversity (10-61%) in the overall germplasm collection. The study also revealed var. *pruriens* was genetically more diverse than var. *utilis*.

Subsequently, Gupta and Kak (2007) used seed protein profiling for genetic diversity analysis among 20 *M. pruriens* accessions. The analysis carried out using SDS-PAGE resulted in total of 11 bands with molecular weight range between 15 KDa - 200 KDa. A similarity coefficient of 65% was recorded in the study.

Kalidass and Mohan (2010) analyzed genetic diversity of five *M. pruriens* collection from Western Ghats of India using five RAPD markers. Out of 43 amplified products, 28 showed polymorphism (65.12%). Similarity coefficient grouped these accessions into two clusters. The study indicated sufficient diversity within these clusters.

Leelambika et al. (2010) collected 18 *M. pruriens* accessions from different geographical locations across India and NBPGR, New Delhi and carried out comparative genetic diversity analysis using morphometric, biochemical, isozyme and

RAPD approaches. The results from DNA/protein markers were in consensus with the morphometric analysis. The clustering was largely based on taxonomic affiliations.

Sathyanarayana et al. (2011) analyzed genetic diversity among 25 accessions of Indian *Mucuna* species using AFLP. The Jaccard's similarity coefficient values (0.087 to 0.332) suggested good variability among the accessions indicating their utility in the breeding programs. The UPGMA and PCoA results showed grouping based on both genetic as well as phonetic relationships.

Mahesh and Sathyanarayana (2015) collected 35 *M. pruriens* accessions from diverse geographical locations of India and studied intra-specific variability for salinity tolerance using combination of trait phenotyping and AFLP markers. Superimposing diversity estimates with the phenotyping data revealed 19 divergent parental combinations useful for the breeding as well as mapping of this trait.

Patil et al. (2016) reported genetic diversity among 59 accessions sampled from six species and three botanical varieties of *M. pruriens* using RAPD, ISSR and their combined datasets. The results demonstrated high inter-specific variation than intra-specific variation in the genus *Mucuna*.

Leelambika et al. (2016) combined metabolic and genomic profiling of selected *M. pruriens* accession to identify parents for L-Dopa breeding. Based on the results, the authors identified seven promising parental combinations that can be used for genetic improvement as well as developing mapping population.

In a first ever report on the development of co-dominant markers in *M. pruriens*, Shetty et al. (2015) used EST sequence data from related legume species to develop microsatellite markers. For this, they analyzed total of 2,86,488 EST sequences from four legume species *viz*., *Vigna unguiculata*, *Glycine max*, *Phaseolus vulgaris* and

Cicer arietinum which generated 22,457 SSRs comprising of 29.08% di-, 46.41% tri-, 7.29% tetra-, and 17.22% penta-nucleotide repeats. Based on 22,457 SSR containing sequences, 522 primer combinations were designed and 50 were screened against a diverse panel of 25 genotypes of which three produced polymorphic profiles with an average PIC of 0.65.

Recently in an effort that can provide greater impetus to genomics assisted breeding in *M. pruriens*, Sathyanarayana et al. (2017) reported *de novo* transcriptome assembly and differential gene expression among leaf, root and pod tissues. In addition they have also proposed large number of microsatellite markers which are validated on a panel of diverse *M. pruriens* accessions.

Building on these leads, the present thesis aimed at DNA marker (AFLP) based genetic diversity estimation among *M. pruriens* collection from northeastern region to get preliminary insight into diversity represented in its centre of origin.

2.4 Genetic studies on L-Dopa and other traits

Little information is available on the genetics of L-Dopa content in *M. pruriens*. Initially, L-Dopa trait was considered to be both environmentally and genetically controlled with production highest near the equator, due to high light intensity (Lorenzetti et al. 1998; St.-Laurent et al. 2002). Recent studies contradicted the importance of latitude on L-Dopa expression (Capo-chichi et al. 2003b) and confirmed preponderance of genotypic effects (Mahesh and Sathyanarayana 2011). It has also been shown that, early maturing types show low levels of L-Dopa vis-à-vis late maturing types (Capo-chichi et al. 2003b). The colour of the seed is also related to L-Dopa content; darker seeds possess more L-Dopa than light colored seeds (Siddhuraju el al., 2000).

Krishnamurthy et al. (2002) evaluated 17 parental lines of *M. pruriens* for seed associated characters including L-Dopa content and subjected them to hererosis analysis. After several backcross generations, they shortlisted 37 F1 lines for heterosis (%) analysis over the mid parent, better parent and best parent on the seed yield character. High magnitude of heterosis (40-96%) over the best parent was observed in four F1 lines.

Mahesh et al. (2016) elucidated the genetic control of agronomic traits through an AFLP based intraspecific genetic linkage map of Indian *M. pruriens* using F2 progenies (200) derived from a across between wild and cultivated genotypes. The results have opened up newer avenues for integrating molecular breeding tools into ongoing *M. pruriens* genetic improvement programs.

Recently Chinapolaiah et al. (2017) carried out heterosis and combining ability analysis for yield and yield contributing traits in *M. pruriens* to determine the gene action and to enable selection of superior parents for hybridization. The results revealed that magnitude of *sca* variance was greater than *gca* variance, suggesting the predominance of non-additive gene action for most of the traits, except days of 50% flowering, number of flowering inflorescence, number of pod per bunch and 100 seed weight. The results identified the best parental combinations for heterosis breeding.

In this backdrop, the present study was undertaken with the main objective to detect the trait association and heritability of L-Dopa content using F2 population derived from the cross between early lineages (F1) and those adapted for having low L-Dopa content as well as genotypes not adapted to this system having high L-Dopa content. The aim of the research was to use them in the selection for future *Mucuna* breeding programs on the L-Dopa content.

2.5 Micropropagation

In vitro culture studies in *M. pruriens* started with an effort to produce L-Dopa using cell culture methods. Huizing et al. (1985) first demonstrated presence of L-Dopa in the cell suspension cultures of *M. pruriens* by means of HPLC and Thin Layer chromatography (TLC). Wichers (1985, 1993) found that addition of 2,4-D to the culture medium suppressed L-Dopa production. Subsequently, Huang et al. (1995) reported that high yields of L-Dopa under *in vitro* condition is associated with precise regulation of pH, concentration of Indole-3-Acetic Acid (IAA) and appearance of yellowish or brown cell colour which reflected melanin production. Huang and Chen (1998) found that efficient production of L-Dopa is possible through two stage configuration. In majority of these studies, Murashige and Skoog (MS) medium supplemented with 2,4-D was found to be best medium for callus induction. However, L-Dopa accumulation was more on medium supplemented with 1 mg/L IAA + 1 mg/L BA (Kavitha and Vadivel 2005). Raghavendra et al. (2012) compared difference in L-Dopa production in response to elicitor treatment and precursor feeding. The results showed, precursor feeding triggers higher L-Dopa concentrations and tyrosine activity. Nonetheless, *in vitro* production of L-Dopa from cell suspension culture of *M. pruriens* so far has met only with marginal success.

Earliest work on micropropagation in *M. pruriens* was reported by Chattopadhyay et al. (1995). They cultured hypocotyl, epicotyl and cotyledon explants on revised Tobacco (RT) medium containing NAA (2.7 μM) and 2iP (9.8 μM). Maximum shoot regeneration was achieved from explants of 6-7 day old seedlings. The researchers were successful in achieving rooting on half-strength RT medium in more than 90% of the regenerated shoots by the supplementation of NAA (2.7 μM).

Later, Faisal & Siddique (2006 a and b) developed protocol for multiple shoots from the nodal segments of *M. pruriens*. They used range of cytokinins with cotyledonary node and axillary bud explants from seven day old aseptic seedlings. The authors found that, among the hormones, BA (5.0 µM) was most effective in shoot induction. Media strength and pH were also investigated and highest efficiency of shoot proliferation was observed in BA (5.0 µM) and NAA (0.5 µM) in half-strength MS medium at pH 5.8. The best rooting was obtained on half-strength MS medium supplemented with IBA (2.0 µM).

Sathyanarayana et al. (2008) reported rapid and reliable method for high fidelity micro-propagation in *M. pruriens*. Axillary buds, when cultured on MS + BAP (3.5 µM) produced 6.7 shoots. This number increased up to 16.33 upon re-culturing on the same hormonal medium. The shoots attained elongation on GA_3 (2.89 µM) and produced roots on half-strength MS liquid medium supplemented with NAA (16.20 µM). RAPD method was used to confirm the genetic fidelity of the regenerated shoots.

Raaman et al. (2013) standardized protocol for *in vitro* propagation of *M. pruriens* and extraction of enzyme tyrosinase. Axillary buds, internode, root, seed and leaf were used for callus induction. Results showed maximum green calli were produced using different concentrations of NAA and IBA+BAP. Maximum white calli were produced using different concentration of 2, 4-D in combination with kinetin.

Recently, Vishwakarma et al. (2017) standardized protocol for micropropagation and *Agrobacterium rhizogens* mediated transformation in *M. pruriens* and observed that, kinetin (14µM) produced four shoots after four weeks culture. Rooting was achieved on half strength MS medium with NAA (6 µM). Regenerated plants were hardened

with 75% of survival frequency. Plants were infected by two *A. rhizogens* strains (MTCC-532 and MTCC-2364) for hairy root induction, and effect of manual wounding and acetosyrinzone treatments were tested. Results indicated MTCC-2364 showed hairy roots induction in leaf explants only by manual wounding, whereas in case of acetosyrinzone (50-250µM) mediated method, hairy root induction was observed in all the explants except leaf.

As evident from above results, micropropagation has been considerably successful in *M. pruriens*. Hence, in the present work, we focused on determining best explants and hormonal combination for large-scale multiplication of elite genotypes. We have used accession (500240SK) as representative genotype for our study - the results of which can be extrapolated to other accessions with minor modifications.

3. Objectives

The thesis covered following five major objectives:

1. To establish and characterize *Mucuna pruriens* germplasm from northeast India.
2. To determine L-Dopa variability among these accessions.
3. To estimate their genetic diversity using molecular markers.
4. To identify parental lines for L-Dopa breeding.
5. To develop micropropagation protocol for elite genotypes.

4. Materials and Methods

4.1 Germplasm collection and characterization

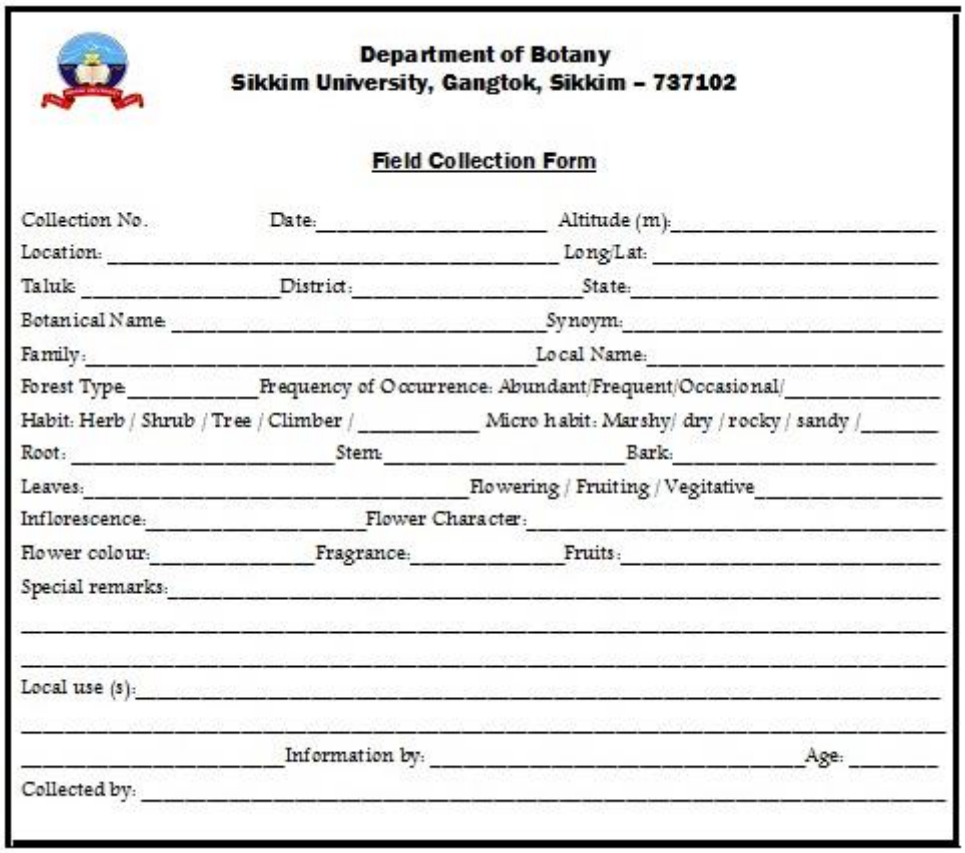
4.1.1 Collection of germplasm

The germplasm collection was carried out during February/March, 2013-2015. For ascertaining the distribution of *M. pruriens* in northeastern region, a detailed assessment of following published floras was carried out;

- Flora of British India (Hooker 1872)
- Legumes of India (Sanjappa 1992)
- Forest flora of Meghalaya (Hardidasan and Rao 1985-87)
- Flora of Arunachal Pradesh (Hajra et al.1996)
- Flora of Tripura (Deb 1981)
- Flora of Assam (Kanjilal 1938)
- Flora of Sikkim (Hooker 1849)
- Flora of Manipur (Deb 1961)
- Flora of Manipur Volume I (B.S.I 2000)
- Flora of Mizoram (Singh 2002)
- Department of Forests, Ecology, Environment & Wildlife, Nagaland, Kohima (Annual Administrative Report 2014-15).

Further, herbarium available with the Botanical Survey of India (BSI) regional offices at Gangtok (Sikkim), Shillong (Meghalaya) and Central National Herbarium (Howrah, West Bengal) were examined for the description of the species. A distribution map of *M. pruriens* in northeast India was prepared and collection locations were identified accordingly. Accessions were collected from

natural growing locations, herbal product vendors, traditional healers, and the backyards of home gardens, both in the form seeds as well as matured pods. Each collection was given a separate collection number and/or donor's ID for proper identification (Fig. 4). A passport data pertaining to name of the collector/donor, site of collection (province, district, village, longitude and latitude) was recorded.



Department of Botany
Sikkim University, Gangtok, Sikkim – 737102

Field Collection Form

Collection No. Date:_____ Altitude (m):_____
Location: _____ Long/Lat: _____
Taluk: _____District:_____State:_____
Botanical Name: _____Synoym:_____
Family:_____Local Name:_____
Forest Type_____Frequency of Occurrence: Abundant/Frequent/Occasional/_____
Habit: Herb / Shrub / Tree / Climber / _____ Micro habit: Marshy/ dry / rocky / sandy /_____
Root: _____Stem:_____Bark:_____
Leaves:_____Flowering / Fruiting / Vegitative_____
Inflorescence:_____Flower Character:_____
Flower colour:_____Fragrance:_____Fruits:_____
Special remarks:_____
_____
_____
Local use (s):_____
_____
_____Information by: _____Age: _____
Collected by: _____

Fig. 4: Passport data format

4.1.2 Establishment of germplasm

All the collections were planted in the Sikkim University botanical garden and established as per the standard package of practices. The site is located in the eastern part of Sikkim state situated in the eastern Himalayan tract at an altitude of 1600 m (5410 ft) above the mean sea level (MSL) along 27^0 33'N longitude and 88^0 62' E latitude. The plants were initially raised in pots and arranged as per

the germplasm layout prepared. For evaluation, we designed a row and block system with each block comprising four replicates of a single accession. The germplasm was laid down to have 8-10 such blocks in each row separated by 3 ft between blocks and 5 ft between rows. The experiments were laid out in a randomized complete block design (RCBD) (Onyishi et al. 2015) (Fig. 5).

Fig. 5: Germplasm of *M. pruriens* maintained at Sikkim University

4.1.3 Germplasm evaluation

Totally 50 *M. pruriens* accessions collected from all over the northeast India were subjected to evaluation. For seedling establishment, five seeds from each accession were initially planted in each pot and after germination only single plant was retained per pot. The potting mixture was prepared as soil: sand: manure in 2:2:1 ratio. No external fertilizer was applied at any stage. The study

was conducted during two seasonal periods: April - December 2015 and 2016 coinciding with the legume cultivation spell in the Sikkim-Himalayan hills.

The data on 10 qualitative characters *viz*., leaf texture/surface, leaflet shape, flower colour, pod trichome colour, pod sensation, pod curvature, plant hairiness, seed shape, seed colour, seed coat pattern and 10 quantitative characters such as: days to emergence, days to flowering, days to first mature pod, number of pods per plant, days to maturity, 100 seed weight, seed length, seed thickness, seed area and seed width were recorded adopting *Mucuna* germplasm descriptor developed earlier by our group (Table 7). Each qualitative descriptor was scored by observing three tagged plants per accession. The data was recorded initially using a scale based on rank descriptors and then converted into binary data (1, 0). Quantitative traits were scored as mean value integers of four measurements taken on four replicates of a single accession. Among the qualitative characters, vegetative growth data was recorded after 45 days, flower and pod data at 50-60% maturity, and days to maturity when 90% of the pods were about to dry and seed characteristics and L-Dopa content after harvesting the seeds from mature pod.

Table 7 List of descriptors used for *M. pruriens* characterization

Vegetative Characters	
Descriptor	**Investigation Method**
Leaf texture/surface	Recoded from the leaflets at the third node, use both surfaces: Membranous/Semi – coriaceous/Coriaceous
Leaflet shape	Measured on the terminal leaflet of third trifoliate leaf according to the ratio of length (l) to width (w) Round (<1.5) / Ovate (1.5-2) / Ovate-lanceolate (2-3)
Floral and Fruit Characteristics	
Descriptor	**Investigation Method**
Flower color	The color of fully expanded flower was noted using RHS color chart

Duration of flowering	From first flowers to stage when 60% of plants have flowered
Pod curvature	Of fully expanded immature pod (Curved / S-shaped)
Pod trichome color	Velvety black/ Creamish white/ Golden orange
Pod sensation	Itching/non-itching
Plant hairiness	Overall plant hairiness: Very low/Low/ Moderate/Dense
Seed color	When dried at harvest
Seed shape	Round / Oval / Cuboid / Other
Seed coat pattern	Absent/Non mottled/Black mottled/Brown mottled/Dark brown mottled/ Grey mottled/Dark brown mottled
Morpho-agronomic Characters	
Descriptor	**Investigation Method**
Days to emergence	Number of days from sowing to 50% seedling emergence
Days to flowering	Number of days from emergence to stage when 50% of plants have begun to flower. Recorded for plants with the same sowing date, at the same location
Days to first mature pod	From the emergence to stage when 50% of plants have matured pods
Days to maturity	From emergence to stage when 90% of pods are ripe
No. of pods per plant	Average from 4 random normal plants
Seed length (mm)	Mean length of 10 mature seeds excluding those from the extremities of pods
Seed width (mm)	Mean width from hilum to keel of 10 mature seeds measured for Length
Seed thickness (mm)	Mean thickness of the 10 mature seeds measured for length: measured perpendicular to length and width
Seed weight (g)	Weight of 100 seeds
Seed area (mm)	Measurement of dried seed length X seed thickness (mm)

4.1.4 Data analysis

The binary and mean value data obtained for the qualitative and quantitative traits were subjected to following four different analyses.

4.1.4.1 Descriptive statistics and frequency distribution

Qualitative traits were recorded on a scale based on rank descriptors. For 10 quantitative characters - range, mean, standard deviation, standard error and coefficient of variation were calculated using Graph Pad - Prism 5.01 (Graph Pad software Inc. 2014) and XLSTAT software V2010.4.01 (XLSTAT 2010). The data

on quantitative traits, after normalization, were subjected to frequency distribution analysis to ascertain the number of allocations in each frequency class which was later used to determine the genotypes falling under extreme phenotypic classes. For seed characteristics, measurements were made using digital balance and vernier calipers.

4.1.4.2 ANOVA

The data on quantitative traits were subjected to one way analysis of variance after normalization of the data-sets. All 50 accessions were initially tested for differences among their mean values using Tukey-Kramer HSD test (Tukey 1953; Kramer 1956) at an alpha level 0.05. For 10 yield traits showing significant ANOVA, Dunnett's test (Dunnett 1955) was conducted at 5% significance level. The cumulative results from all the traits were compared to obtain the most contrasting parental combinations to use as mapping population parents. The results were obtained by using software JMP V10 (SAS Inst. Inc. 2008, 2011), Graph Pad - Prism 5.01 (Graph Pad software Inc. 2014) and Analyse-it V2.2 (Analyse-it, 2009) respectively.

4.1.4.3 Correlation test

Both qualitative and quantitative traits were subjected to correlation analysis using Pearson's correlation coefficient test by employing Graph Pad - Prism 5.01 (Graph Pad software Inc. 2014) or XLSTAT V2010.4.01 (XLSTAT 2010) software. For this, the qualitative data was first converted into binary (1, 0) data matrix based on presence/absence of the trait. For quantitative trait, observations were converted into frequency classes using statistical formula:

$$\text{Number of classes} = 1 + 3.322 \log (N)$$

Where N is the number of observations. Each obtained class was considered as individual marker and used for developing binary data.

4.1.4.4 Cluster analysis

For both qualitative and quantitative traits, cluster analysis was performed using Agglomerative Hierarchical Clustering (AHC) based on Euclidean distances of dissimilarity with Unweighted Pair Group Average (UPGA) aggregation method using XLSTAT software V2010.4.01 (XLSTAT 2010). For dendrogram construction, data was organized into Jaccard's similarity matrix and analyzed by Un-weighted Pair Group Method with Arithmetic averages (UPGMA) algorithm using Free Tree V0.9.1.50 software (Pavlicek, et al. 1999). Multi-scale bootstrapping values of the resulting clusters were determined using the same software and the trees were generated using Tree View (Win32) V-1.0.0.0 software.

4.1.4.5 Principal Component Analysis (PCA)

PCA was performed for both qualitative and quantitative data using XLSTAT software V2010.4.01 (XLSTAT 2010). For this, first similarity matrix was used to generate Eigen values and scores for accessions. The first two principal components which accounted for the highest variation were then used to plot the structure.

4.2 Assessment of L-Dopa variability

4.2.1 L-Dopa isolation

L-Dopa was isolated from dried seed samples as per the method described by Upadhyay et al. (2012) with minor modification. First, the air dried seeds were

powdered and passed through a mesh sieve. The sieved material (100 g) so obtained was extracted on water bath with 400 ml 40% ethyl alcohol, 60% water and 1.5% glacial acetic acid at 60-70^{0}C for 1-2 h. The material was then filtered and marc was further refluxed three times with 300 ml ethyl alcohol (40%). Following this, all the extracts were pooled, concentrated up to 40% under the vacuum using RotaVac (Heidolph, Schwalbach, Germany). It was then kept overnight to settle down, filtered and air dried by using SpeedVac System (Thermo Scientific Savant DNA 120) for jelly extract which was used for the assay.

4.2.1.1 Validation of equipment

Repeatability was checked by using varied concentrations (100-1000 ppm) of standard L-Dopa (Sigma Aldrich, USA) solutions in 10 replicates and the peaks were determined. For each sample, value was obtained in triplicates.

4.2.1.2 Preparation of standard solution

Standard solution of pure L-Dopa was prepared by dissolving 4.0 mg in 10 ml (1000 ppm) of 0.1N HCl in a volumetric flask (Stock solution) to determine the limit of detection (LOD) and limit of quantification (LOQ). One to ten ml of these stock solutions were diluted to 10 ml to get a concentration range of 100-1000 ppm for linearity study.

4.2.1.3 Preparation of sample solution

Approx 50 mg grounded extract was taken and dissolved with 15 ml 0.1 N aqueous HCl separately. The sample was then sonicated for 20 min. After sonication, the volume was made up to 50 ml with 0.1N HCl and filtered through 0.22 µm membrane filter.

4.2.2 L-Dopa estimation

4.2.2.1 HPLC procedure

L-Dopa content was analyzed by High Performance Liquid Chromatography (HPLC, Thermo Scientific DIONEX Ultimate 3000). The sample was manually injected and detected with double-beam photometer detector operated at a wavelength of 283 nm. The data was acquired and analyzed using Chromaleon software. A C18 column (250 x 4.6 mm, 5 µm SS) fitted with guard was used for separation using an isocratic mixture with 0.1M KH_2PO_4, pH 3.0 by Ortho-phosphoric acid as mobile phase. The mobile phase was filtered through a 0.45 µm membrane filter (Sartorius USA) and was then degassed by ultrasonication for 30 min. Analysis was run at a flow rate of 1.0 ml/min and quantification was by peak height. Injection volume was adjusted to 20 µl and detection was made at 283 nm. Standard L-Dopa (Sigma Aldrich) was used for the preparation of calibration curve and retention time analysis. Each sample was analyzed in quadruplicates and L-Dopa concentration was finally represented as percentage (%) composition. All the experiments were performed in triplicates.

4.2.2.2 Calibration curve

Ten different concentrations of stock solutions after dilution (100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 ppm) with mobile phase were injected in triplicates.

4.2.2.3 Linearity

Linearity was determined by using 10 concentrations of the standard solution (100-1000 ppm). The calibration curve was obtained by plotting the absorbance versus the concentrations of the standard solution.

4.3 Genetic diversity analysis

4.3.1 Plant material

A total of 50 *M. pruriens* accessions representing both var. *pruriens* (38) and var. *utilis* (12) collected from northeastern region of India (Table 8) were used.

4.3.2 DNA isolation

One gram of young leaf tissue was harvested from 2-3 week old seedlings from 10 individual plants and bulked separately for each accession. Total genomic DNA was isolated following Doyle and Doyle (1990) method using Cetyl Trimethyl Ammonium Bromide (CTAB). Ground leaves were homogenized in CTAB-extraction buffer containing 0.5% charcoal plus 0.2% β-mercaptoethanol and incubated at 60^0C for 1 h. Purification was carried out twice with SEVAG (24-chloroform: 1-isoamlyalcohol) and once with phenol: chloroform: isoamylalcohol (25:24:1). After centrifugation at 12,000 rpm for 2 min, the DNA was pelleted using 0.67 volume of isopropanol followed by ethanol (70%) wash. Air dried pellets were re-suspended in 0.5 ml of 1X Tris-EDTA buffer (pH 8.0) and quantified fluorometrically on agarose gel (0.8%). Ethidium bromide staining was used to visualize the bands.

Table 8 List of *M. pruriens* accessions collected from northeastern region of India

Sl. No.	Variety	Accession number	Place of collection	Lat. & Long.
1	*Mucuna pruriens* var. *pruriens*	500197WB	West Bengal	$25^0$41ˈN,$85^0$13ˈE
2	*Mucuna pruriens* var. *pruriens*	500199WB	West Bengal	21º15'N, 81º41'E
3	*Mucuna pruriens* var. *utilis*	500203MN	Manipur	$24^0$44' N, 93^0 58' E
4	*Mucuna pruriens* var. *pruriens*	500204AS	Assam	$26^0$11' N, 91^0 44' E
5	*Mucuna pruriens* var. *pruriens*	500205AS	Assam	$26^0$11' N, 91^0 44' E
6	*Mucuna pruriens* var. *utilis*	500210MN	Manipur	25°41'N,94°47'E
7	*Mucuna pruriens* var. u*tilis*	500211NL	Nagaland	$25^0$67'N, 94^0 12'E
8	*Mucuna pruriens* var. *pruriens*	500212AS	Assam	$26^0$11'N, 91^0 44'E
9	*Mucuna pruriens* var. *utilis*	500213MN	Manipur	$24^0$44'N, 93^0 58'E
10	*Mucuna pruriens* var. *utilis*	500215AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
11	*Mucuna pruriens* var. *pruriens*	500216ML	Meghalaya	$25^0$69'N, $91^0$97'E
12	*Mucuna pruriens* var. *utilis*	500217MN	Manipur	25°68'N,93°03'E
13	*Mucuna pruriens* var. *pruriens*	500219TR	Tripura	$23^0$50'N, 91^0 25'E
14	*Mucuna pruriens* var. *pruriens*	500220AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
15	*Mucuna pruriens* var. *pruriens*	500221AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
16	*Mucuna pruriens* var. *pruriens*	500222AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
17	*Mucuna pruriens* var. *pruriens*	500223AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
18	*Mucuna pruriens* var. *pruriens*	500224AR	Arunachal Pradesh	$27^0$08'N, 93^0 40'E
19	*Mucuna pruriens* var. *pruriens*	500225AR	Arunachal Pradesh	$28^0$06'N, 95^0 03'E
20	*Mucuna pruriens* var. *pruriens*	500226AR	Arunachal Pradesh	$28^0$06'N, 95^0 03'E
21	*Mucuna pruriens* var. *pruriens*	500227AR	Arunachal Pradesh	$28^0$06'N, 95^0 03'E
22	*Mucuna pruriens* var. *pruriens*	500228AR	Arunachal Pradesh	$28^0$06'N, 95^0 03'E
23	*Mucuna pruriens* var. *pruriens*	500231AR	Arunachal Pradesh	$28^0$06'N, 95^0 03'E
24	*Mucuna pruriens* var. *pruriens*	500232AS	Assam	$26^0$11'N, 91^0 44'E
25	*Mucuna pruriens* var. *pruriens*	500233AS	Assam	$26^0$11'N, 91^0 44'E
26	*Mucuna pruriens* var. *pruriens*	500234MZ	Mizoram	23°44″N92°43″E

27	*Mucuna pruriens* var. *pruriens*	500235TR	Tripura	23º50" N, 91º25'E
28	*Mucuna pruriens* var. *pruriens*	500236AS	Assam	$26^0$11'N,91^0 44'E
29	*Mucuna pruriens* var. *pruriens*	500237AS	Assam	$26^0$11'N, 91^0 44'E
30	*Mucuna pruriens* var. *pruriens*	500239MZ	Mizoram	23°44″N92°43″E
31	*Mucuna pruriens* var. *pruriens*	500240SK	Sikkim	$27^0$33'N, 88^0 62'E
32	*Mucuna pruriens* var. *pruriens*	500241SK	Sikkim	$27^0$33'N, 88^0 62'E
33	*Mucuna pruriens* var. *pruriens*	500242SK	Sikkim	$27^0$33'N, 88^0 62'E
34	*Mucuna pruriens* var. *pruriens*	500243SK	Sikkim	$27^0$33'N, 88^0 62'E
35	*Mucuna pruriens* var. *pruriens*	500244SK	Sikkim	$27^0$33'N, 88^0 62'E
36	*Mucuna pruriens* var. *pruriens*	500245TR	Tripura	$23^0$50'N, 91^0 25'E
37	*Mucuna pruriens* var. *pruriens*	500246TR	Tripura	$23^0$84'N,$91^0$28'E
38	*Mucuna pruriens* var. *pruriens*	500248TR	Tripura	$23^0$50'N,$91^0$25'E
39	*Mucuna pruriens* var. *pruriens*	500249SK	Sikkim	$27^0$33'N,$88^0$62'E
40	*Mucuna pruriens* var. *utilis*	500255AR	Arunachal Pradesh	$28^0$06'N,$95^0$03'E
41	*Mucuna pruriens* var. *utilis*	500263NL	Nagaland	26°00'N,94°20'E
42	*Mucuna pruriens* var. *utilis*	500266NL	Nagaland	$26^0$22'N,$94^0$47'E
43	*Mucuna pruriens* var. *utilis*	500267NL	Agaland	$25^0$68'N, $94^0$08'E
44	*Mucuna pruriens* var. *utilis*	500268NL	Nagaland	$24^0$57'N, $94^0$43'E
45	*Mucuna pruriens* var. *pruriens*	500269SK	Sikkim	$27^0$5'N, $88^0$53'E
46	*Mucuna pruriens* var. *pruriens*	500271SK	Sikkim	$27^0$33'N, 88^0 62'E
47	*Mucuna pruriens* var. *pruriens*	500274SK	Sikkim	$27^0$33'N,$88^0$62'E
48	*Mucuna pruriens* var. *pruriens*	500275ML	Meghalaya	$25^0$69'N, $91^0$97'E
49	*Mucuna pruriens* var. *pruriens*	500276ML	Meghalaya	$25^0$69'N, $91^0$97'E
50	*Mucuna pruriens* var. *utilis*	500277NL	Nagaland	26°00'N,94°20'E

Note: SK: Sikkim; AS: Assam; AR: Arunachal Pradesh; NL: Nagaland; MN: Manipur; TR: Tripura; L: Meghalaya; WB: West Bengal; MZ: Mizoram

4.3.3 AFLP data generation

Totally 22 AFLP primer combinations (PCs) were used. The details of the AFLP primers are listed in Tables 9a & 9b. Prior to PCR amplification, the quantity of DNA were checked using a Nano-Drop 2000 spectrophotometer (Thermo Scientific, USA) by taking the absorbance at 260/280 nm.

4.3.4 AFLP analysis

Twenty two AFLP primers were tested as per the method described by Vos et al. (1995) with minor modification. The genomic DNA (400–500 ng) was digested with 5U *Eco*RI and 1U *Mse*I restriction endonucleases for 2 h at 37^0C. The *Eco*RI and *Mse*I adapters (AFLP® plant mapping kit, Applied Biosystems, Foster City, CA, USA) were ligated to digested DNA using 1U T4 DNA ligase (New England Biolabs, USA). The restricted–ligated DNA was then diluted by 20 fold with Tris-EDTA (TE) buffer (20 mMTris-HCL, 0.1 mM EDTA, pH 8.0) to perform pre-selective amplification using PCR master mix (2x Taq polymerase included) obtained from Thermo Scientific. The pre-selective amplification profile included: 30 cycles of denaturation at 94^0C for 20 s, annealing at 56^0C for 30 s and extension at 72^0C for 2 min and at last hold at 60^0C for 30 min. The pre-selective amplification was confirmed by running PCR products on 1.5 % agarose gel. Later, the amplification products were further diluted 10 fold with TE buffer and used as a template for selective amplification. The selective amplification was performed in 10 µl reaction volume using selective primer pairs with amplification profile of denaturation at 94^0C for 20 s, annealing at 66^0C for 30 s decrease by 1^0C/cycle up to 56^0C, extension at 72^0C for 2 min for 10 cycles, followed by 94^0C for 20 s, 56^0C for 30 s, 72^0C for 2 min for 30 cycles followed by 60^0C for 30 min. Following selective amplification, post PCR multiplex sets were made based on fluorescence

labeled primers. For post PCR multiplexing, 1 µl each fluorescent dye of 6-FAM (blue), JOE (green) and NED (Yellow) labeled PCR products representing different selective primer-pairs were combined with 7 µl of Hi-Di formamide containing 0.3 µl GeneScan^{TM} 500 ROX® as internal size standard. The multiplexed PCR mixture was denatured for 5 min at 95^{0}C, quick chilled on ice for 5 min and loaded on ABI 3730xl DNA Analyzer for electrophoresis. The fragment analysis was performed by GeneMapper v 4.0 software (Applied Biosystems, USA).

4.3.5 Data analysis

Allele frequency and polymorphism for each AFLP fragments were scored automatically on GeneMapper as well as manually by visualizing the electropherogram. The data was scored as "1" (presence of fragment), "0" (absence of fragment) and "." for (missing data). The genotypic data were used to calculate different population genetic parameters such as polymorphic information content (PIC), marker index (MI) and resolving power (RP). The PIC value and mean gene diversity for each AFLP primer combination was calculated using software Powermarker v 3.25 and the scores were averaged for each PC.

PIC for each AFLP PC was calculated according to Roldan-Ruiz et al. (2000) formula: $PIC_i = 2f_i(1-f_i)$, where PIC_i is the polymorphic information content of marker i; f_i is the frequency of the fragments which were present and $1-f_i$ is the frequency of the fragments which were absent. PIC was averaged over the fragments for each PC. Marker index (MI) was calculated following Powell et al. (1996) as PIC × EMR, where EMR (effective multiple ratio, EMR = β × n) is defined as the product of the fraction of polymorphic loci (β) and the number of

polymorphic loci (n). For each PC, the resolving power was calculated according to Prevost and Wilkinson (1999) as $Rp=\Sigma Ib$, where Ib is the fragment informativeness and calculated as $Ib=1–[2 \times |0.5−p|]$, where p is the proportion of the genotypes containing the fragment.

Diversity indicators were calculated in total, per population and per AFLP fragment in GenAlEx 6.5 (Peakall 2012), POPGENE 1.31 (Yeh 1997) or Arlequin 3.1 (Excoffier 2005) software. Number of different alleles (Na), effective number of alleles (Ne), Shannon's information index (I), gene diversity (h), expected heterozygosity (He) (Nei 1973) and molecular variance (AMOVA) (Excoffier et al. 1992) within and among the populations were calculated in GenAlEx 6.5 (Peakall 2012). POPGENE/Arlequin 3.1 was used to calculate overall diversity in collection (total gene diversity = Ht), diversity within populations (Hs), genetic differentiation (Gst = 1-HS/HT) and gene flow (Nm) which was estimated from Gst as Gst (Nm = 0.5 (1- Gst)/Gst. The Jaccard's similarity coefficient values were used to estimate genetic similarity between the accessions. Based on the similarity matrix, dendrograms were constructed based on UPGMA by using FreeTree V1.0.0.0 and TreeView (Win32) V1.6.6 softwares (Pavlicek et al. 1999).

Genetic structure was investigated using STRUCTURE v 2.3.4. (Pritchard 2000; Falush 2003) which uses Markov Chain Monte Carlo (MCMC) algorithm. An admixture model with correlated allele frequencies was used. The K value was set from 1 to 10, and at least 20 runs were performed for each value of K. The length of the burn-in period was set to 50,000, and the MCMC chains after burn-in were run for an additional 100,000 times. The optimal value of K was determined by examination of the ΔK statistic (Evanno 2005) using StructureHarvester (Earl 2012).

Table 9a Details of pre-selective primer used for AFLP analysis

Sl. No.	AFLP Pre-selective Primer combination	Primer sequence (5' to 3')
1	E-A/M-C	GACTGCGTACCAATTC-A GATGAGTCCTGAGTAA-C

Table 9b Details of selective primers used for AFLP analysis

Sl. No.	AFLP Selective Primer combinations	Selective primer sequence (5' to 3')
1	E-ACT/M-CTA	GACTGCGTACCAATTC-ACT (FAM) GATGAGTCCTGAGTAA-CTA
2	E-AAC/M-CAG	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CAG
3	E-AAC/M-CTA	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CTA
4	E-AAC/M-CTC	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CTC
5	E-ACA/M-CTT	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CTT
6	E-AGG/M-CAG	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CAG
7	E-ACG/M-CAC	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAC
8	E-ACA/M-CAT	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CAT
9	E-AAG/M-CAC	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAC
10	E-ACA/M-CAA	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CAA
11	E-ACG/ M-CAA	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAA
12	E-AAG/ M-CAT	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAT
13	E-AAG/ M-CTG	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CTG
14	E-AAG/ M-CTC	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CTC
15	E-AAG/ M-CAA	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAA
16	E-ACT/ M-CAA	GACTGCGTACCAATTC-ACT (FAM) GATGAGTCCTGAGTAA-CAA
17	E-ACG/ M-CAG	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAG
18	E-AGC/ M-CTA	GACTGCGTACCAATTC-AGC (NED) GATGAGTCCTGAGTAA-CTA
19	E-ACG/ M-CTG	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CTG
20	E-AGG/ M-CAT	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CAT
21	E-AGG/M-CTA	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CTA
22	E-AGC/M-CTG	GACTGCGTACCAATTC-AGC (NED) GATGAGTCCTGAGTAA-CTG

4.4 Genetic analysis of L-Dopa trait

4.4.1. Hybridization of contrasting parents

For genetic analysis of L-Dopa trait, initially four different crosses were initiated from the parental accessions which contained extreme L-Dopa contents (Table 10). These accessions were identified based on our L-Dopa analysis. The hybridization was carried out following hand emasculation, dusting of donor pollens and bagging for protected pollination. The experiments were conducted during the months of September-October, 2015 between 9.00-10.00 AM and seeds from fully matured pods were recovered during November/December 2016. The progenies consisted total of 59 F1 seeds. From these, single F1 seed from the cross between 500240SK (High L-Dopa content; 9.03%) and 500212AS (Low L-Dopa content; 1.55%) was planted to raise F2 progenies (Fig. 6).

Table 10 L-Dopa content among different parental combinations used for crossing

Sl. No.	Parents for F1
1	**500240SK ♂ (9.03%) X 500212AS ♀ (1.55%)**
2	500212AS ♂ **(1.55%)** X 500240SK ♀ **(9.03%)**
3	500175MH ♂ **(7.77%)** X 500212AS ♀ **(1.55%)**
4	500212AS ♂ **(1.55%)** X 500175MH ♀ **(7.77%)**

4.4.2 Development of F2 population

Total 959 F2 seeds were obtained from which 400 random seeds were sown for raising the F2 population. The seeds were sown during the month of April-May, 2017. Seedlings were raised in a pot measuring in 8”x 3” size. All the F2 plants were maintained as per the standard package and practices and subjected to further evaluation. Totally 272 surviving F2 individuals were analyzed.

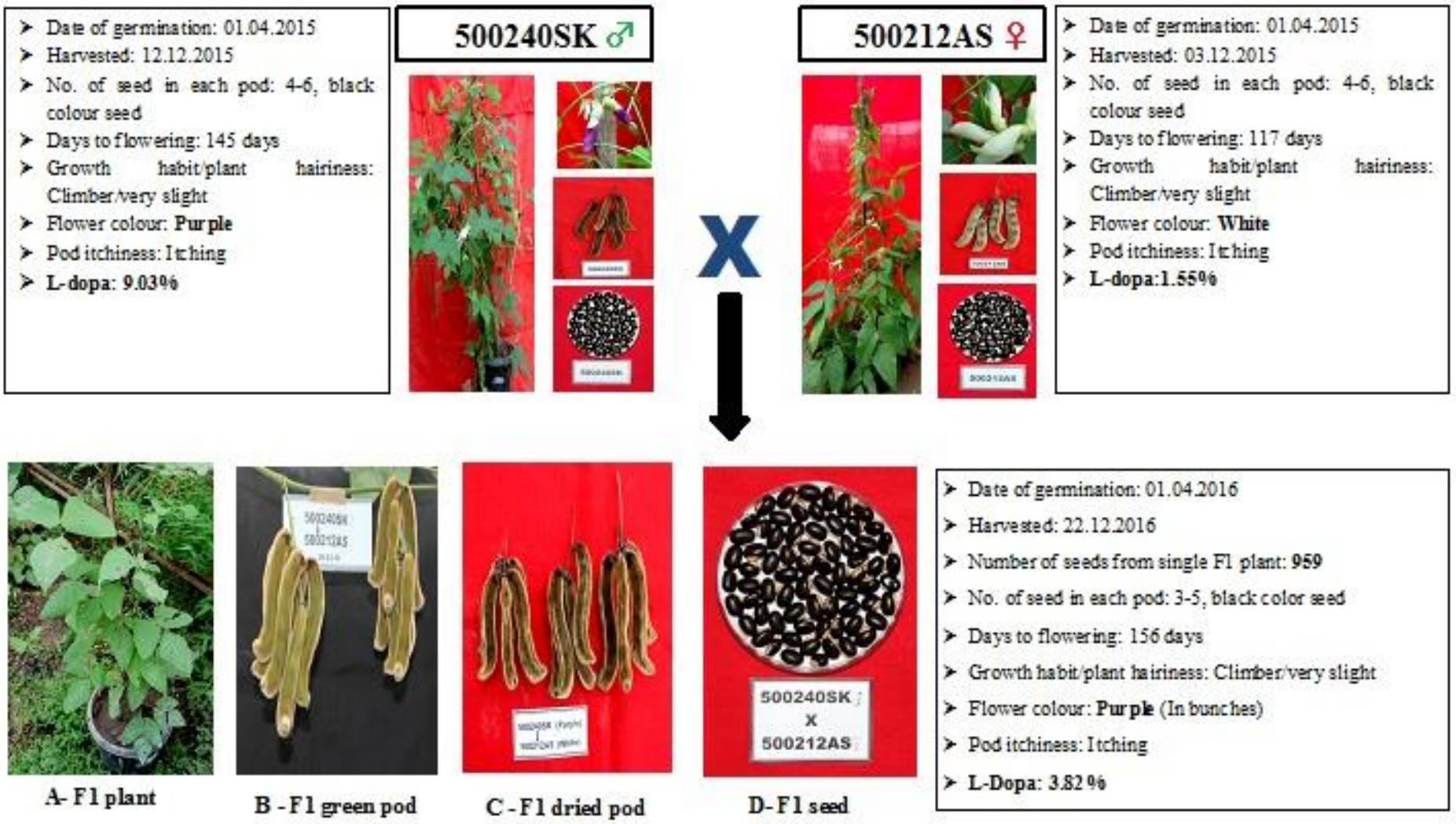


Fig. 6: Development of L-Dopa segregating population

4.4.3 Phenotyping of F2 individuals

Qualitative traits including flower color, fruit color, presence of trichome and pod sensation were recorded visually. Phenotyping data on length (mm), width (mm), thickness of seeds (mm), weight of single seed (g) and 100 seed weight were recorded as per the method described earlier. The data was recorded in five replicates and then averaged.

4.4.4 Variance and heritability analysis

4.4.4.1 Analysis of variance

The data collected for L-Dopa content among parents, F1 and F2 progenies were subjected to analysis of variance (ANOVA) to test the variations among genotypes and progenies. The analysis of variance was calculated using Statistical Analysis System (SAS) software version 9.2 (SAS, 2008).

4.4.4.2 Estimation of variance components

The variability present in the population was estimated by measuring mean, phenotypic and genotypic variance and coefficient of variation. To estimate the phenotypic and genotypic variance, genotypic and phenotypic coefficients of variation were estimated based on the formula of Syukur et al. (2012) as follows:

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

$$\sigma^2P = [\sigma^2G + (\sigma^2E/r)]$$

$$\sigma^2E = \frac{\text{Var. of P1+ Var. of P2+2 x Var. F1}}{4}$$

Where: σ^2G = Genotypic Variance; σ^2P = Phenotypic Variance; σ^2E = Environmental Variance; MSG = Mean Square of Genotypes; MSE = Error Mean Square from ANOVA)

$$\text{Genotypic coefficient of variation (GCV)} = [\sigma^2 G)^{1/2} / \bar{x}] \times 100$$

$$\text{Phenotypic coefficient of variation (PCV)} = [\sigma^2 P)^{1/2} / \bar{x}] \times 100$$

Where: $\sigma^2 G$ = Genotypic variance; $\sigma^2 P$ = Phenotypic variance; is grand mean of a character.

4.4.4.2 Heritability estimation in broad sense

Broad sense heritability (h^2) of all the traits was calculated according Allard (1960) using the formula:

$$h^2 = \sigma [(^2 G) / (\sigma^2 P)\} \times 100$$

Where, h^2 = heritability in broad sense.

4.4.4.3 Estimation of genetic advance

Genetic advance (GA) was determined as described by Johnson et al. (1955b) as:

$$GA = K (\sigma P) h^2$$

Where K = selection differential (K = 2.06 at 5% selection intensity); σP = the phenotypic standard deviation of the character; h^2 = heritability in broad sense.

The genetic advance as percentage of mean (GAM) was calculated as described by Johnson et al. (1955a) as given below:

$$GAM (\%) = GA / \bar{x} \times 100$$

Where GAM = Genetic advance as percentage of the mean; GA = Genetic advance and $\bar{x}$ = Grand mean of character.

4.4.4.4 Normal distribution curve

Normal distribution curve can be characterized by two statistics, the mean or average (μ) of the distribution and the standard deviation (σ), a measure of the spread of the distribution. It was plotted by using the following formula:

$$\mu = \bar{x} = (x_1 + x_2 + x_3 + \ldots\ldots\ldots + x_n) /n = \Sigma^n_i (x_i) / n$$

Where $\bar{x}$ = sample mean, $\Sigma^n_i (x_i)$ is the sum of x values from i=1 to n.

4.5 Micropropagation of elite genotypes

4.5.1 Seed material and sterilization

Mature seeds of the genotype 500240SK collected from well dried pods of six months old plant was used as a seed source. The seeds were initially washed for 8-10 min with detergent - Extran$^{@}$ MA 02 (Merck) followed by 20 min wash in running tap water. The seeds were then surface sterilized in a mixture of 0.2% $HgCl_2$ +0.05% cetrimide + 0.06% bavistin for 5 min. The seeds were then thoroughly washed with autoclaved distilled water.

4.5.2 Media and culture condition

Murashige and Skoog's (1962) salt with 3% (W/V) sucrose was used as basal medium except in case of seed germination medium which lacked sucrose. After adding the growth regulators, the pH of the medium was adjusted to 5.6 ± 0.1 followed by gelling with 0.8% agar. The media was then autoclaved at 121^0C and 1.06 kg/cm^2 pressure for 20 min.

4.5.3 Explants preparation

The surface sterilized seeds were germinated on solid MS basal medium without sucrose and the seedlings were allowed to grow for 12-14 days until the axillary buds became prominent. The explants were then prepared as per the method described by Jayanand et al. (2003). Axillary buds measuring 0.8 - 1.0 cm were aseptically inoculated onto MS medium supplemented with shoot induction hormones. Apical buds and cotyledonary nodes were prepared as per the method described earlier (Sathyanarayana et al. 2008).

4.5.4 Multiple shoots induction

The explants - apical bud, axillary bud and cotyledonary node were cultured on MS medium supplemented with BAP in various concentration (0.00-22.20 µM), Kn (0.464-23.2 µM) and combinations of BAP (4.44 µM) + Kn (0.93-4.65 µM), BAP (4.44 µM) + NAA (1.08-5.40 µM) for multiple shoot induction. The induced shoots were allowed to grow for 30 days and at the end of 30 days, the explants producing maximum number of multiple shoots were sub-cultured on MS + BAP (0.44-4.44 µM) for 30 more days for proliferation of shoots.

4.5.5 Elongation and rooting

Micro-shoots grown on subculture medium were transferred on to MS media supplemented with GA_3 (2.89 - 14.43 µM) for elongation. The shoots of 3-4 cm length were then transferred to half strength MS medium supplemented with NAA (0.00 – 16.2 µM) for rooting.

4.5.6 Hardening of plantlets

After 20 days, plantlets (3-4 cm) with good rooting were removed and thoroughly washed under running tap water for 2-3 min to remove traces of left-out medium. They were then transferred to plastic pots containing autoclaved sand: soil: peat moss (1:1:1) mixture and watered. All the plants were covered with polythene bags to maintain humidity. The plants were initially maintained in culture room conditions *viz*., 25 ± 2^0 C with 60-70% relative humidity (RH) and a photoperiod of 16:8: light: dark with an illumination of 50 µmol $m^{-2}s^{-1}$ flux for a period of one week. After this, polybags were removed and the plants were transferred to greenhouse with 25 ± 2^0 C temperature, and relative humidity 55±5%. After one month, the plants were transferred to field for establishment.

4.5.7 Statistical analysis

The cultures were examined at regular intervals of seven days and all the experiments were repeated at least thrice. Data analysis was done using one way ANOVA and mean values of treatments were compared by Tukey's HSD test with 5% significance using JMP software version 9 (SAS Inst. Inc. 2008) with UNIVARIATE procedure.

5. Results

5.1 Germplasm collection and characterization

5.1.1 Collection of germplasm

Totally 50 *M. pruriens* accessions were collected covering 18 districts of 8 northeastern states of India namely Assam, Manipur, Tripura, Mizoram, Nagaland, Arunachal Pradesh, Meghalaya and Sikkim (Fig. 7). The collection included both the botanical varieties of *M. pruriens* available in the region *viz*., var. *pruriens* (38) and var. *utilis* (12). The third variety, var. *hirsuta* is endemic to peninsular India. All the accessions were new collections from the region.

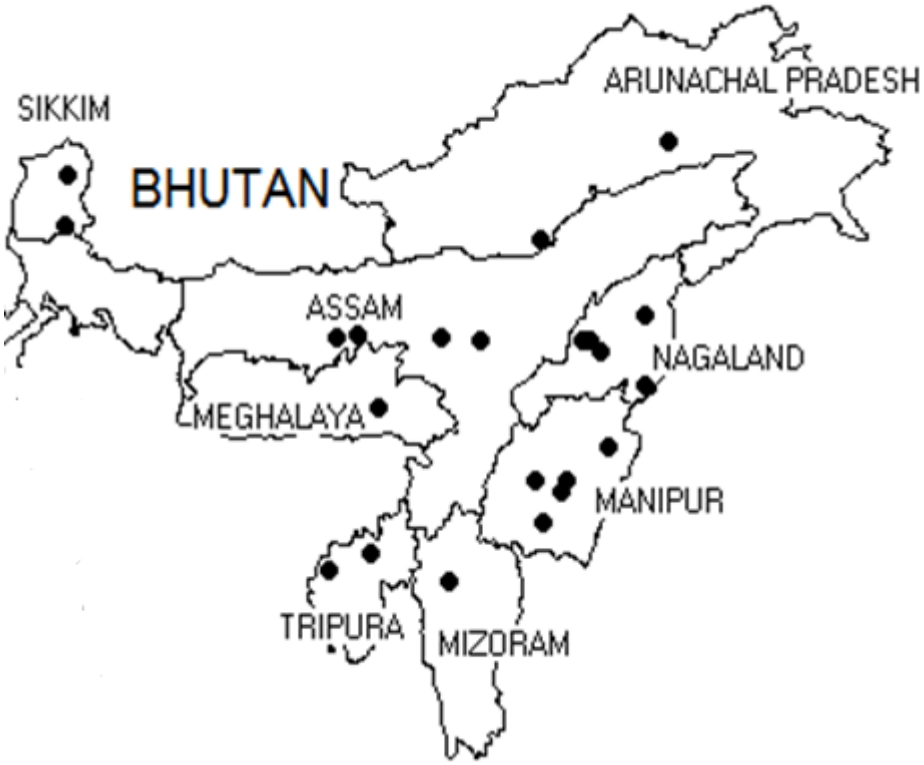


Fig. 7: Map showing geographical locations of germplasm collection
Note: Points in the map indicate collection locality
Map Courtesy: www.mapsofindia.com

5.1.2 Germplasm evaluation

5.1.2.1 Qualitative characters

The results of evaluation carried out on 10 qualitative characters *viz*., leaf texture/surface, leaflet shape, flower colour, pod trichome colour, pod sensation, pod curvature, plant hairiness, seed shape, seed colour, and seed coat pattern revealed at least five of them *viz*., flower colour, pod trichome colour, pod sensation, seed coat colour and seed coat pattern are highly variable. The details are given in Table 11.

In case of leaf texture, majority of the accessions showed membranous (56%) or coriaceous (40%) leaf texture, while very few showed semi-coriaceous (4%) texture. All three kinds of leaf textures were observed in var. *pruriens* whereas var. *utilis* exhibited only membranous leaf. The leaf shape was predominantly ovate (72%) except in some accessions of Arunachal Pradesh, Nagaland and Assam (28%) which showed ovate-lanceolate shape.

In case of flower colour, two accessions, one each belonging to var. *pruriens* (500212AS) and var. *utilis* (500267NL) possessed white flower colour which added to 4% variability to the total collection of which 2.63% was within var. *pruriens* and 8.33% was within var. *utilis*. The, purple or dark violet flower emerged as most common (96%) flower colour in both the varieties.

For pod trichome colour, in var. *pruriens*, 97.36% of the accessions showed golden orange colour and 2.63% were creamish white. In case of var. *utilis*, 75% of accessions showed creamish white and 25% velvety black trichomes.

The pod sensation was categorized as either itching or non-itching. Of the 50 germplasm accessions, 76% exhibited itching which is typical of var. *pruriens* and remaining 24% belonging to var. *utilis* showed non-itching character. Pod curvature also showed two variants: curved (28%) which was predominant in var. *utilis* and S- shaped (72%) which was common in var. *pruriens*, although a few curved accessions (5.26%) were observed even in case var. *pruriens.* For plant hairiness, 10.52% of the accessions had low, 18.42% moderate and 71.05% accession had dense pod hair in var. *pruriens*. In case of var. *utilis*, 25% had very low and 75% had low pod hairs. Among seed traits, one (2%) accession produced round seed, 34% were oval and 64% showed cuboid shape.

In comparison, seed coat colour and pattern showed considerable variability in both the varieties. In var. *pruriens*, seed coat colour ranging from black, dark brown, brown and grayish brown were observed except white. In case of var. *utilis*, only two types of seed coat colour - white and grayish-brown were observed. From the total accessions, 54% produced black seed (mostly var. *pruriens*), 22% produced white seed (only var. *utilis*), remaining 24% produced combination of both (Fig. 8a and b).

In case of seed coat pattern, 44% of the accessions showed black mottled seeds followed by 26% non-mottled ones. Along with this, 4-8% accessions showed brown/dark brown mottled/ striped and grey mottled seed coat patterns. Large variability for this trait was observed in both the var. *pruriens* and var. *utilis*.

Table 11 Variability for different qualitative characters

Qualitative characters	Number of accessions representing the characters					
	Count			%		
Leaf texture/ surface	**var. *pruriens***	**var. *utilis***	**Total**	**var. *pruriens***	**var. *utilis***	**Total**
Membranous	16	12	28	42.10	100.00	56.00
Semi-coriaceous	2	0	2	5.26	0.00	4.00
Coriaceous	20	0	20	52.63	0.00	40.00
Leaflet shape						
Ovate	31	5	36	81.58	41.67	72.00
Ovate-lanceolate	7	7	14	18.42	58.33	28.00
Flower colour						
White	1	1	2	2.63	8.33	4.00
Purple	37	11	48	97.37	91.67	96.00
Pod trichome colour						
Creamish white	1	9	10	2.63	75.00	20.00
Velvety black	0	3	3	0.00	25.00	6.00
Golden orange	37	0	37	97.36	0.00	74.00
Pod sensation						
Itching	38	0	38	100.00	0.00	76.00
Non-itching	0	12	12	0.00	100.00	24.00
Pod curvature						
Curved	2	12	14	5.26	100.00	28.00
S-shaped	36	0	36	94.73	0.00	72.00
Plant hairiness						
Very low	0	3	3	0.00	25.00	6
Low	4	9	13	10.52	75.00	26
Moderate	7	0	7	18.42	0.00	14
Dense	27	0	27	71.05	0.00	54
Seed shape						
Round	0	1	1	0.00	8.33	2.00
Oval	14	3	17	36.84	25.00	34.00
Cuboid	24	8	32	63.16	66.67	64.00
Seed colour						
Black	27	0	27	71.10	0.00	54.00
White	0	11	11	0.00	91.67	22.00
Dark brown	4	0	4	10.53	0.00	8.00
Brown	4	0	4	10.53	0.00	8.00
Grayish brown	3	1	4	7.89	8.33	8.00
Seed coat pattern						
Absent/non-mottled	7	6	13	18.42	50.00	26.00
Black mottled	22	0	22	57.89	0.00	44.00
Brown mottled	3	1	4	7.89	8.33	8.00
Dark brown mottled	3	3	6	7.89	25.00	6.00
Grey mottled	1	1	2	2.63	8.33	4.00
Dark brown striped	2	1	3	5.26	8.33	6.00

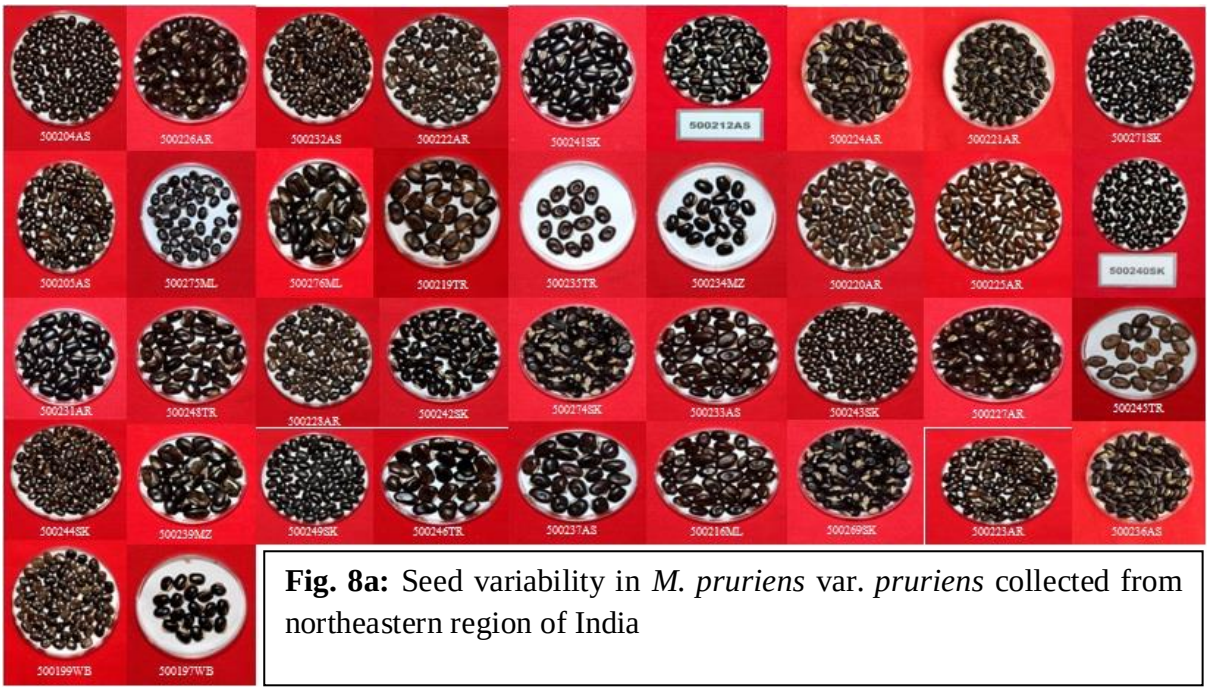


Fig. 8a: Seed variability in *M. pruriens* var. *pruriens* collected from northeastern region of India

Fig. 8b: Seed variability in *M. pruriens* var. *utilis* from northeastern region of India

Correlation analysis

The binary data of the qualitative characters was used to calculate the degree of linear correlation between each pair of characters. Pearson's correlation coefficient (r^2) corresponding to classical linear plot was used to determine how much of the variability of one qualitative trait was explained by the other. The p value for each coefficient was used to test significantly different correlations at 5% probability level and 95% confidence. The results showed that pod curvature and leaf texture were correlated at r^2= 0.463. Pod sensation and leaflet shape were highly correlated at r^2= 0.693. The seed characters were positively correlated with pod characters, specially the seed coat colour and pod sensation (r^2= 0.609) and seed shape and pod hairiness (r^2= 0.565). But significant negative correlation was found between the seed colour and pod curvature as well as seed coat pattern and seed colour. The summary of this analysis is given in Table 12.

Table 12 Pearson's correlation analysis for 10 qualitative characters

Qualitative traits	LT	LS	FC	PTC	PS	PC	PH	SS	SC	SCP
LT	**1**	**-0.373**	0.181	**0.342**	**-0.498**	**0.463**	0.224	0.127	0.071	0.066
LS	**-0.373**	**1**	**-0.327**	**-0.690**	**0.693**	**-0.603**	-0.218	0.089	**0.318**	-0.240
FC	0.181	**-0.327**	**1**	**0.408**	-0.124	0.100	**0.378**	-0.029	-0.016	-0.121
PTC	**0.342**	**-0.690**	**0.408**	**1**	**-0.656**	**0.690**	**0.295**	-0.071	**-0.341**	0.160
PS	**-0.498**	**0.693**	-0.124	**-0.656**	**1**	**-0.901**	-0.252	-0.254	**0.609**	**-0.414**
PC	**0.463**	**-0.603**	0.100	**0.690**	**-0.901**	**1**	0.218	0.229	**-0.497**	**0.341**
PH	0.224	-0.218	**0.378**	**0.295**	-0.252	0.218	**1**	**0.565**	-0.105	0.042
SS	0.127	0.089	-0.029	-0.071	-0.254	0.229	**0.565**	**1**	-0.155	-0.085
SC	0.071	**0.318**	-0.016	**-0.341**	**0.609**	**-0.497**	-0.105	-0.155	**1**	**-0.551**
SCP	0.066	-0.240	-0.121	0.160	**-0.414**	**0.341**	0.042	-0.085	**-0.551**	**1**

Cluster analysis

Agglomerative Hierarchical Clustering (AHC) based on UPGA provides grouping of accessions using mean scores of dissimilarity (Euclidean distance). In our study, cluster analysis based on 10 qualitative characters divided 50 accessions into three major classes (Fig. 9) depicting within class variance of 61.53% and between class variance of 38.47%. It clearly distinguished var. *pruriens* and var. *utilis* accessions except one var. *utilis* accession *viz*., 500210MN which separated in class-3. This accession possesses unique morphological characters including semi-determinate plant height. The Euclidean distances between different classes are shown in Table 13. The different accessions under each class along with variance explained by them are shown in Table 14.

The UPGMA dendrogram constructed based on Jaccard's similarity coefficient divided 50 *M. pruriens* accessions into two major clusters i.e. cluster I (var. *pruriens*) and cluster II (var. *utilis*) with significant bootstrap support between the two major clusters (Fig. 10). The Jaccard's similarity coefficient varied from 0.05 to 1 with an average of 0.39 suggesting good diversity for the qualitative characters.

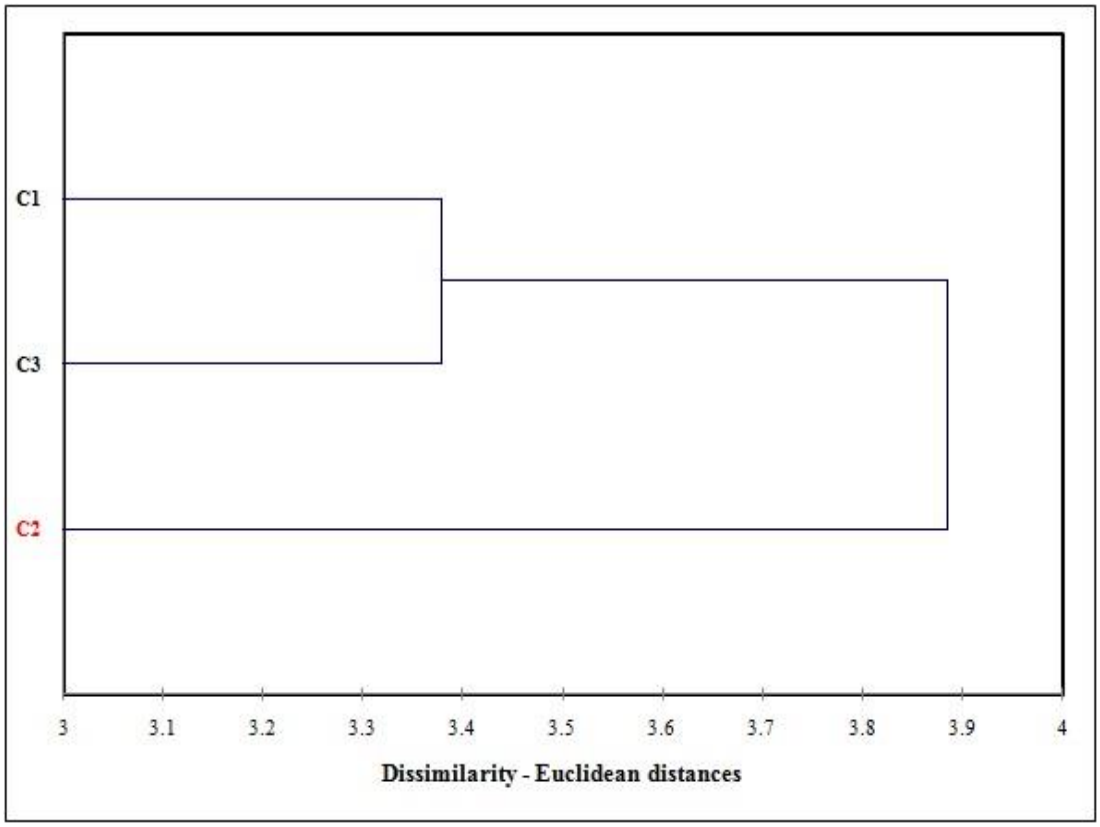


Fig. 9: AHC dendrogram based on 10 qualitative characters

Table 13 Euclidean distance between different AHC classes

Class Number	C1	C2	C3
C1	0	3.194	3.086
C2	3.194	0	3.428
C3	3.086	3.428	0

Table 14 Affiliation of accessions among different AHC classes

Class	C1	C2	C3
Objects	11	38	1
Sum of weights	11	38	1
Within-class variance	2.127	3.175	0.000
Minimum distance to centroid	0.891	1.058	0.000
Average distance to centroid	1.340	1.720	0.000
Maximum distance to centroid	2.169	2.793	0.000
	500203MN	500204AS	500210MN
	500277NL	500205AS	
	500211NL	500276ML	

	500215AR	500275ML	
	500217MN	500212AS	
	500255AR	500216ML	
	500263NL	500219TR	
	500267NL	500220AR	
	500268NL	500221AR	
	500213MN	500222AR	
	500266NL	500223AR	
		500224AR	
		500225AR	
		500226AR	
		500227AR	
		500228AR	
		500231AR	
		500232AS	
		500233AS	
		500234MZ	
		500235TR	
		500236AS	
		500237AS	
		500239MZ	
		500240SK	
		500241SK	
		500242SK	
		500243SK	
		500244SK	
		500245TR	
		500246TR	
		500248TR	
		500274SK	
		500269SK	
		500271SK	
		500197WB	
		500199WB	
		500249SK	

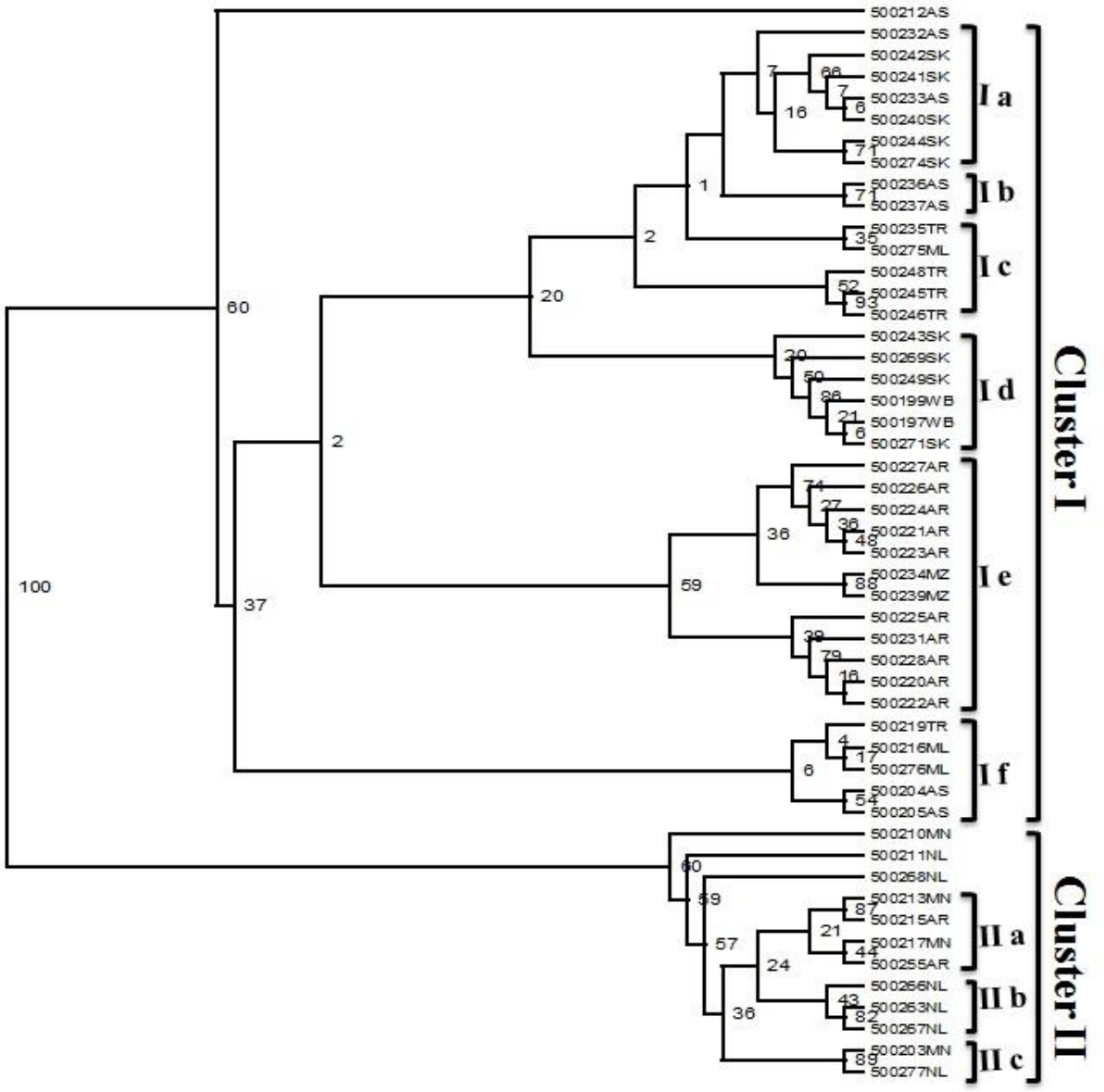


Fig. 10: UPGMA dendrogram based on 10 qualitative characters

Principal Component Analysis (PCA)

Principal component analysis of the data on qualitative characters revealed that first two PCs explained total variation of 43.22% of which 31% was in PC1 and 12.22% in PC2 (Fig. 11). Even though the total variance explained was <50%, the two PCs grouped accessions based on botanical varieties with group A representing var. *utilis* and group B representing var. *pruriens*. The distribution and factor loading of various characters across the two principal axes are shown in Fig. 12.

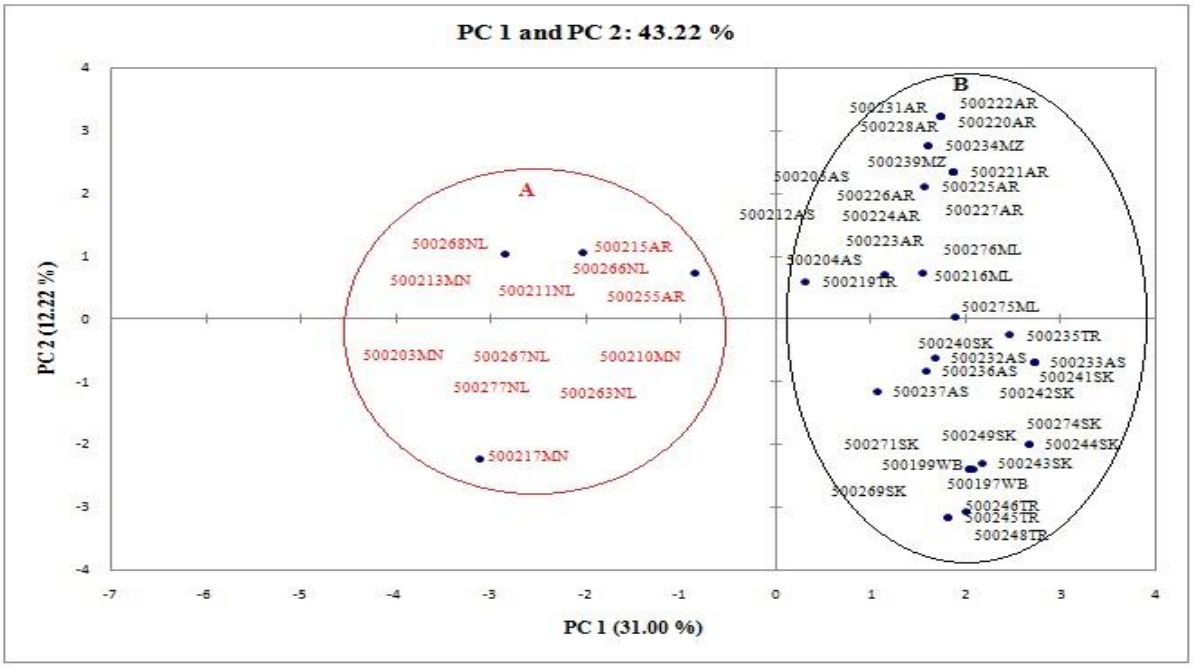


Fig. 11: PCA plot based on 10 qualitative characters (Red - var. *utilis*; Black - var. *pruriens*)

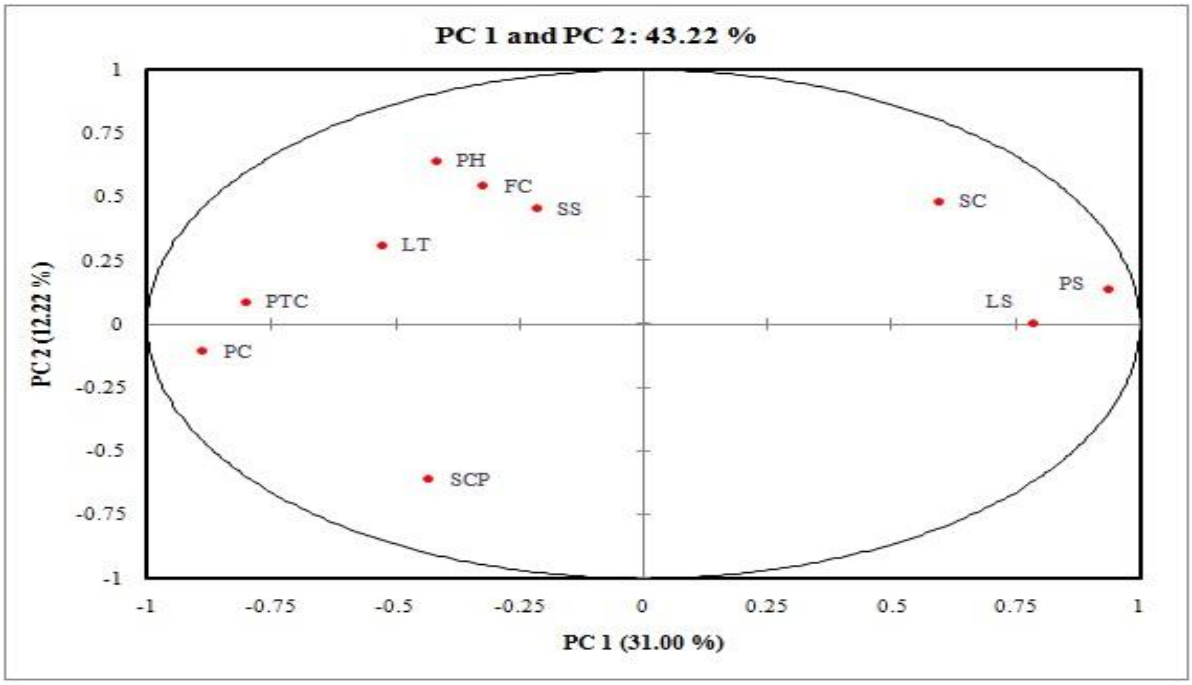


Fig. 12: PCA plot depicting dispersion of qualitative characters contributing to variation

5.1.2.2 Quantitative traits

A total of 10 quantitative traits comprising six yield and four non-yield based traits were analyzed among 50 germplasm accessions. An average of 28% variation was observed within yield based traits such as days to emergence, number of pod per plant, days to flowering, days to first mature pod, days to maturity and 100 seed weight (Table 15a). In case of non-yield based traits such as seed length, seed width, seed thickness and seed area, 33% variation was observed (Table 15b). The maximum and minimum values, range as well as coefficient of variation exhibited by different yield based traits are shown in the Table 15a and b.

Further, we analyzed variability for 10 quantitative traits separately for var. *pruriens* and var. *utilis* and the results are summarized in Tables 16 and 17 respectively. The average coefficient of variation (CV %) for the six yield based traits in 38 wild var. *pruriens* accessions ranged from 8.17 % to 31.73 % with an average of 19.32%, while in case of non-yield based traits, it ranged from 16.92 to 27.16 with an average of 22.10%. In case of 12 cultivated var. *utilis* accessions, it ranged from 6.59 to 20.55 % with a mean of 12.10% for six yield based traits and 11% to 39.46% with a mean of 22.55% in case of four non-yield based traits. In addition, each quantitative trait was subjected to frequency distribution analysis to determine the number of accessions falling into distinct classes (Fig. 13a and b). We also determined variability within each cluster of UPGMA dendrogram. The outputs are captured separately for var. *pruriens* and var. *utilis*.

Table 15a Descriptive statistics of six yield based traits in 50 *M. pruriens* accessions

Variable	Minimum	Maximum	Range	Mean	SD	SE	% CV
Days to emergence	500255 AR (4.00)	500205AS (13.00)	9.00	7.44	2.07	0.30	27.85
Days to Flowering	500221AR (93.00)	500269SK (132.00)	39.00	111.84	10.32	1.46	9.23
Days to first mature pods	500210MN (103.00)	500271SK (163.00)	60.00	138.46	12.27	1.74	8.86
No. of pods per plant	500212AS (14.00)	500248TR (68.00)	54.00	33.08	13.44	1.90	40.62
Days to maturity	500210MN (137.00)	500271SK (197.00)	60.00	170.48	14.31	2.02	8.39
Hundred seed weight (g)	500216ML (18.88)	500266NL (173.36)	154.48	57.94	40.55	5.73	69.98
						Min	**8.39**
						Max	**69.98**
						Ave	**28.00**

Table 15b Descriptive statistics of four non-yield based traits in 50 *M. pruriens* accessions

Variable	Minimum	Maximum	Range	Mean	SD	SE	%CV
Seed length (mm)	500222AR (5.47)	500266NL (17.83)	12.36	10.50	3.01	0.43	28.67
Seed thickness (mm)	500242SK (3.84)	500266NL (10.71)	6.87	6.62	1.84	0.26	27.78
Seed area	500222AR (27.24)	500266NL (173.10)	145.86	70.49	36.07	5.10	51.18
seed width (mm)	500232AS (4.48)	500266NL (12.75)	8.27	8.19	2.09	0.30	25.56
						Min	**25.56**
						Max	**51.17**
						Ave	**33.30**

Table 16 Descriptive statistics of yield and non-yield based traits among *M. pruriens* var. *pruriens* accessions

Yield based traits	Minimum	Maximum	Mean	SD	SE	%CV
Days to emergence	500220AR (4)	500205AS (13)	7.84	2.13	0.34	27.1
Days to Flowering	500221AR (93)	500269SK (132)	112.82	11.03	1.79	10
Days to first mature pods	500221AR (117)	500271AS (163)	140.87	12.08	1.96	8.57
No. of pods per plant	500212AS (14)	500248TR (68)	37.76	11.98	1.94	31.73
Days to maturity	500237AS (145)	500271SK (197)	172.03	14.06	2.28	8.17
Hundred seed weight (g)	500216ML (18.88)	500212AS (67.06)	36.68	11.14	1.81	30.37
					Min % CV	**8.17**
					Max % CV	**31.73**
					Average % CV	**19.32**
Non yield based traits	**Minimum**	**Maximum**	**Mean**	**SD**	**SE**	**%CV**
Seed length (mm)	500222AR (5.47)	500212AS (13.54)	9.43	1.60	0.26	16.92
Seed thickness (mm)	500242SK (3.84)	500221AR (9.89)	6.03	1.64	0.27	27.16
Seed area	500222AR (27.24)	500212AS (85.3)	55.77	12.65	2.05	22.68
seed width (mm)	500232AS (4.48)	500223AR (11.8)	7.37	1.59	0.26	21.59
					Min % CV	**16.92**
					Max % CV	**27.16**
					Average % CV	**22.10**

Table 17 Descriptive statistics of yield and non-yield based traits among *M. pruriens* var. *utilis* accessions

Yield based traits	Minimum	Maximum	Mean	SD	SE	%CV
Days to emergence	500255AR (4)	500268NL (8)	6.167	1.27	0.37	20.55
Days to Flowering	500210MN (94)	500217MN (119)	108.75	7.18	2.07	6.59
Days to first mature pods	500210MN (103)	500213MN (141)	130.83	9.84	2.84	7.52
No. of pods per plant	500263NL (14)	500203MN (22)	18.25	2.38	0.69	13.04
Days to maturity	500210MN (137)	500217MN (189)	165.58	14.61	4.21	8.82
Hundred seed weight (g)	500203MN (100.74)	500266NL (173.36)	125.26	20.02	5.78	15.98
					Min % CV	**6.59**
					Max % CV	**20.55**
					Average % CV	**12.10**
Non yield based traits	**Minimum**	**Maximum**	**Mean**	**SD**	**SE**	**%CV**
Seed length (mm)	500211NL (8.15)	500266NL (17.83)	13.86	3.94	1.14	28.42
Seed thickness (mm)	500211NL (7.48)	500266NL (10.71)	8.52	0.96	0.28	11.33
Seed area	500211NL (56.32)	500266NL (173.1)	117.13	46.22	13.34	39.46
seed width (mm)	500213MN (9.33)	500266NL (12.75)	10.78	1.17	0.34	11
					Min % CV	**11**
					Max % CV	**39.46**
					Average % CV	**22.55**

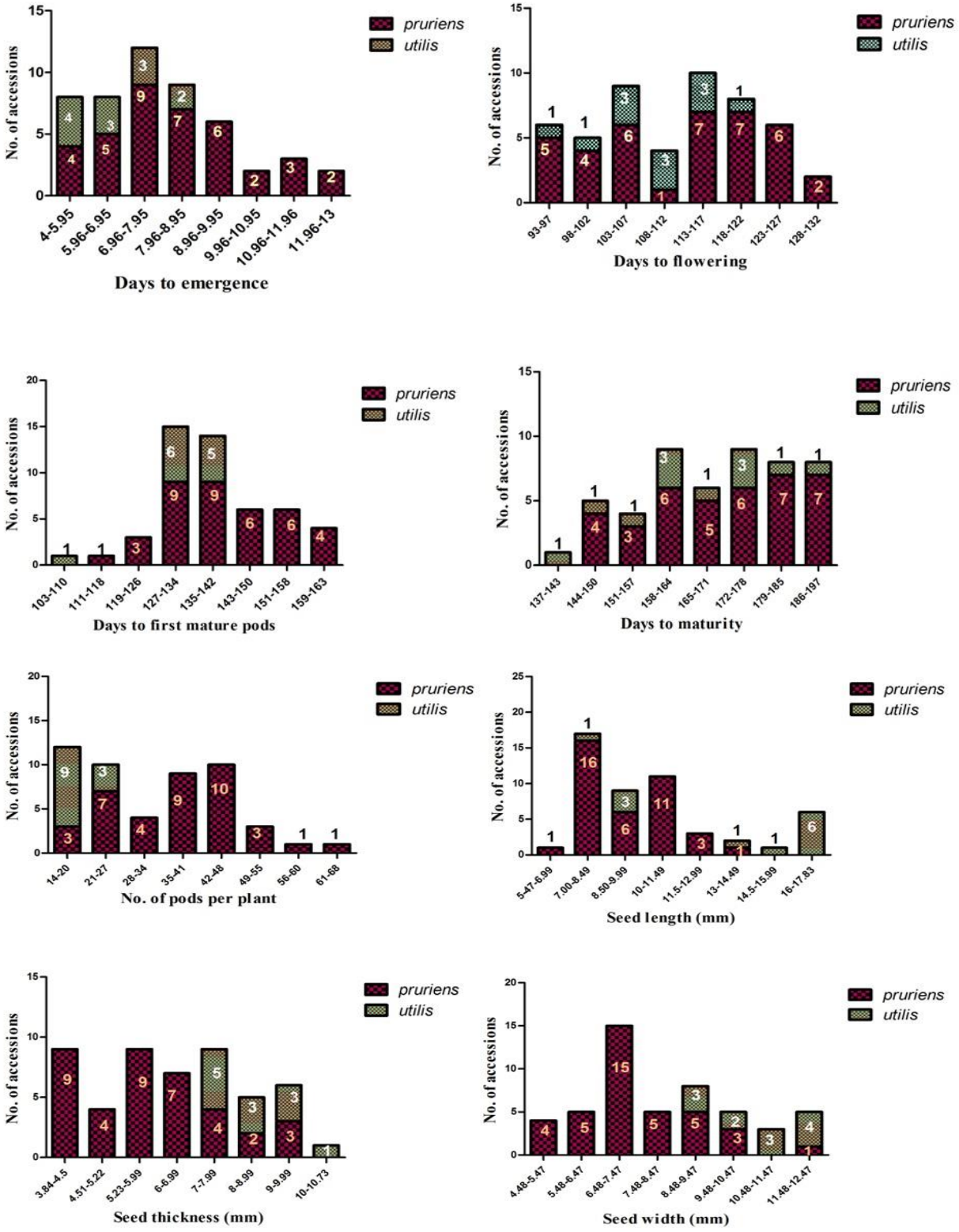


Fig. 13a Results of frequency distribution analysis for different quantitative traits

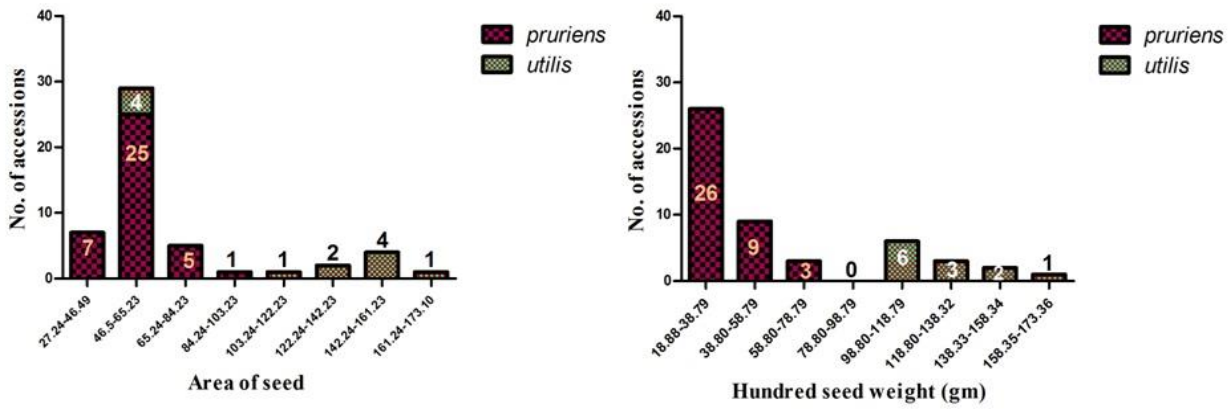


Fig. 13b: Results of frequency distribution analysis for different quantitative traits

ANOVA

Results of one way analysis of variance of different quantitative traits are shown in Table 18. The F ratio obtained from each trait at prob>F classified them into significant, highly significant and very highly significant classes in terms of their contribution to the total variation at $\alpha = 0.05$.

In case of yield based traits, significant variation was observed for traits such as days to emergence, days to flowering, days to first mature pod, number of pod per plant. Moderate variation was recorded for 100 seed weight. Likewise, in case of non-yield based traits, seed length, seed thickness, seed width was found to have minimal variation, whereas seed area showed moderate variation. Multiple comparisons of mean values of traits across different accessions using Tukey's HSD test determined the group of accessions which were significantly different from other.

Table 18 Result of ANOVA for 10 quantitative traits

Traits	F- ratio	F-test	Prob>F
Days to emergence	205.53	***	< 0.0001
Days to flowering	215.84	***	< 0.0001
Days to first mature pod	256.20	***	< 0.0001
Number of pod per plant	103.70	***	< 0.0001
Days to maturity	264.8	***	< 0.0001
Seed length (mm)	1.46	NS	1
Seed thickness (mm)	0.59	NS	1
Seed width(mm)	0.87	NS	1
Seed area	51.34	**	< 0.0001
100 seed weight (g)	42.65	***	< 0.0001

*Significant, ** highly significant, *** very highly significant, NS- Non-significant at probability levels of 5%

Correlation analysis

Correlation tests were carried out separately for both yield and non-yield based traits (Table 19). Highest positive correlation was observed between days to flowering and days to first mature pod at $r^2 = 0.872$ followed by days to maturity at $r^2= 0.52$. Days to first mature pod was positively correlated to days to maturity at $r^2= 0.627$. Likewise, positive correlation were observed between days to emergence and days to flowering ($r^2= 0.350$) and days to first mature pod ($r^2= 0.385$). However 100 seed weight was negatively correlated with days to emergence ($r^2= -0.399$), days to flowering ($r^2= -0.284$), days to first mature pod ($r^2= -0.407$), number of pod per plant ($r^2= -0.606$) and days to maturity ($r^2= -0.188$). In case of non-yield based traits (Table 20), significant positive correlations were observed in for all the traits. Seed length with highest being with seed area ($r^2 = 0.941$), followed by seed width ($r^2 = 0.593$) and seed thickness ($r^2 = 0.457$). Seed thickness was also positively correlated with seed area ($r^2 = 0.564$) and seed width ($r^2 = 0.695$).

Table 19 Pearson's correlation matrix for six yield based quantitative traits

Variables	DE	DF	DFMP	NPPP	DM	HSW
Days to emergence	**1**	**0.0128**	**0.0057**	0.5751	0.9718	**0.0041**
Days to Flowering	**0.3498**	**1**	**< 0.0001**	0.2288	**0.0001**	**0.0452**
Days to first mature pods	**0.3851**	**0.8723**	**1**	0.1606	**< 0.0001**	**0.0033**
No. of pods per plant	-0.0812	0.1733	0.2015	**1**	**0.0062**	**< 0.0001**
Days to maturity	0.0051	**0.5220**	**0.6270**	**0.3819**	**1**	0.1899
Hundred seed weight (g)	**-0.3995**	**-0.2845**	**-0.4071**	**-0.6056**	-0.1885	**1**

Note: The bottom triangular matrix gives the Pearson's correlation coefficient between each pair of traits and the upper triangular matrix gives the p-values for significance. Values in bold are different from 0 with significant level alpha=0.05.

Table 20 Pearson's correlation matrix for four non-yield based quantitative traits

Variables	SL	ST	SA	SW
Seed length (mm)	**1**	**0.0009**	**0.0000**	**0.0000**
Seed thickness (mm)	**0.4566**	**1**	**< 0.0001**	**< 0.0001**
seed width (mm)	**0.5934**	**0.6948**	**1**	**< 0.0001**
Seed area	**0.9405**	**0.5638**	**0.6373**	**1**

Note: The bottom triangular matrix gives the Pearson's correlation coefficient between each pair of traits and the upper triangular matrix gives the p-values for significance. Values in bold are different from 0 with significant level alpha=0.05.

Cluster analysis

AHC clustering based on Euclidean distance of dissimilarity classified 10 quantitative traits into three major clusters (Fig. 14). The variance between the classes was high at 74.26% when compared to within class variance which was at 25.74%. The results indicate each cluster was significantly different from the other. Within dendrogram, var. *utilis* accessions were separated in two classes' i.e class 2 and class 3 and var. *pruriens* separated clearly in class one except one accession 500212AS, which clustered with class 2 representing var. *utilis*. The Euclidean distance between different AHC classes are shown in Table 23 and the affiliation of accessions for each cluster is represented in Table 24.

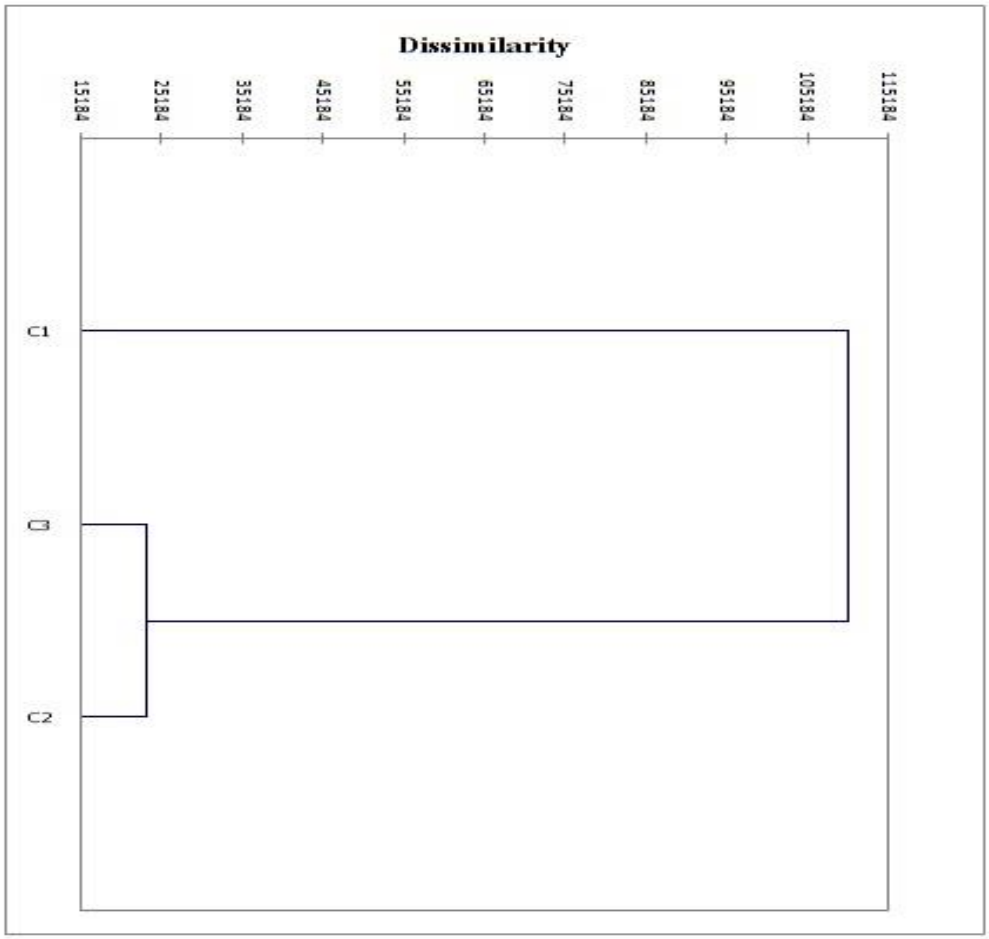


Fig. 14: Illustration of the AHC based on 10 quantitative traits

Table 21 Euclidean distances between different AHC classes

Class	C1	C2	C3
C1	0	77.911	139.129
C2	77.911	0	84.731
C3	139.129	84.731	0

Table 22 Affiliation of accessions for different AHC classes

Class	1	2	3
Objects	37	6	7
Sum of weights	37	6	7
Within-class variance	844.151	1256.199	1169.639
Minimum distance to centric	11.263	13.099	15.730
Average distance to centric	27.584	29.859	29.371
Maximum distance to centric	45.697	46.588	47.426
	500197WB	500212AS	500210MN
	500199WB	500203MN	500213MN
	500204AS	500217MN	500263NL
	500205AS	500211NL	500266NL
	500232AS	500268NL	500267NL
	500233AS	500277NL	500215AR
	500236AS		500255AR
	500237AS		
	500216ML		
	500275ML		
	500276ML		
	500219TR		
	500248TR		
	500235TR		
	500246TR		
	500245TR		
	500234MZ		
	500239MZ		
	500220AR		
	500221AR		
	500222AR		
	500223AR		
	500224AR		
	500225AR		
	500226AR		
	500227AR		
	500228AR		
	500231AR		

	500240SK		
	500241SK		
	500242SK		
	500243SK		
	500244SK		
	500249SK		
	500269SK		
	500271SK		
	500274SK		

Jacquard's similarity coefficients for quantitative traits were found in the range of 0.05 to 1 with an average of 0.54 indicating moderate to high diversity (46%) for quantitative traits among the studied accessions. The diversity within var. *pruriens* was 39%, which was higher than var. *utilis* which was 31%. The dendrogram based on Jaccard's similarity coefficient divided 50 accessions into two main clusters i.e. cluster I and cluster II (Fig. 15). The grouping pattern indicated partition based on botanical varieties as the two cluster, I and II comprised var. *pruriens* and var. *utilis* accessions distinctly. Each major cluster was further divided into sub-clusters namely Ia, Ib, Ic and Id in case of cluster I and 2a and 2b in case of cluster II. Clear cut differences were observed with respect to studied quantitative traits in terms of their mean value in each sub-cluster obtained (Table 23). However, no evidence of grouping based on trait similarities was observed in our UPGMA analysis which indicated that the accessions under each sub-cluster were quite diverse in terms of their average trait values.

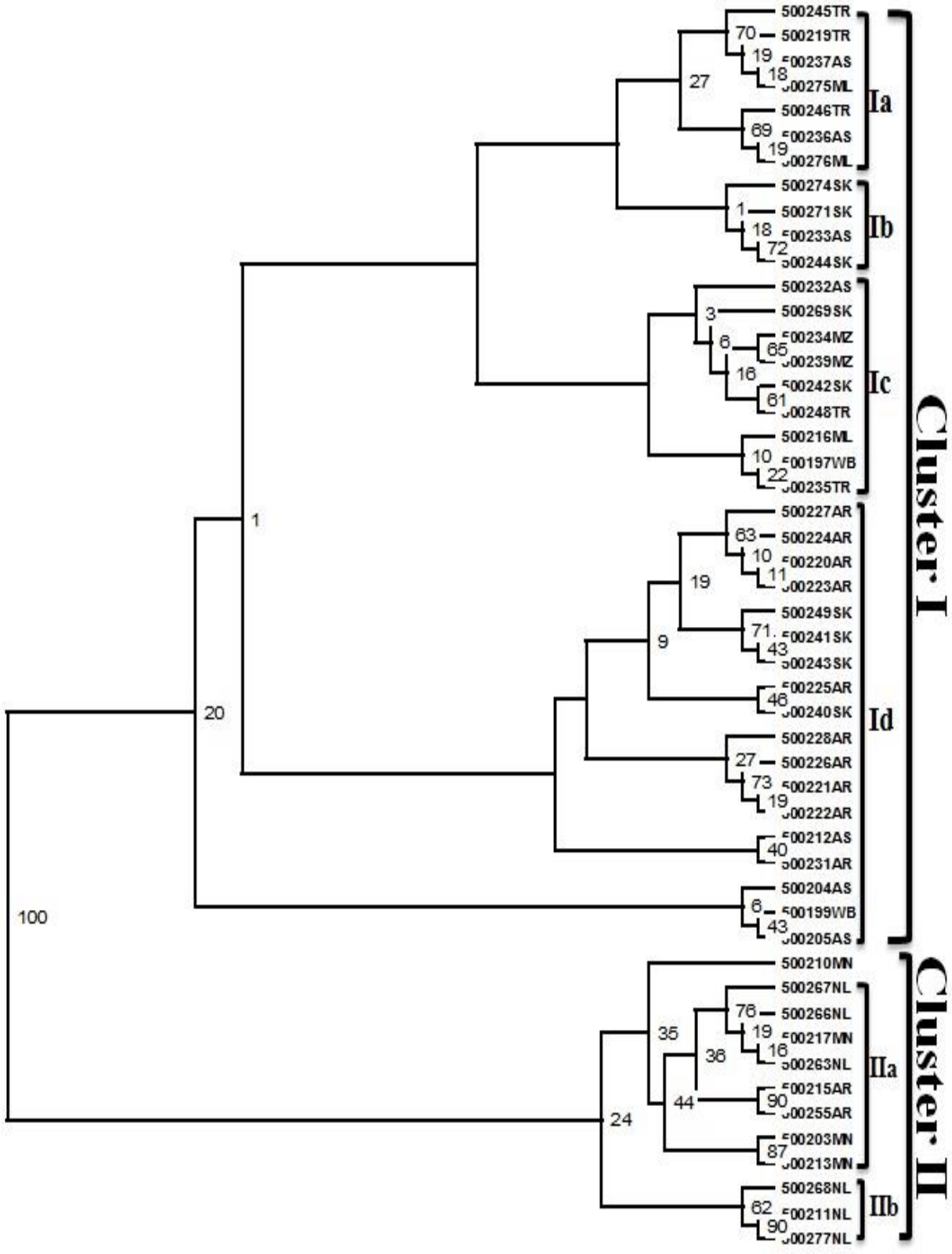


Fig. 15: UPGMA dendrogram based on 10 quantitative characters

Table 23 Quantitative traits mean value from six individual clusters of UPGMA dendrogram

Traits	Cluster (number of accessions)					
	1a (7)	1b (4)	1c (9)	1d (18)	2a (9)	2b (3)
DE	8.71 ± 1.38	9.5±2.10	8±1.58	7.11±2.36	6.11±1.45	6.33±0.58
DF	119.43±5.94	122.5±5.10	118±8.13	105.5±10.27	109.22±8.32	107.33±1.53
DFMP	147.29±7.25	156.75±5.91	142.67±10.42	133.94±10.63	131.56±11.41	128.67±1.53
NPPP	34.14±12.5	39.75±11.24	38.33±14.60	38.44±11.30	18.33±2.34	18.00±3.00
DM	167±20.1	181.25±14.25	175±10.76	170.44±12.65	167.89±15.87	158.67±8.39
SL	8.76±0.60	9.27±1.57	8.38±0.47	10.26±1.83	15.58±2.82	8.7±0.48
ST	7.52±1.34	6.10±1.26	4.30±0.27	6.29±1.52	8.63±1.01	8.18±0.90
SA	58.27±5.92	50.59±8.59	49.42±0.27	59.11±15.10	136.68±34.87	58.47±2.30
SW	7.52±0.78	6.16±1.10	5.85±0.81	8.35±1.48	10.71±1.16	11±1.45
HSW	30.01±6.68	34.39±4.26	29.21±6.60	43.52±11.49	127.90±22.72	117.32±3.77

Note: DE=days to emergence; DF= days to flowering; DFMP= days to first mature pod; NPPP = number of pod per plant; DM= days to maturity; SL= seed length (mm); ST= seed thickness (mm); SA= seed area; SW= seed width (mm); HSW=hundred seed weight (g).

Principal Component Analysis (PCA)

PCA indicated first two PCs i.e. PC1 and PC2 explained a total variation of 61.63% of which 44.5% was from PC1 alone. The distribution and factor loading of various characters across two principal axes is shown in the Fig. 16 and 17. A clear grouping of var. *pruriens* and var. *utilis* accessions into two major axes was observed which is in accordance with the results of UPGMA analysis.

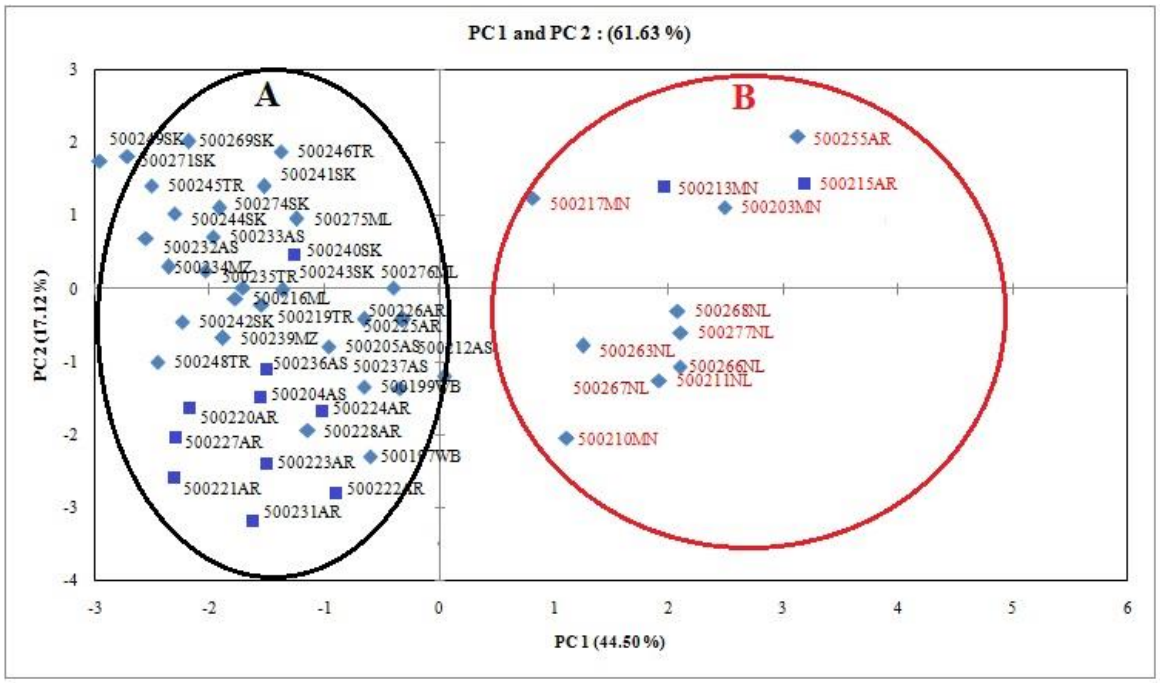


Fig. 16: PCA plot based on quantitative characters
(Blue- *M. pruriens* var. *pruriens*; Red- *M. pruriens* var. *utilis*)

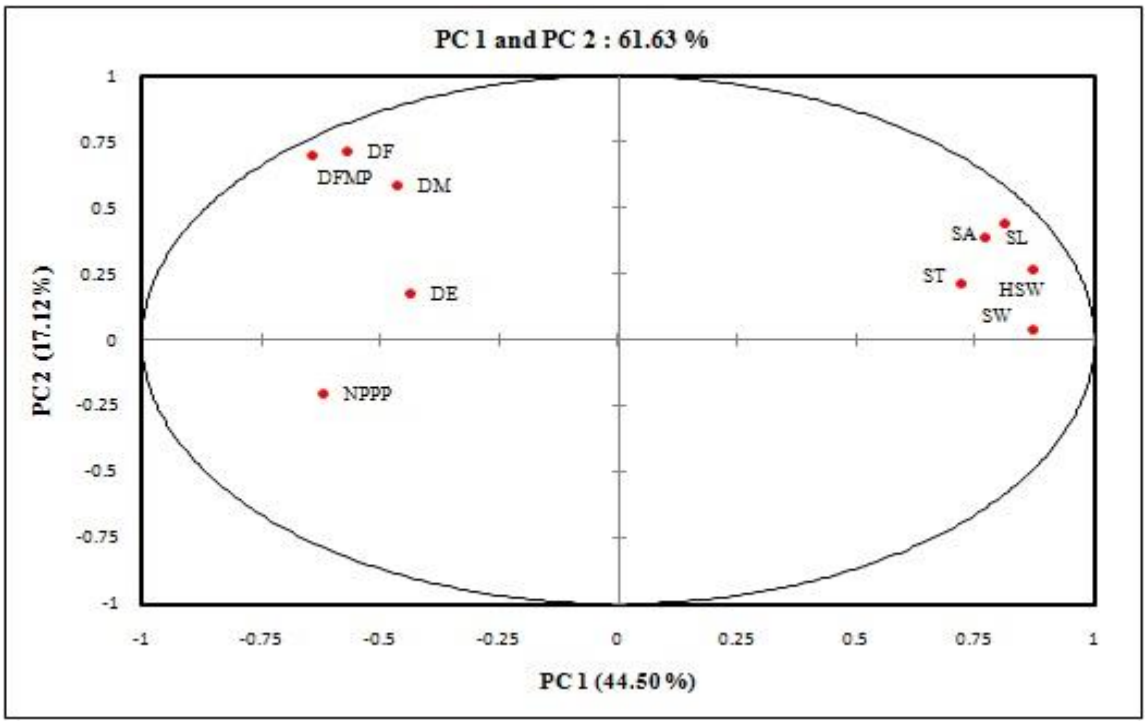


Fig. 17: PCA plot depicting dispersion of quantitative characters contributing to variation

5.1.2.3 Identification of elite stocks and parents of mapping population

We evaluated totally 10 quantitative traits which are important for breeding superior varieties. Thus a detailed analysis was carried out to ascertain the variability for these traits among the on hand accessions and identify key variants in each group based on mean difference of their trait scores.

Days to emergence (DE) showed significant variation ranging from 4-13 days among the different accessions. After evaluation, 20% of accessions were found to be early emerging (<5 days) and 80% were late emerging (>5 days). From the overall results, var. *pruriens* accessions: 500220AR, 500222AR and 500225AR (4 days), 500231AR (5 days) and var. *utilis* accessions: 500255AR (4 days), 500210MN, 500215AR (5 days), 500203MN (6 days) were identified as early emerging genotypes.

Number of days to flowering varied from 93 to 132 days. About 24% (9) of the accessions were under early flowering category and remaining 76% (29) belonged to

late flowering types. Based on the data obtained, var. *utilis* accessions: 500210MN (94 days), 500215AR (98 days) and var. *pruriens* accessions: 500212AS (97 days), 500240SK (119 days) were identified as early flowering stocks.

The days to first matured pod (DFMP) varied from 103 to 163 days. About 14% of the accessions fell under early category (103-130 days), 74% in the medium range (130-150 days) and 12% showed tendency for late first pod maturity (>150 days). About 55% of var. *pruriens* and 45% var. *utilis* showed early first pod maturity. Nearly, 63% of var. *pruriens* and 58% of var. *utilis* accessions were in the medium range. Based on the overall data it is inferred that, among var. *utilis* accessions: 500210MN (103 days), 500215MN (130 days) and 500211NL (127 days) and in var. *pruriens*, accessions: 500221AR (117 days) and 500222AR (121 days) might be promising for the breeding program.

Days to maturity (DM) among the studied accessions varied from 137-197 days. About 40% of the accessions exhibited tendency for early maturity (137-160 days), and about 60% were late maturing (160-197 days). Based on the above data, var. *pruriens* accessions: 500236AS, 500276ML, 500237AS and var. *utilis* accession: 500210MN were identified as early maturing ones and can be used as parental stock for breeding maturity period.

Number of pod per plant (NPPP) varied from 14-68 pods. In this category, all the accessions of var. *utilis* showed minimum (14-22) NPPC values. In var. *pruriens* 11 accessions (29%) were in minimum (14-30 pod per plant) and remaining 27 (71%) were found to be high pod yielding (31-67 pod) accessions. Accessions: 500248TR, 500224AR and 500240SK which showed maximum number of pod per plant may constitute good breeding stock for this trait.

5.2 L-Dopa variability and correlation with seed traits

5.2.1 Standardization of method for L-Dopa analysis

The individual as well as mean values of the trials in triplicates obtained for all the known concentration of L-Dopa are given in Table 24. The mean values at different trials produced minimum standard deviation in each trail confirming the accuracy of the instrument.

5.2.2 L-Dopa screening in germplasm

We applied the above standardized protocol to analyze the L-Dopa content in 50 *M.pruriens* accessions from northeast India and the results are given in Table 25 and Fig. 18. The L-Dopa contents ranged from 1.55% to 9.03% with the coefficient of variation of 46.81%. The highest L-Dopa content was recorded in accession 500240SK (9.03%) and the lowest was in 500212AS (1.55%). Frequency distribution analysis classified 50 accessions into four frequency classes. Majority (>50%) of the var. *utilis* accessions were classified within 1.55 to 5% (low to moderate) category whereas accessions belonging to var. *pruriens* were present in all the frequency classes. Only three accessions figured in high L-Dopa content (5-9%) category. The number of genotypes falling under each frequency class is shown in the Fig. 19.

Table 24 Summary table on validation of HPLC equipment

Sl. no.	Conc. of Std L-Dopa in ppm (Total 1ml)	Vol. of Std. L-Dopa (4mg/ml)	Vol. of MiliQ	Area in triplicate			Mean ± SD (mAU)
				A	B	C	
1	100	100ul	900ul	9.6	8.91	9.02	9.18±0.07
2	200	200ul	800ul	19.869	17.96	18.89	18.91±0.95
3	300	300ul	700ul	28.832	27.26	26.34	27.48±1.26
4	400	400ul	600ul	39.893	39.02	37.64	38.85±1.14
5	500	500ul	500ul	48.276	44.41	46.78	46.48±1.95
6	600	600ul	400ul	59.454	60.12	58.02	59.20±1.07
7	700	700ul	300ul	69.533	71.02	68.57	69.71±1.23
8	800	800ul	200ul	79.368	80.54	76.52	78.81±2.07
9	900	900ul	100ul	90.384	91.56	89.22	90.39±1.17
10	1000	1000ul	0ul	98.685	100.12	97.56	98.78±1.28

Table 25 Variability for L-Dopa content (%) among 50 *M. pru*riens accessions

Sl. No.	Accession No.	L-Dopa (%)	Sl. No.	Accession No.	L-Dopa (%)
1	500197WB	2.84 ± 0.02	**26**	500234MZ	2.32 ± 0.01
2	500199WB	2.69 ± 0.03	**27**	500235TR	3.96 ±0.04
3	500203MN	1.73 ± 0.01	**28**	500236AS	2.91 ± 0.01
4	500204AS	4.02 ± 0.10	**29**	500237AS	2.87 ± 0.03
5	500205AS	2.63 ± 0.02	**30**	500239MZ	2.19 ± 0.04
6	500210MN	1.79 ±0.01	**31**	500240SK	9.03 ± 1.02
7	500211NL	2.13 ± 0.04	**32**	500241SK	2.60 ± 0.02
8	500212AS	1.55 ± 0.10	**33**	500242SK	4.54 ± 0.03
9	500213MN	1.62 ± 0.12	**34**	500243SK	3.68 ± 0.04
10	500215AR	2.88 ± 0.04	**35**	500244SK	2.22 ± 0.02
11	500216ML	3.29 ± 0.04	**36**	500245TR	4.72 ± 0.05
12	500217MN	1.83 ± 0.06	**37**	500246TR	3.72 ± 0.03
13	500219TR	2.54 ± 0.04	**38**	500248TR	2.48 ± 0.05
14	500220AR	2.06 ± 0.04	**39**	500249SK	3.64 ± 0.04
15	500221AR	2.54 ± 0.03	**40**	500255AR	2.31 ± 0.03
16	500222AR	2.38 ± 0.03	**41**	500263NL	1.98 ± 0.06
17	500223AR	2.79 ± 0.04	**42**	500266NL	3.11 ± 0.01
18	500224AR	2.87 ± 0.03	**43**	500267NL	2.64 ± 0.02
19	500225AR	9.02 ± 0.21	**44**	500268NL	3.74 ± 0.04
20	500226AR	3.91 ± 0.01	**45**	500269SK	5.13 ± 0.06
21	500227AR	3.58 ± 0.01	**46**	500271SK	2.93 ± 0.01
22	500228AR	2.98 ± 0.03	**47**	500274SK	2.59 ± 0.01
23	500231AR	3.26 ± 0.01	**48**	500275ML	2.30 ± 0.03
24	500232AS	2.18 ± 0.04	**49**	500276ML	2.58 ± 0.04
25	500233AS	3.68 ± 0.01	**50**	500277NL	2.74 ± 0.01

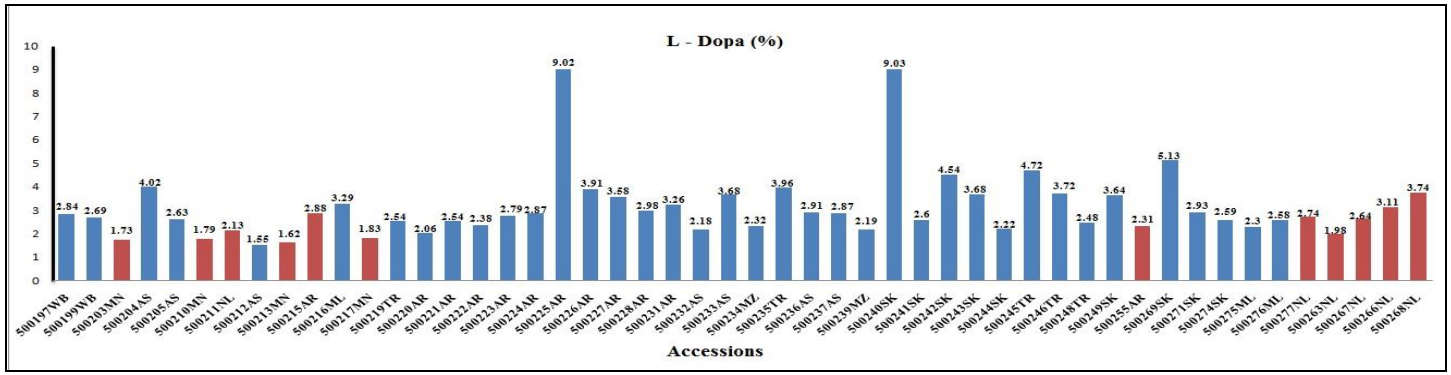


Fig. 18: Diagram showing variability for L-Dopa content in 50 *M. pruriens* accessions

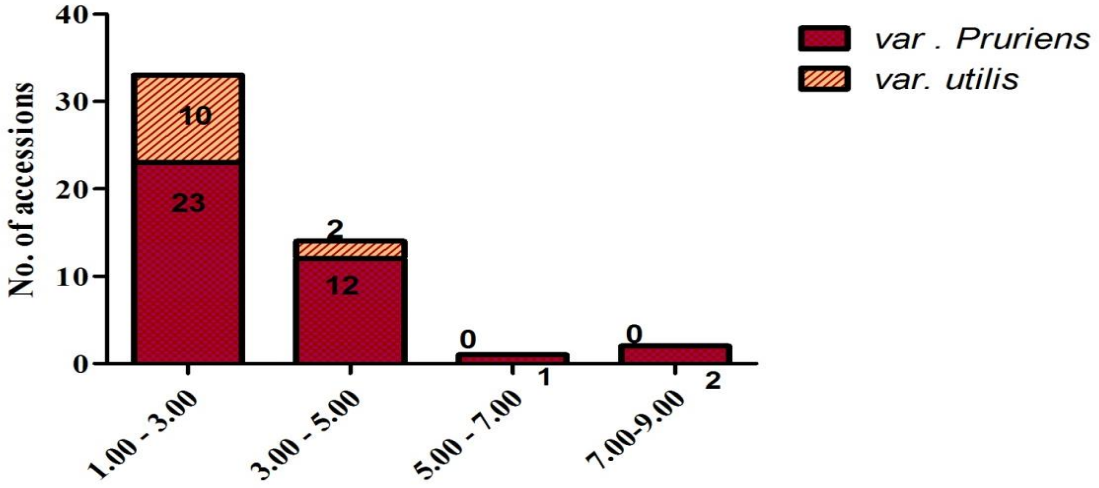


Fig. 19: Frequency distribution analysis based on L-Dopa content

5.2.3 Correlation analysis with seed traits

We analyzed correlation between L-Dopa content with different seed traits. It was revealed that the L-Dopa content (%) was negatively correlated to all the seed characters *viz*., seed length (r^2 = -0.136), seed thickness (r^2 = -0.213), seed area (r^2 = -0.16), seed width (r^2 = -0.153) and hundred seed weight (r^2 = -0.271) (Table 26).

Table 26 Pearson's correlation matrix of L-Dopa content with seed traits

Variables	SL	ST	SW	SA	HSW	LD
Seed length (mm)	**1**	**0.457**	**0.593**	**0.941**	**0.705**	**-0.136**
Seed thickness (mm)	**0.457**	**1**	**0.695**	**0.564**	**0.599**	**-0.213**
Seed width (mm)	**0.593**	**0.695**	**1**	**0.637**	**0.749**	**-0.152**
Seed area	**0.941**	**0.564**	**0.637**	**1**	**0.777**	**-0.160**
Hundred seed weight (g)	**0.705**	**0.599**	**0.749**	**0.777**	**1**	**-0.271**
L- Dopa (%)	**-0.136**	**-0.213**	**-0.153**	**-0.160**	**-0.271**	**1**

Value in bold are different from 0 with significant level alpha=0.95

5.3 Genetic diversity analysis

5.3.1 Marker polymorphism

Of the 26 PCs selected, 22 (84.61%) successfully amplified. These 22 PCs were used to investigate 50 *M. pruriens* accessions. All the PCs generated discrete AFLP profiles - a snapshot of which from the GeneMapper V 4.0 software (Applied Biosystems, USA) is presented in Fig. 20. They generated a total of 1208 fragments in which 1156 (95.70%) were polymorphic. The total number of fragments ranged from 12 (E-CAC/ M-GCT) to 137 (E-ACT/ M-CTC) with an average of 55 fragments per PC. Polymorphic fragments ranged from 11 (E-AAG/ M-CAT) to 127 (E-ACA/ M-CAA) with an average of 52.55. The percentage of polymorphism varied from 82% (E-ACT/M-CAA) to 100% (E-AAC/M-CAG; E-AAC/M-CTA; E-AAC/M-CTC; E-ACA/M-CTT; E-AGG/M-CAG; E-ACG/M-CAC; E-ACA/M-CAT; E-AAG/ M-CAC; and E-ACA/M-CAA) with an average of 97.27%. The fraction of

polymorphic loci was in the range of 0.88 to 1.00 with an average of 0.96. Mean gene diversity for different PCs varied from 0.17 to 0.37 with an average of 0.23.

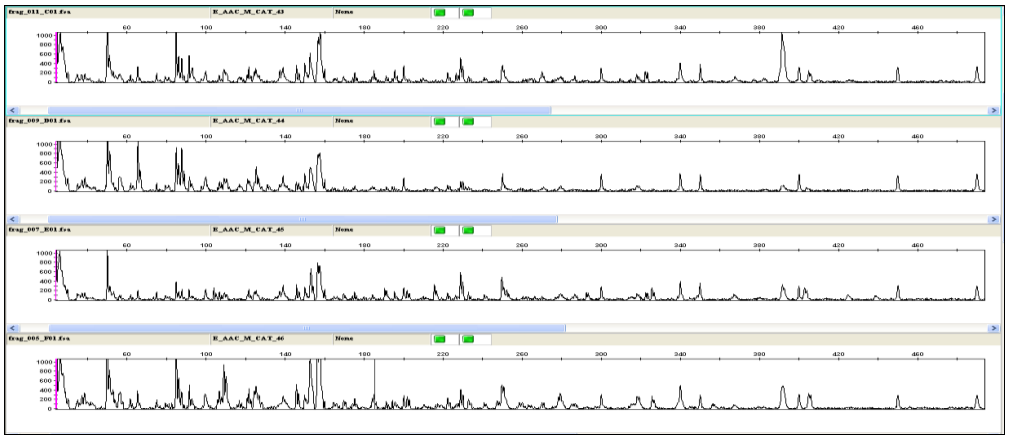


Fig. 20: Representative electropherogram showing AFLP profiles

AFLP profile mainly depends on three major parameters which are enlisted as PIC (polymorphic information content), MI (marker index) and RP (resolving power) of a particular PC (Chandrawati et al. 2014). The calculated PIC values ranged from 0.14 to 0.30 with an average of 0.23 per fragment. The highest PIC value (0.30) was observed for the primer combination: E-ACT/M-CTA and the lowest value (0.14) were recorded for E-ACG/ M-CAG. Based on the PIC values, the AFLP-PCs: E-ACT/M-CTA (0.30), E-ACG/M-CAC; E-AAG/ M-CTC (0.29), E-AGG/M-CTA (0.28), and E-AGG/ M-CAT (0.27) were determined to be more informative than others. The EMR varied from 11.04 (E-AAG/M-CAT) to 127.41 (E-ACA/M-CAA) with an average of 52.51. The MI values varied between 2.42 (E-AAG/ M-CAT) to 28.11 (E-ACG/M-CAC) with an average of 12.20. RP varied from 0.37 (E-AAG/ M-CAT) to 20.30 (E-ACT/M-CTA) with an average of 5.19 (Table 27). The various marker features showed significant correlation to each other. PIC and MI showed a correlation (r^2) of 0.83, $p < 0.005$; MI and RP showed $r^2 = 0.99$, $p < 0.005$ while $r^2 = 0.85$, $p < 0.005$ was the correlation between RP and PIC.

Table 27 Details of marker attributes obtained for different AFLP primer combinations

Sl. No.	AFLP Primer combination	n	nP	β	Gene diversity	PIC	EMR	MI	RP	%P
1	E-ACT/M-CTA	69	62	0.89	0.26	0.30	61.41	18.59	20.30	98.00
2	E-AAC/M-CAG	43	43	1.00	0.37	0.24	43.00	10.32	14.92	100
3	E-AAC/M-CTA	39	36	0.92	0.30	0.23	35.88	8.28	7.87	100
4	E-AAC/M-CTC	50	47	0.94	0.28	0.26	47.00	12.22	9.59	98.00
5	E-ACA/M-CTT	77	68	0.88	0.32	0.22	67.76	14.96	7.60	100
6	E-AGG/M-CAG	37	35	0.95	0.27	0.27	35.15	9.44	3.42	100
7	E-ACG/M-CAC	99	97	0.98	0.34	0.29	97.02	28.11	9.68	100
8	E-ACA/M-CAT	91	89	0.98	0.37	0.23	89.18	20.47	7.41	100
9	E-AAG/M-CAC	99	91	0.92	0.28	0.24	91.08	21.84	4.93	100
10	E-ACA/M-CAA	137	127	0.93	0.29	0.19	127.41	24.13	4.89	100
11	E-ACG/ M-CAA	28	27	0.96	0.22	0.22	26.88	5.94	1.18	96.00
12	E-AAG/ M-CAT	12	11	0.92	0.28	0.22	11.04	2.42	0.37	96.00
13	E-AAG/ M-CTG	53	51	0.96	0.27	0.22	50.88	11.22	2.67	100
14	E-AAG/ M-CTC	68	68	1.00	0.27	0.29	68.00	19.72	2.34	98.00
15	E-AAG/ M-CAA	39	39	1.00	0.36	0.16	39.00	6.24	1.56	98.00
16	E-ACT/ M-CAA	40	39	0.98	0.18	0.14	39.20	5.46	0.56	82.00
17	E-ACG/ M-CAG	38	38	1.00	0.17	0.14	38.00	5.32	0.65	98.00
18	E-AGC/ M-CTA	72	71	0.99	0.24	0.20	71.28	14.20	1.95	94.00
19	E-ACG/ M-CTG	20	20	1.00	0.25	0.21	20.00	4.20	0.53	98.00
20	E-AGG/ M-CAT	41	41	1.00	0.33	0.27	41.00	11.07	1.44	94.00
21	E-AGG/M-CTA	31	31	0.97	0.36	0.28	30.07	8.40	17.34	92.00
22	E-AGC/M-CTG	25	25	1.00	0.28	0.23	25.00	5.75	10.3	98.00
	Min	**12**	**11**	**0.88**	**0.17**	**0.14**	**11.04**	**2.42**	**0.37**	**82.00%**
	Max	**137**	**127**	**1.00**	**0.37**	**0.30**	**127.41**	**28.11**	**20.30**	**100%**
	Avg.	**54.91**	**52.55**	**0.96**	**0.29**	**0.23**	**52.51**	**12.20**	**5.98**	**97.27%**
		1208	**1156**							

n: Total number of fragments; **nP**: Number of polymorphic fragments; **β**: Fraction of Polymorphic Loci; **PIC**: Polymorphic information content; **EMR**: Effective multiplex ratio; **Rp**: Resolving power; **MI**: Marker index; **P**: Percent Polymorphism.

4.5.7 Genetic diversity and population structure

Measures of genetic diversity are provided in Table 28. The effective number alleles (Ne) for different PCs ranged from 1.16 to 1.60 with mean of 1.38. Pair-wise similarity measures estimated using Jaccard's coefficient and Shannon's index (I) produced an average values of 0.37 and 0.34 indicating high genetic diversity in our collection. This is also reflected in the values of h and Ht which produced mean scores of 0.27 and 0.22, respectively. On the contrary, the indices of genetic differentiation (Gst) was low (0.08) and corresponding gene flow value (Nm) was high (6.78) suggesting high levels of gene flow between population groups. As expected, among the two botanical varieties, genetic diversity within var. *pruriens* (wild) was more (I: 0.44) as compared to cultivated var. *utilis* (0.37) (Table 29).

The dendrogram based on UPGMA revealed two major clusters (cluster I and II) accommodating 96% of the accessions; two accessions 500221AR and 500239MZ formed an out-group (Fig. 21a). Among the two clusters, cluster I was the largest with 27 (54%) accessions and grouped mostly collection from Assam plains (AP). This was further divided into three sub-clusters namely IA, IB and IC with 18, 5 and 4 accessions respectively.

Table 28 Diversity indicators for different AFLP markers across all the populations

Primer combinations	Na*	Ne*	h*	I*	Ht*	Hs*	Gst*	Nm*
E-ACT/M-CTA	1.39	1.36	0.07	0.33	0.21	0.08	0.16	2.56
E-AAC/M-CAG	1.67	1.48	0.14	0.42	0.28	0.13	0.14	3.02
E-AAC/M-CTA	1.42	1.44	0.18	0.38	0.25	0.16	0.10	4.57
E-AAC/M-CTC	1.47	1.40	0.20	0.35	0.23	0.16	0.12	3.58
E-ACA/M-CTT	1.69	1.35	0.29	0.35	0.22	0.25	0.07	6.23
E-AGG/M-CAG	1.32	1.35	0.27	0.31	0.21	0.24	0.05	8.69
E-ACG/M-CAC	1.67	1.45	0.27	0.41	0.27	0.23	0.08	5.92
E-ACA/M-CAT	1.69	1.47	0.30	0.42	0.28	0.28	0.08	5.45
E-AAG/M-CAC	1.45	1.34	0.32	0.32	0.21	0.27	0.08	6.07
E-ACA/M-CAA	1.34	1.30	0.31	0.29	0.19	0.29	0.04	12.20
E-ACG/M-CAA	1.28	1.33	0.32	0.29	0.19	0.29	0.08	5.90
E-AAG/M-CAT	1.10	1.23	0.29	0.23	0.14	0.27	0.07	7.17
E-AAG/M-CTG	1.49	1.43	0.30	0.38	0.25	0.28	0.05	9.60
E-AAG/M-CTC	1.38	1.36	0.30	0.32	0.21	0.27	0.07	7.17
E-AAG/M-CAA	1.62	1.43	0.29	0.39	0.26	0.27	0.05	10.06
E-ACT/M-CAA	1.07	1.16	0.30	0.19	0.11	0.26	0.04	13.33
E-ACG/M-CAG	1.07	1.22	0.29	0.21	0.14	0.27	0.07	6.55
E-AGC/M-CTA	1.28	1.30	0.32	0.28	0.18	0.30	0.05	8.66
E-ACG/M-CTG	1.28	1.35	0.29	0.30	0.20	0.27	0.07	6.79
E-AGG/M-CAT	1.62	1.47	0.32	0.41	0.28	0.31	0.05	10.00
E-AGG/M-CTA	1.97	1.60	0.34	0.50	0.34	0.29	0.14	3.09
E-AGC/M-CTG	2.00	1.49	0.29	0.44	0.29	0.24	0.17	2.49
Mean	**1.47**	**1.38**	**0.27**	**0.34**	**0.22**	**0.25**	**0.08**	**6.78**
Std. Dev	**0.26**	**0.10**	**0.07**	**0.08**	**0.06**	**0.05**	**0.03**	**3.02**

***Na**: Number of different alleles; **Ne**: Effective no. of alleles; **h**: Nei's (1973) gene diversity; **I**: Shannon's Information index; **Ht**: Diversity in overall collections total gene diversity; **Hs**: Sub divided population; **Gst**: Genetic differentiation; **Nm**: Estimate of gene flow from Gst or Gcs, Nm= 0.5(1-Gst)/Gst.

Table 29 Gene diversity estimates for groups based on botanical varieties and geographical distribution

Botanical varieties	N*	Na*	Ne*	I*	h*	He*	P (%)
var. *pruriens*	38	1.91	1.48	0.44	0.29	0.25	95.67
var. *utilis*	12	1.45	1.40	0.37	0.24	0.22	71.54
Average	**25**	**1.68**	**1.44**	**0.40**	**0.26**	**0.23**	**83.61**
Geographical distribution							
Eastern Himalayas	11	1.52	1.39	0.35	0.26	0.23	75.26
Assam plains	39	1.88	1.41	0.38	0.28	0.24	93.77
Average	**25**	**1.70**	**1.40**	**0.37**	**0.27**	**0.24**	**84.52**

N: Number of population**; Na**: Number of different alleles; **Ne**: Effective no. of alleles; **I**: Shannon's Information index; **h**: Nei's (1973) gene diversity; **He**: Expected heterozygosity; **P (%)**: Percent polymorphism

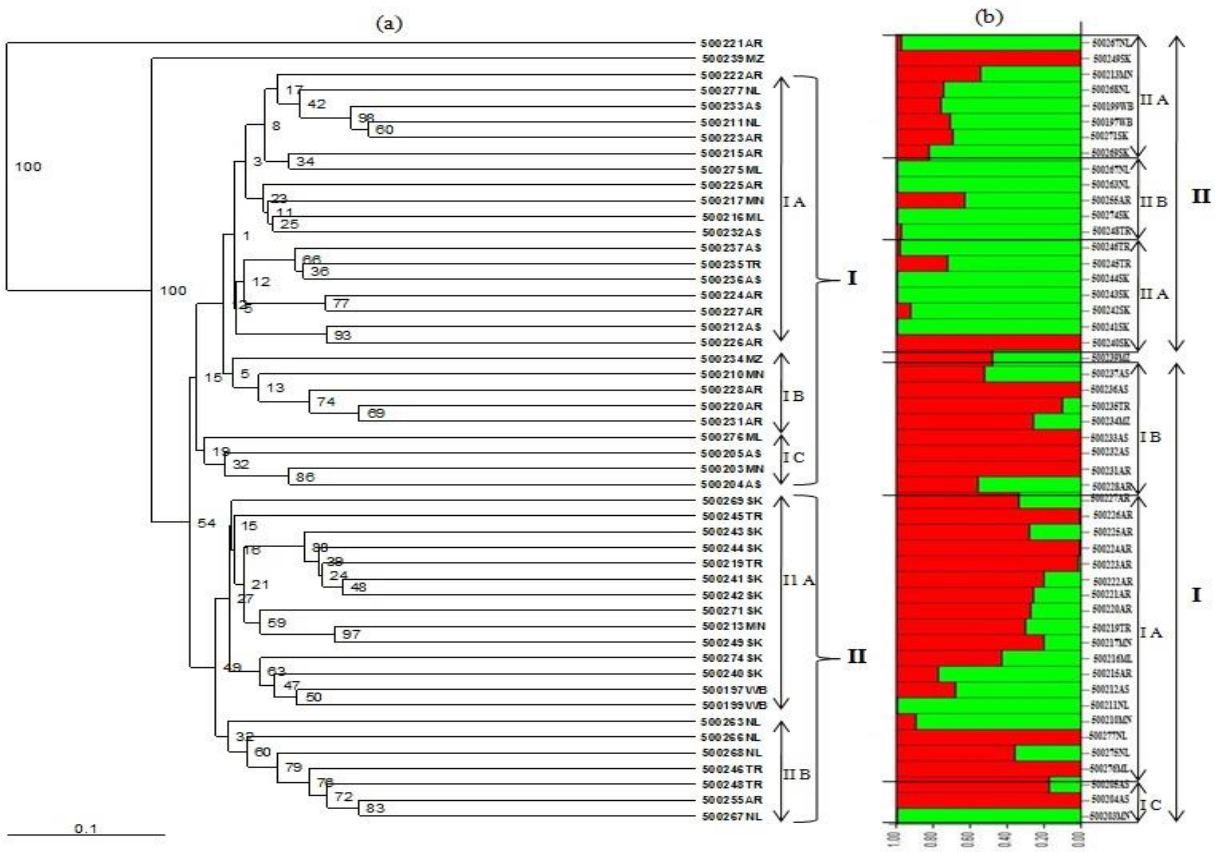


Fig. 21 (a): UPGMA dendrogram of 50 *M. pruriens* accessions based on Jaccard's coefficient and **(b):** hierarchical organization of genetic relatedness analyzed by STRUCTURE V 2.3.4 program. Each colour represents one population and length of segment shows estimated membership proportion of each accession.

Sub-cluster IA contained mix of var. *pruriens* and var. *utilis* accessions while sub-clusters IB and IC grouped majorly var. *pruriens* accessions except one var. *utilis* in each group (500210MN, 500203MN). However, no location specific (states) grouping was observed within cluster-I. Likewise, Cluster II contained 21 (42%) accessions which was further divided into two sub-clusters namely IIA and IIB representing 14 and 7 accessions, respectively. Of these, sub-cluster IIA grouped mostly var. *pruriens* accessions from eastern Himalayan region (EH) except one accessions of var. *utilis* (500213MN) and sub-cluster IIB mostly separated var. *utilis* accessions from AP except two accessions (500246TR and 500248TR) belonging to var. *pruriens*. Thus, though not entirely, both varietal and geographical affiliations emerged to have played

some role in groping of the accessions. The results of PCoA (Fig. 22) also supported the latter findings although total variation explained by PCoA I and II were significantly low.

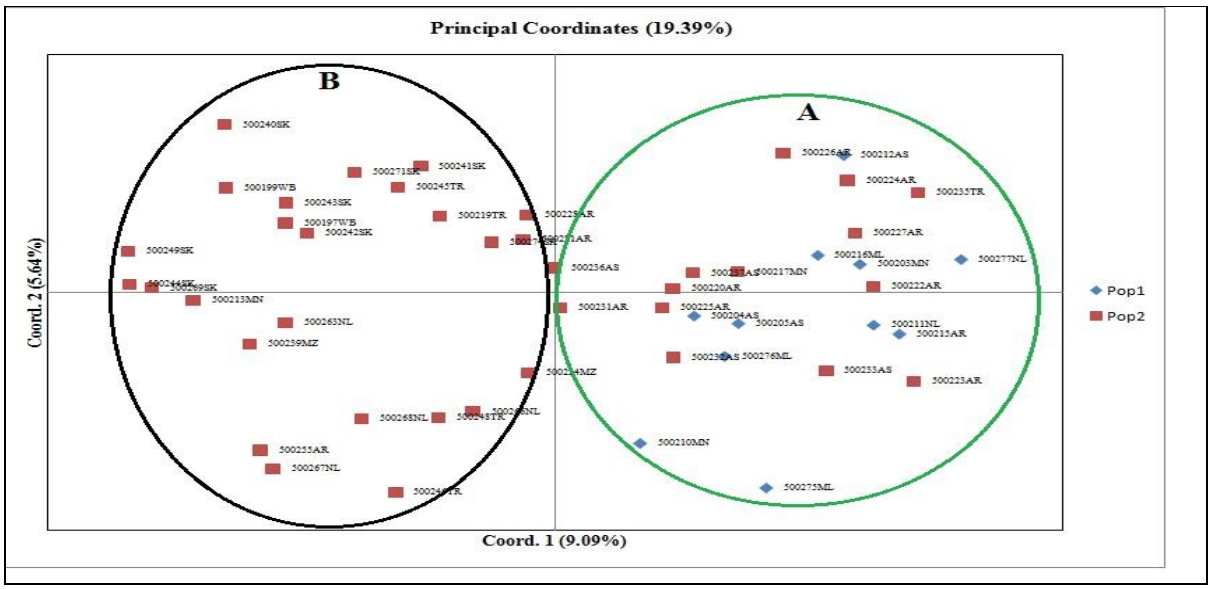


Fig. 22: PCoA plot of 50 *M. pruriens* accessions based on AFLP

(**EH** = Eastern Himalayas; **AP** Assam plains)

Close examination of the UPGMA data also revealed configuration of two independent gene pools representing AP (cluster I) and EH (cluster IIB). Therefore, for AMOVA analysis, we considered two different categories (a) Assam plains and eastern Himalayan region based on geographical distribution and (b) var. *pruriens* and var. *utilis* based on botanical varieties. Partitioning of genetic variations within and between population groups revealed, 7% and 8% of the total genetic variations was accrued from inter-population variation and 92% and 93% variance was contributed by intra-population variations for botanical varieties and geographical distribution categories, respectively (Table 30). Low Fst values (0.08 and 0.07) and high Nm estimates (3.04 and 3.50) for both the categories further supported lack of population structure as a factor of high gene flow.

Table 30 AMOVA results based on botanical varieties as well as geographical distribution

Botanical varieties	Df	SS	MS	Est. Var.	%	Fst	Nm
Among Pops	1	416.02	416.02	13.69	8%	0.08 & P>0.001	3.04
Within Pops	48	7982.26	166.30	166.30	92%		
Total	**49**	**8398.28**		**179.99**	**100%**		
Geographical distribution							
Among Pops	1	374.45	374.45	12.08	7%	0.07 & p>0.001	3.50
Within Pops	48	8023.83	167.16	167.16	93%		
Total	**49**	**8398.28**		**179.24**	**100%**		

Df: Degree of freedom; **SS**: Sum of squares; **MS**: Mean squares; **Fst**: Genetic differentiation among populations; **Nm**: Estimate of gene flow from Gst or Gcs, Nm= 0.5(1-Gst)/Gst.

To corroborate these findings, we carried out population structure analysis based on Bayesian statistics. The results revealed presence of two sub-populations as ΔK was maximum for K=2. The grouping pattern based on StructureHarvester results (Fig. 21b) was broadly in conformity with UPGMA and PCoA results (Fig. 21a and 22). For instance, in case of both UPGMA and PCoA, we could recognize formation of two broad clusters representing Assam plains and eastern Himalayas. The maximum ΔK value obtained for K-2 in structure analysis possibly represented this. However three sub-clusters within cluster-I lacked any pattern of geographical structuring. But significant gene flow was evident as the structure plot revealed several admixture taxa in our collection. Likewise, majority of the var. *utilis* accessions from AP formed part of cluster-IIB even in structure plot.

5.4 Genetic analysis of L-Dopa trait

5.4.1. Analysis of F1 hybrid

Summary of phenotyping data of the parents along with F1 hybrids for different qualitative and quantitative characters are given in Table 31. F1 hybrids flowered late as compared to parents. Consequently, maturity period for pods (DFMP and DM)

were prolonged. L-Dopa content was found to be significantly reduced in F1 hybrid (3.82%) as compared to high L-Dopa parent (9.03%) and was found to be leaning towards lower L-Dopa parent (1.55%). The entire F1 plants produced purple flower colour.

Table 31 Morpho-agronomic characters of parents, F1 hybrid and F2 progenies

Traits	*M. pruriens* var. *pruriens*		F1 (Hybrid)
	500240SK (♂)	500212AS (♀)	
Leaf texture (LT)	Membranous	Membranous	Membranous
Leaf colour (LC)	Dark green	Green	Dark green
Days to flowering (DF)	120	97	137
Flower colour (FC)	Dark purple	White	Dark purple
Pod trichome colour (PTC)	Golden orange	Creamish white	Golden orange
Days to first mature pod (DFMP)	131	118	154
Pod itchiness (PI)	Highly itching	Itching	Highly itching
Days to maturity (DM)	203	193	217
Seed colour (SC)	Black	Black	Black
Seed coat pattern (SP)	Dark black mottled	Dark brown mottled	Dark black mottled
Seed shape (SS)	Rhomboid	Rhomboid	Ovoid
Number of seed per pod (NSPP)	3-4	5-6	4-5
Hundred seed weight (HSW)	33.97 ± 2.47	67.06 ± 2.47	38.43±8.45
Seed area (SA)	56.01	85.30	66.56
Seed length (SL)	10.49±0.82	13.54±0.62	11.34±0.87
Seed thickness (ST)	5.34±0.11	6.30±0.30	5.67±0.48
L-Dopa (LD)	9.03%	1.55%	3.82%

LT, leaf texture; LC, leaf colour; DF, days to flowering; FC, flower colour; PTC, plant trichome colour; DFMP, days to first mature pod; PI, pod itchiness; DM, days to mature; SL, seed length; SW, seed width; ST, seed thickness; SA, seed area; SC, seed colour; SP, seed coat pattern; SS, seed shape; NSPP, no. of seed per plant; HSW, hundred seed weight; LD (L-Dopa).

5.4.2 Analysis of F2 population

The data on segregation of different qualitative characters among the F2 progenies are given in Table 32. The inheritance pattern of most of the qualitative traits confirmed to polygenic inheritance. However in case of flower colour, the segregation ratio from the chi square test (1:5) did not confirm to otherwise monogenic nature of this trait. Column statistics of different qualitative traits are given in Table 33.

Table 32 Segregation of qualitative characters in F2 population

Qualitative Characters	Number of F2 progenies	Qualitative Characters	Number of F2 progenies
Leaf Texture/ Surface		**Pod Sensation**	
Membranous	163	Low- Itching	42
Semi-coriaceous	37	Itching	161
Coriaceous	72	High- Itching	69
Leaflet Shape		**Pod Hairiness**	
Lanceolate	140	Dense	126
Ovate- lanceolate	132	Slight	146
Plant Hairiness		**Pod Line**	
Low	183	Yes	133
Semi dense	24	No	139
Dense	65	**Flower Colour**	
Pod Curvature		Purple	230
Curved	56	White	42
S- shaped	216		
Pod Trichome Colour			
Creamish White (CW)	181		
Both(CW & GO)	18		
Golden Orange(GO)	73		

Table 33 Chi square test for qualitative traits in F2 population

Traits	t test	Df	Significant ($\alpha = 0.05$)?	Coefficient of variation (%)
LS	34.00	1	Yes	4.16%
PC	1.70	1	No	83.199%
PH	13.60	1	Yes	10.40%
PL	45.33	1	Yes	3.12%
FC	1.45	1	No	97.75%

Among the quantitative characters, the F2 progenies were evaluated mostly for seed traits and the data is given in the Table 34. All the traits showed range of values typical of quantitative traits. Among these, the seed length ranged between 8.53 mm to 12.93 mm with an average of 11.24 mm. Seed width varied from 4.93 mm to 8.97 mm with an average of 7.47 mm. Seed thicknesses ranged between 2.42 mm to 6.74 mm with an average of 5.13 mm. The 100 seed weight ranged between 23.22 g to 68.16g with an average of 46.02g. The percentage of coefficient was 7.96% for seed length, 9.37% for seed width, 15.56% for seed thickness, 20.12% for 100 seed weight.

Table 34 Inheritance of seed traits among different F2 population of L-Dopa cross

F2 code	Seed length (SL) in mm	Seed width (SW) in mm	Seed thickness (ST) in mm	Hundred seed weight (HSW) in g	Seed Area (SA) = SL x ST
SMP001	12.02	8.97	6.10	40.61	73.42
SMP002	11.12	7.68	6.37	50.82	70.87
SMP003	11.23	7.15	6.31	47.43	70.79
SMP004	11.18	7.94	5.66	48.81	63.31
SMP005	11.52	7.65	5.72	50.21	65.87
SMP007	11.65	7.56	5.51	49.73	64.16
SMP008	11.59	7.52	6.16	47.82	71.39
SMP009	11.53	6.01	5.39	48.91	62.28
SMP011	11.52	7.15	5.96	49.21	68.69
SMP012	11.64	6.87	5.44	50.63	63.36
SMP013	11.41	7.71	5.68	49.81	64.87
SMP015	11.23	7.96	5.88	50.01	65.86
SMP016	11.78	7.52	5.57	51.22	65.71
SMP017	10.72	7.77	6.06	48.81	64.93
SMP018	11.42	7.88	5.52	48.82	62.94
SMP019	10.86	7.86	6.15	50.21	66.89
SMP021	11.55	7.72	5.39	49.61	62.35
SMP024	12.11	7.52	5.96	50.52	72.16
SMP026	11.41	6.33	5.44	50.23	62.12
SMP027	10.99	7.04	5.68	48.52	62.53
SMP028	10.89	7.25	5.66	49.11	61.66
SMP029	11.01	8.03	6.17	51.51	67.92
SMP030	12.31	6.71	4.69	36.32	57.86
SMP031	11.53	6.51	4.23	36.71	48.48
SMP032	11.86	6.61	4.21	37.52	49.96
SMP033	11.82	6.93	4.33	39.83	50.74
SMP035	11.71	6.85	4.29	38.11	50.34
SMP036	12.07	8.32	4.58	38.02	55.29
SMP037	11.35	6.77	5.51	48.81	62.54
SMP038	11.29	6.54	6.15	50.22	69.53
SMP039	11.11	6.53	5.39	49.63	59.93
SMP041	11.67	6.56	5.96	48.91	69.58
SMP043	11.37	6.43	5.44	50.32	61.92
SMP044	10.64	7.15	5.68	49.21	60.51
SMP045	10.54	7.14	5.66	50.62	59.66
SMP046	10.83	7.52	6.17	50.22	66.91
SMP047	11.73	7.78	4.39	47.03	51.62
SMP048	12.19	5.71	3.35	36.43	40.84
SMP050	10.34	8.32	4.58	38.01	47.37
SMP051	11.89	6.01	3.52	46.62	49.16

SMP053	12.07	7.48	4.39	63.72	53.15
SMP054	8.81	7.87	5.17	37.95	45.58
SMP059	9.23	7.56	4.02	68.22	36.96
SMP060	9.12	7.45	3.73	40.43	34.06
SMP062	8.92	7.45	3.77	64.65	33.65
SMP063	8.86	7.34	3.79	37.95	33.62
SMP068	9.12	7.51	4.08	65.26	37.22
SMP069	8.83	7.77	3.82	38.34	33.82
SMP071	8.81	7.48	4.03	63.56	35.57
SMP073	8.85	7.76	4.95	38.88	43.65
SMP075	8.72	7.85	4.79	63.02	41.84
SMP076	11.43	8.25	4.16	50.61	47.58
SMP077	12.08	7.43	3.46	61.33	41.89
SMP078	12.14	7.65	3.62	38.04	43.77
SMP079	12.09	6.47	3.14	64.95	38.06
SMP080	12.02	7.41	4.45	38.64	53.53
SMP082	11.98	7.86	3.91	60.34	46.93
SMP083	8.93	6.65	4.41	28.67	39.31
SMP084	9.78	6.85	3.32	63.91	32.54
SMP085	12.59	8.23	4.88	36.43	61.53
SMP086	10.99	8.13	5.67	39.33	62.34
SMP087	10.62	8.08	5.71	37.42	60.71
SMP088	9.52	7.98	5.32	37.24	50.71
SMP089	10.27	7.98	5.44	39.30	55.47
SMP094	11.48	8.28	5.09	38.22	58.54
SMP095	10.91	8.21	5.25	38.13	57.35
SMP096	12.34	7.66	4.92	34.22	60.57
SMP099	10.03	6.43	5.07	36.61	50.92
SMP100	10.78	6.98	5.56	37.84	60.02
SMP101	9.24	5.39	2.92	27.86	26.98
SMP103	11.83	7.66	3.91	60.81	46.34
SMP105	8.79	6.65	4.45	28.62	38.69
SMP107	9.58	6.41	3.32	52.61	31.87
SMP108	11.15	7.97	5.46	38.42	60.95
SMP109	10.99	7.98	5.22	38.12	57.42
SMP110	10.58	7.85	5.67	37.32	60.03
SMP111	9.28	7.83	5.75	37.71	53.41
SMP114	9.84	7.91	5.65	38.22	55.68
SMP117	11.15	7.85	5.67	37.84	63.25
SMP118	10.22	7.95	5.69	38.45	58.23
SMP119	10.78	6.98	5.45	36.52	58.83
SMP120	12.28	8.78	5.39	39.23	66.25
SMP122	11.51	8.54	4.55	38.31	51.75

SMP123	11.07	8.44	5.09	39.63	56.4
SMP124	10.79	8.12	5.03	37.62	54.27
SMP126	10.88	7.95	5.02	40.44	54.68
SMP127	10.29	8.44	4.82	39.64	49.71
SMP129	10.04	8.31	5.21	41.22	52.42
SMP130	10.59	8.12	5.18	40.23	54.96
SMP131	10.28	8.15	5.33	40.42	54.86
SMP132	10.51	8.22	5.04	39.21	52.97
SMP133	10.71	8.35	5.52	38.72	59.15
SMP134	10.66	8.32	4.82	42.03	51.44
SMP137	10.89	8.33	4.98	38.41	54.32
SMP138	10.17	5.62	3.83	24.24	39.01
SMP141	10.69	8.25	5.09	39.82	54.49
SMP143	10.72	7.88	5.15	40.03	55.2
SMP144	10.84	8.45	4.67	40.45	50.67
SMP146	11.17	8.09	5.26	39.24	58.75
SMP147	10.91	8.04	4.97	38.41	54.25
SMP148	11.11	8.11	5.09	39.82	56.06
SMP149	10.81	8.01	5.21	40.03	56.35
SMP150	10.59	7.43	4.71	40.46	49.94
SMP151	10.73	7.89	5.46	39.25	58.64
SMP152	11.48	7.93	4.54	58.64	59.84
SMP153	10.82	8.17	5.22	38.81	56.48
SMP154	10.32	8.26	4.74	42.03	48.94
SMP155	10.72	8.13	4.65	39.32	49.97
SMP157	12.28	6.34	3.89	31.61	47.87
SMP160	10.69	7.41	5.28	40.4	55.58
SMP163	11.79	7.12	5.15	47.71	60.71
SMP164	11.06	8.29	4.71	38.72	52.17
SMP165	10.84	7.36	5.28	41.63	57.23
SMP168	9.77	7.23	5.24	42.6	51.24
SMP169	12.93	8.76	6.74	68.16	87.26
SMP171	11.81	7.38	5.11	37.62	60.25
SMP173	10.04	6.74	4.37	23.2	43.94
SMP176	10.43	6.98	6.37	50.81	66.48
SMP178	10.97	7.97	6.34	47.4	69.19
SMP179	10.99	7.92	5.66	48.82	62.23
SMP180	11.45	7.93	5.72	49.91	65.51
SMP184	11.64	7.52	5.51	49.63	64.11
SMP185	12.06	6.11	6.15	49.41	74.25
SMP187	11.9	7.15	5.39	51.75	64.25
SMP188	11.79	7.03	5.96	49.23	70.31
SMP189	11.71	7.49	5.44	50.66	63.79

SMP190	10.99	7.63	5.71	50.24	62.87
SMP191	11.45	7.97	5.56	49.65	63.11
SMP192	11.64	7.52	6.15	49.31	71.65
SMP193	12.06	6.32	5.39	50.22	65.11
SMP194	11.91	7.15	5.96	49.23	70.97
SMP195	11.79	7.33	5.44	50.61	64.21
SMP196	11.71	8.13	5.68	49.85	66.6
SMP197	11.12	7.71	6.37	50.82	70.87
SMP200	12.04	7.09	4.85	49.63	58.52
SMP201	11.91	7.73	5.96	49.82	70.97
SMP202	11.51	7.89	5.08	62.01	63.52
SMP203	11.79	7.17	5.96	50.55	70.31
SMP204	10.26	7.49	4.58	27.31	47.09
SMP205	11.71	7.49	5.44	50.63	63.79
SMP206	10.99	7.72	5.71	49.44	62.87
SMP208	11.45	7.79	5.51	49.64	63.11
SMP209	11.64	7.07	6.15	48.54	71.65
SMP211	12.06	6.73	5.39	50.25	65.11
SMP213	11.15	8.5	4.91	63.65	65.94
SMP214	11.12	8.45	4.58	48.84	51.01
SMP215	11.01	7.95	5.97	50.22	65.76
SMP216	11.73	8.11	5.84	48.33	68.58
SMP219	10.99	8.33	5.76	47.75	63.73
SMP220	12.58	7.35	4.75	41.24	59.85
SMP221	11.07	7.83	4.15	41.66	46.03
SMP222	10.94	8.08	5.79	50.27	63.45
SMP223	11.73	7.87	5.57	49.25	65.39
SMP226	12.47	6.23	3.95	34.43	49.36
SMP228	10.54	7.82	5.59	50.2	59.02
SMP229	10.34	5.88	3.32	34.23	34.42
SMP231	11.12	7.93	4.58	27.05	51.01
SMP232	11.79	7.52	5.96	51.02	70.31
SMP233	11.01	7.42	5.49	50.82	60.54
SMP234	11.73	7.91	5.72	47.44	67.15
SMP235	10.54	8.22	5.66	48.86	59.69
SMP237	12.06	6.33	4.97	49.22	60.04
SMP238	11.45	7.81	5.71	49.93	65.51
SMP240	11.64	7.52	5.55	49.61	64.11
SMP241	12.06	7.18	5.64	49.22	68.02
SMP242	11.92	7.15	5.39	48.53	64.25
SMP243	11.71	7.49	5.38	50.65	63.04
SMP244	10.78	8.32	5.61	55.42	60.58
SMP245	11.45	8.14	5.71	50.23	65.51

SMP246	11.64	7.52	5.55	49.61	64.11
SMP247	11.72	7.53	2.93	36.83	34.37
SMP248	11.25	6.45	3.85	32.42	43.4
SMP249	11.12	8.45	4.58	27.01	51.01
SMP250	11.01	7.85	6.37	50.87	70.19
SMP251	11.73	7.94	6.31	47.45	73.95
SMP252	10.26	8.33	5.66	48.86	58.13
SMP253	11.64	7.52	5.52	49.74	64.11
SMP254	10.59	6.73	6.15	49.33	65.22
SMP255	11.93	7.15	5.39	50.53	64.25
SMP256	11.79	7.66	5.96	49.52	70.31
SMP257	11.71	7.49	5.43	50.64	63.72
SMP258	10.39	8.33	5.71	50.23	59.42
SMP259	11.45	8.24	5.71	50.32	65.51
SMP261	11.45	8.24	5.51	49.42	63.11
SMP262	11.64	7.52	6.15	49.22	71.65
SMP263	11.71	6.73	5.39	50.82	63.25
SMP264	11.96	7.15	5.96	49.72	70.97
SMP265	11.79	7.66	5.37	50.62	63.35
SMP267	11.71	8.34	5.68	49.81	66.62
SMP269	10.26	7.49	4.58	28.22	47.09
SMP273	11.37	8.45	4.73	50.81	53.55
SMP274	11.01	7.15	5.78	47.42	63.67
SMP275	11.73	7.97	5.66	48.81	66.41
SMP276	10.99	7.78	5.71	51.12	62.87
SMP277	11.45	7.85	5.51	49.61	63.11
SMP278	12.06	6.47	5.39	51.23	65.13
SMP279	11.9	7.15	5.96	50.14	70.97
SMP280	11.79	7.46	5.44	50.64	64.21
SMP281	11.63	7.86	5.68	49.85	66.12
SMP282	11.49	8.32	5.88	50.2	67.61
SMP283	11.42	7.74	5.54	27.42	63.31
SMP284	11.53	7.72	5.494	28.72	63.36
SMP285	11.64	7.52	6.15	49.51	71.65
SMP286	11.57	7.69	5.44	28.71	63.02
SMP287	11.72	7.67	5.39	28.51	63.28
SMP288	11.12	7.65	5.35	29.12	59.41
SMP290	10.43	7.62	5.31	31.06	55.32
SMP291	11.48	8.26	5.72	62.16	65.67
SMP292	11.16	8.36	5.72	54.13	63.63
SMP293	11.69	7.27	5.71	64.45	66.72
SMP294	10.99	8.06	5.84	57.66	64.23
SMP297	11.35	6.82	5.53	64.68	62.85

SMP299	12.55	6.54	4.18	44.41	52.59
SMP300	12.26	8.03	6.02	64.28	73.87
SMP301	12.59	6.16	4.17	28.46	52.62
SMP302	11.92	8.06	5.35	64.31	63.8
SMP303	12.19	8.22	5.23	32.65	63.85
SMP304	10.03	6.39	5.12	45.47	51.44
SMP305	11.21	7.28	5.12	32.68	57.39
SMP306	12.76	5.34	4.11	38.69	52.56
SMP307	11.28	5.24	3.21	26.12	36.23
SMP308	12.74	7.09	4.39	38.45	56.04
SMP309	10.97	7.34	6.06	50.26	66.54
SMP311	11.63	7.94	5.51	49.65	64.08
SMP313	10.75	7.66	6.15	49.13	66.21
SMP314	11.64	7.44	4.22	51.12	49.19
SMP315	12.06	7.72	3.11	48.55	37.46
SMP316	12.65	6.17	4.16	38.86	52.63
SMP317	11.31	7.46	6.06	50.24	68.62
SMP318	11.93	7.94	5.52	49.66	65.74
SMP319	10.41	7.44	6.156	49.23	64.13
SMP322	11.71	7.45	4.226	50.27	49.46
SMP323	11.88	6.87	3.41	49.28	40.39
SMP324	9.22	6.11	2.42	32.85	22.34
SMP325	10.97	8.27	6.06	24.23	66.56
SMP326	11.86	7.94	5.51	63.81	65.32
SMP327	10.33	8.33	6.15	29.66	63.62
SMP328	12.74	6.53	4.47	50.82	57.06
SMP329	12.93	5.41	3.53	51.23	45.75
SMP330	8.65	7.23	4.22	25.25	36.55
SMP333	8.53	4.93	3.13	24.22	26.52
SMP334	12.85	7.43	5.09	63.81	65.46
SMP335	10.71	5.45	3.43	29.63	36.82
SMP336	10.86	6.31	3.19	36.45	34.64
SMP343	10.97	8.27	6.06	50.23	66.56
SMP344	10.86	8.33	6.15	49.22	66.89
SMP345	11.64	8.24	5.39	50.21	62.83
SMP346	12.06	7.52	5.96	49.21	71.91
SMP347	11.71	7.94	5.51	49.63	64.54
SMP349	11.73	6.73	5.44	50.63	63.86
SMP351	10.99	7.11	5.68	50.65	62.52
SMP352	12.43	6.94	4.69	54.42	58.33
SMP353	11.37	7.32	4.73	53.64	53.84
SMP355	12.91	7.36	4.71	53.22	60.85
SMP356	12.78	7.23	4.76	54.14	60.92

SMP359	12.74	7.02	4.76	52.82	60.75
SMP360	11.66	7.24	5.35	52.52	62.46
SMP361	11.56	7.36	5.23	52.82	60.58
SMP362	11.71	7.28	5.23	54.34	61.26
SMP364	11.93	7.19	4.73	55.21	56.52
SMP367	11.69	7.18	5.52	55.11	64.64
SMP368	11.63	7.14	5.39	54.62	62.75
SMP370	11.79	7.38	4.73	54.31	55.79
SMP371	12.19	7.36	4.72	53.72	57.62
SMP375	11.69	7.41	4.84	54.43	56.66
SMP376	11.61	7.41	4.82	50.74	55.81
SMP377	11.67	7.36	4.82	53.34	56.27
SMP378	12.14	7.28	4.73	52.42	57.08
SMP379	12.32	7.28	4.72	55.53	58.11
SMP380	12.59	7.29	4.78	54.91	60.22
SMP381	12.53	8.04	4.77	55.32	59.82
SMP384	9.92	7.84	4.75	34.22	47.23
Min.	**8.53**	**4.93**	**2.42**	**23.22**	**22.34**
Max.	**12.93**	**8.97**	**6.74**	**68.16**	**87.26**
Range	**4.40**	**4.04**	**4.32**	**44.96**	**64.82**
Ave.	**11.24**	**7.47**	**5.13**	**46.02**	**58.03**
SD	**0.89**	**0.70**	**0.79**	**0.25**	**0.27**
% CV	**7.96**	**9.37**	**15.56**	**20.12**	**17.71**
P value	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Significant (α = 0.05) ?	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Skewness (Pearson)	**-0.89**	**-0.95**	**-0.87**	**-0.23**	**-0.93**

5.4.3 L-Dopa analysis in F2 population

The data on L-Dopa analysis in different F2 progenies is given in Table 35. The content of L-Dopa ranged from 1.44% -5.84% with mean at 3.11%. The normal distribution curve showed sigmoid distribution (Fig. 23). Further, the correlation analysis of L-Dopa content with other seed traits showed significantly negative correlation except moderate correlation with seed width (0.0797) and seed thickness (0.010) (Table 36).

Table 35 Inheritance of L-Dopa content among different F2 progenies

Sl. No.	F2 Code	% L-Dopa ± SD	Sl. No.	F2 Code	% L-Dopa ± SD	Sl. No.	F2 Code	% L-Dopa ± SD	Sl. No.	F2 Code	% L-Dopa ± SD
1	SMP001	2.62±0.02	41	SMP051	3.63±0.02	81	SMP118	1.51±0.01	121	SMP179	5.34±0.10
2	SMP002	4.67±0.03	42	SMP053	3.65±0.05	82	SMP119	2.84±0.38	122	SMP180	3.28±0.22
3	SMP003	4.52±0.08	43	SMP054	3.82±0.15	83	SMP120	3.68±0.03	123	SMP184	3.57±0.12
4	SMP004	2.72±0.01	44	SMP059	2.59±0.12	84	SMP122	3.46±0.05	124	SMP185	5.42±0.02
5	SMP005	2.98±0.01	45	SMP060	3.51±0.23	85	SMP123	4.48±0.82	125	SMP187	2.94±0.12
6	SMP007	4.80±0.14	46	SMP062	1.68±0.22	86	SMP124	2.85±0.14	126	SMP188	3.26±0.07
7	SMP008	1.70±0.01	47	SMP063	2.50±0.06	87	SMP126	3.17±0.71	127	SMP189	2.54±0.20
8	SMP009	3.57±0.13	48	SMP068	5.61±0.09	88	SMP127	3.78±0.41	128	SMP190	2.47±0.12
9	SMP011	4.27±0.06	49	SMP069	1.84±0.09	89	SMP129	2.81±0.33	129	SMP191	5.25±0.02
10	SMP012	3.68±0.07	50	SMP071	4.98±0.11	90	SMP130	4.40±0.11	130	SMP192	3.28±0.11
11	SMP013	4.37±0.19	51	SMP073	2.81±0.14	91	SMP131	3.81±0.23	131	SMP193	1.58±0.04
12	SMP015	4.16±0.02	52	SMP075	4.68±0.01	92	SMP132	4.88±0.13	132	SMP194	3.95±0.14
13	SMP016	2.72±0.16	53	SMP076	3.36±0.82	93	SMP133	3.91±0.02	133	SMP195	4.11±0.21
14	SMP017	3.96±0.14	54	SMP077	2.79±0.02	94	SMP134	2.17±0.05	134	SMP196	1.94±0.01
15	SMP018	1.85±0.21	55	SMP078	5.27±0.70	95	SMP137	3.74±0.07	135	SMP197	4.97±0.03
16	SMP019	5.24±0.24	56	SMP079	3.88±0.22	96	SMP138	5.20±0.09	136	SMP200	2.74±0.11
17	SMP021	4.03±0.02	57	SMP080	3.64±0.03	97	SMP141	1.49±0.11	137	SMP201	5.12±0.02
18	SMP024	1.89±0.05	58	SMP082	3.43±0.09	98	SMP143	2.45±0.05	138	SMP202	4.98±0.08
19	SMP026	4.77±0.01	59	SMP083	1.64±0.03	99	SMP144	5.39±0.27	139	SMP203	4.25±0.06
20	SMP027	2.36±0.03	60	SMP084	5.02±0.82	100	SMP146	3.66±0.02	140	SMP204	3.67±0.01
21	SMP028	5.17±0.04	61	SMP085	1.98±0.57	101	SMP147	2.17±0.04	141	SMP205	1.69±0.10
22	SMP029	1.50±0.02	62	SMP086	2.44±0.12	102	SMP148	3.87±0.17	142	SMP206	4.61±0.27
23	SMP030	1.62±0.06	63	SMP087	3.20±0.07	103	SMP149	2.83±0.12	143	SMP208	4.07±0.32
24	SMP031	1.66±0.09	64	SMP088	1.54±0.23	104	SMP150	3.63±0.36	144	SMP209	1.71±0.21
25	SMP032	4.42±0.04	65	SMP089	2.13±0.12	105	SMP151	2.16±0.08	145	SMP211	2.85±0.22
26	SMP033	4.20±0.01	66	SMP094	5.23±0.02	106	SMP152	3.39±0.04	146	SMP213	1.88±0.15
27	SMP0.35	1.81±0.11	67	SMP095	3.08±0.10	107	SMP153	2.59±0.70	147	SMP214	2.89±0.30
28	SMP036	1.53±0.07	68	SMP096	4.53±0.34	108	SMP154	5.84±0.07	148	SMP215	3.96±0.05
29	SMP037	2.56±0.08	69	SMP099	3.81±0.04	109	SMP155	2.29±0.21	149	SMP216	4.58±0.31
30	SMP038	3.42±0.03	70	SMP100	2.34±0.13	110	SMP157	2.24±0.09	150	SMP219	1.83±0.04
31	SMP039	3.42±0.03	71	SMP101	1.72±0.02	111	SMP160	5.08±0.01	151	SMP220	2.12±0.12
32	SMP040	1.71±0.02	72	SMP103	4.13±0.12	112	SMP163	1.78±0.21	152	SMP221	3.51±0.30
33	SMP041	4.58±0.05	73	SMP105	1.44±0.21	113	SMP164	1.52±0.11	153	SMP222	2.29±0.16
34	SMP043	1.87±0.08	74	SMP107	4.92±0.70	114	SMP165	3.45±0.04	154	SMP223	3.36±0.02
35	SMP044	3.12±0.20	75	SMP108	3.85±0.13	115	SMP168	2.30±0.02	155	SMP226	1.83±0.14
36	SMP045	3.92±0.14	76	SMP109	2.31±0.54	116	SMP169	4.04±0.14	156	SMP228	2.31±0.06
37	SMP046	2.74±0.34	77	SMP110	5.52±0.51	117	SMP171	4.88±0.11	157	SMP229	3.22±0.19
38	SMP047	4.26±0.04	78	SMP111	1.63±0.10	118	SMP173	2.87±0.09	158	SMP231	4.59±0.12
39	SMP048	1.74±0.70	79	SMP114	4.51±0.27	119	SMP176	1.55±0.14	159	SMP232	2.43±0.07
40	SMP050	3.56±0.01	80	SMP117	4.08±0.70	120	SMP178	2.75±0.02	160	SMP233	3.49±0.01

Table 35 Conti.....

Sl. No.	F2 Code	% L-Dopa ± SD	Sl. No.	F2 Code	% L-Dopa ± SD	Sl. No.	F2 Code	% L-Dopa ± SD
161	SMP234	3.79±0.41	201	SMP282	3.29±0.14	241	SMP330	4.31±0.16
162	SMP235	2.86±0.33	202	SMP283	2.71±0.04	242	SMP333	4.09±0.02
163	SMP237	1.64±0.11	203	SMP284	3.53±0.13	243	SMP334	3.78±0.13
164	SMP238	4.46±0.21	204	SMP285	1.84±0.18	244	SMP335	3.03±0.06
165	SMP240	3.84±0.24	205	SMP286	3.69±0.01	245	SMP336	3.33±0.35
166	SMP241	3.81±0.02	206	SMP287	1.71±0.15	246	SMP343	2.01±0.18
167	SMP242	2.67±0.01	207	SMP288	2.51±0.14	247	SMP344	2.37±0.08
168	SMP243	1.63±0.30	208	SMP290	4.38±0.01	248	SMP345	2.76±0.04
169	SMP244	1.81±0.09	209	SMP291	4.22±0.71	249	SMP346	2.17±0.21
170	SMP245	2.37±0.18	210	SMP292	2.66±0.02	250	SMP347	1.81±0.29
171	SMP246	3.39±0.11	211	SMP293	2.81±0.10	251	SMP349	1.72±0.04
172	SMP247	2.23±0.10	212	SMP294	1.73±0.04	252	SMP351	2.26±0.14
173	SMP248	3.48±0.01	213	SMP297	3.28±0.02	253	SMP352	3.86±0.02
174	SMP249	2.71±0.09	214	SMP299	2.73±0.13	254	SMP353	2.08±0.03
175	SMP250	2.47±0.12	215	SMP300	2.35±0.16	255	SMP355	4.15±0.01
176	SMP251	3.77±0.32	216	SMP301	2.60±1.10	256	SMP356	2.55±0.15
177	SMP252	1.62±0.11	217	SMP302	3.09±0.14	257	SMP359	3.48±0.31
178	SMP253	2.65±0.05	218	SMP303	1.61±0.06	258	SMP360	4.17±0.21
179	SMP254	5.49±0.27	219	SMP304	2.65±0.17	259	SMP361	3.61±0.24
180	SMP255	4.54±0.02	220	SMP305	2.25±0.04	260	SMP362	1.87±0.04
181	SMP256	1.79±0.21	221	SMP306	1.63±0.19	261	SMP364	2.16±0.02
182	SMP257	2.49±0.12	222	SMP307	2.57±0.03	262	SMP367	3.48±0.06
183	SMP258	3.59±0.11	223	SMP308	1.76±0.14	263	SMP368	2.48±0.06
184	SMP259	2.56±0.03	224	SMP309	1.82±0.23	264	SMP370	5.46±0.07
185	SMP261	3.62±0.14	225	SMP311	2.08±0.01	265	SMP371	2.34±0.11
186	SMP262	1.69±0.06	226	SMP313	2.17±0.24	266	SMP375	1.83±0.14
187	SMP263	2.77±0.04	227	SMP314	3.04±0.06	267	SMP376	3.57±0.12
188	SMP264	3.43±0.12	228	SMP315	2.06±0.01	268	SMP377	4.12±0.22
189	SMP265	3.61±0.01	229	SMP316	4.38±0.01	269	SMP378	3.52±0.01
190	SMP267	1.85±0.05	230	SMP317	2.34±0.25	270	SMP379	1.84±0.14
191	SMP269	4.33±0.03	231	SMP318	2.39±0.17	271	SMP380	3.13±0.26
192	SMP273	2.43±0.14	232	SMP319	1.84±0.11	272	SMP384	3.47±0.28
193	SMP274	3.32±0.75	233	SMP322	2.31±0.21		**Min.**	**1.44**
194	SMP275	2.11±0.04	234	SMP323	1.85±0.20		**Max.**	**5.84**
195	SMP276	1.93±0.23	235	SMP324	3.98±0.18		**Range**	**4.40**
196	SMP277	2.55±0.02	236	SMP325	1.77±0.09		**Ave.**	**3.11**
197	SMP278	1.81±0.16	237	SMP326	2.74±0.14		**SD**	**1.11**
198	SMP279	3.63±0.12	238	SMP327	1.82±0.24		**% CV**	**35.41**
199	SMP280	1.62±0.03	239	SMP328	2.25±0.01		**P value**	**< 0.0001**
200	SMP281	4.18±0.18	240	SMP329	4.01±1.30		**Significant (α = .05)?**	**Yes**
							Skewness (Pearson)	**0.37**

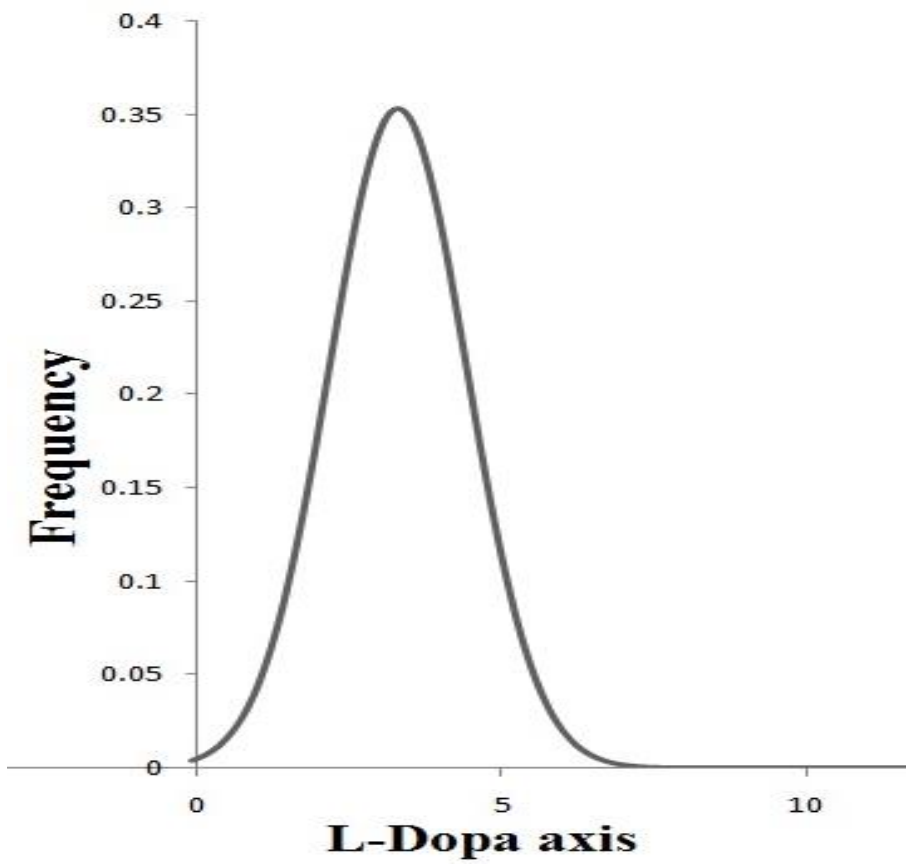


Fig. 23: Normal distribution curve of L-Dopa content among F2 progenies

Table 36 Correlation for L-Dopa content with seed traits in F2 progenies

Traits	SL	SW	ST	HSW	SA	LD
SL	1					
SW	-0.065	1				
ST	0.129	0.437	1			
HSW	0.282	0.127	0.238	1		
SA	0.514	0.348	0.909	0.344	1	
LD	-0.119	0.0797	0.010	-0.056	-0.032	1

Values in bold are different from 0 with a significance level alpha=0.01

5.4.4 Variance and heritability analysis

The data on genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability in narrow sense (h^2), genetic advance (GA) and genetic advance as percentage of means (GAM) estimated for L-Dopa content in F2 generation is presented in Table 37. Phenotypic coefficient of variation was higher (34.26) than genotypic coefficient of variation (24.44) for L-Dopa content. Heritability in broad sense value was high at 71.35%. Moderate value was recorded for genetic advance which was 8.60. Genetic advance as percentage of mean was 196.23.

Table 37 Estimates of PCV, GCV, heritability, GA and GA % mean in F2 population

Analysis	L- Dopa trait values in F2
Range	1.55 - 9.03
Estimates of range, mean and standard error (X±SE)	4.38 ± 1.61
% Coefficient variation : 500240SK (% CV : P1)	3.65
% Coefficient variation : 500212AS (% CV : P2)	2.59
% Coefficient variation : P1 x P2 (% CV : F1)	6.3
Environment variance (σ^2 E)	9.815
Phenotypic variance (σ^2 P)	34.26
Genotypic variance (σ^2 G)	24.445
Genotypic coefficient of variation (GCV (%)	112.77
Phenotypic coefficient of variation (PCV (%)	133.51
Heritability (h^2)	71.35
Genetic advance (GA)	8.6
Genetic advance as percentage of mean (GAM)	196.23

5.5 Micropropagation of elite genotypes

5.5.1 Seed germination

In vitro inoculated seeds germinated within 5-7 days on Murashige & Skoog (MS) basal medium without sucrose. Presence of seed coat inhibited rate and percentage of germination by nearly 40% (Table 38). Its removal favoured germination, which rose to 90%. All the shoots attained a height of 6-7 cm within eight days exhibiting synchronized seed germination pattern (Fig. 24). Axillary buds became prominent after 12-14 days (Fig. 24A).

Table 38 Effect of seed coat on *in vitro* germination of seeds

Sl. No.	% Seed germination (MS basal medium)	
	With seed coat	Without seed coat
1	60	90.90
2	55.56	81.80
3	63.63	86.70
4	63.63	90.00
5	54.50	91.70

5.5.2 Multiple shoots induction

After inoculation, three explants *viz*., axillary bud, cotyledonary node (CN) and apical bud showed differential response for shoot emergence. The CN and axillary buds cultured on MS basal medium did not yield any shoots; but apical buds gave rise to single shoot. The number of shoots increased when the plants were cultured on growth regulator supplemented medium singly or in combination.

The mean number of shoots produced by different explants on MS medium supplemented with different hormonal combinations is shown in Table 39. Of the various hormones tested, BAP was more effective over Kn and other combination of

hormones in all the three explants. The axillary buds produced a maximum of 6.4±0.55 shoots on MS+BAP (22.2 µM) (Fig. 24B).

Table 39 Effect of growth regulators on multiple shoot induction from three different explants

Sl. no.	MS+Hormone combination (µM)	No. of shoots induced[y]		
		Axillary buds	Cotyledonary node	Apical bud
1.	CONTROL	0^{h}	0^{k}	1.0 ± 0.0^{f}
2.	BAP 2.22	2.6 ± 0.54^{efg}	2.6 ± 0.55^{fghij}	3.8 ± 0.84^{b}
3.	BAP 4.44	3.6 ± 0.54^{bcdef}	3.2 ± 0.45^{defghi}	3.8 ± 0.45^{b}
4.	BAP 8.88	3.6 ± 0.55^{bcdef}	3.8 ± 0.45^{bcdef}	5.8 ± 0.84^{a}
5.	BAP 13.32	4.6 ± 0.55^{bc}	4.2 ± 0.45^{bcde}	5.6 ± 0.89^{a}
6.	BAP 17.76	4.6 ± 0.55^{bcd}	4.6 ± 0.55^{bc}	2.8 ± 0.84^{bcd}
7.	BAP 22.20	6.4 ± 0.55^{a}	6.4 ± 0.55^{a}	2.6 ± 0.24^{bcd}
8.	KN 4.65	2.4 ± 0.89^{efg}	2.2 ± 0.45^{b}	2.6 ± 0.24^{bcd}
9.	KN 9.30	3.0 ± 0.89^{defg}	2.5 ± 0.55^{ghij}	2.17 ± 0.4^{cdef}
10.	KN 13.95	3.2 ± 0.45^{cdefg}	3.6 ± 0.55^{bcdefg}	1.6 ± 0.55^{def}
11.	KN 18.60	4.0 ± 0.7^{bcde}	4.4 ± 0.55^{bcd}	3.2 ± 0.84^{bc}
12.	KN 23.25	5.0 ± 0.71^{ab}	4.8 ± 0.84^{b}	2.0 ± 0^{cdef}
13.	2ip 2.47	2.25 ± 0.5^{fg}	1.75 ± 0.5^{j}	1.0± 0.0ef
14.	2ip 14.93	1.8 ± 0.84^{g}	2.0 ± 0^{ij}	1.6 ± 0.54^{def}
15.	BAP 4.44+ KN 2.33	3.6 ± 0.89^{bcdef}	3.2 ± 0.45^{defghi}	2.0 ± 0^{cdef}
16.	BAP 4.44+ KN 4.65	4.4 ± 0.55^{bcd}	3.4 ± 0.89^{cdefgh}	2.0 ± 0^{bcd}
17.	BAP 4.44+ NAA 1.32	3.6 ± 1.15^{bcdef}	3.0 ± 1.0^{efghij}	2.4 ± 0.89^{cde}
18.	BAP 4.44+ NAA 2.63	4.4 ± 0.55^{bcd}	2.6 ± 0.89^{fghij}	1.6 ± 0.55^{def}

[y]Values are mean ± standard deviation of three independent experiments with each treatment containing 20 replicates. Means followed by same letters are not significantly different α=0.05 significance level, as determined by Tukey's HSD test.

Similar results were also observed in case of CN except for quality of shoots, which was better in axillary buds compared to CN. Kinetin (23.25 µM) on the other hand produced maximum of 5.0±0.71 and 4.8±0.84 shoots in axillary buds and CN respectively. Steady increase in number of shoots was noticed up to 22.2 µM in case of BAP and 23.20 µM in case of Kn, both in axillary bud and CN explants. Cytokinin supplementation beyond this concentration adversely affected the shoot development, as the regenerated shoots became stunted and dense. The combination of BAP (4.44 µM) + NAA (1.32–2.63 µM) and BAP (4.44 µM) + Kn (2.33-4.65 µM) resulted in 3-5 plantlets in both the explants. Only 2-3 shoots were formed in 2ip (2.5-5 µM).

On the other hand, the apical bud showed strikingly different response compared to other two explants. In this case, BAP at much lower concentration of 13.32 µM induced maximum of 5.6±0.8 shoots, whereas 2ip, Kn and other combinations resulted in poor response. Since the best response, both in terms of quality and quantity of shoots were obtained in axillary bud explants, we used this in all our further experiments.

5.5.3 Elongation and rooting

Elongation of shoots was found to be best on GA_3 containing medium. Transferring the micro shoots to MS + GA_3 (2.89-14.43 µM) showed varied degrees of elongation with GA3 (2.89 µM) producing best response (4.0 cm). The elongated shoots were rooted on half-strength liquid MS medium supplemented with NAA (0.00-16.2 µM). Best rooting was obtained on MS+NAA (5.40 µM) (Fig 24D). The details of the results obtained for elongation and rooting are shown in Table 40.

Table 40 Effect of different concentrations of NAA on root induction

Sl. No.	MS+NAA (µM)	Response (%)	No. of roots/ explant[y]	Root length (cm)[y]
1.	Control[x]	0.00	0.00± 0.00[c]	0.00± 0.00
2.	NAA 2.70	60.00	3.67 ± 0.58[bc]	4.61± 0.16
3.	NAA 5.26	100.00	5.00± 1.00[b]	2.62± 0.74
4.	NAA 10.8	100.00	6.67± 1.15[b]	2.27± 0.84
5.	NAA 16.2	100.00	16.67± 2.89[a]	2.78± 0.51

[x]Control = Hormone free MS medium.

[y]Values are mean ± standard deviation of three independent experiments with each treatment containing 20 replicates. Means followed by same letters are not significantly different α=0.05 significance level, as determined by Tukey's HSD test.

5.5.5. Hardening and field transfer

Plantlets directly transferred to sand: soil mixture (1:1) exhibited average survival of 60-70% due to frequent fungal contamination. To subvert this, plantlets were initially transferred to autoclaved mixture of sand: soil: peat moss (1:1:1) and maintained in growth chamber for 14 days. The acclimatized plants were then transferred to sand: soil (1:1) mixture and allowed to harden for 30 days (Fig. 24E & F) before finally transferring to field. This resulted in 95% survival. All the survived plants were successfully transplanted to the field condition.

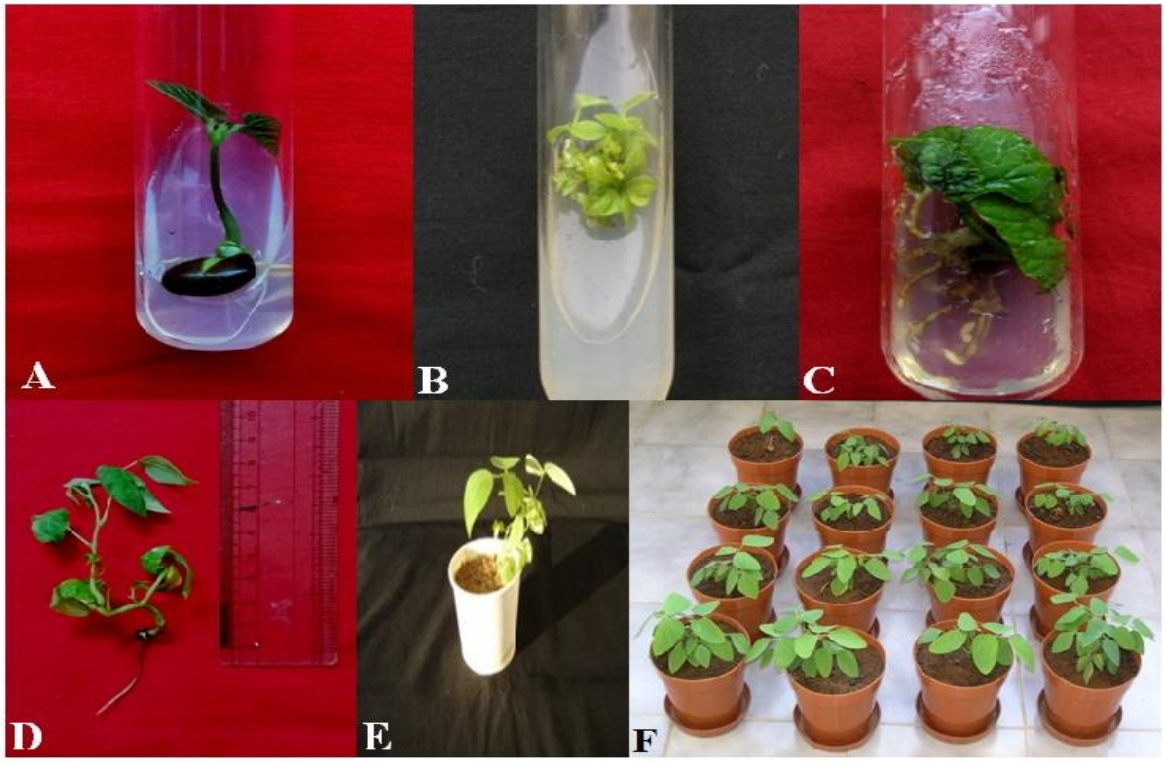


Fig. 24: Multiple shoot induction from axillary bud explants of *M. pruriens*. **A:** 12 day old *in vitro* germinated *M. pruriens* seedling on MS basal without sucrose. **B:** Multiple shoots induction from axillary bud explant on MS+BAP (22.2 µM). **C:** Proliferation of shoots on MS+ BAP (2.67) **D:** Rooting of the shootlets on MS+ NAA (5.26 µM) supported on filter paper wad. **E& F:** Hardened plantlets.

6. Discussions

6.1 Germplasm collection and characterization

The centre of origin of a crop plant is generally associated with broad genetic diversity contributed by the wild and weedy relatives (Harlan 1971; Guo et al. 2014). *M. pruriens* is native to eastern India (Burkill 1966; Duke 1981; Wilmot-Dear 1984) which includes parts of northeast India. This region showcase abundant distribution of wild variety represented by var. *pruriens* and there is a long history of cultivation of velvet bean (var. *utilis*) by several indigenous groups (Arora 1991). Recently, after the discovery of L-Dopa from the plant, the prospects for large-scale cultivation of *M. pruriens* are growing in the region. As a result, efforts to develop improved varieties with better agronomic, nutritional and medicinal properties are gaining momentum (Sathyanarayana et al. 2017).

The success of any plant breeding program depends on determining the variability in the germplasm collection. Gene pool evaluation can pave a way for identifying valuable genetic resources as well as defining relevant breeding strategies. Morphometric evaluation provides early insight on it by capturing the variability of the agronomic traits (Ghafoor et al. 2003). This is reinforced by the deployment of multivariate analysis - which provides improved resolution on the relationship among the accessions, germplasm structure and elite lines (Aremu et al. 2012). Such efforts have contributed to progress in breeding works in several legumes species such as alfalfa (Jenczewski 1999); lima bean (Asante 2008); bambara groundnut (Ouedraogo 2008); lentils (Toklu 2009) and cowpea (Adewale 2011).

On the other hand, germplasm characterization of *M. pruriens* in India in general, and northeast India in particular, are lacking. Information available so far are mostly derived from the evaluation of small number of accessions from a limited geographical regions (Gurumoorthi et al. 2003; Pugalenthi and Vadivel, 2007a; Mamatha et al. 2010). Also majority of them have focused on a limited number of agronomic or nutritional traits (Krishnamurthy et al. 2005). In this background, the present study focused on collection and characterization of variability represented in *M. pruriens* germplasm of northeast India using multivariate analysis as a primary objective for this thesis work.

6.1.1 Variability in total germplasm collection

Totally 50 accessions were collected from different northeast Indian states and evaluated. Yield based traits contributed in the range of 8.39% to 69.98% while non-yield based traits varied from 25.56 to 51.17%. The Jaccard's similarity coefficient revealed moderate to high genetic diversity for the qualitative (61%) and quantitative (56%) traits. The results suggest that the germplasm collection from northeast India embrace good diversity with regard to both qualitative and quantitative traits. This observation is in line with previous reports in this crop from other regions of India (Sathyanarayana et al. 2011). However, earlier works reported only mean score along with variation for few selected quantitative traits (Krishnamurthy et al. 2005; Pugalenthi and Vadivel 2007b; Mamatha et al. 2010). In comparison, the present study, which includes several qualitative and quantitative traits, provides most comprehensive information for the first time with regard to diversity of *M. pruriens* in the northeast India.

Many seed based traits showed considerable variability among the germplasm accessions. In particular, HSW which has direct bearing on the seed yield ranged from 18.88 g (500216ML - var. *pruriens*) to 173.36 g (500266NL - var. *utilis*) indicating substantial diversity for this trait. Similarly, other traits such as days to flowering showed variation from 93 days (500221AR – var. *pruriens*) to 132 (500269SK – var. *pruriens*) days and days to maturity from 137 days (500210MN - var. *utilis*) to.197 days (500271SK - var. *pruriens*). Good variability was also observed for days for emergence which ranged from 4 (500225AR) to 13 days (500205AS). Even though number of pods per plant also showed good variability, total number of pods recorded per plant (14 - 68 in var. *pruriens* and 14-22 in var. *utlis*) were low as compared to other studies. This might be due to fact that the evaluation experiments for the present study were conducted in potted plants.

Another important character that showed promising variability was days to maturity which ranged in between 137 days to 197 days. Earliness determined by the plant's ability to reach flowering and maturity cycle in a short period is closely associated with crop's productivity (Shavrukov et al. 2017). Reducing the crop cycle, in many cases, has also helped to avoid unfavorable temperatures and low air humidity during the flowering and pod formation stages (Martinez-Calvo 2008). Thus, early maturing accessions identified in the present study serves as an important genetic stock for breeding earliness in this crop.

Overall, large differences were observed for different morphological and phenetic characters in *M. pruriens* accession of northeast India. This might be due to extreme climatic and edaphic variations experienced in the region which comprises some of the large rivers and mighty mountains including Himalayas. Similar variation in morphometric traits have also been recorded in other crops including rice varieties

(Sarhadi et al. 2009; Joshi et al. 2011; Dikshit et al. 2014; Debbarma et al. 2017) and *Jatropha curcas* (Saikia et al. 2009; Adreeja et al. 2017) grown in northeast India.

6.1.2 Diversity among var. *pruriens* accessions

In case of wild variety (var. *pruriens*), the average coefficient of variation (CV %) for the quantitative traits ranged from 8.17 % to 46.79 % with an average of 21.59%. Genetic diversity indices based on Jaccard's coefficient were high both in case of qualitative (47%) and quantitative (40%) traits. This is in consensus with earlier reports by Krishnamurthy et al. (2005) and a Ph.D work by Mahesh (2015) submitted to Visvesvaraya Technological University, Belagum. Several genotypes also showed promising trait values and can be a source of important genes for economic traits. It is well known that the wild germplasm are important reservoirs of resistance genes against biotic and abiotic stresses (Sudha et al. 2013). Thus, these genotypes might serve as an important source of breeding lines for introducing wild germplasm into ongoing breeding programs of *M. pruriens*.

6.1.3 Diversity among var. *utilis* accessions

On the other hand, the average coefficient of variation for quantitative traits among the cultivated var. *utilis* accessions ranged from 6.59% to 39.46 % with a mean average of 16.369%. The genetic diversity values based on Jaccard's similarity value for qualitative traits was 41% and quantitative traits was 32% indicating low to moderate genetic diversity for both the groups of traits. This observation is in accordance with morphological diversity in velvet bean germplasm reported in earlier study (Sathyanarayana et al. 2012).

Among the evaluated traits, seed yield and days to maturity (DM) showed considerable variability ranging from 100.74 g to 173.36 g (seed yield) and 137 days to 189 days (DM). Generally var. *utilis* accessions are late maturing types (Sathyanarayana et al. 2011). However, few accessions in the present study showed early maturity (500210MN-137 days; 500211NL-149 days and 500263NL-151 days) than those reported earlier (Sathyanarayana et al. 2012). This might be due selection of early maturing genotypes by indigenous groups owing to long history of *M. pruriens* cultivation in northeast India (Watt 1883; Piper and Tracy 1910; CSIR 1962; Arora 1991). Thus, these accessions might offer significant promise for future breeding program on developing early maturing genotypes.

6.1.4 Identification of elite genotypes and contrasting parental lines

As can be seen from the foregoing discussion, evaluation of *M. pruriens* collection from northeast India based on selected morpho-agronomic traits revealed considerable variability in our germplasm with potential applications in breeding programs. It has also revealed good scope for introducing wild germplasm into ongoing breeding programs, as improvement efforts so far have focused only around cultivated genotypes. Though crop breeding began with the domestication of wild plants, it has grown and continued with the improvement of both the landraces and the elite varieties (Lyu et al. 2013). In this background, the set of accessions comprising both the cultivated and wild varieties identified as elite genotypes in this study serves as an important source of breeding material. Further, the accessions possessing extreme phenotypes can fulfill the necessary pre-requisite for development of mapping population for linkage mapping (Collard et al. 2005) and other genetic analysis as attempted in subsequent chapters of this thesis work.

Considering the fact that the isolated use of morphological markers might lead to incomprehensive conclusions due to inherent limitations of these markers (Smith and Smith, 1992; Tatikonda et al. 2009), it is important to supplement morphometric analysis with that from DNA based markers to gain more realistic insights into diversity pattern in our germplasm which is attempted as another objective of this thesis work.

6.2 L-Dopa variability and correlation with seed traits

For breeding L-Dopa content in *M. pruriens*, generation of reliable data on its variability in the germplasm collection is an important endeavor. In the present study, pure L-Dopa was isolated from seed samples using extraction method described by Upadhyay et al. (2012). The efficacy of this method has been tested in earlier studies as well (Patel et al. 2016). We determined purity of the isolated L-Dopa using HPLC method after due validation of the equipment. The results showed single clear peak with RT (6.319 min.) matching with that of standard L-Dopa (RT 6.345 min.) attesting to the purity of the isolated sample. The same protocol was further applied to assess L-Dopa variability among 50 *M.pruriens* accessions.

The results of L-Dopa analysis revealed broad variability ranging from very low (1.55%) to very high (9.03%). The observation on high variability for L-Dopa content in our study is in conformity with other earlier works. For instance, Bell and Jenzen (1971) reported 5.9-9.0% L-Dopa among six accessions. Daxenbichler et al. (1971, 1972) reported 2.2 – 7.2% in their screening of 36 accessions. Lorenzetti et al. (2000) found L-Dopa content in the range of 1.9-7.6% among 38 accessions. Few Indian studies have also showed similar trend with L-Dopa concentration reported in the range of 3.63-5.05% (Siddhuraju et al. 2000); 2.30-6.12% (Krishnamurthy et al.

2005); 1.51-6.29% (Singh et al. 2008); 4 - 6.5% (Vijayambika et al. 2010); 4.11 - 6.61% (Mamatha et al. 2010); 5.4-7.0% (Kalidass and Mohan 2011) and 3.29-5.44 % (Raina et al. 2012). The highest variability among Indian accessions found in the present study might be due availability of large altitudinal gradient across northeastern region of India (Chatterjee 2008). Lorenzetti et al. (2000) suggests that L-Dopa content may vary with altitude due to changes in the intensity of light and scattered ultraviolet radiation. With regard to correlation analysis, L-Dopa content was found to be negatively correlated with all other seed characters indicating these traits are not useful as selection criteria for identifying high/low L-Dopa containing genotypes. This observation is in conformity with the reports by Raina et al. (2012, 2018) on the same species.

Among the different accessions studied, we determined accessions: 500212AS (1.55%) and 500240SK (9.03%) as genotype showing extreme phenotype for L-Dopa content. These accessions were further used as contrasting parents for generating segregating population to study variability and heritability of the L-Dopa trait which is discussed in subsequent chapter.

6.3 Genetic diversity analysis

The center of origin of a crop plant represents vital source of genes to farmers and professional breeders responding to biotic and abiotic stresses and climate change (Dyer et al. 2014). Medicinal legume *M. pruriens* is native of southern China, Malaysia or eastern India (Burkill 1966; Duke 1981; Wilmot-Dear 1984) with considerable antiquity in these regions (Arora et al. 1991). It is reported that *Mucuna* was widely grown in the foothills and lower-hills of the eastern Himalayas during the eighteenth and nineteenth centuries and was eventually replaced due to introduction

of more palatable legumes (CSIR 1962; Piper and Tracy 1910; Watt 1883). However, its cultivation as a minor crop continues even today by many native people who grow it as home garden crop for dry seeds and green pods (Arora et al. 1991). Recently, with the discovery of L-Dopa and other agronomic benefits from the plant, the opportunities for its large-scale cultivation in the region are growing necessitating development of improved cultivars with enhanced nutritional value and resistance to biotic and abiotic stresses (Sathyanarayana et al. 2017). Even though, northeast India represents most important source of gene pool for *M. pruriens,* little data is available on the genetic diversity and population structure of this species from the region. Thus, the present investigation was carried out to meet this objective using AFLP markers. AFLP was particularly chosen as it offers pan genome coverage comprising large number of loci from different chromosomal regions, and regarded as choice marker for diversity analysis (Karp and Edwards 1995; Mueller and Wolfenbarger 1999; Mikulášková et al. 2012; Patrick et al. 2012). Earlier, AFLPs have been successfully used for analyzing genetic variation, phylogenetic relationships and characterization of natural populations as well as breeding lines across the range of taxa (Hill et al. 1996; Kardolus et al. 1998, Roldan-Ruiz et al. 2000, De Riek et al. 2001; Dubey et al. 2010). Their utility for genetic diversity analysis in *Mucuna* species is also well recognized (Capo-chichi et al. 2004; Sathyanarayana et al. 2011).

6.3.1 Marker attributes

Majority of the PCs produced high quality AFLP profiles generating an average of 54.91 total and 52.55 (97.27%) polymorphic fragments per PC. A high rate of polymorphism indicated good genetic variability among the investigated accessions. In diversity studies, PIC values are used as primary measure of discriminatory power or informativeness of PC (Chandrawati et al. 2014). The average PIC value of 0.23

obtained in the present study although was on lower side, is superior to PIC value (0.16) reported in earlier AFLP studies on *Mucuna* spp. (Sathyanarayana et al. 2011). Higher PIC values obtained in the present investigation might be due to new AFLP markers employed in the present study. Out of 22 PCs, 9 (41%) produced 100% polymorphic fragments. Thus, these informative PCs could be used for large scale germplasm characterization and genotyping. The other two parameters, RP and MI provide complementary attributes to the efficiency of the marker system and are used along with PIC in diversity studies (Gupta et al. 2013; Pecina-Quintero et al. 2013). Considering combination of PIC, MI and RP parameters, the PCs: E-ACG/M-CAC, E-ACA/M-CAA, E-AAG/M-CAC and E-ACA/M-CAT and RP PCs E-ACT/M-CTA, E-AGG/M-CTA and E-AAC/M-CAG are found to be more useful for future diversity studies. Among these, the PCs: E-ACT/M-CTA, E-AAC/M-CAG, E-AAC/M-CTA and E-AAC/M-CTC are already reported to be informative in earlier AFLP study on *Mucuna* species (Sathyanarayana et al. 2011), and thus our result is in conformity with the above finding.

6.3.2 Gene diversity and population structure

A pair wise similarity matrix based on Jaccard's coefficient was constructed to understand the extent of genetic diversity among all the *M. pruriens* accessions. A high level of diversity was noticed (average similarity coefficient: 0.37; average Shannon's information index: 0.34) within our collection which is reflected even in the values of h and Ht with mean scores of 0.27 and 0.22, respectively. This is significant as the accessions used here correspond to neighboring areas of eastern India from where *M. pruriens* is reported to have come from (Capo-chichi et al. 2003). So far, only anecdotal evidences were available to support this theory including large-scale presence of perennial *M. bracteata* - a probable ancestor of this

species (Jaheer et al. 2015) in the region. Thus, high allelic diversity among *M. pruriens* of northeast India, reported for the first time in the present study, provides earliest empirical evidence for the eastern Indian origin of *M. pruriens* and paves a way for future investigation on migration and colonization pattern of this species as well as events that shaped its domestication. Among the two botanical varieties, var. *pruriens* which represents wild variety exhibits higher diversity (I: 0.44) as compared to var. *utilis* (I: 0.37) which is cultivated. The phenomenon of wild *M. pruriens* varieties exhibiting greater diversity vis-à-vis cultivated var. *utilis* is a common observation reported in many earlier studies (Leelambika and Sathyanarayana 2011, Sathyanarayana et al. 2012, 2016).

On the other hand, the population genetic structure investigated using UPGMA, AMOVA and STRUCUTRE revealed absence of geographical structure in populations of northeast India. In case of AMOVA - which determines contribution of individual populations to the total diversity, the results revealed 93% of the variance was accrued by intra-population variation while only 7% contributed to inter-population variation pointing towards absence of barriers for the gene flow. To corroborate this, we estimated Fst and Nm, to measure the extent of gene flow. The moderate estimate of Fst (0.08.) and high Nm value (3.04) substantiated AMOVA results of low divergence as a factor of high gene flow. It is well known that gene flow values (Nm) >1 is strong enough to prevent substantial differentiation due to genetic drift (Slatkin and Barton 1989). It is likely that *M. pruriens* has experienced little geographical isolation in this region due to predominant distribution in the foothills extending up to plains in this otherwise geographically/climatically diverse region comprising some of the large mountains and mighty rivers that offers rich scope for speciation by isolation. Such a situation might have favored long distance

pollen/seed dispersal allowing frequent exchange of alleles between different populations leading to low differentiation as noted here. Even though out-crossing is reported to be rare in *M. pruriens* (Duke 1981), it cannot be entirely ruled out owing to presence of numerous natural hybrids as reported in earlier works (Piper and Tracy 1910; Bailey 1947; Burkill 1966; Capo-chichi et al. 2001; Sathyanarayana et al. 2016). It is possible that genetic compatibility between wild and cultivated *M. pruriens* population has lead to gene flow and genetic introgression from wild populations to domesticated ones and vice-versa resulting in inter-varietal hybrids as reported in common bean cultivars of Andes (Freyre et al. 1996; Beebe et al.1997) and Mesoamerica (Papa and Gepts 2003).

Further, close examination of the UPGMA and STRUCTURE results points towards configuration of two independent gene pools possibly representing Assam plains (AP) and eastern Himalayas (EH). However sub-clusters within cluster-I (Assam plains) lacked state specific grouping pattern. This might be due to the fact that majority of these accessions were sampled from the areas adjoining Assam in the neighbouring states of Meghalaya, Tripura and Mizoram in south, Manipur and Nagaland in east and foothills of Arunachal Pradesh in the extreme of north. This topography forms a continuous range with considerable similarity in species distribution (Chattarjee 2008). Thus, *M. pruriens* distributed along this range perhaps correspond to single continuous population as gene flow has amalgamated neighboring populations into one, with common genetic structure. On the contrary, the accessions grouped in sub-cluster IIA likely represent divergent gene pool representative of eastern-Himalayas as these accessions were acquired from the lower altitudes of Sikkim-Darjeeling hills along the Teesta River which represents distinct floristic element. It is also possible that the latter might symbolize migrated elements from north Bengal plains through

anthropogenic activities or animal movements as this region (foothills of Sikkim-Darjeeling) is located adjoining to Siliguri-Malda plains of north Bengal. The grouping of two Tripura accessions along with cluster-IIA possibly can be explained based on this. That said, not even single *M. pruriens* population was to be found along the Sikkim-Darjeeling Himalayan region until few years back. However in recent years, continuous migration of several tropical plants including *M. pruriens* has been observed along the altitudinal gradient of EHs possibly signifying changing floristic composition of EHs due to climate change which is another interesting work for future investigation.

The other sub-cluster within cluster II (IIB) mostly represents accessions belonging to var. *utilis* - a cultivated accession. Alignment of var. *utilis* from AP along with cluster-II is little difficult to explain owing to following reasons: Firstly, these accessions were obtained as seed samples from private vendors/researchers and thus their precise origin could not be authentically established. Secondly, the variety itself presents a chequered background with no evidence available on its origin or genetic history from anywhere in the world and thus it is hard to pinpoint contributing factors for its atypical distribution. Thirdly, the fact that it freely hybridizes with other varieties of *M. pruriens* (var. *pruriens* and var. *hirsuta*) has resulted in several natural hybrids rendering precise genotypic identification difficult. This is evident even from the results of structure analysis in our study which confirms presence of several admixture taxa in our collection. The latter problem is common to all the three varieties of *M. pruriens* and thus has been a subject of debate in literature on *Mucuna* species (Capo-chichi et al. 2001; Eilitta et al. 2002; Sathyanarayana et al. 2016). Thus future work should focus on genetic history of *M. pruriens* varieties especially the cultivated var. *utilis* to reliably address questions related to pattern of diversity in

addition to migration, domestication and improvement of this promising underutilized crop.

6.4 Genetic analysis of L-Dopa trait

Any targeted genetic improvement program begins with the understanding of genetic mechanism underlying the trait of interest. Such pre-requisites include information on the nature of inheritance (monogenic or polygenic), variability in the gene pool and components - both genetic as well as non-genetic influencing the trait of interest (GXE interactions etc.). These information forms a basis for developing realistic and meaningful breeding program. However such data are generally lacking in large number of medicinal plants species and *M. pruriens* is no exception to this.

To determine the genetic nature of the L-Dopa trait, we analyzed its content in F1 hybrids and 272 F2 individuals. The F1 hybrids showed much lower L-Dopa content (3.82%) almost leaning towards low L-Dopa parent. This suggests, low L-Dopa content may be dominant over high L-Dopa content in *M. pruriens*. The results also suggests that there exists complete association of decreasing alleles in dominant parent (low L-Dopa) and increasing alleles in recessive parent (high L-Dopa) resulting in the above pattern of inheritance.

Among the F2 progenies, L-Dopa distribution was in the range of 1.44% to 5.84% (Table 37). The highest value obtained (5.84%) in any single F2 individual was less than the high L-Dopa parent suggesting extensive crossing-over along the genomic region controlling L-Dopa. We obtained sigmoid normal distribution curve for the L-Dopa content. This suggests L-Dopa content is a quantitative trait controlled by multiple loci. Such observations have been used earlier to draw inference on QTLs in other crop species as well (Kumar and Gupta 2015). However, this is first any such

report on the genetic nature of L-Dopa trait in *M. pruriens* or any other plant species where L-Dopa is present.

We also analyzed segregation of different seed traits in F2 progeny to determine their correlation with L-Dopa content. The P value was significant between all the seed traits (Table 38). However, most of them showed significant negative correlation with the L-Dopa content except seed thickness (0.010) and seed width (0.0797) which showed moderate correlation with L-Dopa content. This might be due to involvement of novel genes, which expresses independent of seed development events in L-Dopa production pathway. That said, the biosynthetic pathway and/or genetic regulation of L-Dopa production in *M. pruriens* has not been mapped so far. In mammals, L-Dopa production takes place via hydroxylation of tyrosine residues by tyrosine hydroxylase (Tabrez et al. 2012). In plants, it is believed to be analogous to that in mammals (Soares et al. 2014). Nonetheless, the overall correlation analysis suggests seed variable parameters may not be a selection criterion for identifying the genotype with high L-Dopa content.

It is widely regarded that the estimates of genetic and phenotypic variances and covariance are important for the prophecy of breeding values and for the prediction of expected genetic response of selection programs. In addition, genetic variability, heritability and genetic advance measures the relative degree to which a character is transmitted to the progeny (Eid, 2009; Ajayi et al. 2014). Therefore a breeder must have inside knowledge about the nature and magnitude of heritability (h^2) and genetic advances (GA) before launching any breeding program. Since biochemical traits are often inherited in a complex manner and influenced by the environment, path coefficient analysis will be an added advantage (Kahani and Hittalmani 2015). The study on the above aspects is essential to identify superior genotypes (Kahani and

Hittalmani 2015). Further, the crosses between genotypes with highest genetic divergence would be liable for improvement because, they have higher probability to yield desirable recombinants in the progeny. Keeping in view these factors, genetic studies in *M. prureins* was undertaken to estimate the genetic component of variance for L-Dopa content and further to compute the heritability, coefficients of variability and genetic advance in F2 segregating population.

On an expected line, the results showed higher phenotypic coefficient of variation (PCV; 133.51) than genotypic coefficient of variation (GCV; 112.77) for L-Dopa content. The difference between PCV and GCV is probably due to environmental effects (Kahani and Hittalmani 2015). Similar results are reported by Yadeta et al. (2011) and Sharma et al. (2010). However, since GCV provides only genetic variability information present in quantitative traits, it is not possible to determine the amount of heritable variation only from the GCV value. In this situation, GCV along with heritability estimates results in the improved outcome for the amount of advancement to be expected from the selection (Burton and Devane 1953).

Broad sense heritability (h^2) was thus calculated to analyze the heritable portion of the variation. Broad sense heritability is the ratio of genotypic variance to total variance (Hanson et al. 1956) and provides measure of effectiveness of selection. Johnson et al. (1955a) reported that heritability in broad sense can be categorized as low (0-30%), medium (30-60%) and high (>60%). Further, Singh (2001) has reported, heritability value greater than 80% can be regarded as very high, values between 60 to 79% are moderately high, values between 40-59% are medium and values less than 40% are low. From this view point, moderate to high heritability estimate (71.3%) was obtained for L-Dopa content in the present study. The result suggests that the environmental factors did not affect greatly the phenotypic performance of this trait

and selection based on the phenotypic performance will be reliable and effective for L-Dopa content. Lesser effects of environmental factors on L-Dopa production in *M. pruriens* is also established in earlier studies involving GXE interaction effects reported by Capo-chichi et al. (2003) and Mahesh and Sathyanarayana (2011). High heritability estimates leading to better selection efficiency have been reported in many earlier studies involving several crop species (Sreelathakumary and Rajamony 2004; Lestari et al. 2006; Yadeta et al. 2011; Qasim et al. 2013; Syukur and Rosidah 2014).

However heritability in broad sense includes both the additive and epistatic gene effects (Burton, 1953; Johnson et al. 1955a). Hence, the heritability estimates becomes more significant when accompanied by estimates of genetic advance (GA). The GA determines the improvement that may be made in a fastidious character by applying certain amount of selection intensity. Therefore we estimated genetic advance in this trait. The moderate value of GA (8.60) obtained for L-Dopa content in the present study indicate the trait is under additive gene action and thus its improvement could be achieved through mass selection (Ibrahim and Hussein 2006; Khatun et al. 2015).

In summary, it can be concluded that high heritability with moderate genetic advance values obtained for L-Dopa content reflect the presence of additive gene action for the expression of this trait which is fixable for next generations and selection in next population based on this character would be reliable.

6.5 Micropropagation of elite genotypes

Micropropagation is an important biotechnological tool used for multiplying elite genotypes in various plant species (Yasodha et al. 2004). Propagation of plants using *in vitro* methods has five main advantages over conventional method namely

multiplication of elite clones, multiplying plant throughout the year, regeneration of pathogen free plant even from the infected mother plants, cloning of male sterile lines and production of large number of plants in short time (Rani and Raina 2000). Selection of mother plant source for *in vitro* multiplication is an essential prerequisite for large scale propagation of the elite plant. Micropropagation of a superior/elite genotypes is usually carried to meet the production requirement of a quality and quantity plant material to enable availability at large commercial scale (Tesfa 2016). Hence it is imperative to employ efficient propagation system that realizes mass multiplication of newly evolved genotypes in a short period of time. Sanger et al. (2011) carried out similar study to ensure high quantity and quality of sugarcane planting material within a short period of time and space. Such studies have also been reported in medicinal plant *Artimisia annua* (Gupta et al. 1996).

6.5.1 Seed germination

In velvet bean, the dehusked seeds showed better germination than the seeds with seed coat. Removal of seed coat was found to increase the germination percentage by almost 50%. Decrease in germination with seed coat is attributed to the presence of phenolic compounds and their subsequent oxidation during germination process. Inhibition of germination due to oxidation of phenolic compounds in seed coat is a common phenomenon in legume plants as reported in *Phaseolus vulgaris* (Deshpande et al. 1982), pigeon pea (Singh 1993). Therefore, their removal of might have triggered better germination responses in the present study.

6.5.2 Micropropagation

Micropropagation, which is also known as as clonal propagation, is the technique of triggering the meristematic tissues of plants predetermined to form shoot to result in

multiple shoots by the influence of hormones *in vitro*. This has been the technique of widespread use, as the process is simple and involves lesser effort. The major benefits include rapid multiplication, maintenance of uniformity, disease free plant regeneration and propagation of sexually derived sterile hybrids. It has been proved unique in case of virus free plant establishment (Biswas et al. 2007) and genetic transformation through *Agrobacterium tumefaciens* and particle bombardment techniques (Gondo et al. 2009).

In the present study, of the three explants tried, the axillary bud and CN produced equal response, which was higher compared to apical buds on BAP (2.2-22.2µM) containing medium. The result indicates explant specific nature of response to the plant growth regulators due to differential expression of genes in different plant parts (Chandra and Pental 2003). The superiority of axillary buds over CN and the apical bud for clonal propagation is also reported in *Dolichos* (Sujatha et al. 2007). In Turkish chickpea, the apical buds produced shoot numbers as low as 1–3 (Aasim et al. 2008). However, failure of cotyledonary nodes to produce good quality shoots is not clearly understood and remains unexplained.

Of the various hormones tested for clonal propagation, BAP was found to be superior over Kn and other hormonal combinations in evincing better response from CN and axillary bud explants. Superiority of BAP over Kn for multiple shoot induction is in accordance with *Eclipta alba* (Franca et al. 1995), *Sapium sebiferum* (Siril and Dhar 1997), and *Pterocarpus marsupium* (Suresh and Ajay 2004). The stunted nature of shoot formation corresponding to increased concentration of BAP in the medium is also reported in *Orthosiphon* (Keng et al. 2004) and *Eupatorium* (Martin 2004).

Nonetheless, earlier report in *M. pruriens* (Chattopadhyay et al. 1995; Faisal and Siddique 2006 a & b) showed maximum shoot induction on NAA + 2ip and NAA + BAP, while this hormone combination produced only 3-5 shoots inducing basal callus in the present study. These differential responses, revealed by different *M. pruriens* varieties indicate possible existence of genotype specific responses within *M. pruriens* varieties. Similar results are reported even in Mulberry cultivars (Tewary et al. 1996). It is well established that, in cultured tissues, the requirement for exogenous hormone depends on the endogenous levels in the plant tissue which varies with organ, plant genotype, and the phase of the growth (Suresh and Ajay 2004). Thus, the result establishes requirement of independent standardization in each of the varieties of *M. pruriens* after careful consideration of varietal identity.

Axillary bud explants showed better elongation on GA_3 containing medium which were induced for optimum rooting on NAA supplemented medium. The inductive effect of GA_3 on elongation of shoots is well-established in plant tissue culture (Jayanand et al. 2003). Successful elongation of shoots using GA_3 has been achieved even in pigeon pea (Villiers et al. 2008). NAA as the key hormone for rooting is also reported in several micropropagation protocols like *M. pruriens* var. *pruriens*, *Pisonia alba*, *Hyptis suaveolens*, *Jatropha curcas* and *M. pruriens* var. *utilis* (Chattopadhyay et al. 1995; Chandra et al. 1999; Britto et al. 2001; Rajore et al. 2002; Sathyanarayana et al. 2008).

Thus, the comparative evaluation of axillary bud, apical bud and cotyledonary node reveals that the axillary bud and the cotyledonary node explants produced best response on MS medium supplemented with BAP (22.2 µM) of which, axillary bud proves better over the CN in terms of quality of shoots in *M. pruriens* and thus can be used in large scale micropropagation of elite genotype (500240SK).

7. Summary and Conclusions

Among the handful of plants that produce L-Dopa, *M. pruriens* is the best known source in terms of both quality and efficacy of the drug (Ghosal et al., 1971; Misra and Wagner, 2004, 2007). Notwithstanding the agronomic and medicinal potential offered by this species, it has remained largely underexploited owing to poor research efforts, particularly in the direction of genetic improvement. In view of this, the present thesis work was carried out to (a) collect and characterize the *M. pruriens* germplasm from northeast India - part of centre of origin of this crop and (b) perform genetic and heritability analysis of L-Dopa trait using F2 segregating population. In the end, we have proposed mass multiplication protocol for elite genotype for large-scale propagation.

7.1 Germplasm collection and characterization

Fifty *M. pruriens* accessions belonging to both var. *pruriens* (38) and var. *utilis* (12) were collected from different parts of northeast India and established in Sikkim University botanical garden. All the accessions were evaluated for 10 qualitative and quantitative traits each, as per the standard descriptors. The results showed qualitative characters exhibiting mean diversity of 61%, and quantitative traits showing mean diversity of 56%. Wild germplasm represented by var. *pruriens* was more diverse for both qualitative and quantitative characters. The UPGMA clustering grouped the accession into major clusters based on taxonomic and phenetic affinities. The work also suggested good scope for inclusion of wild germplasm in the breeding programs. The multivariate analyses helped identification of several elite and contrasting parental lines for inclusion in the breeding program.

7.2 Assessment of L-Dopa variability

For L-Dopa analysis, HPLC based method was first standardized using several validation experiments. The standardized protocol thus developed was used to screen variability in the germplasm collection. The results revealed L-Dopa content in the range of 1.55 to 9.03% at 46.80% coefficient of variation suggesting existence of high variability for this trait in the germplasm collection.

7.3 Genetic diversity estimation

Almost all the primer combinations produced high quality AFLP profile. They generated an average of 54.91 fragments suggesting good genetic variability among the studied accessions. Of the 22 PCs, 9 (41%) produced 100% polymorphic fragments. Based on the combination of PIC, MI and RP parameters, the PCs - E-ACG/M-CAC, E-ACA/M-CAA, E-AAG/M-CAC and E-ACA/M-CAT and RP PCs E-ACT/M-CTA, E-AGG/M-CTA and E-AAC/M-CAG have been suggested as useful PCs for diversity studies. The Jaccard's similarity coefficients based genetic diversity estimates showed good diversity (0.37) which is reflected even in the values of h and Ht with mean scores of 0.27 and 0.22, respectively. The population structure investigated using UPGMA, AMOVA and STRUCUTRE revealed absence of geographical structure. In case of AMOVA, the results revealed 93% of the variance was accrued by intra-population variation, while only 7% contributed to inter-population variation pointing towards absence of barriers for gene flow.

Thus, it can be concluded that, *M. pruriens* population of northeast India exhibits good diversity providing credence to northeast Indian origin theory of this species. Pair-wise estimates of genetic differentiation were low and AMOVA revealed no evidence of genetic structuring, suggesting species did not experience severe

bottleneck events. However, clustering based on UPGMA and STRUCTURE broadly point to presence of two gene pools representing Assam plains and eastern Himalayas. The future work should focus on the mechanisms underlying such variation besides addressing key questions related to migration and domestication history of *M. pruriens* varieties in general and var. *utilis* in particular.

7.4 Genetic analysis for L-Dopa trait

To determine the genetic nature of L-Dopa trait, we analyzed L-Dopa content in 51 F1 hybrids and 272 F2 individuals derived from a single cross. All the F1 hybrids showed lower L-Dopa content (average: 3.82%), suggesting low L-Dopa may be dominant over high L-Dopa content. In F2 populations, the highest L-Dopa value obtained (5.84%) for any single F2 individual was less than the high L-Dopa parent (9.03%) suggesting extensive crossing-over along the genomic region controlling the L-Dopa production. Further, we obtained sigmoid normal distribution curve for L-Dopa content suggesting quantitative nature of the trait. This is first any such report on the genetic nature of L-Dopa trait in *M. pruriens* or any other plant species. Further, high heritability (71.3%) with moderate genetic advance as percentage of mean values (196.23) obtained for L-Dopa content reflect the presence of additive gene action which is fixable for next generations and selection based on this trait would be reliable and effective.

7.5 Micropropagation of elite genotypes

Mass multiplication of representative elite genotype 500240SK was attempted using three different explants. Of these, the auxiliary bud and CN produced equal response (6.4± 0.55) and was higher compared to apical buds (2.6± 0.24) on BAP containing medium. The result indicates explants specific nature of response to the plant growth

regulators due to differential expression of genes in different plant parts. BAP (22.2µM) was found to be superior over Kn and other hormonal combinations. Axillary bud explants showed better elongation on GA_3 (2.89 µM) containing medium which were induced for optimum rooting on NAA (5.40 µM). The protocol proposed will be useful for large scale production of viable, uniform and healthy plants with maximum survival rate.

In conclusion, investigation encompassed in this thesis is expected to trigger research works on extended germplasm collection, development of core collections, linkage map development etc. in an immediate future. In a longer run, this will form a strong starting point for both conventional as well as molecular breeding regimes involving Marker Assisted Selection (MAS) and other biotechnological interventions for L-Dopa breeding. Further, a comparative RNASeq analysis between contrasting L-Dopa parents identified during the study and their correlation with metabolomics is likely to provide key information and leads on the biosynthetic pathway of L-Dopa production in *M. pruriens* and the gene candidates involved. All these will go a long way in targeted genetic improvement of L-Dopa content and sustainable cultivation of this important medicinal plant.

8. Future Prospects

The anticipated future works as an extension of this thesis work are as follows:

1. Widening of germplasm base using collection from extended geographical regions for categorizing trait specific germplasm.
2. Development of core collection sets for L-Dopa trait.
3. Validating genetic nature of L-Dopa trait through 6^{th} generation mean analysis.
4. Development of RIL (F6) population for L-Dopa content.
5. QTL mapping of L-Dopa trait using co-dominant markers such as SSR or SNPs.
6. Identification of gene involved in L-Dopa synthesis pathway through functional genomics approaches.

9. Bibliography

Aasim M, Khawar KM, Ozcan S (2008). *In vitro* micropropagation from shoot meristems of Turkish cowpea (*Vigna unguiculata* L.) cv Akkiz. Bangladesh Journal of Botany 37 (2) 149-154

Abbott A (2010). Levodopa: The story so far. Nature (466) 6–7

Abraham J (2011). Effect of *M. pruriens* seeds on fertility of male albino rats *Rattus norvegicus*. Journal of Pharmacy Research 4 (1) 233–236

Adewale BD, Adeigbe OO, Aremu C (2011). Genetic distance and diversity among some cowpea (*Vigna unguiculata* L. Walp.) genotypes. International Journal of Plant Science (1) 9–14

Adreeja B, Lokanadha RG, Lingaraj S (2017). Morphometric characterization of *Jatropha curcas* germplasm of northeast India. African Journal of Biotechnology (16) 648–656.

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

Ajayi AT, Adekola MO, Taiwo BH, Azuh VO (2014). Character expression and differences in yield potential of 10 genotypes of cowpea (*Vigna unguiculata* L. Walp.). International Journal of Plant Research (4) 63–71

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

Alekel DL, St-Germain A, Pererson CT, Hanson KB, Stewart JW, Toda T (2000). Isoflavone-rich soy protein isolate attenuates bone loss in the lumbar spine of perimenopausal women. American Journal of Clinical Nutrition (72) 844-852

Ali S, Shultz, JL, Qadeer, MA (2007). High performance microbiological transformation of L-tyrosine to L-Dopa by *Yarrowia lipolytica* NRRL-143. BMC Biotechnology 7 (50)

Alin MH (1997). *In vitro* susceptibility of Tanzanian wild isolates of *Plasmodium falciparum* to artemisinin chloroquine sulfadoxine/pyrimethamine and mefloquine. Parasitology (114) 503–506

Allard RW (1960). Principles of Plant breeding. John Willey and Sons Inc. New York

Allen RS, Millgate AG, Chitty JA (2004). RNAi-mediated replacement of morphine with the non-narcotic alkaloid reticuline in opium poppy. Nature Biotechnology (22) 1559-1566

Analyse-it Software Ltd (2009). http://wwwanalyse-itcom/

Annual Administrative Report (2014-15). Department of Forests Ecology Environment & Wildlife, Kohima, Nagaland

Aremu OT, Emelue GU, Osayimwen FE, Obasogie FO (2012). Estimate of habitat quality of white throated monkey (*Ceropithecus erythrogaster)* in Okomu National Park, Nigeria. Nigerian Journal of Agriculture, Food and Environment 8 (1) 47–51

Arora KR (1991). Native food plants of the northeastern India. In: Contributions to Ethnobotany of India (Eds. SK Jain). Scientific Publishers, Jodhpur, India 137-152

Arora RK, KPS Chandel (1972). Botanical source areas of wild herbage legumes in India. Tropical Grassland (6) 213-221

Arulmozhi M, Janardhanan K (1992). The biochemical composition and nutritional potential of the tribal pulse *M. monosperma* DC. Ex. Wight. Plant Foods for Human Nutrition (42) 45-53

Asante K, Offei SK, Addy R, Carson AG (2008). Phenotypic and seed protein analysis in 31 lima bean (*Phaseolus lunatus*) cultivars in Ghana. West African Journal of Allied Ecology (12) 129-138

Asgarpanah J (2012). Phytopharmacology and medicinal properties of *Salix aegyptiaca* L. African Journal of Biotechnology (11) 7145–7150

Awang D, Buckles D, Arnason JT (1997). The phytochemistry, toxicology and processing potential of the covercrop velvet bean (cowhage, cowitch) (*Mucuna* Adans. s, Fabaceae). International Workshop on Green Manure – Cover Crop Systems for Smallholders in Tropical and Subtropical Regions 6–12 April, Chapeco, Santa Catarina, Brazil. Santa Catarina, Brazil: Rural Extension and Agricultural Research Institute of Santa Catarina.

Azhaguvel P, Vidya SD, Sharma A, Varshney RK (2006). Methodological advancement in molecular markers to delimit the genes for crop improvement. Floriculture Ornamental and Plant Biotechnology. Global Science Books, London (50) 460–468

Bailey LH (1947). The standard cyclopedia of horticulture. Macmillan New York, USA

Baluchnejadmojarad T, Roghani M, Jalali MR, Bagheri NM (2009). Neuroprotective effect of genistein in 6-hydroxydopamine hemiparkinsonian rat model. Phytotherapy Research (23)132–135

Barbeau A (1980). Lecithin in Parkinson's disease. Journal of Neural Transmission (16) 187–193

Barron B, Sovik E, Cornish JL (2010). The roles of dopamine and related compounds in reward seeking behavior across animal phyla. Frontiers in Behavioral Neuroscience 4 (2) 163

Beebe S, Toro O, Gonzales AV, Chacon MI, Debouck D (1997). Wile-weed-crop complexes of common bean (*Phaseolous vulgaris* L. Fabaceae) in the Andes of Peru and Colombia and their implications for conservation and breeding. Genetic Resources and Crop Evolution (44) 73–91

Bell CJ, Dixon RA, Farmer AD, Flores R, Inman J, GonZales RA, Harrison MJ, Paiva NL, Scott AD, Weller JW (2001). The *Medicago* genome initiative: A model legume database. Nucleic Acids Research 29 (1) 114–117

Bell EA, Janzen DH (1971). Medical and ecological considerations of L-Dopa and 5- HTP in seeds. Nature (229) 136-137

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

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) 303-309

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

Biswas MK, Hossain M, Islam R (2007). Virus free plantlets production of strawberry through meristem culture. World Journal of Agricultural Sciences 3 (6) 757- 763

Blum E, Mazourek MO,'Connell M, Curry J, Thorup T, Liu K, Jahn M, Paran I (2003). Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. Theoretical and Applied Genetics (108) 79–86

Bousquet M, Saint-Pierre M, Julien C, Salem N, Cicchetti F, Calon F (2008). Beneficial effects of dietary omega-3 polyunsaturated fatty acid on toxin-induced neuronal degeneration in an animal model of Parkinson's disease. FASEB Journal (22) 1213–1225

Brain KR (1976). Accumulation of L-Dopa in cultures from *M. pruriens*. Plant Science Letters (7) 157-161

Bravo L, Siddhuraju P, Saura-Calixto F (1998). Effect of various processing methods on the *in vitro starch* digestibility and resistant starch content of Indian pulses. Journal of Agriculture and Food Chemistry (46) 4667–4674

Bressani R (2000). Factors influencing nutritive value in food grain legumes: *Mucuna* compared to other grain legumes. In: Flores M, Eilittä M, Myhrman R, Carew LB, Carsky RJ (Eds.) *Mucuna* as a Food and Feed: Current uses and the way forward workshop held April 26-29 2000 in Tegucigalpa, Honduras. CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, Honduras, CIDICCO, Honduras 2002 164-188

Brink M, Belay G (2006). Plant resources of Tropical Africa 1. Cereals and Pulses-PROTA Foundation, Backhuys Publishers, Wageningen, Netherlands

Britto SJ, Robinson JJ, Natarajan E, Arockiasamy DI (2001). Micropropagation of *Hyptis suaveolens* (L.) Poit through nodal culture. Advances in Plant Sciences 14 (11) 561-565

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 *M. pruriens* for tumerogenic substances. Tropical and Subtropical Agroecosystems (1) 287-293

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

Burton GW, Devane EH (1953). Estimating heritability in tall fescue (*Festuca arundinacea*) from replicated clonal material. Agronomy Journal 45 (10) 478–481

Caetamo-Anollés G, Gressholff PM (1998). DNA analysis of turf grass genetic diversity. Crop Science (38) 1415-1424

Calne DB, Sandler M (1970). L-Dopa and Parkinsonism. Nature (226) 21–24

Cannon SB, McKain MR, Harkess A, (2015). Multiple polyploidy events in the early radiation of nodulating and non-nodulating legumes. Molecular Biology and Evolution 32 (1) 193–210

Canter PH, Thomas H, Ernst E (2005). Bringing medicinal plants into cultivation : opportunities and challenges for biotechnology. Trends in Biotechnology (23) 180-185

Capo-chichi LJA, DBW, CMM (2003a). The use of molecular markers to study genetic diversity in *Mucuna*. Tropical and Subtropical Agroecosystems (1) 309–318

Capo-chichi LJA, Eilittä M, Carsky R, Gilbert R, Maasdorp B (2003b). Effect of genotype and environment on L-Dopa concentration in *Mucuna*’s (*Mucuna* sp.) seeds. Tropical and Subtropical Agroecosystems (1) 319-328

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) 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) 1180-1188

Carlen C (2011). Breeding and cultivation of medicinal plants. Herbal Medicine 79–91

Carsky RJ, Ndikawa R (1998). Identification of cover crops for the semi-arid savanna zone of West Africa. In: Buckles D, Eteka A, Osiname M, Galiba M, Galiano G (Eds.) Cover crops in West Africa - contributing to sustainable agriculture. IDRC, IITA, Sasakawa Global 2000, Otawa, Canada, Ibadan, Nigeria, Cotonou, Benin 179-187

Cerenius L, Söderhäll K (2004). The prophenoloxidase-activating system in invertebrates. Immunology Review (198) 116–126

Chakraborty P (2018). Herbal genomics as tools for dissecting new metabolic pathways of unexplored medicinal plants and drug discovery. Biochimie Open (6) 9-16

Chandra A, Pental D (2003). Regeneration and genetic transformation of grain legumes: An overview. Current Science 84 (3) 381-387

Chandra KS, Rachaaji JS, Gowda KRD, Saraswathi TKJ (1999). *In vitro* propagation of *Pisonia alba* (Linn.) Sapanogae (Lettuce Tree) A threatened species. Phytomorphology (49) 43-47

Chandrawati MR, Singh PK, Ranade SA, Yadav HK (2014). Diversity analysis in Indian genotypes of linseed (*Linum usitatissimum* L.) using AFLP markers. Gene (549) 171–178

Charlton C, Crowell B (1992). Parkinson's disease-like effects of S-adenosyl-L-methionine: Effects of L-Dopa. Pharmacology Biochemistry and Behavior (43) 423–431

Chatterjee S (2008). Biodiversity conservation issues of northeast India. International Forestry Review 10 (2) 315-324

Chattopadhyay S, Datta SK, Mahato SB (1995a). Production of L-Dopa from cell suspension culture of *M. pruriens* f. *pruriens*. Plant Cell Reports (13) 519-522

Chattopadhyay S, Datta SK, Mahato SB (1995). Rapid micropropagation for *M. pruriens* f. *pruriens* L. Plant Cell Reports (15) 271-273

Chauhan HS, Khanuja SPS, Kalra A, Gupta AK, Shasany AK, Darokar MP, Mathavan RE, Singh HP, Patra DD, Singh HN, Singh J, Gupta ML, Singh A, Lal RK, Bansal RP, Dhawan OP, Singh AK, Kumar B, Garg SN (2006). A high rhizome and oil yielding variety "Cim-Balya" of Vach (*Acorus calamus*). Journal of Medicinal and Aromatic Plant Science (28) 632-634

Chem Abstr (1990). (113) 29157

Chem Abstr (1991). (114) 108913

Chen S, Xu J, Liu C (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. Nature Communications (3) 913-922.

Chen W, Kui L, Zhang G, Zhu S, Zhang J, Wang X, Yang M, Huang H, Liu Y, Wang Y, Li Y, Zeng L, Wang W, He X, Dong Y, Yang S (2017). Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. Molecular Plant (10) 899–902

Chikagwa-Malunga SK, Adesogan AT, Salawu MB, Szabo NJ, Littell RC, Kim SC, Phatak SC (2009). Nutritional characterization of *M. pruiriens*. 2: *In vitro* ruminal fluid fermentability of *M. pruriens, Mucuna* L-Dopa and soybean meal incubated with or without L-Dopa. Animal Feed Science and Technology (148) 51-67

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 (36) 1–8

Collard BCY, 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) 169–196

CSIR (1962). The Wealth of India: A dictionary of Indian raw materials and industrial products. CSIR, New Delhi

Damodaran M, Ramaswamy R (1937). Isolation of L-Dopa from the seeds of *M. pruriens*. Biochemistry (31) 2149-2151

Darokar MP, Khanuja SPS, Gupta AK, Kumar S, Gupta MM, Singh AK, Shasany AK, Raju BV, Mishra PK, Rahman L, Kalra A, Verma RK, Mishra S, Gangwar A, Mishra HO, Chauhan HS, Singh S, Bahl JR, Kumar B, Lal RK, Singh AK, Bansal RP, Tandon S, Dhawan OP (2007). 'CIM-Jagriti' a variety of Brahmi (*Bacopa monnieri*) with high herbage and bacoside yield potential. Journal of Medicinal and Aromatic Plant Science (29) 90-93

Davidson-Hunt I (2000). Ecological Ethnobotany: Stumbling toward new practices and paradigms. MASA Journal (16) 1–13

Daxenbichler ME, Kleiman R, Weisleder D, Van Etten CH, Carlson KD (1972). A new amino acid, (-)-1-methyl-3- carboxy-6,7-dihydroxy-1,2,3, 4-tetrahydroisoquinoline, from velvet bean. Tetrahedron Letters (18) 1801-1802

Daxenbichler ME, Van Etten CH, Hallinan EA, Earle FR, Barclay SA (1971). Seeds as sources of L-Dopa. Journal of Medicinal Chemistry (14) 463- 465

De Riek J, Calsyn E, Everaert I, Van Bockstaele E, De Loose M (2001). AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theoretical and Applied Genetics (103) 1254–1265

Deb DB (1961). Dicotyledonous plants of Manipur territory. Botanical Survey of India, Shilong, India (3) 3-4

Deb DB (1981). Flora of Tripura state. ASTRAL Publishers, India

Debbarma M, Pala NA, Kumar M, Bussmann RW (2017). Traditional knowledge of medicinal plants in tribes of Tripura in northeast India. African Journal of Traditional, Complementary and Alternative Medicines 14 (4) 156-168

Debrunner N, Rauber AL, Schwarz A, Michel V (2000). First report of St. John's-Wort anthracnose caused by *Colletotrichum gloeosporioides* in Switzerland. Plant Disease 84 (2) 203

Deshpande SS, Sathe SK, Salunkhe DK, Cornforth DP (1982). Effects of dehulling on phytic acid, polyphenols and enzyme-inhibitors of dry beans *(Phaseolus vulgaris* L.). Journal of Food Science (47) 1846-1850

Dewick DM (2002). Medicinal natural products, a biosynthetic aroach. John Wiley & Sons Inc, New York 1

Dhawan SS, Rai GK, Darokar MP, Lal RK, Misra HO, Khanuja SPS (2011). Comparative genetic analysis of trichome-less and normal pod genotypes of *M. pruriens* (Fabaceae). Genetics and Molecular Research 10 (3) 2049-2056

Diallo OK, Kante S, Myhrman R, Soumah M, Cissé NY, Berhe T (2002). Increasing farmer adoption of *M. pruriens* as human food and animal feed in the republic of Guinea In: Flores BM, Eilittä M, Myhrman R, Carew LB, Carsky RJ, (Eds.) Food and feed from *Mucuna*: Current uses and the way forward workshop, (April 26-29 2000) CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, Honduras 60-72

Dikshit N, Sivaraj N, Das AB (2014). Genotypic variations in agro-morphological and physico-chemical traits in ethnic aromatic rice (*Oryza sativa*) landraces of Odisha. Indian Journal of Agricultural Sciences 84 (8) 987-992

Ding Y, Kinjo J, Yang C, Nohara T (1991). Triterpenes from *M. birdwoodiana*. Phytochemistry (30) 3703–7

Dixon RA, Ferreira D (2002). Molecules of interest: Genistein. Phytochemistry (60) 205–211

Dixon RA, Sumner LW (2003). Legume natural products : Understanding and manipulating complex pathways for human and animal health. Plant Physiology (131) 878–885

Doaa AR, Magda A, Bendary EL (2010). Production of 3, 4-dihydroxyphenyl-L-alanine (L-Dopa) by Egyptian halophilic black yeast. World Journal of Microbiology and Biotechnology (27) 39-46

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

Drewes SE, Horn MM, Munro OQ, Dhlamini JTB, Meyer JJM, Rakuambo NC (2002). Pyrano-isoflavones with erectile-dysfunction activity from *Eriosema kraussianum*. Phytochemistry (59) 739–747

Dubey MK, Shasany AK, Dhawan OP, Shukla AK, Shanker K, Khanuja SPS (2010). Detection of DNA polymorphism in *Papaver somniferum* genotypes differing in straw morphinan alkaloid content. Plant Biosystematics (144) 513–517

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

Dunne EL, Moss SJ, Smart TG (1998). Inhibition of GABAA receptor function by tyrosine kinase inhibitors and their inactive analogues. Molecular and Cellular Neuroscience (12) 300–310

Dunnett CW (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of Acoustical Scciety of America (50) 1096-1121

Dyer GA, Lopez-Feldman A, Yunez-Naude A, Taylor JE (2014). Genetic erosion in maize's centre of origin. Proceedings of the National Academy of Sciences U.S.A 111 (39) 14094-14099

Earl DA, vonHoldt BM (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetic Resources (4) 359–361

Egounlety M (2003). Processing of Velvet bean (*M. pruriens* var. *utilis*) by fermentation. Tropical and Subtropical Agroecosystems (1) 173-181

Eid MH (2009). Estimation of heritability and genetic advance of yield traits in wheat (*Triticum aestivum* L.) under drought condition. International Journal of Genetics and Molecular Biology (1) 115–120

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

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: Flores BM, Eilittä M, Myhrman R, Carew LB, Carsky RJ, (Eds.). Food and feed from *Mucuna*: Current uses and the way forward workshop. CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, Honduras (April 26-29, 2000) 18-47

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

El-Khishin DA, Belatus EL, El-Hamid AA, Radwan KH (2009). Molecular characterization of banana cultivars (*Musa* Spp.) from Egypt using AFLP. Research Journal of Agricultural and Biological Science 5 (3) 272–279

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology (14) 2611–2620

Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics (1) 47–50

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131 (2) 479–491

Faisal M, Siddique I, Anis M (2006a). An efficient plant regeneration system for *M. pruriens* L. (DC.) using cotyledonary node explants. In Vitro Cellular & Developmental Biology - Plant 42 (1) 59-64

Faisal M, Siddique I, Anis M (2006b). *In vitro* rapid regeneration of plantlets from nodal explants of *M. pruriens* - A valuable medicinal plant. Annals of Applied Biology 148 (1) 1–6

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics (164) 1567-1587

Fathima KR, Tresina Soris P, Mohan VR (2010). Nutritional and antinutritional assessment of *M. pruriens* (L.) DC. var. *pruriens* an underutilized tribal pulse. Advances in Bioresearch (1) 79-89

First EA (2011). L-Dopa ropes in tRNA. Chemical Biology (18) 1201-1202

Flora of Manipur (2000). Volume-I. Ranunculaceae - Asteraceae by the BSI

Franca SC, Bertoni BW, Pererira AMS (1995). Antihepatoxic agent in micropropagated plantlets of *Eclipta alba*. Plant Cell Tissue and Organ Culture (40) 297-299

Freyre R, Rı'os R, Guzma'n L, Debouck D, Gepts P (1996). Ecogeographic distribution of *Phaseolus* spp. (Fabaceae*)* in Bolivia. Economic Botany 50 (2) 195–215

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

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) 254-259

Garcia AAF, Benchimol LL, Barbosa AMM, Geraldi IO, Souza Jr CL, Souza AP de (2004). Comparison of RAPD, RFLP, AFLP, and SSR markers for diversity studies in tropical maize inbred lines. Genetics and Molecular Biology (27) 579–588

Gautam M, Chandel M, Azmi W (2012). Therapeutic role of L-Dopa produced as a secondary metabolite from different legumes and plant sources. Annals of Phytomedicine (1) 1–8

Gautam SS, Navneet KS (2012). The antibacterial and phytochemicals aspects of *Viola odorata* Linn. extracts against respiratory tract pathogens. Proceedings of the National Academy of Sciences, India 82 (4) 567-572

Genevieve M, Hoopes, JPH, Jeongwoon K, Zhao D, Wiegert-Rininger K, Crisovan E, Buell CR (2018). Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. G3: Genes, Genomes, Genetics 8 (2) 385-391

George EF, Sherrington PD (1984). Plant propagation by tissue culture: Hand book and directory of commercial laboratories. Journal of Basic Microbiology 25 (7)

George JS (2006). Basic neurochemistry: Molecular Cellular and Medical Aspects (7th ed). Elsevier Academic Press, USA

Gepts P, Beavis WD, Brummer EC (2005). Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through Genomics Conference. Plant Physiology (137) 1228–1235

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Molecular Ecology (9) 1037-1048

Ghafoor MA, Anwar R (2003). Diversity in food legumes for sustainable utilization of plant genetic resources In: Sustainable utilization of plant genetic resources for agricultural production. Anwar R, Takahashi J, Bhatti MS, Masood S, (Eds.) PARC/IPGRI/JICA, Islamabad, Pakistan 238-250

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

Ghose G (2017). Traditional use of plants against snakebite in Indian subcontinent : A review of the recent literature. African Journal of Traditional, Complementry and Alternative Medicines (4) 55–64

Golisz A, Sugano M, Hiradate S, Fujii Y (2011). Microarray analysis of *Arabidopsis* plants in response to allelochemical L-Dopa. Planta (40) 231- 240

Gondo T, Matsumoto J, Tsuruta SI, Yoshida M, Kawakami A, Terami F, Ebina M, Yamada T, Akashi R (2009). Particle inflow gun-mediated transformation of multiple-shoot clumps in Rhodes grass (*Chloris gayana*). Journal of Plant Physiology 166 (4) 435-441

Gonzalez-Castejon M, Rodriguez-Casado A (2011). Dietary phytochemicals and their potential effects on obesity: A review. Pharmacological Research (64) 438-455

Gowen SR, Quenherve P, Fogain R (2005). Nematode parasites of bananas and plantains. 611-643. In: Luc M, Sikora RA, Bridge J. (Eds.). Plant parasitic Nematodes in Subtropical and Tropical Agriculture, 2nd Edition. CAB International Edit, Wallingford Oxon, UK

Graf E, Eaton JW (1990). Antioxidant activity of phytic acid. Free Radical Biology and Medicine (8) 61–69

Graph Pad Software Inc. (2014). GraphPad Statistics Guide 1-402

Guerranti R, Ogueli IG, Bertocci E, Muzzi C, Aguiyi JC, Cianti R, Armini A, Bini L, Leoncini R, Marinello E, Pagani R (2008). Proteomic analysis of the pathophysiological process involved in the antisnake venom effect of *M. pruriens* extract. Proteomics 8 (2) 402–412

Guggenheim M (1913). Dihydroxyphenylalanine a new amino acid from *Vicia faba*. Hoppe-Seyler´s Zeitschrift für physiologische Chemie (88) 276

Gullo VP, McAlpine J, Lam KS, Baker D, Petersen FJ (2006). Drug discovery from natural products. Industrial Microbiology and Biotechnology (33) 523

Guo Y, Chen S, Li Z, Cowling WA (2014). Centre of origin and centres of diversity in an ancient crop *Brassica rapa* (Turnip Rape). Journal of Heredity 105 (4) 555–565

Gupta A, Maurya R, Roy RK, Sawant SV, Yadav HK (2013). AFLP based genetic relationship and population structure analysis of *Canna*-An ornamental plant. Scientia Horticulturae (154) 1–7

Gupta M, Jain C, Mathur DA, Singh AK, Verma RK, Kumar S (1996). Isolation of a high artemisinic acid containing plant of *Artemisia annua*. Planta Medica (62) 280-281

Gupta V, Kak A (2007). Genetic diversity analysis in *M. pruriens* germplasm using SDS- PAGE. International Journal of Plant Sciences (2)124-127

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) 105-115

Hadapad B, Ravi CS, Shivaprasad M (2018). Genetic variability and correlation studies for quantitative and qualitative traits in velvet bean (*M. pruriens* L.) genotypes in rubber plantation under hill zone of Karnataka. Journal of Pharmacognosy and Phytochemistry (3) 86–90

Hajra PK, Verma DM, Giri GS (1996). Materials for the flora of Arunachal Pradesh. Botanical Survey of India - I

Haneda K, Watanabe S, Takeda I (1971). Synthesis of 3,4-dihydroxyphenyl L-alanine from L-tyrosine by microorganisms. Applied Microbiology (22) 721-722

Hanson CH, Robinson HF, Comstock RE (1956). Biometrical studies in yield in segregating populations of Korean lespedeza. Agronomy Journal (48) 214-318

Hao D, Ma P, Mu J, Chen S, Xiao P, Peng Y, Huo L, Xu L, Sun C (2012). *De novo* characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. Science China Life Sciences 55(5) 452–466

Hao da C, Chen SL, Osbourn A, Kontogianni VG, Liu LW, Jordán MJ (2015). Temporal transcriptome changes induced by methyl jasmonate in *Salvia sclarea*. Genetics 558 (1) 41–53

Hao da C, Ge G, Xiao P, Zhang Y, Yang L (2011). The first insight into the tissue specific *Taxus* transcriptome via illumina second generation sequencing. PLoS One 6 (6) 1-15

Haq N (1983). New Food legume crops for the tropics. In: Nugent J, Cormor MO (Eds.). Better crops for the food, Pitman Books, Cuba Foundation Symposium, London 144-160

Hardidasan K, Rao RR (1985). Forest flora of Meghalaya. Bishen Singh Mahendra Pal Singh, Dehra Dun, India

Harlan JR (1971). Agricultural Origins: Centers and Non-centers. Science 174 (4) 468-474

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1996). PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theoretical and Applied Genetics (93) 1202–1210

Hinz M (2009). Depression. In: Kohlstadt I (Eds.) Food and nutrients in disease management. Baton Rouge, CRC Press 465–481

Hollister JD, Greiner S, Wang W (2015). Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. Molecular Biology and Evolution 32 (4) 896–905

Hooker JD (1872- 1897). The flora of British India. London 1-7

Hooker WJ (1849). The Rhododendrons of Sikkim-Himalaya (2nd ed). Reeve Benham and Reeve, London

Horbovitz M, Brenac P, Obendorf RL (1998). Fagopyritol B1 O-α- D-galactopyranosyl-(1→2)-D-chiro-inositol a galactosylcyclitol in maturing buckwheat seeds is associated with desiccation tolerance. Planta (205) 1–11

Hornykiewicz O (2002). L-Dopa: From a biologically inactive amino acid to a successful therapeutic agent. Amino Acids 23 (1-3) 65-70

Huang LZ, Campos C, Ly J, Carroll FI, Quik M (2011). Nicotinic receptor agonists decrease L-Dopa-induced dyskinesias most effectively in partially lesioned parkinsonian rats. Neuropharmacology (60) 861–868

Huang SY, Chen SY (1998). Efficient L-Dopa production by *Stizolobium hassjoo* cell culture in a two stage configuration. Journal of Biotechnology 62 (2) 95-103

Huang SY, Chen SY, Wu KL, Taung WT (1995). Strategy for inducing pertinent cell line and optimization of the medium for *Stizolobium hassjoo* producing L-Dopa. Journal of Fermentation and Bioengineering 79 (4) 342-347

Huang SY, Hu JJ (2001). Effects of gas composition on the cell growth and L-Dopa production in the suspension culture of *Stizolobium hassjao.* Journal of the Chinese Institute of Chemical Engineers (32) 1-11

Huizing HJ, Wijnsma R, Batterman S, Malingre M, Wichers HJ (1985). Production of L-Dopa by cell suspension cultures of *M. pruriens*. 1. Initiation and maintenance of cell suspension cultures of *M. pruriens* and identification of L-Dopa. Plant Cell Tissue and Organ Culture (4) 61-73

Hussain G, Manyam BV (1997). *M. pruriens* proves more effective than L-Dopa in Parkinson's disease animal model. Phytotherapy Research (11) 419-423

Ibrahim MM, Hussein RM (2006). Variability, heritability and genetic advance in some genotypes of roselle (*Hibiscus sabdariffa* L.). World Journal of Agriculture Science (2) 340- 345

Inamdar S, Joshi S, Jadhav J, Bapat V (2012). Innovative use of intact seeds of *M. monosperma* wight for improved yield of L-Dopa. Natural Products and Bioprospect (2) 16-20

Ingle PK (2003). L-Dopa bearing plants. Natural Product Radiance (2) 126–133

Iyayi A, Egharevba JI (1998). Biochemical evaluation of seeds of an underutilized legume (*M. utilis*). Nigerian Journal of Animal Production (25) 40-45

Jaheer M, Chopra R, Kunder KR, Bhat D, Rashmi KV, Sathyanarayana N (2015). Cytogenetic and ITS-*psbA-trnH* sequence analysis for phylogenetic inference in *Mucuna* sp. of India. Tropical Plant Biology (8) 108–116

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) 141-152

Janardhanan K, Lakshmanan KK (1985). Studies on the pulse *M. utilis*: Chemical composition and antinutritional factors. Journal of Food Science and Technology (22) 369-371

Jayanand B, Sudarsanam G, Sharma K (2003). An efficient protocol for regeneration of whole plants of chickpea (*Cicer arietinum* L.) by using meristem explants derived from *in vitro* germinated seedlings. In Vitro Cellular & Developmental Biology - Plant (39) 171-179

Jenczewski E, Prosperi JM, Ronfort J (1999). Differentiation between natural and cultivated populations of *Medicago sativa* (Leguminosae) from Spain: Analysis with random amplified polymorphic DNA (RAPD) markers and comparison to allozymes. Molecular Ecology (8) 1317–1330

Jianab S, Shib S (2009). Genetic variation in *Heritiera littoralis* (Malvaceae) from east and south Asia revealed by AFLP markers. Plant Biosystems (143) 56-62

Johnson HW, Robinson HF, Comstock LE (1955a). Genotypic and phenotypic correlation in soybean and their implications in selection. Agronomy Journal (47) 447-483

Johnson HW, Robinson HF, Comstock RE (1955b). Estimation of genetic and environmental variation in soybeans. Journal of Agronomy (47) 314–318

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

Joshi A, Agrawal RC, Chawla HS (2011). Assessment of distinctiveness, uniformity and stability of indigenous aromatic rice (*Oryza sativa*) varieties based on morphological descriptors. Indian Journal of Agricultural Sciences 81 (7) 595-601

Joshi LD, Pant MC (1970). Hypoglycaemic effect of *Glycine soja, Dolichos biflorus* and *M. pruriens* seed diets in albino rats. Indian Journal of Pharmacology (2) 29–35

Joshi SP, Prabhakar K, Ranjekar PK, Gupta VS (2011). Molecular markers in plant genome analysis. Current Science 77 (2) 230-240

Kahani F, Hittalmani S (2015). Genetic analysis and traits association in F2 intervarietal populations in rice under aerobic condition. Rice research (3) 3-7.

Kala BK, Mohan VR (2010). Nutritional and antinutritional potential of three accessions of itching bean (*M. pruriens* (L.) DC. var. *pruriens*): an under-utilized tribal pulse. International Journal of Food Science and Nutrition (61) 497-511

Kalidass C, Mohan VR (2010). Genetic diversity of (velvet bean) *M. pruriens* (L.) DC. from southern western ghats, India. International Journal of Biological Technology (1) 31–37

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) 279-293

Kamble SY, Patil SR, Sawant PS, Sangita Sawant, Pawar SG, Singh EA (2009). Studies on plants used in traditional medicine by *Bhilla* tribe of Maharashtra. Indian Journal of Traditional Knowledge 9 (3) 591-598

Kanjilal UK, Das A, DC RN (1938). Flora of Assam. Bishen Singh and Mahendra Pal Singh, Dehra Dun, India

Kardolus JP, Van-Eck HJ, Van den Berg RG (1998). The potential of AFLPs in biosystematics a first application in *Solanum* taxonomy (*Solanaceae*). Plant Systematics and Evolution (210) 87–103

Kari AJ, Conn PJ, Niswender CM (2009). Glutamate receptors as therapeutic targets for Parkinson's disease. CNS and Neurological Disorders-Drug Targets (8) 475-491

Karobath M, Diaz J, Huttunen M (1971). The effect of L-Dopa on the concentrations of tryptophan, tyrosine and serotonin in the rat brain. European Journal of Pharmacology (14) 393–396

Karp A, Edwards KJ (1995). Molecular techniques in the analysis of the extent and distribution of genetic diversity. In: Molecular genetics techniques for plant genetic resources. Report of an IPGRI workshop, Rome, Italy 9-11

Kasso M, Balakrishnan M (2013). *Ex situ* conservation of biodiversity with particular emphasis to Ethiopia. International Scholarly Research Notices (2) 1-11

Kasture S, Mohan M, Kasture V (2013). *M. pruriens* seeds in the treatment of Parkinson's disease: Pharmacological review. Oriental Pharmacy and Experimental Medicine 1-11

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

Kavitha C, Vadivel E (2005). *In vitro* production of L- Dopa from *M. pruriens* (L.) DC. Biochemical and Cellular Archives 5 (2) 161-168

Kebriaee-zadeh A (2003). Overview of national drug policy of Iran. Iranian Journal of Pharmaceutical Research (2) 1–2

Kellner A (2015). Genome sequence of *Catharanthus roseus.* http:// medicinalplantgenomics.msu.edu

Kellner FJ, Kim BJ, Clavijo JP, Hamilton KL, Childs (2015). Genome-guided investigation of plant natural product biosynthesis. Plant Journal (82) 680–692

Keng L, Leng W (2004). Plant regeneration from stem nodal segments of *Orthosiphon stamineus* benth., a medicinal plant with diuretic activity. In Vitro Cellular and Developmental Biology-Plant (40) 115-118

Keurentjes JJB, Fu J Ric de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006). The genetics of plant metabolism. Nature Genetics (38) 842–849

Khatun T, Mohamed M, Hanafi MRY, Wong MY (2015). Genetic variation, heritability and diversity analysis of upland rice (*Oryza sativa* L.) genotypes based on quantitative traits. BioMed Research International 1-7

Khor SP, Hsu A (2007). The pharmacokinetics and pharmacodynamics of levodopa in the treatment of Parkinson's disease. Current Clinical Pharmacology (2) 234–243

Kidd PM (1997). Glutathione: Systemic protectant against oxidative and free radical damage. Alternative Medicine Review (2) 155–176

Kirmayer LJ, Simpson C, Cargo M (2003). Healing traditions: culture community and mental health promotion with Canadian aboriginal peoples. Australasian Psychiatry (11) 15–23

Kliebenstein DJ, Pedersen D, Barker B, Mitchell-Olds T (2002). Comparative quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. Genetics (161) 325–332

Kramer CY (1956). Extension of multiple range tests to group means with unequal number of replications. Biometrics (12) 307-310

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

Krishnan Marg KS (2003). L-Dopa bearing plants. Natural Product Radiance 2 (3) 126-133

Kuklin AI, Conger BV (1995). Catecholamines in plants. Journal of Plant Growth Regulator (14) 91

Kulkarni RN, Baskaran K, Chandrashekra RS, Khanuja SPS, Darokar MP, Shasany AK, Uniyal GC, Gupta MM, Kumar S (1997). Improved varieties and genetic research in medicinal and aromatic plants (MAPs). In: Proceedings of second national interactive meet on medicinal & aromatic plants 53-61

Kulkarni RN, Baskaran K, Jhang T (2016). Breeding medicinal plant, periwinkle [*Catharanthus roseus* (L) G. Don]: A review. Plant Genetic Resources: Charecterization and Utilization 14 (4) 283–302

Kulma A, Szopa J (2007). Catecholamines are active compounds in plants. Plant Science (172) 433–440

Kumar Anil (2016). Assessment of genetic diversity in selected accessions of *M. pruriens* using various morphometric traits. Journal of Science and Technology Research (6) 1–3

Kumar B, Kumar MS, Annupurna K, Maheshwari DK (2006). Genetic diversity of plant growth promoting rhizobia isolated from a medicinal legume *M. pruriens* Linn. Current Science (91) 1524-1529

Kumar J, Gupta P (2008). Molecular aroaches for improvement of medicinal and aromatic plants. Plant Biotechnology Reports 1-21

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 S, Bahl JR, Bansal RP, Garg SN, Naqvi AA, Khanuja SPS, Shasany AK, Singh V, Singh SC (1999). CSIR-CIMAP's Released Varieties

Kwang-Hoon K (1998). Purification and characterization of the tyrosinase isozymes of pine needles. Biochemistry and Molecular Biology International 45 (4) 717-724

Lal RK, Chandra R, Chauhan HS, Misra HO, Sangwan RS, Gupta MM, Verma RK, Singh AK, Yadav AK, Dhawan OP, Kalra A, Bahl JR, Singh HP, Gupta AK, Rai SK, Kumar B, Dubey B, Jhang T, Singh S, Singh VR, Pandey R (2011). Registration of new high yielding variety "CIMAP-Pratap" of *Withania somnifera* (L.) dunal suitable for cultivation in drought prone areas in India. Journal of Medicinal and Aromatic Plant Science 34 (4) 178-182

Lal RK, Chandra R, Mishra HO, Shankar K, Singh AK, Dhawan OP, Gupta AK, Kumar B, Gupta MM, Verma RK, Alam M, Singh SK Kalra A, Singh HP, Chauhan HS, Singh HN (2010). Registration of high yielding variety "CIMAP Ajay" morphine rich in concentrate of poy straw of Opium poppy (*Papaver somniferum*). Journal of Medicinal and Aromatic Plant Science 33(2) 204-208

Lal RK, Khanuja SPS, Chandra R, Misra HO, Singh A, Patra NK, Kumar B, Shasany AK, Chauhan HS, Darokar MP, Gupta AK, Bahl JR, Kalra A, Singh HN, Dhawan OP, Yadav A, Alam M, Gupta ML, Singh AK, Pandey R (2006). 'CIM-Sujal': High yielding variety of fennel *Foeniculum vulgare* (Miller) with better anethol content CIMAP Newsletter (2) 20

Lal RK, Shasany AK, Chandra R, Singh S, Sarkar S, Sattar A, Lahri R, Singh SC, Chanotiya CS, Dhawan OP, Gupta AK, Rai SK, Bahl JR, Kumar B, Jhang T, Dhawan SS, Zaim M, Gupta MM, Kalra A, Singh VR, Chauhan HS, Panday R, Singh S, Dubey B, Ujagir R, Yadav R (2014). Improved Varieties of Medicinal and Aromatic Plants CSIR-CIMAP's Contribution. CSIR-CIMAP's Released Varieties in 2015

Lal RK, Shasany AK, Singh S, Chandra R, Sarkar S, Lahri R, Yadav R, Dhawan OP, Sattar A, Gupta AK, Chanotiya CS, Dhawan SS, Gupta P, Bahl JR, Kumar B, Jhang T, Zaim M, Kalra A , Singh VR, Chauhan HS, Pandey R, Singh SC, Shukla S, Singh S (2015). Improved Varieties of Medicinal and Aromatic Plants CSIR-CIMAP's Contribution. CSIR-CIMAP's Released Varieties in 2015

Lall GK, Darby AC, Nystedt B, Macleod ET, Bishop RP, Welburn SC (2010). Amplified fragment length polymorphism (AFLP) analysis of closely related wild and captive tsetse fly (*Glossina morsitans morsitans*) populations. Parasite and Vectors 1–8

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

Lauritzen I, Blondeau N, Heurteaux C, Widmann C, Romey G, Lazdunski M (2000). Polyunsaturated fatty acids are potent neuroprotectors. Journal of European Molecular Biology Organization (19) 1784–1793

Lavania UC, Santosh Kumar Rai, Lavania Seshu, Surochita Basu, Basant Kumar Dubey, Ram jagir (2016). A novel autotetraploid clone 'CIMAP-Khus 40' of vetiver (*Vetiveria zizaniodes*) useful for carbon sequestration and soil conservation. Patent No US/26474/0832016

Lawal OS, Adebowale KO (2004). Effect of acetylation and succinylation on solubility profile, water absorption capacity, oil absorption capacity and emulsifying properties of *Mucuna* bean (*M. pruriens*) protein concentrate. Food/Nahrung 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 aroaches. World Journal of Agricultural Sciences (6) 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) 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

Lestari AD, Dewi W, Qosim WA, Rahardja M, Rostini N, Setiamihardja R (2006). Genetic variability and heritability of yield components and yield characters of 15 red pepper genotypes. Zuriat (17) 94-102

Lewis G, Schrire B, Mackinder B, Lock M (2005). Legumes of the World. The Royal Botanical Gardens, Kew

Li Y, Xu C, Lin X, (2014c). *De novo* assembly and characterization of the fruit transcriptome of Chinese Jujuba (*Zizipus jujuba* Mill) using 454 pyrosequencing and the development of novel trinucleotide SSR markers. PLoS One 9 (9) 1-17

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, Souza Jr CL, Souza AP (2002). Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.) Theoretical and Applied Genetics (104) 30-38

Liu Y, Dao Z, Liu Y, Long C (2009). Medicinal plants used by the Tibetan in Shangrila, Yunnan, China. Journal of Ethnobiology and Ethnomedicine 5 (15) 1-10

Lopez M, Pedro P, Garzon de la Mora W, Wysocka B, Maiz- tegui ME, Alzugaray H, Del Zotto MI, Borelli (2004). Quinolizidine alkaloids isolated from *Lupinus* species enhance insulin secretion. European Journal of Pharmacology (504) 139–142

Lorenzetti E, MacIsaac S, Anasson JT, Avang DVC, Buckles D (1998). The phytochemistry, toxicology and food potential of velvet bean (*Mucuna* Adans s. Fabaceae). In: Buckles D, Osiname O, Galliba MM, Galiano G, (Eds.). Cover crops of West Africa - contributing to sustainable agriculture. IDRC, Ottawa, Canada & IITA Ibadan, Nigeria 57

Lorenzetti F, MacIsaac S, Arnason JT, Awang DVC, Buckles D (2000). The phytochemistry toxicology and food potential of *M. pruriens* (*Mucuna* Adans s. Fabaceae). In: Buckles D, Eteka A, Osiname M, Galiba M, Galiano G, (Eds.) Cover crops in West Africa - Contributing to sustainable agriculture IDRC, IITA, Sasakawa Global, Otawa, Canada; Ibadan, Nigeria; Cotonou, Benin 1998 67-84

Lu Z, Nie G, Belton PS, Tang H, Zhao B (2006). Structure–activity relationship analysis of antioxidant ability and neuroprotective effect of gallic acid derivatives. Neurochemistry International (48) 263–274

Lubis SHA, Sastrapradja S, Lubis I, Sastrapradja D (1978). Genetic variation of *M. pruriens*. II: Inheritance of flower color. Annals of Bogorienses (4) 187-191

Lubis SHA, Sastrapradja S, Lubis I, Sastrapradja D (1979). Genetic variation of *M. pruriens* (L.) DC. III: Inheritance of pod hairs. Annals of Bogorienses (1) 1-10

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

Lyu J, Zhang S, Dong Y, He W, Zhang J, Deng X, Zhang Y, Li X, Li B, Huang W, Wan W, Yu Y, Li Q, Li J, Liu X, Wang B, Tao D, Zhang G, Wang J, Xu X, Hu F, Wang W (2013). Analysis of elite variety tag SNPs reveals an important allele in upland rice. Nature Communications (4) 1–9

Maesen LJG, Somaatmadja S (1992). Eries: PROSEA: Plant Resources of southeast Asia. 12 (1) 708

Mahajani SS, Doshi VJ, Parikh KM (1996). Bioavailability of L-Dopa from HP-200-a formulation of seed powder of *M. pruriens* (Bak): A pharmacokinetic and pharmacodynamic study. Phytotherapy Research 10 (3) 254-256

Maher P, Davis JB (1996). The role of monoamine metabolism in oxidative glutamate toxicity. Journal of Neuroscience (16) 6394–6401

Mahesh S (2015). Identification of elite germplasm, DNA fingerprinting and development of framework linkage map in Indian velvetbean (*M. pruriens*). Ph. D thesis, Visvesvaraya Technological University, Belgaum, Karnataka

Mahesh S, Leelambika M, Anithakumari AM, Sathyanarayana N (2016). Genetic maing 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 (2011b). The genotype x environment interaction and stability analysis for L-Dopa trait in velvet bean (*M. pruriens*) seeds. Indian Journal of Genetics 71 (3) 279–282

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

Mahima, Rahal A, Deb R, Latheef SK, Samad HA, Tiwari R, Verma AK, Kumar A, Dharma K (2012). Immunomodulatory and therapeutic potential of herbal, traditional/indigenous and ethnoveterinary medicines. Pakistan Journal of Biological Sciences (15) 754-774

Mallick SR, Jena RC, Samal KC (2012). Rapid *in vitro* multiplication of an endangered medicinal plant sarpgandha (*Rauwolfia serpentina*). American Journal of Plant Science (3) 437–442

Mamatha B, Shivananda TN, Siddaramaa R (2006). Nitrogen fixing ability of 13 genotypes of *M. pruriens*. In: Proceddings of 18th world congress of soil science, Philadelphia, Pennsylvania USA 9-15

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) 1297–1300

Mandel S, Weinreb O, Amit T, Youdim MBH (2005). Mechanism of neuroprotective action of the anti-Parkinson drug rasagiline and its derivatives. Brain Research Reviews (48) 379–387

Manyam BV, Dhanasekaran M, Hare TA (2004a). Effect of antiparkinson drug HP-200 (*M. pruriens*) on the central monoaminergic neurotransmitters. Phytotherapy research 18 (2) 97-101

Manyam BV, Dhanasekaran M, Hare TA (2004b). Neuroprotective effects of the antiparkinson drug *M. pruriens*. Phytotherapy Research (18) 706-712

Martin KP (2004). Rapid axillary bud proliferation and *ex vitro* rooting of *Eupatorium triplinerve*. Biologia Plantarum (47) 589-591

Martinez-Calvo J, Gisbert AD, Alamar MC, Hernandorena R, Romero C, Llacer G, Badenes ML (2008). Study of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.) by multivariate analysis. Genetic Resources and Crop Evolution 55 (5) 695-703

Mary JR, Janardhanan K (1992). Studies on chemical composition and antinutritional factors in three germplasm seed materials of the tribal pulse *M. pruriens* (L.) DC. Food Chemistry (43) 13-18

McCallum J (2007). Onion In: C Kole (ed.) Genome mapping and molecular breeding in plants. Springer (5) 331- 42

McCallum J, Andrew C, Meeghan PJ, Martin S, Ruth B, John S (2006). Genetic mapping of a major gene affecting onion bulb fructan content. Theoretical and Applied Genetics (112) 958-96

Mehta JC, Majumdar DN (1994). Indian medicinal plants - V. *Mucuna pruriens* bark (Papilionaceae). Indian Journal of Pharmacology (6) 92-94

Mewes S, Krüger H, Pank F (2008). Physiological, morphological, chemical and genomic diversities of different origins of thyme (*Thymus vulgaris* L.). Genetic Resources and Crop Evolution (55) 1303–1311

Mikulášková E, Fér T, Kučabová V (2012). The effect of different DNA isolation protocols and AFLP fingerprinting optimizations on error rate estimates in the bryophyte *Campylopus introflexus*. Lindbergia (35) 7–17 2012

Miller ER (1920). Dihydroxyphenylalanine, a constituent of the velvet bean. Journal of Biological Chemistry (44) 481-486

Misra A, Shasany AK, Shukla AK, Darokar MP, Singh SC, Sundaresan V, Singh J, Bagchi GD, Jain SP, Saikia D, Khanuja SP (2010). AFLP markers for identification of *Swertia* species (Gentianaceae). Genetics and Molecular Research (9) 1535–1544

Misra HO, Khanuja SPS, Lal RK, Singh A, Gupta MM, Darokar MP, Shasany AK, Jain N, Patra DD, Verma RK, Tripathi RS, Singh M, Dhawan OP, Singh AK, Bansal RP (2003). High yielding variety of kalmegh (*Andrographis paniculata*) 'CIM-Megha'. Journal of Medicinal and Aromatic Plant science 27 (3) 525-527

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

Misra L, Wagner H (2007). Extraction of bioactive principles from *M. pruriens* seeds. Indian Journal of Biochemistry and Biophysics (44) 56–60

Modi KP, Patel NM, Goyal RK (2008). Estimation of L-Dopa from *M. pruriens* Linn. and formulations containing *M. pruriens* by HPTLC method. Chemistry and Pharmaceutical Bulletin 56 (3) 357-359

Mohan VR, Janardhanan K (1995). Chemical analysis and nutritional assessment of lesser known pulses of the genus *Mucuna*. Food chemistry (52) 275-280

Moor N, Klipcan L, Safro MK (2011). Bacterial and eukaryotic phenylalanyl-tRNA synthetases catalyze misaminoacylation of $tRNA^{Phe}$ with 3,4-dihydroxy-L-phenylalanine. Chemistry and Biology (18) 1221–1229

Moura DJ, Richter MF, Boeira JM, Henriques JAP, Saffi J (2007). Antioxidant properties of β-carboline alkaloids are related to their antimutagenic and antigenotoxic activities. Mutagenesis (22) 293–302

Mueller UG, Wolfenbarger LL (1999). AFLP genotyping and fingerprinting. Trends in Ecology and Evolution (14) 389–394

Muller T, Welnic J, Muhlack S (2007). Acute levodopa administration reduces cortisol release in patients with Parkinson's disease. Journal of Neural Transmission 114 (3) 347-350

Mungikar R, Khandalkar M, Jagtap S (2018). Bio-resources market chain linkages and an opinion for its sustainable utilization in India. Advances in Complementary and Alternative Medicine 1 (3) 1-2

Muñoz A, Li Q, Gardoni F, Marcello E, Qin C, Carlsson T, Kirik D, Di Luca M, Björklund A, Bezard E, Carta M (2008). Combined 5- HT1A and 5-HT1B receptor agonists for the treatment of L-Dopa- induced dyskinesia. Brain (131) 3380–3394

Murashige T, Skoog F (1962). A revised medium for rapid growth and bio assay with tobacco tissue culture. Physiology Plant (15) 473–497

Nadkarni AK (1982). The Indian materia medica. Popular Prakashan Private Limited, Bombay, India 250

Nagatsu I, Sudo Y, Nagatsu T (1972). Tyrosine hydroxylation in the banana plant. Enxymologia (43) 25- 31

Nalawade SM, Sagare AP, Lee CY, Kao CL, Tsay HS (2003). Studies on tissue culture of Chinese medicinal plant resources in Taiwan and their sustainable utilization. Botanical Bulletin of Academia Sinica (44) 79–98

Nei M (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences USA (70) 3321–3323

Nei M (1987). Molecular evolutionary genetics. Columbia University Press, New York 75 (3)

Nishihara E, Parvez MM, Araya H, Fujji Y (2004). Germination growth response of different plant species to the allelochemical L-Dopa. Plant Growth Regulator (42) 18-19

Obata T (2003). Phytic acid suppresses 1-methyl-4-phenylpyridinium ion-induced hydroxyl radical generation in rat striatum. Brain Research (978) 241–244

Onyishi GC, Ngwuta AA, Ekpe II, Anaele CO (2015). Linking pod and seed yield with other agronomic traits in *M. sloanei* "Ukpo." Journal of Global Biosciences (4) 1151–1155

Oudhia P (2002). Kapikachu or cowhage (*M. pruriens*). Crop Fact Sheet. Version of 5-9

Ouedraogo B (2008). Les Determinants De L'Internsification Du Volume De L'Epargne Dans Le Systeme Financier Decentralise Au Burkina Faso: Cas Des Caisses Populaires De Oragadougou Revue Tiers-Monde Frmand Colin (4) 901-926

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

Pank F (2010). Conventional breeding of medicinal and aromatic plants - fundamentals and examples. Acta Horticulturae (860) 135-146

Papa R, Gepts P (2003). Assymetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolous vulgaris* L.) from Mesoamerica. Theoretical and Applied Genetics (106) 239-250

Patel NL, Patel AB, Vikani MK V, Shah RJ (2016). Analytical method development and validation of aphrodisiac herbal formulations. Indo-American Journal of Pharmcology Research 6 (4) 5096-5102

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) 207–217

Patil VS, Sadashiv NG, Shetty RK, Bindu H, Ganapathi M (2018). Evaluation of cowhage (*M. pruriens* L.) genotypes for growth and flowering characters in areca nut plantation under hill zone of Karnataka. Legume Research-An International Journal 4(1) 48-52

Patrick KC, Carola G, Bernhard M, France G (2012). Automated masking of AFLP markers improves reliability of phylogenetic analyses. PLoS One 7 (11) e49119

Pavlícek A, Hrdá S, Flegr J (1999). Free-Tree--freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus *Frenkelia*. Folia Biologica (45) 97–9

Peakall R, Smouse PE (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - An update. Bioinformatics (28) 2537–2539

Pecina-Quintero V, Anaya-Lopez JL, Nunez-Colin CA, Zamarripa-Colmenero A, Montes-Garcia N, Solís-Bonilla JL, Aguilar-Rangel MR (2013). Assessing the genetic diversity of castor bean from Chiapas, México using SSR and AFLP markers. Industrial Crops and Products (41) 134–143

Pieris N, Jansz ER, Dharmadasa HM (1980). Studies on *Mucuna* species of Sri Lanka. 1. The L-Dopa content of seeds. Journal of Natural Science (8) 35-40

Piper CV, Tracy SM (1910). Type of *Stizolobium pachylobium* S. Department of Agriculture Bureau of Plant Industry Bulletin (179) 18–19

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for gemplasm analysis. Molecular Breeding (2) 225-238

Prakash D, Gupta C (2011). Role of phytoestrogens as nutraceuticals in human health. Pharmacology online (1) 510– 523

Prakash D, Tewari SK (1999). Variation on L-Dopa content in *Mucuna* species Journal of Medicinal and Aromatic Plant Science (21) 343-346

Prevost A, Wilkinson MJ (1999). A new system of comparing PCR primers alied to ISSR fingerprinting of potato cultivars. Theoretical and Applied Genetics (98) 107–112

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

Pugalenthi M, Vadivel V (2007a). Agro biodiversity of 11 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) 1117–1124

Pugalenthi M, Vadivel V (2007b). L-Dopa (L-3, 4-Dihydroxyphenylalanine): A non-protein toxic amino acid in *M. pruriens* seeds. Food (1) 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) 201-218

Pulikkalpura H, Kurup R, Mathew PJ, Baby S (2015). Levodopa in *M. pruriens* and its degradation. Scientific Reports (5) 1–9

Pytliak M, Vargová V, Mechírová V, Felšöci M (2011). Serotonin receptors–from molecular biology to clinical applications. Physiology Research (60) 15–25

Qasim WA, Rachmadi M, Hamdani JS, Nuri I (2013). Phenotypic performance, variability and heritability of 32 high yielding red pepper genotypes. Indonesian Journal of Agronomy (41) 140–146

Quik M, Huang LZ, Parameswaran N, Bordia T, Campos C, Perez XA (2009). Multiple roles for nicotine in Parkinson's disease. Biochemistry and Pharmacology (78) 677–685

Raaman N, Mummoorthy C, Swaminathan A (2013). Micropropagation of *M. pruriens* (L.) DC. and determination of tyrosinase. Medicinal Plants (5) 179–186

Raghavendra S, Kumar V, Ramesh CK, Khan MHM (2012). Enhanced production of L-Dopa in cell cultures of *M. pruriens* L. and *M. prurita* H. Natural Product Research (26) 792–801

Rai PP, Saidu M (1977). Characterization of L-Dopa in seeds of *M. sloanei.* Current Science 46 (22) 778

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

Raina AP, Khatri R (2011). Quantitative determination of L-Dopa in seeds of *M. pruriens* germplasm by High Performance Thin Layer Chromatography. Indian Journal of Pharmacological Science (73) 456–459

Raina AP, Misra RC (2018). Chemical evaluation of *Mucuna* species for L-Dopa content-An anti-Parkinson's drug yielding medicinal plant from India. Indian Journal of Traditional Knowledge (17) 148–154

Rajaram N, Janardhanan K (1991). The biochemical composition and nutritional potential of the tribal pulse *M. gigantea* (Wild) DC. Plant Foods for Human Nutrition (41) 45-51

Rajore S, Sardana J, Batra A (2002). *In vitro* cloning of *Jatropha curcus* (L.) Journal of Plant Biology 29 (2) 195-198

Ramya KB, Thaakur S (2007). Herbs containing L- Dopa: An update. Ancient Science of Life 27 (1) 50-55

Rani V, Raina SN (2000). Genetic fidelity of organized meristem-derived micropropagated plants: A critical rearaisal. In Vitro Cellular & Developmental Biology-Plant (36) 319-330

Rao BRR, Rajput DK (2010). Global Scenario of Medicinal Plants. National Conference on Conservation of Medicinal Plants. CMP-HPU, Hyderabad

Rathod BG, Patel NM (2014). Development of validated RP-HPLC method for the estimation of L-Dopa from *M. pruriens* its extracts and in Aphrodisiac formulation. International Journal of Pharmacology and Science Research (5) 508–513

Riahi G, Morissette M, Parent M, Di-Paolo T (2011). Brain 5-HT2A receptors in MPTP monkeys and levodopa-induced dyskinesias. European Journal of Neuroscience (33) 1823–1831

Roldan-Ruiz I, Dendauw J, Bockstaele V, Depicker A, Loose M (2000). AFLP markers reveal high polymorphic rates in rye grasses (*Lolium* s.). Molecular Breedin*g* (6) 125–134

Ronning CM, Stommel JR, Kowalski SP, Sanford LL, Kobayashi RS, Pineada O (1999). Identification of molecular markers associated with leptine production in a population of *Solanum chacoense* Bitter. Bitter. Theoretical and Applied Genetics (98) 39–46

Rout GR, Samantaray S, Das P (2000a). *In vitro* manipulation and propagation of medicinal plants. Biotechnology Advances (18) 91120

Ruscher K, Rzeczinski S, Thein E, Freyer D, Victorov IV, Lam TT, Dirnagl U (2007). Neuroprotective effects of the beta-carboline abecarnil studied in cultured cortical neurons and organotypic retinal cultures. Neuropharmacology (52) 1488–1495

Russell RJ, Fuller JD, Macaulay M, Hats BG, Jahoor A, Powell W, Waugh R (1997). Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics (95) 714-722

Sahi CJ, Foss P, Rosazza J, Lembagar M (1969). Microbial synthesis of L,3-4-dihydroxy phenylalanine. Journal of the American Chemical Society (91) 6204-6207

Saikia SP, Bhau BS, Rabha A (2009). Study of accession source variation in morpho-physiological parameters and growth performance of *Jatropha curcas* Linn. Current Science (96) 1631–1636

Sameri MJ, Sarkaki A, Farbood Y, Mansouri SM (2011). Motor disorders and impaired electrical power of pallidal EEG improved by gallic acid in animal model of Parkinson's disease. Pakistan Journal of Biological Science (15) 1109–1116

Sanger RS, Kalpana S, Garg SK (2011). Biotechnological approaches for high sugarcane yield. Plant Sciences Feed 1 (7) 101-111

Sanjappa M, (1992). Legumes of India (2nd ed). Bishen Singh Mahendra Pal Singh, Dehra Dun, India

Sarhadi WA, Ookawa T, Yoshihashi T, Madadi AK, Yosofzai W, Oikawa Y, Hirata Y (2009). Characterization of aroma and agronomic traits in Afghan native rice cultivars. Plant Production Sciences 12 (1) 63-69

SAS Institute Inc (2008). JMP statistics and graphics guide version 8.0 SAS Institute Inc

SAS Institute Inc (2012). Using JMP 10 Cary NC: SAS Institute Inc

Saslis-Lagoudakis C, Williamson H, Savolainen EMV, Hawkins JA (2011). Cross-cultural comparison of three medicinal floras and implications for bioprospecting strategies. Journal of Ethnopharmacology (135) 476–487

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

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

Sathyanarayana N, Bharath TN, Vikas PB, Rajesha R (2008). *In vitro* clonal propagation of *M. pruriens* var. *utilis* and its evaluation of genetic stability through RAPD markers. African Journal of Biotechnology (7) 973–980

Sathyanarayana N, Leelambika M, Mahesh S, Jaheer M (2011). AFLP assessment of genetic diversity among Indian *M.* accessions. Physiology and Molecular Biology of Plants (17) 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 morpho-agronomical characters. Tropical and Subtropical Agroecosystems (15) 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 Charecterization and Utilization (14) 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 *M. pruriens*: *De novo* transcriptome assembly annotation identification and validation of EST-SSR markers. BMC Genomics (18) 1-18

Saxena A, Cramer CS (2013). Metabolomics : A potential tool for breeding nutraceutical vegetables. Advance in Crop Science and Technology (1) 2–5

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology (24) 447–454

Schintu N, Frau L, Ibba M, Caboni P, Garau A, Carboni E, Carta AR (2009). PPAR-gamma-mediated neuroprotection in a chronic mouse model of Parkinson's disease. European Journal of Neuroscience 29 (5) 954–963

Sen S, Chakraborty R (2016). Revival modernization and integration of Indian traditional herbal medicine in clinical practice: Importance, challenges and future. Journal of Traditional and Complementary Medicine (7) 234–244

Serrano-Dueñas M, Cardozo-Pelaez F, Sánchez-Ramos JR (2001). Effects of *Banisteriopsis caapi* extract on Parkinson's disease. Scientific Review of Alternative Medicine (5) 129–134

Sharma G, Srivastava AK, Prakash D (2011). Phytochemicals of nutraceutical importance: Their role in health and diseases. Pharmacology online (2) 408–427

Sharma JR, Lal RK, Misra HO, Gupta MM (1987). Improved varieties and genetic research in medicinal and aromatic plants (MAPs). In: Proceedings of second national interactive meet on medicinal and aromatic plants. (2005) 53-61

Sharma K, Semwal CS, Uniyal SP (2010). Genetic variability and character association analysis in bell pepper (*Capsicum annuum* L.). Journal of Horticulture and Forestry (2) 58-65

Shavrukov Y, Kurishbayev A, Jatayev S, Shvidchenko V, Zotova L, Koekemoer F, Groot S, De-Soole K, Langridge P (2017). Early flowering as a drought escape mechanism in plants : How can it aid wheat production ? Frontiers in Plant Science 17 (8) 1-8

Shepherd M, Chaparro JX, Teasdale R (1999). Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. Theoretical and Applied Genetics 99 (7-8) 1207–1215

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

Shukla KK, Mahdi AA, Ahmad MK, Jaiswar SP, Shankwar SN, Tiwari SC (2010). *M. pruriens* reduces stress and improves the quality of semen in infertile men. Evidence Based Complementry and Alternative Medicine 7 (1) 137–144

Shults CW, Oakes D, Kieburtz K, Beal MF, Haas R, Plumb S, Juncos JL, Nutt J, Shoulson I, Carter J, Kompoliti K, Perlmutter JS, Reich S, Stern M, Watts RL, Kurlan R, Molho E, Harrison M, Lew M (2002). Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Archives of Neurology (59) 1541– 1550

Siddhuraju P, Becker K (2001a). Effect of various indigenous processing methods on the ∝- galactoside and mono- and disaccharide content of an Indian tribal pulse, *M. 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) 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 underutilized tropical legume *M. pruriens* var. *utilis.* Journal of Agriculture and Food Chemistry (48) 6048-6060

Simopoulos AP (2004). Omega-3 fatty acids and antioxidants in edible wild plants. Biology Research 37(2) 263-277

Singh U (1993). Protein quality of pigeon pea (*Cajanus cajan* [L.] Mill sp.) as influenced by seed polyphenols and cooking process. Plant Foods for Human Nutrition (43) 171- 179

Singh AK, Malik SS, Tomar YS (2008). Studies on the medicinal compound L-Dopa in *M. pruriens* (Bak). Indian Journal of Plant Genetic Resources 21(3) 217-220

Singh AP, Sarkar S, Tripathi M, Rajender S (2013). *M. pruriens* and its major constituent L-Dopa recover spermatogenic loss by combating ROS loss of mitochondrial membrane potential and apoptosis. PLoS ONE (8) 1–10

Singh DN (2001). Heritability and genetic advances in linseed (*L. usiataissimum* L.). Journal of Research Birsa Agricultural University 13 (1) 73–74

Singh K, Bansal C, Koundal KR (2010). SSR and AFLP based genetic diversity of soybean germplasm differing in photoperiod sensitivity. Genetics and Molecular Biology 33(2) 319-324

Singh NP, Singh KP, Singh DK (2002). Flora of Mizoram. BSI, Kolkata

Singh RK, Bhatia VS, Bhat KV, Mohapatra T, Nagendra K, Ramya KB, Thaakur S (2007). Herbs containing L- Dopa: An update. Ancient Science of Life 27(1) 50-55

Singh SK, Dhawan SS (2017). Analyzing trichomes and spatio-temporal expression of a cysteine protease gene Mucunain in *M. pruriens* L. (DC.). Protoplasma (255) 575–584

Singh SP, Sharma JR, Mishra HO, Lal RK, Gupta MM, Tajuddin (1997). Development of new variety Sona of Senna (*Cassia angustifolia*). Journal of Medicinal and Aromatic Plant Science (19) 446

Singh SP, Tomar VKS, Singh AK, Singh JP, Sundaresan V, Singh VR, Singh HP (2007). High yielding variety of Shatawar (*Asparagus racemosus*) with high root yield and high saponin content. CIMAP annual report 2008 113-114

Siril EA, Dhar U (1997). Micropropagation of mature Chinese tallow tree (*Sapium sebiferum* Roxb.). Plant Cell Reports 16 (9) 637-640

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

Smith JSC, Smith OS (1992). Measurement of genetic diversity among maize hybrids: A comparison of isozymic RFLP pedigree and heterosis data. Maydica 37 (5) 53-60

Soares AR, Marchiosi R, Siqueira-Soares R, de C, Barbosa-de-Lima R, Dantas-do-Santos W, Ferrarese-Filho O (2014). The role of L-Dopa in plants. Plant Signaling and Behavior (9) 1–7

Sofowora A (1982). Medicinal plants in traditional medicine in West Africa (1st Edn.), John Wiley and Sons, London

Southon IW, Bisby FA, Buckingham J, Harborne JB (1994). Phytochemical dictionary of the leguminosae. Volume 1. Plants and their constituents. Volume 2. Chemical constituent. Chapman and Hall, London

Sreelathakumary I, Rajamony L (2004). Variability, heritability and genetic advance in chilli (*Capsicum annuum* L.). Journal of Tropical Agriculture (42) 35- 37

St-Laurent L, Livesey J, Arnason JT, Bruneau A (2002). Variation in L-Dopa concentration in accessions of *M. pruriens* (L.) DC. and in *M. brachycarpa*. Rech. In: Flores M, Eilittä M, Myhrman R, Carew LB, Carsky RJ (Eds.). Food and Feed from *Mucuna*: Current uses and the way forward. Proceedings of an International Workshop, CIDICCO, CIEPCA, World Hunger Research Center, Tegucigalpa,, Honduras 352- 375

Sudha M, Anusuya P, Mahadev NG, Karthikeyan A, Nagarajan P, Raveendran M, Balasubramanian P (2013). Molecular studies on mungbean (*Vigna radiate* (L.) Wilczek) and rice bean (*Vigna umbellata* (Thunb.) inter-specific hybridization for Mungbean yellow mosaic virus resistance and development of species-specific SCAR marker for rice bean. Arch. Phytopathol, Plant Protection (46) 503-517

Sudupak MA, Akkaya MS, Kence A (2004). Genetic relationships among perennial and annual *Cicer* species growing in Turkey assessed by AFLP fingerprinting. Theoretical and Applied Genetics (108) 937-944

Sujatha G, Jayabalan N, Ranjitha Kumari BD (2007). Rapid *in vitro* micropropagation of *Cicer arietinum* L. Horticulture Science 34 (1) 1–5

Suresh C, Ajay KS (2004). *In vitro* shoot regeneration from cotyledonary node explants of a multipurpose leguminous tree *Pterocarpus marsupium* Roxb. In Vitro Cellular & Developmental Biology - Plant (40) 167-170

Swiedrych A, Lorenc K, Skirycz A, Szopa J (2004). The catecholamine biosynthesis route in potato is affected by stress. Plant Physiology and Biochemistry 42 (7-8) 593-600

Syukur M, Rosidah S (2014). Estimation of genetic parameter for quantitative character of *Capsicum annum*. Journal of Tropical Crop Science 1 (1) 4-6

Syukur MS, Sujiprihati R, Yunianti (2012). Teknik Pemuliaan Tanaman Penebar Swadaya Jakarta

Tabrez S, Jabir NR, Shakil S (2012). A synopsis on the role of tyrosine hydroxylase in Parkinson's disease. CNS and Neurological Disorder-Drug Targets (11) 395–409

Tarawali G, Manyong VM, Carsky RJ, Vissoh PV, Osei-Bonsu P, Galiba M (1999). Adoption of improved fallows in West Africa: Lessons from *Mucuna* and *Stylo* case studies. Agroforestry Systems (47) 93-122

Tatikonda L, Suhas P, Kannan WS, Thakur N, Beerelli K, Sreedevi, Hoisington DA, Devi P, Varshney RK (2009). AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L.: A biofuel plant. Plant Science (176) 505-513

Teixeira AA, Rich EC, Szabo NJ (2003). Water extraction of L-Dopa form *Mucuna* bean. Tropical and Subtropical Agroecosystem (1) 159 – 171

Tesfa MBA, KB (2016). *In vitro* shoot multiplication of elite sugarcane (*Saccharum officinarum* L.) genotypes using liquid shake culture system. Journal of Biology 6 (1) 35–40

Tewary PK, Raghunath MK, Venkateshwarulu M, Sarkar A (1996). Genotypic difference in response to *in vitro* shoot development of mulberry (*Morus* spp.). Indian Journal of Sericulture 35 (2) 104-106

Tharakan B, Dhanasekaran M, Mize-Berge J, Manyam BV (2007). Anti-Parkinson botanical *M. pruriens* prevents levodopa induced plasmid and genomic DNA damage. Phytotherapy Research (21) 1124-1126

Thiel T, Michalek W, Varshney RK, Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics (106) 411- 422

Toklu F, Karaköy T, Hakli E, Bicer T, Brandolini A, Kilian B, Özkan H (2009). Genetic variation among lentil (*Lens culinaris* Medik) landraces from southeast Turkey. Journal of Plant Breeding (128) 178–186

Tukey JW (1953). The problem of multiple comparisons. Unpublished manuscript

Tyagi BR, Singh V, Singh H, Naqvi AA (1988). Improved varieties and genetic research in medicinal and aromatic plants (MAPs). In: Proceedings of Second National Interactive Meet on Medicinal and Aromatic Plants. CSIR-CIMAP, Lucknow 53-61

Umezawa H, Tobe H, Shibamoto N, Nakamura F, Nakamura K, Matsuzaki M, Takeuchi T (1975). Isolation of isoflavones inhibiting Dopa decarboxylase from fungi and streptomyces. The Journal of Antibiotics (28) 947–952

Upadhyay V, Tiwari AK, Sharma N, Joshi HM, Singh BP and Kalakoti BS (2012). Alication of high performance liquid chromatography to the determination and validation of levodopa in *M. pruriens* L. International Journal for Pharmaceutical Research Scholars. 1 (4) 135-140

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

Vachhani UD, Trivedi MN, Bajaj A, Shah CP, June A (2011). A HPTLC method for quantitative estimation of L-Dopa from *M. pruriens* in polyherbal aphrodisiac formulation. Research Journal of Pharmaceutical, Biological and Chemical Sciences (2) 389–396

Vadivel V, Biesalski HK (2012). Total phenolic content, *in vitro* antioxidant activity and type II diabetes relevant enzyme inhibition properties of methanolic extract of traditionally processed underutilized food legume *Acacia nilotica* (L.) Willd ex. Delile. International Food Research Journal 19 (2) 593-601

Venkatesha SC, Ramanjini PH, Gowda KN, Ganapathy M, Byre Gowda R, Ramachandra G, Girish V, Channamallikarjuna L, Shantala GTKS (2010). Genetic fingerprinting in dolichos bean using AFLP markers and morphological traits. International Journal of Biotechnology and Biochemistry 6 (3) 395-404

Vijayambika C, Jegadeesan M, Ganthi AS (2010). Comparative L-Dopa and antinutritional contents in seed materials of market samples of *M. pruriens* (L.) DC. International Journal of Research in Ayurveda and Pharmacy 1 (2) 480-483

Villiers SD, Emongor Q, Njeri R, Gwata E, Hoisington D, Njagi I, Silim S, Sharma S (2008). Evaluation of the shoot regeneration response in tissue culture of pigeonpea (*Cajanus cajan* [L.] Mill sp.) varieties adapted to eastern and southern Africa. African Journal of Biotechnology 7 (5) 587-590.

Vishwakarma KS, Mohammed SI, Chaudhari AR, Salunkhe NS, Maheshwari VL (2017). Micropropagation and *Agrobacterium rhizogenes* mediated transformation studies in *M. pruriens* (L.) DC. Indian Journal of Natural Product Resources (8) 172–178

Vos P, Hogers R, Bleeker M, Reijans M, Van-de-Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research (23) 4407–4414

Wang ZJ, Li GM, Tang WL, Yin M (2006). Neuroprotective effects of stearic acid against toxicity of oxygen/glucose deprivation or glutamate on rat cortical or hiocampal slices. Acta Pharmacologica Sinica 27 (2) 145–150

Watt G (1883). Economic Products of India. Superintendent of government printing. Calcutta, India. (6) 188993

Weerakoon SR, Jagathpriya Weerasena OVDS, Kelum P, Peiris D, Somaratne S (2010). Assessment of genetic variation within mustard (*Brassica juncea* (L.) Czern & Coss) germplasm in Sri Lanka using fluorescent-based amplified fragment length polymorphic DNA markers. International Journal of Biotechnology and Biochemistry (6) 757–768

Wichers HJ, Pras N, Huizing HJ (1989). *M. pruriens*: *In vitro* production of L-Dopa. In: Biotechnology in Agriculture and Forestry. Medicinal and Aromatic plants (Eds. YPS Bajaj), Springer- Verleag, Berlin, Heidelberg, Germany

Wichers HJ, Visser JF, Huizing HJ, Niesko P (1993). Occurrence of L -Dopa and dopamine in plants and cell cultures of *M. pruriens* and effects of 2,4-D and NaCl on the compounds. Plant Cell Tissue and Organ Culture (33) 259-264

Wichers HJ, Wijnsma R, Visser JF, Malingre M, Huizing HJ (1985). Production of L -Dopa by cell suspension cultures of *M. pruriens*. II. Effect of environmental parameters on the production of L- Dopa. Plant Cell Tissue and Organ Culture (4) 75-82

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research (18) 6231–6235

Williams JT, Ahmad Z (1999). Priorities of medicinal plants research and development in Pakistan. Medicinal and Aromatic Plants Program in Asia (MAPPA), New Delhi, India 1-64

Wilmot-Dear CM (1984). A revision of *Mucuna* (Leguminosae-Phaseoleae) in China and Japan. Kew Bulletin (39) 23-65

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

Wink M (2013). Evolution of secondary metabolites in legumes (Fabaceae). South African Journal of Botany (89) 164–175

XLSTAT (2010). Add in soft USA, New York Yamasaki S, Dillenburg LC (1999). Measurement of leaf relative water content in *A. angustifolia*. The Revista Brasileira de Fisiologia Vegetal (11) 69-75

Yadeta B, Belew D, Gebreselassie W, Marame F (2011). Variability heritability and genetic advance in hot pepper (*C. annuum* L.) genotypes in West Shoa Ethiopia. American-Eurasian Journal of Agricultural & Environmental Sciences (10) 587–592

Yang Y, Moore MJ, Brockington SF, (2015). Dissecting molecular evolution in the highly diverse plant clade caryophyllales using transcriptome sequencing. Molecular Biology and Evolution 32 (8) 2001-2014

Yasodha R, Sumathi R, Gurumurthi K (2004). Micropropagation for quality propagule production in plantation forestry. Indian Journal of Biotechnology (3) 159–170

Yeh, FC, Yang RC, Boyle TB, Ye Z, Mao JX (1997). POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada

Yencho GC, Kowalski SP, Kobayashi RS, Sinden SL, Bonier bale MW, Deahl KL (1998). QTL maing of foliar glycoalkaloid aglycones in *Solanum tuberosum* and *S. berthaultii* potato progenies: quantitative variation and plant secondary metabolism. Theoretical and Applied Genetics (97) 563–574

Zeevalk G, Manzino L, Sonsalla PK, Bernard LP (2007). Characterization of intracellular elevation of glutathione (GSH) with glutathione monoethyl ester and GSH in brain and neuronal cultures: relevance to Parkinson's disease. Experimental Neurology (203) 512–520

Zhang F, Gao Q, Khan G, Luo K, Chen S (2014). Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida* an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. Gene 553 (2) 90–97

Zhao D, Hamilton JP, Pham GM (2017). *De novo* genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. Giga Science (6) 1–7

Zietkiewicz E, Rafalski A, Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR) anchored polymerase chain reaction amplification. Genomics (20) 176-193

Zimnoch-Guzowska E, Marczewski W, Lebecka R, Flis B, Schäfer-Pregl R, Salamini F, Gebhardt C (2000). QTL analysis of new sources of resistence to *Erwinia carotovora* ssp. atroseptica in potato done by AFLP RFLP and resistence-gene-like markers. Crop Science 40 (4) 1156-1167

Plant Gene 15 (2018) 6–14

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plantgene

High levels of gene flow constraints population structure in *Mucuna pruriens* L. (DC.) of northeast India

Pankaj Kumar Tripathi[a], Satya Narayan Jena[b], Tikam Singh Rana[b], N. Sathyanarayana[a,*]

[a] Department of Botany, Sikkim University, Gangtok 737102, Sikkim, India
[b] CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow 226001, India

ARTICLE INFO

Keywords:
Mucuna pruriens
Northeast India
AFLP marker
Genetic diversity
Population structure

ABSTRACT

Medicinal legume *Mucuna pruriens* L. (DC.) is an ecologically and economically important species known worldwide for anti-Parkinson's drug L-Dopa. The plant is native to eastern India which includes parts of northeast India where it has long history of cultivation. Recently, with the discovery of novel medicinal properties and agronomic benefits, the prospects for its large-scale cultivation in the region is growing necessitating development of improved cultivars. However, little data is available on genetic diversity and population structure from its native range to use as resource base for future molecular breeding works and conservation planning. Here we describe first empirical study on genetic diversity and gene flow of *M.pruriens* from northeast India based on AFLP markers.

The high estimates of Jaccard's coefficient and Shannon's index reveal good genetic diversity supporting centre of origin theory. However, pair-wise estimates of genetic differentiation (F_{ST}, Nei's D) were low and AMOVA revealed no evidence for genetic structuring suggesting high levels of gene flow and absence of genetic drift or bottleneck events. Clustering based on UPGMA and STRUCTURE analyses were in conformity with these findings, but point to existence of two independent gene pools representing perhaps Assam plains and eastern Himalayas – the two disparate components of geographically diverse northeast India. The mechanisms underlying such variation besides implications on genetic diversity along with scope for future investigations are discussed.

1. Introduction

Mucuna pruriens (L.) DC. is a diploid tropical legume (2n = 2x = 22) classified within the phaseoloid clade of leguminosae. It is perhaps the most important domesticated species of the genus *Mucuna* grown world-wide in a range of environments including temperate conditions (Duke, 1981; Russell et al., 1992; CABI, 2015) as food, feed, fodder and cover crop (Siddhuraju and Becker, 2001; Pulikkalpura et al., 2015). The seeds contain L-Dopa (L-3,4 dihydroxy phenylalanine) - a non-protein amino acid well-known for the symptomatic treatment of Parkinson's disease. The species comprise three botanical varieties of which wild var. *pruriens* and var. *hirsuta* (itching beans) shows orange or silvery-grey trichomes which induce highly itching skin reactions upon contact due to presence of proteinaceous substances, mucunain and serotonin (Agharkar, 1991). They often colonize open fields, hedgerows, by woodland margins or on lake shores (Duke, 1981; Lampariello et al., 2012). The cultivated var. *utilis* (velvet bean) bears smooth non-itching trichomes and is mostly grown as cover-crop for superior biomass, nitrogen fixing ability and in homestead gardens for edible pods and nutritious seeds (Haridas et al., 2013).

The centre of origin of a crop plant is normally associated with broad genetic diversity contributed by wild and weedy relatives (Harlan, 1971; Guo et al., 2014). *M.pruriens* is thought to be native of eastern India or southern China (Burkill, 1966; Duke, 1981; Wilmot-Dear, 1984) which includes parts of northeast India; but concrete data to support this theory is not yet available and wild ancestor of the species is yet to be discovered. The region however showcase abundant distribution of wild variety represented by var. *pruriens* and there is a long history of cultivation of velvet bean (var. *utilis*) in the region by several indigenous groups of northeast India such as Khasi, Naga, Kuki, Jaintia, Chakma and Mizo etc. who have used it as a minor food for centuries (Arora, 1991). Recently, with the discovery of highly important medicinal properties, the prospects for its large-scale cultivation in the region are growing. As a result, widening the genetic resource base to develop improved cultivars with better agronomic and nutritional potential is gaining momentum (Sathyanarayana et al., 2017). In this background, the present investigation was carried out to analyze genetic diversity and population structure of *M. pruriens* from

* Corresponding author at: Department of Botany, Sikkim University, 6th Mile, Samdur, P.O. Tadong, 737102 Gangtok, Sikkim, India.
E-mail address: nsathyanarayana@cus.ac.in (N. Sathyanarayana).

https://doi.org/10.1016/j.plgene.2018.05.005
Received 1 February 2018; Received in revised form 9 May 2018; Accepted 24 May 2018
Available online 25 May 2018

Fig. 1. Collection locations of the germplasm used in the study (Note: Points in the map indicate collection locality). Map Courtesy: www.mapsofindia.com

this region to (a) examine the northeast Indian origin theory of *M.pruriens* and (b) generate baseline data for future molecular breeding works. AFLPs were particularly chosen as it covers large area of the genome (Karp and Edwards, 1995) and are extremely proficient in revealing diversity at the infra-specific level (Karp and Edwards, 1995; Capo-chichi et al., 2001).

2. Materials and methods

2.1. Germplasm collection and genomic DNA isolation

A totally 50 *M. pruriens* accessions representing two botanical varieties viz., var. *pruriens* and var. *utilis* were collected from different parts of northeast India during Feb–April 2015 and 2016. Subsequently, voucher specimens were deposited in the herbarium of the Botany department, Sikkim University, Gangtok. The collection areas covered and the list of accessions acquired are depicted in Fig. 1 and Table 1, respectively. The seedlings were raised in triplicate from all the accessions in the Botanical Garden of Sikkim University during July–August 2016 and the total genomic DNA was isolated from pooled young leaf tissues from each accession following CTAB method (Doyle and Doyle, 1990). The quality and quantity of DNA were assessed using a NanoDrop 2000 spectrophotometer.

2.2. AFLP data generation

Twenty two AFLP primers (Tables 2a and 2b) were tested as per the method described by Vos et al. (1995) with minor modifications. The genomic DNA (400–500 ng) was digested with 5U *Eco*RI and 1U *Mse*I restriction endonucleases for 2 h at 37 °C. The *Eco*RI and *Mse*I adapter (AFLP® plant mapping kit, Applied Biosystems, Foster City, CA, USA) were ligated to digested DNA using 1U T4 DNA ligase (New England Biolabs, USA). The restricted–ligated DNA was then diluted by 20 fold with TE buffer (20 mMTris–HCL, 0.1 mM EDTA, pH 8.0) to perform pre-selective amplification. The pre-selective amplification was carried out using ready PCR master mix including Taq polymerase (Thermo Scientific, USA). The cycling profile included: 30 cycles of denaturation at 94 °C for 20 s, annealing at 56 °C for 30 s and extension at 72 °C for 2 min and at last hold at 60 °C for 30 min. The pre-selective amplification was confirmed by running PCR product on 1.5% agarose gel and amplification product was further diluted by 10-folds with TE buffer and used as a template for selective amplification. The selective amplification was performed in 10 µl reaction volume using selective primer pairs with amplification profile of denaturation at 94 °C for 20 s, annealing at 66 °C for 30 s decrease by 1 °C/cycle up to 56 °C, extension at 72 °C for 2 min for 10 cycles, followed by 94 °C for 20 s, 56 °C for 30 s, 72 °C for 2 min for 30 cycles followed by 60 °C for 30 min. Following selective amplification, post PCR multiplex sets were made based on fluorescence labelled primers. For post PCR multiplexing, 1 µl each of 6-FAM, JOE and NED labelled PCR products representing different selective primer pairs were combined with 7 µl of Hi-Di formamide

Table 1
List of *M. pruriens* accessions used in the study.

Sl. no.	Variety	Accession no.	Place of collection
1	*Mucuna pruriens* var. *pruriens*	500197WB	West Bengal
2	*Mucuna pruriens* var. *pruriens*	500199WB	West Bengal
3	*Mucuna pruriens* var. *utilis*	500203MN	Manipur
4	*Mucuna pruriens* var. *pruriens*	500204AS	Assam
5	*Mucuna pruriens* var. *pruriens*	500205AS	Assam
6	*Mucuna pruriens* var. *utilis*	500210MN	Manipur
7	*Mucuna pruriens* var. *utilis*	500211NL	Nagaland
8	*Mucuna pruriens* var. *pruriens*	500212AS	Assam
9	*Mucuna pruriens* var. *utilis*	500213MN	Manipur
10	*Mucuna pruriens* var. *utilis*	500215AR	Arunachal Pradesh
11	*Mucuna pruriens* var. *pruriens*	500216ML	Meghalaya
12	*Mucuna pruriens* var. *utilis*	500217MN	Manipur
13	*Mucuna pruriens* var. *pruriens*	500219TR	Tripura
14	*Mucuna pruriens* var. *pruriens*	500220AR	Arunachal Pradesh
15	*Mucuna pruriens* var. *pruriens*	500221AR	Arunachal Pradesh
16	*Mucuna pruriens* var. *pruriens*	500222AR	Arunachal Pradesh
17	*Mucuna pruriens* var. *pruriens*	500223AR	Arunachal Pradesh
18	*Mucuna pruriens* var. *pruriens*	500224AR	Arunachal Pradesh
19	*Mucuna pruriens* var. *pruriens*	500225AR	Arunachal Pradesh
20	*Mucuna pruriens* var. *pruriens*	500226AR	Arunachal Pradesh
21	*Mucuna pruriens* var. *pruriens*	500227AR	Arunachal Pradesh
22	*Mucuna pruriens* var. *pruriens*	500228AR	Arunachal Pradesh
23	*Mucuna pruriens* var. *pruriens*	500231AR	Arunachal Pradesh
24	*Mucuna pruriens* var. *pruriens*	500232AS	Assam
25	*Mucuna pruriens* var. *pruriens*	500233AS	Assam
26	*Mucuna pruriens* var. *pruriens*	500234MZ	Mizoram
27	*Mucuna pruriens* var. *pruriens*	500235TR	Tripura
28	*Mucuna pruriens* var. *pruriens*	500236AS	Assam
29	*Mucuna pruriens* var. *pruriens*	500237AS	Assam
30	*Mucuna pruriens* var. *pruriens*	500239MZ	Mizoram
31	*Mucuna pruriens* var. *pruriens*	500240SK	Sikkim
32	*Mucuna pruriens* var. *pruriens*	500241SK	Sikkim
33	*Mucuna pruriens* var. *pruriens*	500242SK	Sikkim
34	*Mucuna pruriens* var. *pruriens*	500243SK	Sikkim
35	*Mucuna pruriens* var. *pruriens*	500244SK	Sikkim
36	*Mucuna pruriens* var. *pruriens*	500245TR	Tripura
37	*Mucuna pruriens* var. *pruriens*	500246TR	Tripura
38	*Mucuna pruriens* var. *pruriens*	500248TR	Tripura
39	*Mucuna pruriens* var. *pruriens*	500249SK	Sikkim
40	*Mucuna pruriens* var. *utilis*	500255AR	Arunachal Pradesh
41	*Mucuna pruriens* var. *utilis*	500263NL	Nagaland
42	*Mucuna pruriens* var. *utilis*	500266NL	Nagaland
43	*Mucuna pruriens* var. *utilis*	500267NL	Nagaland
44	*Mucuna pruriens* var. *utilis*	500268NL	Nagaland
45	*Mucuna pruriens* var. *pruriens*	500269SK	Sikkim
46	*Mucuna pruriens* var. *pruriens*	500271SK	Sikkim
47	*Mucuna pruriens* var. *pruriens*	500274SK	Sikkim
48	*Mucuna pruriens* var. *pruriens*	500275ML	Meghalaya
49	*Mucuna pruriens* var. *pruriens*	500276ML	Meghalaya
50	*Mucuna pruriens* var. *utilis*	500277NL	Nagaland

MN: Manipur; NL: Nagaland; AS: Assam; AR: Arunachal Pradesh; ML: Meghalaya; SK: Sikkim; MZ: Mizoram; TR: Tripura; WB: West Bengal.

Table 2a
Details of pre-selective primers used for AFLP analysis.

Sl. no.	AFLP pre-selective primer combination	Pre-selective primer sequence (5′ to 3′)
1	E-A/M-C	GACTGCGTACCAATTC-A GATGAGTCCTGAGTAA-C

Table 2b
Details of selective primers used for AFLP analysis.

Sl. no.	AFLP selective primer combinations	Selective primer sequence (5′ to 3′)
1	E-ACT/M-CTA	GACTGCGTACCAATTC-ACT (FAM) GATGAGTCCTGAGTAA-CTA
2	E-AAC/M-CAG	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CAG
3	E-AAC/M-CTA	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CTA
4	E-AAC/M-CTC	GACTGCGTACCAATTC-AAC (NED) GATGAGTCCTGAGTAA-CTC
5	E-ACA/M-CTT	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CTT
6	E-AGG/M-CAG	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CAG
7	E-ACG/M-CAC	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAC
8	E-ACA/M-CAT	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CAT
9	E-AAG/M-CAC	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAC
10	E-ACA/M-CAA	GACTGCGTACCAATTC-ACA (FAM) GATGAGTCCTGAGTAA-CAA
11	E-ACG/M-CAA	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAA
12	E-AAG/M-CAT	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAT
13	E-AAG/M-CTG	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CTG
14	E-AAG/M-CTC	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CTC
15	E-AAG/M-CAA	GACTGCGTACCAATTC-AAG (JOE) GATGAGTCCTGAGTAA-CAA
16	E-ACT/M-CAA	GACTGCGTACCAATTC-ACT (FAM) GATGAGTCCTGAGTAA-CAA
17	E-ACG/M-CAG	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CAG
18	E-AGC/M-CTA	GACTGCGTACCAATTC-AGC (NED) GATGAGTCCTGAGTAA-CTA
19	E-ACG/M-CTG	GACTGCGTACCAATTC-ACG (JOE) GATGAGTCCTGAGTAA-CTG
20	E-AGG/M-CAT	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CAT
21	E-AGG/M-CTA	GACTGCGTACCAATTC-AGG (JOE) GATGAGTCCTGAGTAA-CTA
22	E-AGC/M-CTG	GACTGCGTACCAATTC-AGC (NED) GATGAGTCCTGAGTAA-CTG

containing 0.3 µl GeneScanTM 500 ROX® as internal size standard. The multiplexed PCR mixture was denatured for 5 min at 95 °C, quick chilled on ice for 5 min and loaded on ABI 3730xl DNA Analyzer for electrophoresis. The fragment analysis was performed by GeneMapper v 4.0 software (Applied Biosystems, USA).

2.3. Data analysis

Allele frequency and polymorphism for each AFLP fragments were scored automatically on GeneMapper as well as manually by visualizing electropherogram. The data was scored as "1" (presence of fragment) and "0" (absence of fragment) and "." for (missing data). The genotypic data were used to calculate different population genetic parameters such as polymorphic information content (PIC), marker index (MI) and resolving power (RP). The PIC value and mean gene diversity for each AFLP primer combination was calculated as part of summary statistics using software: Powermarker v 3.25 and was averaged over the fragments for each primer combination (PC).

PIC for each AFLP PC was calculated according to Roldan-Ruiz et al. (2000) formula: PICi = 2fi(1 − fi), where PICi is the polymorphic information content of marker i; fi is the frequency of the fragments which were present and 1 − fi is the frequency of the fragments which were absent. PIC was averaged over the fragments for each PC. Marker index (MI) was calculated following Powell et al. (1996) as: =PIC × EMR, where EMR (effective multiple ratio, EMR = β × n) is defined as the product of the fraction of polymorphic loci (β) and the number of polymorphic loci (n). The resolving power (RP) of each PC was calculated according to Prevost and Wilkinson (1999) as: Rp = ΣIb, where Ib is the fragment informativeness and calculated as: Ib = 1–[2 × |0.5 − p|], where p is the proportion of the genotypes containing the fragment.

Diversity indicators were calculated in total, per population and per AFLP fragment in GenAlEx 6.5 (Peakall and Smouse, 2012), POPGENE 1.31 (Yeh et al., 1997) or Arlequin 3.1 (Excoffier et al., 2005) softwares. Number of different alleles (Na), effective number of alleles ($1/(\Sigma pi^2)$) (Ne), Shannon's information index ($-1*\Sigma(pi*\ln(pi))$), gene diversity/expected heterozygosity (Nei, 1973) ($1 - \Sigma pi^2$), (where pi is the frequency of the i^{th} allele and Σpi^2 is the sum of the squared allele frequencies) and molecular variance (AMOVA) (Excoffier et al., 1992) within and among the populations were calculated in GenAlEx (Nei, 1987). POPGENE/Arlequin 3.1 was used to calculate overall diversity in collections (total gene diversity = Ht), diversity within populations (Hs), genetic differentiation (Gst = 1 − HS/HT) and Gene flow (Nm) which was estimated from Gst as Gst (Nm = 0.5 (1 − Gst)/Gst). The Jaccard's similarity coefficient was used to estimate genetic similarity between the accessions. Based on the similarity matrix, dendrograms were constructed using the unweighted pair group method with arithmetic mean (UPGMA) by using FreeTree V1.0.0.0 and TreeView (Win32) V1.6.6 softwares (Pavlicek et al., 1999).

Genetic structure was investigated using STRUCTURE v 2.3.4. (Pritchard et al., 2000; Falush et al., 2003), which applies the Markov Chain MonteCarlo (MCMC) algorithm. An admixture model with correlated allele frequencies was used. The K value was set from one to ten, and at least twenty runs were performed for each value of K. The length of the burn-in period was set to 50,000, and the MCMC chains after burn-in were run for an additional 100,000 times. The optimal value of K was determined by examination of the ΔK statistic (Evanno et al., 2005) using StructureHarvester (Earl and von Holdt, 2012). The output of structure analyses was visualized using the software CLUMPP v1.1.2 (Kopelman et al., 2015) and DISTRUCT v1.1 (Rosenberg, 2016).

3. Results

3.1. Marker polymorphism and AFLP features

Of the 26 PCs selected, 22 (84.61%) successfully amplified. These 22 PCs were used to investigate 50 *M. pruriens* accessions. All the PCs generated discrete AFLP profiles - a snapshot of which from the GeneMapper v 4.0 software (Applied Biosystems, USA) is presented in Fig. 2. They generated a total of 1208 fragments in which 1156 (95.70%) were polymorphic. The total number of fragments ranged from 12 (E-CAC/M-GCT) to 137 (E-ACT/M-CTC) with an average of 55 fragments per PC. Polymorphic fragments ranged from 11 (E-AAG/M-CAT) to 127 (E-ACA/M-CAA) with an average of 52.55. The percentage of polymorphism varied from 82% (E-ACT/M-CAA) to 100% (E-AAC/M-CAG; E-AAC/M-CTA; E-AAC/M-CTC; E-ACA/M-CTT; E-AGG/M-CAG; E-ACG/M-CAC; E-ACA/M-CAT; E-AAG/M-CAC; and E-ACA/M-CAA) with an average of 97.27%. The fraction of polymorphic loci was in the range of 0.88 to 1.00 with an average of 0.96. Mean gene diversity for different PCs varied from 0.17 to 0.37 with an average of 0.23.

The discriminatory power of the informative AFLP profile mainly depends upon three parameters namely polymorphic information content (PIC), marker index (MI) and a resolving power (RP) of a particular PC (Chandrawati et al., 2014). The calculated PIC values ranged from 0.14 to 0.30 with an average of 0.23 per fragment. The highest PIC value (0.30) was observed for the primer combination: E-ACT/M-CTA and the lowest value (0.14) was recorded for E-ACG/M-CAG. Based on the PIC values, the AFLP-PCs: E-ACT/M-CTA (0.30), E-ACG/M-CAC; E-AAG/M-CTC (0.29), E-AGG/M-CTA (0.28), and E-AGG/M-CAT (0.27)

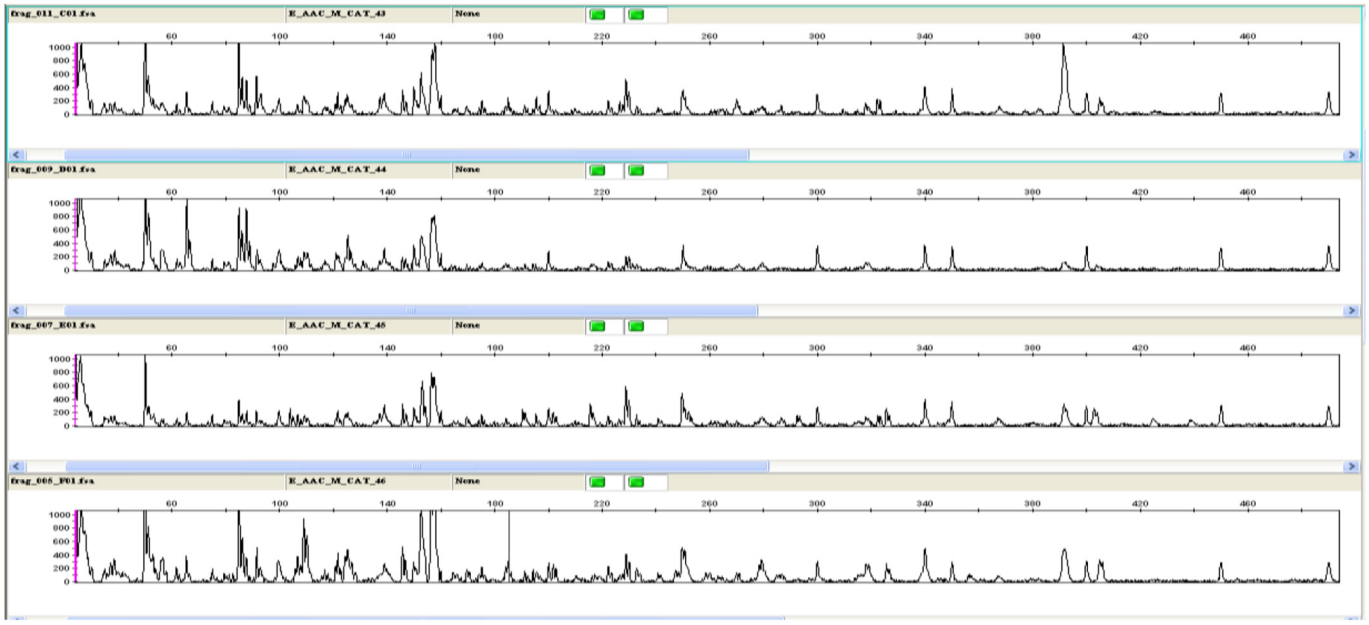


Fig. 2. Representative electropherogram showing AFLP profiles.

Table 3
Details of marker attributes obtained for different AFLP primer combinations.

Sl. no.	AFLP primer combination	n	nP	β	Gene diversity	PIC	EMR	MI	RP	%P
1	E-ACT/M-CTA	69	62	0.89	0.26	0.30	61.41	18.59	20.30	98.00
2	E-AAC/M-CAG	43	43	1.00	0.37	0.24	43.00	10.32	14.92	100
3	E-AAC/M-CTA	39	36	0.92	0.30	0.23	35.88	8.28	7.87	100
4	E-AAC/M-CTC	50	47	0.94	0.28	0.26	47.00	12.22	9.59	98.00
5	E-ACA/M-CTT	77	68	0.88	0.32	0.22	67.76	14.96	7.60	100
6	E-AGG/M-CAG	37	35	0.95	0.27	0.27	35.15	9.44	3.42	100
7	E-ACG/M-CAC	99	97	0.98	0.34	0.29	97.02	28.11	9.68	100
8	E-ACA/M-CAT	91	89	0.98	0.37	0.23	89.18	20.47	7.41	100
9	E-AAG/M-CAC	99	91	0.92	0.28	0.24	91.08	21.84	4.93	100
10	E-ACA/M-CAA	137	127	0.93	0.29	0.19	127.41	24.13	4.89	100
11	E-ACG/M-CAA	28	27	0.96	0.22	0.22	26.88	5.94	1.18	96.00
12	E-AAG/M-CAT	12	11	0.92	0.28	0.22	11.04	2.42	0.37	96.00
13	E-AAG/M-CTG	53	51	0.96	0.27	0.22	50.88	11.22	2.67	100
14	E-AAG/M-CTC	68	68	1.00	0.27	0.29	68.00	19.72	2.34	98.00
15	E-AAG/M-CAA	39	39	1.00	0.36	0.16	39.00	6.24	1.56	98.00
16	E-ACT/M-CAA	40	39	0.98	0.18	0.14	39.20	5.46	0.56	82.00
17	E-ACG/M-CAG	38	38	1.00	0.17	0.14	38.00	5.32	0.65	98.00
18	E-AGC/M-CTA	72	71	0.99	0.24	0.20	71.28	14.20	1.95	94.00
19	E-ACG/M-CTG	20	20	1.00	0.25	0.21	20.00	4.20	0.53	98.00
20	E-AGG/M-CAT	41	41	1.00	0.33	0.27	41.00	11.07	1.44	94.00
21	E-AGG/M-CTA	31	31	0.97	0.36	0.28	30.07	8.40	17.34	92.00
22	E-AGC/M-CTG	25	25	1.00	0.28	0.23	25.00	5.75	10.3	98.00
	Min	12	11	0.88	0.17	0.14	11.04	2.42	0.37	82.00%
	Max	137	127	1.00	0.37	0.30	127.41	28.11	20.30	100%
	Avg.	54.91	52.55	0.96	0.29	0.23	52.51	12.20	5.98	97.27%
		1208	1156							

n: Total number of fragments; nP: Number of polymorphic fragments; β: Fraction of Polymorphic Loci; PIC: Polymorphic information content; **EMR**: Effective multiplex ratio; Rp: Resolving power; MI: Marker index; P: Percent Polymorphism.

were determined to be more informative than others. The EMR varied from 11.04 (E-AAG/M-CAT) to 127.41 (E-ACA/M-CAA) with an average of 52.51. The MI values varied between 2.42 (E-AAG/M-CAT) to 28.11 (E-ACG/M-CAC) with an average of 12.20. RP varied from 0.37 (E-AAG/M-CAT) to 20.30 (E-ACT/M-CTA) with an average of 5.19 (Table 3). The various marker features showed a significant correlation to each other. PIC and MI showed a correlation (r^2) of 0.83, $p < 0.005$; MI and RP showed $r^2 = 0.99$, $p < 0.005$ while $r^2 = 0.85$, $p < 0.005$ was the correlation between RP and PIC.

3.2. Genetic diversity and population structure analysis

Measures of genetic diversity are provided in Table 4. The effective number alleles (Ne) for different PCs ranged from 1.16 to 1.60 with mean of 1.38. Pair-wise similarity measure estimated using Jaccard's coefficient and Shannon's index (I) produced average values of 0.37 and 0.34 indicating high genetic diversity in our collection. This is also reflected in the values, h and Ht which produced mean scores of 0.27 and 0.22, respectively. On the contrary, the indices of genetic differentiation (Gst) was low (0.08) and corresponding gene flow value (Nm) was high (6.78) suggesting high levels of gene flow between different

Table 4
Diversity indicators for different AFLP markers across all the populations.

Primer combinations	Na*	Ne*	h*	I*	Ht*	Hs*	Gst*	Nm*
E-ACT/M-CTA	1.39	1.36	0.07	0.33	0.21	0.08	0.16	2.56
E-AAC/M-CAG	1.67	1.48	0.14	0.42	0.28	0.13	0.14	3.02
E-AAC/M-CTA	1.42	1.44	0.18	0.38	0.25	0.16	0.10	4.57
E-AAC/M-CTC	1.47	1.40	0.20	0.35	0.23	0.16	0.12	3.58
E-ACA/M-CTT	1.69	1.35	0.29	0.35	0.22	0.25	0.07	6.23
E-AGG/M-CAG	1.32	1.35	0.27	0.31	0.21	0.24	0.05	8.69
E-ACG/M-CAC	1.67	1.45	0.27	0.41	0.27	0.23	0.08	5.92
E-ACA/M-CAT	1.69	1.47	0.30	0.42	0.28	0.28	0.08	5.45
E-AAG/M-CAC	1.45	1.34	0.32	0.32	0.21	0.27	0.08	6.07
E-ACA/M-CAA	1.34	1.30	0.31	0.29	0.19	0.29	0.04	12.20
E-ACG/M-CAA	1.28	1.33	0.32	0.29	0.19	0.29	0.08	5.90
E-AAG/M-CAT	1.10	1.23	0.29	0.23	0.14	0.27	0.07	7.17
E-AAG/M-CTG	1.49	1.43	0.30	0.38	0.25	0.28	0.05	9.60
E-AAG/M-CTC	1.38	1.36	0.30	0.32	0.21	0.27	0.07	7.17
E-AAG/M-CAA	1.62	1.43	0.29	0.39	0.26	0.27	0.05	10.06
E-ACT/M-CAA	1.07	1.16	0.30	0.19	0.11	0.26	0.04	13.33
E-ACG/M-CAG	1.07	1.22	0.29	0.21	0.14	0.27	0.07	6.55
E-AGC/M-CTA	1.28	1.30	0.32	0.28	0.18	0.30	0.05	8.66
E-ACG/M-CTG	1.28	1.35	0.29	0.30	0.20	0.27	0.07	6.79
E-AGG/M-CAT	1.62	1.47	0.32	0.41	0.28	0.31	0.05	10.00
E-AGG/M-CTA	1.97	1.60	0.34	0.50	0.34	0.29	0.14	3.09
E-AGC/M-CTG	2.00	1.49	0.29	0.44	0.29	0.24	0.17	2.49
Mean	1.47	1.38	0.27	0.34	0.22	0.25	0.08	6.78
Std. Dev	0.26	0.10	0.07	0.08	0.06	0.05	0.03	3.02

*Na: Number of different alleles; Ne: Effective no. of alleles; h: Nei's (1973) gene diversity; I: Shannon's Information index; Ht: Diversity in overall collections total gene diversity; **Hs**: Sub divided population; Gst: Genetic differentiation; **Nm**: Estimate of gene flow from Gst or Gcs, Nm = 0.5(1 − Gst)/Gst.

Table 5
Gene diversity estimates based on botanical varieties and geographical distribution.

Population Groups	N*	Na*	Ne*	I*	h*	He*	P (%)
Botanical varieties							
var. *pruriens*	38	1.91	1.48	0.44	0.29	0.25	95.67
var. *utilis*	12	1.45	1.40	0.37	0.24	0.22	71.54
Average	25	1.68	1.44	0.40	0.26	0.23	83.61
Geographical distribution							
Eastern Himalayas	11	1.52	1.39	0.35	0.26	0.23	75.26
Assam plains	39	1.88	1.41	0.38	0.28	0.24	93.77
Average	25	1.70	1.40	0.37	0.27	0.24	84.52

*N: Number of population; Na: Number of different alleles; Ne: Effective no. of alleles; I: Shannon's Information index; h: Nei's (1973) gene diversity; He: Expected heterozygosity; P (%): Percent polymorphism

population groups. On an expected line, among the two botanical varieties, genetic diversity within var. *pruriens* (wild) was more (I: 0.44) as compared to cultivated var. *utilis* (0.37) (Table.5).

The dendrogram based on UPGMA revealed two major clusters (cluster I and II) accommodating 96% of the accessions; two accessions 500221AR and 500239MZ formed an out-group (Fig. 3a).

Among the two clusters, cluster I was the largest with 27 (54%) accessions and grouped mostly collection from Assam plains (AP). This was further divided into three sub-clusters namely IA, IB and IC with 18, 5 and 4 accessions respectively. Sub-cluster IA contained mix of var. *pruriens* and var. *utilis* accessions while sub-clusters IB and IC grouped majorly var. *pruriens* accessions except one var. *utilis* in each group (500210MN, 500203MN). However, no location specific (states) grouping was observed within cluster-I. Likewise, Cluster II contained 21 (42%) accessions which were further divided into two sub-clusters namely IIA and IIB representing 14 and 7 accessions, respectively. Of these, sub-cluster IIA grouped mostly var. *pruriens* accessions from eastern Himalayan region (EH) except one accessions of var. *utilis* (500213MN) and sub-cluster IIB mostly separated var. *utilis* accessions from AP except two accessions (500246TR and 500248TR) belonging to var. *pruriens*. Thus, though not entirely, both varietal and geographical affiliation emerged to have played some role in groping of the accessions.

Close examination of the UPGMA data also revealed configuration of two independent gene pools representing AP (cluster I) and EH (cluster IIB). Therefore, for AMOVA analysis we considered two different categories (a) Assam plains and eastern Himalayan regions based on geographical distribution and (b) var. *pruriens* and var. *utilis* based on botanical varieties. Partitioning of genetic variations within and between population groups revealed, 7% and 8% of the total genetic variations was accrued from inter-population variation and 92% and 93% variance was contributed by intra-population variations for botanical varieties and geographical distribution categories, respectively (Table 6). Low Fst values (0.08 and 0.07) and high Nm estimates (3.04 and 3.50) for both the categories further supported lack of population structure as a factor of high gene flow.

The results revealed presence of 2 sub-populations as ΔK was maximum for K = 2. The grouping pattern based on StructureHarvester results (Fig. 3b) was broadly in conformity with UPGMA results (Fig. 3a). For instance, in case of UPGMA, we could recognize formation of two broad clusters representing Assam plains and eastern Himalayas. The maximum ΔK value obtained for K-2 in structure analysis possibly represent this. However three sub-clusters formed within cluster-I lacked any pattern of geographical structuring. But significant gene flow between population groups was evident as the structure plot revealed several admixture taxa in our collection. Likewise, majority of the var. *utilis* accessions from AP formed part of cluster-IIB even in structure plot.

4. Discussion

The center of origin of a crop plant is normally associated with broad genetic diversity due to presence of wild and weedy relatives (Guo et al., 2014). They represent vital source of gene pool to farmers and professional breeders responding to biotic and abiotic stresses and climate change (Dyer et al., 2014). Medicinal legume *M. pruriens* is thought be native of southern China, Malaysia or eastern India (Burkill, 1966; Duke, 1981; Wilmot-Dear, 1984) with considerable antiquity in these regions (Arora, 1991). It is reported that *Mucuna* was widely grown in the foothills and lower hills of the eastern Himalayas during the eighteenth and nineteenth centuries and was eventually replaced due to introduction of more palatable legumes (CSIR, 1962; Piper and Tracy, 1910; Watt, 1883). However, its cultivation as a minor crop continues even today by many native people who grow it as home garden crop for dry seeds and green pods (Arora, 1991). Recently, with the discovery of L-Dopa and other agronomic benefits from the plant, the opportunities for its large-scale cultivation in the region are growing necessitating development of improved cultivars with enhanced nutritional value and resistance to biotic and abiotic stresses (Sathyanarayana et al., 2017). Even though northeast India represents important source of gene pool for *M.pruriens,* little data is available on genetic diversity and population structure of this species from the region. Thus, the present investigation was carried out to meet this objective using AFLP markers. AFLP was particularly chosen as it offers pan genome coverage comprising large number of loci from different chromosomal regions, and regarded as choice marker for diversity analysis (Karp and Edwards, 1995; Mueller and Wolfenbarger, 1999; Eva Mikuláškova et al., 2012; Patrick et al., 2012). Earlier, AFLPs have been successfully used for analyzing genetic variation, distance based phylogenetic relationships and characterization of natural populations as well as breeding lines across range of taxa (Hill et al., 1996; Kardolus et al., 1998; Roldan-Ruiz et al., 2000; De Riek et al., 2001; Dubey et al., 2010). Their utility for genetic diversity analysis in *Mucuna* species is also well recognized (Capo-chichi et al., 2004; Sathyanarayana et al., 2011).

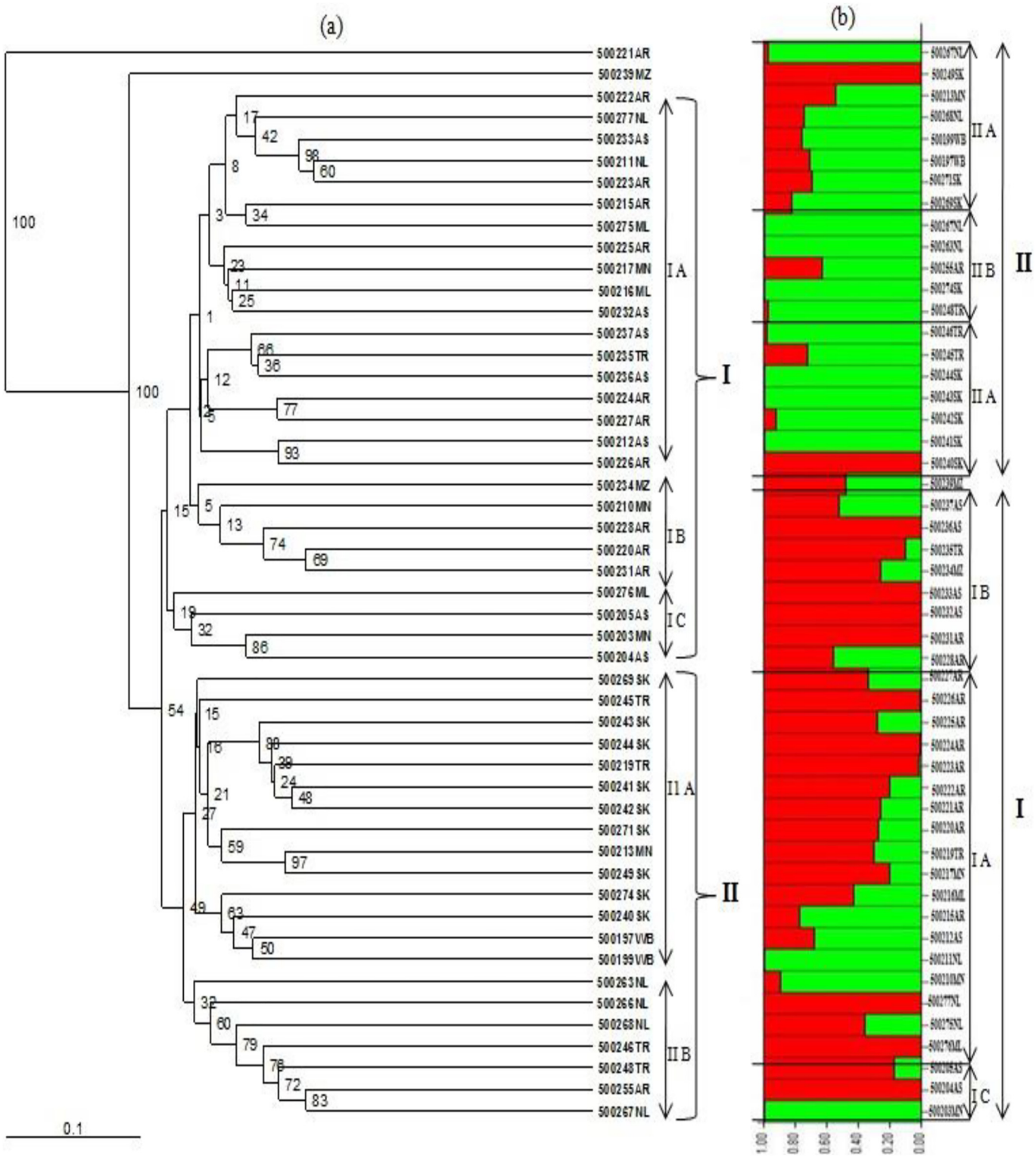


Fig. 3. (a) UPGMA-dendrogram of 50 *M. pruriens* accessions based on Jaccard's coefficient and (b) hierarchical organization of genetic relatedness analyzed by STRUCTURE V 2.3.4 program. Each colour represents one population and length of segment shows estimated membership proportion of each accession.

4.1. Marker attributes

Majority of the PCs produced high quality AFLP profiles generating an average of 54.91 total and 52.55 (97.27%) polymorphic fragments per PC. A high rate of polymorphism indicated good genetic variability among the accessions investigated. In diversity studies, PIC values are used as primary measure of discriminatory power or informativeness of PC (Chandrawati et al., 2014.). The average PIC value of 0.23 obtained in the present study although was on lower side, is superior to PIC value (0.16) reported in earlier AFLP studies on *Mucuna* spp. (Sathyanarayana et al., 2011). Higher PIC values obtained in the present investigation might be due to new AFLP markers employed in the present study. Out of 22 PCs, 9 (41%) produced 100% polymorphic fragments. Thus, these informative PCs could be used for large scale germplasm

Table 6
AMOVA based on botanical varieties as well as geographical distribution.

Population group	Df	SS	MS	Est. Var.	%	Fst	Nm
Botanical varieties							
Among Pops	1	416.02	416.02	13.69	8%	0.08 & $p > 0.001$	3.04
Within Pops	48	7982.26	166.30	166.30	92%		
Total	49	8398.28		179.99	100%		
Geographical distribution							
Among Pops	1	374.45	374.45	12.08	7%	0.07 & $p > 0.001$	3.50
Within Pops	48	8023.83	167.16	167.16	93%		
Total	49	8398.28		179.24	100%		

Df: Degree of freedom; SS: Sum of squares; MS: Mean squares; Fst: Genetic differentiation among populations; Nm: Estimate of gene flow from Gst or Gcs, $Nm = 0.5(1 - Gst)/Gst$.
To corroborate this, we carried out population structure analysis based on Bayesian statistics.

characterization and genotyping. The other two parameters, RP and MI provide complementary attributes to the efficiency of the marker system and are used along with PIC in diversity studies (Gupta et al., 2013; Pecina-Quintero et al., 2013). Considering combination of PIC, MI and RP parameters, the PCs: E-ACG/M-CAC, E-ACA/M-CAA, E-AAG/M-CAC and E-ACA/M-CAT and RP PCs E-ACT/M-CTA, E-AGG/M-CTA and E-AAC/M-CAG are found to be more useful for future diversity studies. Among these, the PCs: E-ACT/M-CTA, E-AAC/M-CAG, E-AAC/M-CTA and E-AAC/M-CTC are already reported to be informative in earlier AFLP study on *Mucuna* species (Sathyanarayana et al., 2011), and thus our result is in conformity with the above finding.

4.2. Gene diversity and population structure

A pair wise similarity matrix based on Jaccard's coefficient was constructed to understand the extent of genetic diversity among all the *M. pruriens* accessions. A high level of diversity was noticed (average similarity coefficient: 0.37; average Shannon's information index: 0.34) within our collection which is reflected even in the values of h and Ht with mean scores of 0.27 and 0.22, respectively. This is significant as the accessions used here correspond to neighboring areas of eastern India and southern China from where *M. pruriens* is reported to have come from (Capo-chichi et al., 2003). So far, only anecdotal evidences were available to support this theory including large-scale presence of perennial *M. bracteata* - a probable ancestor of this species (Jaheer et al., 2015) in the region. Thus, high allelic diversity among *M. pruriens* of northeast India, reported for the first time in the present study provides earliest empirical evidence for the eastern Indian origin of *M. pruriens* and paves a way for future investigation on migration and colonization pattern of this species as well as events that shaped domestication of this species. Among the two botanical varieties, var. *pruriens* which represents wild variety exhibit higher diversity (I: 0.44) as compared to var. *utilis* (I: 0.37) which is a cultivated variety. The phenomenon of wild *M. pruriens* varieties exhibiting greater diversity vis-à-vis cultivated var. *utilis* is a common observation reported in many earlier studies (Leelambika and Sathyanarayana, 2011; Sathyanarayana et al., 2012, 2016).

On the other hand, the population genetic structure investigated using UPGMA, AMOVA and STRUCUTRE revealed absence of geographical structure in *M. pruriens* population of northeast India. In case of AMOVA - which determines contribution of individual populations to the total diversity, the results revealed 93% of the variance was accrued by intra-population variation while only 7% contributed to inter-population variation pointing towards absence of barriers for gene flow. To corroborate this, we estimated Fst and Nm to measure the extent of gene flow. The moderate estimate of Fst (0.08.) and high Nm value (3.04) substantiated AMOVA results of low divergence as a factor of high gene flow. It is well known that gene flow values (Nm) > 1 is strong enough to prevent substantial differentiation due to genetic drift (Slatkin and Barton, 1989). It is likely that *M. pruriens* has experienced little geographical isolation in this region due to predominant distribution in the foothills extending up to plains in this otherwise geographically and climatically diverse region comprising some of the large mountains and mighty rivers that offers rich scope for speciation by isolation. Such a situation might have rather favored long distance pollen/seed dispersal allowing frequent exchange of alleles between different populations leading to low differentiation as noted here. Even though out-crossing is reported to be rare in *M. pruriens* (Duke, 1981), it cannot be entirely ruled out owing to presence of numerous natural hybrids as reported in earlier works (Piper and Tracy, 1910; Bailey, 1947; Burkill, 1966; Capo-chichi et al., 2001; Sathyanarayana et al., 2016). It is possible that genetic compatibility between wild and cultivated *M.pruriens* population has lead to gene flow and genetic introgression from wild populations to domesticated ones and vice-versa resulting in inter-varietal hybrids as reported in Andes (Freyre et al., 1996; Beebe et al., 1997) and Mesoamerican cultivars of common bean (Papa and Gepts, 2003).

Further, close examination of the UPGMA and STRUCTURE results points towards configuration of two independent gene pools possibly representing Assam plains (AP) and eastern Himalayas (EH). However sub-clusters within cluster-I (Assam plains) lacked specific grouping pattern. This might be due to the fact that majority of these accessions were sampled from the areas adjoining Assam in the neighboring states of Meghalaya, Tripura and Mizoram in south, Manipur and Nagaland in east and foothills of Arunachal Pradesh in the extreme of north. This topography forms a continuous range with considerable similarity in species distribution (Chatterjee, 2008). Thus, *M. pruriens* distributed along this range perhaps correspond to single continuous population as gene flow has amalgamated neighboring populations into one with common genetic structure. On the contrary, the accessions grouped in sub-cluster IIA likely signify divergent gene pool representative of eastern-Himalayas as these accessions were acquired from the lower altitudes of Sikkim-Darjeeling hills along the Teesta River which represent distinct floristic elements. It is also possible that the latter might symbolize migrated elements from the North Bengal plains through anthropogenic activities or animal movements as this region (foothills of Sikkim-Darjeeling) is located adjoining to Siliguri-Malda plains of North Bengal. The grouping of two Tripura accessions along with cluster-IIA possibly can be explained based on this. That said, not even single *M. pruriens* population were to be found along the Sikkim-Darjeeling Himalayan region until a few years back. However, in recent years continuous migration of several tropical plants including *M. pruriens* has been observed along the altitudinal gradient of EHs possibly signifying changing floristic composition of EHs due to climate change which forms another interesting piece of work for future investigation.

The other sub-cluster within cluster II (IIB) mostly represents accessions belonging to var. *utilis* - a cultivated accession. Alignment of var. *utilis* from AP along with cluster-II is a little difficult to explain owing to following reasons: Firstly, these accessions were obtained as seed samples from private vendors/researchers and thus their origin could not be authentically established. Secondly, the variety itself presents a chequered background with no evidence available on its genetic history from anywhere in the world and thus it is hard to pinpoint contributing factors for its atypical distribution. Thirdly, the fact that it freely hybridizes with other varieties of *M. pruriens* (var. *pruriens* and var. *hirsuta*) has resulted in several natural hybrids rendering precise genotypic identification difficult. This is evident even from the results of structure analysis in our study which confirms presence of several admixture taxa in our collection. The latter problem is common to all the three varieties of *M. pruriens* and thus has been a subject of debate in literature on *Mucuna* species (Capo-chichi et al., 2001; Eilitta

et al., 2002; Sathyanarayana et al., 2016). Thus future work should focus on genetic history of *M. pruriens* varieties especially the cultivated var. *utilis* to reliably address questions related to pattern of diversity in addition to migration, domestication and improvement of this promising underutilized crop.

5. Conclusion

It is concluded that *M. pruriens* population of northeast India exhibits high diversity providing credence to northeast Indian origin theory of this species. Pair-wise estimates of genetic differentiation were low and AMOVA revealed no evidence of genetic structuring, suggesting species did not experience severe bottleneck events or extensive genetic drift in the region. However, clustering based on UPGMA and STRUCTURE broadly point to presence of two independent gene pools representing Assam plains and eastern Himalayas. Large topography related genetic variation may be maintained by strong selection in these gene pools despite high gene flow observed within this species. Thus future work should focus on the mechanisms underlying such variation besides addressing key questions related to migration and domestication history of *M. pruriens* varieties in general and var. *utilis* in particular.

Conflict of interest

None.

Funding

All the funding agencies and support levied are acknowledged.

Acknowledgments

Authors thank Sikkim University, Gangtok and CSIR-National Botanical Research Institute, Lucknow for the facilities. NS thank Science and Engineering Research Board (SERB), Govt. of India for the financial support. Authors also thank the reviewers of the manuscript for their support in publishing this work.

References

Agharkar, S.P., 1991. Medicinal Plants of Bombay Presidency. Scientific Publication, Jodhpur, India, pp. 1–2.

Arora, K.R., 1991. Native food plants of the northeastern India. In: Jain, S.K. (Ed.), Contributions to Ethnobotany of India. Scientific Publishers, Jodhpur, India, pp. 137–152.

Bailey, L.H., 1947. The Standard Cyclopedia of Horticulture. Macmillan, New York, NY, USA.

Beebe, S., Toro, O., Gonzales, A.V., Chacon, M.I., Debouck, D., 1997. Wile-weed-crop complexes of common bean (*Phaseolous vulgaris* L., *Fabaceae*) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet. Resour. Crop. Evol. 44, 73–91. http://dx.doi.org/10.1023/A:1008621632680.

Burkill, I.H., 1966. A Dictionary of the Economic Products of the Malay Peninsula. Ministry of Agriculture and Cooperatives, Kuala Lumpur, Malaysia.

CABI, 2015. *M. pruriens* (Velvet bean). https://www.cabi.org/isc/datasheet/35134http://dx.doi.org/10.3402/jmahp.v3.29679.

Capo-chichi, L.J.A., Weaver, D.B., Morton, C.M., 2001. AFLP assessment of genetic variability among velvet bean (*Mucuna* sp.) accessions. Theor. Appl. Genet. 103, 1180–1188. http://dx.doi.org/10.1007/s001220100722.

Capo-chichi, L.J.A., Weaver, D.B., Morton, C.M., 2003. The use of molecular markers to study genetic diversity in *Mucuna*. Trop. Subtrop. Agroecosyst. 1, 309–318.

Capo-chichi, L.J.A., Morton, C.M., Weaver, D.B., 2004. An intraspecific genetic map of velvetbean (*Mucuna* sp.) based on AFLP markers. Theor. Appl. Genet. 108, 814–821. http://dx.doi.org/10.1007/s00122-003-1493-8.

Chandrawati, Maurya, R., Singh, P.K., Ranade, S.A., Yadav, H.K., 2014. Diversity analysis in Indian genotypes of linseed (*Linum usitatissimum* L.) using AFLP markers. Gene 549, 171–178. http://dx.doi.org/10.1016/j.gene.2014.07.067.

Chatterjee, S., 2008. Biodiversity conservation issues of Northeast India. Int. For. Rev. 10 (2), 315–324.

Council of Scientific and Industrial Research, New Delhi, 1962. The Wealth of India: A Dictionary of Indian Raw Materials and Industrial Products. CSIR, New Delhi.

De Riek, J., Calsyn, E., Everaert, I., Van Bockstaele, E., De Loose, M., 2001. AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet. 103, 1254–1265. http://dx.doi.org/10.1007/s001220100710.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Dubey, M.K., Shasany, A.K., Dhawan, O.P., Shukla, A.K., Shanker, K., Khanuja, S.P.S., 2010. Detection of DNA polymorphism in *Papaver somniferum* genotypes differing in straw morphinan alkaloid content. Plant Biosyst. 144, 513–517.

Duke, J.A., 1981. Handbook of Legumes of World Economic Importance. Plenum Press, New York.

Dyer, G.A., Lopez-Feldman, A., Yunez-Naude, A., Taylor, J.E., 2014. Genetic erosion in maize's centre of origin. Proc. Natl. Acad. Sci. U. S. A. 111 (39), 14094–14099. http://dx.doi.org/10.1073/pnas.1407033111.

Earl, D.A., von Holdt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361. http://dx.doi.org/10.1007/s12686-011-9548-7.

Eilitta, M., Bressani, R., Carew, L.B., Carsky, R.J., Flores, M., Gilbert, R., Huyck, L., St. Laurent, L., Szabo, N.J., 2002. *Mucuna* as a food and feed crop: an overview. In: Flores, B.M., Eilittä, M., Myhrman, R., Carew, L.B., Carsky, R.J. (Eds.), Food and Feed from *Mucuna*: Current Uses and the Way Forward, Workshop, CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, Honduras (April 26–29, 2000), pp. 18–47.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 2005 (14), 2611–2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x.

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479–491.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinforma. 1, 47–50.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multi locus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.

Freyre, R., Ri'os, R., Guzma'n, L., Debouck, D., Gepts, P., 1996. Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. Econ. Bot. 50, 195–215. http://dx.doi.org/10.1007/BF02861451.

Guo, Y., Chen, S., Li, Z., Cowling, W.A., 2014. Center of origin and centers of diversity in an ancient Crop, *Brassica rapa* (Turnip Rape). J. Hered. 105 (4), 555–565. http://dx.doi.org/10.1093/jhered/esu021.

Gupta, A., Maurya, R., Roy, R.K., Sawant Samir, V., Yadav, H.K., 2013. AFLP based genetic relationship and population structure analysis of *Canna* - an ornamental plant. Sci. Hortic. 154, 1–7. http://dx.doi.org/10.1016/j.scienta.2013.02.005.

Haridas, P., Baby, M., Mathew, P.J., Mathew, P.M., 2013. Cytological studies in *Mucuna pruriens* species complex occurring in peninsular India. J. Cytol. Genet. 14, 9–14.

Harlan, J.R., 1971. Agricultural origins: centers and noncenters. Science 174, 468–474. http://dx.doi.org/10.1126/science.174.4008.468.

Hill, M., Witsenboer, H., Zabeau, M., Vos, P., Kesseli, R., Michelmore, R., 1996. PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor. Appl. Genet. 93, 1202–1210. http://dx.doi.org/10.1007/BF00223451.

Jaheer, M., Chopra, R., Kunder, K.R., Bhat, D., Rashmi, K.V., Sathyanarayana, N., 2015. Cytogenetic and ITS-psbA-trnH sequence analysis for phylogenetic inference in *Mucuna* sp. of India. Trop. Plant Biol. 8, 108–116. http://dx.doi.org/10.1007/s12042-015-9154-z.

Kardolus, J.P., Van-Eck, H.J., Van den Berg, R.G., 1998. The potential of AFLPs in biosystematics a first application in *Solanum* taxonomy (*Solanaceae*). Plant Syst. Evol. 210, 87–103. http://dx.doi.org/10.1007/BF00984729.

Karp, A., Edwards, K.J., 1995. Molecular techniques in the analysis of the extent and distribution of genetic diversity. In: Molecular Genetic Techniques for Plant Genetic Resources. Report of an IPGRI workshop, 9–11 October, Rome, Italy, http://dx.doi.org/10.2466/pms.1995.81.3.955.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across. Mol. Ecol. Resour. 15, 1179–1191. http://dx.doi.org/10.1111/1755-0998.12387.

Lampariello, L.R., Cortelazzo, A., Guerranti, R., Sticozzi, C., Valacchi, G., 2012. The magic bean of *M. pruriens*. J. Tradit. Complement Med. 2 (4), 331–339. http://dx.doi.org/10.1016/S2225-4110(16)30119-5.

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

Mikulášková, Eva, Fér, Tomáš, Kučabová, Veronika, 2012. The effect of different DNA isolation protocols and AFLP fingerprinting optimizations on error rate estimates in the bryophyte *Campylopus introflexus*. Lindbergia 35, 7–17.

Mueller, U.G., Wolfenbarger, L.L., 1999. AFLP genotyping and fingerprinting. Trends Ecol. Evol. 14, 389–394. http://dx.doi.org/10.1016/S0169-5347(99)01659-6.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. 7 (0), 3321–3323.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Papa, R., Gepts, P., 2003. Assymetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (Phaseolous vulgaris L.) from Mesoamerica. Theor Appl Genet. 106, 239–250.

Patrick, K.C., Carola, G., Bernhard, M., France, G., 2012. Automated masking of AFLP markers improves reliability of phylogenetic analyses. PLoS One 7 (11), e49119.

Pavlicek, A., Hrda, S., Flegr, J., 1999. FreeTree - freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of the genus *Frenkelia*. Folia Biol. (Praha) 45, 97–99.

Peakall, R., Smouse, P., 2012. GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research - an update. Bioinformatics 28, 2537–2539. http://dx.doi.org/10.1093/bioinformatics/bts460.

Pecina-Quintero, V., Anaya-Lopez, J.L., Nunez-Colin, C.A., Zamarripa-Colmenero, A., Montes-Garcia, N., Solís-Bonilla, J.L., Aguilar-Rangel, M.R., 2013. Assessing the genetic diversity of castor bean from Chiapas, México using SSR and AFLP markers. Ind. Crop. Prod. 41, 134–143. http://dx.doi.org/10.1016/j.indcrop.2012.04.033.

Piper, C.V., Tracy, S.M., 1910. Type of *Stizolobium pachylobium*.S. Department of Agriculture Bureau of Plant Industry Bulletin. 179. pp. 18–19 (pl. 4b, 7).

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., Rafalasky, A., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. (3), 225–238. http://dx.doi.org/10.1007/BF00564200.

Prevost, A., Wilkinson, M.J., 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor. Appl. Genet. 98, 107–112. http://dx.doi.org/10.1007/s001220051046.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multi locus genotype data. Genetics 155, 945–959.

Pulikkalpura, H., Kurup, R., Mathew, P.J., Baby, S., 2015. L-Dopa in *Mucuna pruriens* and its degradation. Sci. Rep. 5, 11078. http://dx.doi.org/10.1038/srep11078.

Roldan-Ruiz, I., Dendauw, J., Bockstaele, V., Depicker, A., Loose, M., 2000. AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Mol. Breed. 6, 125–134. http://dx.doi.org/10.1023/A:1009680614564.

Rosenberg, N.A., 2016. Admixture models and the breeding systems of H. S. Jennings: a GENETICS connection. Genetics 202, 9–13. http://dx.doi.org/10.1534/genetics.115.181057.

Russell, K.M., Charles, H.J., William III, W.A., Edward, M.D., Gary, M.S., Ray, F., 1992. Relationships among isoprene emission rate, photosynthesis, and isoprene synthase activity as influenced by temperature. Plant Physiol. 98, 1175–1180 (0032-0889/92/98/11 75/06/$01 .00/0).

Sathyanarayana, N., Leelambika, M., Mahesh, S., Jaheer, M., 2011. AFLP assessment of genetic diversity among Indian *Mucuna* Accessions. Physiol. Mol. Biol. Plants 17 (2), 171–180. http://dx.doi.org/10.1007/s12298-011-0058-6.

Sathyanarayana, N., Mahesh, S., Jaheer, M., Leelambika, M., 2012. Geneic diversity of wild and cultivated *Mucuna pruriens* (L.) DC accessions analysed using thirty morpho-agronomical charecters. Trop. Subtrop. Agroecosyst. 15, 249.

Sathyanarayana, N., Mahesh, S., Leelambika, M., Jaheer, M., Chopra, R., Rashmi, K.V., 2016. Role of genetic resources and molecular markers in *M. pruriens (L.)*DC improvement-a review. Genetic resources: characterization & utilization. Plant. Genet. Resour. 1–13.

Sathyanarayana, N., Pittala, Ranjith Kumar, Tripathi, Pankaj Kumar, Chopra, Ratan, Singh, Heikham Russiachand, Belamkar, Vikas, Bhardwaj, Pardeep Kumar, Doyle, Jeff J., Egan, Ashley N., 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.

Siddhuraju, P., Becker, K., 2001. Rapid reversed-phase high performance liquid chromatographic method for the quantification of L-Dopa (L-3,4-dihydroxyphenylalanine), non-methylated and methylated tetra hydro isoquinoline compounds from *Mucuna* beans. Food Chem. 72, 389–394. http://dx.doi.org/10.1016/S0308-8146(00)00281-8.

Slatkin, M., Barton, N.H., 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43, 1349–1368. http://dx.doi.org/10.1111/j.1558-5646.1989.tb02587.x.

Vos, P., Hogers, R., Bleeker, M., Reijanas, M., van de Lee, T., Hornes, M., Freijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23, 4407–4414. http://dx.doi.org/10.1093/nar/23.21.4407.

Watt, G., 1883. Economic Products of India. Superintendent of Government Printing, Calcutta, India.

Wilmot-Dear, C.M., 1984. A revision of *Mucuna* (*Leguminosae, Phaseoleae*) in China and Japan. Kew Bull. 39 (1), 23–65. http://dx.doi.org/10.2307/4107853.

Yeh, F.C., Yang, R.C., Boyle, T.B., Ye, Z., Mao, J.X., 1997. POPGENE, the user-friendly shareware for population genetic analysis. In: Molecular Biology and Biotechnology Centre. vol. 10 University of Alberta, Canada.

Sathyanarayana et al. BMC Genomics (2017) 18:409
DOI 10.1186/s12864-017-3780-9

BMC Genomics

RESEARCH ARTICLE Open Access

Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *de novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers

N. Sathyanarayana[1*], Ranjith Kumar Pittala[1], Pankaj Kumar Tripathi[1], Ratan Chopra[2], Heikham Russiachand Singh[3], Vikas Belamkar[4], Pardeep Kumar Bhardwaj[5], Jeff J. Doyle[6] and Ashley N. Egan[7*]

Abstract

Background: The medicinal legume *Mucuna pruriens* (L.) DC. has attracted attention worldwide as a source of the anti-Parkinson's drug L-Dopa. It is also a popular green manure cover crop that offers many agronomic benefits including high protein content, nitrogen fixation and soil nutrients. The plant currently lacks genomic resources and there is limited knowledge on gene expression, metabolic pathways, and genetics of secondary metabolite production. Here, we present transcriptomic resources for *M. pruriens*, including a *de novo* transcriptome assembly and annotation, as well as differential transcript expression analyses between root, leaf, and pod tissues. We also develop microsatellite markers and analyze genetic diversity and population structure within a set of Indian germplasm accessions.

Results: One-hundred ninety-one million two hundred thirty-three thousand two hundred forty-two bp cleaned reads were assembled into 67,561 transcripts with mean length of 626 bp and N50 of 987 bp. Assembled sequences were annotated using BLASTX against public databases with over 80% of transcripts annotated. We identified 7,493 simple sequence repeat (SSR) motifs, including 787 polymorphic repeats between the parents of a mapping population. 134 SSRs from expressed sequenced tags (ESTs) were screened against 23 *M. pruriens* accessions from India, with 52 EST-SSRs retained after quality control. Population structure analysis using a Bayesian framework implemented in fastSTRUCTURE showed nearly similar groupings as with distance-based (neighbor-joining) and principal component analyses, with most of the accessions clustering per geographical origins. Pair-wise comparison of transcript expression in leaves, roots and pods identified 4,387 differentially expressed transcripts with the highest number occurring between roots and leaves. Differentially expressed transcripts were enriched with transcription factors and transcripts annotated as belonging to secondary metabolite pathways.

(Continued on next page)

* Correspondence: nsathyanarayana@cus.ac.in; egana@si.edu; ashegan2@gmail.com
[1]Department of Botany, Sikkim University, 6th Mile, Tadong-737102, Gangtok, Sikkim, India
[7]Department of Botany, Smithsonian Institution, National Museum of Natural History, US National Herbarium, 10th and Constitution Ave NW, Washington, DC 20013, USA
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The *M. pruriens* transcriptomic resources generated in this study provide foundational resources for gene discovery and development of molecular markers. Polymorphic SSRs identified can be used for genetic diversity, marker-trait analyses, and development of functional markers for crop improvement. The results of differential expression studies can be used to investigate genes involved in L-Dopa synthesis and other key metabolic pathways in *M. pruriens*.

Keywords: Velvet bean, *Mucuna pruriens*, Transcriptomics, Differential gene expression, EST-SSRs, Population structure, Leguminosae, Fabaceae

Background

There are many minor food legumes whose potential is underexploited and untapped. Adzuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi], velvet or itching bean (*Mucuna* spp.), bambara groundnut (*Vigna subterranea* L.), faba bean (*Vicia faba* L.), horse gram [*Macrotyloma uniflorum* (Lam.) Verdc.], hyacinth bean *(Lablab purpureus* L.*)*, grass pea (*Lathyrus sativus* L.), moth bean [*Vigna aconitifolia* (Jacq.)], rice bean [*Vigna umbellata* (Thunb.) Ohwi and Ohashi] and winged bean [*Psophocarpus tetragonolobus* (L.) DC.] are prominent members of this group [1]. Many of them possess rich nutritional value and form an important source of protein, vitamins and minerals in low-income, food-deficit countries [2]. Being well adapted to marginal conditions, they may serve as storehouses of vital genes related to biotic and abiotic stress tolerance. Developing genomic resources and characterizing these important agronomic traits would help in identifying genes which could potentially be used in targeting other legumes to increase their tolerance.

Mucuna Adans. comprises 105 species [3] and is classified within the phaseoloid clade of Fabaceae, which also includes soybean, common bean, mung bean and other relatives [4]. *Mucuna pruriens* (L.) DC. (velvet bean) is reported to be native to China and East India [5] but has now attained a pantropical distribution with a major niche in the Indian subcontinent [6]. Like common bean and mung bean, velvet bean has a chromosome number of 2n = 2x = 22 [7], but with a much larger estimated genome size of between 1,281 to 1,361 Mbp/C (A.N. Egan & N. Sathyanarayana, unpublished data). The plant exhibits a climbing habit, hairy aerial parts and a long inflorescence of white or dark purple flowers. Pods are mostly green or brown in color with 4-6 seeds. In wild plants, the pods are thickly covered with soft or stiff orange bristles that cause intense allergic irritation to human skin in vars. *pruriens* and *hirsuta* whereas the cultivated var. *utilis* has non-irritant hairs. Botanically, it is represented by two varieties, var. *utilis* (cultivated) and var. *pruriens* (wild) - while the presence of a third group, var. *hirsuta* (wild), is also reported [8]. Owing to a wide-ranging geographical and climatic distribution, the species exhibits rich phenotypic diversity, especially in the Indian subcontinent.

The proximate nutritional composition, total protein content and in vitro protein digestibility of *M. pruriens* seeds are similar to other edible legumes [9]. Consequently, it is used as a minor food crop by native peoples of India and Africa while its use continues as livestock feed – a common use in the early 1900s in the USA [5]. It is known to produce seed yield of 2,000 kg/hectare [5], perform well under dry farming and low soil fertility conditions [10], exhibits allelopathic properties [11], and is effective in lowering pathogenic nematode populations [12]. Positive impacts of *M. pruriens* as a green manure cover crop are well documented in earlier studies [13]. The fast-growing habit of *M. pruriens* allows groundcover within 60-90 days, producing large biomass vis-à-vis other cover crops [12]. This, coupled with high nitrogen (N_2) fixing ability, has led the species to be regarded as a *"featured example of green manure's contribution to the sustainable agricultural system"* [5].

Seeds of *M. pruriens* contain high levels (1–9%) of L-Dopa (L-3,4 dihydroxy phenylalanine) [14, 15] - a precursor of dopamine used in the treatment of Parkinson's disease [16]. Daxenbichler et al. [17] screened 1000 species in 135 plant families and found only *Mucuna* spp. to contain sufficient L-Dopa for commercial use. Biochemically, L-Dopa is a non-protein amino acid produced as an intermediate product in the enzymatic synthesis of dopamine from L-tyrosine [18]. Although its efficacy for the treatment of Parkinson's disease is widely recognized [19], ingestion of large amounts of *Mucuna* plant parts, particularly its nutritionally rich seeds, is discouraged due to potential toxicity and associated side effects from long-term consumption of L-Dopa [20]. Nonetheless, it is has been reported safe to consume 500 g/day of *Mucuna*-based food with L-Dopa content ≤ 0.1% [15].

Notwithstanding these benefits, the agronomic potential of *M. pruriens* has remained largely underexploited. Much needs to be done in terms of breeding efforts especially to develop improved varieties not only for high or low L-

Dopa content, but also for enhanced nutritional value, resistances against biotic and abiotic stresses, and self-supporting, determinate cultivars. With the advent of genomic tools that can aid in developing molecular markers, genetic maps etc., the genetic improvement of underutilized crops has been greatly facilitated, enabling the development of improved genotypes or varieties with enhanced trait values [21, 22]. However, studies focusing on the development of genomic resources of *M. pruriens* are lacking, with only a few reports available on the use of molecular markers such as RAPDs and AFLPs (reviewed in [6]). In recent years, transcriptome sequencing has emerged as an efficient method to generate genomic-level data, large expressed sequence tag (EST) sequences, and molecular markers. Next generation sequencing (NGS) technologies are providing cutting-edge approaches for high-throughput sequence generation [22], allowing rapid and comprehensive analyses of plant genomes and cost-effective means of analyzing transcripts [23, 24]. Transcriptome sequencing has been successfully used for marker development in many underutilized legume plants, such as faba bean [25], adzuki bean [26], rice bean [27], hyacinth bean, grass pea and Bambara groundnut [28], and winged bean [28, 29], but has not yet been applied to research on *Mucuna*.

The present study reports the first transcriptome sequencing of *M. pruriens* genotypes. The objectives of this study were to (a) generate ESTs through whole transcriptome sequencing of two *M. pruriens* accessions; (b) develop and annotate a *de novo* transcriptome assembly; (c) discover and validate polymorphic microsatellite markers; (d) compare transcript expression in leaves, roots, and pods; and (e) perform genetic diversity and population structure analyses in a set of germplasm accessions for simple sequence repeat (SSR) marker validation. This study sets the stage for future molecular breeding, population and adaptation genomic studies, and provides a valuable resource for ongoing research into this agronomically and medicinally important legume species.

Materials And Methods

Plant material & RNA isolation

Two accessions representing different botanical varieties of *M. pruriens* and exhibiting contrasting phenotypes (Additional file 1: Table S1) were chosen for the study: *M. pruriens* var. *utilis* (IC0620620; collector's ID: 500108KA) is a cultivated accession and *var. pruriens* (IC0620622; collector's ID: 500113MH) is a wild accession. Accessions are available in the germplasm center of India. The plants were grown in the greenhouse facility of Sikkim University, Gangtok, India and a mapping population was developed. Young leaf, root and immature pod tissues were harvested for RNA isolation. Leaf, root, and pod tissues were chosen to maximize the number of genes expressed across tissues of different developmental processes and involved in key metabolic processes such as photosynthesis and respiration to 1) enable comparison with other transcriptomes utilizing the same tissue types (esp. Leaf) and 2) maximize sequencing across a range of developmental and metabolic processes.

Total RNA was isolated from each of the tissues using the method described by Ghawana et al. [30]. Pooled samples for both the accessions were prepared by combining equimolar concentration of total RNA for each of the tissues. Total RNA quality was assessed using NanoDrop-ND 2000C spectrophotometer and bioanalyzer. Samples with RIN (RNA integrity number) greater than 8.0 along with the ratios of 1.9-2.1 (260/280) and the ratios of 2.0-2.5 (260/230) were selected for sequencing.

Synthesis of cDNA library and Illumina sequencing

RNA-seq library preparation and sequencing was carried out at Next Generation Genomics Facility (NGGF), Centre for Cellular and Molecular Platforms (C-CAMP), Bangalore, India. For tissues extracted from accession IC0620620, tissues were pooled prior to sequencing by combining equimolar concentration of total RNA for each of the tissues and sequenced as a single library. Four paired-end cDNA libraries were generated representing each tissue for IC0620622. The paired-end 2x100 bp library preparation was done following the protocol of the Illumina TruSeq RNA sample preparation kit (Illumina Inc.) as per the manufacturer's instruction. One paired-end cDNA library was separately generated from the pooled RNA samples of genotype IC0620622. In total, five libraries were prepared and sequenced on a single lane of a 2×100 paired-end run by Illumina HiSeq^{TM} 1000.

De novo assembly and redundancy removal

To obtain a robust overview of the transcripts in the *Mucuna* species, we generated a *de novo* assembly combining the filtered reads of the two accessions. Raw reads were filtered using quality value (Q) ≥30 and demultiplexed using an option of one mismatch in index. A total of 191 million reads from both the genotypes were used to develop the combined *de novo* assembly (referred to subsequently as 620–22). We used Trinity [31] with default parameters and a minimum contig length of 200 bp for assembly generation. To generate non-redundant transcripts, highly similar fragments were clustered using CD-HIT v. 4.6 [32] with 95% identity as cut-off, resulting in a total of 67,561 transcripts from the total of 72,561 Trinity assembled transcripts. Separate genotype-wise assembly was not performed at this stage due to the six-fold difference in the number of reads

obtained between the two genotypes (1:6 relative ratios between IC0620620: IC0620622).

Annotation and gene ontology

For protein functional annotation, transcripts longer than 200 bp were searched against non-redundant protein databases of NCBI, Swiss-Prot and Uniref90 using BLASTX with an E-value cut-off of $1e^{-05}$. BLASTX searches were also performed on the assembled transcripts using the legume database (http://plantgrn.noble.org/LegumeIP) and only the top hits were considered. The Annocript v1.1.2 pipeline [33] was employed to obtain Gene Ontology (GO) terms for describing biological process, cellular components, and molecular functions. The Enzyme Classes (EC) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were also annotated using the Annocript pipeline.

Sequence similarity with other legume species

To compare the complement of genes characterized in the *M. pruriens* transcriptome assembly against gene assemblies in other legume species, protein sequences (*Medicago truncatula* Gaertn. Mt 4.0; *Glycine max* (L.) Merr. release 1.1; *Lotus japonicus* (Regel) K. Larsen release 2.5; *Phaseolus vulgaris* L. release 1.0; *Cicer arietinum* L. and *Cajanus cajan* (L.) Millsp. version 5.0) were downloaded from the NCBI database. BLASTX searches were performed on the 620–22 transcripts with an E-value cut-off of $1e^{-05}$, and the top hit for each transcript was used for further analysis.

Mining transcription factor families

For mining transcription factor gene families, we downloaded the plant transcription factors database (PlnTFDB) 3.0 (http://plntfdb.bio.uni-potsdam.de/v3.0) [34] and queried *M. pruriens* transcripts against the PlnTFDB using BLASTX with an E-value cut-off of $1e^{-05}$.

Mining of SSRs and detection of polymorphic SSRs

SSRs were searched in the 620-22 transcripts using the Perl script MISA (http://pgrc.ipk-gatersleben.de/misa) [35]. The microsatellite unit size and minimum number of repeats assigned was as follows: mono-nucleotide repeats more than 10 times, di-nucleotide repeats more than 6 times, tri-, tetra-, penta-, hexa-nucleotide repeats more than 5 times. The program was run up to deca-nucleotide repeats, but the results presented here are up to hexa- repeats only. For detecting the polymorphic SSRs between the mapping population parents, reads from each of the two genotypes were mapped to the 620-22 assembly. We used lobSTR v 3.0.3 [36] to identify polymorphic SSRs. This program provides a unique advantage over the conventional MISA pipeline as it can simultaneously compare two or more samples during the alignment process. We built a custom STR reference and the raw reads were passed through the program to be aligned around the SSR regions. BAM files from the alignment were sorted and indexed using samtools. Sorted BAM files were genotyped using the program allelotype within lobSTR. LobSTR identified 3,865 polymorphic SSRs, 787 of which were retained after selecting those having a quality value ≥10 and the alternative allele present in one of the samples. Of the 787 polymorphic SSRs, a subset of 134 SSRs was selected randomly and primers were designed using Batch-Primer3 (http://probes.pw.usda.gov/batchprimer3/) [37].

Genic SSR amplification and validation

A total of 25 *M. pruriens* accessions (Table 1) representing different geographical locations in India (Fig. 1) were selected from the Sikkim University germplasm collection for marker validation using the 134 EST-SSRs. The collection comprised representative taxa from all the three botanical varieties of *M. pruriens* viz. var. *utilis* ($n = 6$), var. *pruriens* ($n = 13$), and var. *hirsuta* ($n = 6$). Genomic DNA was isolated from young leaves using a modified cetyltrimethylammonim bromide (CTAB) method [38]. PCR amplification was performed in a final volume of 25 μl containing 50 ng/μl of template DNA, 1X PCR buffer, 1.5 mM $MgCl_2$, 2.5 mM dNTPs, primers (1 μM each) and 1U of Taq polymerase. The PCR conditions were as follows: initial denaturation at 94°C for 3 min followed by 35 cycles of 30s at 94°C, 30s at the annealing temperature (*Tm*) and 20s at 72°C with a final extension of 7 min at 72°C. The amplification was visualized using a UV illumination gel documentation system (Uvi-Tech DOL-008.XD, England). Subsequently, PCR products from different dye-labeled primers were pooled in equal volumes and 1.0 μl each of amplicons were mixed with 7 μl of formamide, 0.05 μl of the GeneScan™ 500 LIZ® Size Standard (Applied Biosystems, USA) and 2.95 μl of distilled water. DNA fragments were denatured and size fractioned using capillary electrophoresis on an ABI 3730 DNA Genetic Analyzer (Applied Biosystems, USA).

We applied stringent filtering criteria (minor allele frequency > 0.05 and missing percentage less than 20% for each marker band) for the marker bands produced, which resulted in 52 primer-pairs and 125 high-quality alleles/bands for all the downstream analysis. For data analysis, each marker-band was scored as a dominant marker with presence of a band in a genotype indicated as "1" and absence "0." Further, the genotyping data were converted into bi-allelic format (e.g., 1 denoted as AA and 0 as GG) and a hapmap file was generated. The hapmap file was converted into Variant Call Format (VCF) using TASSEL v 5.2.29 [39]. The quality control on the dataset was performed using VCFtools [40]. Markers with minor allele frequency (MAF) >0.05, max-missing

Table 1 Details of accessions used for EST-SSR validation

Sample Number	Accession number/ Collector ID	Variety	Latitude (N)	Longitude (E)	Altitude (AMSL) (m)	State of Origin
1	500101KA*	var. *utilis*	13°14′	77°62′	911	Karnataka
2	IC0620620**	var. *utilis*	13°14′	77°62′	911	Karnataka
3	IC0620622**	var. *pruriens*	20°00′	73°77′	745	Maharashtra
4	500120TN*	var. *hirsuta*	9°55′	78°07′	138	Tamil Nadu
5	IC0620624**	var. *pruriens*	14°48′	74°12′	7	Karnataka
6	500136TN*	var. *hirsuta*	10°04′	77°45′	298	Tamil Nadu
7	500147AP*	var. *hirsuta*	18°39′	78°10′	383	Telangana
8	500154AP*	var. *hirsuta*	16°04′	78°52′	434	Andhra Pradesh
9	500186MH*	var. *hirsuta*	19°09′	77°27′	373	Maharashtra
10	500192OR*	var. *pruriens*	20°18′	85°62′	63	Odisha
11	500193OR*	var. *pruriens*	21°94′	86°72′	51	Odisha
12	500194OR*	var. *pruriens*	21°94′	86°72′	51	Odisha
13	500195OR*	var. *pruriens*	21°63′	85°58′	650	Odisha
14	500196OR*	var. *pruriens*	20°47′	85°12′	121	Odisha
15	500197WB*	var. *pruriens*	26°71′	88°43′	125	West Bengal
16	500199WB*	var. *pruriens*	26°70′	88°80′	65	West Bengal
17	500202TN*	var. *hirsuta*	12°57′	79°56′	43	Tamil Nadu
18	500210MN*	var. *utilis*	25°41′	94°47′	782	Manipur
19	500211NL*	var. *utilis*	25°67′	94°12′	1333	Nagaland
20	500212AS*	var. *pruriens*	26°11′	91°44′	61	Assam
21	500217MN*	var. *utilis*	25°68′	93°03′	776	Manipur
22	500219TR*	var. *pruriens*	23°50′	91°25′	64	Tripura
23	500221AR*	var. *pruriens*	27°08′	93°40′	1035	Arunachal Pradesh
24	500224AR*	var. *pruriens*	27°08′	93°40′	296	Arunachal Pradesh
25	500267NL*	var. *utilis*	25°68′	94°08′	1360	Nagaland

*Collectors ID of newly collected accessions; **National genebank ID

sites <20%, and accessions with maximum missing site <20% were retained. Estimates of expected heterozygosity (He), observed heterozygosity (Ho), effective number of alleles (Ne) [41], polymorphism information content (PIC), gene diversity (h) [42], and Shannon's Information index (I) [43] were calculated using the software GenAlex 6.5 [44] and POPGENE [45]. Phylogenetic tree construction using the neighbor-joining (NJ) algorithm and principal component analysis (PCA) were performed using TASSEL. The resulting tree was visualized as a mid-point rooted tree using FigTree v1.4.2 [46]. A plot of PC1 versus PC2 was made in R (http://www.r-project.org/) [47] using the ggplot2 package (ggplot2.org) and the geom_text_repel function in ggrepel (http://github.com/slowkow/ggrepel) to plot accession names. Population structure was investigated using a Bayesian framework implemented in the program fastSTRUCTURE [48] with the following commands: –prior simple –full –seed = 100 –cv = 5 for subgroups 2 to 10 (K = 2 to 10). The output was investigated with "choosing model complexity" script included in the program fastSTRUCTURE and by plotting marginal likelihood and cross-validation error against the number of subgroups to determine the possible range of subgroups. The range of subgroups identified were then inspected with proportion of membership of genotypes to respective subgroups (coefficient of ancestry values) and known geographical origin information to determine the precise number of subgroups in the collection. The results were visualized using a plot made in R with ggplot2, reshape (http://had.co.nz/reshape) and RColorBrewer (http://colorbrewer2.org) packages.

Expression analysis

For analyzing differentially expressed transcripts, raw reads obtained from transcriptome sequencing of leaf, root, and pod tissues of the *M. pruriens* var. *utilis* (IC0620620) were aligned separately to the 620-22 assembly using bowtie aligner version 1.1.1 [49]. We normalized the gene expression level in each library to

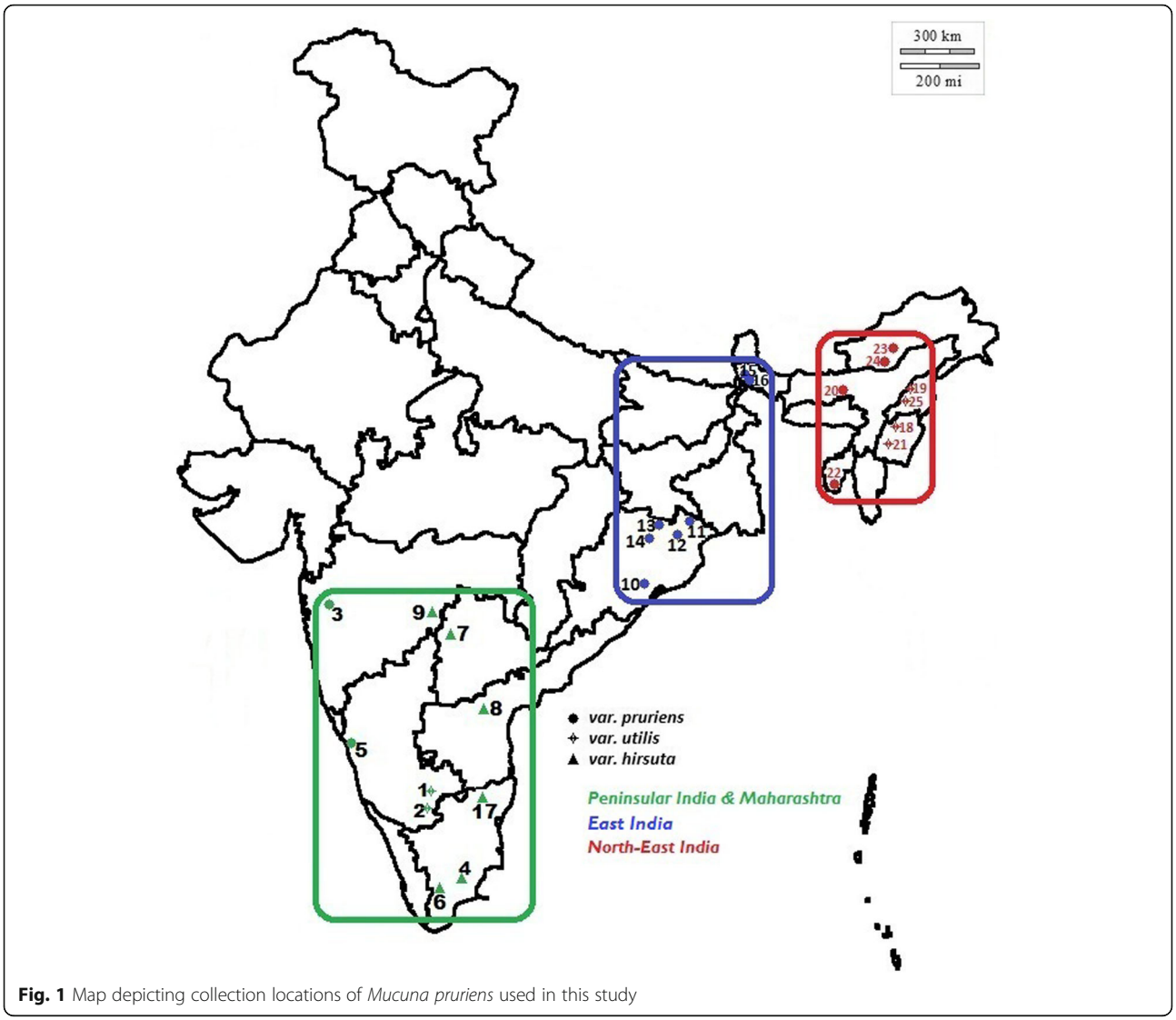


Fig. 1 Map depicting collection locations of *Mucuna pruriens* used in this study

produce an effective library size for use in calculating read counts. We used RSEM version 3.0 [50] to calculate the read count and the estimated expression levels as fragments per transcript kilobase per million fragments mapped (FPKM) using edgeR software in R [51]. A dispersion value of 0.1 was used in the expression analysis. Differentially expressed transcripts (DETs) were determined with log-fold expression change ≥4 and a statistically significant p-value of 0.001. Pair-wise comparisons among the three tissues were conducted by comparing the sequenced samples to identify common DETs across all the tissue types. The top fifty DETs from pair-wise comparisons across tissues were extracted to generate a heat map and demonstrate the dynamic expression patterns in different tissues. In addition, we also extracted the top differentially expressed transcripts from secondary metabolite biosynthesis classes based on annotation to determine expression patterns related to secondary metabolite pathways.

Results

Sequencing and de novo assembly

Illumina sequencing generated 18.24 GB of data containing 167,986,452 and 27,801,324 raw reads for genotypes IC0620620 and IC0620622, respectively (Table 2). The combined assembly 620-22 produced 72,561 transcripts. After clustering, 67,561 transcripts were retained with an N50 length of 987 bp and a mean transcript length of 641 bp (Table 3; Additional file 1: Figure S1).

Functional annotation and characterization of *M. pruriens* transcripts

A total of 49,925 (73.9%), 35,535 (52.6%) and 54,450 (80.6%) transcripts showed significant hits with NCBI-NR, Swiss-Prot and UniRef proteins, respectively, with

Table 2 Summary of data generated for *Mucuna pruriens* transcriptome. G1 is *Mucuna pruriens* var. *utilis* (IC0620620; collector's ID: 500108KA); G2 is *M. pruriens var. pruriens* (IC0620622; collector's ID: 500113MH)

Sample	fastq file size (GB)	Total number of paired end reads	Total number of reads after quality filtering
G1 Leaf	1.86	19,406,426	18,997,424
G1 Pod	5.42	58,585,008	57,166,422
G1 Root	2.69	28,623,354	28,046,508
G1 Pooled	5.68	61,341,664	59,885,295
G2 Pooled	2.59	27,801,324	27,137,593
Total	18.24	195,757,776	191,233,242

34,686 transcripts having conserved domains and 6,248 with hits against the Rfam database. Broadly, the putative orthologs of genes involved in various pathways and cellular processes were found to be conserved in *M. pruriens*. Further, GO terms were assigned to *M. pruriens* transcripts that showed significant similarity with annotated proteins from other plant species (Additional file 2). A total of 30,575 (45.3%) transcripts were assigned at least one GO term in the biological process category, 46,961 (69.51%) in the molecular function category and 30,199 (44.70%) in the cellular component category. Among the various biological processes, genes coding for proteins involved in transcription (3.56%) and transcription regulation (3.50%) were highly represented. The genes involved in other important biological processes such as carbohydrate metabolism, signal transduction, response to stress, transport, cell wall organization and protein folding were also identified through GO annotations. Similarly, ATP, DNA as well as different metal ion binding activities were most represented among the molecular functions; and integral membrane, nucleus and cytoplasm related activities were most represented among the cellular component categories (Fig. 2).

Table 3 Statistics of non-redundant set of *Mucuna pruriens* transcripts obtained from Trinity assembly

Total number of assembled bases	46,525,999
Number of transcripts	72,561
The total number of transcripts after clustering	67,561
The mean sequence length	626
Average % of N	0.00
Average % of GC content	44.58
N50	987
Maximum transcript length	17,978
Average transcript length	641
Number of putative non coding sequences	1,493
Length of the longest ORF (bp)	2,362
Number of ORFs ≥ 100 bp	36,228
Number of ORFs on plus (+) strand	36,421
Number of ORFs on minus (-) strand	31,140

Enzyme classes were obtained for 3,963 assembled sequences, whereas associated KEGG classification was obtained for 3,492 assembled sequences (Additional file 3). The top 20 abundant enzyme classes observed for the *M. pruriens* transcriptome are listed in Additional file 1: Figure S2A. The greatest number of assembled transcripts belonged to the serine/threonine protein kinase enzyme class (38.4%). Besides this, Additional file 1: Figure S2B displays the top 20 KEGG pathways represented by the assembled transcriptome sequences. The highest number of sequences belonged to protein modification pathways (37.5%) followed by lipid metabolism and glycan metabolism. As evident from the results, the highest represented groups included several pathways associated with housekeeping processes as well as plant development and secondary metabolism.

Sequence similarity with other legume species

A comparison of assembled transcripts against proteomes of chickpea, pigeon pea, soybean, common bean, mung bean, garden pea, barrel medic and Lotus showed that 58,208 of 67,561 transcripts (86.2%) from the 620–22 assembly had significant similarity to sequences in one or more legumes (Additional file 4). About 71% of these transcripts had ≥ 70% sequence identity (Additional file 1: Figure S3A). The largest number of *M. pruriens* transcripts showed significant similarity with soybean transcripts followed by *Medicago*, *Phaseolus, Vigna*, *Cicer* and the least similarity with *Pisum* and *Cajanus* (Additional file 1: Figure S3B). The lack of strong correlation with taxonomy (e.g. greater similarity to *Medicago* than to the phaseoloid, *Cajanus*) is presumably due to varying quality of the different genome assemblies.

Mining transcription factor families

In total, 2,223 putative *M. pruriens* transcription factors distributed in at least 55 families were identified representing 3.29% of *M. pruriens* assembled transcripts (Additional file 5). Among these, the basic/helix-loop-helix (bHLH; 227), C2H2-type (151), MYB (146), MYB related (130), NAC (126) and WRKY (122) were among the top categories (Fig. 3). However, almost all the families showed minor species-specific differences in relation to TF gene families reported for *Cicer*, *Lotus*, *Medicago* and *Glycine* (Table 4).

Detection of genic SSR markers

We detected a total of 6,284 transcripts (Additional file 6) within which 7,943 potential EST-SSRs (Additional

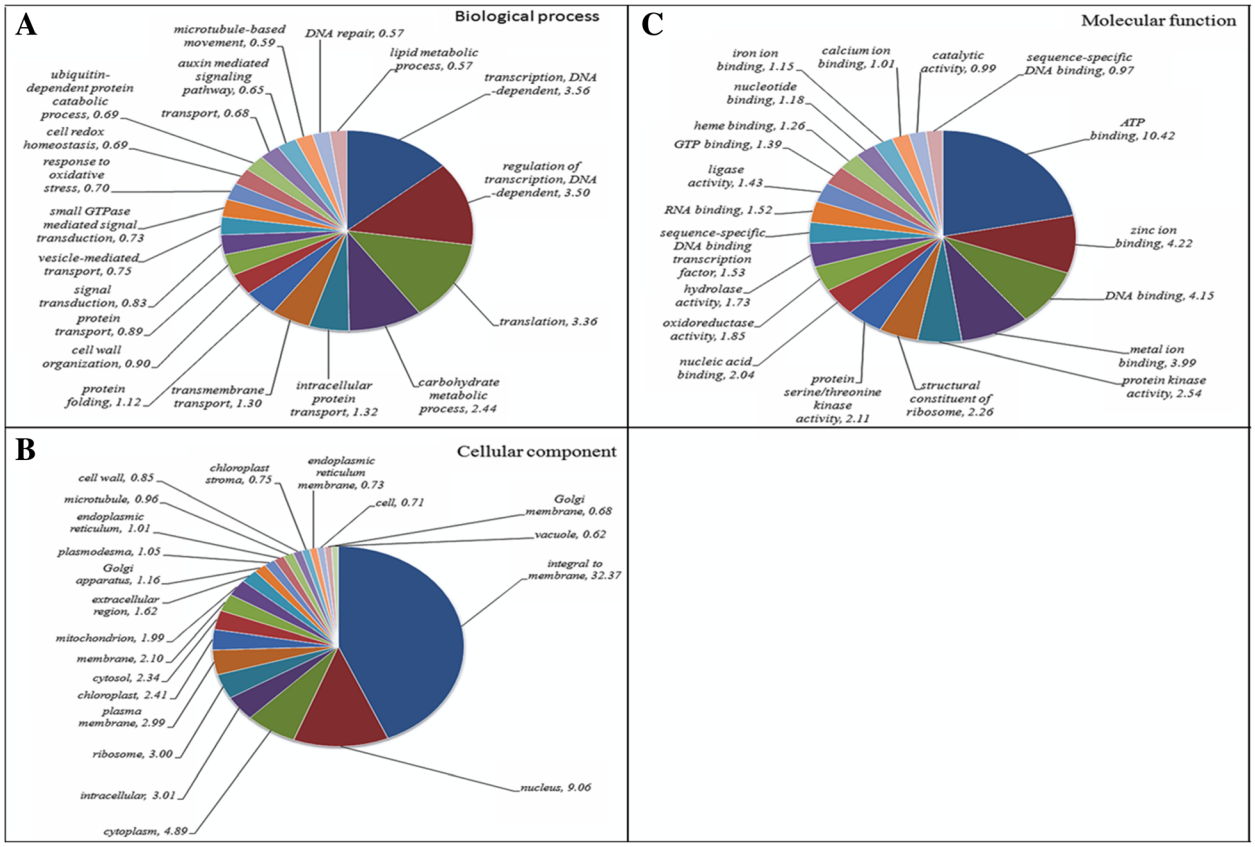


Fig. 2 Functional annotation of Mucuna pruriens transcripts. Gene ontology term assignments to transcripts in different categories of **a** biological process, **b** cellular component, and **c** molecular function. Numbers are percentage of function for each major category

file 7) were discovered. The mono-nucleotide SSRs represented the largest fraction (3,638), with the vast majority (92%) comprising A or T repeats, which likely represent remnants of mRNA poly-A tails (Fig. 4). Only a small fraction of tetra-nucleotide (146), penta- (64) and hexa- (100) nucleotide SSRs were identified in *M. pruriens* transcripts (Table 5). Of the 6,284 SSR-containing sequences, 1,174 transcripts contained more than one SSR. Further, EST-SSRs with five tandem repeats were most common, followed by ten, six, eleven, seven, and twelve tandem repeats, whereas the remaining tandem repeats each accounted for less than 5% of the our EST-SSRs (Additional file 1: Table S2).

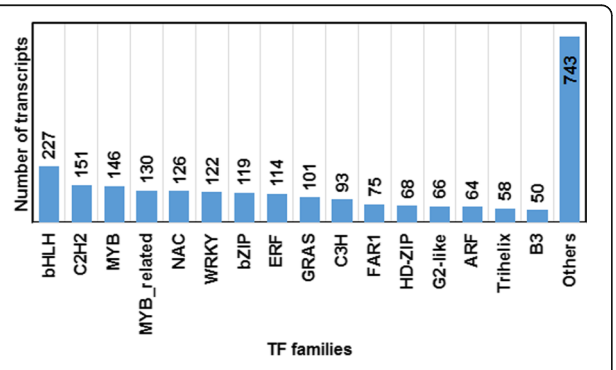


Fig. 3 Distribution of *Mucuna pruriens* transcripts in different transcription factor families

Table 4 Number of transcripts encoding for transcription factor families in *Mucuna pruriens* compared to other legumes. The data on *M. pruriens* is from our study; data for soybean, *Medicago* and *Lotus* is from Libault *et al* [69]; data for Chickpea is from Garg *et al* [70]

TF family	*M. pruriens*	Chickpea	Soybean	*Medicago*	*Lotus*
bHLH	227	488	393	71	64
AUX/IAA-ARF	64	216	129	24	36
C2C2-CO-like	16	15	72	15	21
C2C2-GATA	44	49	62	29	16
C2C2-YABBY	13	8	18	6	4
C3H	93	594	147	41	50
CAMTA	18	26	15	6	4
MYB	146	528	791	171	191
PHD	10	489	222	45	47

Screening for SSRs using lobSTR identified a total of 3,865 SSRs polymorphic between the parents (Additional file 8). Aligning IC0620622 (G2) reads against the combined assembly yielded 3,075 SSR calls, with 1,339 of these present in areas with greater than 5x coverage and a mean coverage of 9.54x. Alternatively, alignment of IC0620620 (G1; 500108) reads to the combined assembly yielded 3,517 SSR calls, 2,092 of which were present in areas with greater than 5x coverage and with a mean coverage of 19.15x. After filtering the SSRs based on the parameters mentioned earlier, we obtained a total of 787 polymorphic repeats (Additional file 8). The details of the SSR repeats from the lobSTR analysis are given in Additional file 1: Table S3 and the distribution of different SSR motifs from the lobSTR output is presented in Additional file 1: Figure S4.

Table 5 Statistics of SSRs identified in *Mucuna pruriens* transcripts

SSRs mining	
Total number of sequences examined	67,561
Total size of examined sequences (bp)	42,340,968
Total number of identified SSRs	7,943
Number of SSR containing sequences	6,284 (9.3%)
Number of sequences containing more than one SSR	1,174
Number of SSRs present in compound formation	963
Frequency of SSRs	One per 5.3 kb
Distribution of SSRs in different repeat types	
Mono-nucleotide	3,638 (45.80%)
Di-nucleotide	1,674 (21.07%)
Tri-nucleotide	2,240 (28.20%)
Tetra-nucleotide	146 (1.83%)
Penta-nucleotide	64 (0.80%)
Hexa-nucleotide	100 (1.25%)

EST-SSR validation and population structure

Of the 134 primer pairs selected, 98 (73.13%) successfully amplified the genomic DNA of 25 *M. pruriens* accessions. From this, consistently amplified marker-bands from 82 primer pairs were chosen for further analysis. Nearly 2,000 marker bands were amplified by these 82 primer pairs and 125 high-quality marker-bands representing 52 primer pairs across 23 accessions were retained after quality control (Additional file 1:Table S4). Accessions 500267NL and 500101KA were dropped due to >20% missing marker information. Various measures of genetic diversity for each primer pair are reported in Additional file 1: Table S4. Parameters of genetic diversity were also estimated between the population groups representing different geographical locations (3) and botanical varieties (3). The genetic diversity index (h) between different geographical locations and botanical varieties ranged from 0.35 to 0.37 and 0.34 to 0.36 with mean values of 0.35 and 0.35 respectively (Table 6), whereas h was much lower for genetic subgroups, ranging from 0.16 to 0.21. In all groups, the total gene diversity, Ht, was higher than the gene diversity within the groups, Hs. The coefficient of gene differentiation (Gst) was 0.04, which indicated very less genetic differentiation among different population groups as compared to within group variations. Gene flow indices (Nm) were relatively high, ranging from 1.83 to 4.09.

Population structure analysis using fastSTRUCTURE on the 23 genotypes suggested the presence of 4 to 6 subgroups using the "choosing model complexity" script, and 4 or 8 subgroups based on likelihood score (Additional file 1: Figure S5), with K = 4 being the most probable. For K = 4 subgroups, 21 genotypes had >80% proportion of membership to a respective subgroup as determined by the coefficient of coancestry value of each genotype (Fig. 5a). Subgroup 1 is composed of nine individuals, mostly of var. *pruriens* ($n = 5$) and only one var. *utilis*, with mixed representation from peninsular and northeast India (each $n = 4$) but only a single

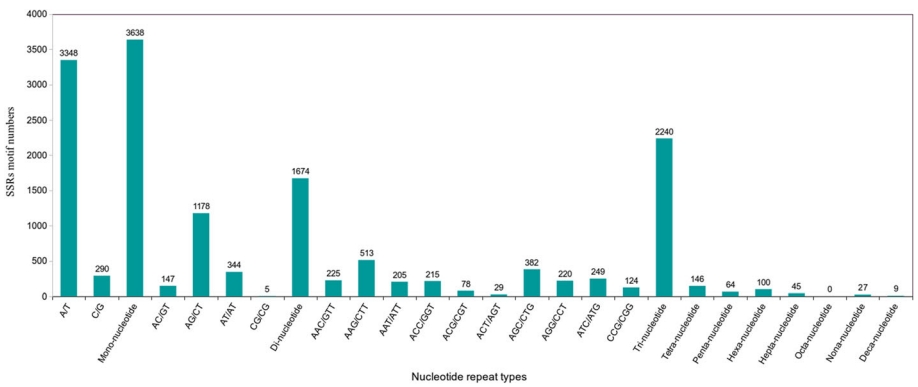


Fig. 4 Simple sequence repeat length distribution across different motif classes in *Mucuna pruriens* transcriptome

Table 6 Gene diversity estimates for groups based on botanical varieties, geographical distribution and population structure analysis

	Population group	Na	Ne	I	h
Geographical distribution	East India	2.23	1.78	0.59	0.37
	North East India	2.21	1.68	0.52	0.36
	Peninsular India	2.98	1.95	0.72	0.35
		Ht	Hs	Gst	Nm
	Mean	0.41	0.36	0.04	4.09
	SD (±)	0.19	0.18		
Botanical varieties	var. *pruriens*	2.67	1.83	0.64	0.36
	var. *hirsuta*	2.46	1.78	0.60	0.36
	var. *utilis*	2.10	1.84	0.59	0.34
		Ht	Hs	Gst	Nm
	Mean	0.43	0.36	0.04	2.57
	SD (±)	0.19	0.17		
Population groups based on K = 4 sub grouping	SG1	2.54	1.91	0.66	0.19
	SG2	1.87	1.53	0.41	0.16
	SG3	2.36	1.77	0.59	0.21
	SG4	2.10	1.84	0.60	0.18
		Ht	Hs	Gst	Nm
	Mean	0.41	0.34	0.04	1.83
	SD (±)	0.19	0.17		

Na- Number of alleles; Ne- Effective no. of alleles [41]; I- Shannon information content; h- Nei's gene diversity [42]

accession hailing from eastern India. Subgroup 2 is exclusively composed of var. *pruriens* from eastern ($n = 4$) and northeastern ($n = 1$) India. Subgroup 3 is exclusively from peninsular India with most accessions of classified as var. *hirsuta* ($n = 3$), but with one each from the other varieties. Subgroup 4 is the most heterogeneous group and includes two var. *pruriens* accessions from eastern India and two var. *utilis* accessions from northeast India.

The genotypes in the NJ tree and PCA are color coded from the information of 4 subgroups in the fastSTRUCTURE. NJ and PCA revealed similar clustering of genotypes as fastSTRUCTURE with at least three out of the four groups clustering largely according to their geographical origin (Fig. 5). Principal component 1 (PC1; Fig. 5b) accounted for 12.8% of genetic variability and separated the majority of northeastern and eastern accessions from peninsular + Maharashtra accessions, with five northeastern or eastern accessions falling on the left side. No peninsular accessions fell right of PC1. Furthermore, all *hirsuta* and all *utilis* accessions but one were found left of PC1. The *utilis* accession (500217MN, northeast India) is one of two individuals that could be interpreted as having a hybrid ancestry, evidenced by a ~50% split assignment to subgroup1 and subgroup 4, the other accession being IC0620620 (var. *utilis*, peninsular India) and assigned to subgroups 1 and 3 with near equal probability. Subgroup 2 is entirely separated from subgroup 3 by PC1. Principal component 2 (PC2; Fig. 5b) accounted for another 10.1% of genetic variability. Subgroup 1, the largest and most diverse subgroup, clustered exclusively below PC2.

The neighbor joining algorithm produced two main clades (Fig. 5c), one equating to a similar grouping separated in PCA by PC1 which is also the same as subgroup 2 plus two accessions from subgroup 4. This clade includes only var. *pruriens* from the northeastern and eastern areas with the exception of the potential hybrid 500217MN, var. *utilis*. The second clade includes subgroups 1 and 3 plus two accessions of subgroup two, with no clear clustering based on variety.

Expression analysis

A total of 4,387 transcripts were differentially expressed among three tissues of IC0620620 with log-fold expression change ≥4 and a statistically significant p-value of 0.001, of which 1,897 were commonly expressed in all three tissue types; 191 to 372 were shared between tissue types and 25 to 1489 were unique to tissues, respectively (Fig. 6a). In leaf, 1,606 transcripts exhibited up-regulation and 2,361 transcripts exhibited down regulation as compared to roots, followed by 182 up-regulated and 550 down-regulated against the pod tissue. Similarly, pairwise comparison of pods and roots exhibited 555 and 946 transcripts were up-and down- regulated respectively (Fig. 6b). The top fifty differentially expressed transcripts in each of the three tissues are visualized to show varying expression

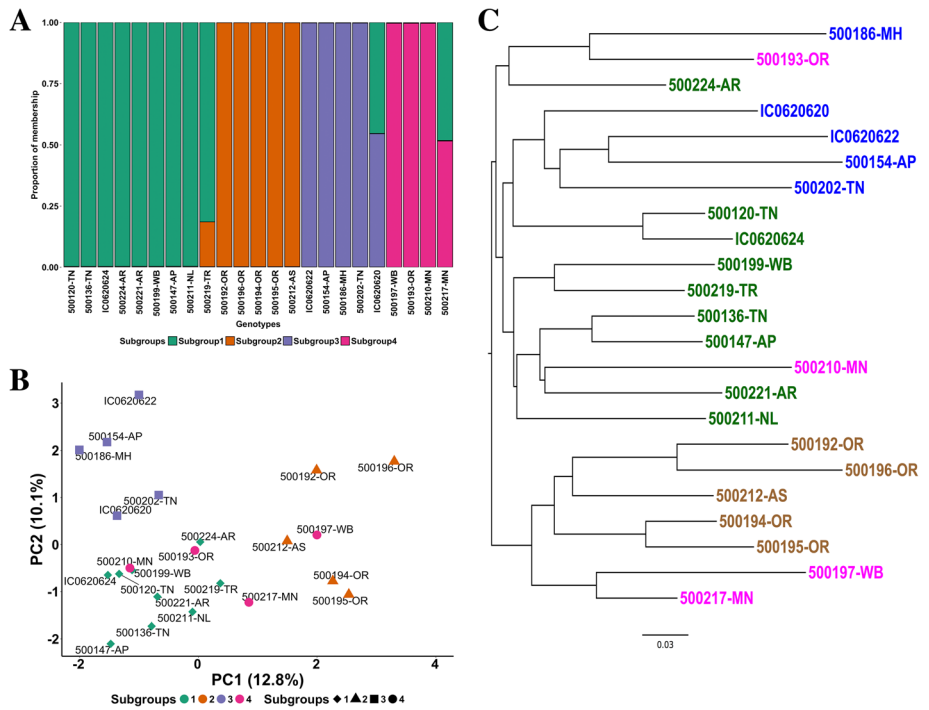


Fig. 5 Population Structure analysis of the 23 Indian *Mucuna pruriens* accessions. **a** Bayesian clustering (fastSTRUCTURE, *K* = *4*); **b** Scatter plot from principal component analysis (PCA); **c** Neighbor-joining tree generated for all accessions

patterns (Additional file 1: Figure S6). Among the differentially expressed transcripts, 223 were found to encode for TFs representing 43 different families, including MYB, MADS, WRKY, and bHLH families, some of whose members are involved in secondary metabolite biosynthesis. We also investigated the expression of other genes involved in secondary metabolism/biosynthesis and observed that the top 47 transcripts in this category showed varying differential expression patterns (Fig. 6c). Among these, highly expressed transcripts included those belonging to flavonoid, isoprenoid, phenylpropanoid, and wax pathways (Additional file 9).

Discussion

Developing genetic resources for *Mucuna* through transcriptomics

The legume family is second only to the grass family (Poaceae) in economic importance, with a number of species utilized as crops, fodder, industrial agents, construction materials, and medicines [52]. The genomes of legume species including soybean [53], the common bean (*Phaseolus vulgaris* L.) [54], cowpea (*Vigna unguiculata* (L.) Walp.) [55], *Medicago sativa* L. [56], *Lotus japonicus* L. [57], pigeonpea (*Cajanus cajan* (L.) Millsp.) [58], and Lupin (*Lupinus angustifolius* L.) [59] have been sequenced, providing critical genomic resources across Fabaceae. However, a number of economically important legumes, including orphan crop species or those used for purposes other than human consumption, are in need of efforts to develop genetic and genomic resources to act as intellectual infrastructure upon which to build a broader scientific future. To our knowledge, we present the first *de novo* transcriptome assembly described for *Mucuna pruriens*, the velvet bean, an orphan legume crop used both for human consumption and medicine, providing an important genetic resource base for future genetic studies and breeding efforts.

Mucuna is a monophyletic genus [60] that represents an early-branching, distinct evolutionary lineage within the phaseoloid legumes. *Mucuna* is variously allied with monotypic *Haymondia wallichii* (DC.) A.N.Egan & B.Pan bis [4, 61] or with tribe Desmodieae [4, 62–64]. The addition of our transcriptomic data representing the *Mucuna* lineage thus fills a void within the developing resource base of transcriptomic data available for comparative evolutionary studies across legumes.

Within our species-level transcriptome assembly, we recovered 67,561 transcripts and annotated over 86% of these against one or more legume proteomes (Additional file 1: Figure S3), presenting a large collection of expressed genes that can be used in downstream analyses for genetic study and crop improvement programs. Our total number of transcripts recovered is quantitatively similar to a number of recent legume transcriptomic studies. As an example, Ištvánek et al. [65] recovered 64,761 transcripts from red clover (*Trifolium pratense* L.), ~73% of which were annotated. From a

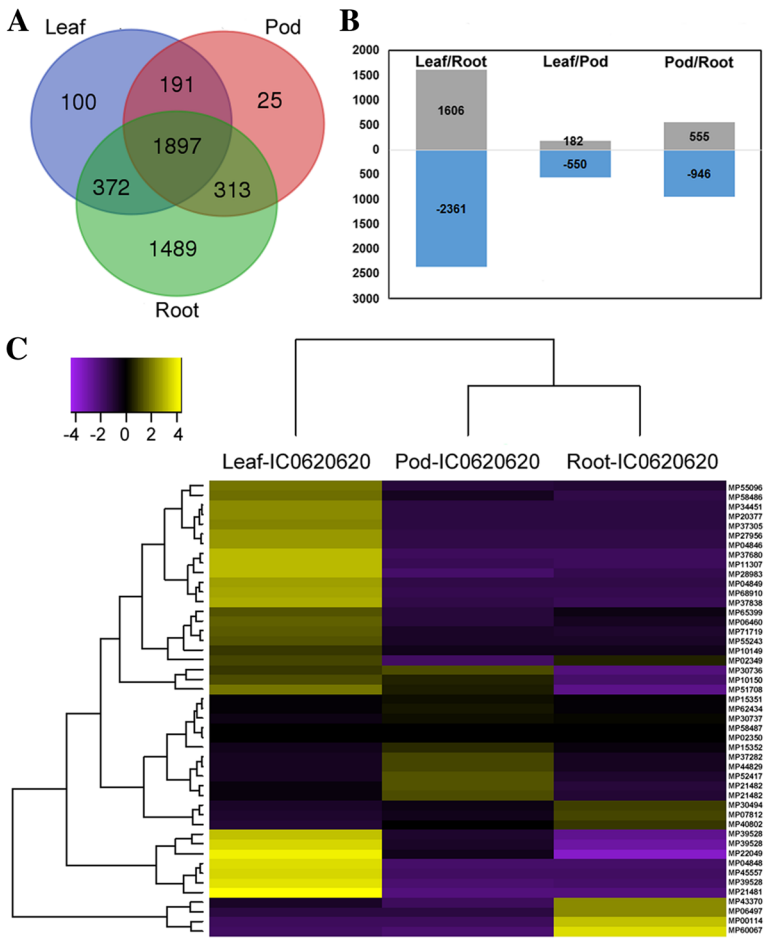


Fig. 6 Differential transcript expression in leaf, root, and pod tissues. **a** Diagram showing overlap of genes between leaf, root, and pod tissues showing differential transcript expression. **b** Pairwise comparisons across tissues showing differentially expressed transcripts. Those above the line are transcripts up-regulated and those below are down-regulated within the pairwise comparison. **c** Heat map of secondary metabolite associated differentially expressed genes of leaf, pod, and root transcriptomes. The various shades in the boxes showed similar tendencies of gene expression. Labels along the right side correspond to transcript names (see Additional file 9)

genomic perspective, our complement of annotated genes may seem high for a diploid species. The common bean, *Phaseolus vulgaris* L., is estimated to have ~28,000 coding genes [54], Chinese licorice, *Glycyrrhiza uralensis* Fisch., has ~ 35,500 genes [66], whereas soybean, *Glycine max* (L.) Merr., a recent polyploid crop [67], has between ~46,000 and 56,000 protein-coding genes (*Glycine max* Wm82.a2.v1 build at phytozome.jgi.doe.gov; [53]). The higher number of annotated transcripts in our study may be due to the presence of multiple isoforms, alternatively transcribed transcripts, and/or portions of genes that did not completely assemble with the settings we used. Our higher number of annotated genes could also include an increased complement of coding genes necessary for plants that produce a high degree of secondary metabolites [68].

Transcription factors (TFs) play important roles in gene regulation and function. We assessed the number and distribution of TF gene families within *M. pruriens* (Fig. 3; Table 4) and found that the overall distribution of transcription factor encoding transcripts among the various known protein families is similar to that of soybean and other legumes [69]. Transcription factors constituted ~3.3% of annotated transcripts, a number similar to that in *Medicago truncatula* and *Lotus japonicus* but slightly higher than that of *Psophocarpus tetragonolobus* (Table 4, Additional file 5; [29]) and considerably lower than that estimated for soybean [54] or chickpea [70]. Soybean likely has a higher number of TFs than most legumes due polyploidy. The MYB and basic-Helix-Loop-Helix (bHLH) TF gene families are two of the most prevalent TF gene families in our data, both of which play important roles in secondary metabolite biosynthesis, particularly for flavonoid and anthocyanin compounds [71]. In terms of differences within TF gene families of *Mucuna* relative to other legumes, we noted several events of expansion (e.g. C2C2-

GATA, CAMTA) and contraction (e.g. PHD), evidence of the evolving nature of TF across lineages within legumes.

Differential transcript expression across tissues within *Mucuna*

Differential transcript expression analysis performed by pair-wise comparison among the three tissues of genotype IC0620620 (Fig. 6 and Additional file 9) found thousands of differentially expressed transcripts across leaves, roots, or pods. Transcripts highly expressed in one tissue versus others suggest tissue-preferred expression, which can be helpful for further studies (Additional file 1: Figure S6). For example, transcripts involved in anthocyanin biosynthesis were upregulated in leaves relative to pods or roots (Fig. 6c). Uniquely expressed genes in leaf, root or pod tissues will be of importance to understand their contributions towards economically important traits in *Mucuna*. Genes involved in secondary metabolism are especially important as *Mucuna* is a natural source of mucunain and serotonin, chemicals found in pod hairs that promote itching [72], and of high levels of L-dopa found in seeds [73]. L-dopa is the precursor to dopamine, norepinephrine, and epinephrine (adrenaline), important neurotransmitters in the brain, and is widely used in the treatment of Parkinson's disease [74]. Recent studies have shown that some Parkinson's patients better tolerate taking ground *Mucuna* seeds as a source of L-DOPA and that this natural source may be more effective and neuroprotective than L-DOPA itself while lessening adverse side effects [19]. We have initiated further investigation on the expression of genes specific to L-Dopa and other important secondary metabolites in this plant to gain further understanding on the regulation of genes involved, the results of which will be reported in a secondary paper.

Tissue-preferred expression of transcripts involved in secondary metabolism has been found in other plants, especially those used for medicinal purposes (e.g. in citronella, *Cymbopogon winterianus* Jowitt [75]). Beyond L-Dopa, many other important secondary compounds have evolved within legumes [76]. For comparisons between root and leaf, those transcripts upregulated in the root relative to leaf transcriptomes were mostly related to secondary metabolism. A similar result was found for comparison of root and shoot tissues in *Leucaena leucocephala* (Lam.) de Wit [77]. In our study, the most highly upregulated gene involved in secondary metabolism in the root transcriptome was isoflavone reductase which exhibits a ~6-fold increase relative to pods and leaves (unigene MP60067 in Fig. 6c and Additional file 9). Isoflavone reductase is an enzyme unique to the plant kingdom involved in the isoflavonoid phytoalexin biosynthesis pathway and is suggested to play important roles in stress responses. Overexpression of isoflavone reductase in soybean was shown to enhance resistance to the oomycete *Phytophthora sojae* and induce antioxidant activity in the plant [78]. As root secondary metabolites have been less investigated in *M. pruriens*, this work may enable new areas of research and lead to discovery of novel secondary compounds of pharmaceutical interest.

Other transcripts related to secondary metabolites, such as anthocyanins, showed differential expression in leaf tissue relative to pods and roots. These plant pigments produce dark colors, particularly blue and purple, in above-ground plant tissues and also provide important antioxidant properties [79]. The strongest secondary metabolism-related transcript upregulated in the leaves was chalcone synthase, the first step in the phenylpropanoid pathway that leads to production of many flavonoid secondary metabolites, including anthocyanins [80]. NAD(P)H-dependent 6' deoxychalcone synthase was moderately upregulated in the pod transcriptome and is an enzyme involved in synthesis of isoliquiritigenin, a secondary compound known primarily from licorice (*Glycyrrhiza spp.*), a related legume genus. Isoliquiritigenin provides a number of useful pharmacological properties such as anti-inflammatory, anti-viral, antimicrobial, and cardioprotective effects and has shown remarkable anti-cancer properties [81]. In combination with other transcriptomic and genomic resources, our transcriptome provides a useful resource for genetic studies related to secondary metabolites of medicinal application and interest.

Detection and validation of microsatellite (SSR) markers in *Mucuna pruriens*

The medicinal potential afforded by secondary chemistry within *Mucuna* has no doubt led to its popularity throughout India where it is a component in over 200 indigenous Ayurvedic drug formulations used against a wide range of disorders, such as menstrual discomfort, neurological issues, sexual dysfunctions, tuberculosis and even elephantiasis [14]. Consequently, *M. pruriens* is found throughout India in both cultivated and wild forms. Microsatellites, or SSRs, are excellent genetic markers to aid in construction of genetic linkage maps and association analysis. As such, development of a database of SSR markers known to be polymorphic within *M. pruriens* may be useful for future genetic improvements of this important medicinal plant. Within our transcriptome, we detected over 4,000 EST-SSRs of dinucleotide or higher repeats (Table 5; Fig. 4). Of these, tri-nucleotide repeats were the most abundant, a sensible result given the coding nature of the transcriptome [82]. Other legumes exhibit the same trend, including the winged bean [29] and peanut [83].

Certain repeat motifs were more prevalent than others in our ~3,800 EST-SSRs polymorphic between parents (summarized in Additional file 1: Table S3), a finding noted in other legumes previously [29, 84]. Within our polymorphic SSR set, $(AG)_n$, $(AAG)_n$, and $(AAAG)_n$ are most prevalent in each repeat class (Table S3), a bias that was first recognized in *Arabidopsis* [85]. Within our full set of detected SSRs, motif type $(AG/CT)_n$ comprises 70.4% of all di-nucleotide repeats, with the $(CG/GC)_n$ motif nearly nonexistent (Fig. 4). The bias towards AG and against CG repeats has been demonstrated across eukaryotes [86], including within other legumes such as *Phaseolus* [87] and winged bean [29]. Prior studies have suggested that AG repeat motifs are commonly found in 5' untranslated regions [84] and, as such, may be involved in transcription and regulation [85]. In the full set of detected SSRs, the $(AAG/GTT)_n$ repeat motifs and their complements are the most prevalent. The ranking of tri-nucleotide repeat classes closely mirrors that found in winged bean [29].

Validation of SSRs discovered via transcriptome sequencing is the next step to building a working marker set for genetic improvement efforts. Of the 134 primer pairs we screened, over 73% successfully amplified genomic DNA across 25 *Mucuna* accessions. Our success rate is comparable to or somewhat lower than other efforts to validate transcriptome-derived SSR markers in legumes: Dutta *et al* [88] had a 80% success rate within pigeon pea; Liu *et al* [89] found 82% success in Alfalfa, whereas Jhanwar *et al* [90] found a high (98%) success rate in the cultivated chickpea. For genic-derived SSRs, marker dropout could be cause by chimeric primers, by the creation of primers across intron/exon splice sites, or by creation across alternative splice sites or chimeric transcripts.

Assessing genetic diversity and population structure within Indian *Mucuna pruriens* accessions

The relative influence of different barriers to gene flow, such physical or geographical separation or incipient genetic or morphological changes that impact the ability to crossbreed, against promoters of gene flow, such as migration or the movement or interbreeding of individuals by human mediation, ultimately impacts genetic diversity within a species, particularly cultivated species [91]. Given the economic, medicinal and ethnobotanical importance of *M. pruriens*, assessing the genetic diversity within *M. pruriens* is an important endeavor. Our efforts to create and validate a database of potential SSR markers ultimately yielded 52 polymorphic markers for genetic diversity assessment within 23 Indian *M. pruriens* accessions representing all three *M. pruriens* varieties and sourced across India. All but two of the 52 markers showed adequate to high ability to discern ancestry based on the Shannon Information Content (I) [43] using the suggested cutoff of $I \geq 0.3$ [92], attesting to the utility of these markers for genetic diversity assessment. The average I across all our markers is 0.78, a value much higher than the ancestral discerning power of RAPD markers in *M. pruriens* (average of 0.62 across 15 primer pairs) [93]. The average polymorphic information content (PIC) across our 52 markers in *M. pruriens* is 0.24, a value similar to that obtained by Leelambika et al. [94] but higher than all estimates using AFLP or RAPD data which ranged from 0.166 [95] to 0.174 [94], attesting to the appropriate choice of SSR markers for assessing genetic diversity in *Mucuna*.

We explored genetic diversity across different subpopulation divisions based on geography (East, Northeast, and Peninsular India), variety (*M.* vars. *pruriens, hirsuta,* and *utilis*), and empirical genetic structure (fastSTRUCTURE subgroups 1-4) (Table 6). East India showed a slightly higher gene diversity than other geographical areas, a finding somewhat surprising given that all accessions for East India are of a single variety, *M. p.* var. *pruriens*. Peninsular India had the highest average number of alleles, which could be attributed to this area comprising accessions from all three varieties, leading to an overall higher average number of alleles through inclusion of the allelic diversity specific to variety. Many cultivated crops are genetically depauperate compared to their wild relatives [96], and *M. p.* var. *utilis* is no exception. Wild accessions (*M.* vars. *pruriens* and *hirsuta*) had higher gene diversity and a higher average number of alleles than cultivated (*M. p.* var. *utilis*) accessions (Table 6), corroborating previous studies in *Mucuna* [93, 94]. Similar trends are found throughout legumes, for example within *Phaseolus vulgaris* [97]. However, estimates of gene flow were high regardless of subdivision type (i.e. geography, variety, genetic structure), suggesting significant mixing among germplasm. All that said, caution is warranted in interpreting these comparative results given our low number of accessions examined.

Estimation of population genetic substructure revealed $K = 4$ subgroups (Fig. 5a) with coefficient of ancestry placing most individuals strongly within a particular subgroup. Both geography and variety are somewhat correlated to groupings within genetic substructure, PCA, and clustering analyses, indicating their impact on structure of genetic diversity. For instance, subgroups 2, 3 and 4 contain genotypes mostly collected from east India, peninsular India + Maharashtra and northeast India, respectively, except for a few accessions whose placement was variable. Subgroup 2 comprises only var. *pruriens* accessions that are mostly from eastern India whereas subgroup 3 is mainly var. *hirsuta* with one accession each of var. *pruriens* and var. *utilis*.

Neither geography nor variety correlated completely with clades produced by the NJ algorithm, in contrast to

previous cluster analyses based on ISSR data which were largely associated by taxonomy [94]. That said, a recent analysis using ISSR and RAPD markers across several species of *Mucuna*, including all *pruriens* varieties, found evidence via UPGMA cluster analyses for varietal cohesion of var. *utilis* and var. *pruriens*, as well as their sister relationship, but found that var. *hirsuta* clustered apart from the others, suggesting a separate evolutionary trajectory for this variety [98]. The lack of strict varietal clustering across our three population structure assessment methods (Fig. 5) coupled with high estimates of migration (Table 6) and clear suggestion of at least two hybrid individuals as determined by mixed coancestry within fastSTRUCTURE analyses suggests that hybridization can take place easily among varieties, a conclusion also suggested previously [94]. Alternatively, these variations may represent ancestral states perpetuated into extant populations. However, the small sample size used in this study resulted in limited power to accurately identify subgroups containing consistent genotypes across all the three methods tested. Thus, further studies involving larger samples derived from extended geographical regions are needed to make generalized conclusions on the divergence and population structure of Indian *M. pruriens*.

Additional files

Additional file 1: Figures and tables highlighting the analysis and results obtained in this study. **Figure S1.** Length distribution of *Mucuna pruriens* transcripts in Trinity assembly. **Figure S2.** Functional characterization and abundance of *Mucuna pruriens* transcriptome for enzyme classes (A) and KEGG pathways (B). Transcripts were classified in the top 20 abundant enzyme classes and KEGG pathways; area under each pie slice represents the value in percent. **Figure S3.** Legume sequence similarity analysis. Percentage identity of transcripts against other legume protein databases (A) and relative numbers of transcripts that had significant sequence similarity by species (B). The percentage of transcripts showing similarity value (E-value ≤ 1E-05) in BLASTX searches is shown. **Figure S4.** Repeat distribution in *Mucuna pruriens* transcriptome discovered using lobSTR program. **Figure S5.** Results of fastSTRUCTURE analysis across K = 2 to K = 10 subgrounds. Subgroup number K plotted against marginal likelihood (A) and cross validation error (B). **Figure S6.** Heat map and complete linkage hierarchical clustering of differentially expressed transcripts of leaf, pod, root, and pooled transcriptomes. The various shades in the boxes showed similar tendencies of gene expression. **Table S1.** Contrasting phenotypes of the two genotypes used for transcriptome sequencing. **Table S2.** Length distribution of the EST-SSRs based on the number of nucleotide repeats. **Table S3.** Summary of the repeats in *Mucuna pruriens* transcriptome based on lobSTR. **Table S4.** Polymorphism information of 52 EST-SSR markers for 23 *Mucuna pruriens* accessions. Ho: Observed heterozygosity; He: Expected heterozygosity (Kimura & Crow, 1964); PIC: Polymorphism information content; Na: No. of alleles; Ne: Effective no. of alleles; I: Shannon information content; h: Nei's (1973) gene diversity. (DOCX 4530 kb)

Additional file 2: Number of transcripts assigned to Biological, Cellular and Metabolic processes from GO analysis of the annotated transcripts of the *Mucuna* assembly. (XLSX 249 kb)

Additional file 3: Number of transcripts assigned for enzyme classes and KEGG pathways of the M.pruriens transcripts. (XLSX 26 kb)

Additional file 4: BLASTP analysis of the transcripts against the legume protein databases. (XLSX 4558 kb)

Additional file 5: Transcription factors identified in the *Mucuna* assembly using PlnTFDB (http://plntfdb.bio.uni-potsdam.de/v3.0, Pérez-Rodríguez et al. 2009) (XLSX 81 kb)

Additional file 6: Transcripts with the repeats in the Mucuna assembly identified using Perl script MISA (MicroSAtellite; http://pgrc.ipk-gatersleben.de/misa; Thiel et al. 2003) (XLSX 2581 kb)

Additional file 7: Description of the repeats identified in the assembly using MISA and the percentage of each class are reported. In addition the position of the start/end of the 7,943 repeat sequences is also reported. (XLSX 302 kb)

Additional file 8: Polymorphic sequence repeats identified between the *Mucuna* genotypes using lobSTR program, including the full list and the filtered list. (XLSX 545 kb)

Additional file 9: Fold change expression values of the transcripts in each of the tissue analyzed related to the secondary metabolite pathways. (XLSX 227 kb)

Abbreviations

AFLP: Amplified fragment length polymorphism; ATP: Adenosine triphosphate; bHLH: Basic helix-loop-helix; BLAST: Best local alignment search tool; bp: Base pair; C: Chromosome; CAMTA: Calmodulin-binding transcription activator; cDNA: Coding deoxyribonucleic acid; CTAB: Cetyltrimethylammonium bromide; DET: Differentially expressed transcript; DNA RNA: Ribonucleic acid; DNA: Deoxyribonucleic acid; dNTP: Deoxynucleotide triphosphate; EC: Enzyme class; EST-SSR: Expressed sequence tag-simple sequence repeats; FPKM: Fragments per kilobase per million; GB: Gigabase; GO: Gene ontology; Gst: Coefficient of gene differentiation; h: Gene diversity; He: Expected heterozygosity; Ho: Observed heterozygosity; Hs: Gene diversity within groups; Ht: Total gene diversity; I: Shannon's information index; K: Number of subpopulations; KEGG: Kyoto encyclopedia of genes and genomes; MADS: MCM1, AGAMOUS, DEFICIENS, and SRF box genes; MAF: Minor allele frequency; Mbp: Megabase pair; mRNA: Messenger RNA; MYB: Myeloblastosis; NAC: NAM, ATAF, and CUC; Ne: Effective number of alleles; NGS: Next generation sequencing; NJ: Neighbor-joining; Nm: Gene flow index; PC: Principal component; PCA: Principal component analysis; PCR: Polymerase chain reaction; PHD: Plant homeodomain; PIC: Polymorphic information content; RAPD: Random amplified polymorphic; SSR: Simple sequence repeat; TF: Transcription factor; VCF: Variant call format

Acknowledgements

NS thanks Sikkim University, Gangtok, India for facilities; C-CAMP, Bangalore, India for NGS data production; ICRISAT, Hyderabad, India for genotyping services.

Funding

This research was supported by funding from the Dept. of Biotechnology (DBT), Govt. of India (Grant No. BT/PR3489/PBD/16/945/2011) to NS and from the US National Science Foundation to ANE (DEB-1352217) and JJD (DEB-0948800).

Availability of data and materials

The Illumina paired-end reads as well as the assembly have been submitted to the U.S. National Center for Biotechnology Information (NCBI) sequence read archive (SRA) database: SRR3453110.

Authors' contributions

All authors contributed to various aspects of this work (ordered by degree of contribution): conceived the study (NS), aided in study design (NS, ANE, JJD), obtained research funds (NS, JJD, ANE), coordinated activities (NS, ANE), obtained and grew plants, RNA isolation and quantification, diversity analysis (PRK, TPK, PKB), bioinformatics and differential transcript expression analyses (RC, HRS), population structure analyses (VB), contributed to preparation of the manuscript (ANE, NS, RC, VB, JJD). All authors read and approved the final manuscript.

Competing interests

The author(s) declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note

Author details
[1]Department of Botany, Sikkim University, 6th Mile, Tadong-737102, Gangtok, Sikkim, India. [2]United States Department of Agriculture, Agriculture Research Service, 3810 4th St., Lubbock, TX 79415, USA. [3]Department of Plant Science, McGill University, Raymond Building, 21111 Lakeshore Road, Ste. Anne de Bellevue, Quebec H9X 3V9, Canada. [4]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA. [5]Institute of Bioresources and Sustainable Development, ikkim Centre, Tadong-737102, Gangtok, Sikkim, India. [6]Section of Plant Breeding and Genetics, School of Integrative Plant Science, Cornell University, 412 Mann Library, Ithaca, NY 14853, USA. [7]Department of Botany, Smithsonian Institution, National Museum of Natural History, US National Herbarium, 10th and Constitution Ave NW, Washington, DC 20013, USA.

Received: 22 December 2016 Accepted: 10 May 2017
Published online: 25 May 2017

References

1. Bhat R, Karim A. Exploring the nutritional potential of wild and underutilized legumes. Compr Rev Food Sci Food Saf. 2009;8(4):305–31.
2. Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Nigam SN, Saxena K, Vadez V, Sethy NK, Bhatia S, Aruna R, et al. Molecular genetics and breeding of grain legume crops for the semi-arid tropics. In: Varshney RK, Tuberosa R, editors. Genomics Assisted Crop Improvement: Genomics Applications in Crops, vol. 2: Springer: Netherlands; 2007. p. 207–41.
3. Schrire BD. Phaseoleae. In: Lewis G, Schrire BD, MacKinder B, Lock M, editors. Legumes of the World. Kew: Royal Botanic Gardens; 2005.
4. Egan AN, Vatanparast M, Cagle W. Parsing polyphyletic *Pueraria*: Delimiting distinct evolutionary lineages through phylogeny. Mol Phylogenet Evol. 2016;104:44–59.
5. Buckles D. Velvetbean: A "new" plant with a history. Econ Bot. 1995;49(1):13–25.
6. Sathyanarayana N, Mahesh S, Leelambika M, Jaheer M, Chopra R, Rashmi K. Role of genetic resources and molecular markers in *M. pruriens* (L.) DC. improvement. Plant Genet Resources. 2016;1:1–13.
7. Bairiganjan G, Patnaik S. Chromosomal evolution in Fabaceae. Cytologia. 1989;54(1):51–64.
8. Wilmot-Dear C. A revision of *Mucuna* (Leguminosae: Phaseoleae) in the Indian subcontinent and Burma. Kew Bull. 1987;42(1):23–46.
9. Bressani R. Factors influencing nutritive value in food grain legumes: *Mucuna* compared to other grain legumes. In: Food and feed from Mucuna: Current uses and the way forward Proceedings of an International Workshop: 2002; Tegucigalpa, Honduras. 2002. p. 164–88.
10. Carsky RJ, Becker M, Hauser S. Mucuna cover crop fallow systems: potential and limitations. In: Sustaining Soil Fertility in West Africa. Edited by Carksy R, vol. 58. Wisconsin, USA: Soil Sciences Society of America; 2001: 111-135
11. Fujii Y, Shibuya T, Yasuda T. L-3, 4-dihydroxyphenylalanine as an allelochemical candidate from *Mucuna pruriens* (L.) DC. var. *utilis*. Agric Biol Chem. 1991;55(2):617–8.
12. Carsky J, Ndikawa R. Identification of cover crops for the semi-arid savanna zone of West Africa. In: Buckles D, Eteka A, Osiname O, Galiba M, Galiano G, editors. Cover Crops in West Africa Contributing to a Austainable Agriculture. Ottawa: International Development Research Center; 1998. p. 179–87.
13. Jorge MA, Eilittä M, Proud FJ, Maasdorp V, Beksissa H, Sarial K, Hanson J. Mucuna species: recent advances in application of biotechnology. In: Fruit, Vegetable and Cereal Science and Biotechnology, Global Science Books, vol. 1. 2007. p. 80–94.
14. Kavitha C, Thangamani C. Amazing bean *Mucuna pruriens*: A comprehensive review. J Med Plant Res. 2014;8(2):138–43.
15. Lorenzetti E, MacIsaac S, Anasson J, Avang D. The phytochemistry, toxicology and food potential of velvet bean (Mucuna Adans spp. Fabaceae). In: Cover crops of West African contributing to sustainable agriculture 1998: IDRC, Ottawa, Canada & IITA, Ibadan, Nigeria. 1998. p. 57.
16. St Laurent L, Livesey J, Arnason J, Bruneau A. Variation in L-Dopa concentration in accessions of *Mucuna pruriens* (L.) DC and in Mucuna brachycarpa Rech. In: Food and Feed from Mucuna: Current Uses and the Way Forward, Proceedings of an International Workshop: 2002; Tegucigalpa, Honduras. 2002. p. 352–73.
17. Daxenbichler ME, VanEtten CH, Hallinan EA, Earle FR, Barclay AS. Seeds as sources of L-DOPA. J Med Chem. 1971;14(5):463–5.
18. Hornykiewicz O. A brief history of levodopa. J Neurol. 2010;257(2):249–52.
19. Katzenschlager R, Evans A, Manson A, Patsalos P, Ratnaraj N, Watt H, Timmermann L, Van der Giessen R, Lees A. *Mucuna pruriens* in Parkinson's disease: a double blind clinical and pharmacological study. J Neurol Neurosurg Psychiatry. 2004;75(12):1672–7.
20. Szabo N, Tebbett I. The chemistry and toxicity of *Mucuna species*. In: *Food and feed from Mucuna: Current uses and the way forward Proceedings of an International Workshop: 2002; Tegucigalpa, Honduras*. 2002. p. 120–41.
21. Sharma KK, Dumbala SR, Bhatnagar-Mathur P. Biotech approaches for crop improvement in the semi-arid tropics. In: Ricroch A, et al, editors. Plant Biotechnology: Experience and Future Prospects. Spring International Publishing Switzerland; 2014. p. 193-207.
22. Egan AN, Schlueter J, Spooner DM. Applications of next-generation sequencing in plant biology. Am J Bot. 2012;99(2):175–85.
23. Wei F, Luo S, Zheng Q, Qiu J, Yang W, Wu M, Xiao X. Transcriptome sequencing and comparative analysis reveal long-term flowing mechanisms in *Hevea brasiliensis* latex. Gene. 2015;556(2):153–62.
24. Wen J, Egan AN, Dikow RB, Zimmer EA. Utility of transcriptome sequencing for phylogenetic inference and character evolution. In: Hörandl E, Appelhans MS, editors. Next-generation sequencing in plant systematics. Koeltz Scientific Books. Regnum Vegetabile. 2015;158:51–91.
25. Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J, Materne M, Forster JW. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics. 2012;13(1):1.
26. Chen H, Wang L, Wang S, Liu C, Blair MW, Cheng X. Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers. PLoS One. 2015;10(4):e0120273.
27. Chen H, Chen X, Tian J, Yang Y, Liu Z, Hao X, Wang L, Wang S, Liang J, Zhang L. Development of gene-based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. PLoS One. 2016;11(3):e0151040.
28. Chapman MA. Transcriptome sequencing and marker development for four underutilized legumes. Appl Plant Sci. 2015;3(2):1400111. apps.
29. Vatanparast M, Shetty P, Chopra R, Doyle JJ, Sathyanarayana N, Egan AN. Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep. 2016;6:29070.
30. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj PK, Rani A, Singh RS, Raizada J, Singh K, et al. An RNA isolation system for plant tissues rich in secondary metabolites. BMC Res Notes. 2011;4:85.
31. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
32. Li W, Fu L, Niu B, Wu S, Wooley J. Ultrafast clustering algorithms for metagenomic sequence analysis. Brief Bioinform. 2012;13(6):656–88.
33. Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R. Annocript: a flexible pipeline for the annotation of transcriptomes also able to identify putative long noncoding RNAs. Bioinformatics. 2015;31(13):2199–201. doi:10.1093/bioinformatics/btv106.
34. Pérez-Rodríguez P, Riano-Pachon DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res. 2009;38(Database issue):D822–7.
35. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;106(3):411–22.
36. Gymrek M, Golan D, Rosset S, Erlich Y. lobSTR: a short tandem repeat profiler for personal genomes. Genome Res. 2012;22(6):1154–62.
37. You FM, Huo N, Gu YQ, Luo M-c, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinf. 2008;9(1):1.
38. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
39. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.

40. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
41. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;49:725–38.
42. Nei M. Analysis of gene diversity in subdivided populations. PNAS. 1973; 70(12):3321–3.
43. Lewontin RC. The apportionment of human diversity. In: Evolutionary Biology. vol. 6: Springer: USA; 1972. p. 381–98.
44. PE PRS. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics. 2012;28:2537–9.
45. Yeh FC, Yang R, Boyle T, Ye Z, Mao JX. POPGENE, version 1.32: the user friendly software for population genetic analysis. Edmonton, AB, Canada: Molecular Biology and Biotechnology Centre, University of Alberta; 1999.
46. Rambaut A: FigTree v1.4.2. In. University of Edinburgh, Edinburgh, UK Available at: http://tree.bio.ed.ac.uk/software/figtree; 2014
47. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
48. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
50. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 2011;12(1):1.
51. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
52. Wiersema JH, Leon B. World economic plants: a standard reference. New York: CRC press; 2016.
53. Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, et al. Genome sequence of the paleopolyploid soybean. Nature. 2010;463(7278):178–83.
54. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46(7):707–13.
55. Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, et al. Genome resources for climate-resilient cowpea, an essential crop for food security. Plant J. 2017; 89(5):1042–54. doi:10.1111/tpj.13404.
56. Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature. 2011;480(7378):520–4.
57. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al. Genome Structure of the Legume, Lotus japonicus. DNA Res. 2008;15(4):227–39.
58. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol. 2012;30(1):83–9.
59. Hane JK, Ming Y, Kamphuis LG, Nelson MN, Garg G, Atkins CA, Bayer PE, Bravo A, Bringans S, Cannon S, et al. A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. Plant Biotechnol J. 2017;15(3):318–30. doi:10.1111/pbi.12615.
60. Moura TM, Vatanparast M, Tozzi AM, Forest F, Wilmot-Dear CM, Simon MF, Mansano VF, Kajita T, Lewis GP, Herendeen PS. A molecular phylogeny and new infrageneric classification of Mucuna Adans. (Leguminosae-Papilionoideae) including insights from morphology and hypotheses about biogeography. Int J Plant Sci. 2016;177(1):76–89.
61. Egan AN, Pan B. Resolution of polyphyly in *Pueraria* (Leguminosae, Papilionoideae): The creation of two new genera, *Haymondia* and *Toxicopueraria*, the resurrection of *Neustanthus*, and a new combination in Teyleria. Phytotaxa. 2015;218(3):201–26.
62. Li H, Wang W, Lin L, Zhu X, Li J, Zhu X, Chen Z. Diversification of the phaseoloid legumes: effects of climate change, range expansion and habit shift. Front Plant Sci. 2013;4:386. doi:10.3389/fpls.2013.00386.
63. Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. Syst Bot. 2009;34(1):115–28.
64. de Queiroz LP, Pastore JFB, Cardoso D, Snak C, Lima ALC, Gagnon E, Vatanparast M, Holland AE, Egan AN. multilocus phylogenetic analysis reveals the monophyly of a recircumscribed papilionoid legume tribe Diocleae with well-supported generic relationships. Mol Phylogenet Evol. 2015;90:1–19.
65. Ištvánek J, Jaroš M, Křenek A, Řepková J. Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). Am J Bot. 2014; 101(2):327–37.
66. Mochida K, Sakurai T, Seki H, Yoshida T, Takahagi K, Sawai S, Uchiyama H, Muranaka T, Saito K. Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. Plant J. 2017;89(2):181–94. doi:10.1111/tpj.13385.
67. Egan AN, Doyle JJ. A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). Syst Biol. 2010;59(5):534–47.
68. Pichersky E, Gang DR. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci. 2000;5(10):439–45.
69. Libault M, Joshi T, Benedito VA, Xu D, Udvardi MK, Stacey G. Legume transcription factor genes: what makes legumes so special? Plant Physiol. 2009;151(3):991–1001.
70. Garg R, Patel RK, Tyagi AK, Jain M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res. 2011;18(1):53–63.
71. Vom Endt D, Kijne JW, Memelink J. Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry. 2002;61(2):107–14.
72. Reddy V, Lerner E. Plant cysteine proteases that evoke itch activate protease-activated receptors. Br J Dermatol. 2010;163(3):532–5.
73. Brain K. Accumulation of L-DOPA in cultures from *Mucuna pruriens*. Plant Sci Lett. 1976;7(3):157–61.
74. (UK) NCCfCC. Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care. London: Royal College of Physicians (UK); 2006.
75. Devi K, Mishra SK, Sahu J, Panda D, Modi MK, Sen P. Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of *Cymbopogon winterianus*. Sci Rep. 2016;6:21026.
76. Wink M, Mohamed GIA. Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the *rbcL* gene. Biochem Syst Ecol. 2003;31(8):897–917.
77. Ishihara KL, Honda MD, Pham DT, Borthakur D. Transcriptome analysis of *Leucaena leucocephala* and identification of highly expressed genes in roots and shoots. Transcriptomics. 2016;4:135.
78. Cheng Q, Li N, Dong L, Zhang D, Fan S, Jiang L, Wang X, Xu P, Zhang S. Overexpression of soybean isoflavone reductase (GmIFR) enhances resistance to *Phytophthora sojae* in soybean. Front Plant Sci. 2015;6:1024.
79. Konczak I, Zhang W. Anthocyanins—more than nature's colours. J Biomed Biotechnol. 2004;2004(5):239–40.
80. Husain Q. Chemistry and biochemistry of some vegetable flavors. In: Hui YH, editor. Handbook of Fruit and Vegetable Flavors. Hoboken: John Wiley & Sons, Inc; 2010. p. 575–625.
81. Peng F, Du Q, Peng C, Wang N, Tang H, Xie X, Shen J, Chen J. A review: the pharmacology of isoliquiritigenin. Phytother Res. 2015;29(7):969–77.
82. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 2005;23(1):48–55.
83. Bosamia TC, Mishra GP, Thankappan R, Dobaria JR. Novel and stress relevant EST derived SSR markers developed and validated in peanut. PLoS One. 2015;10(6):e0129127.
84. Mun J-H, Kim D-J, Choi H-K, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudez A-M. Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. Genetics. 2006;172(4):2541–55.
85. Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K. Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. Bioinformatics. 2004;20(7):1081–6.
86. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 2000;10(7):967–81.
87. Blair MW, Torres MM, Giraldo MC, Pedraza F. Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.). BMC Plant Biol. 2009;9(1):100.
88. Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol. 2011;11(1):17.

89. Liu Z, Chen T, Ma L, Zhao Z, Zhao PX, Nan Z, Wang Y. Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. PLoS One. 2013;8(12):e83549.
90. Jhanwar S, Priya P, Garg R, Parida SK, Tyagi AK, Jain M. Transcriptome sequencing of wild chickpea as a rich resource for marker development. Plant Biotechnol J. 2012;10(6):690–702.
91. Ellstrand NC, Rieseberg LH. When gene flow really matters: gene flow in applied evolutionary biology. Evol Appl. 2016;9(7):833–6.
92. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E. A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet. 2004;74(5):1001–13.
93. Padmesh P, Reji J, Dhar MJ, Seeni S. Estimation of genetic diversity in varieties of *Mucuna pruriens* using RAPD. Biol Plant. 2006;50(3):367–72.
94. Leelambika M, Mahesh S, Jaheer M, Tripathi P, Kumar PR, Sathyanarayana N. Targeted metabolic and genomic profiling reveals parents for L-dopa breeding in *Mucuna pruriens* (L.) DC. Trop Plant Biol. 2016;9(4):239–51.
95. Sathyanarayana N, Leelambika M, Mahesh S, Jaheer M. AFLP assessment of genetic diversity among Indian *Mucuna* accessions. Physiol Mol Biol Plants. 2011;17(2):171–80.
96. Flint-Garcia SA. Genetics and consequences of crop domestication. J Agric Food Chem. 2013;61(35):8267–76.
97. Papa R, Gepts P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet. 2003;106(2):239–50.
98. Patil RR, Pawar KD, Rane MR, Yadav SR, Bapat VA, Jadhav JP. Assessment of genetic diversity in *Mucuna* species of India using randomly amplified polymorphic DNA and inter simple sequence repeat markers. Physiol Mol Biol Plants. 2016;22(2):207–17.

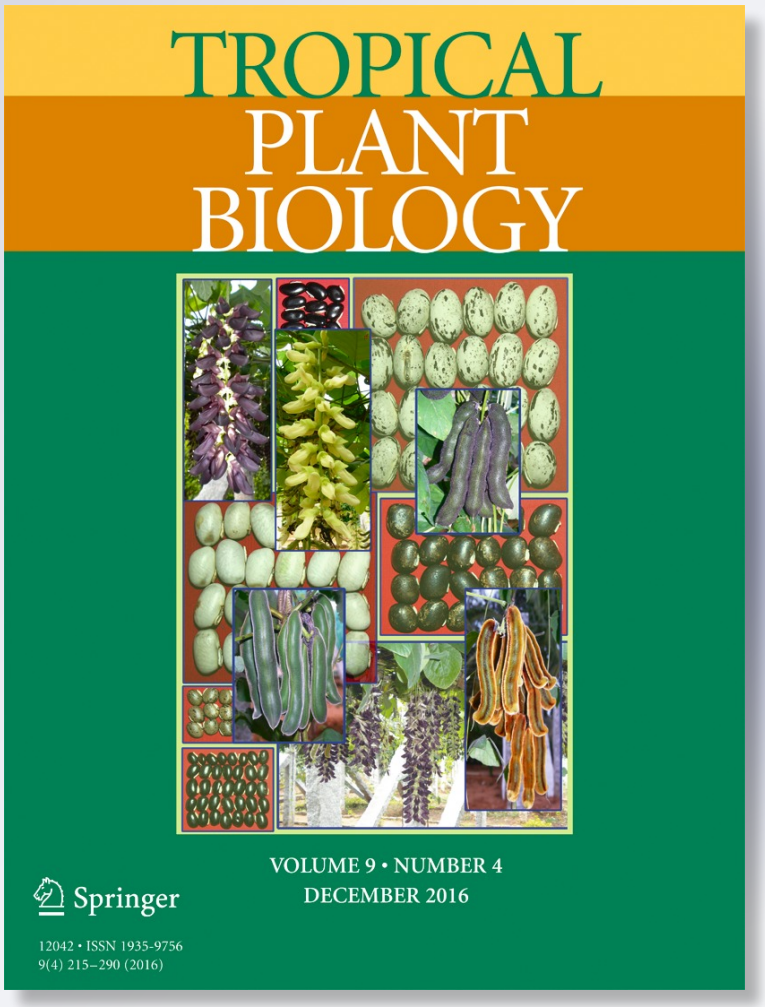
Targeted Metabolic and Genomic Profiling Reveals Parents for L-Dopa Breeding in Mucuna pruriens *(L.) DC.*

M. Leelambika, S. Mahesh, M. Jaheer, P. K. Tripathi, P. Ranjith Kumar & N. Sathyanarayana

Tropical Plant Biology
An International Journal devoted to original research in tropical plants

ISSN 1935-9756
Volume 9
Number 4

Tropical Plant Biol. (2016) 9:239-251
DOI 10.1007/s12042-016-9174-3



Springer

Tropical Plant Biol. (2016) 9:239–251
DOI 10.1007/s12042-016-9174-3

Targeted Metabolic and Genomic Profiling Reveals Parents for L-Dopa Breeding in *Mucuna pruriens* (L.) DC.

M. Leelambika[1] · S. Mahesh[1] · M. Jaheer[1] · P. K. Tripathi[2] · P. Ranjith Kumar[2] · N. Sathyanarayana[2]

Received: 9 November 2015 / Accepted: 26 May 2016 / Published online: 8 June 2016

Abstract Integrating targeted metabolic profiling with genetic diversity estimates is vital for selecting genetic stocks for breeding as well as mapping a biochemical trait. In *Mucuna pruriens* (L.) DC., copious amount of L-Dopa in seeds is viewed both as a boon and bane. Our objective of this work was to assess in conjunction both L-Dopa and genomic diversity in a selected set of germplasm to elucidate their relationship and identify parents for L-Dopa breeding/mapping. The findings revealed good genetic as well as metabolic (L-Dopa) diversity among the studied accessions. However, we could not establish direct relationship between these two as the marker data measured using AFLP and ISSR were not correlated with the seed L-Dopa contents. Based on the consensus information from both the data sets, seven parental combinations have been suggested. These findings are expected to pave way for genetic improvement as well as genetic mapping of L-Dopa trait besides integration of molecular markers in *M. pruriens* breeding programs.

Keywords *Mucuna pruriens* · L-Dopa · ISSR · AFLP · Genetic mapping · Correlation study

Communicated by: Paulo Arruda

✉ N. Sathyanarayana
sathyan_dixit@yahoo.in; nsathyanarayana@cus.ac.in

[1] Department of Biotechnology, Sir M Visvesvaraya Institute of Technology, Bangalore -562157, India

[2] Department of Botany, Sikkim University, 6th Mile Samdur, P.O. Tadong, Gangtok, Sikkim 737102, India

Introduction

Genetic diversity among the members of a species is critical prerequisite for successful plant breeding programs. Its two major data sources: trait phenotype and DNA markers denote different facets of genome polymorphism, and are sometime applied independently in parental selection. However, when neutral DNA markers are employed for this purpose, integration of phenotypic data with genotypic information is strongly recommended. This consideration is even more critical in case of biochemical traits where the trait expression is restricted by, among other things, the fitness it confers to the plant. In such cases, especially for the crops with a long history of domestication that possess commercially relevant chemical phenotypes, complementation of DNA markers with chemical profiling is considered necessary to achieve the desired end results (Laurentin et al. 2008).

Mucuna pruriens L. (DC.) - is a self-pollinated tropical legume classified within the phaseoloid clade of Leguminosae which also includes soybean, mung bean and relatives. It is the source of several important pharmaceuticals (Warrier et al. 1996), the most prominent being 3, 4 dihydroxy-L-phenylalanine (L-Dopa), present in copious quantity in seeds (1.4–9.1 %). Daxenbichler et al. (1971) screened 1000 species in 135 plant families, and found only *Mucuna* sp. to contain sufficient L-Dopa for commercial use. Biochemically a non-protein amino acid produced as an intermediary product in the enzymatic synthesis of dopamine from L-tyrosine, its efficacy for the treatment of Parkinson's disease is well established (Soares et al. 2014). However prolonged consumption of drug or preparations containing it is known to induce severe side effects in humans as well as diminish performance and health in livestock (Katzenschlager et al. 2004). Due to this, it is regarded as greater risk among all the anti-nutritional substances present in this otherwise protein rich

legume sp. (Szabo and Tebbett 2002). Nonetheless, a healthy individual can safely consume 500 g of *Mucuna* based food/day with 0.1 % L-Dopa (Lorenzetti et al. 1998). Teixeira et al. (2003) further confirmed this dose based on research in Faba bean and Broad bean (*Vicia faba*) where it is present in 0.2–0.5 % respectively and still consumed safely worldwide. Breeding varieties with low and high L-Dopa content therefore offers significant promise for large scale cultivation of *M. pruriens*.

Variability for L-Dopa content in *M. pruriens* germplasm is reported in number of earlier studies (Mary Josephine and Janardhanan 1992; Krishnamurthy et al. 1996; Kalidass and Mohan 2011; Fathima et al. 2010 etc.). Besides, quite a few workers analyzed genetic diversity in natural population using DNA markers (Capo-Chichi et al. 2001; Capo-Chichi et al. 2003a; Padmesh et al. 2006; Sathyanarayana et al. 2011). However, so far, comprehensive L-Dopa phenotypes have not been included systematically in diversity surveys in *M. pruriens* germplasm. Our objective in this study was therefore to assess L-Dopa diversity in *M. pruriens* by targeted profiling to elucidate the relationship between metabolic and genome diversity in this plant to identify promising genetic stocks for breeding/mapping this trait.

Materials and Methods

The Plant Material

Totally 40 *M. pruriens* accessions were used in this study. The collection comprised representative genotypes from both the botanical varieties of *M. pruriens* viz.: var. *utilis* (Joy et al. 2007) and var. *pruriens (Lorenzetti et al. 1998)* (Table 1). These accessions originated from different geographical locations in India and included both cultivated (var. *utilis*; velvetbean) and wild (var. *pruriens*, itching bean) varieties.

L-Dopa Isolation and HPLC

Healthy seeds recovered from 10 replicates of each assayed accession were pooled separately and pure L-Dopa was extracted following modified Daxenbichler et al. (1971) and Brain (1976) methods. Finely ground seed powder was defatted using petroleum ether and L-Dopa was extracted with 0.1 N HCl added with 80 % ethanol and 0.1 % ascorbic acid. The purity of the isolated L-Dopa was further confirmed by subjecting the random samples to TLC analysis (Gunendi and Pamuk 1999) with L-Dopa from the Sigma Aldrich chemicals used as a standard.

The isolated L-Dopa samples were analyzed on High Performance Liquid Chromatography (HPLC, Agilent Compact LC1120 system). The sample was manually injected and detected with double-beam photometer detector operated at a wavelength of 283 nm. The data was acquired and analyzed using EZ-Chrom Elite compact software. A C18 column (250 x 4.6 mm, 5 μm SS column,) fitted with a guard column was used for the separation using an isocratic mixture of acetonitrile as mobile phase. The mobile phase was filtered through a 0.45 μm membrane filter (Sartorius USA) and was then degassed by ultrasonication for 30 min. Analysis was run at a flow rate of 1.0 ml/min and quantification was by peak height. Injection volume was adjusted to 20 μl and detection was made at 283 nm. Standard L-Dopa (Sigma Aldrich) was used for preparation of calibration curve and retention time analysis. Each sample was analyzed in quadruplicates and L-Dopa concentration was finally represented as percentage composition. All the experiments were performed in triplicates.

L-Dopa Data Processing and Analysis

The L-Dopa data was subjected to one- way ANOVA and the mean separation was achieved based on significant Tukey's HSD test at 5 % significance using JMP software version 10 (SAS 2012). Later, the L-Dopa content was set into frequency distribution classes to transform the quantitative data to a binary matrix. Thus obtained binary matrix was used to generate Jaccard's similarity coefficient matrix and the cluster analysis was conducted with the same using NTSYS-pc version 2.1 (Exerter software; Rohlf 2009) to visualize the relationship among accessions based on their L-Dopa content. The correlation between the cluster tree data matrix and the similarity matrix was calculated using Mantel's test (1000 permutations).

DNA Isolation

About 1 g of young leaf tissue was harvested from 2–3 week old seedlings from ten individual plants and bulked separately for each *Mucuna* accession. Total genomic DNA was isolated following modified Doyle and Doyle (1990) method using cetyl trimethyl ammonium bromide (CTAB). Ground leaves were homogenized in CTAB-extraction buffer containing 0.5 % charcoal along with 0.2 % β-mercaptoethanol and incubated at 60 °C for 1 h. Purification steps were carried out twice with SEVAG (24-chloroform: 1-isoamlyalcohol) and once with phenol: chloroform: isoamylalcohol (25:24:1). After centrifugation at 12,000 rpm for 2 min, the DNA was pelleted using 0.67 volume of isopropanol followed by wash with ethanol (70 %). Air dried pellets were re-suspended in 0.5 ml of 1X Tris- EDTA buffer (pH 8.0) and quantified fluorometrically on 0.8 % agarose gel using ethidium bromide staining.

ISSR Genotyping

Amplification of ISSRs was carried out in a 25 μl reaction mixture containing 0.3 mM dNTP's, 10 mM Tris-HCl, 3 mM

Table 1 List of *M. pruriens* accessions used in the study

Sl. No.	Accession Name	Accession Number	Place of collection	Latitude and longitude
1.	*M. pruriens* var. *utilis*	500102KA	Karnataka	-
2.	*M. pruriens* var. *utilis*	500101KA	Karnataka	-
3.	*M. pruriens* var. *utilis*	500108KA	Hunasamaranahalli, Karnataka	13°14′ N, 77°62′ E
4.	*M. pruriens* var. *utilis*	IC385925	NBPGR, Dumka, Jharkand	24°26′ N, 87°24′ E
5.	*M. pruriens* var. *utilis*	IC385928	NBPGR	-
6.	*M. pruriens* var. *utilis*	IC392850	NBPGR, Moudi, Orissa	22°06′ N, 86°40′ E
7.	*M. pruriens* var. *utilis*	IC471876	NBPGR, Delhi	28°62′ N, 77°23′ E
8.	*M. pruriens* var. *utilis*	IC185926	NBPGR	-
9.	*M. pruriens* var. *utilis*	IC385841	NBPGR, Pakud, Jharkand	24°62′ N, 87°84′ E
10.	*M. pruriens* var. *utilis*	IC369144	NBPGR, Latehar, Jharkand	23°74′ N, 84°50′ E
11.	*M. pruriens* var. *utilis*	IC395793	NBPGR, Narmada, Gujarat	21°89′ N, 73°48′ E
12.	*M. pruriens* var. *utilis*	IC392241	NBPGR, Sawalpur, Jharkand	23°78′ N, 86°43′ E
13.	*M. pruriens* var. *utilis*	IC385926	NBPGR, Dhangadih, Jharkand	24°26′ N, 87°24′ E
14.	*M. pruriens* var. *utilis*	IC385842	NBPGR, Mohanpur, Jharkand	24°48′ N, 86°69′ E
15.	*M. pruriens* var. *utilis*	IC326953	NBPGR, Jaunaji, Solan, HP	30°90′ N, 77°09′ E
16.	*M. pruriens* var. *utilis*	IC471870	NBPGR, Delhi	28°62′ N, 77°23′ E
17.	*M. pruriens* var. *utilis*	500155AP	Tandur, Andhra Pradesh	-
18.	*M. pruriens* var. *pruriens*	500112KA	Mysore, Karnataka	12°18′ N, 76°42′ E
19.	*M. pruriens* var. *pruriens*	500113MH	Triambakeshwar, Maharashtra	20°00′ N, 73°77′ E
20.	*M. pruriens* var. *pruriens*	500110KA	Devarayanadurga, Karnataka	13°33′ N, 77°10'E
21.	*M. pruriens* var. *pruriens*	500116TN	Raw drug vendor, Madurai	-
22.	*M. pruriens* var. *pruriens*	IC265577	NBPGR, Njeezhoor, Kerala	9°58′ N, 76°52′ E
23.	*M. pruriens* var. *pruriens*	IC391941	NBPGR, Gaatogaon, Orissa	21°65′ N, 85°63′ E
24.	*M. pruriens* var. *pruriens*	IC391885	NBPGR, Anandpur, Orissa	21°65′ N, 85°63′ E
25.	*M. pruriens* var. *pruriens*	500138TN	Chittanavasal, Tamil Nadu	10°38′ N, 78°82′ E
26.	*M. pruriens* var. *pruriens*	500123KL	Seed company, Kerala	-
27.	*M. pruriens* var. *pruriens*	500109KA	Shimoga, Karnataka	13°56′ N, 75°38′ E
28.	*M. pruriens* var. *pruriens*	500111KA	Shivaganga, Tumkur, Karnataka	13°33′ N, 77°10′ E
29.	*M. pruriens* var. *pruriens*	500115TN	Raw drug vendor, Madurai	-
30.	*M. pruriens* var. *pruriens*	500120TN	Raw drug vendor, Madurai	-
31.	*M. pruriens* var. *pruriens*	500121TN	Raw drug vendor, Madurai	-
32.	*M. pruriens* var. *pruriens*	500122TN	Raw drug vendor, Madurai	-
33.	*M. pruriens* var. *pruriens*	500143KA	Basarikatte, Karnataka	13°18′ N, 75°49′ E
34.	*M. pruriens* var. *pruriens*	500144AP	Pathur, Andhra Pradesh	18°03′ N, 78°18′ E
35.	*M. pruriens* var. *pruriens*	500146AP	Ghanapur, Andhra Pradesh	18°39′ N, 78°10′ E
36.	*M. pruriens* var. *pruriens*	500147AP	Dhurzgaon, Andhra Pradesh	18°39′ N, 78°10′ E
37.	*M. pruriens* var. *pruriens*	500148AP	Mancheppa, Andhra Pradesh	18°40′ N, 78°10′ E
38.	*M. pruriens* var. *pruriens*	500149AP	Pochampad Dam, Andhra Pradesh	18°89′ N, 78°10′ E
39.	*M. pruriens* var. *pruriens*	500150AP	Kanakapura, Andhra Pradesh	19°05′ N, 78°45′ E
40.	*M. pruriens* var. *pruriens*	500151AP	Kadam, Andhra Pradesh	19°37′ N, 78°30′ E

$MgCl_2$, 50 mM KCl, 0.1 % Triton X-100, 1.0 U *Taq DNA polymerase*, 10 pmol primer and 50 ng of genomic DNA. Amplification was performed in PTC-200 Peltier thermal cycler (MJ Research, USA) using the thermal profile: initial cycle of 94 °C for 2 min, 36 °C for 2 min and 72 °C for 2 min followed by 38 cycles of 94 °C for 1 min, 36 °C for 1 min and 72 °C for 2 min with a final extension of 72 °C for 7 min. The products were held at 4 °C. Reaction mixture in which template DNA was replaced by distilled water was used as negative control. Amplified products were resolved on 1.5 % agarose gel (0.5X TBE) followed by ethidium bromide staining. The gel was then visualized, photographed and archived using BIO-RAD™ gel documentation system (BIO-RAD, USA).

AFLP Genotyping

AFLP fingerprinting was carried out as per Capo-Chichi et al. (2001) with minor modifications. Genomic DNA (500 ng) was digested with 10 U of *Eco* RI and 4 U of *Mse* I (New England Biolabs, USA) at 37 °C for 3 h. Without inactivating the restriction enzymes, adapters [*Eco* RI (5 pmol) and *Mse* I (50 pmol)] were ligated to the restricted DNA fragments in ligation buffer [1x T4 DNA ligase buffer, 1 μl of T4 DNA ligase (NEW ENGLAND BIOLABS, USA)] and incubated at 37 °C for 12 h. Pre-amplification of the diluted (2-fold) - ligated DNA was carried out with the primers complimentary to the *Eco* RI and *Mse* I adapters with two sets of selective nucleotides, one with cytosine and guanine and the other with adenine and cytosine respectively in PTC-200™ thermocycler (MJ RESEARCH INC. USA) using the following cycling parameters: 20 cycles of 94 °C for 30 s, 56 °C for 60 s, 72 °C for 60 s. The diluted (4-fold), amplified products were used as template for selective amplification. The second amplification was carried out with twelve selective primer combinations of *Eco* RI and *Mse* I each with three selective nucleotides in a total volume of 10 μl. The PCR program consisted of two segments. The first segment comprised of 12 cycles with one cycle at 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 60 s. The annealing temperature was then lowered by 0.7 °C per cycle during the first 12 cycles to reach an optimum temperature of 56 °C. The second segment comprised of 23 cycles at 94 °C for 30 s, 56 °C for 60 s and 72 °C at 60 s.

Gel electrophoresis was carried out using Sequegen DNA sequencer (BIORAD™, USA). Following amplification, the PCR products were mixed with 8 μl of formamide stop/loading buffer and denatured at 94 °C for 5 min, then chilled immediately to 4 °C. Eight microliters of each reaction mixture was loaded onto a 6 % denaturing PAGE. The gel was pre-run in 1X TBE freshly prepared from a 10X TBE stock solution. The key electrophoresis parameters included voltage set at 1200 V and temperature at 45 °C. The DNA bands were visualized using silver staining.

Data Analysis

Marker Attributes/ Features

Genotyping data obtained from each marker system was used for assessing its discriminatory power by evaluating: (i) Polymorphism information content (PIC), (ii) Marker index (MI) and (iii) Resolving power (Rp). The PIC value and mean gene diversity for each primer / enzyme combination was calculated as part of summary statistics using Power Marker version 3.25 (Liu and Muse 2005) and was averaged over the fragments for each primer combination. The marker index was calculated as given in Varshney et al. (2007) as MI = PIC x EMR where, EMR is "the effective multiplex ratio and is defined as the product of total number of loci/fragments per primer (n) and the fraction of polymorphic loci/fragments (β) (E = n* β)".

Resolving power of each primer was calculated according to Prevost and Wilkinson (1999) which is given as $Rp = \sum I_b$ where I_b represents the band informativeness. The I_b can be represented into a 0–1 scale by the following formula: $I_b = 1\text{-}[2^* |0.5\text{-}P|]$ where, p is the proportion of the accessions containing the band/ fragment.

Cluster Analysis

Each ISSR/AFLP marker was treated as a unit character and scored as a binary code (1/0). The results were analyzed using NTSYS-pc version 2.1 (Exerter software; Rohlf 2009) with SIMQUAL option on the basis of Jaccard's coefficient to generate genetic similarity coefficients among all the possible pairs and ordered in a similarity matrix (Jaccard 1908). The resulting matrices were subjected to clustering by UPGMA (Sokal and Michener 1958). The goodness of fit of the clustering to the data matrix was calculated by the COPH and MXCOMP programs. The robustness of the phenograms was tested by bootstrap analyses of 1000 iterations using Winboot software (Yap and Nelson 1996).

Multivariate and Correlation Analysis

UPGMA clustering pattern of all the three data sets: AFLP, ISSR and L-Dopa were analyzed and the data from the marker analysis were compared with L-Dopa profiling using NTSYS-pc version 2.1 (Exerter software; Rohlf 2009). Comparison of clustering obtained by AFLP: L-Dopa and ISSR: L-Dopa was based on Pearson's correlation and a Mantel test between the matrices with 1000 permutations. The data sets also were compared by tree plots to visualize the correlation.

Results

L-Dopa Variability

TLC analysis produced single clear band matching in colour and R_f (0.387) with that of standard L-Dopa sample confirming high purity isolations. The L-Dopa screening revealed moderate to high variability among different accessions which ranged from 1.09 % (IC369144) to 3.59 % (500185MH) (Table 2). Frequency distribution analysis generated seven major frequency classes (Fig. 1) with majority of them falling in the middle range of 1.50–1.90 % (Doyle and Doyle 1990); 1.91–2.31 % (Doyle and Doyle 1990) and 2.32–2.72 % (Capo-chichi et al. 2003a).

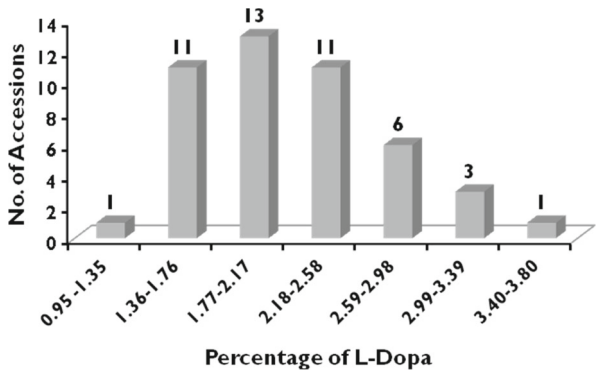
Table 2 L-Dopa content (%) among 40 *M.pruriens* accessions based on HPLC analysis

Sl. No.	Accessions	L-Dopa (%) Mean ± SD[y]	Sl. No.	Accessions	L-Dopa (%) Mean ± SD[y]
1.	IC369144	1.091 ± 0.041[t]	21.	500122TN	2.03 ± 0.018[n]
2.	IC385928	1.481 ± 0.010[rs]	22.	500111KA	2.08 ± 0.030[n]
3.	IC391885	1.52 ± 0.010[rs]	23.	500108KA	2.099 ± 0.007[kl]
4.	IC385841	1.542 ± 0.020[rs]	24.	500110KA	2.1 ± 0.006[kl]
5.	500115TN	1.57 ± 0.020[rs]	25.	500144AP	2.107 ± 0.007[k]
6.	IC385926	1.626 ± 0.020[q]	26.	500112KA	2.153 ± 0.010[k]
7.	500101KA	1.672 ± 0.016[q]	27.	500150AP	2.16 ± 0.020[k]
8.	500146AP	1.68 ± 0.013[q]	28.	500148AP	2.31 ± 0.053[hi]
9.	IC385842	1.633 ± 0.025[q]	29.	500109KA	2.321 ± 0.049[hi]
10.	IC392850	1.69 ± 0.030[q]	30.	IC392241	2.351 ± 0.011[hi]
11.	IC385925	1.73 ± 0.026[q]	31.	IC395793	2.36 ± 0.023[hi]
12.	500143KA	1.75 ± 0.019[q]	32.	500123KL	2.37 ± 0.021[hi]
13.	IC391941	1.85 ± 0.040[o]	33.	IC265577	2.4 ± 0.009[hi]
14.	IC326953	1.863 ± 0.048[o]	34.	IC471870	2.397 ± 0.019[h]
15.	500155AP	1.878 ± 0.048[op]	35.	500147AP	2.4 ± 0.024[h]
16.	500102KA	1.93 ± 0.015[n]	36.	IC471876	2.415 ± 0.018[h]
17.	500138TN	1.939 ± 0.010n	37.	500149AP	2.5 ± 0.031[h]
18.	IC185926	1.939 ± 0.00[6n]	38.	500113MH	2.76 ± 0.017[ef]
19.	500120TN	2.01 ± 0.004[n]	39.	500121TN	2.95 ± 0.025[e]
20.	500151AP	2.01 ± 0.020[n]	40.	500185MH	3.585 ± 0.030[a]

[Y] Values are mean ± standard deviation of 10 independent experiments. Means followed by same letter are not significantly different at 5 % significance level as determined by Tukey's HSD test

ISSR Analysis

A total of 42 ISSR primers (Table 3) were used for estimation of inter and intra-varietal variations out of which 27 showed scorable markers. A total of 264 markers were generated of which 243 were polymorphic with number of products ranging from 1 (829) to 17 (P3). On an average single primer generated 9.7 products per primer with 92.04 % polymorphism. The results of marker attributes are detailed in Table 4. The PIC values ranged from 0.14 (819) to 0.49 (829) with an average of 0.26 per fragment. The marker index (MI) for different primers ranged from 0.49 (829) to 3.91(P3). Positive correlation was observed between PIC and MI values ($r^2 = 0.45$, $p < 0.005$). Resolving power (Rp) of the primers ranged from 0.0 (829) to 5.34 (ISSR2) with an average of 2.45 per primer. The Rp values were also found positively correlated with MI ($r^2 = 0.69$, $p < 0.005$).



Note: The values above each bar indicate the frequency of accessions in that particular class

Fig. 1 Frequency distribution of *M. pruriens* accessions differing in L-Dopa content. Note: The values above each bar indicate the frequency of accessions in that particular class

The Jaccard's similarity coefficient based on ISSR data revealed moderate genetic variability that ranged from 0.60 to 0.96 with an average at 0.61. At an inter-varietal level the extent of variability was more in wild "itching bean" (var. *pruriens*) as similarity index is from 0.60 to 0.72 (mean value of 0.63) vis-à-vis cultivated velvetbean (var. *utilis*) where similarity index ranged from 0.80 to 0.96 with mean value of 0.85.

The UPGMA dendrogram (Fig. 2) categorized 40 accessions into seven clusters at approximately 60 % confidence levels. Cluster II supported at 75.2 % confidence interval limits along with cluster IV (BS59.2 %) and VII (50.7 %) separated wild accessions belonging to var. *pruriens*; while cluster-III (66.1 %) and V (BS68.7 %) majorly partitioned cultivated accessions belonging to var. *utilis* except IC391941, IC391818 and IC265577 which belonged to var. *pruriens*. Within Cluster I and cluster VI accessions belonging to all the three varieties were intermixed. Several sub-

Table 3 ISSR primer sequences used in the study

Sl. No.	Primer Name	Sequence 5′-3′
1.	807	AGAGAGAGAGAGAGAGT
2.	808	AGAGAGAGAGAGAGAGC
3.	811	GAGAGAGAGAGAGAGAC
4.	814	CTCTCTCTCTCTCTCTYA
5.	817	CACACACACACACACAA
6.	821	GTGTGTGTGTGTGTGTT
7.	827	ACACACACACACACACG
8.	829	TGTGTGTGTGTGTGTGC
9.	830	TGTGTGTGTGTGTGTGG
10.	835	AGAGAGAGAGAGAGAGYC
11.	842	GAGAGAGAGAGAGAGAYG
12.	845	CTCTCTCTCTCTCTCTRG
13.	846	CACACACACACACACART
14.	847	CACACACACACACACARC
15.	856	ACACACACACACACACYA
16.	861	ACCACCACCACCACCACC
17.	862	AGCAGCAGCAGCAGCAGC
18.	863	AGTAGTAGTAGTAGTAGT
19.	865	CCGCCGCCGCCGCCGCCG
20.	880	GGAGAGGAGAGGAGA
21.	818	CACACACACACACACAG
22.	819	GTGTGTGTGTGTGTGTA
23.	826	ACACACACACACACACT
24.	838	TATATATATATATATARC
25.	841	GAGAGAGAGAGAGAGAYC
26.	855	ACACACACACACACACYT
27.	857	CACACACACACAAGCT
28.	866	CTCCTCCTCCTCCTCCTC
29.	873	GACAGACAGACAGACA
30.	887	DYDTCTCTCTCTCTCTC
31.	898	CACACACACACAGCT
32.	901	GTGTGTGTGTGTYR
33.	ISSR-1	CGTAGTCGTTCCTCCTCCTCCTCC
34.	ISSR-2	ACTACGACTTCCTCCTCCTCCTCC
35.	ISSR-5	TCCTCCTCCTCCTCCRY
36.	ISSR-7	TAATAATAATAATAATAATAATAA
37.	ISSR-8	CACCACCACCACCACCAC
38.	P1	GTAGCACTCC
39.	P2	TCGGCACGCA
40.	P3	CTGATACGCC
41.	P4	GTGTCTCAGG
42.	P5	CCCCGGTAAC

groups were also found within two major clusters II and III. A cophenetic-value (ultra-metric) matrix was generated from the coefficients of SAHN's similarity matrix. The cophenetic correlation coefficient between the cophenetic matrix and the data matrix of ISSR data was 0.82. Such a high value is considered good fit (Romesburg 1990) and shows that the original matrix was well represented by cluster analysis.

Table 4 Marker attributes for different ISSR primers

Sl.No.	Primer Name	PIC[a]	EMR[b]	MI[c]	Rp[d]
1.	811	0.24	8.00	1.92	1.86
2.	814	0.37	6.02	2.23	3.22
3.	817	0.33	5.00	1.65	0.42
4.	819	0.14	10.00	1.40	4.32
5.	827	0.19	5.01	1.52	3.60
6.	829	0.49	1.00	0.49	0.00
7.	830	0.19	8.00	1.52	1.44
8.	841	0.24	9.00	2.16	0.92
9.	842	0.22	4.95	1.09	0.74
10.	845	0.22	10.01	2.20	2.86
11.	846	0.32	1.98	0.63	0.90
12.	847	0.24	7.04	1.69	2.08
13.	855	0.28	8.00	2.24	1.82
14.	856	0.15	4.97	0.75	2.42
15.	857	0.19	11.00	2.09	2.50
16.	866	0.21	11.00	2.31	1.68
17.	873	0.19	13.00	2.47	2.38
18.	880	0.25	11.00	2.75	1.94
19.	898	0.29	12.00	3.48	1.70
20.	901	0.25	11.05	2.76	3.18
21.	907	0.33	9.00	2.97	3.34
22.	ISSR1	0.22	13.00	2.86	2.02
23.	ISSR2	0.33	10.00	3.30	5.34
24.	P1	0.32	7.00	2.24	1.94
25.	P2	0.23	15.00	3.45	4.54
26.	P3	0.23	17.00	3.91	5.32
27.	P4	0.28	11.00	3.08	3.58
Minimum		0.14	1.00	0.49	0.00
Maximum		0.49	17.00	3.91	5.34
Average		0.26	8.89	2.19	2.45

[a] PIC-Polymorphic Information Content

[b] EMR- Effective Multiplex Ratio

[c] MI- Marker Index

[d] Rp- Resolving Power

AFLP Analysis

Analysis of 40 *Mucuna* accessions with 12 selective primer pairs (Table 5) identified a total of 1612 fragments with number of products ranging from 82 (*Eco* RI + CAC & *Mse*I + GCT) to 172 (*Eco* RI + ACT & *Mse*I + CTC) confirming high multiplex ratio produced by the AFLP markers. On an average single primer generated 134 products that produced a total of 1609 polymorphic fragments with a percentage of polymorphic loci equal to 99.16 %. The results of marker attributes are detailed in

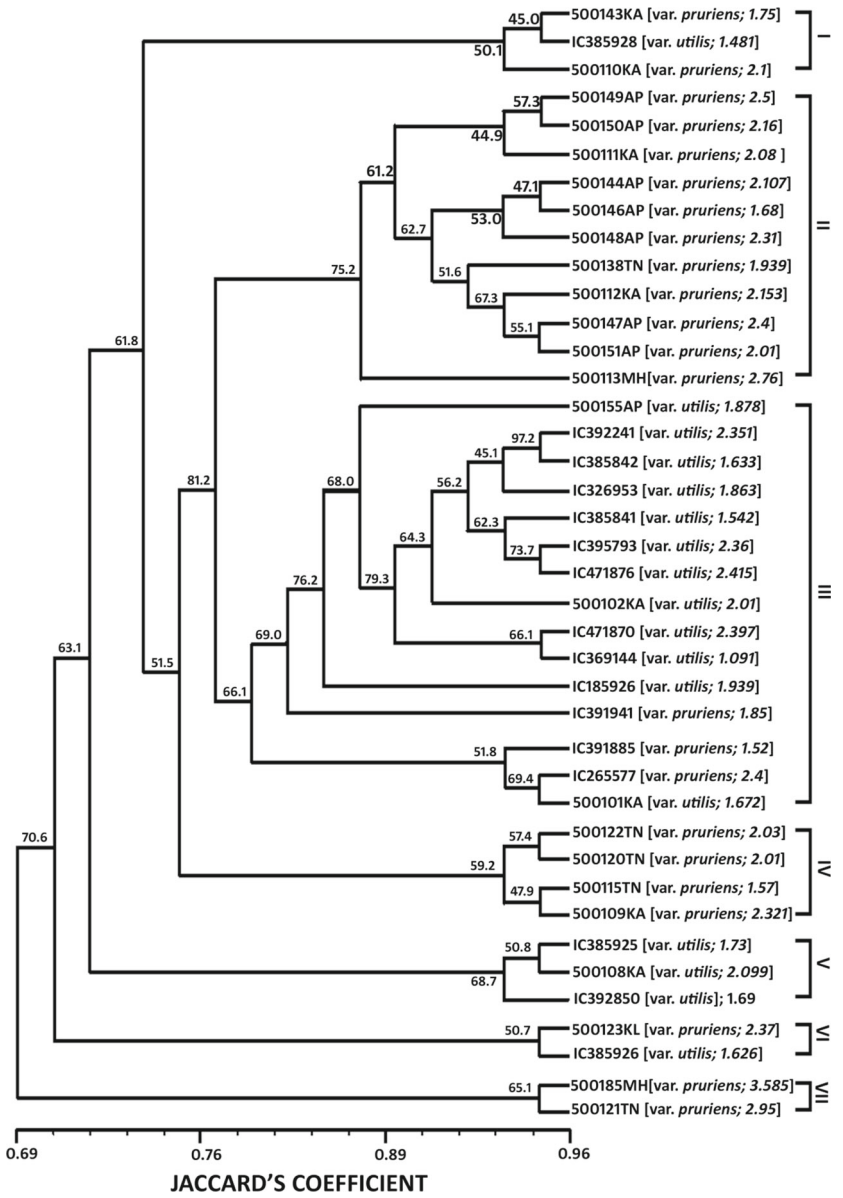


Fig. 2 UPGMA dendrogram based on ISSR and L-Dopa data sets. Note: The data indicated in the braces adjacent to the accession name contains information on varietal type and L-Dopa (%). Also, the values at the branch nodes of the dendrogram indicates the bootstrap support for the clustering pattern obtained

Note: The data indicated in the braces adjacent to the accession name contains information on varietal type and L-Dopa (%). Also, the values at the branch nodes of the dendrogram indicates the bootstrap support for the clustering pattern obtained

Table 6. The PIC values ranged from 0.145 to 0.214 with an average of 0.174 per fragment. The highest PIC value (0.214) was observed for the primer combination E-ACT/M-CAT and the lowest (0.145) was recorded for E-CAA/M-GCT. The MI values ranged from 14.58 to 34.03 with an average of 23.42 per primer combination. Highest value (34.03) was scored with the primer pair for E-ACT/ M-CAT and the lowest value (14.58) for the primer pair E-CAC/M-GCT. A positive correlation was observed between MI and PIC values ($r^2 = 0.98$, $p < 0.005$). Resolving power (R_P) ranged from 22.50 to 49.06 with an average of 34.8 per primer combination. Highest value (49.06) was scored with the primer combination E-ACT/ M-CTC and the lowest value (22.50) for E-CAC/ M-GCT. The R_P values were also found positively correlated with MI ($r^2 = 0.60$, $p < 0.005$).

Similarity index in case of AFLP markers ranged 0.10 to 0.81 with an average of 0.19. At inter-varietal level, wild accessions showed greater variability as similarity index ranged from 0.10 to 0.81(mean value of 0.19) vis-à-vis cultivated velvetbean (var. *utilis*) where it was 0.13 to 0.37 with mean value of 0.22. The comparison of marker attributes and genetic diversity indices between the two marker systems used in the present study is given in Figs. 3 and 4.

The UPGMA dendrogram (Fig. 5) grouped 40 accessions into three clusters at approximately 70 % confidence level.

Table 5 Oligonucleotide adapters and primers used for AFLP analysis

Name of the oligonucleotide	Code	Sequence
Eco RI adapter	E-0	5′-AAT TGG TAC GCA GTC TAC-3′
		3′-CC ATG CGT CAG ATG CTC-5’
Mse I adapter	M-0	5′-TAC TCA GGA CTC AT-3′
		3′-G AGT CCT GAG TAG CAG-5’
Eco RI primer	E-A00	5′-GAC TGC GTA CCA ATT C A-3’
Mse I primer	M-C00	5′-GAT GAG TCC TGA GTA A C-3’
Eco RI primer	E-C00	5′-GAC TGC GTA CCA ATT C C-3’
Mse I primer	M-G00	5′-GAT GAG TCC TGA GTA A G-3’
Eco RI + 3-CAC	E-CAC	5′- GAC TGC GTA CCA ATT C CAC-3’
Eco RI + 3-CAA	E-CAA	5′-GAC TGC GTA CCA ATT C CAA-3’
Eco RI + 3- ACT	E-ACT	5′-GAC TGC GTA CCA ATT C ACT-3’
Eco RI + 3- AAC	E-ACC	5′-GAC TGC GTA CCA ATT C AAC-3’
Mse I + 3-GCT	M-GCT	5′-GAT GAG TCC TGA GTA A GCT-3’
Mse I + 3-GCA	M-GCA	5′-GAT GAG TCC TGA GTA A GCA-3’
Mse I + 3-CAT	M-CAT	5′-GAT GAG TCC TGA GTA A CAT-3’
Mse I + 3-CAG	M-CAG	5′-GAT GAG TCC TGA GTA A CAG-3’
Mse I + 3-CTA	M-CTA	5′-GAT GAG TCC TGA GTA A CTA-3’
Mse I + 3-CTC	M-CTC	5′-GAT GAG TCC TGA GTA A CTC-3’

Cluster I and III supported at 89 % and 66 % confidence interval limits separated wild itching bean accessions belonging to var. *pruriens*; while cluster-II (74.6 %) majorly separated cultivated accessions belonging to var. *utilis* except 500138TN which belonged to var. *pruriens*. The high r-value of 0.91 from the Mantel's test suggest that the similarity coefficient data obtained for AFLP markers is well represented by the cophenetic matrix generated from the tree data.

Table 6 Marker attributes for AFLP primer combinations

Primer combination	PIC[a]	EMR[b]	MI[c]	Rp[d]
E-CAC/M-GCT	0.180	81.02	14.58	22.50
E-CAC/M-GCA	0.180	101	18.18	27.18
E-CAA/M-GCT	0.145	113	16.39	23.44
E-CAA/M-GCA	0.188	120	22.56	34.00
E-ACT/M-CAT	0.214	159	34.03	55.32
E-ACT/M-CAG	0.172	116	19.95	27.68
E-ACT/M-CTA	0.164	146	23.94	35.82
E-ACT/M-CTC	0.196	172	33.71	49.06
E-AAC/M-CAT	0.175	156	27.30	39.60
E-AAC/M-CAG	0.158	140	22.12	34.16
E-AAC/M-CTA	0.159	149	23.69	32.06
E-AAC/M-CTC	0.158	155.95	24.64	36.78
Maximum	0.214	172	34.03	49.06
Minimum	0.145	81.02	14.58	22.50
Average	0.174	134.08	23.42	34.80

[a] PIC- Polymorphic Information Content

[b] EMR- Effective Multiplex Ratio

[c] MI- Marker Index

[d] Rp- Resolving Power

Correlation among the Marker Data and L-Dopa Diversity and Identification of Contrasting Parental Lines

The correlation between the markers was assessed using Mantel's test (r-statistic). AFLP data was weakly correlated with ISSR (r = 0.15) along with variation in the grouping of accessions. Likewise data from both the marker types were poorly correlated to L-Dopa content [AFLP: L-Dopa (r = 0.33); ISSR: L-Dopa (r = 0.29)] which was very well

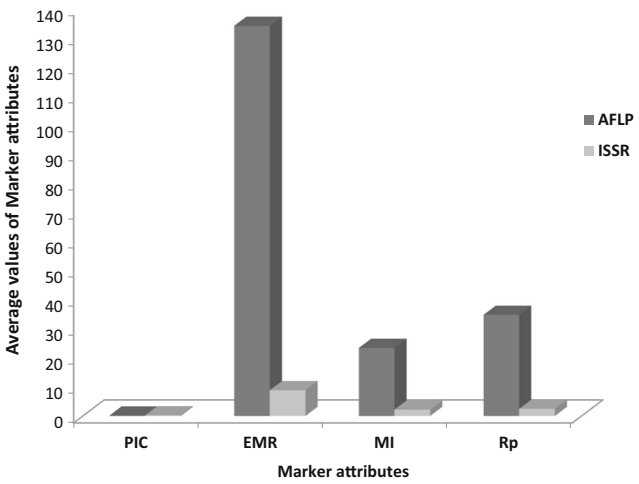


Fig. 3 Comparison of marker attributes between AFLP and ISSR

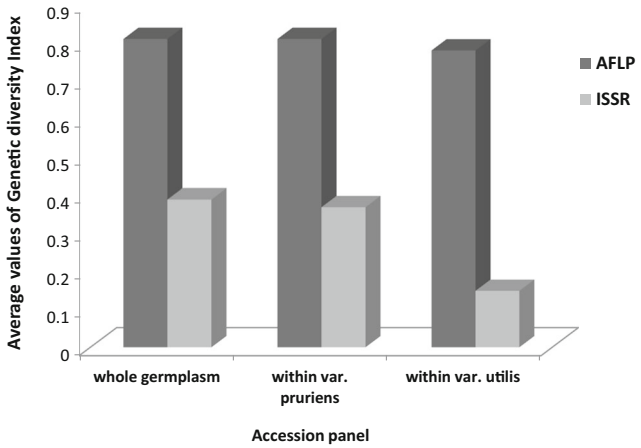


Fig. 4 Comparison of genetic diversity indices obtained from AFLP and ISSR

reflected in the clusters obtained. The clustering pattern did not correspond to the frequency distribution classes of L-Dopa. However, overlaying the genetic distances with L-Dopa frequency classes paved way for identification of several contrasting pairs of accessions. To achieve this, genetic diversity threshold of >80 % in case of AFLP and >25 % in case of ISSR between any two accessions with L-Dopa difference of ≥1.0 % was considered. Based on this, 7 parental genotype combinations were elucidated. Among these, 500108KA (2.1 %) × 500185MH (3.59 %) showed highest genetic distance 88 % (AFLP) and 33 % (ISSR) followed by 500121TN (2.95 %) × IC369144 (1.09 %) which shared a genetic distance of 86 % (AFLP) and 30 % (ISSR). The highest L-Dopa difference observed among these accessions was 2.5 % between 500185MH (3.59 %) and IC369144 (1.09 %) with genetic distance of 85 % (AFLP) and 30 % (ISSR) respectively. Other parental combinations are listed in Table 7.

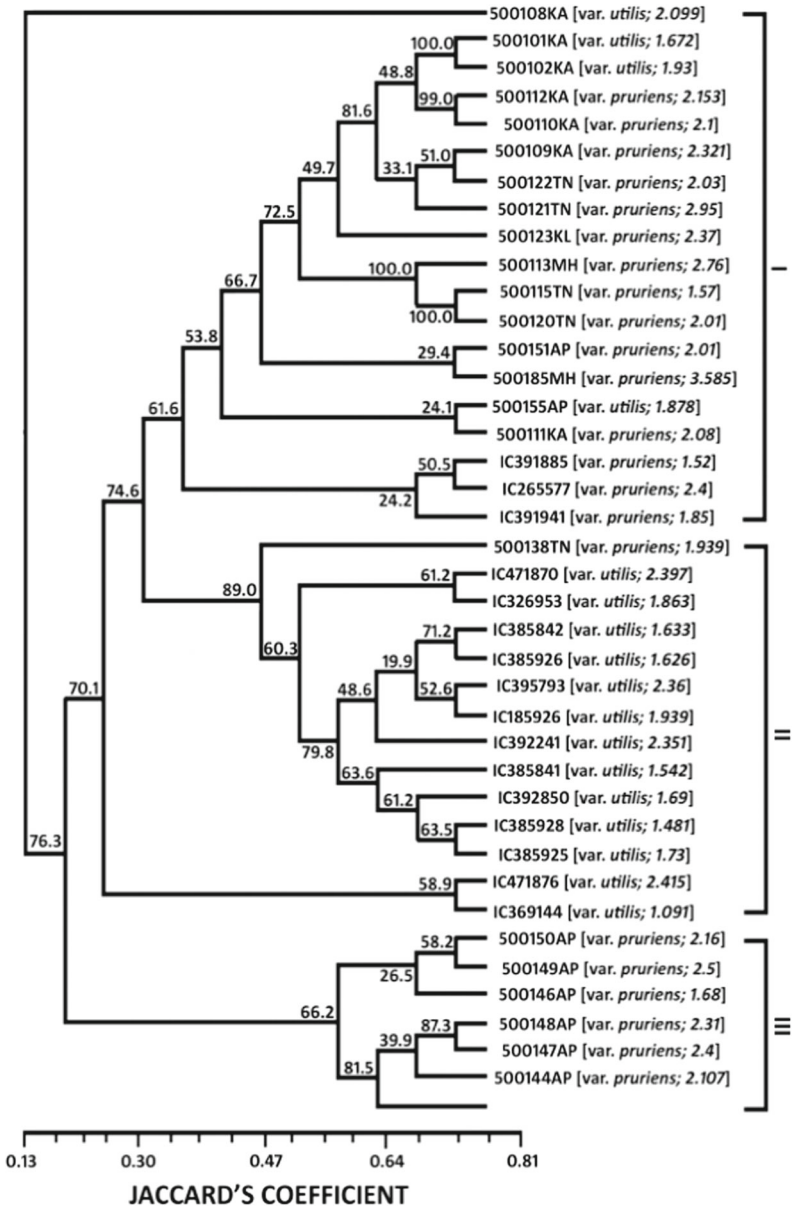
Discussion

Diversity estimates provides a rationale for conservation strategies and support selection of starting material for plant breeding programs (Laurentin et al. 2008). Conventionally these are obtained by measuring polymorphism at morphological or DNA level. Biochemical traits represent different facet of genome polymorphism normally not covered by neutral DNA markers. In such cases, complementation of targeted/non targeted metabolic profiling with DNA markers has been shown to be essential to attain pragmatic diversity estimate. This aspect is well demonstrated in pepper (Joy et al. 2007; Tam et al. 2005), mulberry (Bhattacharya and Ranade 2001; Awasthi et al. 2004; Bhattacharya et al. 2005) and fenugreek (Dangi et al. 2004) - the commercial value of many of which is largely affected by their secondary metabolite content. Therefore, our objective of this work was to assess L-Dopa and genomic diversity in *M.pruriens* germplasm together to elucidate their relationship and suggest parents for breeding/mapping programs particularly targeting L-Dopa content.

The results of L-Dopa screening revealed moderate to high variability in Indian germplasm. Padmesh et al. (2006) found homogeneous distribution among the genotypes from the Western Ghats of India. High variability revealed in the present study is explained by broad genetic base encompassed in our germplasm which represent pan India collection. However, significantly higher as well as lower L-Dopa contents than presented in this work are also reported in literature (Mary Josephine and Janardhanan 1992; Krishnamurthy et al. 1996; Fathima et al. 2010; Kalidass and Mohan 2011). This suggests possible role of environmental factors on synthesis. On the other hand, diverse analytical platform used for L-Dopa estimation might have also influenced this variability. It has been recorded that the plants grown near equator account for high L-Dopa synthesis; Lorenzetti et al. (1998) explains this as due to latitude related factors such as radiation. However, the fact that synthesis as function of irradiance could be either stimulatory (Wichers et al. 1983) or inhibitory (Brain 1976; Pras et al. 1993) for the active principle production suggests that even factors other than latitude might be responsible for variability in L-Dopa content in *M. pruriens*.

The Jaccard's similarity coefficient based on ISSR and AFLP data revealed moderate to high genetic variability – the degree of which was more in wild (var. *pruriens*) compared to cultivated (var. *utilis*) accessions. This observation corroborates with Padmesh et al. (2006) who found var. *utilis* (mean similarity index of 0.82) to be genetically less diverse than var. *pruriens* (*MSI* = 0.70) as well as Capo-Chichi et al. (2003a). Narrow genetic base among the cultivated accessions might be due to hybridization involving closely related parents. Thus efforts are needed to broaden the genetic base of cultivated accessions to reinforce breeding program. Nonetheless, high overall genetic diversity observed in this otherwise self-pollinated species suggests diversification as a factor of adaption to different environmental conditions.

Cluster analysis based on ISSR data partitioned accessions into seven groups majorly determined by varietal affiliation with a few intermixes. Similar pattern was also observed in case of AFLP data where the resultant three major clusters were mainly based on taxonomic association. *M. pruriens* varieties interbreed with ease producing hybrids which forms significant portion of its natural population (Padmesh et al. 2006). Quite often they even fashion transitional morphotypes presenting considerable taxonomic ambiguities. Accessions positioned as intermixes in our dendrograms conceivably represent such hybrids. Wilmot Dear (1987) suggests presence of three botanical varieties in *M. pruriens* viz. var. *pruriens*, var. *hirsuta* (both wild forms) and var. *utilis* (cultivated form). Of these, var. *hirsuta* and *pruriens* are nearly identical with only nature of indumentum



Note: The data indicated in the braces adjacent to the accession name contains information on varietal type and L-Dopa (%). Also, the values at the branch nodes of the dendrogram indicates the bootstrap support for the clustering pattern obtained

Fig. 5 UPGMA dendrogram based on AFLP and L-Dopa data sets. Note: The data indicated in the braces adjacent to the accession name contains information on varietal type and L-Dopa (%). Also, the values at the branch nodes of the dendrogram indicates the bootstrap support for the clustering pattern obtained

serving as point of distinction between them. Leelambika and Sathyanarayana (2011), after examining several accessions belonging to these varieties confirmed close genetic similarities between these two suggesting merger of var. *pruriens* and var. *hirsuta* in one subgroup under the name var. *pruriens* thus allowing only two groups viz. var. *utilis* (cultivated + non-itching trichomes) and var. *pruriens* (wild + itching trichomes) to be recognized within *M. pruriens*. The results of the present investigation further support this view point.

The Mantel test between the two marker data sets revealed $r = 0.237$ implying non-significant correlation between AFLP and ISSR data sets. This is plausible as the genomic coverage as well as diversity detection system of the two markers vary drastically. Such observations are reported even in *Momordica charantia* L. sp. (Behera et al. 2008a) and *Dioscorea* sp. (Velasco-Ramírez et al. 2014).

We further examined correlation between the chemical phentoyping data (L-Dopa) with that from marker analysis (ISSR and ALFP) to determine parental genotypes for L-

Table 7 Putative parental combinations identified for L-Dopa breeding/mapping

Putative parental combination	% Genetic distance		% L-Dopa difference
	AFLP	ISSR	
500185MH (W, 3.59) X 500108KA (C, 2.1)	88	33	1.49
500121TN (W, 2.95) X IC369144 (C, 1.09)	86	30	1.86
500143KA (W, 1.75) X 500185MH (C, 3.59)	86	32	1.84
500143KA (W, 1.75) X 500121TN (C, 2.95)	85	31	1.20
500185MH (W, 3.59) X IC369144 (C, 1.09)	85	30	2.50
500113MH (W, 2.76) X IC369144 (C, 1.09)	83	27	1.67
500113MH (W, 2.76) X IC385928 (C, 1.48)	81	37	1.28

The value in parenthesis represents percentage of L-Dopa. The accessions indicated as W-wild: var. *pruriens*; C-cultivated: var. *utilis* respectively

Dopa breeding. The idea was: positive correlation between the two would imply DNA markers used in the study are correlated to genetic loci controlling this biochemical trait. This proposition makes parental selection fairly straight forward as the integration of both the datasets provide stringent genetic diversity estimate for parental selection (Laurentin et al. 2008). However, Mantel test between marker data sets and L-Dopa content in our study returned non-significant correlation [AFLP: L-Dopa (r = 0.33); ISSR: L-Dopa (r = 0.29)]. In sesame (*Sesamum indicum* L.) such lack of correlation is attributed to higher diversification rate at the metabolic level (Laurentin et al. 2008). Similar observations are also recorded in case of *Rhizobia* (Wolde-Meskel et al. 2004). On the contrary, strong correlation between genome variation and metabolite diversity was observed in a study involving endophytic fungi (Seymour et al. 2004). The results of the present investigation possibly point to lower penetration of the marker data, both ISSR and AFLP to the changes in the L-Dopa levels suggesting the need for in depth marker analysis using extended primer collection. On the other hand, it is quite possible that the marker patterns and metabolic profiles reflect different facets of genetic polymorphism in *M. pruriens* as neutral markers like ISSR and AFLPs are particularly insensitive to gene expression and may occur most frequently in non-coding portions while variation in L-Dopa profiles might be due to changes in biosynthetic pathway related genes influenced by different selection pressures. Nonetheless, Dai-fu et al. (2009) suggests that in such situations where the distances determined by fingerprinting does not reflect the differences in the metabolic profiling, cluster analysis based on marker data can be employed as a tool to identify genetically diverse lines which also tend to appear contrasting for the desired trait. Based on this, we have identified 7 putative parents which are contrasting at both genomic and metabolic levels using independent criterion and initiated segregating population from one of them to identify QTLs linked to L-Dopa trait.

Conclusions

This research represents first comprehensive investigation on relationship between DNA and L-Dopa diversity in Indian *M. pruriens* germplasm. To the best of our knowledge, this is also first report on identification of parents for L-Dopa mapping/breeding. The contrasting parents revealed here will form a basis for future mapping studies on L-Dopa trait. Together with ongoing efforts on species specific microsatellites and SNP marker development, this will pave a way for successful integration of molecular markers in breeding programs starting from development of saturated linkage maps and identification of markers/QTLs linked to L-Dopa and other economic traits in *M. pruriens*.

Acknowledgments Authors thank Department of Biotechnology (DBT), Govt. of India for the financial support; Sikkim University and Sri Krishnadevaraya Educational Trust (Sri KET), Bangalore, India for the facilities; Dr. S A Ranade, National Botanical Research Institute (NBRI), Lucknow, India and Dr. G Ravikanth, Ashoka Trust for Research in Ecology and the Environment (ATREE), Bangalore India for ISSR markers.

Compliance with ethical standards

Conflict of Interest The authors declare no conflict of interest.

Funding All the funding agencies and support levied are acknowledged

References

Awasthi AK, Nagaraja GM, Naik GV, Kanginakudru S, Thangavelu K, Nagaraju J (2004) Genetic diversity and relationships in mulberry (genus *Morus*) as revealed by RAPD and ISSR marker assays. BMC Genet 5:1

Behera TX, Gaikward AB, Singh AK, Staub JE (2008a) Relative efficiency of DNA markers (RAPD, ISSR and AFLP) in detecting genetic diversity of bitter gourd (*Momordica charantia* L.). J Sci Food Agric 88:733–737

Bhattacharya E, Ranade SA (2001) Molecular distinction amongst varieties of mulberry using RAPD and DAMD profiles. BMC Plant Biol 1:3

Bhattacharya E, Dandin SB, Ranade SA (2005) Single primer amplification reaction methods reveal exotic and indigenous mulberry varieties are similarly diverse. J Biosci 30:669–677

Brain KR (1976) Accumulation of L-DOPA in cultures from *Mucuna pruriens*. Plant Sci Lett 7:157–161

Capo-chichi LJA, Weaver DB, Morton CM (2001) AFLP assessment of genetic variability among velvetbean (*Mucuna* sp.) accessions. Theor Appl Genet 103:1180–1188

Capo-chichi LJA, Weaver DB, Morton CM (2003a) The use of molecular markers to study genetic diversity in *Mucuna*. Trop Subtrop Agroecosystems 1:309–318

Dai-fu MA, Qiang LI, Xiu-ying LI, Hong-min LI, Zhong-hou TANG, Ling HU, Qing-he AO, Yi-ping XIE, Xin WANG (2009) Selection of parents for breeding edible varieties of Sweetpotato with high carotene content. Agric Sci China 8:1166–1173

Dangi RS, Lagu MD, Choudhary LB, Ranjekar PK, Gupta VS (2004) Assessment of genetic diversity in *Trigonella foenum-graecum* and *Trigonella caerulea* using ISSR and RAPD markers. BMC Plant Biol 4:13

Daxenbichler ME, Van Etten CH, Hallinan EA, Earle FR, Barclay SA (1971) Seeds as sources of L-Dopa. J Med Chem 14:463–465

Doyle JJ, Doyle LJ (1990) Isolation of DNA from plant tissue. Focus 12: 13–15

Fathima KR, Tresina Soris P, Mohan VR (2010) Nutritional and antinutritional assessment of *Mucuna pruriens* (L.) DC. Var. *pruriens* an underutilized tribal pulse. Adv BioRes 1:79–89

Gunendi G, Pamuk F (1999) Thin layer chromatographic separation and quantitation of L-Dopa and L-tyrosine in mixtures. Turk J Chem 23: 269–273

Jaccard P (1908) Nouvelles recherches Sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Joy N, Abraham Z, Soniya EV (2007) A preliminary assessment of genetic relationships among agronomically important cultivars of black pepper. BMC Genet 8:42

Kalidass C, Mohan VR (2011) Nutritional and antinutritional composition of itching bean (*Mucuna pruriens* L. DC. Var. *pruriens*), an underutilized tribal pulse in western Ghats, Tamil Nadu. Tropical and subtropical agroecosystems 14:279–293

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. J Neurol Neurosurg Psychiatry 75:1672–1677. doi: 10.1136/jnnp.2003.028761

Krishnamurthy R, Suvagia P, Pathak JM, Parikh KM, Bhikne NS (1996) Selection of *Mucuna* lines for higher L-Dopa and improved yield through conventional breeding. In: Proceedings of the Conference on Breeding Research on Medicianal and Aromatic Plants, Quedlinburg, Germany, Bundesanst Zuchtungsforschung Kulturpflanzen 2: 231–234.

Laurentin H, Ratzinger A, Karlovsky P (2008) Relationship between metabolic and genomic diversity in sesame (*Sesamum indicum* L.). BMC Genomics 9(1):250

Leelambika M, Sathyanarayana N (2011) Genetic characterization of Indian Mucuna (*Leguminoceae*) species using morphometric and random amplification of polymorphic DNA (RAPD) approaches. Plant Biosystems 1–12. doi:10.1080/11263504.2011.610924

Liu K, Muse SV (2005) Power marker. Integrated analysis environment for genetic marker data. Bioinformatics 21(9):2198–2129

Lorenzetti F, MacIsaac S, Arnason JT, Awang DVC, Buckles D (1998) The phytochemistry, toxicology, and food potential of *M. pruriens* (*Mucuna* Adans. Spp., Fabaceae), In: D. Buckles, A. Eteka, M. Osiname, M. Galiba and G. Galiano (eds.) Cover crops in West Africa - contributing to sustainable agriculture, IDRC, IITA, Sasakawa global 2000, Otawa, Canada; Ibadan, Nigeria; Cotonou, Benin, pp 67–84

Mary Josephine R, Janardhanan K (1992) Studies on chemical composition and antinutritional factors in three germplasm seed materials of the tribal pulse, *Mucuna pruriens* (L.) DC. Food Chem 43:13–18

Padmesh P, Reji JV, Jinish Dhar M, Seeni D (2006) Estimation of genetic diversity in varieties of *M. pruriens* using RAPD. Biol Plant 50: 367–372

Pras N, Woerdenbag HJ, Batterman J, Visser F, VanUden W (1993) *Mucuna pruriens*: improvement of the biotechnological production of the anti-Parkinson drug L-DOPA by plant cell selection. Pharm World Sci 15:263–268

Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Genet 98:107–112

Rohlf FJ (2009) NTSYS-pc Numerical Taxonomy System version 2.21c. Exeter software, Setauket, New York

Romesburg HC (1990) The case method of teaching research methods. Proceedings of IUFRO Subject Group S6.09: Philosophy and Methods of Forest Research

SAS Institute Inc, Using JMP 10 (2012) Cary, NC: SAS Institute Inc.

Sathyanarayana N, Leelambika M, Mahesh S, Jaheer M (2011) AFLP assessment of genetic diversity among Indian *Mucuna* accessions. Physiol Mol Biol Plants 17(2):171–180

Seymour F, Cresswell J, Jack P, Lappin-Scott H, Haag H, Talbot N (2004) The influence of genotypic variation on metabolite diversity in populations of two endophytic fungal species. Fungal Genet Biol 41: 721–734

Anderson Ricardo Soares, Rogerio Marchiosi, Rita de Cassia Siqueira-Soares, Rogerio Barbosa de Lima, Wanderley Dantas dos Santos and Osvaldo Ferrarese-Filho (2014) The role of L-Dopa in plants. Plant Signal Behav 9:e28275-1-7

Sokal R, Michener C (1958) A statistical method for evaluating statistical relationships. Kans Univ Sci Bull 38:1409–1438

Szabo NJ, Tebbett IR (2002) The chemistry and toxicity of *Mucuna* species, In M. Flores, M. Eilittä, R. Myhrman, L. B. Carew and R. J. Carsky (eds.), *Mucuna* as a Food and Feed: Current Uses and the Way Forward, Workshop held April 26–29, 2000 in Tegucigalpa, Honduras. CIDICCO, CIEPCA, and World Hunger Research Center, Tegucigalpa, Honduras, CIDICCO, Honduras, pp 120–141

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon- based SSAP, AFLP and SSR. Theor Appl Genet 110:819–831

Teixeira AA, Rich EC, Szabo NJ (2003) Water extraction of L-Dopa form *Mucuna* bean. Trop Subtrop Agroecosyst 1:159–171

Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173:638–649

Velasco-Ramírez AP, Torres-Morán MI, Molina-Moret S, de Jesús Sánchez-González J, Santacruz-Ruvalcaba F (2014) Efficiency of RAPD, ISSR, AFLP and ISTR markers for the detection of polymorphisms and genetic relationships in camote de cerro (*Dioscorea* spp.). Electron J Biotechnol 17:65–71

Warrier PK, Nambiar VPK, Ramankutty C (1996) Indian medicinal plants, a compendium of 500 species, Orient Longman Limited, Madras, India, Vol 4

Wichers HJ, Malingre TM, Huizing HJ (1983) The effect of some environmental factors on the production of L-DOPA by alginate-entrapped cells of *Mucuna pruriens*. Planta 158:482–486

Wilmot Dear CM (1987) A revision of *Mucuna* (Leguminosae-Phaseolae) in the Indian subcontinent and Burma. Kew Bull 42(1): 23–46

Wolde-meskel E, Terefework Z, Lindstrom K, Frostegard A (2004) Metabolic and genomic diversity of rhizobia isolated from field standing native and exotic woody legumes in southern Ethiopia. Syst Appl Microbiol 27:603–611

Yap VI, Nelson RJ (1996) WINBOOT- A Program for Performing Bootstrap Analysis of Binary Data to Determine the Confidence Limits of UPGMA-Based Dendrograms.- IRRI Discussion paper series No. 14, Manila, Philippines; International Rice Research Institute