DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS IN *MUCUNA PRURIENS* (L.) DC.

A Thesis Submitted

To

Sikkim University

In Partial Fulfilment of the Requirements for the

Degree of Doctor of Philosophy

By

Pittala Ranjith Kumar

Department of Botany
School of Life Sciences

May 2019

CERTIFICATE

This is to certify that the Ph.D. thesis entitled **"DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS 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. PITTALA RANJITH KUMAR** 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

CERTIFICATE

This is to certify that the Ph.D. thesis entitled **"DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS 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. PITTALA RANJITH KUMAR** 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.

PROF. SHANTI S. SHARMA
Head
Dept. of Botany
School of Life Sciences
Sikkim University, Gangtok

DECLARATION

I declare that the Ph.D. thesis entitled **"DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS IN *MUCUNA PRURIENS* (L.) DC."** submitted by me for the award of the degree of **DOCTOR OF PHILOSOPHY** is a bona fide research work carried out by me at the Department of Botany, Sikkim University under the supervision of **Prof. N. SATHYANARAYANA.**

The thesis contains no material which has been accepted for a degree or diploma of any other University or Institution, except by way of background information and duly acknowledged in the thesis. Further, to the best of my knowledge and belief, the thesis does not contain any material previously published or written by another person, except where due acknowledgement is made in the text of the thesis, nor does it contain any material that infringes copyright.

PITTALA RANJITH KUMAR
Registration No: 13SU11967
Ph.D. Scholar
Department of Botany
School of Life Sciences
Sikkim University, Gangtok

PLAGIARISM CERTIFICATE

This is to certify that the plagiarism check has been carried out for the following Ph.D. thesis with help of **URKUND SOFTWARE** and the result is **5%** tolerance rate, within the permissible limit (below 10% tolerance rate) as per the norm of the Sikkim University.

DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS IN *MUCUNA PRURIENS* (L.) DC."

Submitted by **PITTALA RANJITH KUMAR** under the supervision of **Prof. N. SATHYANARAYANA**, Department of Botany, School of Life Sciences, Sikkim University.

PITTALA RANJITH KUMAR
Doctoral Student

PROF. N. SATHYANARAYANA
Research Supervisor

Vetted by Librarian

ACKNOWLEDGEMENTS

This thesis marks the end of my passage in obtaining Ph.D. degree, but fabrication of this research represents the culmination of determinations of several people. Only my name appears on the cover of this thesis, but it has been kept on track and been brought to completion with the support and encouragement of my family, friends, teachers, well-wishers equals and more. At the end of my thesis writing, it is a matter of immense pleasure for me to present my thanks to all those who contributed to the completion of this study in one way or the other and made it an unforgettable experience for me.

At first, I bow my head before "**The God Sai**" who blessed me with the health and confidence to undertake and complete this work successfully.

I feel proud and privileged to express my deepest sense of gratitude towards my Ph. D supervisor, **Prof. N. Sathyanarayana**, Department of Botany, Sikkim University for his constant support, worthy guidance, valuable suggestions, encouragement and helping attitude throughout the course of present investigation. Thank you sir for being there for me and I hope to relish it for the lifetime.

I wish to express my sincere thanks to **Prof. Shanti S. Sharma**, 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 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 acknowledgments to **Department of Biotechnology (DBT),** Govt. of India, New Delhi for providing financial support in the form of Senior Research Fellowship under the Project **"**Development and characterization of microsatellite markers in Mucuna pruriens – A medicinally and agronomically important Legume species" undertaken by **Prof. N. Sathyanarayana.**

I also thank Dr. **Pradeep Kumar** B, Scientist, Regional Centre of Institute of Bioresources and Sustainable Development (DBT), for his help in RNA isolation. 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.

I would like to thank my dear colleague **Dr. Pankaj Kumar Tripathi** who was my friend and roommate. He shared his valuable time with me, and helped greatly in my research days. I thank you for your help in both professional and personal life.

I extend my thanks to **Mr. Dawa Lhendup Lepcha** and **Mr. Lalit Chaurasia** for their timely help in every aspect, when I was truly in need of it, in both professional and personal matters.

I consider it as a pleasure to express my deep sense of gratitude to – **Sai Sundeep, Patrush Lepcha, Nima Tshering Lepcha, Aditya Pradhan, Shubhakar Gurung, Ashish Rai, Hemant Sharma, Aita Rani Subba, Alice Rachel Subba, Raksha Mukhia, Puja Sashankar, Abhijeet Chhetri, Sandhya Gurung** and **Bhusan Gurung** for creating a friendly and homely environment with lots of scientific interactions as well as personal attachments. I also extend my thanks to the authors of the books and research articles referred during this research work.

What I am today is due to my family, the engineers of my wisdom, my pillar of support, my source of strength and inspiration. I take this precious moment to express my heartiest thanks to my grandparents Late Shri **Pittala Narsaiah** and **Venkatamma** whose affection, encouragement and blessings have always been a source of light to me.

I am extremely grateful to my lovable mother Late **Jhansi Rani** and my father **Stalin** for their prayers, caring and sacrifices for educating and preparing me for my future.My deepest sense of gratitude, reverence and special thanks to my brother **Mr. Bhagath** for his endless support, motivation and care in all aspects of my life. I express my special thanks to my sister **Mrs. Rajyalaxmi** for her motivational words and encouragement.

This Ph.D. work would not have been a success without the endless understanding, support and unconditional love from my beloved wife, **Mrs. Roopali (Swathi).** Your patience, attention, wisdom and encouragement helped me to overcome all the setbacks and stay focussed on my research study. Thank you for always loving me and having accepted my choice.

I also extend my affectionate appreciation and exceptional thanks to my sister-in-laws **Swetha** and **Shruthi** for their uninterrupted support to focus on completing this research work. I honestly express my gratitude to you both, for your support in professional and personal matters.

It may be last but, a very special acknowledgment to my dearest son, lovable charm Master **Riyansh Sai (Vijju)** for his cheerful smiling, unconditional love and innocence which always made me happy and gave me strength during hard days of my research life.

Words are inadequate to express my gratitude and appreciation to everyone who has supported me in any aspect during the tenure of this work. I thank all my teachers and friends who have come across in this journey of my research career directly or indirectly.

Not everyone is mentioned, but none is forgotten, I humbly acknowledge all.

Place: Gangtok
Date: May, 2019 **Pittala Ranjith Kumar**

CONTENTS

CHAPTERS

List of Tables

List of Figures

Additional files (Refer: Sathyanarayana et al. 2017, BMC Genomics)

File No.	Title
Additional file 1:	Tables and figures are incorporated in the main text
Additional file 2:	Number of transcripts assigned to Biological, Cellular and Metabolic processes from GO analysis of the annotated transcripts of the *Mucuna* assembly
Additional file 3:	Number of transcripts assigned for enzyme classes and KEGG pathways of the *M. pruriens* transcripts
Additional file 4:	BLASTP analysis of the transcripts against the legume protein databases
Additional file 5:	Transcription factors identified in the *Mucuna* assembly using PlnTFDB
Additional file 6:	Transcripts with the repeats in the *Mucuna* assembly identified using Perl script MISA
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 7943 repeat sequences is also reported
Additional file 8:	Polymorphic sequence repeats identified between the *Mucuna* genotypes using lobSTR program, including the full list and the filtered list
Additional file 9:	Fold change expression values of the transcripts in each of the tissue analyzed related to the secondary metabolite pathways

Note: All the data from the additional files (2-8) generated during the data analysis could not be incorporated into the main text due to large data size. Hence, links to these files/data sets are provided from our published paper, *N Sathyanarayana et al. 2017, BMC Genomics* wherever cited. The list is provided in the above table.

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
bHLH	basic-Helix-Loop-Helix
C-CAMP	Centre for Cellular and Molecular Platforms
CDD	Conserved Domain Database
cDNA	Complementary Deoxyribose Nucleic Acid
CIAT	Combat Intelligence Augmentation Teams
CTAB	Cetyltrimethylammonium Bromide
DEG	Differential Expression Analysis
DEPC	Diethylpyrocarbonate
DET	Differentially Expressed Transcripts
DNA	Deoxyribose Nucleic Acid
EC	Enzyme Classes
EDTA	Ethylenediaminetetraacetic acid
EMR	Effective Multiplex Ratio
EST	Expressed Sequence Tag
EtBr	Ethidium Bromide
FAO	Food and Agriculture Organization
GO	Gene Ontology
Gst	Gene differentiation
H	Gene Diversity
He	Expected Hetrozygosity
Ho	Observed Hetrozygosity
I	Shannon's Information Index
IFDA	International Fund for Agricultural Development
KEGG	Kyoto Encyclopaedia of Genes and Genomes
L-DOPA	3, 4 - Dihydroxy-L-Phenylalanine
MISA	MIcroSAtellite identification search tool
NCBI	National Center for Biotechnology Information
Ne	Effective Number Of Alleles
NGGF	Next Generation Genomics Facility
NGS	Next-Generation Sequencing
Nm	Gene flow indices

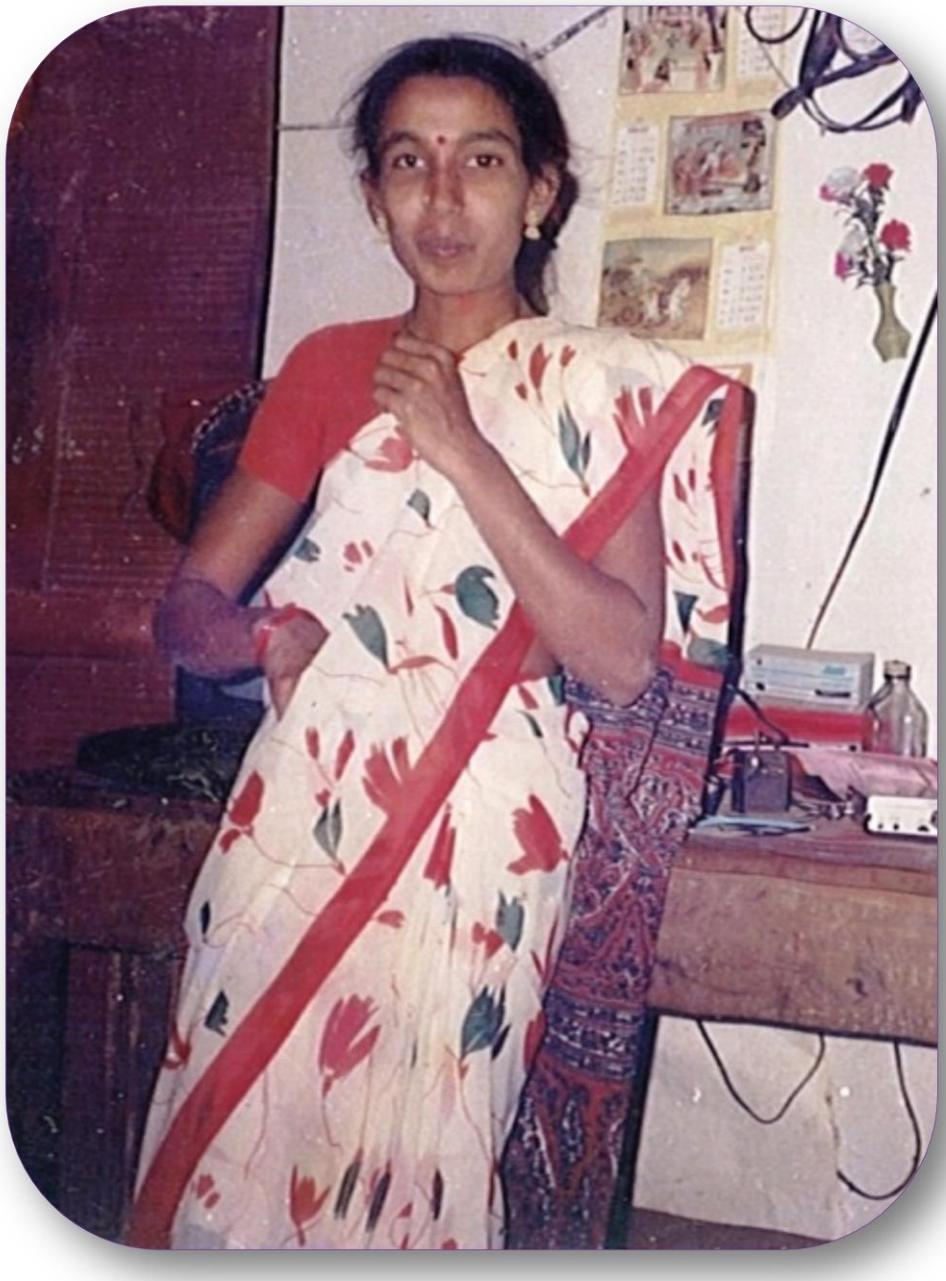
PCR	Polymerase Chain Reaction
PEM	Protein Energy Malnutrition
PlnTFDB	plant Transcription Factors Database
PoU	Prevalence of Undernutrition
pSSR	Polymorphic SSRs
QTL	Quantitative Trail Loci
RAPD	Random Amplification of Polymorphic DNA
RIL	Recombinant Inbred Line
RIL	Recombination Inbred Lines
RIN	RNA Integrity Number
RNA	Ribose Nucleic Acid
Rp	Primer Resolving Power
SCLS	Systemic Capillary Leak Syndrome
SSR	Simple Sequence Repeat
STR	Short Tandem Repeats
Tm	Melting Temperature
UNICEF	United Nation International Children's Emergency Fund
UPGMA	Unweighted Pair Group Method With Arithmetic Mean
UTR	Untranslated Regions
VCF	Variant Call Format
WHO	World Health Orginization

Symbols and units

cm	Centimeter
cM	centi-Morgan
ft	Feet
gm	Gram
h	Hour
”	Inches
kg	Kilogram
m	Meter
MT	Metric ton
µg	Microgram
µl	Microlitre
µM	Micromolar
MQ	MiliQ Water
mg	Milligram
ml	Millilitre
mM	Millimolar
min	Minutes
nm	Nanometer
ppm	Parts Per Million
%	Percentage
RT	Retention Time
s	Second
SD	Standard Deviation
v/v	Volume by volume
w/v	Weight by volume

Dedicated to

My Beloved Mother



Jhansi Rani

1. Introduction

1.1. Food and nutritional security - Global scenario

Nutritional security is termed as physical and economic access to enough safe and nutritious food by all people, always (FAO 2013). International Fund for Agricultural Development (IFAD) estimates roughly 1.2 billion people around the world are deprived of this minimum need for their day to day survival (IFAD 2002). Moreover, the world population is expected to reach 9.7 billion by 2050 (United Nations 2017), which means almost 3 billion more mouths are to be fed from the current or possible decrease in food productivity in next twenty years or so. Adding to this, the risks posed by the climate change, population growth, rising food prices, and degrading environment are further expected to complicate the food security scenario in the years ahead (Schmidhuber and Tubiello 2007).

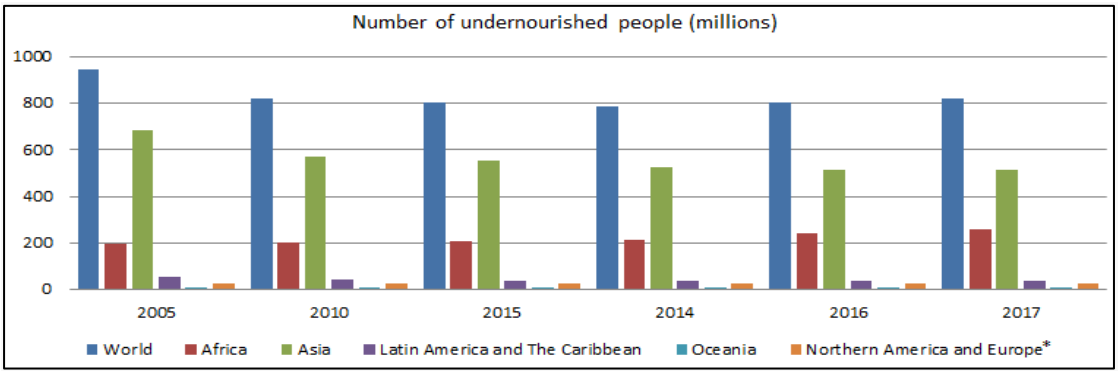


Figure 1: Number of undernourished people in the World (2005–2017) (Source: FAO. 2018). Note: * refer to less than 2.5% of the population each year

Food and Agriculture Organization (FAO) and the World Health Organization (WHO) reports shows roughly 821 million people across the globe are suffering from different forms of malnutrition (FAO 2018) (Figure 1). Roughly 50% of the child mortality, reported globally, is linked to undernutrition (Grantham and McGregor 1995). Of the 195 million undernourished children in developing countries, 90% live

in Asia and Africa (UNICEF 2009). Almost 50% of them are in five countries of which four are in Asia (India - 31.2%, China - 6.5%, Pakistan - 5.1%, and Indonesia - 3.9%) (UNICEF 2009). Prevalence of undernutrition (PoU) in these countries has led to significant health problems as there exists direct relationship between the undernutrition and infectious diseases such as diarrhea, malaria, HIV, measles as well as intestinal helminths, malabsorptive state and anorexia (Baqui et al. 2002; Maleta 2006). Poor families are particularly affected by the PoU which is significant in low-income groups (Harishankar et al. 2004). In a country like India, where 70% of its population lives in villages and depends on subsistence farming or other menial jobs for livelihood, providing safe and nutritious food should become the top priority of the government (Mehra 2008). This message is well conveyed in the findings of the Global Nutrition Report (2016) which ranks India in 114th position out of 132 countries on under five stunting; 120th out of 130 countries in the category of under five wasting and 170th out of 185 countries on the prevalence of anemia - the debilities which are mainly linked to malnutrition.

Protein Energy Malnutrition (PEM) is regarded as most serious among different forms of malnutrition with long-lasting impacts (Grover and Ee 2009) (Figure 2). PEM is defined as the disparity between the protein and energy supply and the body's demand for them to ensure optimal growth and function (De Onis and Blossner 1997). PEM can be the primary cause for several health complications such as hypoalbuminemia, systemic capillary leak syndrome (SCLS), secondary immune deficiency, etc. (Waterlow 1997; Lehman and Ballow 2014). The mortality rate of children suffering from PEM can range from 5% to 40% depending on the underlying cause and economic conditions (Chao et al. 2018). South-East Asian countries have the highest prevalence of protein-energy malnutrition (Total: 25%; Children: 15%) (De Onis et al.

1993) when compared to other countries, perhaps due to cultural reasons (Chaparro et al. 2014).

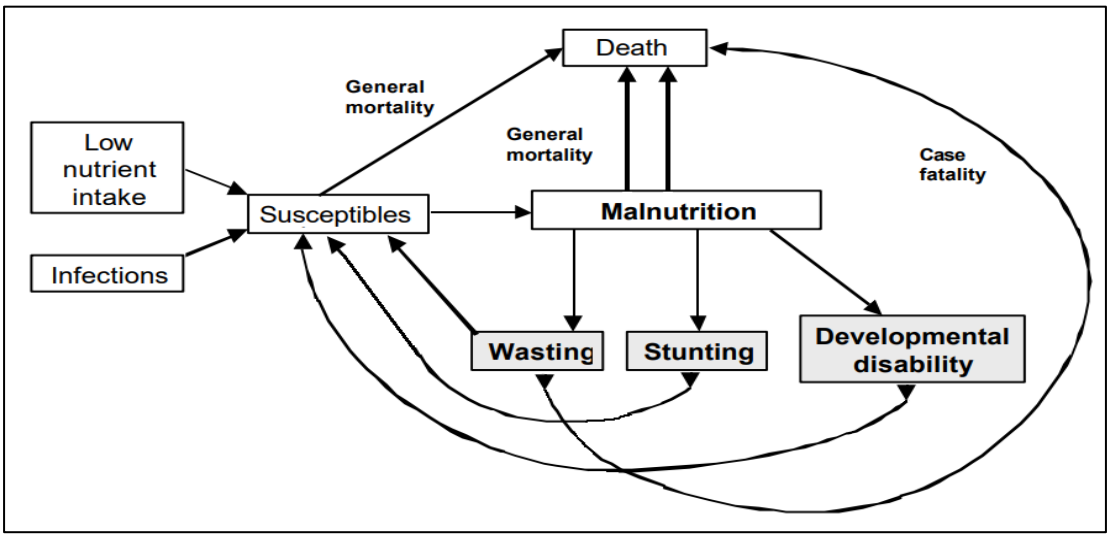


Figure 2: Protein-energy malnutrition disease model (Source: Stein 2000)

Thus, undernourishment with its two constituents - PEM and micronutrient deficiency emerge as the major health burdens around the world, particularly in the developing countries, on both societies and individuals. Although several international organizations are working on mitigating these problems, the progress achieved so far is far from satisfactory. Among the various factors contributing towards it, the lack of availability of low-cost protein diet stands as the predominant one. Thus, there is a need to develop inexpensive alternative sources of protein diet for the poor and marginal population of the world for addressing the burgeoning issue of PEM. Exploiting protein-rich underutilized crop species can be one of the key solutions for this.

1.2. Underutilized crops

There are about 3,74,000 plant species known so far (Christenhusz and Byng 2016) of which 30,000 are reported to be edible and safe for human consumption (Shelef et al. 2017). Only 7000 of them have been either cultivated or collected as food so far

(Khoshbakht and Hammer 2008). Almost 50% of the human calorie intake comprises only three plants - wheat, maize and rice (Keyzer et al. 2005; Ebert 2014) which together occupies 40% or 555 million ha of all arable land globally (Tilman et al. 2011; Stamp et al. 2012). Moreover, 90% of the world's food requirement is met by only 50 crop species (Collins and Hawtin 1999; FAO 2018).

Consequently, thousands of edible plant species remain relatively "neglected" and/or "underutilized", even though they have the potential to contribute for food security, nutrition, and income generation (Dansi et al. 2012; Chivenge et al. 2015). A good number of them, however, are not adopted to large-scale cultivation due to lack of research and development efforts. Dansi et al. (2012) observes that, the majority of the research grants in developing countries are supporting works on major crops alone. Adding to this, a perception among the farmers that these crops offer fewer returns on investment have greatly influenced their choice of major crop for cultivation (Nelson et al. 2004).

A glimpse of some nutritionally endowed underutilized plant species is provided in Table-1. With appropriate R&D inputs, popularization and market linkages, these and other underutilized crops (UUCs) can provide an attractive alternative for addressing the malnutrition issues and improve food security (Padulosi et al. 1999; Ghane et al. 2010). In addition, UUCs also adds to a variety in the human diet and serves as a cheap source of supplementary protein for large human population in developing countries where availability of animal meat is limited and expensive (Kumar et al. 2016; Considine et al. 2017). Among different UUCs, legume species emerge as an attractive candidate for addressing PEM.

Table 1: List of promising underutilized plant species with their nutritional value

Name	Protein* (Nx6.25)	Crude fat*	Dietary fibre*	Carbohydrates*	Reference
Lupinus luteus L. (Yellow lupin)	33-42	4-12	25-40	10.9	A, C
Vicia faba. L. (Broad bean)	26-34	2-4	15-24	18.93	A, D
Pennisetum glaucum L. (Pearl millet)	11.6	5	11.3	67.5	B
Eleusine coracana L. (Finger millet)	7.7	1.5	11.5	72.6	B
Fagopyrum esculentum Moench. (Buck wheat)	12.3	2.3	10.1	66.0	B
Ziziphus mauritiana Lam. (Indian jujube)	7.9-8.7	0.8-1.5	4.9-7.3	79.5-83.2	E
Ullucus tuberosus Loz. (Olluco)	13.2	0.7	4.9	75.2	F
Annona squamosal L. (sugar-apples)	1.96	0.176	3.21	18.2	G
Panicum miliaceum L. (Proso millet)	12.5	3.1	14.2	70.4	H

*g/100g; (A) Mlyneková et al. 2014 (B) Kumar et al. 2016 (C) Gdala and Buraczewska 1996 (D) Mejri et al. 2018 (E) Nyanga et al. 2013 (F) Sperling and King 1990 (G) Albuquerque et al. 2016 (H) Habiyaremye et al. 2017.

1.3. Underutilized legumes for PEM

Grain legumes by far are the primary source of protein in a vegetarian diet and therefore consumed by a large section of the human population (Luse and Okwuraiwe 1975). They have long been recognized as a cost-effective substitute for animal protein (Famurewa and Raji 2005) and a metaphor "*Poor man's meat*" (Messina 1999) aptly describes their potentials. Nonetheless, many grain legumes are not affordable to a large section of the population owing to their inhibitive costs and limited production potential (Graham and Vance 2003; Foyer et al. 2016). The production data suggests that total global grain legume production in 2013 was only 121 Mt, except soybean (278 Mt), leaving much to be desired in terms of their accessibility and affordability to resource-poor population of the world.

On the other hand, wild and underutilized legume species possess an adequate amount of protein, polyunsaturated fatty acids (PUFAs), essential amino acids, dietary fibre, vitamins and essential minerals akin to common legumes. Besides, underutilized legumes are more climate resilient and thrive well in hard climatic conditions (Ebert et al. 2014). Having adapted to marginal conditions, they are the storehouse of resistant genes to biotic and abiotic stresses (Stoddard et al. 2006; Bhat et al. 2008; Bhat and Karim 2009; Ebert et al. 2014).

Table 2: List of prominent tropical underutilized legume species and their nutritional value

Legume	**Protein**	**Fat**	**Carbohydrates**	**Fiber**	**Reference**
Adzuki bean [a]	25.2	*	57.1	13.5	Chau et al. 1998
M. pruriens [b]	28.0	7.8	33.0	8.0	Janardhanan & Vadivel 2000
Sword bean [a]	35.0	*	69.0	13.5	Bhat and Karim 2009
Scarlet bean [b]	20.0	7.9	53.1	10.5	Aremu et al. 2010
Lima bean [a]	23.2	0.2	71.1	18.4	Ezeagu and Ibegbu 2010
Rice bean [b]	21.0	1.9	67.0	6.4	Sharma 2014
Lablab [a]	27.0	1.7	60.0	9.0	Gulzar and Minnaar 2017
Winged bean [b]	34.0	17.0	22.0	12.1	Lepcha et al. 2017

*Not detected; [a] (g per 100 g); [b] (%)

The tropical regions, which covers almost 40% of the world's agricultural area and being a belt of poverty, need an extra focus in ensuring nutritional security to a growing world population. Some of the underutilized legumes of this region have shown great potential in terms of nutritional content and key agronomic traits such as drought tolerance, resistance to pests and diseases, and adaptation to semi-arid and arid environments (Mayes et al. 2011; Chivenge et al. 2015; Chimonyo et al. 2016; Chibarabada et al. 2017; Hadebe et al. 2017; Mabhaudhi et al. 2017a; 2017b). Some such promising lesser-known tropical legume species and their nutritional attributes are listed in Table 2. In otherwise less-fertile tropical soil, their ability to fix

atmospheric nitrogen can boost soil fertility thus reducing the need and dependency on expensive nitrogenous fertilizers (Gebrelibanos et al. 2013). Additionally, a few of them are a source of novel bioactive compounds that can be tapped for the production of therapeutic drugs and functional foods (Bhat and Karim 2009; Maphosa and Jideani 2017).

Thus, the identification of promising unexploited legume resources for tropical agriculture can pay rich dividends. The success from such an effort, integrated with a sustained crop improvement program involving modern plant breeding know-how, can go a long way in finding a long-lasting solution to PEM and providing nutritional security for this region. In this context, we have reviewed below the prospects of four selected underutilized legume species, in addition to velvet bean, covered in this thesis, which are suited to varied requiremettns of tropical agriculture and to place them in perspective.

1.3.1. Lablab bean

Lablab purpureus (L.) Sweet - popularly known as lablab bean is one of the traditional vegetables originated in India (Deka and Sarkar 1990; Fuller 2003). Despite being labeled as 'underutilized' (Hammer et al. 2001), lablab bean is now cultivated in substantial areas in the tropical regions either as a single crop or in mixed production systems. It is mainly grown as an intercrop with maize (Pengelly and Maass 2001; Maass et al. 2005). The schematic representation of different uses of lablab bean is presented in Figure 3. According to Rai et al. (2006), Indo-China region is the center of diversity of this crop. It is endowed with great variability in terms of growth habit, morphological characters, maturity including shape, size, the colour of fruit and seed in this region.

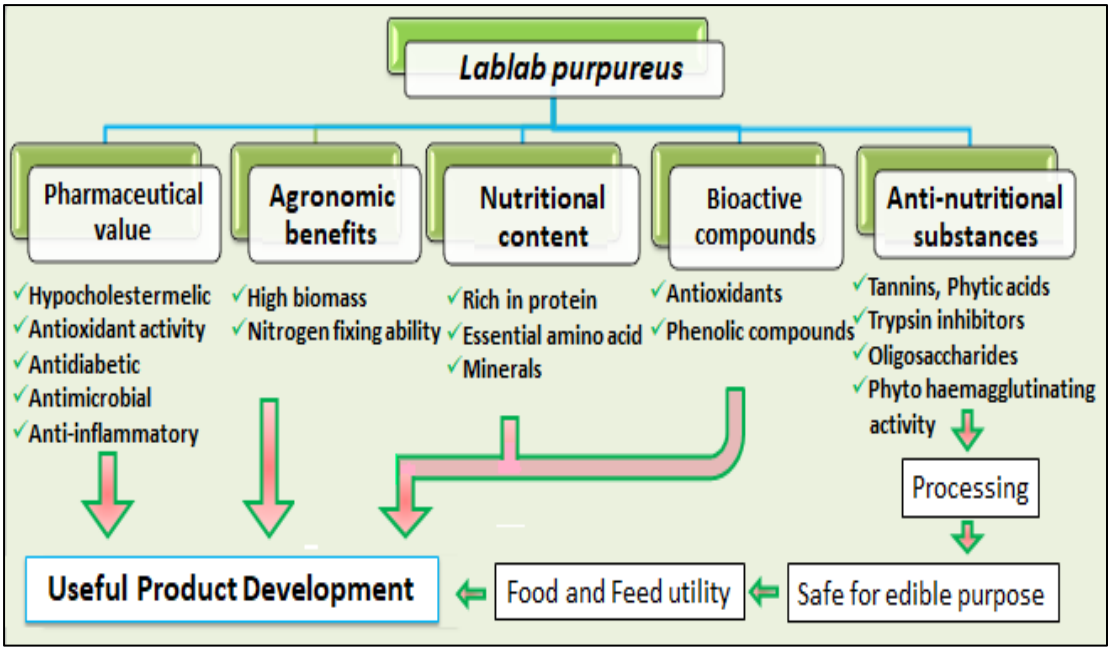


Figure 3: Overview on uses of Lablab bean (Bhat and Karim, 2009; Naeem et al. 2016; Lepcha et al. 2019)

From the tropical agriculture viewpoint, this species can be one of the excellent candidates for the fight against malnutrition due to a high content of protein, carbohydrate (Chau et al. 1998), minerals (Duke 2008) and energy (Deka and Sarkar 1990). Duke (2008) identified many amino acids, micro, and macronutrients as well as vitamins in lablab which are helpful for human health. Lablab pods and seeds have other uses including vegetables, forage crop and possess promising nutraceutical and pharmaceutical properties (Morris 2003). Adaptation to dry conditions and resistance to major diseases and pests (Amole et al. 2013), makes it an even more attractive candidate for tropical agriculture. However, little studies have been carried out on genetic improvement and breeding except for few reports on genetic diversity estimation using molecular markers (Maass 2005; Wang et al. 2007; Rai et al. 2016). From the genomics viewpoint, lablab bean is far removed from the genomics revolution with significant gaps in genomic resource development (Lepcha et al. 2019).

1.3.2. Rice bean

Rice bean (*Vigna umbellata* (Thunb.) Ohwi and Ohashi) commonly called climbing bean, mountain bean, mambi bean, and oriental bean is another promising tropical legume lesser known, less researched and barely exploited. It belongs to a group of minor food and fodder crops and is habitually grown as a mixed crop or intercrop with maize (*Zea mays*), sorghum (*Sorghum bicolor*), and cowpea (*V. unguiculata*) in addition to an individual crop in upland areas (Khadka and Acharya 2009). According to available reports, Asia is the center of diversity of this species, where it is distributed from the Himalayan highlands to South, Southeast and East Asia (Siriwardhane et al. 1991; Tomooka et al. 2002; Tun and Yamaguchi 2007). It is mainly used as a dried pulse, besides vegetable, green manure, and fodder crop (Arora et al. 1980; Isemura et al. 2010).

Rice bean plays an important role in human nutrition, animal feed, and soil improvement (Mal and Joshi 1991), especially in upland agriculture. Nutritionally, it has high content of protein, essential amino acids, minerals and essential fatty acids (Mohan and Janardhanan 1994). The young pods and leaves are used as a vegetable (Kongjaimun et al. 2012a). The overall potential of rice bean is presented in Figure 4. It is also resistant to major pests, diseases and does not need fertilizer or extra care during the growth (Dahipahle et al. 2017). Rice bean shows tolerance to high temperature with good yields, requiring 18-30° C and 1000-1500 mm annual rainfall. It is tolerant to extreme cold/frost conditions (Khanal et al. 2007) which makes it an ideal candidate for cold regions.

Research efforts so far in this crop have provided basic insights into genetic diversity and relationship among the germplasm collection and, identification of a few high

yielding (Rana et al. 2014) as well as insect resistant varieties (Edwards 2006). Genomic research is still in the nascent stage. Recently *de novo* assembly from RNA-Seq data has been developed (Wang et al. 2016; Chen et al. 2016) paving a way for marker development and other breeding-oriented applications.

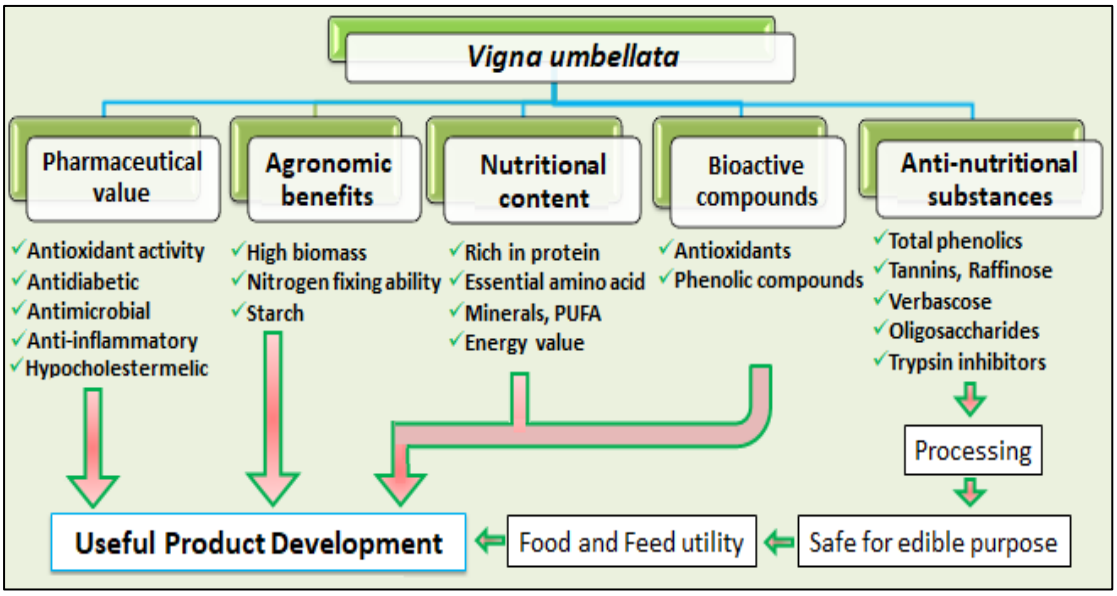


Figure 4: Overview on uses of Rice bean (Bhat and Karim 2009; Sharma 2014; Dahipahle et al. 2017)

1.3.3. Winged bean

Psophocarpus tetragonolobus (L.) DC., popularly known as winged bean or Asparagus pea or Goa bean is a tropical legume classified within the phaseoloid clade of Leguminosae. A predominantly self-pollinated crop, winged bean displays twining habit, tuberous roots, longitudinally winged pods and both annual and perennial growth form (Hymowitz and Boyd 1977). It has a diploid genome of 9 pairs of chromosomes (2n=2x=18, Harder 1992) and an estimated genome size of 1.22 Gbp (Vatanparast et al. 2016). The plant grows in abundance in hot, humid, equatorial countries such as Malaysia, Indonesia, Thailand, India, The Philippines, Bangladesh, Myanmar and Sri Lanka (Harder and Smartt 1992). The exceptional nutritional quality of this plant, and the fact that these are available at all stages of life cycle

makes winged bean a promising crop for widespread use in protein deficient areas of the World (Harder and Smartt 1992). If the yields of both seed and tuber are combined, the winged bean can outdo many crop legumes that are conventionally grown in the tropics. Consequently, it is projected as a promising alternative for soybean in these regions.

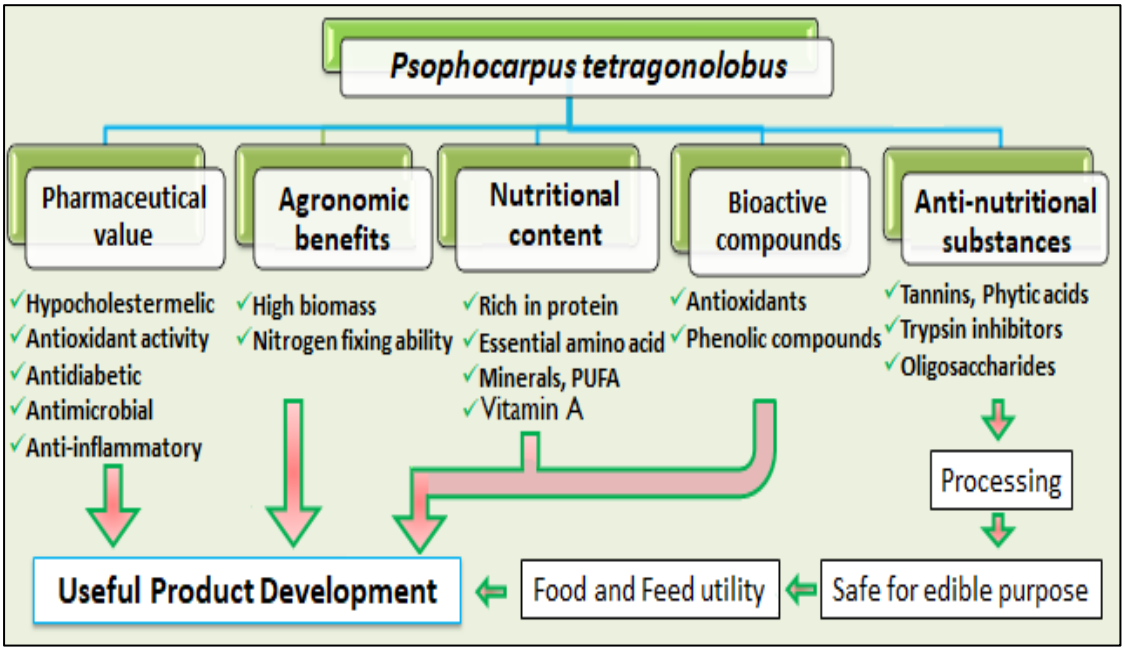


Figure 5: Overview on uses of Winged bean (Bhat and Karim 2009; Lepcha et al. 2017, 2019; Tanzi et al. 2019)

Tropical soils are rich in rhizobia and winged bean, native to the cowpea cross-inoculation group, makes perfect mutualisms with it (National Academy of Sciences 1975; Elmes 1976). In developing countries like India, the single most cause for childhood blindness is vitamin A deficiency (Rahi et al. 1995). Nearly 254 million children are in the risk of blindness due to the deficiency of vitamin A (West 2002). According to National Academy of Sciences (1975), the amount of vitamin A present in winged bean leaves is up to 20,000 IU per 100 g of edible portion, which is highest ever recorded in green leaves of tropical plants. This implies that winged bean is immensely helpful to prevent vitamin A deficiency and its consequences. In

summary, the nutritional eminence along with tropical adaptation makes winged bean a gifted candidate for cultivation in these less endowed parts of the world (Kadam et al. 1984; Mohanty et al. 2015). Figure 5 describes the potential attributes of the winged bean.

1.3.4. Scarlet bean

Phaseolus coccineus L. (syn. *Phaseolus multiflorus* Wild; 2n=2x=22) - well known as scarlet runner bean or runner bean is the third-most important cultivated *Phaseolus* species worldwide after *P. vulgaris* (common bean) and *P. lunatus* (butter bean) (Santalla et al. 2004; Spataro et al. 2011). However, as compared to common bean, it is less researched and cultivated. It is a perennial twinning vine which, in transitional temperate climatic conditions, shows an annual growth habit (Labuda 2010). Runner bean originated in the mountains of Central America (Spataro et al. 2011) and is mainly distributed in temperate and tropical areas (Delgado Salinas 1988). Because of sustained nature towards cooler spells, scarlet bean becomes an important crop in the Northern part of Europe and in the mountainous areas of southern Europe (Spataro et al. 2011; Bitocchi et al. 2017). Compared to other *Phaseolus* species, this species is more amended to cold temperatures (cold tolerance) and cool summers (Delgado Salinas 1988; Rodino et al. 2007) and can be promising crop for colder Himalayan regions.

The beneficial attributes of runner bean are presented in Figure 6. Recently, there has been a renewed interest in runner bean cultivation, particularly after it has been shown that the seeds of *P. coccineus* (runner bean) are more digestible than those of *P. vulgaris* (common bean) (Carlderon et al. 1992; Acampora et al. 2007).

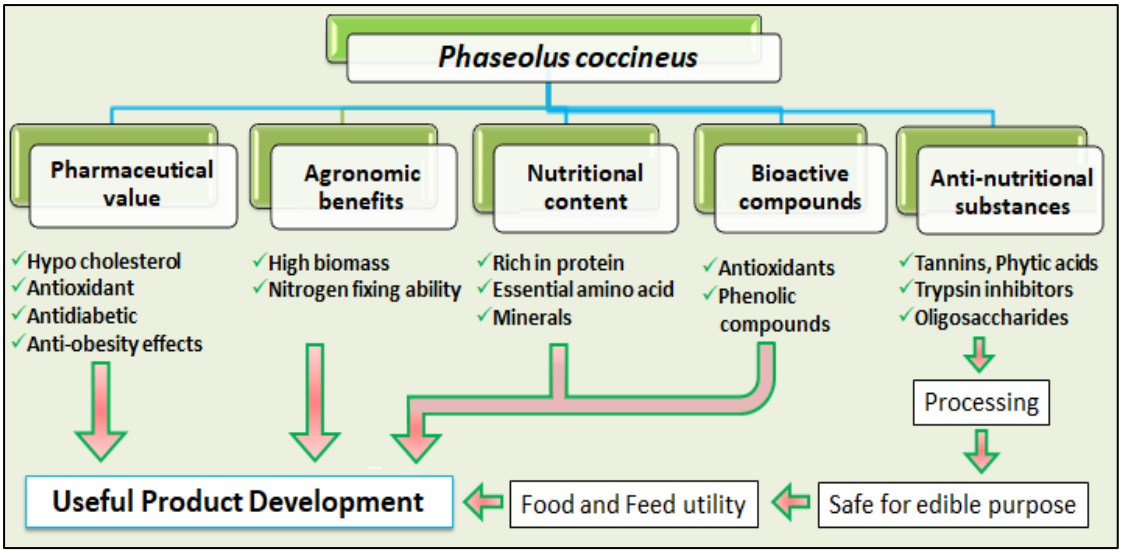


Figure 6: Overview on uses of Common bean (Acampora et al. 2007; Bhat and Karim 2009; Schwember et al. 2017)

From a genomic viewpoint, runner bean has been little studied, notwithstanding the development of molecular markers and genome sequencing of other legumes in the last decade. The high genetic diversity of runner bean is vital for breeding purposes, especially as a source of cold tolerance and disease resistance genes. But the lack of germplasm characterization limits its utilization as donor species in case of inter-specific hybridizations, and therefore restricts its use in other *Phaseolus* breeding programs. Developing broadly adapted, self-determinate cultivars that enable mechanical harvesting, evaluating heterosis for hybrid production, and germplasm characterization are great challenges and opportunities in the future that would increase its cultivation on a broader scale worldwide.

1.4. Genomic resources in underutilized legume crops

Molecular breeding approaches need genomic resources such as species-specific molecular markers, genetic/linkage maps, and DNA/RNA sequence information, which were unavailable even in case of major legume species barring soybean, until recent past. Of late, speedy developments have been witnessed in legume genomics

due to developments in next-generation sequencing (NGS) and high-throughput genotyping methods. Consequently, genome sequences and draft assembly for some important legume crops have been reported during the last 5 years. This includes, *Cajanus cajan* (Pigeon pea - 833 Mb genome; Varshney et al. 2012), *Cicer arietinum* (Chickpea - 738 Mb; Varshney et al. 2013), *Lotus japonicus* (Lotus - 472 Mb; Sato et al. 2008), *Medicago truncatula* (Barrel clover - 373 Mb; Young et al. 2011) and *Phaseolus vulgaris* (Common bean - 588 Mb; McClean et al. 2004) etc. More recently, efforts are on for sequencing of few other legume species such as: *Trifolium pratense* (Red clover - 440 Mb; Sato et al. 2005), *Arachis hypogaea* (Groundnut - 2800 Mb; Duan et al. 2012), *Lupinus angustifolius* (Lupin - 924 Mb; Yang et al. 2013), and *Pisum sativum* (Garden pea - 4450 Mb; Yang et al. 2015).

Nevertheless, underutilized species continue to suffer from a dearth of molecular-genetic and genomic resources and thus remain 'orphans' from the genome revolution. However, decreasing sequencing and genotyping costs are fuelling new hopes for these crops. As a consequence, RNA-Seq data has been increasingly applied for genetic studies even in some underutilized legume plants such as faba bean (*Vicia faba* - Kaur et al. 2012), adzuki bean (*Vigna angularis* - Chen et al. 2015a), rice bean (*Vigna umbellata* - Chen et al. 2016), hyacinth bean (*Lablab purpureus*), grass pea (*Lathyrus sativus*) bambara groundnut (*Vigna subterranea* - Chapman et al. 2015), winged bean (*Psophocarpus tetragonolobus* - Chapman et al. 2015; Vatanparast et al. 2016) and field pea (*Pisum sativum* - Kerr et al. 2017).

Towards marker development, Chen et al. (2016), sequenced the whole transcriptome of rice bean (*V. umbellata)* and developed first species-specific SSR markers for this crop. In winged bean, sufficient progress has been made towards the development of

genic markers based on RNA-Seq data (Wong et al. 2014; Chapman et al. 2015; Vatanparast et al. 2016). Hane et al. (2017) developed a comprehensive assembly, draft genome sequence for the narrow-leaf lupin (NLL) (*Lupinus angustifolius* L.) (genome size 951Mb) - the first representative of the Genistoid clade of Papilionoid legumes (Hane et al. 2017). *De novo* transcriptomic assembly and analysis of cowpea (*V. unguiculata* L. Walp.) (620Mb) transcriptome has also been done by Chen et al. (2017). In adzuki bean (*Vigna angularis*), Kang et al. (2015) identified 26,857 high confidence protein-coding genes by transcriptome sequencing. This group identified a total of 7947 EST-SSRs from which 500 SSR markers have been validated. The summary of microsatellite marker development in some different UULS is given in Table 3.

Table 3: Summary of SSR marker development efforts in different UULS

S. No	UULS	No of SSRs	Accessions (used/for)	Study/ Developed/ Generated	Reference
1	*Vicia faba* (Faba bean)	11	29	Generation and characterization	Gong et al. 2010
		30	40	Genetic diversity	Tekalign et al. 2018
2	*Vigna angularis* (Adzuki bean)	326	Cross between cultivated and wild rice bean	SSR-based linkage map	Isemura et al. 2010
		500	38	112 million high quality reads; Genetic diversity	Chen et al. 2015a*
3	*V. subterranea* (Bambara groundnut)	22	240	Genetic diversity	Somta et al. 2011
		68	24		Molosiwa et al. 2015
		20	45	Heterozygosity	Ho et al. 2016
4	*V. unguiculata* (Cowpea)	226	Cross between ((JP81610 × TVnu457) × JP81610)	SSR-based genetic linkage map	Kongjaimun et al. 2012b

5	*V. radiate* (Mung bean)	430	Cross between cultivated and wild Mung bean	SSR-based genetic linkage map	Isemura et al. 2012
		66	31	103 million high quality reads	Chen et al. 2015b*
6	*Lablab purpureus* (Hyacinth bean)	22	24	Genetic diversity	Zhang et al. 2013
		134	143	Transferability and genetic diversity	Rai et al. 2016
		55	16		Keerthi et al. 2018
7	*V. umbellate* (Rice bean)	13	472	Genetic diversity	Tian et al. 2013
		24	230	Genetic diversity	Wang et al. 2015
		300	32	Genetic diversity	Chen et al. 2016
		220	22	Genetic diversity	Wang et al. 2016
		53	65	Genetic diversity	Iangrai et al. 2017
8	*Pisum sativum* (Garden pea)	1139	*De novo* transcriptome assembly	15,779,854 reads	Chapman 2015*
	L. purpureus	2567		16,190,774 reads	
	V. subterranea	1884		7,887,745 reads	
	Psophocarpus tetragonolobus	1305		16,625,155 reads	
9	*Psophocarpus tetragonolobus* (Winged bean)	12,956	*De novo* transcriptome assembly	804,757 reads	Vatanparast et al. 2016*
		18	6	Genetic diversity	Wong et al. 2017*
10	*V. unguiculata*	500	32, *De novo* transcriptome assembly	57,214,890 paired reads; Genetic diversity	Chen et al. 2017*
		1336	Cross between salt tolerant and salt sensitive accession	SSR-based linkage map	Marubodee et al. 2015
		1071	50	Transferability	Gupta and Gopalakrishna 2010
		49	22	Genetic relationship	Badiane et al. 2012

11	*V. aconitifolia* (Moth bean)	172	Wild moth bean (TN67) and cultivated moth bean (ICPMO056)	SSR-based linkage map	Yundaeng et al. 2018
12.	*V. subterranea* *L. purpureus*	NA	*De novo* transcriptome assembly	Draft genome information	Chang et al. 2019*

*Whole transcriptome sequencing.

The focus of these activities, along with the development of genome-scale data sets that can be used in high-throughput approaches has been to facilitate genomics-assisted breeding in these legumes. However, no such efforts have been made earlier in *M. pruriens*. In view of this, the present thesis work focused on the development and characterization of microsatellite markers in *M. pruriens* using RNA-Seq data.

1.5. Microsatellites – an overview

Microsatellites are considered superior markers for the genetic studies owing to their reproducibility, hyper-variability, multi-allelic nature, co-dominant inheritance, relative abundance and wide genome coverage (Parida et al. 2009; Kalia et al. 2011). Additionally, chromosome-specific location, ease for automation, and high throughput genotyping (Parida et al. 2009; Bhat and Karim 2009) makes them a choice marker in genetic studies including genotypic profiling, marker-assisted selection, diversity, phylogenetic analysis, etc. (Varshney et al. 2007; Mason 2015; Paliwal et al. 2016). After the first paper by Litt and Luty in 1989, as on February 2019, PubMed database search with the quoted keyword “microsatellite” shows almost ~59,000 research publications which have used or discussed microsatellites (Figure 7) mirroring the popularity of these markers for genetic studies.

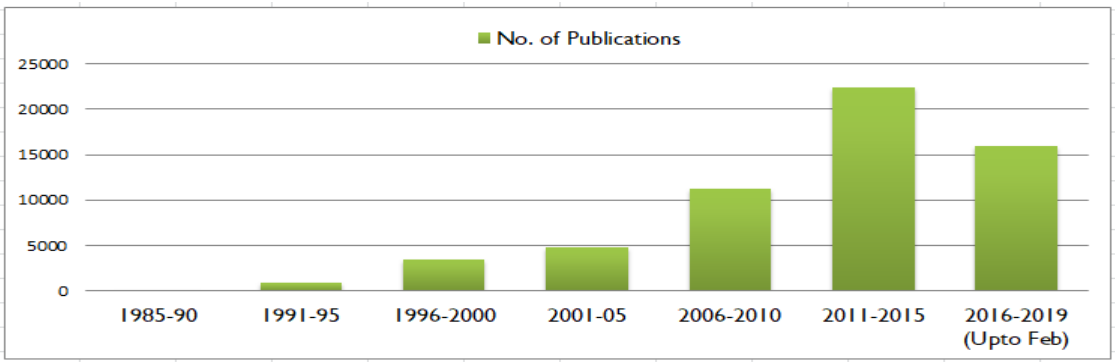


Figure 7: Number of publications retrieved with key word - "Microsatellite" from the PubMed database 2019 (https://www.ncbi.nlm.nih.gov/pmc)

1.5.1. History of microsatellites

Miesfeld et al. (1981) first identified microsatellites by sequence analysis of the β-globin gene locus. Later, Hamada et al. (1982) showed its presence in various eukaryotes ranging from yeast to vertebrates. Delseny et al. (1983) and Tautz and Renz (1984) confirmed its presence in plants. These tandemly repeated short nucleotide motifs that contained 1-6 bases have been termed by different names such as microsatellites (Litt and Luty 1989), simple sequence length polymorphism (SSLP; Tautz 1989), simple sequence repeats (SSR; Jacob et al. 1991) and short tandem repeats (STR; Edwards et al. 199l).

1.5.2 Origin and evolution of microsatellites

Microsatellites are widespread both in prokaryotic and eukaryotic genomes (Field and Wills 1996). Though predominantly located in noncoding regions of the DNA whose origin and functions are unclear (Hancock 1995; Tóth et al. 2000; Oliveira et al. 2006), they are often found even in coding regions (Phumichai et al. 2015; Taheri et al. 2018). The SSR motifs reveal high mutation rates than the rest of the genome ranging from 10^{-7} to 10^{-3} nucleotides per locus per generation (Buschiazzo and Gemmell 2006). The SSR evolution resulting in an increase or decrease in repeat

number is associated with its mutation rate. Several research groups described different mechanisms to explain high mutation rate of microsatellites which include (a) single-strand DNA slippage (b) double-strand DNA recombination (c) mismatch/double-strand break repair, and (d) retrotransposition. A detailed review of this topic is presented by Kalia et al. (2011).

1.5.3 Classification of microsatellites

Depending on the repeat arrangement, size, type, and location in the genome, microsatellites have been classified into various types. Firstly, Weber (1990) described it as perfect, imperfect and compound/composite microsatellites, although Wang et al. (2009) mentioned the terms simple perfect, simple imperfect, compound perfect and compound imperfect. Uninterrupted series of a repeat nucleotide unit is called as perfect microsatellites, ex. $(AT)_{15}$, and the presence of one to three base disruption in the tandem run of repeats are termed as imperfect microsatellite, ex. $(AT)_7C(AT)_{10}$. Lastly, two adjacent distinctive sequence repeats are called compound or composite microsatellites, ex. $(AT)_7(GC)_8$ (Kalia et al. 2011). Compared to imperfect SSRs, longer perfect SSRs contain higher genetic variation in remarkable level and exhibits greater allelic variability (Buschiazzo and Gemmell. 2006; Kelkar et al. 2008; Merritt et al. 2015; Bastías et al. 2016).

Depending on the number of nucleotides per repeat unit, SSR's have been classified as mono-, di-, tri-, tetra-, penta- or hexanucleotides (Kalia et al. 2011). Based on the length of the repeat motif, SSR loci are categorized into two groups. Class I, or hypervariable markers, consisting of SSRs ≥20 bp, and Class II, or potentially variable markers, consisting of SSRs ≥12 bp <20 bp. Microsatellites are also classified based on their location. Most of the genomic SSRs are of nuclear origin

(nuSSRs); however, microsatellites are also distributed on mitochondria (mtSSRs) and chloroplasts (cpSSRs).

1.5.4. SSR development methods

Development of SSR markers follows one of the following three important methods;

(a) From SSR-enriched genomic DNA libraries

In brief, this method uses SSR isolation by constructing genomic libraries either by selective hybridization or primer extension method. It uses repeat-specific probes, which allows selection of repeat motifs based on specific hybridization signals (Senan et al. 2014). The hybridized DNA fragments are either directly amplified and sequenced or cloned into a vector before sequencing and searched for the presence of SSR motifs (Glenn and Schable 2005; Geng et al. 2010). The success of this method hinges on specific bonding between labelled probes and DNA fragments containing SSRs (Senan et al. 2014). Primer extension method, on the other hand, uses specific primers of SSR which allows selective amplification of microsatellite containing genomic DNA (Robic et al. 1994; Paetkau 1999). The process depends on the genomic library construction using a suitable vector to recover the library as a single-stranded DNA (ssDNA) (Ostrander et al. 1992; Paetkau 1999).

(b) From non-enriched genomic DNA libraries

This method uses the screening of non-enriched genomic DNA library for SSR motifs using hybridization, followed by sequencing (Senan et al. 2014). The idea is, cloning DNA fragments before enrichment reduces the redundancy, and provides a wide range of SSR motifs as compared to enriched ones (Golein et al. 2006; Blair et al. 2009a).

(c) Using next generation sequencing methods

Recent advancements in next generation sequencing (NGS) technologies is providing yet another simple and inexpensive method for the microsatellite development (Shendure and Ji 2008; Stapley et al. 2010; Ekblom and Galindo 2011; Duan et al. 2017).

The approach involves sequencing of cDNA libraries using any of the popular platforms, followed by generation and screening of transcript assembly for SSRs (Severin et al. 2010; Zenoni et al. 2010; Taheri et al. 2018). Even in the absence of whole genome data, which is the case in many crop plants, *de novo* assembly has been proved to be good enough and usable for SSR mining (Yates et al. 2014; Garg et al. 2011a, 2011b) with large amounts of expressed sequence data (Li et al. 2014). SSR development based on RNA-Seq data has been attempted in a large number of crop species with considerable success (Li et al. 2018)**.** The schematic view of SSR development using transcriptome data is given in Figure 8. We have used similar approach for the genic microsatellite marker development in the present thesis.

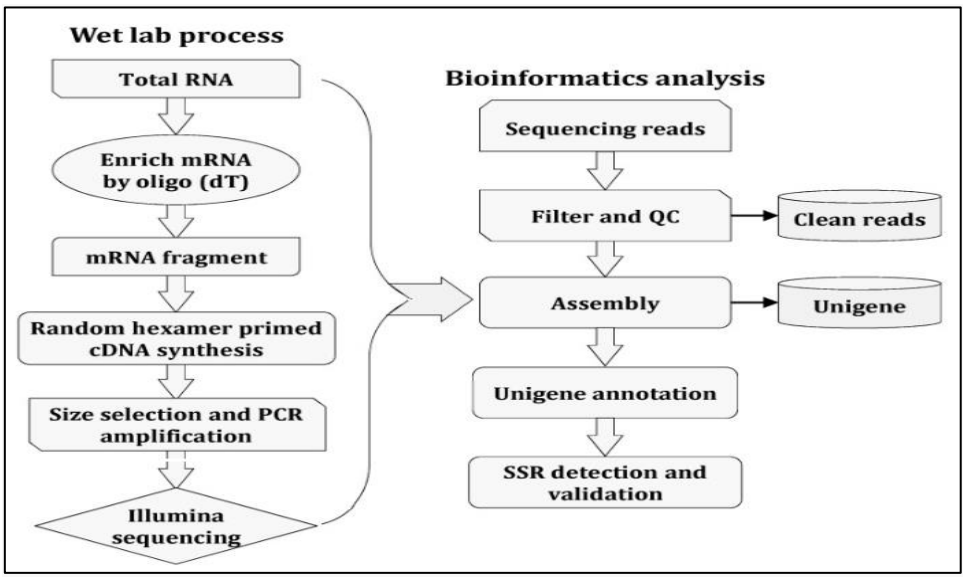


Figure 8: Schematic representation of microsatellite development using RNA-Seq data

1.5.5. Why genic SSRs?

It is widely recognized that the development of genomic SSRs is time-consuming and labour intensive (Chen et al. 2015a). Moreover, they represent non-coding part of the genome and provide little information on the variation at the genic level (Rakoczy-Trojanowska and Bolibok et al. 2004). On the contrary, the unigene-based microsatellite markers, having derived from the transcribed regions of genome (Korir et al. 2013), reflect true functional diversity and thus provide reliable tool for the characterization of gene functions and their biological identity (Ronning et al. 2003). A comparison between genic and genomic SSRs is given in Table 4.

Table 4: Advantages of genic-SSRs over genomic SSRs

	Genomic SSRs	**Genic SSRs**
Time consuming	High	Comparatively less
Development cost	High	Low
Produce data	High	Comparatively less
Function	Not known	Putative function is known
Transferability	Less	More
Cross species analysis	Less	High
Redundancy	Less	High
Robustness	Less	High
Gene function characterization	Less	High
Gene tagging	Difficult	Possible
QTL analysis	Not very accurate	Accurate

(Data source: Ronning et al. 2003; Varshney et al. 2005; Szalma et al. 2005; Shin et al. 2006; Crossa et al. 2007; Park et al. 2009)

1.6. *Mucuna pruriens* (L.) DC.

Mucuna pruriens (L.) DC., also known as ‘cow itch’ or ‘itching bean’ is a promising tropical species classified within the Phaseoloid clade of Leguminosae which also include soybean, common bean, mung bean, and relatives (Pugalenthi et al. 2005; Gepts et al. 2005). The plant contains vital nutrients (Bressani 2002) and is the key

source of 3, 4 - dihydroxy-L-phenylalanine (L-DOPA) - a dopamine precursor, used in the treatment of Parkinson's disease (Katzenschlager et al. 2004). It is also a popular green manure cover crop and offers benefits for soil nutrient improvement (Lampariello et al. 2012).

However, little efforts have been made for the improvement of this plant except in the area of genetic diversity estimation (Sathyanarayana et al. 2011; Sathyanarayana et al. 2016; Patil et al. 2016). Undeniably, *M. pruriens* is one of those legume species that critically lack genomic resources. Hence, we have attempted to generate and validate genic microsatellite from the transcriptome data as part of this thesis work.

A detailed review of the current status and future prospects of *M. pruriens* research and the synthesis of outcome are presented in the review of literature section.

2. Review of Literature

2.1. *Mucuna pruriens* (L.) DC. - Origin, distribution, and cultivation

Mucuna pruriens (L.) is a self-pollinated tropical legume classified within the Phaseoloid clade of Leguminosae (Duke 1981). It is native to China and Eastern India (Burkill 1966) where the plant was once widely cultivated as green manure cover crop (Duke 1981). According to Bort (1909) and Duke (1981), it was initially introduced into the USA in 1876, and later to Mexico and Central America, during 1920's by banana companies (Ceballos et al. 2012). Today it is found in Asia, America and African countries with wide-ranging distribution in both tropical and subtropical parts of the world (Fung et al. 2011; Kumar and Saha 2013). In Asia, the species mostly occurs in Southeast Asian countries like India, Bangladesh, Sri Lanka and Malaysia (Fung et al. 2011; Kumar and Saha 2013).

Among the three varieties of *M. pruriens*, var. *utilis*, commonly called velvet bean, commands considerable economic significance. This plant played an important role in the US agriculture during the 19th century (Buckles et al. 1998), and was extensively used as a rotation crop with corn, cotton, and sugarcane to increase the yield and also for soil fertility in *Citrus* orchards in Florida (Duggar 1899; Bailey 1907; Bort 1909). However, its cultivation decreased in the following decades (1960s and 1970s) owing to development of high-yielding varieties and adoption of chemical fertilizers and farm mechanisation as part of the green revolution (Conway 1997; Urquiaga et al. 1999; Okito et al. 2004) - which saw the decline of usage and the research on cover crops, crop rotation and other traditional soil management practices (Ceballos et al. 2012).

2.2. Taxonomy

Genus *Mucuna* Adans. belongs to the Fabaceae family and includes 100 species of annual and perennial legumes of pantropical distribution (Buckles 1995). In India, it is represented by ten species (Wilmot-Dear 1987; Aitawade et al. 2012) - of which *M. atropurpurea* (Roxb.) DC. ex Wight and Arn is endemic to Peninsular India; *M. imbricata* DC. ex Bak., *M. bracteata* DC. ex Kurz, *M. macrocarpa* Wall., *M. sempervirens* Hemsl and *M. nigricans* (Lour) Steud. are largely distributed in the Eastern Himalayas and *M. pruriens*, *M. monosperma* DC ex Wight and *M. gigantea* (Wild.) DC. are widely disseminated. Except *M. pruriens*, and perhaps *M. sanjappae* (Aitawade et al. 2012) all other species display perennial growth habits.

Table 5: Botanical classification of *Mucuna pruriens* (L.) DC.

Based on APG (Angiosperm Phylogeny Group)-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

M. pruriens is a key member of this genus and exhibits highly promising agronomic potentials (Sathyanarayana et al. 2016). The botanical placement of the species is provided in Table 5. Within this species, two main varieties, var. *utilis* and var. *pruriens*, are widely reported (Dassanayake and Fosberg 1980; Wilmot-Dear 1987; Sasidharan 2004), while the presence of third affiliate viz., var. *hirsuta* is also acknowledged (Wilmot-Dear 1987; Almeida 1996; Leelambika and Sathyanarayana 2011). The latter was earlier regarded as an independent species (Baker 1876; Nair and Henry 1983; Saldanha 1996), but subsequent revisions, especially the one by Wilmot-Dear (1987), categorically suggested its inclusion, along with few others, under the botanical varieties of *M. pruriens*. However, till date, many literature continue to treat this as an independent species. Such anomalies, though predominant in *M. pruriens*, is not uncommon in other taxa of this genus.

The taxonomy of var. *utilis*, in particular, is confused with numerous synonyms at the species levels (Duke 1981; Buckles 1995). Burkill (1966) found *M. nivea* as being synonymous with *M. lyonia* and *M. cochinchinensis*. Genotypes are described either in terms of their growing regions (e.g., *Mucuna* sp. var. Ghana, *M*. sp. var. IRZ, etc.) or by the popular names they are known in different places like *M. cochinchinensis* in Southeast Asia or *M. deeringiana* in Florida. It is hard to recognize that whether the name of a cultivar is true representative of its genotype. This issue has impeded the effective utilization of velvet bean genetic resources in India and other places, and hence it is necessary to conduct research at the species level as well as to assess the genetic diversity as well as relationships among the accessions before any breeding program could be initiated (Capo-chichi et al. 2003).

2.3. Botanical description

M. pruriens L. (DC.) is a vine legume with a life-cycle ranging from 120 to 330 days (Tracy and Coe 1918; Bailey 1947; Buckles 1995). The plant exhibit trifoliate leaves with oblique lateral folioles, 5-20 cm long, 3-15 cm wide. Flowers are white to purple, 4-6 in number and arranged in hanging racemes of 2-3 cm long with wing and keel 3-4 cm in length. Pods are 5-10 cm long, 1-2 cm wide with 3-6 seeds, and covered densely by velvety itching or non-itching pubescence ranging in colour from silvery grey to orange to green. Seeds are white, bright black or mottled brown in colour; hilum 3-5 mm long with long aril (Buckles 1995). Numerous roots, 7-10 m long, possess abundant nodules near the soil surface. The plants accumulate between 2.2 and 10.9 t/ha of dry biomass and produce between 0.24 and 6.12 t/ha of seed yield (Duggar 1899; Tracy and Coe 1918; Scott 1919; Watson 1922; Whyte et al. 1955; Duke 1981; Göhl 1982; Kay 1985; Purseglove 1987; Pugalenthi et al. 2005; Ceballos et al. 2012). It shows better growth at an altitude below 1600 m with warm and moist conditions (Buckles 1995; Aitawade and Yadav 2012). It is also a source of several important pharmaceuticals (Warrier et al. 1996).

The plant occurs both in wild and cultivated forms. The cultivated variety, represented by var. *utilis*, produce non-itching pods with smooth trichome hairs and is grown in household and home-gardens as a green vegetable across the parts of Central, South and Northeast India (Pugalenthi and Vadivel 2007a). The wild varieties are classified under two varieties viz., var. *pruriens* and var. *hirsuta* and are commonly referred to as “itching bean” or “cow itch” due to highly itching persistent trichomes present abundantly all over their pod surface (Buckles 1995; Pugalenthi and Vadivel 2007b; Leelambika and Sathyanarayana 2011). The seeds are mostly rich in L-DOPA and non-edible.

2.4. Economic importance

2.4.1. Food uses

Mucuna beans have been used as food and feed for centuries in Asian countries like in Sri Lanka (Ravindran and Ravindran 1988), The Philippines (Laurena et al. 1994; Jorge et al. 2007), Indonesia (Jorge et al. 2007), and in some African countries like Ghana, Mozambique, Malawi, Zambia (Onweluzo and Eilittä 2003) and Nigeria (Jorge et al. 2007). In India, various ethnic groups use it as a minor food (Janardhanan and Lakshmanan 1985; Rajaram and Janardhanan 1991; Josephine and Janardhanan 1992; Mohan and Janardhanan 1993, 1995). Indegenous groups in Northeast India like Khasi, Naga, Kuki, Jaintia, Chakma, Mizo, etc. consume the seeds and young pods as a vegetable (Lampariello et al. 2012). Due to the presence of relatively high levels of anti-nutritional factors, the seeds are processed with pre-preparation steps like soaking, boiling and/or roasting or mixed with salt (Arora 1981). The nutritional and economic significance of *M. pruriens* are schematically presented in Figure 9.

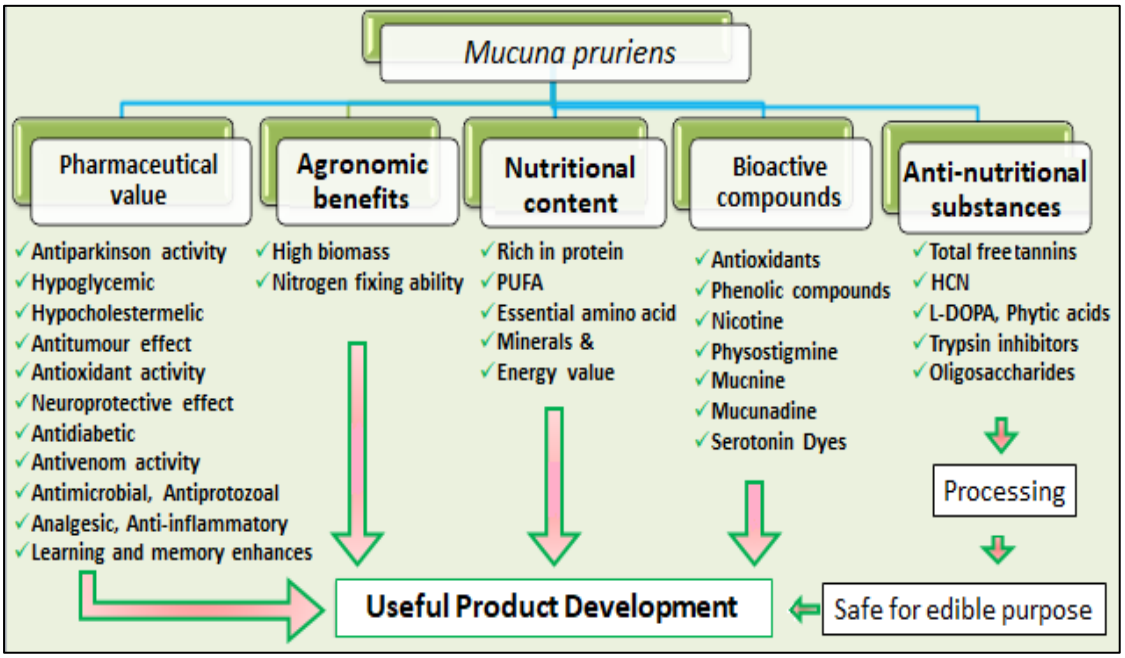


Figure 9: Overview on uses of *M. pruriens* (Mahesh 2015)

2.4.2. Nutritional properties

The proximate chemical composition, minerals and essential amino acids in *M. pruriens* seeds are summarized in Table 6. The protein content is in the range of 27.5% - 28.8%, which is higher than other food legumes such as pigeon pea (Kumar et al. 1991) and chickpea (Srivastav et al. 1990; Hira and Chopra 1995; Srivastava and Ali 2004). However, compared to other oil-rich legumes (ex. soybean, groundnut), the amount of crude fat (7.19%) is low (Table 6). The total dietary fiber (TDF) and ash content range from 6.5% - 9.7% and 3.3% - 5.5% respectively. Both the varieties show higher energy content (1568–1620 kJ $100g^{-1}$) as compared to pulse crops like cowpea, horse gram, moth bean, green gram and peas which are in the range of 1318-1394 kJ $100g^{-1}$ (Tresina and Mohan 2011; Tresina et al. 2014).

Table 6: Proximate chemical composition, minerals and essential amino acids of *M. pruriens*

M. pruriens	**var. *pruriens***	**var. *utilis***
Crude Protein (%)	27.51 - 31.24	26.26 - 28.82
Crude Lipid (%)	6.57 - 7.84	6.45 - 7.19
Total Dietary Fibre (TDF) (%)	6.52 - 9.71	8.46 - 9.73
Calorific Value (KJ 100 g^{-1})	1576.0-1620.0	1568.0-1602.0
Total sulfur amino acids (%)	1.16-1.32	
Moisture (g 100 g^{-1})	6.7-8.5	
Ash (g 100 g^{-1})	3.3-5.5	
Carbohydrate (%)	49.9 - 61.2	
Mineral composition (mg 100 g^{-1})		
Sodium	34.53 - 98.51	98.39 - 128.12
Potassium	1395.0- 1601.0	1628.0- 1846.0
Calcium	584.0 - 741.0	689.0 - 746.15
Magnesium	440.78 - 596.0	298.0 - 341.0
Phosphorus	410.0 - 583.0	327.0 - 456.0
Iron	4.45 - 7.56	12.41 - 14.74
Zinc	1.42 - 2.63	6.12 - 8.25
Essential Amino Acids (g 100 g^{-1})		
Threonine	4.08 - 5.28	3.40 - 3.52

Valine	3.33 - 6.30	3.64 - 4.60
Cystine	1.12 - 2.20	0.38 - 0.78
Methionine	0.54 - 1.10	0.84 - 0.96
Isoleucine	2.54 - 5.30	5.26 - 6.01
Leucine	5.17 - 6.50	5.36 - 7.01
Tyrosine	2.43 - 5.11	4.36 - 5.12
Phenylalanine	3.13 - 4.80	3.42 - 3.78
Lysine	4.85 - 5.40	5.36 - 5.94
Histidine	2.23 - 4.31	2.20 - 3.44
Tryptophan	0.92 - 1.21	0.96 - 1.10
L-DOPA (%)	1 - 9	
Saponins (%)	1.2 - 1.3	
Phytic acid (%)	0.48 - 0.65	0.12 - 0.72
Total phenolics (mg/g)	3.4 - 4.78	3.26 - 4.88
Total tannins (mg/g)	0.14 - 0.23	0.14 - 0.24
Hemagglutinins (HU/mg)	9 - 166	14 - 162

(Source: Pirman et al. 2001; Gurumoorthi et al. 2003; Tresina and Vadivel 2011; Kalidass and Mohan 2010; Kalidass and Mahapatra 2014; Daffodil et al. 2016).

2.4.3. Anti-nutritional factors

One of the major barriers for the promotion of *M. pruriens* as a food or feed is the presence of antinutritional factors, which are present in high concentration as compared to other edible legumes (Bhat et al. 2007). These compounds include phenolics, tannins, lectins, phytic acid, trypsin inhibitors and L-DOPA (L-3,4-dihydroxyphenylalanine) (Ravindran and Ravindran 1988; Siddhuraju et al. 1996). However, pre-treatment or processing methods such as hydrothermal treatments, fermentation, and germination have shown promising results in reducing their content (Siddhuraju and Becker 2001; Wanjekeche et al. 2003). In the case of L-DOPA, a healthy adult can tolerate approximately 500 mg of L-DOPA per day (Lorenzetti et al. 1998). However, this dose may not apply for long term consumption by pregnant women, children and the people under medication (Szabo et al. 2002; Jorge et al. 2007).

2.4.4. Medicinal properties

M. pruriens is an important medicinal plant (Raina et al. 2011). The magic of *M. pruriens* is largely due to the presence of L-DOPA (Pugalenthi and Vadivel 2007b). The therapeutic potential of this drug for the treatment of Parkinson's disease is well recognized (Abbott 2010). Moreover, it can recover spermatogenic loss and has proven aphrodisiac property which makes *M. pruriens* seed the treatment of choice for infertility (Hornykiewicz 2002; Singh et al. 2013). Katzenschlager et al. (2004) assessed clinical effects and pharmacokinetics of natural L-DOPA vis-a-vis standard carbidopa, and reported distinct advantage of natural extracts over the synthetic preparation in the long-term management of Parkinson's disease.

Additionally, *M. pruriens* seed powder is used to treat rubefacient conditions as well as asthma, cancer, cough, dropsy, cholera, diarrhoea, dysuria, insanity, mumps, pleuritis, ringworm, dog bite, snakebite, sores, syphilis, and tumours (Kirtikar and Basu 1985). It has also anti-diabetic (Horbovitz et al. 1998), anti-inflammatory, antimicrobial (Soowora 1982), neuroprotective (Misra and Wagner 2007), anti-venom (Guerranti et al. 2008) and antioxidant properties (Bravo 1998; Sathiyanarayanan and Arulmozhi 2007). In Wistar male rats, the learning skills and memory boosting activities were found to have enhanced on administering seed extracts from *M. pruriens* (Poornachandra et al. 2005). It can stimulate growth hormone (Alleman et al. 2011) and assist in weight loss (Majekodunmi et al. 2011). Presence of hypoglycaemic and hypocholesterolemic effects have also been tested in rat models (Pant et al. 1968). Gupta et al. (1997) reported antineoplastic efficacy in their search for anticancer plants.

2.4.5. Agronomic benefits

M. pruriens has been playing a significant role in agriculture as a pasture crop, green manure, and soil-improving crop. Like other legume species, it can fix atmospheric nitrogen through a symbiotic relationship with soil microorganisms (Buckles 1995). It can effectively control nematode population (Carsky and Ndikawa 1998; Queneherve et al. 1998), shows allelopathic properties (Fujii et al. 1991; Valverde 2003) and exhibits resistance against wide-ranging diseases (Eilitta et al. 2002). Due to high N_2 fixing ability and biomass production, the plant is often regarded as a "characteristic example of green manures contribution to the sustainable agricultural system" (Buckles 1995). The agronomic benefits of *M. pruriens* are dealt in greater detail in earlier works (Carsky and Ndikawa 1998; Tarawali et al. 1999; Jorge et al. 2007; Leelambika and Sathyanarayana 2011).

2.5. Genetic studies

2.5.1 Molecular markers and genetic diversity

Elucidating the pattern of genetic diversity and relationship among the accessions in a germplasm collection is an important prerequisite for any breeding program (Azhaguvel et al. 2006). Diversity assessment based on morphological characters alone may not be fully reliable due to a limited number of markers and environmental effects (Tatikonda et al. 2009). Molecular markers on the other hand are independent of such influences and can be generated reliably using a small amount of DNA from any stage of growth. Earlier studies in *Mucuna* species used RAPD (Random Amplification of Polymorphic DNA) and AFLP (Amplified Fragment Length Polymorphism) markers for genetic diversity analysis (Capo-chichi et al. 2001; Padmesh et al. 2006; Sathyanarayana et al. 2008). Using AFLP method, Capo-chichi

et al. (2001) reported a narrow genetic base (3-13%) among the 40 US landraces of *M. pruriens*. Augmenting of the collection from the CIAT resulted in broadening of the genetic base of this collection (0-32%) (Capo-chichi et al. 2003). In India, Padmesh et al. (2006) observed good diversity (10-61%) among the *M. pruriens* collection from the Western Ghats of India using RAPD. The study reported a narrow genetic base within var. *utilis* (SI-0.82) as compared to var. *pruriens* (SI-0.70). These findings were further reinforced by an extended study by Leelambika et al. (2010) using combined morphometric, biochemical, isozyme and RAPD analysis on a large number of accessions as well other studies from different geographical regions of India (Sathyanarayana et al. 2011; Mahesh and Sathyanarayana 2015; Shetty et al. 2015; Patil et al. 2016; Leelambika et al. 2016). It has also been shown that some isozyme markers can act as a diagnostic markers for taxon delimitation in *Mucuna* species (Leelambika and Sathyanarayana 2011). Further, analysis of accessions involving all the three botanical varieties using AFLP markers established close genetic similarities between var. *pruriens* and var. *hirsuta* (Leelambika et al. 2010). It was thus suggested that these two varieties, along with their natural hybrids, can be merged into one sub-group under var. *pruriens*, thus allowing only two sub-groups viz., var. *utilis* and var. *pruriens* to be recognized in this species. This is expected to reduce nomenclatural issues in breeding programs.

Recently Tripathi et al. (2018) studied genetic diversity among the Northeast Indian accessions and reported moderate diversity, high gene flow and extensive admixture among these accessions. This study assumes significance considering the fact that the species is reported to have originated in Indo-China region which includes parts of Northeast India.

2.5.2. Linkage map and quantitative trait loci (QTL) analysis

Genetic linkage maps act as a critical resource base for important genetic studies and plant breeding programs (Staub et al. 1996). They are also useful in detecting chromosomal locations of genes/markers as well as understanding the effects of gene interactions, especially those governing the complex traits (Hayashi et al. 2001; Sun et al. 2014). Recently genetic maps have been developed in case of some underutilized legume species including azuki bean (Han et al. 2005), lima bean (Bonifácio et al. 2012), bambara groundnut (Ahmad 2012) etc. In case of *M. pruriens*, an AFLP based linkage map using F2 segregant population, published earlier (Capo-chichi et al. 2004), had shown prospects of good genome coverage with AFLP markers. Using this lead, the first framework linkage map of Indian *M. pruriens* indicating QTL positions for floral, pod and seed traits using F2 intraspecific population has been reported by Mahesh et al. (2016), thus creating a new genomic resource for Indian *M. pruriens*.

2.5.3. Genomics and NGS for marker development

Development and deployment of codominant markers such as microsatellites and SNPs for genetic studies signify important landmark in genomic resource development in *M. pruriens*. Microsatellite development using expressed sequence tags (ESTs) is cost-effective in case of underutilized species. Shetty et al. (2015) thus explored this option in *M. pruriens* using publicly available EST databases of four legume species viz., soybean, common bean, chickpea and cowpea which generated 22,457 SSR containing sequences. From these, primers were designed for 522 sequences and 50 were evaluated against a diverse panel of 25 *M. pruriens* genotypes, which produced polymorphic profiles with an average PIC of 0.65. As an extension of this project, the present thesis work was undertaken to transcriptome sequence two contrasting parents representing distinct botanical varieties of *M. pruriens* to generate a database of species-specific genic microsatellite markers in *M. pruriens*.

3. Objectives

The thesis work comprised the following **five** objectives:

1. To transcriptome sequence *M. pruriens* genotypes and construction of *de novo* assembly
2. To discover polymorphic microsatellite markers using transcriptome data
3. To validate the microsatellite markers against a panel of diverse *M. pruriens* genotypes
4. To annotate the functional attributes in *M. pruriens* transcriptome
5. To attempt the microsatellite-based genetic linkage map of *M. pruriens* based on available markers

4. Materials and Methods

4.1. Transcriptome sequencing and *de novo* assembly

4.1.1. Plant material

For transcriptome sequencing, we selected two *M. pruriens* parental genotypes viz., IC0620620 (Figure 10A) and IC0620622 (Figure 10B), belonging to different botanical varieties. Genotype IC0620620 is a cultivated variety representing var. *utilis* and IC0620622 is a wild variety belonging to var. *pruriens*. They revealed contrasting phenotypes (Table 7) and an ideal genetic distance determined based on both morphological and molecular data.

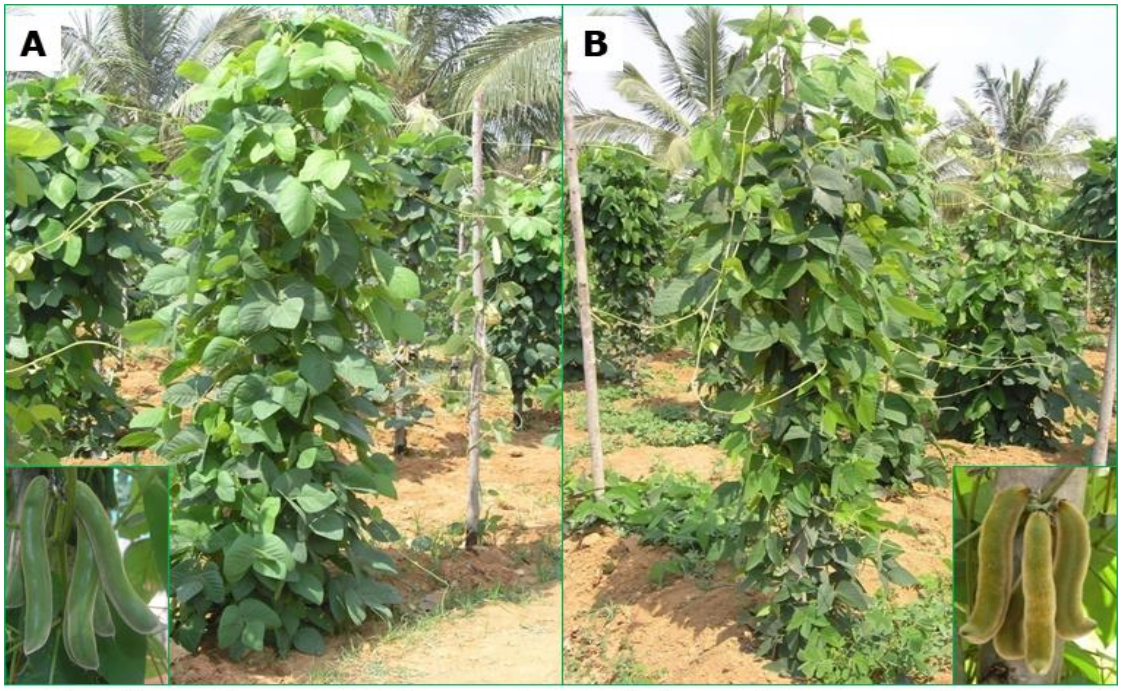


Figure 10: Parents used for transcriptome data and mapping population generation

In addition, a diverse panel of 25 *M. pruriens* accessions from germplasm collection as well as a Recombinant Inbred Line (RIL) population (F6) comprising 161 individuals, developed earlier using the above parents, were also raised for QTL

mapping purpose. All the plants were maintained at the Botanical garden of the Sikkim University. The latitude and the longitude of the experimental site were 27.31° N and 88.60° E respectively, and the site was 1650 m higher up the mean sea level.

Table 7: Comparison of morpho-agronomic characters of the parental accessions

Plant ID / **Parameters**	***M. pruriens***	
	var. *utilis* ♀	**var. *pruriens*** ♂
Place of collection	Hunasamaranahalli Karnataka	Triambakeshwar Maharashtra
NBPGR ID	IC0620620	IC0620622
Collectors ID	500108KA	500113MH
Latitude	13° 14' N	20° 00' N
Longitude	77° 62' E	73° 77' E
Days to emergence	7	5
TLL (cm)	12.837 ± 2.134	7.81 ± 0.378
ALL (cm)	12.475 ± 6.44	7.58 ± 0.577
TLW (cm)	8.562 ± 1.53	3.46 ± 0.374
ALW (cm)	7.825 ± 1.418	3.476 ± 0.447
TLS	Ovate	Ovate–lanceolate
ALS	Ovate	Ovate–lanceolate
PL (cm)	10.925 ± 2.65	3.287 ± 0.652
LT	Membranous	Membranous
LC	Green RHS 137C	Dark green, RHS 137A
DF	127	109
IL (cm)	34.33 ± 4.99	93.00 ± 6.82
NBPC	31.00 ± 3.74	86.75 ± 9.73
FC	Yellow green, RHS 145C	Dark violet, RHS 43A
FL (cm)	5.17 ± 0.287	4.68 ± 0.083
PTC	Creamish white	Golden orange
DFMP	140	131
POL (cm)	11.25 ± 0.45	8.7 ± 0.79
POW (cm)	2.13 ± 0.126	2.15 ± 0.11

POC	Curved	S-shaped
PI	Nonitching	Highly itching
NPPC	20.00 ± 5.02	15.00 ± 5.00
NPPP	203	340
DM	210	170
SL (mm)	15.17 ± 1.592	10.52 ± 1.524
SW (mm)	11.61 ± 0.662	7.63 ± 1.132
ST (mm)	7.32 ± 0.308	3.83 ± 0.616
SC	White	Brown RHS 199C
SP	-	Black mottled
SS	Rhomboid	Rhomboid
NSPP	5-6	4-6
HSW (g)	93.05 ± 5.84	27.75 ± 2.598

Note: TLL- terminal leaflet length; **ALL**- adjacent leaflet length; **TLW**- terminal leaflet width; **ALW**- adjacent leaflet width; **TLS**- terminal leaflet shape; **ALS**- adjacent leaflet shape; **PL**- petiole length; **LT**- leaf texture; **LC**- leaf colour; **DF**- days to flowering; **IL**- inflorescence length; **NBPC**- no. of flower buds per cluster; **FC**- flower colour; **FL**- flower length; **PTC**- pod trichome colour; **DFMP**- days to first mature pods; **POL**- pod length; **POW**- pod width; **POC**- pod curvature; **PI**- pod itchiness; **NPPC**- no. of pods per cluster; **NPPP**- no. of pods per plant; **DM**- days to maturity; **SL**- seed length; **SW**- seed width; **ST**- seed thickness; **SC**- seed colour; **SP**- seed coat pattern; **SS**- seed shape; **NSPP**- no. of seeds per pod; **HSW**- hundred seeds weight; *****not scored; the scale measurements are average value ± standard error; **RHS**- Royal Horticultural Society colour code chart.

4.1.2. RNA isolation

For RNA isolation, tissues harvested from different plant parts and across the developmental stages such as leaf, young flower bud, pod, and root were used. All the tissues, after harvesting, were placed immediately in liquid nitrogen and frozen until use. During RNA isolation, about 0.1 g of each sample tissue was pulverized to a fine powder using pre-chilled RNase-free treated mortar and pestle. To this, 2 ml of denaturing solution-I containing saturated phenol and DEPC treated water (0.8 ml) was added and mixed. The samples were then divided into two parts and transferred to 2 ml RNase-free microcentrifuge tubes and allowed to settle for 5 min at room temperature. To each tube, 0.2 ml of chloroform was then added, mixed well by continuous inversion, vortexed (<10 s) and incubated in room temperature for 10 min

for nucleic acid precipitation. The tubes were centrifuged at 13,000 rpm for 10 min at 4° C and the aqueous phase was transferred to a fresh RNAse free microcentrifuge tubes using sterile pipette. The supernatant was then added with 0.6 volume of isopropanol, vortexed briefly (<10 s) and kept at room temperature for 10 min to allow the nucleic acids to precipitate. The samples were later centrifuged at 13,000 rpm for 10 min at 4° C. The pellets were washed with 70% ethanol, air dried for 2-3 min, and finally dissolved in 20-50 µl DEPC-treated water.

4.1.3. RNA quality check

The purity of the isolated RNA samples was examined by measuring the ratio of absorbance: A260/A280 nm and A260/A230 nm using Nanodrop 2000 spectrophotometer (Thermo Scientific, USA). The RNA samples were further tested using electrophoresis on 1% denaturing formaldehyde-agarose gel by loading 2 µl of RNA along with dye and ladder. The integrity of the isolated RNA was determined through an RNA integrity number (RIN), using Bioanalyzer. Only those samples which have RIN>7.0, purity value of ~2.0 from 260/280 ratio, and 2.0-2.2 from 260/230 ratio, were used for cDNA library construction. In the case of pooled samples, RNAs from different tissues was pooled in equimolar concentration for both the genotypes separately and used for cDNA library construction.

4.1.4. cDNA library construction and illumina sequencing

The cDNA library construction and sequencing were carried out at the Next Generation Genomics Facility (NGGF) at Centre for Cellular and Molecular Platforms (C-CAMP), Bengaluru, India. To obtain good coverage of tissue-specific transcripts, each sample was regulated to contain the same concentration of RNA (~400 ng/µl) from which 20 µg was used for cDNA synthesis. For pooled samples,

equimolar RNA from each bulk was used. First, from the total RNA preparation, poly (A) mRNA was isolated with oligo (dT) (deoxythymine) beads using column chromatography. Following purification, mRNAs were fragmented at high temperature. The first cDNA strand was produced using random hexamer primers and superscript^{TM} III (Invitrogen). The second strand was generated using DNA polymerase I. Short fragments were purified and after end separation poly (A) tails were connected using sequencing adapters. Suitable fragments were recovered for PCR amplification using agarose gel electrophoresis. The paired-end library preparation was done following the protocol of the Illumina TruSeq stranded RNA sample preparation kit (Illumina Inc.) as per the manufacturer's instruction.

For sequencing purpose, 5 paired end 2x-100 bp cDNA libraries were constructed. This included (a) one each from the pooled RNA samples of a var. *utilis* (accession IC0620620) and var. *pruriens* (accession IC0620622) and (b) three separate libraries from the leaf, pod and root RNAs of var. *pruriens*. All the samples were sequenced on a single lane of 2×100 paired-end run by Illumina HiSeq^{TM} 1000. The library was connected to a paired-end flow cell comprising complementary adapters and fragments were amplified to create the overlapping clusters. The adapters were designed to allow selective cutting of the forward DNA strand after resynthesizing of the reverse strand during sequencing. The copied reverse strand was subsequently used to sequence from the other end of the fragment.

4.1.5. Redundancy removal and *de novo* assembly

The so obtained sequence reads were subjected to quality filtering. Initially, we used FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) to filter each read separately. All those sequences which did not match the set criterion - minimum

length of 200 bp (-l 200) were removed by trimming. For this, we used the tool FASTX Clipper and FASTA Trimmer, by setting the first base of 15 (-f 15) and last base of 800 (-l 800) respectively. After removing the adapter contamination, we checked the quality with the minimum phred score of 20 (-q 20) and minimum percentage of included bases of 80 (-p 80). After cleaning steps, the quality of output reads was inspected using FastQC software v. 0.11.3 (Andrews 2010) with the phred score quality value (Q) ≥30 (command -Q 30) and demultiplexed using an option of one mismatch in the index (2016s). By this method, we filtered and recovered a total of 191,233,242 quality reads (including both the genotypes - IC0620620 & 622) from the total of 195,757,776 raw (paired-end) reads. These reads, from both the genotypes, were used to construct *de novo* assembly.

Several tools are available for construction of *de novo* assembly from RNA-Seq reads such as Multiple-k (Surget-Groba et al. 2010), Rnnotator (Martin et al. 2010), Trans-ABySS (Robertson et al. 2010), Velvet-Oases (Schulz et al. 2012), and SOAPdenovo-Trans (http://soap.Genomics.org.cn/SOAPdenovo-Trans.html). In recent years, Trinity (Grabherr et al. 2011), which generates individual de Bruijn graphs, is gaining popularity for *de novo* assembly construction using transcriptome data. We used Trinity assembler for construction of *de novo* assembly in our work. The assembling process generated a total of 67,561 clean transcripts from a total of 72,561 Trinity assembled transcripts with 95% identity as cut-off by using CD-HIT v. 4.6.

Considering six-fold variation in the number of reads obtained between the two genotypes, separate assembly for each genotype was not performed at this stage and a combined assembly was generated using the data from both the genotypes.

4.2. Discovery of polymorphic microsatellite markers

4.2.1. SSR identification

For microsatellite discovery, we used MIcroSAtellite identification search tool, popularly called as MISA, which identifies simple sequence repeat motifs (http://pgrc.ipk-gatersleben.de/misa) (Thiel et al. 2003). The criteria for the microsatellite motifs selection was set as: mono-nucleotide repeats > 10 times, di-nucleotide repeats > 6 times, tri-, tetra-, penta-, hexa-nucleotide repeats > 5 times. From the MISA output, we identified repeats up to deca-nucleotide repeats; but the results here are presented up to hexa- repeats only.

4.2.2. Identification of polymorphic SSRs

Further, for detection of polymorphic SSRs (pSSRs) between the parental genotypes, we used a specialized algorithm called lobSTR (Gymrek et al. 2012) which identifies polymorphic short tandem repeats (STR). For this, reads from each genotype were mapped against combined assembly: IC0620620-22. We developed a conventional STR reference and passed raw reads through the program to be aligned around the SSR regions. By using the allelotype option in the lobSTR, BAM files were genotyped. Using samtools (Li et al. 2009), BAM files generated from the alignment were sorted and indexed for pSSRs.

4.2.3. Primer design

The above approach identified of total 3865 pSSRs from the total 7943 SSRs of which 787 showed quality value ≥10. From this, a portion of 134 SSRs was randomly selected for validation. Primer design was done using Batch-Primer3 program (http://probes.pw.usda.gov/batchprimer3/) (You et al. 2008). The parameters for

primer design were set as product size of 70 to 300 bp; primer size of 18 to 22 bp with optimal length 20 bp; primer melting temperature (Tm) of 50° C to 60° C with an optimum at 55° C; and primers at least 5 bp away from the SSR locus. The designed primers were synthesized at Eurofins Genomics India Pvt. Ltd. Bengaluru for validation.

4.3. Validation of microsatellite markers

4.3.1. Establishment of germplasm

We used a panel of 25 diverse *M. pruriens* accessions from our germplasm for validation. These accessions represented different geographical regions in India including Peninsular, East, and Northeast India. The collection included representative taxa from the three botanical varieties of *M. pruriens* viz. var. *utilis* (n=6), var. *pruriens* (n=13), and var. *hirsuta* (n=6). The list of accessions and the geographic coordinates of their collection sites are given in Table 8, and the locations are shown in Figure 11. For germplasm establishment, the seeds of all the 25 accessions were planted in the Sikkim University botanical garden and established as per the standard package of practices (Figure 12). The experimental site is located at Gangtok in East Sikkim, positioned in the Eastern Himalayan tract with a longitude of 27^0 33’ N, latitude 88^0 62’ E with an altitude of 1600 m AMSL (above the mean sea level).

For seed germination and seedling development, initially, 5 seeds of each accession were planted in each pot. The potting mixture was prepared by adding soil: sand: compost manure in 2:2:1 ratio. No external fertilizer or insecticide was applied at any stage of the plant growth. After successful germination, extra plants were trimmed to retain only a single plant per pot.

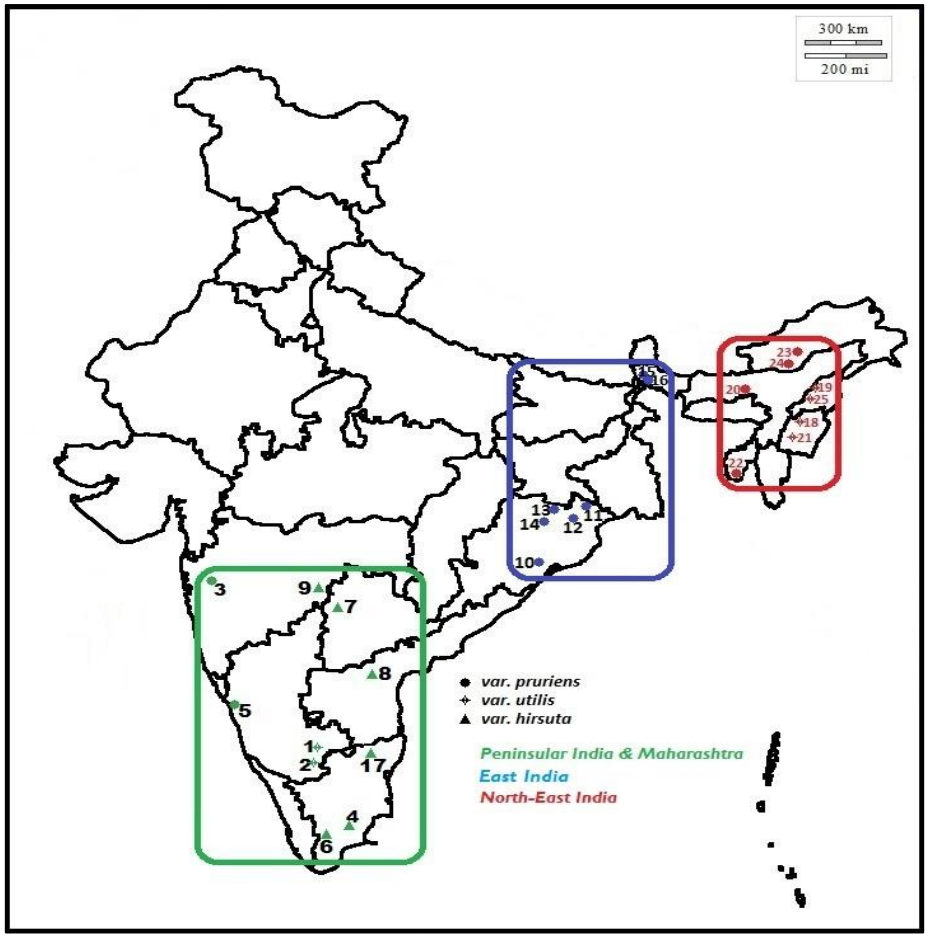


Figure 11: Map depicting collection locations of *M. pruriens* accessions used in this study (Map source: https://d-maps.com/carte.php?num_car=24865&lang=en)

Figure 12: Germplasm of *M. pruriens* maintained at Sikkim University garden

Table 8: Details of accessions used for EST-SSR validation

Sample No.	Accession No.	Variety	Latitude (N)	Longitude (E)	Altitude (AMSL) (m)	State of Origin
1	500101-KA*	var. *utilis*	13° 14'	77° 62'	910	Karnataka
2	IC-0620620**	var. *utilis*	13° 14'	77° 62'	911	Karnataka
3	IC-0620622**	var. *pruriens*	20° 00'	73° 77'	745	Maharashtra
4	500120-TN*	var. *hirsuta*	09° 55'	78° 07'	138	Tamilnadu
5	IC-0620624**	var. *pruriens*	14° 48'	74° 12'	7	Karnataka
6	500136-TN*	var. *hirsuta*	10° 04'	77° 45'	298	Tamilnadu
7	500147-AP*	var. *hirsuta*	18° 39'	78° 10'	383	Telangana
8	500154-AP*	var. *hirsuta*	16° 04'	78° 52'	434	Andhra Pradesh
9	500186-MH*	var. *hirsuta*	19° 09'	77° 27'	373	Maharashtra
10	500192-OR*	var. *pruriens*	20° 18'	85° 62'	31	Odisha
11	500193-OR*	var. *pruriens*	21° 94'	86° 72'	51	Odisha
12	500194-OR*	var. *pruriens*	21° 94'	86° 72'	51	Odisha
13	500195-OR*	var. *pruriens*	21° 63'	85° 58'	650	Odisha
14	500196-OR*	var. *pruriens*	20° 47'	85° 12'	186	Odisha
15	500197-WB*	var. *pruriens*	26° 71'	88° 43'	125	West Bengal
16	500199-WB*	var. *pruriens*	26° 70'	88° 80'	65	West Bengal
17	500202-TN*	var. *hirsuta*	11° 48'	78° 13'	554	Tamilnadu
18	500210-MN*	var. *utilis*	25° 41'	94° 47'	782	Manipur
19	500211-NL*	var. *utilis*	25° 67'	94° 12'	1333	Nagaland
20	500212-AS*	var. *pruriens*	26° 11'	91° 44'	61	Assam
21	500217-MN*	var. *pruriens*	25° 68'	93° 03'	776	Manipur
22	500219-TR*	var. *pruriens*	23° 50'	91°25'	64	Tripura
23	500221-AR*	var. *pruriens*	27° 08'	93°40'	1035	Arunachal Pradesh
24	500224-AR*	var. *pruriens*	27° 08'	93°40'	296	
25	500267-NL*	var. *utilis*	25° 68'	94°08'	1360	Nagaland

*Collector's ID of newly collected accessions; **National genebank ID

4.3.2. Marker data generation

4.3.2.1. Genomic DNA isolation

For genomic DNA isolation, 0.5-1.0 gm of fresh leaf tissue was harvested from 15 days old seedling**.** DNA isolation was carried out by using the cetyltrimethylammonium bromide (CTAB) method described by Doyle and Doyle (1987) with minor modifications. Harvested samples were immediately transferred to liquid nitrogen to freeze the sample until isolation. During the isolation, the leaf samples were pulverized using mortar and pestle, and transferred to Oakridge tube containing 10 ml CTAB extraction buffer along with charcoal (0.5%) and β-mercaptoethanol (0.2%) and incubated at 60^0 C for 1h. The supernatant obtained, after centrifugation was washed twice with SEVAG (24-chloroform : 1-isoamyl alcohol) and once with phenol: chloroform: isoamyl alcohol (25:24:1) mixtures. The nucleic acids were allowed to precipitate in isopropanol (0.67 volume) followed by ethanol (70%) wash. The air-dried DNA pellets were resuspended in 0.5 ml of 1X Tris-EDTA buffer (pH 8.0). After RNase treatment, the quantity of DNA was tested using agarose gel electrophoresis with Ethidium bromide (EtBr) as an intercalating agent.

4.3.2.2. SSR amplification

Totally 134 SSRs were shortlisted for validation. 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, 1 µM each primer and 1 U of *Taq* polymerase. The PCR conditions were as follows: initial denaturation at 94^o C for 3 min followed by 35 cycles of 30 s at 94^o C, 30 s at annealing temperature (Tm) and 20 s at 72^o C with a final extension of 7 min at 72^o C. The amplification was visualized using gel

documentation system (Uvi-Tech DOL-008.XD, England). Next, the PCR products generated by different dye-labeled primers were pooled in equal volume and 1.0 µl of each amplicon was mixed with 7 µl of formamide, 0.05 µl of the GeneScan™ 500 LIZ® Size Standard (Applied Biosystems, USA) and 2.95 µl distilled water. DNA fragments were denatured and size fractioned with capillary electrophoresis on an ABI 3730 DNA Analyzer (Applied Biosystems, USA).

4.3.2.3. Data filtering and analysis

The marker data were filtered using the criteria: minor allele frequency > 0.05 and the missing percentage <20%. This resulted in 52 primer-pairs (Table 9) for downstream analysis. Only reproducible and consistent fragments were considered and scored as dominant marker with the presence of a band indicated as "1", absence as "0" and "**.**" for missing data. The genotyping data were further transformed into a bi-allelic format (e.g., 1 denoted as AA and 0 as GG) and a HapMap file was produced. The HapMap file was converted into Variant Call Format (VCF) using TASSEL v 5.2.29 (Bradbury et al. 2007). The quality control on the dataset was performed using VCF tools (Danecek et al. 2011). Markers with minor allele frequency (MAF) > 0.05, max-missing sites <20%, and accessions with maximum missing site <20% were retained. The genotypic data were used to analyse population genetic parameters such as: polymorphic information content (PIC), marker index (MI), primer resolving power (Rp), effective multiplex ratio (EMR), observed heterozygosity (Ho), expected heterozygosity (He), effective number of alleles (Ne), gene diversity (h) and Shannon's Information index (I).

The PIC value and mean gene diversity (h) of each SSR marker was calculated using software Powermarker V 3.25 (Liu and Muse 2005) and the scores/values were

averaged for each marker. According to Roldan-Ruiz et al. (2000) formula, the PIC for each marker is calculated as= 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. Marker index (MI) is the product of PIC and EMR. By following Powell et al. (1996), MI was calculated as PIC × EMR, where EMR (EMR = β × n) is the product of the fraction of polymorphic loci (β) and the number of polymorphic loci (n). For each PC, the resolving power (Rp) was obtained according to Prevost and Wilkinson (1999) using formula Rp = ∑Ib, where band informativeness (Ib) was calculated by 1 – [2(0.5-p)], where, 'P' is the proportion of accession containing the band/fragment.

Diversity indicators were calculated in total, per population, and per SSR fragment using the software GenAlEx 6.5 (Peakall and Smouse 2012), POPGENE 1.31 (Yeh et al. 1999). Number of different alleles (Na), effective number of alleles (Ne) (Kimura and Crow 1964), Shannon's information index (I) (Lewontin 1972), 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 and Smouse 2012).

POPGENE was used to calculate overall diversity in the 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 Nm = 0.5 (1-Gst)/Gst). Jaccard's similarity coefficient values were used to estimate the genetic similarity between the accessions. Based on the similarity matrix, dendrograms were constructed based on UPGMA, and results were viewed using FreeTree V1.0.0.0 and TreeView (Win32) V1.6.6 (Pavlicek et al. 1999) software.

Table 9: List of 52 EST-SSR primer-pairs used for data generation

S.No.	Primer ID	Forward Primer	Reverse Primer
1	MPU_06	GCATTTCAAATAAACGTCAAC	TTGACACAATTTGACAGAGGT
2	MPU_07	AATGGATCCCTTTTCTCTATG	TATTGGAATAGATCCCCTTGT
3	MPU_15	AGTTAAAACTCACCCACCCTA	TGGAGGTGAGTGAATAGATGA
4	MPU_16	GACTCCAACTCCTCCTTCA	ACTGTTGTTGTTGCTGATGTT
5	MPU_18	GAAGAGGAGGTTCAGAACAGT	GCATCAGATACAACAAAGGAG
6	MPU_21	CCGAGAGTACAAGGGTAAAAA	TTCTCCTTTCCCTGAAGATAC
7	MPU_22	TAAAACTCCTTTTCCTTCTCC	AGTTCCTTCAAAATACGCTTC
8	MPU_23	AAGCTTGTCACTGTCAAAAAG	ATGCAACACATGTCAACACT
9	MPU_24	CATATCCTAGACCGCAGTTG	CGTTTTGGTCATTGTGAAT
10	MPU_27	ATGAGAGTCTGAGGGAGATTG	ATGTTGCATTGTCTCAACAGT
11	MPU_30	AAAAACTTTGCAGTACCGATT	TGATTCATTCAAAATCTCCAC
12	MPU_31	TGTCATTGGCTTGATATTCTT	GGAACGGAAATAAATTCTTGT
13	MPU_33	CCAGCTTACAACAACAGTCTT	AACTTTGGACAGGAATATTGG
14	MPU_37	GTTCTCCTTCTCTGGTGAGTC	CTAAGACAAGCACACCTCAAG
15	MPU_39	AGTGGCAACTATATGACGGTA	GAGTCACAACCAAATAGCTCA
16	MPU_42	CAAAGATGACGAAGATGGATA	TCTTTCAACTCAATCTTCCAA
17	MPU_43	GTCTAGGGTGGCTTTGTAGAC	CAACACCCACCAAATCTAGTA
18	MPU_45	ATCTCTGTCTCTGTCCCTTTC	GCGATCTCTCTCTTTCTCTCT
19	MPU_47	GTTCACCACATCCATTATGAC	ACCACACACACACATACACAC
20	MPU_48	GGTGGAGACCTTACACAATATG	CACCTCCAGTTTATTCACCTC
21	MPU_49	GGAGGATAAAAGTGCTTCATT	ACCGTCCTCTTAATTGTTGAT
22	MPU_52	AAGGAAAGCATATGAAAGGAG	ATTCCTCCAATTCTATCTCGT
23	MPU_57	TATCGATCTCAATTCCCATAA	TCTCTCTCTAAAGCCCAAAAC
24	MPU_58	CACTCTCAGTGAGTCTTCTGG	GAATGCTTGAGGCATAGAAG
25	MPU_59	GAGGTGTTGAAGAAGTGTGAG	ATTCCCACCATTCTCATTAGT

26	MPU_64	GCTTTCTATGAATGACTGCAT	AATTTCCAACCATGGATTAAG
27	MPU_65	CCAGTTTAGCTTGCACATACT	CCAATAGGCGAATTATATGAC
28	MPU_68	GGGTTAGGGTTACGGAAA	GCGATACTTCCTTTTCAAGTC
29	MPU_72	AAGAGCGATAGAGGAAGACAT	TGAAGTAGATTGGCAGGTTAG
30	MPU_78	TGAAGAGCTTTCAAAGTGAAG	GAACGAATTTTTAGGGTGATT
31	MPU_80	ACCAGAAAGGTAAAGCTCTCA	TTTTTGTTACTCTCCAAGCAG
32	MPU_81	TTCTATCGGAGAAGAAGTCG	CACACCTCTCTCTCTCTCTCTC
33	MPU_83	CACTCATGGAAGGTTCTAGGT	CTCATTTTGCTTTTCAAGAGA
34	MPU_88	GGAACTGGGTATGAGGTATTT	AGATGAGACCCCATTTGTTAC
35	MPU_89	ATAGCTCCGATTACCCAAAT	CACCCATTTCTCTCTTTCTCT
36	MPU_90	TTCCAAACAGTGGTCATAATC	TGTATTGCTGTTGGTAAGGAT
37	MPU_92	CGGTCGTAGTAATCGCTTC	GGTTTTGCTCTGTTTTCTCTT
38	MPU_98	TGCCTTATGAGGAATCTTACA	GTATTTCCCCTTGTTCATCTT
39	MPU_99	ACTGCAGAATGGAACTTTGTA	TAAAAGCCCTACTCAAAATCC
40	MPU_102	GAGAGAACTTCAATTCCCCTA	GCTTGTTCATGTTTCTTTGTC
41	MPU_106	GGAGAAATCAAATAGGTCGTC	AGTTCAAACCTAACCGAACTC
42	MPU_107	CGTATTTGTTGAACAGGATGT	GTCTCTCCTACCTCGATCAGT
43	MPU_111	CATCAACCTTATTCGACTCAT	AGTTCCTCCGTTACGTCAT
44	MPU_114	GTCTCTTTCACTCTTCCGTTT	GCAAGGTCTCTTTCACTCTTT
45	MPU_115	AAAGGCAATTGAGATCTAAGAA	CCTTTGTCTCCTCTTTCTTTC
46	MPU_119	GGGTTTGACAGGAAGTACAA	ACTTCCTAGAGCTTCCACTGT
47	MPU_121	TCATCACCATCACCATCAC	GATATCTGCCAGGTCCATC
48	MPU_122	TTCCTCGTGGAGATACAGATA	TGATGCATGCTATGATTAGAA
49	MPU_123	TTGTGCTGTTGTTATGATTGA	CTTTGTTTAAAACTGAACCA
50	MPU_124	CCTTCCCTTTAGATGTGAAAT	ACATTGATAGCAGTGGAGAAA
51	MPU_125	AGGAGAGAGAGTGAAATTGGA	ACAACGTGAACAGAGAGAGAA
52	MPU_126	TCTATGGGTTTGTCCTCATC	CATTTTTCCACAATCACTTTC

Bayesian framework implemented in the fastSTRUCTURE program (Raj et al. 2014) was used to examine the population structure with the following commands: -prior simple -full -seed = 100 -cv=5 for subgroups 2 to 10 (K=2 to 10). By plotting the marginal likelihood, cross-validation errors against the number of subgroups, and with the program fastSTRUCTURE script "choosing model complexity" - the output was generated along with determining the possible range of subgroups. The range of subgroups identified was then inspected for a proportion of affiliation of genotypes to particular subgroups (coefficient of ancestry values) and known geographical origin data to arrive at a precise number of subgroups in the collection.

4.4. Functional annotation of transcriptome

For functional annotation, high-quality reads were separately assembled using Trinity with default parameters. The assembled unigene sets were pooled and assembled into non-redundant unigenes. Finally, to gain insight into protein function in our transcriptome, all non-redundant unigenes were searched against the NCBI's non-redundant (Nr) protein database, Swiss-Prot and UniRef90 as well as non-coding RNA database (ncRNAs) and Conserved Domain Database (CDD) using BLASTX with an E-value threshold 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. Based on the results, Annocript v1.1.2 pipeline (Musacchia et al. 2015) was used to obtain Gene Ontology (GO) terms relating to the biological process, cellular components, and molecular functions. The putative orthologs of genes implicated in various pathways like Enzyme Classes (EC) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were also ascertained.

4.4.1. Sequence similarity with other legume species

We compared the set of genes characterized in *M. pruriens* transcriptome with the gene assemblies in other legume species. For this, protein sequences (*Medicago truncatula* Gaertn. Mt 4.0; *Lotus japonicus* (Regel) K. Larsen release 2.5; *Phaseolus vulgaris* L. release 1.0; *Glycine max* (L.) Merr. release 1.1; *Cicer arietinum* L. and *Cajanus cajan* (L.) Millsp. version 5.0) were downloaded from the NCBI database. We performed BLASTX searches 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.

4.4.2. Mining transcription factor families

For determining the transcription factor gene families, we downloaded the plant transcription factors database (PlnTFDB) 3.0 (http://plntfdb.bio.uni-potsdam.de/v3.0) (Pérez-Rodríguez et al. 2009) and queried *M. pruriens* transcripts against the PlnTFDB using BLASTX with an E-value cut-off of $1e^{-05}$.

4.4.3. Expression analysis

For expression analysis, raw reads obtained from transcriptome sequencing of individual tissues of *M. pruriens* var. *utilis* (IC0620620) were aligned to the 620-22 assembly separately using bowtie aligner version 1.1.1 (Langmead and Salzberg 2012). For calculating the read counts, we normalized the levels of gene expression in each library to create an effective library size. We employed RSEM version 3.0 (Li and Dewey 2011) to compute the read count and estimate/predict expression levels as FPKM (fragments per transcript kilobase per million fragments mapped) using edgeR software in R (Robinson et al. 2010). A dispersion value of 0.1 was used in our expression analysis. Further, differentially expressed transcripts (DETs) were

assessed with log-fold expression change ≥4 and a p-value of 0.001. To identify common DETs across all the tissue types, pair-wise comparisons between the three tissues were carried out by comparing the sequenced samples and top 50 DETs were extracted to generate a heat map and represent the dynamic expression patterns in different tissues. Besides, based on the annotation, we also extracted information related to top differentially expressed transcripts in secondary metabolite biosynthesis classes to determine expression patterns linked to secondary metabolite pathways.

4.5. Microsatellite-based genetic linkage map

4.5.1. Mapping population

The Recombination Inbred Lines (RIL) population of 161 individuals generated earlier by Mahesh et al. (2016) using the parents: IC0620620 (♀, *M. pruriens* var. *utilis*) and IC0620622 (♂, *M. pruriens* var. *pruriens*) were used for linkage map construction. The comparison of morpho-agronomic characters of the parental accessions along with F1 hybrid and F2 population are given table 10. The experiments leading to RIL population generation were conducted by Prof. N. Sathyanarayana – the supervisor of this thesis and his group at Sir M Visvesvaraya Institute of Technology, Bengaluru. The pictures of the crossing experiments are reproduced below with the permission of Dr. S Mahesh (Figure 13).

4.5.2. Polymorphism and segregation analysis

For determining polymorphism, genomic DNA was extracted from the young leaf tissue (2-3 weeks old) of the F6 RIL individuals using CTAB method described earlier. For linkage map construction, a total of 100 microsatellite markers were selected randomly for determining the parental polymorphism. The genotyping were

Table 10: Comparison of morpho-agronomic characters between the parents, F1 hybrid and F2 population (From Mahesh et al. 2016).

Plant ID / Parameters	*M. pruriens*			
	var. *utilis* ♀	**var. *pruriens* ♂**	**F1 (hybrid)**	**F2 (200 random individuals)**
Place of collection	Hunasmaranahalli, Karnataka	Triambakeshwar, Maharashtra	Mucuna field Gene Bank, Bengaluru, Karnataka	
NBPGR ID	IC0620620	IC0620622	--	--
Collectors ID	500108KA	500113MH	--	--
Latitude	13° 14' N	20° 00' N	13° 14' N	13° 14' N
Longitude	77° 62' E	73° 77' E	77° 62' E	77° 62' E
Days to emergence	7	5	11	6–10
TLL (cm)	12.837 ± 2	134 7.81 ± 0.378	12.00 ± 0.238	*
ALL (cm)	12.475 ± 6.44	7.58 ± 0.577	10.00 ± 5.44	*
TLW (cm)	8.562 ± 1.53	3.46 ± 0.374	8.00 ± 0.389	*
ALW (cm)	7.825 ± 1.418	3.476 ± 0.447	6.70 ± 0.347	*
TLS	Ovate	Ovate-lanceolate	Ovate	Both parental types
ALS	Ovate	Ovate-lanceolate	Ovate	Both parental types
PL (cm)	10.925 ± 2.65	3.287 ± 0.652	11.2 ± 0.532	*
LT	Membranous	Membranous	Membranous	Membranous
LC	Green RHS 137C	Dark green, RHS 137A	Green RHS 137C	Both parental types
DF	127	109	193	125–220
IL (cm)	34.33 ± 4.99	93.00 ± 6.82	23.25 ± 5.36	*
NBPC	31.00 ± 3.74	86.75 ± 9.73	22.83 ± 4.87	*
FC	Yellow green, RHS 145C	Dark violet, RHS 43A	Dark violet, RHS 43A	Both parental types

FL (cm)	5.17 ± 0.287	4.68 ± 0.083	4.8 ± 0.073	*
PTC	Creamish white	Golden orange	Dense silvery white	Both parental, F1 types & black velvety
DFMP	140	131	260	150–280
POL (cm)	11.25 ± 0.45	8.7 ± 0.79	10.73 ± 0.22	*
POW (cm)	2.13 ± 0.126	2.15 ± 0.11	1.88 ± 0.083	*
POC	Curved	S-shaped	Curved	Both parental types
PI	Nonitching	Highly itching	Highly itching	Both parental, F1 & extreme phenotypes
NPPC	20.00 ± 5.02	15.00 ± 5.00	17 ± 1.22	*
NPPP	203	340	460	*
DM	210	170	310	180 - 320
SL (mm)	15.17 ± 1.592	10.52 ± 1.524	12.98 ± 1.301	11.95 - 18.7
SW (mm)	11.61 ± 0.662	7.63 ± 1.132	9.29 ± 0.976	9.78 - 14.57
ST (mm)	7.32 ± 0.308	3.83 ± 0.616	6.86 ± 0.924	5.75 - 9.71
SC	White	Brown RHS 199C	Black	Both parental & F1 types, light brown, black, dark brown, black/brown mottled
SP	-	Black mottled	Brown mottled	-
SS	Rhomboid	Rhomboid	Ovoid	Both parental and F1 types
NSPP	5-6	4-6	5-6	5-6
HSW (g)	93.05 ± 5.84	27.75 ± 2.598	73.80 ± 7.158	42.98 ± 143.8

Note: TLL- terminal leaflet length; **ALL**- adjacent leaflet length; **TLW**- terminal leaflet width; **ALW**- adjacent leaflet width; **TLS**- terminal leaflet shape; **ALS**- adjacent leaflet shape; **PL**- petiole length; **LT**- leaf texture; **LC**- leaf colour; **DF**- days to flowering; **IL**- inflorescence length; **NBPC**- no. of flower buds per cluster; **FC**- flower colour; **FL**- flower length; **PTC**- pod trichome colour; **DFMP**- days to first mature pods; **POL**- pod length; **PO**W- pod width; **POC**- pod curvature; **PI**- pod itchiness; **NPPC**- no. of pods per cluster; **NPPP**- no. of pods per plant; **DM**- days to maturity; **SL**- seed length; **SW**- seed width; **ST**- seed thickness; **SC**- seed colour; **SP**- seed coat pattern; **SS**- seed shape; **NSPP**- no. of seeds per pod; **HSW**- hundred seeds weight; *****not scored; the scale measurements are average value ± standard error; **RHS**- Royal Horticultural Society colour code chart.

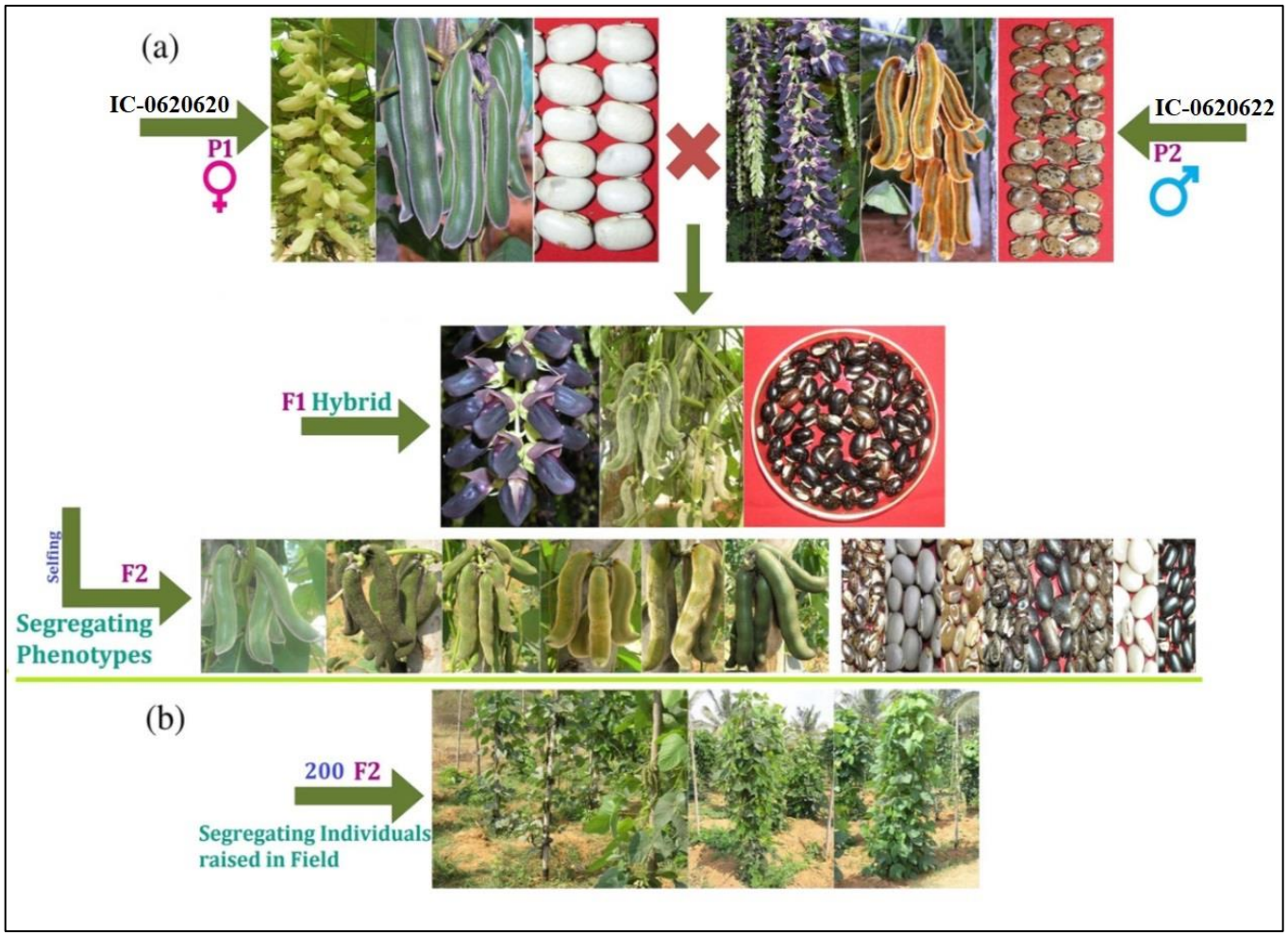


Figure 13: Representation of (a) Fruit, pod & seeds of F1 hybrid and F2 segregating population (b) F2 population in the field

carried out as per the method described earlier. Of the 100 markers used, only 62 showed parental polymorphism (62%). From this, we obtained usable data for only 32 markers, while others either failed to amplify in a large number of individuals or produced extreme linkage distortion in Chi-square test. Consequently, we could use only 32 markers for linkage map construction.

4.5.3. Scoring of SSR markers for linkage map

The SSR fragments were scored as presence versus absence of bands, and therefore for the parents, F6 generation markers were assigned to either parental allele for map construction. Markers originating from each parent were scored according to the standard coding system using A, B, and H of JoinMap v4 analysis (Van Ooijen 2006) where "A" is homozygous for parent 1 allele, "B" is homozygous for parent 2 allele, "H" is heterozygous and the missing data recorded as "**.**".

4.5.4. Linkage map construction

By using the Grouping Module of JoinMap v4, the robustness of the data sets in terms of linkage group was confirmed. Locus order of each linkage groups was determined temporally using the following parameters: A minimum LOD score of 3.0 and maximum recombination fraction (θ) of 0.50 were set as thresholds for linkage group. Maximum recombination fraction (θ), the goodness of fit jump threshold for removal of loci = 5.0, the number of extra loci after which a ripple is performed = 1, and third-round = yes. Recombination fractions were converted into map distances in centimorgans (cM) by means of the Kosambi mapping function.

5. Results

5.1. Transcriptome sequencing and *de novo* assembly

5.1.1. RNA quality check

All the isolated RNA samples (leaf, pod, root and pooled) were subjected to quality check using UV absorbance and RIN determination. For all the samples, the ratio of UV absorbance at 260/280 was in the range of 1.9-2.1 and for 260/230 it was 2.0-2.5. Further, all the samples reported RIN greater >7.0 (Table 11, Figure 14), confirming to the quality standards required for transcriptome sequencing.

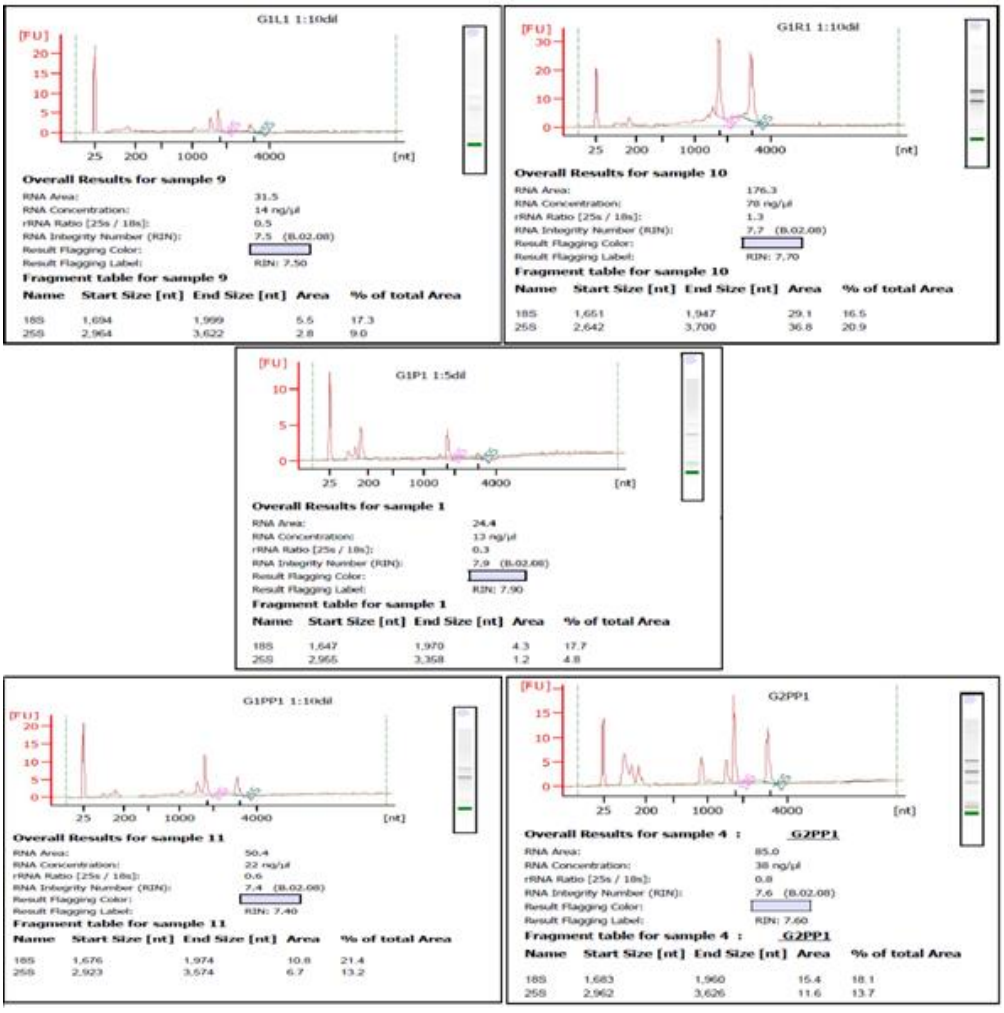


Figure 14: Bioanalyzer profiles of the isolated RNA samples

Table 11: RIN number of isolated RNA samples

RNA Sample		RIN score
Tissue material	**Given ID**	
G1 Leaf	G1 L1	7.5
G1 Pod	G1 P1	7.9
G1 Root	G1 R1	7.7
G1 Pooled	G1 PP1	7.4
G2 Pooled	G2 PP1	7.6

G1: Genotype - IC0620620; **G2**: Genotype - IC0620622.

5.1.2. *De novo* assembly and redundancy removal

Illumina sequencing from both the accessions generated total of 18.24 GB data comprising 167,986,452 and 27,801,324 raw reads from the genotypes: IC0620620 and IC0620622 respectively (Table 12). This included adapter/primer sequences, low-quality reads and an empty/blank reads. After rigorous data filtering and quality check, a total of 191,233,242 (*i.e.* 97.7%) high-quality clean reads were recovered from the total raw reads.

Table 12: Summary of data generated for *M. pruriens* transcriptome (sequencing and quality filtering).

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
G1 Total	**15.65**	**167,956,452**	**164,095,649**
G2 Pooled	2.59	27,801,324	27,137,593
Summary	**18.24**	**195,757,776**	**191,233,242**

G1: Genotype 500108KA - IC0620620; **G2**: Genotype 500113MH - IC0620622.

These reads were used to generate transcriptome assembly using Trinity assembler. The combined assembly - 620-22, generated using all the clean reads produced 72,561 transcripts that included 46,525,999 bases with an average GC content of 44.58%. The mean values of four individual nucleotides - Adenine, Guanine, Thymine,

Cytosine were recorded as 27.35%, 23.31%, 26.98%, and 22.36% respectively, with an N50 length of 987 bp. The Limin et al. (2012) developed software called CD-HIT v. 4.6 was used to generate non-redundant transcripts and cluster highly similar fragments. After clustering, totally 67,561 transcripts were retained with an average transcript length of 641 bp (Figure 15). The longest transcript contained 17,978 bases with a total 1493 putative non-coding sequences. The detailed assembly statistics of the combined assembly 620-622 are given in Table 13. In addition, we generated another assembly exclusively from the reads of the genotype IC0620620 for aligning against 620-622 assembly during lobSTR analysis. The comparision between the two assemblies used for lobSTR analysis is given in Table 14.

Table 13: Statistics of non-redundant set of *M. pruriens* transcripts obtained from Trinity assembly

Total number of assembled bases	46,525,999
Number of transcripts	72,561
The total number of sequences after clustering	67,561
The mean sequence length	626
Average % of Adenine	27.35
Average % of Guanine	23.31
Average % of Thymine	26.98
Average % of Cytosine	22.36
Average % of N	0.00
Average % of GC content	44.58
N50	987
Maximum transcript length (bp)	17,978
Average transcript length (bp)	641
Number of putative non-coding sequences	1,493
Length of longest ORF (bp)	2,362
Number of ORF≥100 bp	36,228
Number of ORFs on (+) strand	36,421
Number of ORFs minus (-) strand	31,140

Finally, we deposited all the high-quality paired-end reads as well as assembly in the U.S. National Center for Biotechnology Information (NCBI) sequence read archive (SRA) database: SRR3453110.

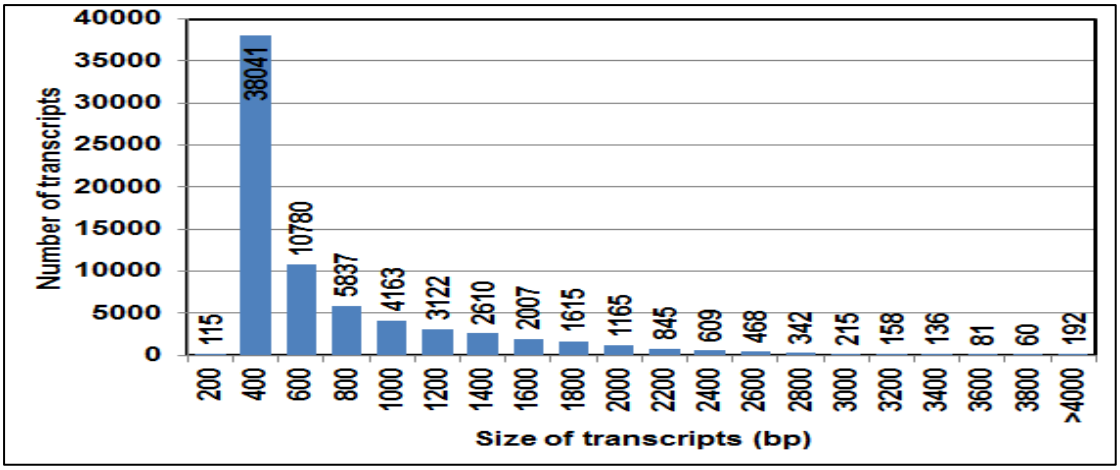


Figure 15: Length distribution of *M. pruriens* transcripts in Trinity assembly

Table 14: Statistics of the transcriptome assemblies generated using RNA-Seq data

Assembly	Total number of assembled bases	No. of transcripts	No. of transcripts after clustering	Average contig size	N50	% GC
IC0620620-622	46,525,999	72,561	67,561	641	987	44.57
IC0620620	51,600,321	51,385	43,768	1,004	1,571	42.55

5.2. Discovery of genic microsatellite markers and validation

5.2.1. SSR identification

We detected a total of 6284 transcripts [*Additional file 6: Transcripts with the repeats in the Mucuna assembly identified using Perl script MISA; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects /12864_2017_3780_MOESM6_ESM.xlsx;* Sathyanarayana et al. 2017; BMC Genomics] within which 7943 potential EST-SSRs [*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 7943 repeat sequences is also reported; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM7_ESM.xlsx*; Sathyanarayana et al. 2017; BMC Genomics] were discovered.

Of these, the mono-nucleotide repeats represented the largest fraction (3638) with the majority (92%) showing A or T repeats (Figure 16). Additionally, about 1674 di-nucleotide, 2240 tri-nucleotide and only a small fraction of tetra- (146), penta- (64) and hexa- (100) nucleotide SSRs were identified in *M. pruriens* transcripts (Table 15). Of the 6284 SSR-containing transcripts, 1174 transcripts contained more than one SSRs. Further, EST-SSRs containing five tandem repeats were most abundant (18.71%), followed by ten (17.43%), six (15.45%), eleven (9.31%), seven (7.51%), and twelve (5.85%) repeats, whereas the remaining tandem repeats, each accounted for less than 5% of our EST-SSRs (Table 16).

Table 15: Statistics of SSRs identified in *M. pruriens* transcripts

SSR 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 5.3kb
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%)

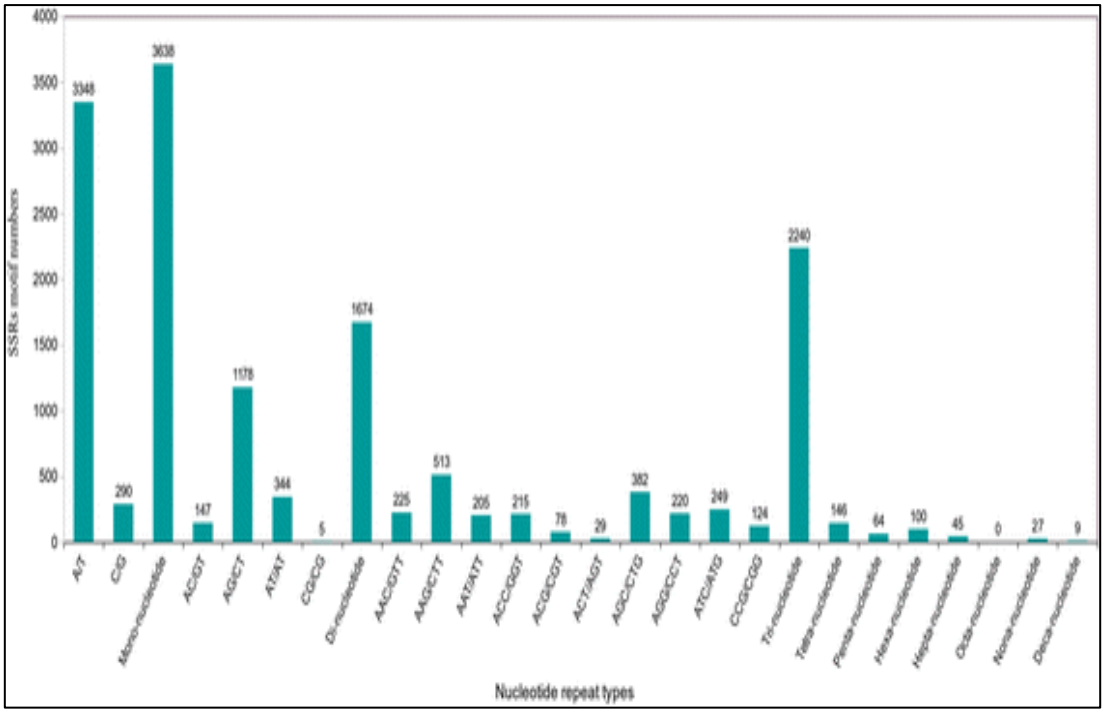


Figure 16: Simple sequence repeats length distribution across different motif classes in *M. pruriens* transcriptome

Table 16: Length distribution of the EST-SSRs based on the number of nucleotide repeats

Repeat Number ⇩	**⇦ Repeat Type ⇨**							
	Mono-	**Di-**	**Tri-**	**Quad-**	**Penta-**	**Hexa-**	**Total***	**Percentage**
5			1269	96	47	59	1471	18.71
6		611	537	36	11	20	1215	15.45
7		329	240	7	3	12	591	7.51
8		260	124	2	2	2	390	4.90
9		163	20	2	0	0	185	2.35
10	1238	108	18	2	0	5	1,371	17.43
11	653	65	13	0	0	1	732	9.31
12	393	58	9	0	0	0	460	5.85
≥13	1354	80	10	1	1	1	1447	18.40
Total	**3638**	**1674**	**2240**	**146**	**64**	**100**	**7862**	
Percentage	45.80	21.07	28.02	1.83	0.80	1.25		

*Data are given only up to hexa- nucleotide.

5.2.2. Detection of polymorphic SSR using lobSTR program

Further, a specialized program lobSTR was employed for identifying the polymorphic SSRs between the contrasting parents. The lobSTR algorithm, with the input of raw sequencing data in FASTA/FASTQ or BAM format, reports the presence of alleles at each profiled STR locus in a custom tab-delimited text format (Gymrek et al. 2012; Cantarella and D'Agostino 2015). The rapidity/speed of the lobSTR and its consistency has surpassed the performance of conventional algorithms hitherto used for STR profiling (Gymrek et al. 2012). Another advantage of lobSTR over the conventional MISA pipeline is, during the alignment process, it simultaneously compares two or more samples and provides output in terms of polymorphic SSRs (pSSR).

This screening for pSSRs in our work recognized a total of 3865 polymorphic SSRs between the parents [*Additional file 8: Polymorphic sequence repeats identified between the Mucuna genotypes using lobSTR program; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM8_ESM.xlsx*; Sathyanarayana et al. 2017; BMC Genomics]. Further, aligning IC0620622 reads against the combined assembly generated 3075 SSR calls of which 1339 were in areas with > 5x coverage and a mean coverage of 9.54x (Table 17).

Table 17: Allelotype statistics obtained from lobSTR program

Sample	**Number of SSR calls**	**Num calls >=5x**	**Mean coverage**
Combined Assembly (IC0620620-622)	3075	1339	9.54
G1 (IC0620620) aligned to Ref1	3517	2092	19.15

Alternatively, alignment of IC0620620 reads (G1; 500108KA) to the combined assembly yielded 3517 SSR calls of which 2092 were in areas with >5x coverage and with a mean coverage of 19.15x. After filtering the SSRs based on the parameters explained earlier, we obtained a total of 787 polymorphic repeats [*Additional file 8: Polymorphic sequence repeats identified between the Mucuna genotypes using lobSTR program; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM8_ ESM.xlsx*; Sathyanarayana et al. 2017; BMC Genomics]. The results from the lobSTR analysis are given in Table 18 and the distribution of different SSR motifs from the lobSTR output is presented in Figure 17.

Table 18: Summary of the repeats in the alignment of *M. pruriens* transcriptome based on lobSTR

SSR Type	Repeat Motif	Repeat Size (bp)	Number of SSRs	Total SSRs (%)
Di-nucleotide	AG	2	166	4.45
	AT	2	27	0.72
	AC	2	25	0.67
Tri-nucleotide	AAG	3	207	5.55
	ATC	3	100	2.68
	AGG	3	86	2.31
	AAC	3	84	2.25
	AGC	3	79	2.12
	ACC	3	61	1.64
	CCG	3	58	1.55
	AAT	3	42	1.13
	ACG	3	23	0.62
	ACT	3	19	0.51
Tetra-nucleotide	AAAG	4	52	1.39
	AAAT	4	29	0.78
	AAAC	4	28	0.75
	AATG	4	17	0.46
	ACTC	4	17	0.46
	ACAT	4	16	0.43
	AGGG	4	12	0.32

	AAGG	4	12	0.32
	AACC	4	10	0.27
	AGAT	4	9	0.24
	AATC	4	9	0.24
	AATT	4	8	0.21
	AAGC	4	7	0.19
	ACCC	4	5	0.13
	AGCT	4	5	0.13
	ACGC	4	4	0.11
	AACT	4	4	0.11
	ATGC	4	4	0.11
	ATCC	4	3	0.08
	AGCG	4	2	0.05
	AGGC	4	2	0.05
	ATCG	4	1	0.03
	ACGT	4	1	0.03
	ACTG	4	1	0.03
	ACCG	4	1	0.03
	ACCT	4	1	0.03
	CCCG	4	1	0.03
Penta-nucleotide	Total	5	606	16.25
Hexa-nucleotide	Total	6	1886	50.56

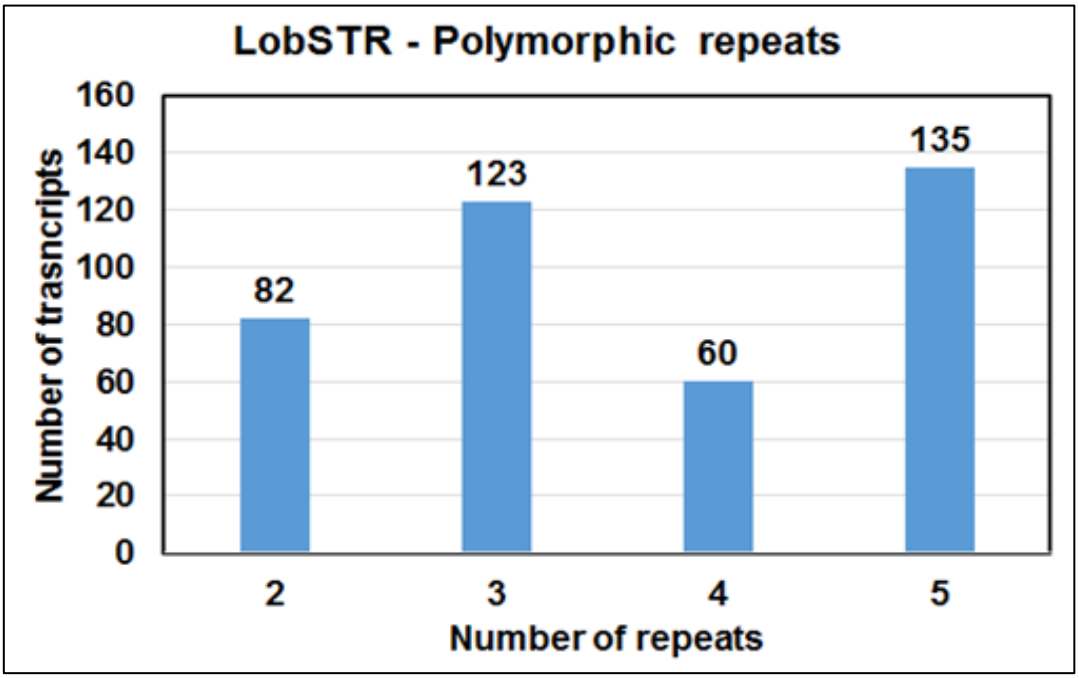


Figure 17: Repeat distribution in *M. pruriens* transcriptome discovered using lobSTR program

5.2.3. Validation of EST-SSR markers

Of the 134 primer pairs shortlisted, 98 (73.13%) produced clear amplification on a panel of 25 *M. pruriens* accessions chosen for validation. From this, consistently amplified marker-bands from 82 primer pairs were selected for analysis. Roughly 2000 marker-bands were produced by these 82 primer-pairs and 125 high-quality marker-bands denoting 52 primer pairs (Table 9) for 23 accessions (Table 8) were retained after quality filtering. Accessions 500101KA and 500267NL were removed due to >20% missing marker data. Different measures of genetic diversity for each primer-pair are reported in Table 19.

We analysed parameters of genetic diversity between population groups representing different botanical varieties (3) and geographical locations (3). The genetic diversity index (h) between the geographical locations and botanical varieties were in the range of 0.35 to 0.37 and 0.34 to 0.36 with mean values at 0.35 and 0.35 respectively (Table 20), where h was significantly lower for genetic subgroups, which ranged from 0.16 to 0.21. In all the groups, the total gene diversity (Ht) was greater than gene diversity within the groups (Hs). The coefficient of gene differentiation (Gst) was 0.04. The higher Gst value suggested less genetic differentiation among the population groups vis-à-vis within group variations. Gene flow indices (Nm) were high and was found in the range of 1.83 to 4.09.

Table 19: Polymorphism information of 52 EST-SSR markers for 23 *M. pruriens* accessions

S. No.	Primer ID	Motif	Alleles	Na	Ne	I	Ho	He	h	PIC
1	MPU_06	AGG	3	2	1.32	0.41	0.1	0.25	0.23	0.19
2	MPU_07	ACA	2	3	1.38	0.5	0.04	0.27	0.29	0.23
3	MPU_15	TC	3	5	2.5	1.16	0.05	0.6	0.26	0.27
4	MPU_16	ACA	4	4	2.32	1.1	0.24	0.57	0.2	0.24
5	MPU_18	GAA	2	2	1.54	0.54	0	0.35	0.35	0.35
6	MPU_21	GAA	1	3	1.2	0.37	0	0.17	0.28	0.15
7	MPU_22	TC	3	4	3.04	1.17	0.09	0.67	0.28	0.33
8	MPU_23	CCT	2	3	1.31	0.47	0	0.23	0.25	0.17
9	MPU_24	ATA	2	3	2.1	0.85	0.04	0.52	0.38	0.35
10	MPU_27	AGA	1	5	1.37	0.6	0.09	0.27	0	0.15
11	MPU_30	AT	2	4	1.77	0.8	0.04	0.44	0.35	0.3
12	MPU_31	CAT	3	6	2.5	1.24	0.24	0.6	0.26	0.28
13	MPU_33	GCT	2	2	1.29	0.39	0	0.23	0.27	0.2
14	MPU_37	GTT	2	2	1.63	0.57	0	0.39	0.37	0.31
15	MPU_39	CA	2	5	1.92	0.96	0	0.48	0.37	0.29
16	MPU_42	AG	6	6	5.56	1.75	0.09	0.82	0.15	0.22
17	MPU_43	ACT	2	3	1.33	0.47	0.05	0.25	0.05	0.08
18	MPU_45	GA	3	3	2.14	0.91	0.04	0.53	0.28	0.28

19	MPU_47	TG	2	2	1.51	0.52	0.05	0.34	0.35	0.28
20	MPU_48	GAG	4	4	2.33	1.01	0.22	0.57	0.22	0.27
21	MPU_49	TG	2	5	2.85	1.23	0	0.65	0.36	0.36
22	MPU_52	GGT	3	6	2.15	1.08	0.33	0.53	0.22	0.25
23	MPU_57	GCG	2	5	1.6	0.8	0.17	0.37	0.29	0.23
24	MPU_58	CAA	2	2	1.29	0.39	0	0.23	0.24	0.2
25	MPU_59	AGA	3	3	1.36	0.49	0.13	0.26	0.2	0.2
26	MPU_64	AT	3	3	2.47	1	0.05	0.6	0.28	0.31
27	MPU_65	CAT	4	4	2.93	1.19	0.4	0.66	0.2	0.26
28	MPU_68	AG	3	4	1.72	0.74	0.18	0.42	0.26	0.27
29	MPU_72	TAG	2	2	1.57	0.55	0.04	0.36	0.33	0.3
30	MPU_78	CAA	2	4	2.24	0.94	0.1	0.55	0.37	0.36
31	MPU_80	GA	1	4	1.22	0.41	0.09	0.18	0.48	0.21
32	MPU_81	GA	2	3	1.36	0.52	0	0.26	0.27	0.19
33	MPU_83	CTT	1	4	1.2	0.4	0.09	0.17	0	0.21
34	MPU_88	TC	2	4	1.52	0.65	0.09	0.34	0.31	0.25
35	MPU_89	TGGGTT	2	3	2.06	0.78	0.09	0.52	0.23	0.21
36	MPU_90	GGT	4	5	4.03	1.45	0.46	0.75	0.16	0.26
37	MPU_92	GGA	3	4	2.4	1.04	0.35	0.58	0.24	0.3
38	MPU_98	AAC	2	3	1.56	0.66	0.05	0.36	0.31	0.26

39	MPU_99	TC	3	4	1.61	0.75	0.09	0.38	0.24	0.26
40	MPU_102	CT	1	6	1.44	0.7	0.13	0.31	0.5	0.28
41	MPU_106	AGA	2	3	1.4	0.53	0.04	0.28	0.27	0.2
42	MPU_107	CTGC	3	6	4.1	1.54	0.25	0.76	0.19	0.26
43	MPU_111	GCC	5	5	3.1	1.32	0.32	0.68	0.17	0.24
44	MPU_114	TTC	2	2	1.31	0.4	0	0.24	0.24	0.21
45	MPU_115	AG	3	4	2	0.95	0.05	0.5	0.23	0.22
46	MPU_119	AGA	1	2	1.1	0.2	0	0.1	0	0.09
47	MPU_121	CAC	2	4	1.44	0.64	0	0.31	0.27	0.2
48	MPU_122	CAG	1	3	1.14	0.28	0.04	0.12	0	0.15
49	MPU_123	AAT	1	4	1.37	0.57	0.04	0.27	0.5	0.25
50	MPU_124	GCA	4	6	5.01	1.7	0.04	0.8	0.2	0.27
51	MPU_125	AGA	1	4	1.31	0.53	0	0.24	0.42	0.2
52	MPU_126	GTG	2	2	1.29	0.39	0.17	0.23	0.28	0.26
	Average	-	**2.4**	**3.73**	**1.98**	**0.78**	**0.1**	**0.41**	**0.26**	**0.24**
	SD	-	**1.07**	**1.25**	**0.96**	**0.37**	**0.11**	**0.19**	**0.11**	**0.06**

Na: Number of alleles; **Ne**: Effective number of alleles; **I**: Shannon information content; **Ho**: Observed heterozygosity; **He**: Expected heterozygosity (Kimura and Crow 1964); **h**: Nei's (1973) gene diversity; **PIC**: Polymorphism information content.

Table 20: Gene diversity estimates for groups based on botanical varieties, geographical distribution and population structure analysis

Geographical distribution	**Population group**	**Na**	**Ne**	**I**	**h**
	East India	2.23	1.78	0.59	0.37
	Northeast 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 verities	**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	**Subgroup 1**	2.54	1.91	0.66	0.19
	Subgroup 2	1.87	1.53	0.41	0.16
	Subgroup 3	2.36	1.77	0.59	0.21
	Subgroup 4	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		

Note: Na = Number of different alleles; **Ne** = Effective no. of alleles; **I** = Shannon's Information index; **h** = Nei's (1973) gene diversity; **Ht** = Diversity in overall collections total gene diversity; **Hs** = Sub divided population; **Gst** = Genetic differentiation; **Nm** = Estimate of gene flow from Fst, (Nm= (1/Fst)-1)/4; Fst = Genetic differentiation among populations).

Population structure analysis on 23 genotypes using fastSTRUCTURE revealed the presence of four to six subgroups based on "choosing model complexity" script, and four or eight subgroups using likelihood score (Figure 18), with K = 4 being the most probable number. For K = 4 subgroups, 21 genotypes showed >80% membership to a particular subgroup as reported by the coefficient of coancestry value for each genotype (Figure 19A).

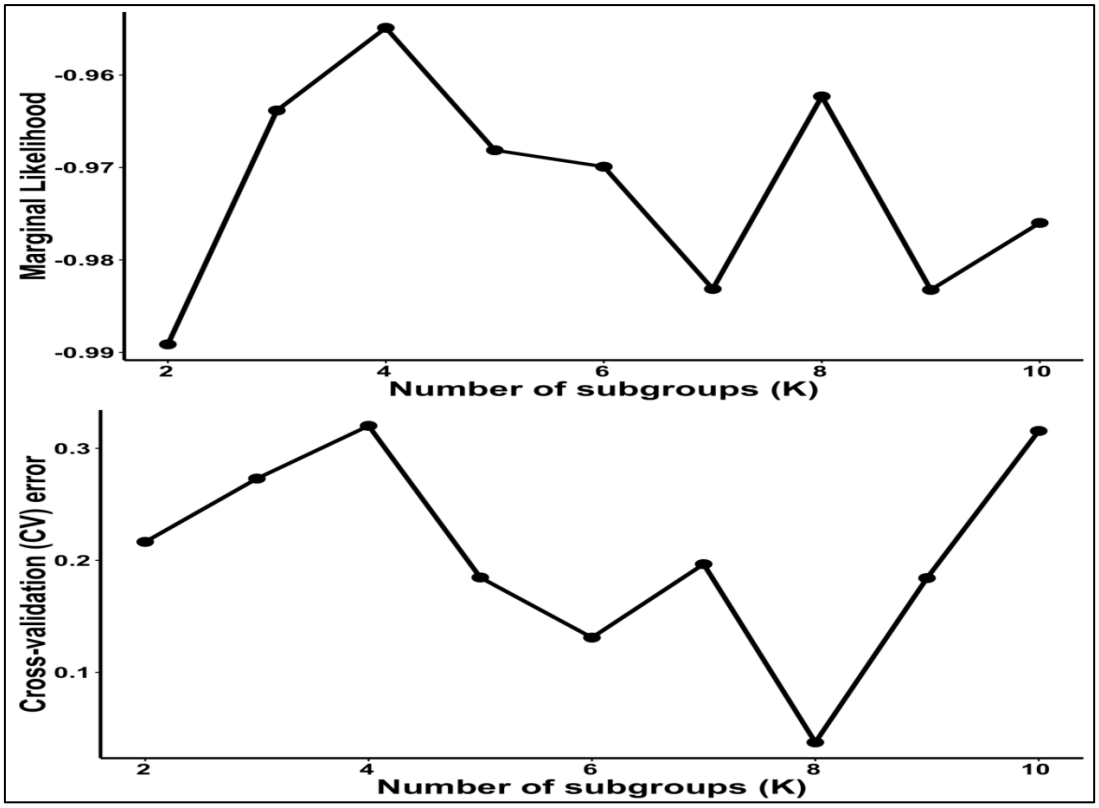


Figure 18: Results of fastSTRUCTURE analysis across K=2 to K=10 sub-groups. **(A)** Subgroup number K plotted against marginal likelihood and **(B)** cross validation error

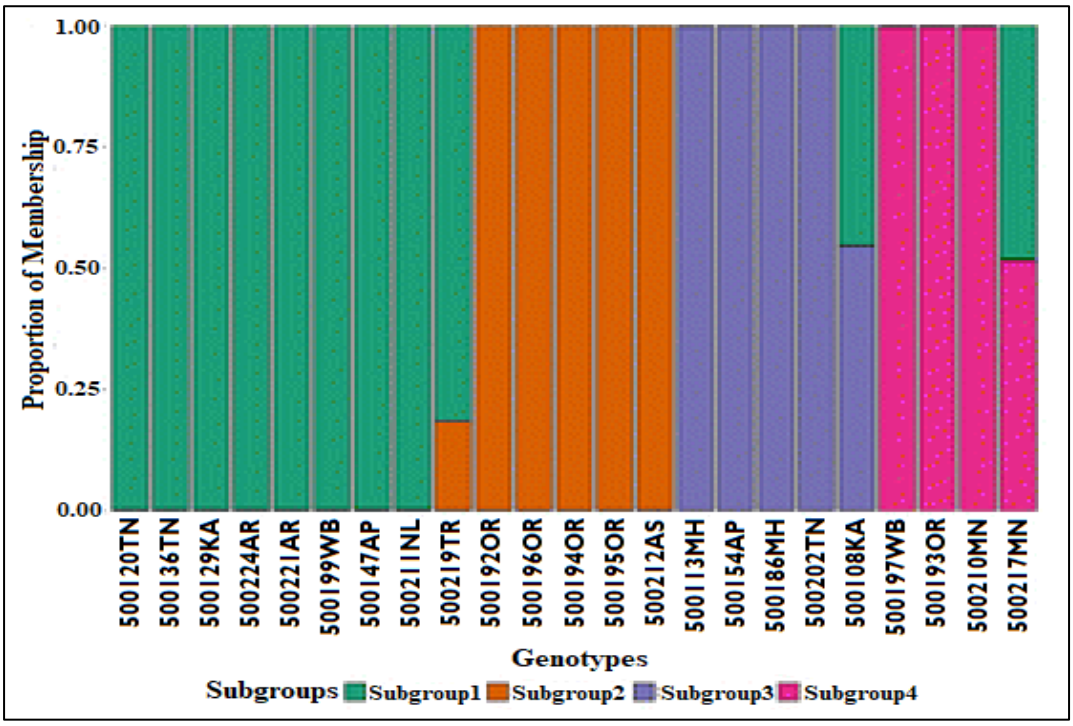


Figure 19A: Population Structure analysis of the 23 Indian *M. pruriens* accessions. Bayesian clustering (fastSTRUCTURE, K = 4)

Subgroup 1 comprised nine individuals, mostly belonging to var. *pruriens* ($n = 5$) along with one var. *utilis* accession with mixed representation from the Northeast and Peninsular India (each $n = 4$) and one accession from East India. Subgroup 2 showed exclusively var. *pruriens* accessions from the Eastern ($n = 4$) and Northeast ($n = 1$) India. Similarly, subgroup 3 separated accessions belonging to Peninsular India with most accessions belonging to var. *hirsuta* ($n = 3$) and one each from the other two varieties. Subgroup 4 was the most heterogeneous group and showed the mixed representation of two var. *pruriens* accessions from East India and two var. *utilis* accessions belonging to Northeast India.

The genotypes clustered in the neighbor-joining (NJ) tree and principal component analysis (PCA) are colour-coded from the information of four subgroups revealed in the fastSTRUCTURE analysis. Both NJ and PCA showed analogous clustering pattern as fastSTRUCTURE with three out of the four groups grouping in accordance with their geographical origin (Figures 19A-C). Principal component 1 (PC1; Figure 19B) accounted for 12.8% of the variability and separated the majority of Eastern and Northeast accessions from those of Maharashtra + Peninsular India accessions, with five Northeast or Eastern accessions falling on the left side of the plot. None of the Peninsular Indian accessions fell right of PC1. Also, all the var. *hirsuta* and var. *utilis* accessions, except one, were positioned to the left of PC1. Of the two individuals that could be interpreted as having a hybrid ancestry, one was 500217MN (var. *utilis*, Northeast India) as demonstrated by a ~50% split assignment to subgroup 1 and subgroup 4, the other being IC0620620 (var. *utilis*, Peninsular India) which was assigned to subgroups 1 and 3 with near equal possibility. Subgroup 2 was completely separated from subgroup 3 by PC1. Principal component 2 (PC2;

Figure 19B) accounted for the rest of 10.1% of genetic variability. Subgroup 1, the largest and most diverse subgroup, clustered entirely below PC2.

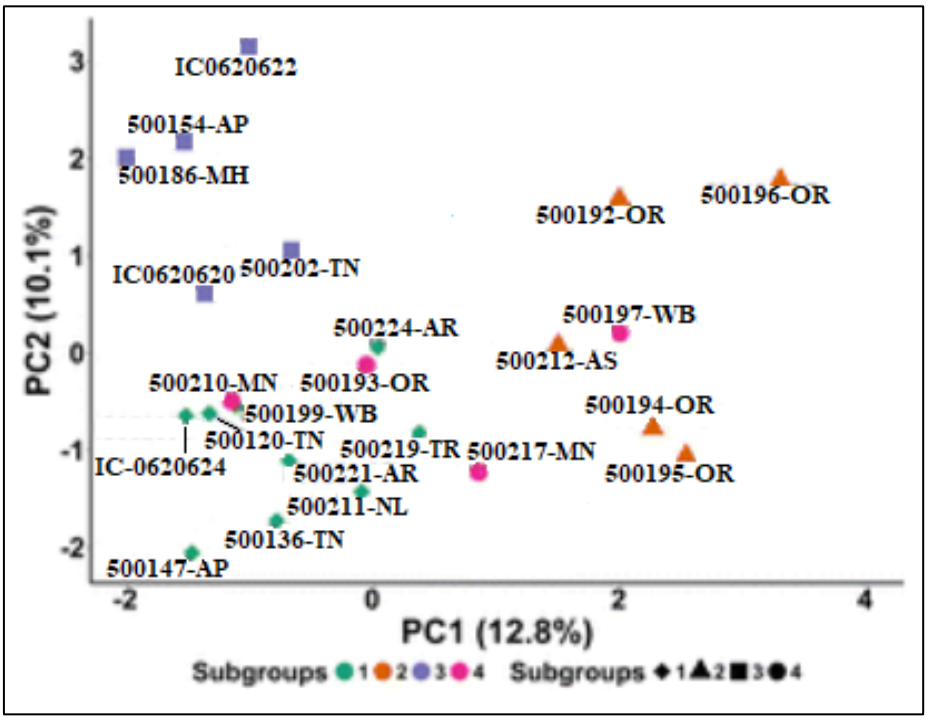


Figure 19B: Population Structure analysis of the 23 Indian *M. pruriens* accessions- Scatter plot from principal component analysis (PCA)

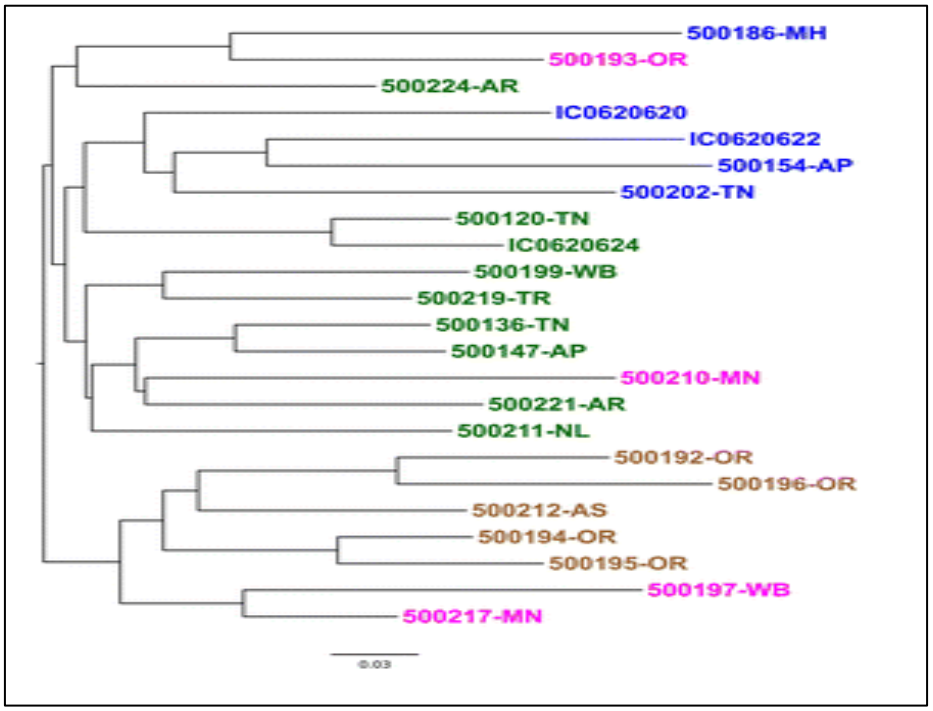


Figure 19C: Population Structure analysis of the 23 Indian *M. pruriens* accessions - Neighbor-joining (NJ) tree

The NJ algorithm generated two main clades (Figure 19C), one showing similar grouping produced in PCA by PC1 which is also the same as subgroup 2 along with two accessions from subgroup 4. This clade included only var. *pruriens* belonging to the Northeast and Eastern India with the exception of the hybrid 500217MN (var. *utilis)*. The second clade included subgroups 1 and 3, along with two accessions of subgroup 2, with no clear-cut clustering pattern based on the variety.

5.3. Functional annotation of transcriptome

A total number of 49,925 (73.9%), 35,535 (52.5%) and 54,450 (80.5%) transcripts showed significant hits with NCBI-NR, Swiss-Prot and UniRef proteins, respectively, with 34,686 transcripts having conserved domains and 6248 with hits against the Rfam database. The details are given in Table 21.

Table 21: Number of *M. pruriens* unigene hits in different non - redundant databases

Database	Total Blast Hits
NCBI-NR	49,925
Swiss-Prot	35,535
UniRef	54,450
Conserved Domains	34,686
Rfam	6,248

Largely, the putative orthologs of genes having roles in different pathways as well as cellular processes were found to be much conserved in *M. pruriens*. Further, Gene Ontology (GO) terms were assigned to *M. pruriens* transcripts, which showed considerable similarity with annotated transcripts from other plant species [*Additional file 2* : *Number of transcripts assigned to Biological, Cellular and Metabolic processes from GO analysis of the annotated transcripts of the Mucuna assembly; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM2_ESM.xlsx*; Sathyanarayana et al. 2017,

BMC Genomics]. In our work, 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, proteins involved in transcription and (3.56%) and transcription regulations (3.50%) were highly expressed. The genes involved in other important biological processes such as signal transduction, carbohydrate metabolism, response to stress, transport etc. were also elucidated through GO annotations. Similarly, genes involved in ATP, DNA as well as different metal ion binding activities were most represented in various molecular functions; and nucleus and cytoplasm related functions were most abundant among the cellular component categories (Figure 20 A-C).

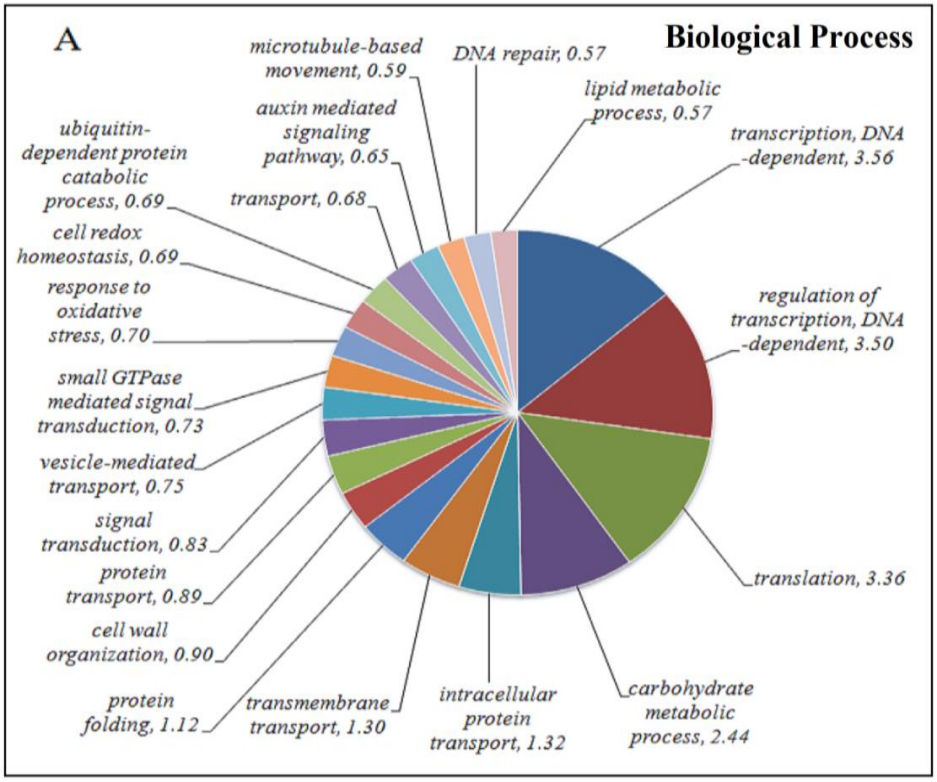


Figure 20A: Functional annotation of *M. pruriens* transcripts. GO term assignment to the *M. pruriens* transcripts in Biological process. (The numbers next to categories are the percentage of each category of biological function.)

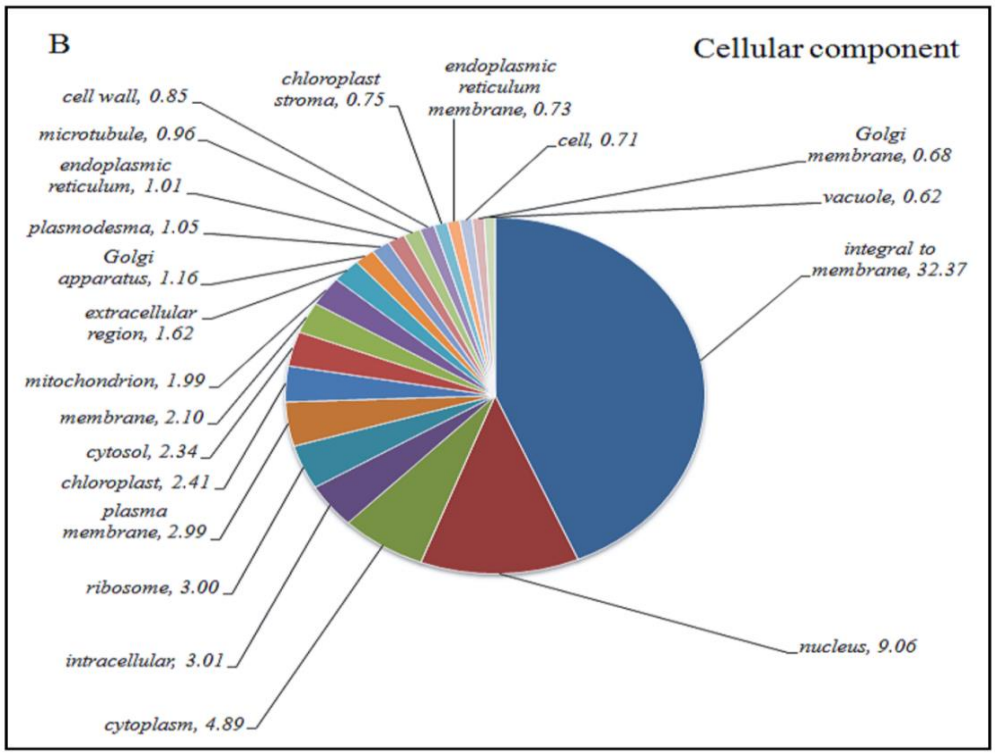


Figure 20B: Functional annotation of *M. pruriens* transcripts. GO term assignment to the *M. pruriens* transcripts in Cellular component. (The numbers next to categories are the percentage of each category of biological function.)

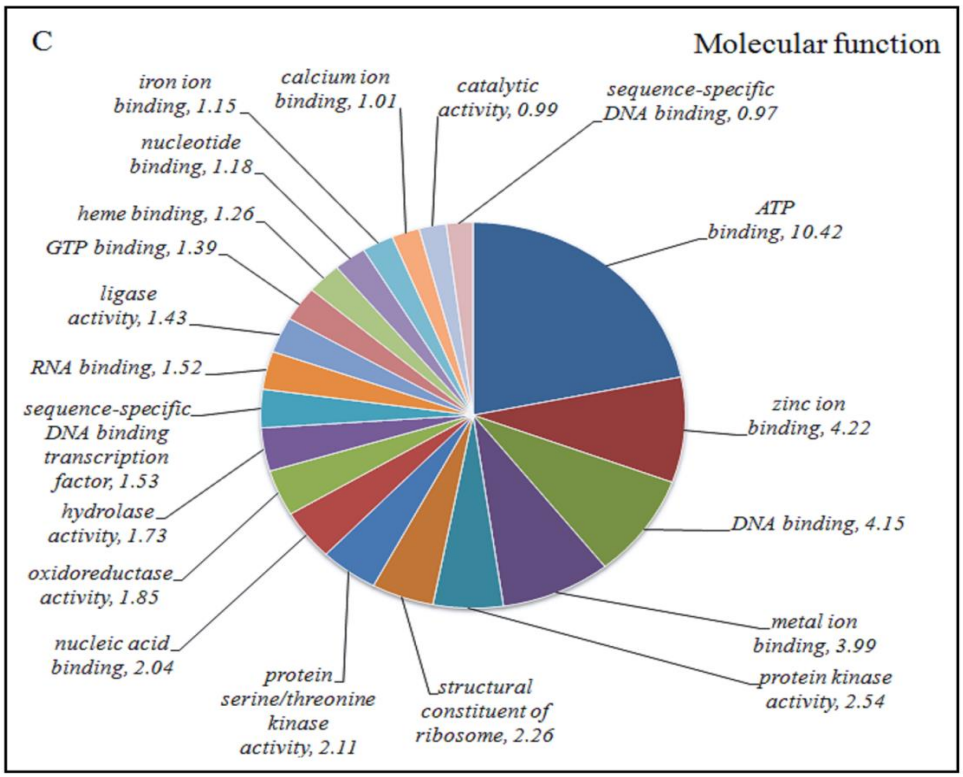


Figure 20C: Functional annotation of *M. pruriens* transcripts. GO term assignment to the *M. pruriens* transcripts in Molecular function. (The numbers next to categories are the percentage of each category of biological function.)

The transcripts were also annotated for Enzyme Classes (EC) and KEGG (Kyoto Encyclopedia of Genes and Genomes) classifications. EC classification was thus obtained for 3963 assembled sequences, whereas KEGG classification was obtained for 3492 assembled sequences (Table 22) (*Additional file 3: Number of transcripts assigned for enzyme classes and KEGG pathways of the M.pruriens transcripts; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM3_ESM.xlsx*; Sathyanarayana et al. 2017, BMC Genomics]. Figure 21A lists top 20 enzyme classes obtained for our annotated transcriptome. Remarkably, a large amount of these assembled transcripts was related to serine/threonine protein kinase enzyme class alone (38.4%). Further, Figure 21B shows top 20 KEGG pathways represented by our transcriptome. The highest number of sequences were related to pathways linked to protein modification (37.5%) followed by lipid metabolism (8.8%), glycan metabolism (6.9%), carbohydrate degradation (5.6%), cofactor biosynthesis (4.7%), glycan biosynthesis (2.8%), carbohydrate metabolism (2.7%), phospholipid metabolism (2.03%), secondary metabolite biosynthesis (1.94%), purine metabolism (1.89%) and so on. As can be seen from the results, the highest represented groups included pathways associated with the housekeeping processes as well as plant development and secondary metabolism.

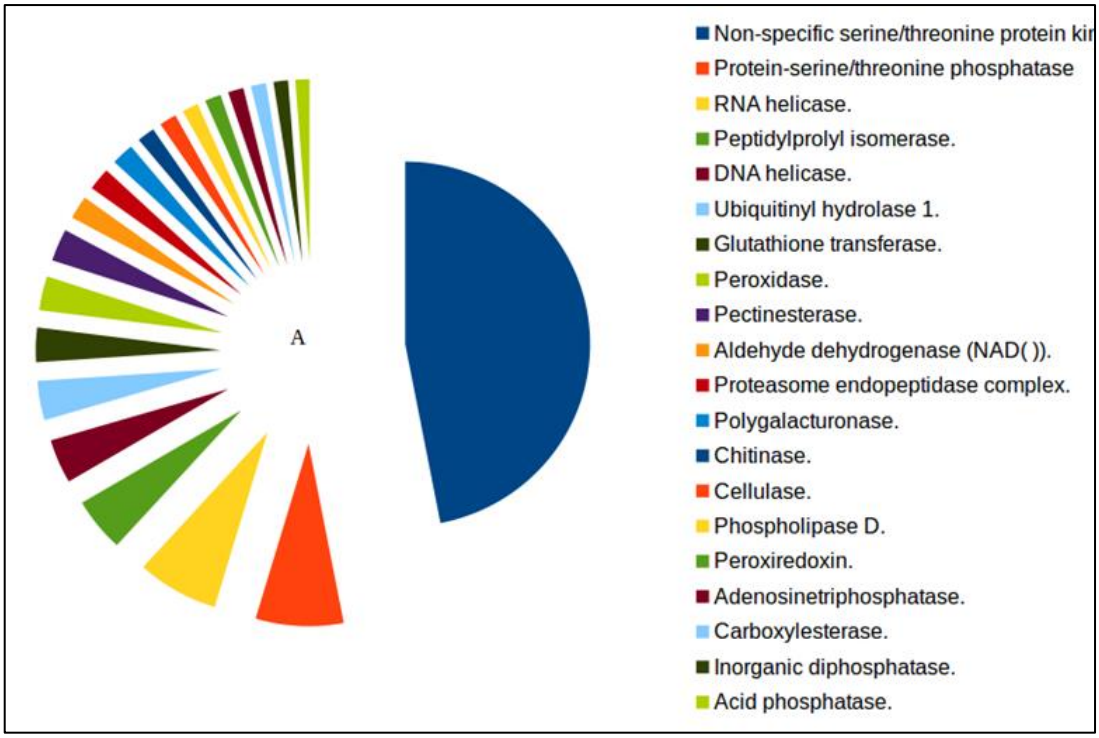


Figure 21A: Functional characterization and abundance of *M. pruriens* transcriptome for enzyme classes. Transcripts were classified in the top 20 abundant enzyme classes and KEGG pathways; area under each pie slice represents the value in percept.

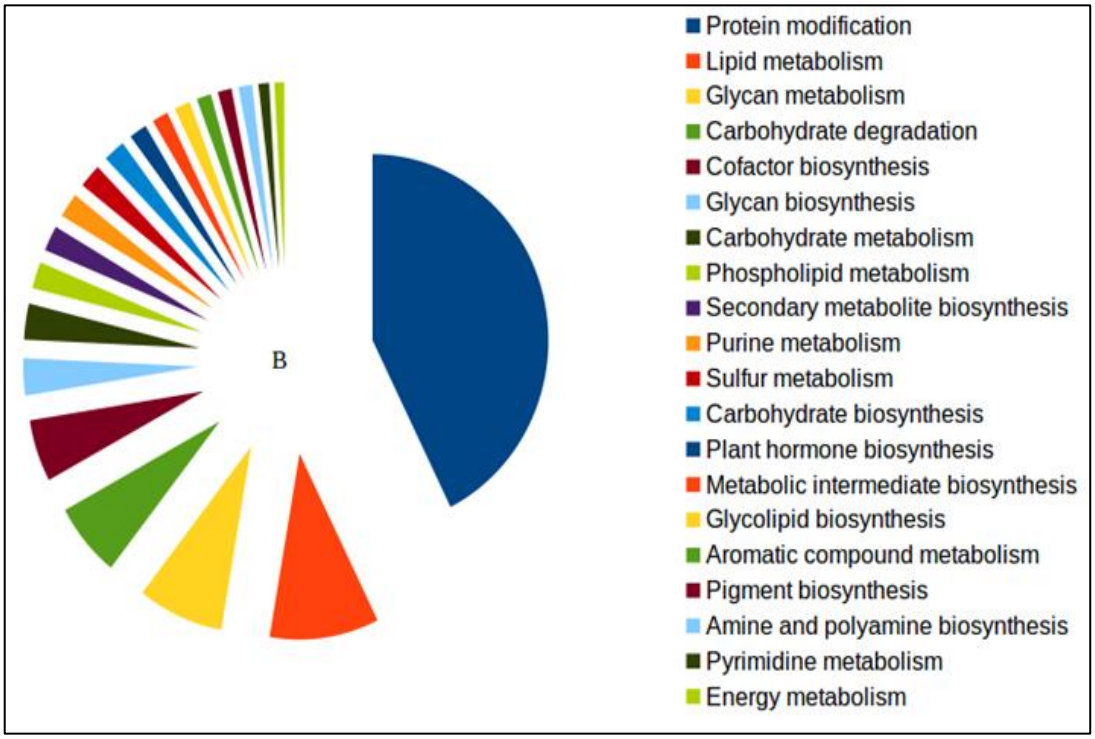


Figure 21B: Functional characterization and abundance of *M. pruriens* transcriptome for KEGG pathways - Transcripts were classified in the top 20 abundant enzyme classes and KEGG pathways; area under each pie slice represents the value in percept.

Table 22: Number of transcripts assigned for enzyme classes and KEGG pathways of the *M. pruriens* transcripts

KEGG Description	Occurrence	KEGG Description	Occurrence
Protein modification	1310	Plant hormone degradation	12
Lipid metabolism	308	Alkene biosynthesis	12
Glycan metabolism	242	Genetic information processing	11
Carbohydrate degradation	199	Protein biosynthesis	9
Cofactor biosynthesis	166	Isoprenoid biosynthesis	9
Glycan biosynthesis	99	Plant hormone metabolism	8
Carbohydrate metabolism	96	Membrane lipid metabolism	8
Phospholipid metabolism	71	Glucan metabolism	6
Secondary metabolite biosynthesis	68	Cofactor metabolism	6
Purine metabolism	66	Cofactor degradation	6
Sulfur metabolism	64	Metabolic intermediate metabolism	5
Carbohydrate biosynthesis	62	Cell wall biogenesis	5
Plant hormone biosynthesis	50	Amine and polyamine metabolism	5
Metabolic intermediate biosynthesis	46	Organosulfur biosynthesis	3
Glycolipid biosynthesis	45	Flavonoid metabolism	3
Aromatic compound metabolism	41	Carbohydrate acid metabolism	3
Pigment biosynthesis	39	Fermentation	3
Amine and polyamine biosynthesis	39	Alkaloid degradation	3
Pyrimidine metabolism	30	Xenobiotic degradation	2
Energy metabolism	28	Steroid hormone biosynthesis	2
Polyol metabolism	27	Siderophore biosynthesis	2
Glycerolipid metabolism	24	Organosulfur degradation	2
tRNA modification	23	Organic acid metabolism	2
Alkaloid biosynthesis	22	Mycotoxin biosynthesis	2
Secondary metabolite metabolism	21	Amine and polyamine degradation	2
Protein degradation	21	Thermoadapter biosynthesis	1

Photosynthesis	21	Steroid metabolism	1
Steroid biosynthesis	18	Sphingolipid metabolism	1
Carotenoid biosynthesis	17	Signal transduction	1
Phytoalexin biosynthesis	16	Quinol/quinone metabolism	1
Alcohol metabolism	16	Nucleoside biosynthesis	1
Phenylpropanoid metabolism	15	Ketone metabolism	1
Nitrogen metabolism	15	Ketone degradation	1
Terpene metabolism	13	Catecholamine biosynthesis	1
Glycan degradation	13	Alkene metabolism	1

5.4. Sequence similarity with other legume species

We performed a comparative study of *M. pruriens* assembled contigs with other related legumes. The results showed 58,208 of 67,561 transcripts (86.2%) from the combined 620–22 assembly had significant similarity to sequences in one or more legumes [*Additional file 4: BLASTP analysis of the transcripts against the legume protein databases; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM4_ESM.xlsx*; Sathyanarayana et al. 2017, BMC Genomics].

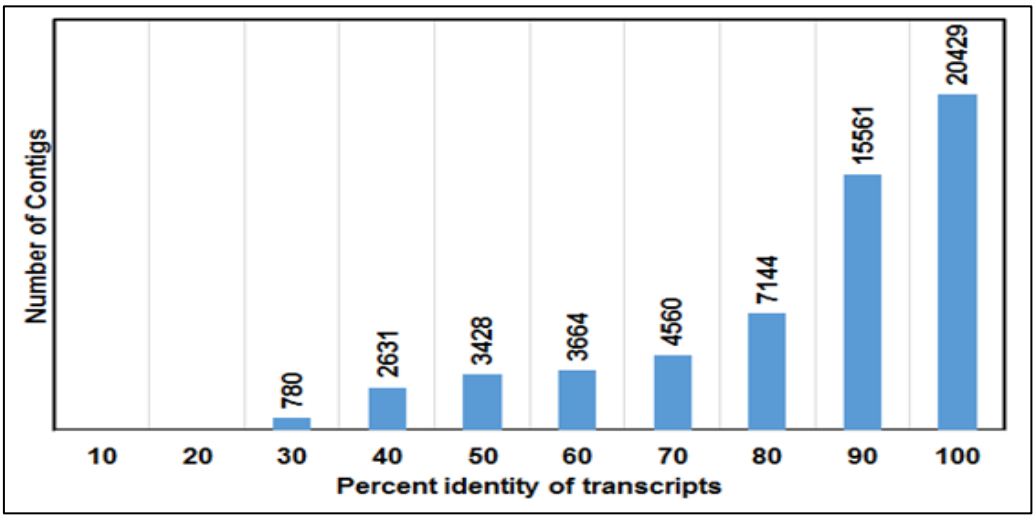


Figure 22: Legume sequence similarity analysis: Percentage identity of transcripts against other legume protein databases

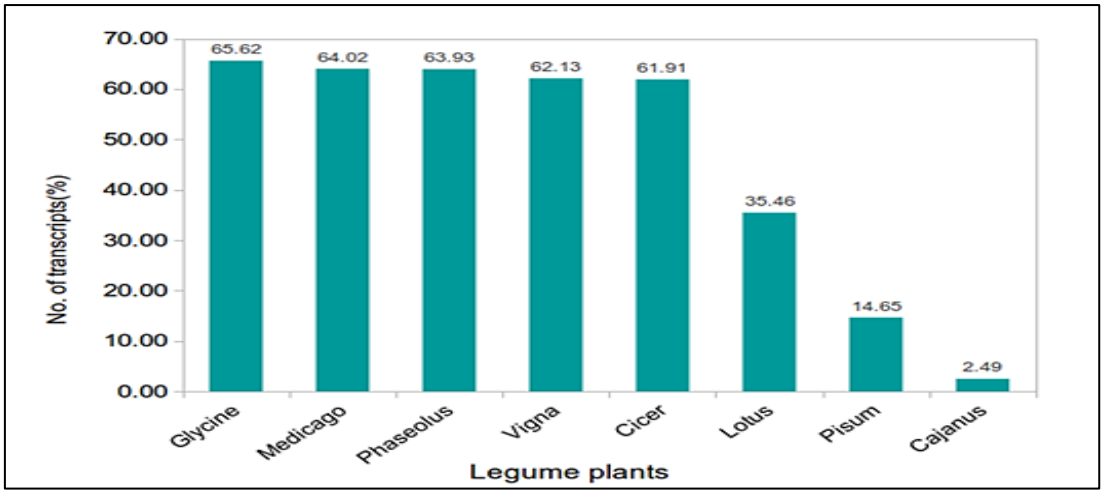


Figure 23: Legume sequence similarity analysis: Relative numbers of transcripts that had significant sequence similarity by species. The percentage of transcripts showing similarity value (E-valule≤1 E^{-05}) in BLASTX searches are shown

About 71% of these transcripts showed ≥ 70% sequence identity (Figure 22). The maximum number (65.62%) of *M. pruriens* transcripts showed significant similarity with *Glycine max* unigenes. This was followed by *Medicago truncatula* (64.02%), *Phaseolus vulgaris* (63.93%), *Vigna* spp. (62.13%) and *Cicer arietinum* (61.91%). These 5 legumes showed almost same amount of similarity shown in BLASTX searches against *M. pruriens*. The least similarity was observed with *Pisum* (14.65%) and *Cajanus* (2.49%). The detailed results are presented in Figure 23.

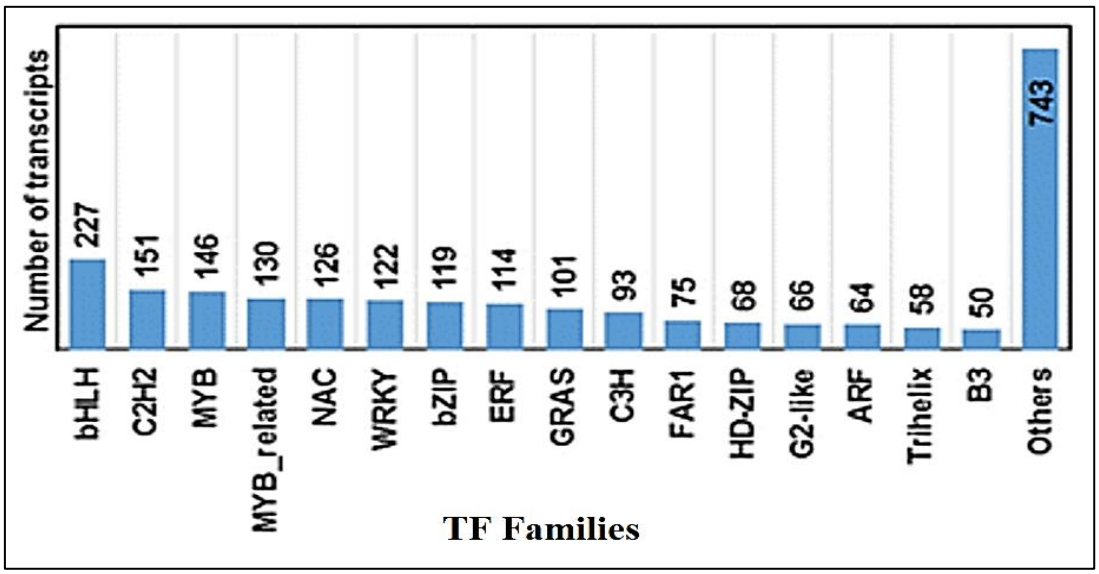


Figure 24: Distribution of *M. pruriens* transcripts in different transcription factor families

5.5. Mining transcription factor families

In our assembly, totally 2223 putative transcription factors distributed in at least 55 families were identified representing 3.29% of *M. pruriens* assembled transcripts [*Additional file 5: Transcription factors identified in the* Mucuna *assembly using PlnTFDB; link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM5_ESM.xlsx*; Sathyanarayana et al. 2017, BMC Genomics]. Among these, the top categories included the basic/helix-loop-helix (bHLH; 227), C2H2-type (151), MYB (146), MYB related (130), NAC (126) and WRKY (122) (Figure 24). Nearly all TF families revealed minor species-specific differences in relation to TF gene families reported for *Lotus, Medicago*, *Glycine* and *Cicer* (Table 23).

Table 23: Number of transcripts encoding for transcription factor families in *M. pruriens* compared to other legumes. The data on *M. pruriens* is from our study; data for *Glycine max*, *Medicago sativa* and *Lotus japonicus* is from Libault et al. (2009)*;* data for *Cicer arietinum* is from Garg et al. (2011).

*]***TF family**	***M. pruriens***	***Cicer arietinum***	***Glycine max***	***Medicago sativa***	***Lotus japonicus***
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

5.6. Expression analysis

The results of differential expression analysis (DEG) showed a total of 4387 differentially expressed transcripts among three tissues of IC0620620 with ≥4 log-fold expression change and significant p-value of 0.001. Among these, 1897 were expressed commonly in all three tissue types; 191 to 372 were shared among different tissue types and 25 to 1489 were unique to particular tissues (Figure 25A).

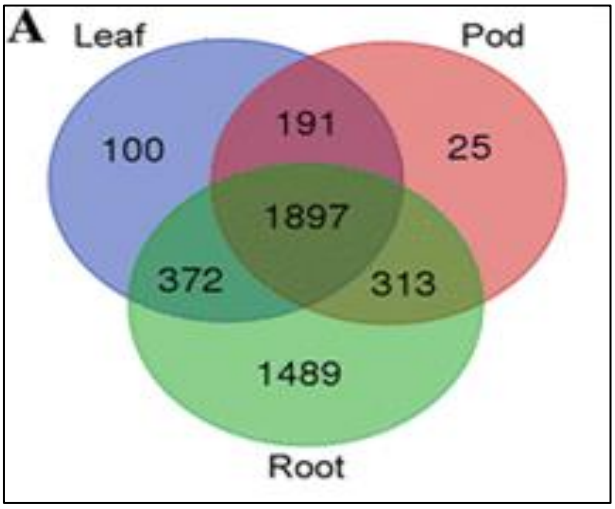


Figure 25A: Differential transcript expression in leaf, root, and pod tissues - diagram showing overlap of genes between leaf, root, and pod tissues showing differential transcript expression

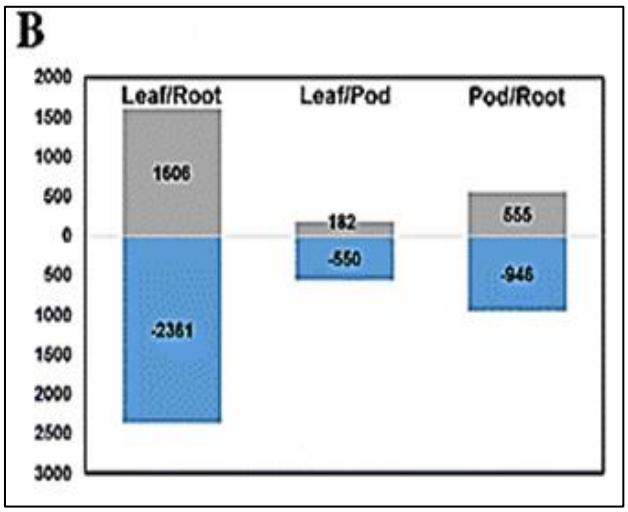


Figure 25B: Differential transcript expression in leaf, root, and pod tissues - pairwise comparisons across tissues showing differentially expressed transcripts. Those above the line are transcripts up-regulated and those below are down-regulated within the pairwise comparison

In case of leaf transcriptome, totally 1606 transcripts showed up-regulation, while 2361 exhibited down regulation as compared to roots, followed by 182 up-regulated and 550 down-regulated against the pod tissue. Likewise, pairwise comparison of pods and roots showed 555 and 946 transcripts as up-and down-regulated, respectively (Figure 25B). The top 50 differentially expressed transcripts in each of the three tissues showing varying expression patterns is shown in Figure 26.

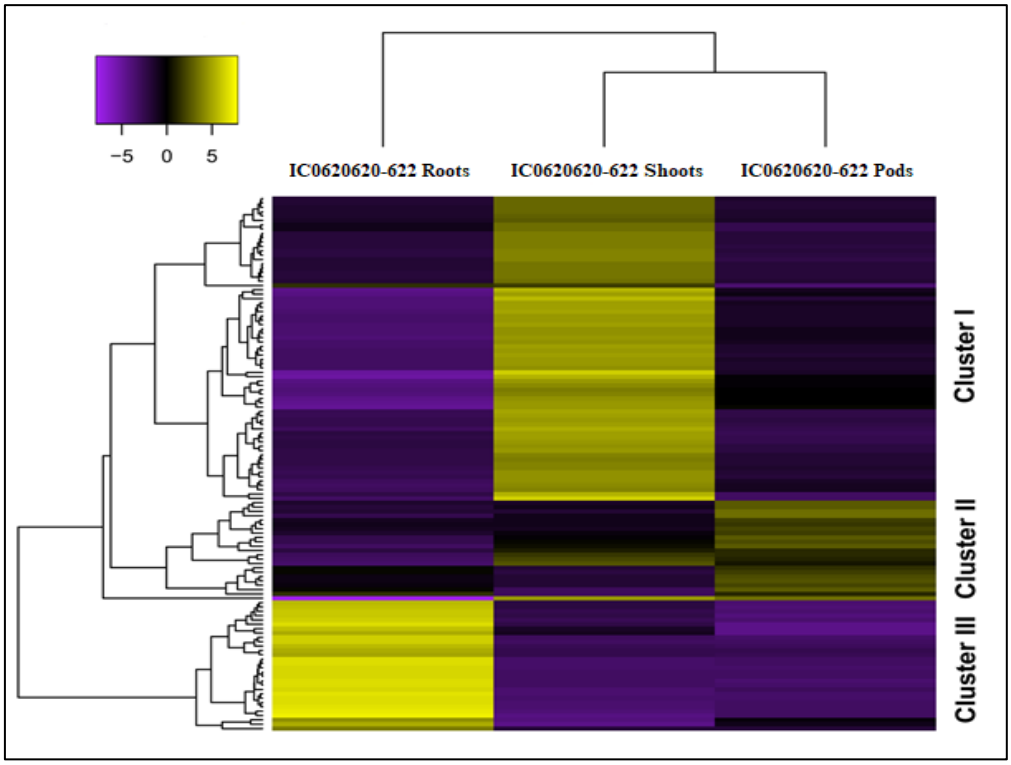


Figure 26: Heat map and complete linkage hierarchical clustering of differentially expressed transcripts of leaf, pod, root, and pooled transcriptomes. The various shades in the boxes showed similar tendencies of gene expression

From among the differentially expressed transcripts, we found 223 were related 43 different TF families such as MYB, MADS, WRKY, and bHLH some of whose members are linked to secondary metabolite pathways. Thus, we further investigated the expression of other genes implicated in secondary metabolism/biosynthesis. The results revealed top 47 transcripts in this category showed varying differential expression patterns (Figure 27). Highly expressed transcripts included those linked to

the flavonoid, phenylpropanoid, isoprenoid, and wax pathways [*Additional file 9: Fold change expression values of the transcripts in each of the tissue analysed related to the secondary metabolite pathways link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM9_ES M.xlsx*; Sathyanarayana et al. 2017, BMC Genomics].

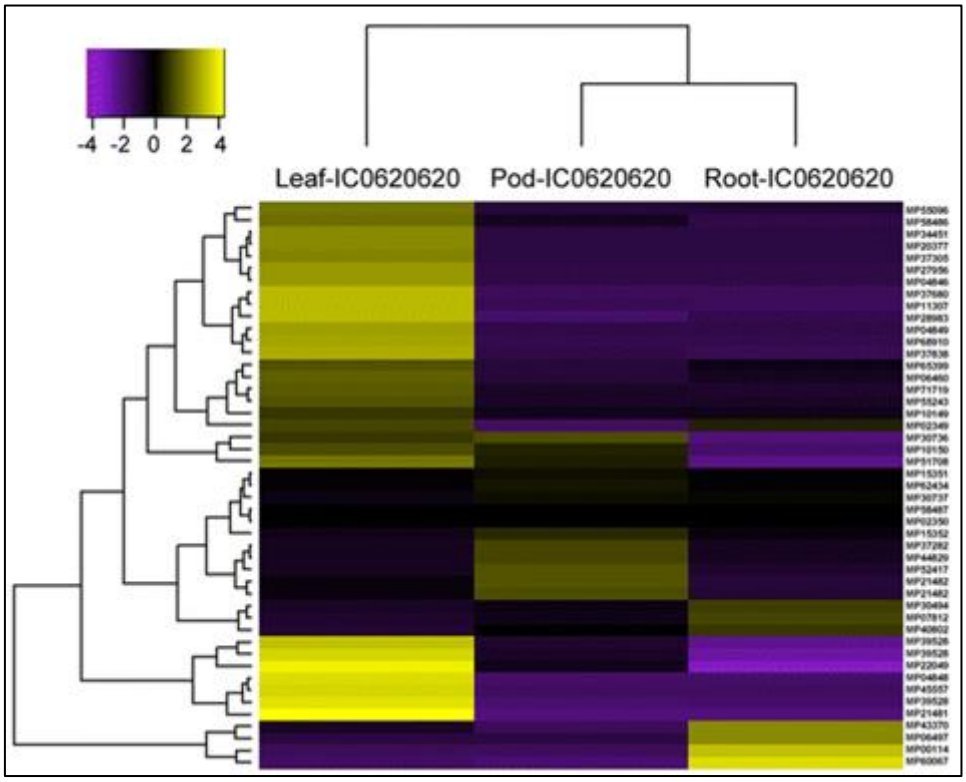


Figure 27: Differential transcript expression in leaf, root, and pod tissues - heat map of secondary metabolite associated differentially expressed genes of leaf, pod, and root transcriptomes. The various shades in the boxes showed similar tendencies of gene expression. Labels along the right side correspond to transcript names.

5.7. Microsatellite based genetic linkage map

Linkage analysis was performed using JoinMap 4.0 software. A minimum LOD score of 3.0 and maximum recombination fraction (θ) of 0.50 were set as thresholds for linkage group. Recombination fractions were converted into map distances in centimorgans using the Kosambi mapping function.

Using the above threshold, 32 SSR markers were mapped into 11 linkage groups (Figure 28) which covered a total distance of 837.04 cM in length. Linkage groups ranged from 4.49 cM to 186.48 cM in length with an average of 76.04 cM per linkage group. While most linkage groups consisted of only two to three linked markers, two linkage groups showed five linked markers. The largest linkage group spanned 186.48 cM (Figure 28).

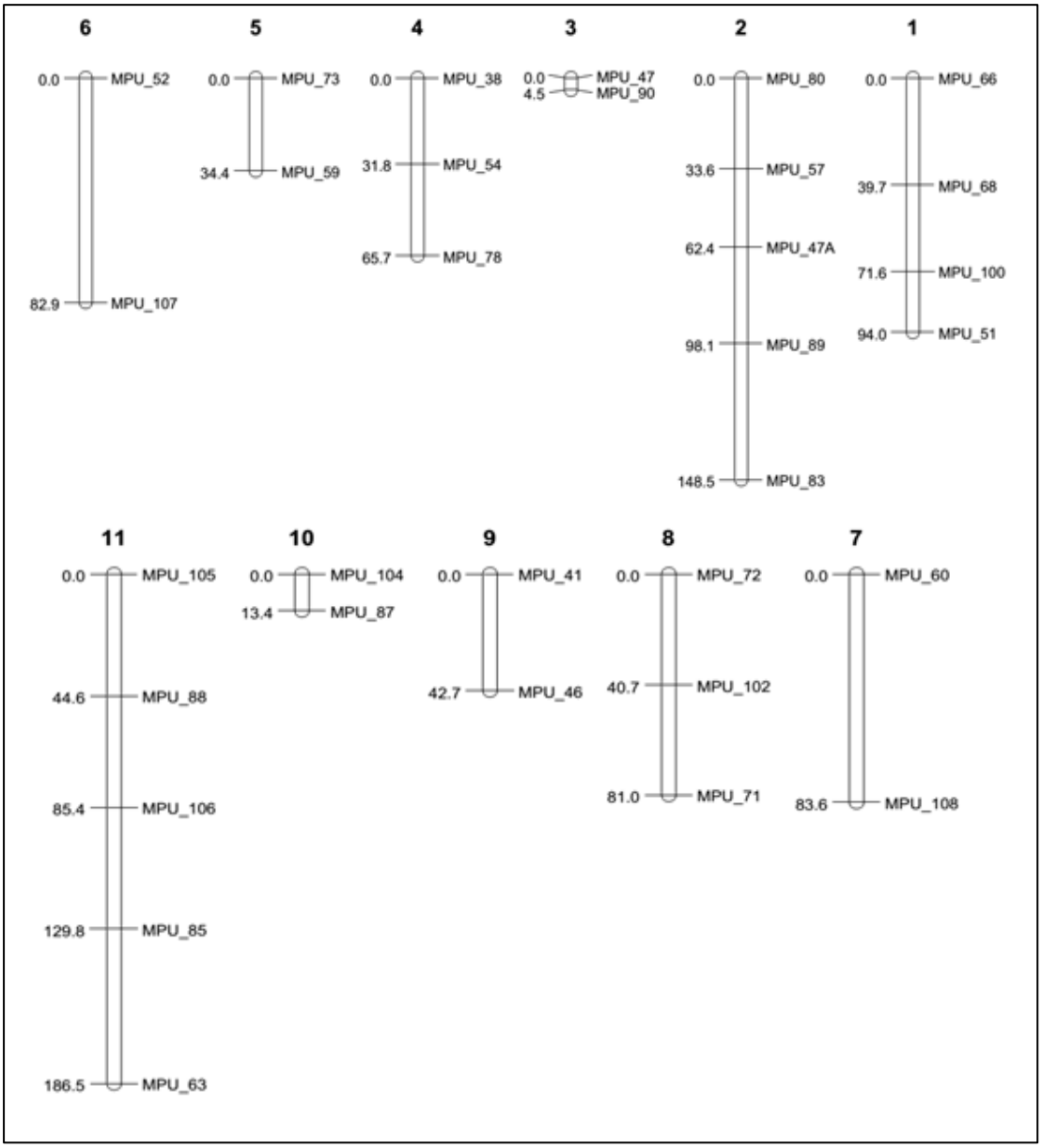


Figure 28: Thirty-two SSR markers incorporated into linkage groups constructed using JoinMap 4 (Van Ooijen 2006) and spanning 837 cM.

6. Discussion

Legumes are an important source of food, fodder, oil and fiber products for human beings (Phillips 1993). They are also the source of vitamins, protein, minerals, iron, calcium, magnesium, zinc in addition to omega-3 fatty acids (Pandey et al. 2016) and valuable pharmaceuticals (Sathyanarayana et al. 2016). Even though quite a few edible legumes are available on the market, their production rate, as compared to consumption requirement, has remained unmet and an ever-increasing demand has been witnessed (Ali and Kumar 2000). Also, changing food habit by a significant portion of the world population to a protein-rich vegetarian based diet has created unwarranted scarcity to plant-based protein resources (Bhat and Karim 2009).

There are several minor food legumes whose potential is underexploited. Bambara groundnut (*Vigna subterranea* L.), faba bean (*Vicia faba* L.), adzuki bean (*Vigna angularis*), velvet bean (*Mucuna* spp.), moth bean (*Vigna aconitifolia*), rice bean (*Vigna umbellata*), horse gram (*Macrotyloma uniflorum*), hyacinth bean (*Lablab purpureus* L.), grass pea (*Lathyrus sativus* L.), and winged bean (*Psophocarpus tetragonolobus* (L.) DC.), are important members of this grouping (Sathyanarayana et al. 2016). Many of them possess exceptional nutritional value and are an important source of protein in resource poor, food-deficit countries. Having adapted to marginal growing conditions, they also serve as a storehouse of important genes linked to biotic and abiotic stress tolerance (Bhat et al. 2008; Bhat and Karim 2009). However, almost all these crops suffer from poorly developed infrastructure, particularly in terms of genetic and genomic resources, thus limiting their application in the context of food security.

Medicinal legume *M. pruriens* (L.) DC. has fascinated the world as a source of the anti-Parkinson's medicine L-DOPA (Lloyd et al. 1975). It is also a well-known green

manure cover crop and offers significant agronomic benefits including high nitrogen fixation, soil nutrient improvement and protein content (Pugalenthi et al. 2005; Ceballos et al. 2012). This Ph.D. thesis describes an effort to develop genomic resources, particularly species-specific microsatellite markers for *M. pruriens* using RNA-Seq data and validate a sub-set of them on a panel of diverse genotypes. Furthermore, a preliminary attempt has been made to develop a frame-work genetic linkage map to act as a basis for future molecular breeding efforts in this legume species.

6.1. Developing genetic resources through transcriptomics

The Fabaceae is second only to the Poaceae in economic importance, with a large number of species utilized as food crops, industrial crops, fodder, medicines and construction materials (Wiersema et al. 2016). The genome sequencing of several legume species including Soybean (*Glycine max* (L.) Merr.; Schmutz et al. 2010), the Common bean (*Phaseolus vulgaris* L.) (Schmutz et al. 2014), Alfalfa (*Medicago sativa* L.) (Young et al. 2011), Lotus (*Lotus japonicus* L.) (Sato et al. 2008), Cowpea (*Vigna unguiculata* (L.) Walp.) (Muñoz-Amatriaín et al. 2017), Pigeon pea (*Cajanus cajan* (L.) Millsp.) (Varshney et al. 2012), and Lupin (*Lupinus angustifolius* L.) (Hane et al. 2017) have been completed, providing basic genomic resources across Fabaceae. However, a number of orphan crop species used as minor food or those used for applications other than human consumption are in need of efforts to generate genetic and genomic resources to act as intellectual capital upon which to build a larger scientific future. We present here the first *de novo* transcriptome assembly for *M. pruriens*, an orphan legume crop used both for human consumption and medicine, providing an important genetic resource base for future genetic studies and breeding efforts.

Within the phaseoloid legumes, *Mucuna* is a monophyletic genus (Moura et al. 2016) that represents an early-branching and distinct evolutionary lineage. It is allied, in many ways, with monotypic *Haymondia wallichii* (DC.) Egan and Pan (Egan and pan 2015; Egan et al. 2016) or with tribe Desmodieae (Egan et al. 2016; Li et al. 2013; Stefanovic et al. 2009; de Queiroz et al. 2015). The inclusion of our transcriptomic data representing the *Mucuna* lineage thus fills a void within the emergent resource base of RNA-Seq data available for comparative evolutionary studies among different legume species.

In our species-level (combined for both the varieties) transcriptome assembly, we recovered 67,561 transcripts of which over 86% have been annotated against one or more legume proteomes. This represents a repertoire of expressed genes that can be used in downstream analyses for genetic analysis and crop improvement efforts. Our total number of recovered transcripts is quantitatively similar to a number of recent legume RNA-Seq studies. For instance, in *Trifolium pratense* L., Ištvánek et al. (2014) recovered 64,761 transcripts of which ~73% were annotated. From a genomic outlook, the annotated complement of genes in our work may seem relatively higher for a diploid species. The *Phaseolus vulgaris* L., (common bean) is estimated to have ~28,000 coding genes (Muñoz-Amatriaín et al. 2017), *Glycyrrhiza uralensis* Fisch., (Chinese licorice) has ~ 35,500 genes (Mochida et al. 2017) whereas *Glycine max* (L.) Merr., (soybean) - a recent polyploid crop (Egan and Doyle 2010) has between ~46,000 and 56,000 protein-coding genes (*Glycine max* Wm82.a2.v1 build at phytozome.jgi.doe.gov; Schmutz et al. 2014). The large number of annotated transcripts in our study may be due to the presence of many isoforms, alternatively transcribed transcripts, and/or gene fragments that failed to assemble completely with the settings used for the *de novo* assembly. It may be also due to a higher complement

of coding genes expressed in plants that produce a high quantity of secondary metabolites (Pichersky and Gang 2000).

In gene regulation and function, transcription factors (TFs) play an important role. Therefore, we assessed the number and distribution of transcription factors gene families within *M. pruriens* (Figure 24; Table 23). The results revealed that the overall distribution of transcripts coding for transcription factors among the various known protein families is akin to that of soybean and other legumes (Libault et al. 2009). Transcription factors constituted ~3.3% of annotated transcripts in our work which is similar to that in *Medicago truncatula* and *Lotus japonicus* but little higher than that of *Psophocarpus tetragonolobus* (Table 25) [*Additional file 5: Transcription factors identified in the Mucuna assembly using PlnTFDB link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM5_ESM.xlsx*; Sathyanarayana et al. 2017; BMC Genomics] (Vatanparast et al. 2016) and significantly lower than that estimated for soybean (*Glycine max* (L.) Merr.) (Schmutz et al. 2014) or chickpea (*Cicer arietinum* L.) (Garg et al. 2011). Understandably, soybean is expected to have more number of TFs than most other legumes due to its polyploid status. In our data, the MYB as well as bHLH (basic-Helix-Loop-Helix) gene families, are two of the most prevalent TF gene families, both of which play vital roles in secondary metabolite biosynthesis, predominantly for flavonoid and anthocyanin compounds (Vom Endt et al. 2002). With regards to differences within TF gene families of *Mucuna* relative to other legumes, our work reported several events of expansion (e.g. C2C2-GATA, CAMTA) and contraction (e.g. PHD), signalling the evolving nature of TF across the legume lineages.

6.2. Differential transcript expression across tissues

In our study, differential expression analysis carried out for three different tissues of genotype IC0620620 (Figure 24A-C) [*Additional file 9: Fold change expression values of the transcripts in each of the tissue analysed related to the secondary metabolite pathways, link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects/12864_2017_3780_MOESM9_ESM.xlsx*; Sathyanarayana et al. 2017; BMC Genomics) found thousands of differentially expressed transcripts between pods, roots and leaves. Transcripts highly expressed in one tissue against others indicate tissue-preferred expression – the information which is highly useful for further analysis on gene expression pattern. For instance, transcripts related to anthocyanin biosynthesis were upregulated in leaves vis-à-vis pods or roots (Figure 27). Uniquely expressed genes in leaf, root or pod tissues will be of great significance to determine their contributions towards commercially important traits in *Mucuna*. Genes related to secondary metabolism are particularly important as *Mucuna* is a natural source of mucunain and serotonin, chemicals present in pod hairs that cause itching (Reddy and Lerner 2010), and of elevated levels of L-DOPA found in seeds (Brain et al. 1976). L-DOPA is the precursor to dopamine, norepinephrine, and epinephrine (adrenaline), key neurotransmitters in the brain, and is extensively used in the treatment of Parkinson's disease (National Collaborating Centre for Chronic Conditions (UK). 2006). Recent studies have revealed that some Parkinson's patients respond better to *Mucuna* seed extract as a L-DOPA source and that this natural source is likely to be more effective as neuroprotective drug than L-DOPA itself while reducing adverse side effects (Katzenschlager et al. 2004). In our laboratory, efforts have been initiated towards further studies on the expression of genes specific to L-DOPA and other important secondary metabolites in this plant to gain more insights on the regulation of genes involved.

Tissue-specific expression of transcripts/genes involved in secondary metabolism has been reported in different plants, especially those possessing medicinal uses (e.g. in citronella, *Cymbopogon winterianus* Jowitt (Devi et al. 2016)). For comparison between root and leaf, we found transcripts upregulated in the root vis-à-vis leaf transcriptomes were mostly related to secondary metabolism. Similar results have been reported in a comparison of root and shoot tissues in *Leucaena leucocephala* (Lam.) de Wit (Ishihara et al. 2016). The highly upregulated gene in the root transcriptome related to secondary metabolism in our study was isoflavone reductase which showed a ~6-fold surge relative to pods and leaves (Figure 27) [*Additional file 9: Fold change expression values of the transcripts in each of the tissue analysed related to the secondary metabolite pathways, link - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-017-3780-9/MediaObjects /12864_2017_3780_MOESM9_ESM.xlsx*; Sathyanarayana et al. 2017, BMC Genomics]. Isoflavone reductase is an enzyme unique to the plant kingdom involved in the isoflavonoid phytoalexin biosynthesis and has been implicated in stress-related responses. In soybean (*Glycine max* (L.) Merr.), overexpression of isoflavone reductase was shown to enhance resistance to the *Phytophthora sojae* and confer antioxidant activity in the plant (Cheng et al. 2015). As root secondary metabolites are very less studied in *M. pruriens,* this information may enable newer research areas and allow identification of novel secondary metabolites of pharmaceutical importance.

Other transcripts associated with plant pigments such as anthocyanins revealed differential expression in leaf tissue compared to pods and roots. These pigments produce dark colors, mostly blue and purple in the above-ground plant tissues and also confer important antioxidant properties (Konczak et al. 2004). The predominant

secondary metabolism-related transcript upregulated in the leaves was chalcone synthase, the first step enzyme in the phenylpropanoid pathway that leads to the biosynthesis of several flavonoid secondary metabolites, including anthocyanins (Husain 2010). NAD(P)H-dependent 6' deoxychalcone synthase was moderately upregulated in the pod transcriptome. This is an enzyme implicated in the biosynthesis of isoliquiritigenin, a secondary metabolite known mainly from licorice (*Glycyrrhiza* spp.), a closely related legume genus. Isoliquiritigenin exhibits a number of beneficial pharmacological properties such as anti-viral, anti-inflammatory, anti-microbial, and cardioprotective effects and has shown significant anti-cancer properties (Peng et al. 2010). Thus, in combination with RNA-Seq and whole genomic resources developed from other medicinal legumes, our transcriptome provides a valuable resource for genetic studies related to secondary metabolites of medicinal application and interest in legume species.

6.3. Detection and validation of microsatellite markers

Microsatellites, or SSRs, are excellent genetic markers to aid genetic diversity assessment, constructing framework genetic maps, mapping useful genes, marker assisted selection and comparative mapping studies as well as association analysis (Gonzalo et al. 2005; Barakat et al. 2011). Thus, development of an SSR database known to be polymorphic within *M. pruriens* was highly desired in the direction of genetic improvement of this medicinal plant. In agreement with this need, we detected, within our transcriptome, over 3800 EST-SSRs of di-nucleotide or higher repeats (Table 15, Figure 16). Among these, tri-nucleotide repeats were the most copious, a sensible result given the coding nature of the transcriptome (Varshney et al. 2005). Other legumes exhibit similar trend, including the peanut (Bosamia et al. 2015) and winged bean (Vatanparast et al.2016).

We found, certain repeat motifs were more prevalent than others in our ~3800 EST-SSRs polymorphic between parents (Table 18), an observation reported even in other legumes previously (Vatanparast et al. 2016, Mun et al. 2006). Within our polymorphic SSR collection, $(AG)_n$, $(AAG)_n$, and $(AAAG)_n$ were most common in each repeat class (Table 18), a bias that was first identified in *Arabidopsis* (Zhang et al. 2004). Inside our full set of detected SSRs, motif type $(AG/CT)_n$ comprised 70.4% of all di-nucleotide repeats, with the $(CG/GC)_n$ motif nearly non-existent (Figure 16). This bias towards AG and against CG repeats has been proved across eukaryotes (Tóth et al. 2000), including other legumes such as *Phaseolus* (Blair et al. 2009b) and *Psophocarpus* (Vatanparast et al. 2016). Earlier works have suggested that AG repeats are commonly found in 5' UTRs (untranslated regions) (Mun et al. 2006) and, as such, may be related to transcription and regulation (Zhang et al. 2004). In the full set of detected SSRs, the $(AAG/GTT)_n$ repeat motifs and their complements were the most prevalent. The order ranking of tri-nucleotide repeat classes closely resembles that found in winged bean (Vatanparast et al. 2016).

Validation of SSRs discovered through transcriptome sequencing is an important step to building a working marker-set for *M. pruriens* genetic improvement efforts. Of the 134 primer pairs we screened, over 73% amplified genomic DNA successfully across 25 *Mucuna* accessions. This is comparable to or little lower than other efforts to validate genic SSR markers in legumes. Dutta et al. (2011) had 80% success rate in pigeon pea; Liu et al. (2013) recorded 82% success in *Alfalfa*, whereas Jhanwar et al. (2012) found a very high (98%) success rate in the chickpea. For transcriptome derived SSRs, marker dropout could be possible by chimeric primers, by the primer creation across intron/exon splice sites, or across alternative splice sites or chimeric transcripts.

6.4. Assessing genetic diversity and population structure

Genetic diversity within a species is subject to the relative influence of different barriers to gene flow such as physical or geographical separation, incipient genetic or morphological changes that impact the ability to crossbreed against promoters of gene flow such as migration or the movement or inter-breeding of individuals by human mediation, etc. (Ellstrand and Rieseberg 2016). Given the medicinal and ethnobotanical importance of *M. pruriens*, assessing the genetic diversity within this species is an important endeavour. Our efforts to create and validate a database of polymorphic EST-SSR markers for *M. pruriens*, discussed in the earlier section, ultimately yielded 52 polymorphic markers within 23 Indian *M. pruriens* accessions representing all three varieties and sourced across India. All but two of the 52 markers we used in the study showed adequate to high ability to discern ancestry based on the Shannon Information Content (I) (Lewontin et al. 1972) using the suggested cutoff of $I \geq 0.3$ (Smith et al. 2004). This attested to the utility of these markers for genetic diversity assessment. The average I across all our markers was 0.78, a value significantly higher than the discerning power of RAPD markers in *M. pruriens* (average of 0.62 across 15 primer pairs) (Padmesh et al. 2006). The average polymorphic information content (PIC) across 52 markers was 0.24, a value close to that obtained by Leelambika et al. (2016) but little higher than all estimates using AFLP or RAPD data which ranged from 0.166 (Sathyanarayana et al. 2011) to 0.174 (Leelambika et al. 2016), confirming to the appropriate choice of SSR markers for assessing genetic diversity in *Mucuna*.

We also explored genetic diversity across different subpopulation based on variety (*M.* vars. *hirsuta*, *pruriens* and *utilis*), geography (Peninsular, East and Northeast India), and empirical genetic structure (fastSTRUCTURE subgroups 1-4) (Table 20).

The results indicated that East India has slightly higher gene diversity than other geographical areas, a finding somewhat surprising considering the fact that all accessions for East India are of a single variety, *M. p.* var. *pruriens*. Peninsular India had the highest average number of alleles. This is attributed to this area comprising accessions from all three varieties, thus leading to an overall higher average number of alleles by the inclusion of the allelic diversity specific to this variety. Many cultivated crops are genetically less diverse as compared to their wild relatives (Flint-Garcia 2013), and *M. pruriens* var. *utilis* is no exception. Wild accessions (*M.* vars. *pruriens* and *hirsuta*) showed higher gene diversity and an average number of alleles as compared to cultivated (*M. p.* var. *utilis*) accessions (Table 20), corroborating previous studies on this aspect in *Mucuna* (Padmesh et al. 2006; Leelambika et al. 2016). Similar trends have been recorded throughout legumes, for instance within *Phaseolus vulgaris* (Papa and Gepts 2003). Nonetheless, estimates of gene flow were high irrespective of subdivision type (i.e. geography, variety, genetic structure), suggesting significant mixing among germplasm. All that said, caution is warranted at this stage in interpreting these results given our low number of accessions examined.

Our population genetic substructure examination revealed $K = 4$ subgroups (Figure 19 A-C) with coefficient of ancestry placing most accessions strongly within a particular subgroup. Both geography and variety were correlated to groupings within genetic substructure, PCA, and clustering analyses suggesting their impact on portioning of genetic diversity. For example, subgroups 2, 3 and 4 contained genotypes mostly collected from the East India, Peninsular India + Maharashtra and Northeast India, respectively, except for a few individuals whose placement was variable. Subgroup 2

grouped only var. *pruriens* accessions that are mostly from Eastern India whereas subgroup 3 mainly separated var. *hirsuta* with one accession each of var. *pruriens* and var. *utilis*.

We found neither geography nor variety correlated fully with clades generated by the NJ algorithm in contrast to earlier cluster analyses based on ISSR data where accessions largely associated based on their taxonomic affiliations (Leelambika et al. 2016). However, a more recent analysis using ISSR and RAPD markers across several *Mucuna* species, 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 quite distantly from the others, suggesting a possible separate evolutionary trajectory for this variety (Patil et al. 2016).

The lack of strict varietal clustering across three population structure assessment methods coupled with high estimates of migration and clear suggestion of at least two hybrid individuals as found by mixed coancestry within fastSTRUCTURE analyses suggests that hybridization takes place easily between varieties in this species, a conclusion also suggested previously (Leelambika et al. 2016). It is also possible that these variations may signify ancestral states perpetuated into extant populations. However, the small population size used in this study resulted in limited power to clearly recognize subgroups containing consistent genotypes across the three methods tested. Thus, a comprehensive study involving higher number of samples derived from extended geographical regions are needed to make generalized inference on the population structure and divergence of Indian *M. pruriens*.

6.5. Genetic linkage map

The 32 polymorphic EST-SSR markers used in our study were mapped into 11 linkage groups which covered a total distance of 837.04 cM in length. Linkage groups ranged from 4.49 cM to 186.48 cM in length with an average of 76.04 cM per linkage group. While most linkage groups consisted of only two to three linked markers, two linkage groups showed five linked markers. The largest linkage group spanned 186.48 cM.

From a genetic perspective, our efforts to develop framework linkage map met with limited success owing to low resolution of the map. This was due to the failure of microsatellite amplification in a large number of progenies and/or extreme linkage distortion. The reasons for such deviation within genic microsatellite markers especially in an inter-varietal cross are well-known (Min et al. 2017). As the map attained is of lower resolution than expected, this result can be at best described as a preliminary result. This also restricts its relevance for larger scientific discussion/inferences. Thus, large number of markers from extended SSR collection must be tried out in the future to convert this into a workable genetic linkage map. Markers developed both from MISA and lobSTR analysis in this project work provides enough scope for accomplishing this and holds significant promise to go over this objective once again in the near future.

7. Summary and Conclusions

M. pruriens is a plant with promising benefits. It is widely known for valued drug L-DOPA (L-3,4-dihydorxyphenylalanine) - a precursor of neurotransmitter dopamine as well as nutritional and soil improving properties. Non-availability of genomic resources, however, has impended effective utilization of its genetic resources for breeding key agronomic traits.

Microsatellites or simple sequence repeats have emerged as choice markers in plant genetics and breeding owing to many desirable attributes including codominant inheritance, reproducibility, relative abundance, wide genome coverage, chromosome specific location and amenability for automation. However, no effort has been made till date towards their development and deployment in *M. pruriens* breeding program. In this background, the current Ph.D. thesis aimed to develop and validate species specific EST-SSR markers for *M. pruriens* through RNA-Seq data analysis.

Our Illumina sequencing efforts on two contrasting genotypes generated ~18.24 GB data comprising 167,986,452 and 27,801,324 raw reads for genotypes IC0620620 and IC0620622 respectively. The Trinity assembly generated after filtering the raw reads yielded 67,561 high quality transcripts. This represents a large collection of genes for downstream analyses and future genetic improvement efforts in this crop. Additionally, the inclusion of our transcriptomic data representing the *Mucuna* lineage fills a gap within the developing resource base of transcriptomic data available for comparative evolutionary studies across the legume species.

In our transcriptome, we assessed the number and distribution of TF gene families within *M. pruriens*. The results indicated that the overall distribution of TF encoding

transcripts among the various known protein families is akin to that of other legumes with MYB and bHLH emerging as two most predominant TF gene families. Differential expression analysis found thousands of differentially expressed transcripts among three tissues of genotype IC0620620. This information will be helpful for further analysis on gene expression pattern in this species particularly of those involved in secondary metabolite pathways including that of L-DOPA. Based on this information, efforts have been initiated in our laboratory for further investigation on the expression of genes specific to L-DOPA as well as other secondary metabolites in this species to gain further understanding on the regulation of genes involved.

Development of an EST-SSR database within *M. pruriens* was highly desired in the direction of genetic improvement of this medicinal plant and therefore was a principal focus of this thesis. From our transcriptome analysis, we successfully detected a total of 6284 transcripts within which 7943 potential EST-SSRs were discovered. Further, for the first time in any legume species, we used a specialized program called lobSTR for identifying polymorphic SSRs between the contrasting parents. This screening identified over 4000 pSSRs which represents the largest collection of polymorphic SSR information in this crop so far.

Validation of discovered marker is yet another important criterion towards building a working marker set for *M. pruriens*. For this, we chose a sub-set of 134 primer pairs, over 73% of which amplified genomic DNA successfully across 25 *Mucuna* accessions. This is comparable to other efforts to validate genic SSR markers in legumes.

Given the medicinal and ethnobotanical importance of *M. pruriens*, assessing the genetic diversity within this species is an important endeavour. We therefore explored genetic diversity across different subpopulation based on variety, geography, and empirical genetic structure. The result indicated that the species has moderately higher gene diversity in East India than other geographical areas. Peninsular India had the highest average number of alleles. Wild accessions (*M. p.* vars. *pruriens* and *hirsuta*) showed greater gene diversity and average number of alleles as compared to cultivated (*M. p.* var. *utilis*) accessions. Estimates of gene flow were high irrespective of subdivision type, suggesting significant mixing among germplasm.

Our population genetic substructure examination revealed $K = 4$ subgroups with coefficient of ancestry placing most accessions strongly within a particular subgroup. Both geography and variety were correlated to groupings within genetic substructure, PCA, and clustering analyses suggesting their impact on portioning of genetic diversity. We found neither geography nor variety correlated fully with clades generated by the NJ algorithm. The lack of strict varietal clustering across three population structure assessment methods coupled with high estimates of migration and clear suggestion of at least two hybrid individuals within fastSTRUCTURE analyses suggests that hybridization takes place easily among the varieties in this species. It is also possible that these variations may signify ancestral states perpetuated into extant populations. However, the small sample size used in this study resulted in limited power to accurately recognize subgroups containing consistent genotypes across all the three methods tested. Thus, further studies involving a large number of samples derived from an extended geographical region are needed to make generalized conclusions on the divergence and population structure of Indian *M. pruriens*.

Towards genetic linkage map construction, 32 polymorphic EST-SSR markers used were mapped into 11 linkage groups which covered a total distance of 837.04 cM in length. Linkage groups ranged from 4.49 cM to 186.48 cM in length with an average of 76.04 cM per linkage group. While most linkage groups consisted only two to three linked markers, two linkage groups showed five linked markers. The largest linkage group spanned 186.48 cM. However, as the map attained was of lower resolution than expected, the result can be at best described as a preliminary outcome. This also restricts its relevance for larger scientific discussion/inferences on this topic. Thus, a large number of markers from extended SSR collection must be tried out in future to convert this into a workable genetic linkage map. Markers developed from this project provide enough scope for accomplishing this and holds significant promise to go over this objective in the near future.

8. Avenues for future research

Medicinal and agronomic potential of *M. pruriens* has remained largely underexploited. Efforts are needed to breed improved varieties not only for high or low L-DOPA content but also for developing cultivars resistant to biotic and abiotic stresses. Varietal development focusing on traits such as early maturing, erect, dwarf, self-supporting determinate cultivars and non-shattering pod with reduced anti-nutritional factors are other major breeding objectives in velvet bean. For initiating molecular breeding efforts towards attaining these goals, characterizing world-wide germplasm and developing functional markers such as EST-SSRs, SNPs, Intron spanning regions etc. are an immediate need. Also, so far, only two genetic linkage maps are available, that too with dominant AFLP markers. More efforts are needed to develop high density QTL-maps based on codominant markers as well as association analysis of target traits.

In addition, diminishing genotyping and sequencing costs are fuelling new hopes of whole genome sequencing as well as sequence-based trait mapping such as genotyping by sequencing (GBS) and/or Genome wide association studies (GWAS) in this crop. These efforts in combination with other 'omics' techniques such as proteomics and metabolomics can foster discovery of candidate genes and pathways involved in important biochemical and agronomic traits. As genome sequencing is becoming increasingly affordable, efforts such as deep sequencing of velvet bean genome should be put forward to enable large scale analyses of gene content, evolution of repetitive elements, linkage and association mapping. Prior to that, it is important to answer key questions concerning domestication and migration history of this orphan legume to bring its evolutionary history into perspective.

9. Bibliography

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

Acampora A, Ciaffi M, De Pace C, Paolacci AR, Tanzarella OA (2007). Pattern of variation for seed size traits and molecular markers in Italian germplasm of *Phaseolus coccineus* L. Euphytica 157 (1-2) 69-82

Ahmad NS, Basu SM, Redjeki ES, Murchie E, Massawe F, Azam-Ali S, Kilian A, Mayes S (2013). Developing genetic mapping and marker-assisted breeding techniques in Bambara groundnut (*Vigna subterranea* L.). Acta Horticulturae 979 (47) 437-449

Aitawade MM, Yadav SR (2012). *Mucuna sanjappae*, a new species from the north-Western Ghats, India. Kew Bulletin 67 (3) 539-543

Albuquerque TG, Santos F, Sanches-Silva A, Oliveira MB, Bento AC, Costa HS (2016). Nutritional and phytochemical composition of *Annona cherimola* Mill. fruits and by-products: Potential health benefits. Food Chemistry (193) 187-195

Ali M, Kumar S (2000). Problems and prospects of pulses research in India. Indian Farming (50) 4-13

Alleman Jr RJ, Canale RE, McCarthy CG, Bloomer RJ (2011). A blend of *Chlorophytum borivilianum* and velvet bean increases serum growth hormone in exercise-trained men. Nutrition and Metabolic Insights (4) 55-63

Almeida MR (1996). Flora of Maharashtra - Volume 1. Orient Press, Mumbai, pp. 294

Amole A, Oduguwa BO, Shittu O, Famakinde A, Okwelum N, Ojo VOA, Dele PA, Idowu OJ, Ogunlolu B, Adebiyi AO (2013). Herbage yield and quality of *Lablab purpureus* during the late dry Season in Western Nigeria. Slovak Journal of Animal Science 46 (1) 22-30

Andrews S (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. Online http://www.bioinformatics.babraham.ac.uk/ projects/fastqc (Mar 2016).

Aremu MO, Olaofe O, Basu SK, Abdulazeez G, Acharya SN (2010). Processed cranberry bean (*Phaseolus coccineus* L.) seed flour for the African diet. Canadian Journal of Plant Science 90 (50) 719-728

Arora RK (1981). Native Food Plants of the North-Eastern Tribals. In Glimpses of Indian Ethonobotany; (Ed. SK Jain) Oxford and IBH Publishing. New Delhi, India, pp. 91-106

Arora RK, Chandel KP, Joshi BS, Pant KC (1980). Rice bean: tribal pulse of eastern India. Economic Botany 34 (3) 260-263

Azhaguvel P, Saraswathi DV, Sharma A, Varshney RK (2006). Methodological advancement in molecular markers to delimit the genes for crop improvement. Floriculture, ornamental and plant biotechnology. 1. London: Global Science books. pp. 460–468

Badiane FA, Gowda BS, Cissé N, Diouf D, Sadio O, Timko MP (2012). Genetic relationship of cowpea (*Vigna unguiculata*) varieties from Senegal based on SSR markers. Genetics and Molecular Research 11 (1) 292-304

Bailey LH (1907). Encyclopedia of American Agriculture: a popular survey of agricultural conditions, practices and ideals in the United States and Canada. Macmillan and Co., London. (2) 656-658

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

Baker JG (1876). Leguminosae. In: Hooker JD, editor. The flora of British India. Kent, London, UK: Rev. & Co. (2) 56-306

Baqui AH, Black RE, El Arifeen S, Yunus M, Chakraborty J, Ahmed S, Vaughan JP (2002). Effect of zinc supplementation started during diarrhoea on morbidity and mortality in Bangladeshi children: community randomised trial. British Medical Journal. 325 (7372) 1059

Barakat MN, Al-Doss AA, Elshafei AA, Moustafa KA (2011). Identification of new microsatellite marker linked to the grain filling rate as indicator for heat tolerance genes in F2 wheat population. Australian Journal of crop science 5 (2) 104

Bastías A, Correa F, Rojas P, Almada R, Muñoz C, Sagredo B (2016). Identification and characterization of microsatellite loci in maqui (*Aristotelia chilensis* [Molina] Stunz) using next-generation sequencing (NGS). PloS One 11 (7) e0159825

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

Bhat R, Sridhar KR, Tomita-Yokotani K (2007). Effect of ionizing radiation on antinutritional features of velvet bean seeds (*Mucuna pruriens*). Food Chemistry 103 (3) 860-866

Bhat R, Sridhar KR, Young CC, Bhagwath AA, Ganesh S (2008). Composition and functional properties of raw and electron beam-irradiated *Mucuna pruriens* seeds. International Journal of Food Science & Technology 43 (8) 1338-1351

Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, Santo D, Nanni L, Attene G, Papa R (2017). Beans (*Phaseolus* spp.) as a model for understanding crop evolution. Frontiers in Plant Science (8) 722

Blair MW, Torres MM, Giraldo MC, Pedraza F (2009b). Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.). BMC Plant Biology 9 (1) 100

Blair MW, Torres MM, Pedraza F, Giraldo MC, Buendía HF, Hurtado N (2009a). Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched, small-insert genomic libraries. Genome 52 (9) 772-782

Bonifácio EM, Fonsêca A, Almeida C, dos Santos KG, Pedrosa-Harand A (2012). Comparative cytogenetic mapping between the lima bean (*Phaseolus lunatus* L.) and the common bean (*P. vulgaris* L.). Theoretical and Applied Genetics 124 (8) 1513-1520

Bort KS (1909). The Florida velvetbean and its history. US Department of Agriculture, Bureau of Plant Industry, Bulletin No (141) 25–32

Bosamia TC, Mishra GP, Thankappan R, Dobaria JR (2015). Novel and stress relevant EST derived SSR markers developed and validated in peanut. PloS One 10 (6) e0129127

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

Brain K (1976). Accumulation of L-DOPA in cultures from *Mucuna pruriens*. Plant Science Letters 7 (3) 157-161

Bravo L, Siddhuraju P, Saura-Calixto F (1998). Effect of various processing methods on the invitro starch digestibility and resistant starch content of Indian pulses. Journal of Agricultural and Food Chemistry 46 (11) 4667-4674

Bressani R (2002). Factros influencing nutritive value in food legumes, *Mucuna* compared to other grain legumes. Food and feed *Mucuna*: Current use and the way forward. Proceedings of an international work shop on food and feed from *Mucuna*: current uses and the way forward. Tegucigalpa Honduras April 26-29, 2000 164-188

Buckles D (1995). Velvetbean: A “new” plant with a history. Economic Botany 49 (1) 13-25

Buckles D, Triomphe B, Sain G (1998). Cover crops in hillside agriculture: farmer innovation with *Mucuna*. Idrc. 1-218

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

Buschiazzo E, Gemmell NJ (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays 28 (10) 1040-1050

Cantarella C, D’Agostino N (2015). PSR: polymorphic SSR retrieval. BMC Research Notes 8 (1) 525

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

Capo-Chichi LJA, Weaver DB, Morton CM (2003). The use of molecular markers in the study of genetic diversity in *Mucuna*. Tropical and Subtropical Agroecosystems 1 (2-3) 309-318

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

Carlderon E, Velasquez L, Bressani R (1992). Comparative study of the chemical composition and nutritive value of runner bean (*Phaseolus coccineus*) and of common bean (*Phaseolus vulgaris*). Archivos Latinoamericanos de Nutrición 42 (1) 64–71

Carsky J, Ndikawa R (1998). Identification of cover crops for the semi-arid savanna zone of West Africa. In: Cover Crops in West Africa Contributing to a Austainable Agriculture. Edited by Buckles D, Eteka A, Osiname O, Galiba M, Galiano G. Ottawa: International Development Research Center 179-187

Ceballos AI, Rivera JR, Arce MM, Valdivia CP (2012). Velvet Bean (*Mucuna pruriens* var. *utilis*) a cover crop as bioherbicide to preserve the environmental services of soil. In Herbicides-Environmental Impact Studies and Management Approaches pp. 167-184

Chang Y, Liu H, Liu M, Liao X, Sahu SK, Fu Y, Song B, Cheng S, Kariba R, Muthemba S, Hendre PS (2019). The draft genomes of five agriculturally important African orphan crops. GigaScience 8 (3) 152

Chao F, You D, Pedersen J, Hug L, Alkema L (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: a systematic assessment. The Lancet Global Health 6 (5) e535-547

Chaparro C, Oot L, Sethuraman K (2014). Overview of the Nutrition Situation in Seven Countries in Southeast Asia. Washington, DC: FHI 360/FANTA; pp. 1-62.

Chapman MA (2015). Transcriptome sequencing and marker development for four underutilized legumes. Applications in Plant Sciences 3 (2) 1400111

Chau CF, Cheung PK, Wong YS (1998). Chemical composition of three underutilized legume seeds grown in China. Food Chemistry 61 (4) 505-509

Chen H, Chen X, Tian J, Yang Y, Liu Z, Hao X, Wang L, Wang S, Liang J, Zhang L (2016). Development of gene-based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. PloS One 11 (3) e0151040

Chen H, Liu L, Wang L, Wang S, Somta P, Cheng X. (2015) Development and validation of EST-SSR markers from the transcriptome of adzuki bean (*Vigna angularis*). PloS One 10 (7) e0131939

Chen H, Liu L, Wang L, Wang S, Wang ML, Cheng X. (2015a) Development of SSR markers and assessment of genetic diversity of adzuki bean in the Chinese germplasm collection. Molecular Breeding 35 (10) 191

Chen H, Wang L, Liu X, Hu L, Wang S, Cheng X (2017). *De novo* transcriptomic analysis of cowpea (*Vigna unguiculata* L. Walp.) for genic SSR marker development. BMC genetics 18 (1) 65

Chen H, Wang L, Wang S, Liu C, Blair MW, Cheng X (2015b). Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers. PloS One 10 (4) e0120273

Cheng Q, Li N, Dong L, Zhang D, Fan S, Jiang L, Wang X, Xu P, Zhang S (2015). Overexpression of soybean isoflavone reductase (GmIFR) enhances resistance to Phytophthora sojae in soybean. Frontiers in Plant Science (6) 1024

Chibarabada T, Modi A, Mabhaudhi T (2017). Expounding the value of grain legumes in the semi-and arid tropics. Sustainability 9 (1) 1-25

Chimonyo VG, Modi AT, Mabhaudhi T (2016). Water use and productivity of a sorghum-cowpea-bottle gourd intercrop system. Agricultural Water Management. 165 (C) 82-96

Chivenge P, Mabhaudhi T, Modi A, Mafongoya P (2015). The potential role of neglected and underutilised crop species as future crops under water scarce conditions in Sub-Saharan Africa. International Journal of Environmental Research and Public Health (6) 5685-711

Christenhusz MJ, Byng JW (2016). The number of known plants species in the world and its annual increase. Phytotaxa 261 (3) 201-217

Collins WW, Hawtin GC (1999). Conserving and using crop plant biodiversity in agroecosystems. In Biodiversity in Agroecosystems; CRC Press, London, UK. 215-236

Considine MJ, Siddique KH, Foyer CH (2017). Nature's pulse power: legumes, food security and climate change. Journal of Experimental Botany 68 (8) 1815-1818

Conway G (1997). The doubly green revolution: food for all in the 21st century. London: Penguin Books; 336p

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics (177) 1889-1913

Daffodil ED, Tresina PS, Mohan VR (2016). Nutritional and antinutritional assessment of *Mucuna pruriens* (L.) DC var. *utilis* (Wall ex. Wight) Bak. Ex Burck and *Mucuna deeringiana* (Bort) Merril: An underutilized tribal pulse. International Food Research Journal 23 (4) 1501-1513

Dahipahle AV, Kumar S, Sharma N, Singh H, Kashyap S, Meena H (2017). Rice Bean-A Multipurpose, Underutilized, Potential Nutritive Fodder Legume-A Review. Journal of Pure and Applied Microbiology 11 (1) 433-440

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST (2011). The variant call format and VCF tools. Bioinformatics 27 (15) 2156-2158

Dansi A, Vodouhè R, Azokpota P, Yedomonhan H, Assogba P, Adjatin A, Loko YL, Dossou-Aminon I, Akpagana K (2012). Diversity of the neglected and underutilized crop species of importance in Benin. The Scientific World Journal (2012) 19

Dassanayake MD, Fosberg FR (1980) editors. A revised handbook to the flora of Ceylon. CRC PressI Llc (2)

De Onis M, Blossner M (1997). World Health Organization (WHO) global database on child growth and malnutrition. Geneva: World Health Organization.

De Onis M, Monteiro C, Akré J, Glugston G (1993). The worldwide magnitude of protein-energy malnutrition: an overview from the WHO Global Database on Child Growth. Bulletin of the World health Organization 71 (6) 703

de Queiroz LP, Pastore JFB, Cardoso D, Snak C, Lima ALdC, Gagnon E, Vatanparast M, Holland AE, Egan AN (2015). A multilocus phylogenetic analysis reveals the monophyly of a recircumscribed papilionoid legume tribe Diocleae with well-supported generic relationships. Molecular Phylogenetics and Evolution (90) 1-19

Deka RK, Sarkar CR (1990). Nutrient composition and antinutritional factors of Dolichos lablab L. seeds. Food Chemistry 38 (4) 239-246

Delgado Salinas A (1988). Variation, taxonomy, domestication, and germplasm potentialities in *Phaseolus coccineus*. In Genetic resources of *Phaseolus* beans: their maintenance, domestication, evolution and utilization (P. Gepts ed.), Kluwer Academic Publishers, Boston, MA pp. 441-463

Delseny M, Laroche M, Penon P (1983). Detection of sequences with Z-DNA forming potential in higher plants. Biochemical and Biophysical Research Communications 116 (1) 113-120

Devi K, Mishra SK, Sahu J, Panda D, Modi MK, Sen P (2016). Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus. Scientific Reports (6) 21026

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin (19) 11-15

Duan X, Schmidt E, Li P, Lenox D, Liu L, Shu C, Zhang J, Liang C (2012). PeanutDB: an integrated bioinformatics web portal for Arachis hypogaea transcriptomics. BMC Plant Biology (12) 94

Duan X, Wang K, Su S, Tian R, Li Y, Chen M (2017). *De novo* transcriptome analysis and microsatellite marker development for population genetic study of a serious insect pest, *Rhopalosiphum padi* (L.) (Hemiptera: Aphididae). PloS One (12) e0172513

Duggar JF. (1899) Velvetbeans. Alabama Agriculture Experiment Station, Auburn, Alabama. USA. Bulletin (104) 113-125

Duke JA (1981). Handbook of Legumes of World Economic Importance. Plenum Press, New York

Duke JA (2008). The phytochemical database (data version January 2008) Retrieved November 5. from http://www.ars-grin.gov/duke/

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeon pea (*Cajanus cajan* (L.) Millspaugh). BMC Plant Biology 11 (1) 17

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

Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics 49 (4) 746-756

Edwards O, Singh KB (2006). Resistance to insect pests: What do legumes have to offer? Euphytica 147 (1-2) 273-285

Egan AN, Doyle JJ (2010). A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). Systematic Biology 59 (5) 534-547

Egan AN, Pan B (2015). 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 218 (3) 201-226

Egan AN, Vatanparast M, Cagle W (2016). Parsing polyphyletic *Pueraria:* Delimiting distinct evolutionary lineages through phylogeny. Molecular Phylogenetics and Evolution (104) 44-59

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 and 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), pp 18-47

Ekblom R, Galindo J (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107 (1) 1-15

Ellstrand NC, Rieseberg LH (2016). When gene flow really matters: gene flow in applied evolutionary biology. Evolutionary Applications 9 (7) 833-836

Elmes RPT (1976). Cross-inoculation relationships of Psophocmpus tetragonolobusand its Rhizobium with other legumes and Rhizobia. Papua New Guinea Agricultural Journal 27 (3) 53-57

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

Ezeagu IE, Ibegbu MD (2010). Biochemical composition and nutritional potential of ukpa: a variety of tropical lima beans (*Phaseolus lunatus*) from Nigeria – a short report. Polish Journal of Food and Nutrition Sciences 60 (3) 231-235

Famurewa JA, Raji AO (2005) Parameters Affecting Milling Qualities of Undefatted Soybeans (Glycine max, l. Merill) (2): Selected Thermal Treatment. International Journal of Food Engineering 1 (5)

FAO, IFAD and WFP (2013). The State of Food Insecurity in the World 2013. The multiple dimensions of food security. FAO, Rome.

FAO, IFAD, UNICEF, WFP and WHO (2018). The State of Food Security and Nutrition in the World 2018. Building climate resilience for food security and nutrition. Rome, FAO

Field D, Wills C (1996). Long, polymorphic microsatellites in simple organisms. Proceedings of the Royal Society of London. Biological Sciences 263 (1367) 209-215

Flint-Garcia SA (2013). Genetics and consequences of crop domestication. Journal of Agricultural and Food Chemistry 61 (35) 8267-8276

Foyer CH, Lam HM, Nguyen HT, Siddique KH, Varshney RK, Colmer TD, Cowling W, Bramley H, Mori TA, Hodgson JM, Cooper JW (2016). Neglecting legumes has

compromised human health and sustainable food production. Nature Plants 2 (8) 16112

Fujii Y, Shibuya T, Yasuda T (1991). L-3,4-Dihydroxyphenylalanine as an allelochemical candidate from *Mucuna pruriens* (L.) DC. var. *utilis*. Agricultural and Biological Chemistry 55 (2) 617-618

Fuller DQ (2003). African crops in prehistoric South Asia: a critical review. In: Neumann K, Butler A, Kahlheber S (eds) Food, fuel, fields-progress in African archaeobotany, Heinrich-Barth- Institut, Köln, Germany. Africa Praehistorica (15) 239-271

Fung SY, Tan NH, Sim SM, Marinello E, Guerranti R, Aguiyi JC (2011). *Mucuna pruriens* Linn. seed extract pretreatment protects against cardiorespiratory and neuromuscular depressant effects of Naja sputatrix (Javan spitting cobra) venom in rats. Indian Journal of Experimental Biology 49 (4) 254-259

Garg R, Patel RK, Tyagi AK, Jain M (2011). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Research 18 (1) 53-63

Gdala J, Buraczewska L (1996). Chemical composition and carbohydrate content of seeds from several lupin species. Journal of Animal and Feed Sciences 5 (4) 403-416

Gebrelibanos M, Tesfaye D, Raghavendra Y, Sintayeyu B (2013). Nutritional and health implications of legumes. International Journal of Pharmaceutical Sciences and Research 4 (4) 1269

Geng J, Li K, Zhang Y, Hu S (2010). A modified enrichment method to construct microsatellite library from plateau pika genome (*Ochotona curzoniae*). Genomics, Proteomics and Bioinformatics 8 (1) 72-76

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005). Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiology 137 (4) 1228–1235

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005). Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference

Ghane SG, Lokhande VH, Ahire ML, Nikam TD (2010). *Indigofera glandulosa* Wendl. (Barbada) a potential source of nutritious food: underutilized and neglected legume in India. Genetic Resources and Crop Evolution 57 (1) 147-153

Glenn TC, Schable NA (2005). Isolating microsatellite DNA loci. Methods in enzymology. Academic Press (395) pp 202-222

Göhl B (1982). Piensos tropicales: Resúmenes informativos sobre piensos y valores nutritivos, Producción y Sanidad Animal 12, FAO, Roma, Italia, p. 550

Golein BE, Talaie A, Zamani Z, Moradi B (2006). Development and characterization of new microsatellite loci from lemon (*Citrus limon*). International Journal of Agriculture and Biology 8 172-174

Gong YM, Xu SC, Mao WH, Hu QZ, Zhang GW, Ding J, Li ZY (2010). Generation and characterization of 11 novel EST derived microsatellites from *Vicia faba* (Fabaceae). American Journal of Botany 97 (7) e69-71

Gonzalo MJ, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, Arus P, Monforte AJ (2005). Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theoretical and Applied Genetics 110 (5) 802-811

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29 (7) 644-652

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

Grantham-McGregor S (1995). A review of studies of the effect of severe malnutrition on mental development. Journal of Nutrition 125 (8 Suppl) 2233S-2238S

Grover Z, Ee LC (2009). Protein energy malnutrition. Pediatric Clinics 56 (5) 1055-1068

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 *Mucuna pruriens* extract. Proteomics 8 (2) 402-412

Gulzar M, Minnaar A (2017). Underutilized protein resources from African legumes. Sustainable Protein Sources 197-208

Gupta M, Mazumder UK, Chakraborti S, Bhattacharya S, Rath N, Bhawal SR (1997). Antiepileptic and anticancer activity of some indigenous plants. Indian Journal of Physiology and Allied Sciences 51 (2) 53-56

Gupta SK, Gopalakrishna T (2010). Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. Genome 53 (7) 508-523

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

Gymrek M, Golan D, Rosset S, Erlich Y (2012). lobSTR: a short tandem repeat profiler for personal genomes. Genome research 22 (6) 1154-1162

Habiyaremye C, Matanguihan JB, D'Alpoim Guedes J, Ganjyal GM, Whiteman MR, Kidwell KK, Murphy KM (2017). Proso millet (*Panicum miliaceum* L.) and its potential for cultivation in the Pacific Northwest, US: a review. Frontiers in Plant Science (7) 1961

Hadebe ST, Modi AT, Mabhaudhi T (2017). Drought tolerance and water use of cereal crops: A focus on sorghum as a food security crop in sub-Saharan Africa. Journal of Agronomy and Crop Science 203 (3) 177-191

Hamada H, Petrino MG, Kakunaga T (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proceedings of the National Academy of Sciences 79 (21) 6465-6469

Hammer K, Heller J, Engels J (2001). Monographs on underutilized and neglected crops. Genetic Resources and Crop Evolution 48 (1) 3-5

Han OK, Kaga A, Isemura T, Wang XW, Tomooka N, Vaughan DA (2005). A genetic linkage map for azuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi). Theoretical and Applied Genetics 111 (7) 1278-1287

Hancock JM (1995). The contribution of slippage-like processes to genome evolution. Journal of Molecular Evolution 41 (6) 1038-1047

Hane JK, Ming Y, Kamphuis LG, Nelson MN, Garg G, Atkins CA, Bayer PE, Bravo A, Bringans S, Cannon S et al. (2017). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. Plant Biotechnology Journal 15 (3) 318-330

Harder D (1992a) Chromosome counts in *Psophocarpus*. Kew Bulletin 47 (3) 529-534

Harder D, Smrtt J (1992). Further evidence on the origin of cultivated winged bean, *Psophocarpus tetragonolobus* (L.) DC. (Fabaceae): Chromosome numbers and the presence of host-specific fingus. Economic Botany 46 (2) 187-191

Harishankar, Dwivedi S, Dabral SB, Walia DK (2004). Nutritional status of children under 6 years of age. Indian Journal of Preventive and Social Medicine (35) 156–62

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

Ho WK, Muchugi A, Muthemba S, Kariba R, Mavenkeni BO, Hendre P, Song B, Van Deynze A, Massawe F, Mayes S (2016). Use of microsatellite markers for the assessment of bambara groundnut breeding system and varietal purity before genome sequencing. Genome 59 (6) 427-431

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

Husain Q (2010). Chemistry and biochemistry of some vegetable flavors. In: Handbook of Fruit and Vegetable Flavors. Edited by Hui YH. Hoboken: John Wiley & Sons, Inc.; 575-625

Hymowitz T, Boyd J (1977) Origin, ethnobotany and agricultural potential of the winged bean- *Psophocarpus tetragonolobus*. Economic Botany 31 (2) 180-188

Iangrai B, Pattanayak A, Khongwir DE, Pale G, Gatphoh EM, Das A, Chrungoo NK (2017). Development and characterization of a new set of genomic microsatellite markers in rice bean (*Vigna umbellata* (Thunb.) Ohwi and Ohashi) and their utilization in genetic diversity analysis of collections from North East India. PloS One 12 (7) e0179801

IFAD (2002). Enabling the Rural Poor to Overcome their Poverty International Fund for Agricultural Development, Rome (2002).

Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T, Jo U, Vaughan DA, Tomooka N (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). PloS One 7 (8) e41304

Isemura T, Kaga A, Tomooka N, Shimizu T, Vaughan DA (2010). The genetics of domestication of rice bean, *Vigna umbellata*. Annals of botany 106 (6) 927-944

Ishihara KL, Honda MD, Pham DT, Borthakur D (2016). Transcriptome analysis of *Leucaena leucocephala* and identification of highly expressed genes in roots and shoots. Transcriptomics 4 (1) 135

Ištvánek J, Jaroš M, Křenek A, Řepková J (2014). Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). American Journal of Botany 101 (2) 327-337

Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao YP, Ganten D, Dzau VJ, Lander ES (1991). Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell 67 (1) 213-224

Janardhanan K, Lakshmanan KK (1985). Studies on the pulse, *Mucuna utilis*: Chemical composition and antinutritional factors. Journal of Food Science and Technology 22 (5) 369-371

Janardhanan V, Vadivel K (2000). Nutritional and anti-nutritional composition of velvet bean: an under-utilized food legume in South India. International Journal of Food Sciences and Nutrition 51 (4) 279-287

Jhanwar S, Priya P, Garg R, Parida SK, Tyagi AK, Jain M (2012). Transcriptome sequencing of wild chickpea as a rich resource for marker development. Plant Biotechnology Journal 10 (6) 690-702

Jorge MAB, Eilittä M, Proud FJ, Maasdorp BV, Beksissa H, Sarial AK, Hanson J (2007). *Mucuna* Species: recent advances in application of biotechnology. Fruit, Vegetable and Cereal Science and Biotechnology 1 (2) 80-94

Josephine RM, Janardhanan K (1992). Studies on chemical composition and antinutritional factors in three germplasm seed materials of the tribal pulse, *Mucuna pruriens* (L.) DC. Food Chemistry 43 (1) 13-18

Kadam SS, Salunkhe DK, Luh BS (1984). Winged bean in human nutrition. Critical Reviews in Food Science and Nutrition 21 (1) 1-40

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

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

Kalidass C, Mohan VR (2010). 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 agroecosystems14 (1) 279-293

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

Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J, Materne M, Forster JW (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics 13 (1) 104

Kay DE (1985). Legumbres alimenticias, Acribia, Zaragoza, España, pp. 437

Keerthi CM, Ramesh S, Byregowda M Vaijayanthi PV (2018). Simple sequence repeat (SSR) marker assay-based genetic diversity among dolichos bean (*Lablab purpureus* L. Sweet) advanced breeding lines differing for productivity per se traits. International Journal of Current Microbiology and Applied Sciences 7 (5) 3736-3744

Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Research 18 (1) 30-38

Kerr SC, Gaiti F, Beveridge CA, Tanurdzic M (2017). *De novo* transcriptome assembly reveals high transcriptional complexity in *Pisum sativum* axillary buds and shows rapid changes in expression of diurnally regulated genes. BMC Genomics 18 (1) 221

Keyzer MA, Merbis MD, Pavel IF, Van Wesenbeeck CF (2005). Diet shifts towards meat and the effects on cereal use: can we feed the animals in 2030?. Ecological Economics 55 (2) 187-202

Khadka K, Acharya BD (2009). Cultivation practices of ricebean. Research and Development (LI-BIRD), Local Initiatives for Biodiversity, Pokhara, Nepal. 1-21

Khanal AR, Khadka K, Poudel I, Joshi KD and Hollington PA (2007). Food Security through Ricebean Research in India and Nepal (FOSRIN). Report 3. Farmers' local knowledge associated with the production, utilization and diversity of ricebean (*Vigna umbellata*) in Nepal. Pokhara, Nepal, Local Initiatives for Biodiversity, Research and Development, Rampur, Chitwan, Nepal: Institute of Agriculture and Animal Sciences, and Bangor, Wales, UK: CAZS Natural Resources, College of Natural Sciences, Bangor University

Khoshbakht K, Hammer K (2008). Species richness in relation to the presence of crop plants in families of higher plants. Journal of Agriculture and Rural Development in the Tropics and Subtropics 109 (2) 181-190

Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. Genetics (49) 725-738

Kirtikar KR, Basu BD (1985). Indian medicinal plants. Allahabad. (3) 1347-1348

Konczak I, Zhang W (2004). Anthocyanins-more than nature's colours. Journal of Biomedicine and Biotechnology 2004 (5) 239-240

Kongjaimun A, Kaga A, Tomooka N, Somta P, Shimizu T, Shu Y, Isemura T, Vaughan DA, Srinives P (2012b). An SSR-based linkage map of yardlong bean (*Vigna unguiculata* (L.) Walp. subsp. *unguiculata* Sesquipedalis Group) and QTL analysis of pod length. Genome 55 (2) 81-92

Kongjaimun A, Kaga A, Tomooka N, Somta P, Vaughan DA, Srinives P (2012a). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *sesquipedalis*. Annals of Botany 109 (6) 1185-1200

Korir NK, Han J, Shangguan LF, Wang C, Kayesh E, Zhang YY, Fang JG (2013). Plant variety and cultivar identification: Advances and prospects. Critical Reviews in Biotechnology 33 (2) 111-125

Kumar A, Metwal M, Kaur S, Gupta AK, Puranik S, Singh S, Singh M, Gupta S, Babu BK, Sood S, Yadav R (2016). Nutraceutical value of finger millet (*Eleusine coracana* (L.) Gaertn.), and their improvement using omics approaches. Frontiers in Plant Sciences (7) 934

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

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

Labuda H (2010). Runner bean (*Phaseolus coccineus* L.)-Biology and use. Acta Scientiarum Polonorum-Hortorum Cultus 9 (3) 117-132

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

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

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9 (4) 357-359

Laurena AC, Revilleza MJ, Mendoza EM (1994). Polyphenols, phytate, cyanogenic glycosides, and trypsin inhibitor activity of several Philippine indigenous food legumes. Journal of Food Composition and Analysis 7 (3) 194-202

Leelambika M, Mahesh S, Jaheer M, Tripathi P, Kumar PR, Sathyanarayana N (2016). Targeted metabolic and genomic profiling reveals parents for L-dopa breeding in *Mucuna pruriens* (L.) DC. Tropical Plant Biology 9 (4) 239-351

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

Leelambika M, Sathyanarayana N (2011). Genetic characterization of Indian *Mucuna* (Leguminoceae) species using morphometric and random amplification of

polymorphic DNA (RAPD) approaches. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology 145 (4) 786-797

Lehman H, Ballow M (2014). Immune Compromise Due to Metabolic Disorders: Malnutrition, Obesity, Stress, and Inborn Errors of Metabolism. In Stiehm's Immune Deficiencies Academic Press pp. 823-834.

Lepcha P, Egan AN, Doyle JJ, Sathyanarayana N (2017). A review on current status and future prospects of winged bean (*Psophocarpus tetragonolobus*) in tropical agriculture. Plant Foods for Human Nutrition 72 (3) 225-235

Lepcha P, Kumar PR, Sathyanarayana N (2019). Exploring Genomics Research in the Context of Some Underutilized Legumes-A Review. In OMICS-Based Approaches in Plant Biotechnology, Scrivener Publishing, pp. 1-8

Lewontin RC (1972). The apportionment of human diversity. In: Evolutionary Biology, Springer (6) 381-398

Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12 (1) 323

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

Li H, Wang W, Lin L, Zhu X, Li J, Zhu X, Chen Z (2013). Diversification of the phaseoloid legumes: effects of climate change, range expansion and habit shift. Frontiers in Plant Sciences (4) 386

Li J, Guo H, Wang Y, Zong J, Chen J, Li D, Li L, Wang J, Liu J (2018). High-throughput SSR marker development and its application in a centipedegrass

(*Eremochloa ophiuroides* (Munro) Hack.) genetic diversity analysis. PloS One 13 (8) e0202605

Li J, Sun X, Yu G, Jia C, Liu J, Pan H (2014). Generation and analysis of expressed sequence tags (ESTs) from halophyte *Atriplex canescens* to explore salt-responsive related genes. International Journal of Molecular Sciences 15 (6) 11172-11189

Libault M, Joshi T, Benedito VA, Xu D, Udvardi MK, Stacey G (2009). Legume transcription factor genes: what makes legumes so special? Plant Physiology 151 (3) 991-1001

Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu and Weizhong Li (2012). CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics 28 (23) 3150-3152

Litt M, Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics 44 (3) 397

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

Liu Z, Chen T, Ma L, Zhao Z, Zhao PX, Nan Z, Wang Y (2013). Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. PloS One 8 (12) e83549

Lloyd KG, Davidson L, Hornykiewicz O (1975). The neurochemistry of Parkinson's disease: effect of L-DOPA therapy. Journal of Pharmacology and Experimental Therapeutics 195 (3) 453-464

Lorenzetti E, MacIsaac S, Anasson J, Avang D (1998). 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 57

Luse RA, Okwuraiwe PE (1975). The Role of Grain Legume in Tropical Nutrition, Proc. 1st African Nutrition Congress

Maass BL, Jamnadass RH, Hanson J, Pengelly BC (2005). Determining sources of diversity in cultivated and wild *Lablab purpureus* related to provenance of germplasm by using amplified fragment length polymorphism. Genetic Resources and Crop Evolution 52 (6) 683-695

Mabhaudhi T, Chimonyo V, Modi A (2017a). Status of underutilised crops in South Africa: opportunities for developing research capacity. Sustainability 9 (9) 1569

Mabhaudhi T, Chimonyo VG, Chibarabada TP, Modi AT (2017b). Developing a roadmap for improving neglected and underutilized crops: a case study of South Africa. Frontiers in Plant Sciences (8) 2143

Mahesh S (2015). Identification of Elite Germplasm, DNA Fingerprinting and development of Framework Linkage Map in Indian Velvetbean (*Mucuna pruriens* L.) (Doctoral dissertation, Sir M Visvesvaraya Institute of Technology, Bangalore).

Mahesh S, Leelambika M, Jaheer M, Anithakumari AM, Sathyanarayana N (2016). Genetic mapping and QTL analysis of agronomic traits in Indian *Mucuna pruriens* using an intraspecific F 2 population. Journal of Genetics 95 (1) 35-44

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

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

Mal B, Joshi V (1991). Under-utilised plant resources: Plant genetic resources conservation and management concepts and approaches. International Board for Plant Genetic Resources, Regional Office for South and Southeast Asia, c/o NBPGR, Pusa Campus, New Delhi 110 012, India

Maleta K (2006). Undernutrition. Malawi medical journal: The Journal of Medical Association of Malawi 18 (4) 189-205

Maphosa Y, Jideani VA (2017). The Role of Legumes in Human Nutrition. InFunctional Food-Improve Health through Adequate Food 2017 Aug 2. IntechOpen.

Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z (2010). Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11 (1) 663

Marubodee R, Ogiso-Tanaka E, Isemura T, Chankaew S, Kaga A, Naito K, Ehara H, Tomooka N (2015). Construction of an SSR and RAD-marker based molecular linkage map of *Vigna vexillata* (L.) A. Rich. PloS One 10 (9) e0138942

Mason AS (2015). SSR Genotyping. In: Batley J, editor. Plant Genotyping. Springer; New York, NY. (1245) 77–89

Mayes S, Massawe FJ, Alderson PG, Roberts JA, Azam-Ali SN, Hermann M (2011). The potential for underutilized crops to improve security of food production. Journal of Experimental Botany 63 (3) 1075-1079

McClean P, Gepts P, Kami J (2004). Genomic and genetic diversity in common bean. In Legume Crop Genomics (Eds RF Wilson, HT Stalker, EC Brummer) pp. 60–82

Mehra R, Rojas MH (2008). Women, food security and agriculture in a global marketplace. International Center for Research on Women (ICRW).

Mejri F, Selmi S, Martins A, Baati T, Chaabane H, Njim L, Serralheiro ML, Rauter AP, Hosni K (2018). Broad bean (*Vicia faba* L.) pods: a rich source of bioactive ingredients with antimicrobial, antioxidant, enzyme inhibitory, anti-diabetic and health-promoting properties. Food and Function 9 (4) 2051-2069

Merritt BJ, Culley TM, Avanesyan A, Stokes R, Brzyski J (2015). An empirical review: Characteristics of plant microsatellite markers that confer higher levels of genetic variation. Applications in Plant Sciences 3 (8) 1500025

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

Miesfeld R, Krystal M, Amheim N (1981). A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes. Nucleic Acids Research 9 (22) 5931-5948

Min X, Zhang Z, Liu Y, Wei X, Liu Z, Wang Y, Liu W (2017). Genome-wide development of microRNA-based SSR markers in *Medicago truncatula* with their transferability analysis and utilization in related legume species. International Journal of Molecular Sciences 18 (11) 2440

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

Mlyneková Z, Chrenková M, Formelová Z (2014). Cereals and Legumes in Nutrition of People with Celiac. International Journal 2 (3) 105-109

Mochida K, Sakurai T, Seki H, Yoshida T, Takahagi K, Sawai S, Uchiyama H, Muranaka T, Saito K (2017). Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. Plant Journal 89 (2) 181-194

Mohan VR, Janardhanan K (1993). Chemical and nutritional evaluation of raw seeds of the tribal pulses *Parkia roxburghii* G. Don. and *Entada phaseoloides* (L.) Merr. International Journal of Food Sciences and Nutrition 44 (1) 47-53

Mohan VR, Janardhanan K (1994). Chemical composition and nutritional evaluation of raw seeds of six rice bean varieties. Journal of the Indian Botanical Society (73) 259-263.

Mohan VR, Janardhanan K (1995). Chemical analysis and nutritional assessment of lesser known pulses of the genus, *Mucuna*. Food Chemistry 52 (3) 275-280

Mohanty CS, Pradhan RC, Singh V, Singh N, Pattanayak R, Prakash O, Chanotiya CS, Rout PK (2015). Physicochemical analysis of *Psophocarpus tetragonolobus* (L.) DC seeds with fatty acids and total lipids compositions. Journal of Food Science and Technology 52 (6) 3660-3670

Molosiwa OO, Aliyu S, Stadler F, Mayes K, Massawe F, Kilian A, Mayes S (2015). SSR marker development, genetic diversity and population structure analysis of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces. Genetic Resources and Crop Evolution 62 (8) 1225-1243

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

Moura TM, Vatanparast M, Tozzi AM, Forest F, Wilmot-Dear CM, Simon MF, Mansano VF, Kajita T, Lewis GP, Herendeen PS (2016). A molecular phylogeny and new infrageneric classification of *Mucuna* Adans. (Leguminosae-Papilionoideae) including insights from morphology and hypotheses about biogeography. International Journal of Plant Sciences 177 (1) 76-89

Mun JH, Kim DJ, Choi HK, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudez AM (2006). Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. Genetics 172 (4) 2541-2555

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

Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R (2015). Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. Bioinformatics 31 (13) 2199-2201

Naeem M, Masroor M, Khan A, Moinuddin (2016). Hyacinth bean (*Lablab purpureus* (L.) Sweet): A hidden treasure of useful phytochemicals. The journal of the International Legume Society (13) 30-32

Nair NC, Henry AN (1983). Flora of Tamil Nadu, India. Series 1: Analysis Vol. 1. Coimbatore: BSI.

National Academy of Sciences (1975). The winged bean: A high-protein crop for the tropics. Washington, DC, USA.

National Collaborating Centre for Chronic Conditions (Great Britain) (2006). Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care. Royal College of Physicians, London.

Nei M (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences 70 (12) 3321-3323

Nelson RJ, Naylor RL, Jahn MM (2004). The role of genomics research in improvement of "orphan" crops. Crop Science 44 (6) 1901-1905

Nyanga LK, Gadaga TH, Nout MJ, Smid EJ, Boekhout T, Zwietering MH (2013). Nutritive value of masau (*Ziziphus mauritiana*) fruits from Zambezi Valley in Zimbabwe. Food chemistry 138 (1) 168-172

Okito A, Alves BJ, Urquiaga S, Boddey RM (2004). Nitrogen fixation by groundnut and velvet bean and residual benefit to a subsequent maize crop. Pesquisa Agropecuária Brasileira 39 (12) 1183-1190

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira ML (2006). Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology 29 (2) 294-307

Onweluzo J, Eilittä M (2003). Surveying *Mucuna's* utilization as a food in Enugu and Kogi states of Nigeria. Tropical and Subtropical Agroecosystems 1 (2-3)

Ostrander EA, Jong PM, Rine J, Duyk G (1992). Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences.

Proceedings of the National Academy of Sciences of the United States of America 89 (8) 3419-3423

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

Padulosi S, Eyzaquirre P, Hodgkin T (1999). Challenges and Strategies in Promoting Conservation and Use of Neglected and Underutilized Crop Species. Alexandria, Va, USA: ASHS Press 1999

Paetkau D (1999). Microsatellites obtained using strand extension: an enrichment protocol. Biotechniques 26 (4) 690-697

Paliwal R, Kumar R, Choudhury DR, Singh AK, Kumar S, Kumar A, Bhatt KC, Singh R, Mahato AK, Singh NK, Singh R (2016). Development of genomic simple sequence repeats (g-SSR) markers in *Tinospora cordifolia* and their application in diversity analyses. Plant Gene (5) 118-125

Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK (2016). Emerging genomic tools for legume breeding: current status and future prospects. Frontiers in Plant Sciences (7) 455

Pant MC, Uddin I, Bhardwaj UR, Tewari RD (1968). Blood sugar and total cholesterol lowering effect of *Glycine soja* (Sieb and Zucc.), *Mucuna pruriens* (DC) and *Dolichos biflorus* (Linn.) seed diets in normal fasting albino rats. Indian Journal of Medical Research, 56 (12) 1808-1812

Papa R, Gepts P (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theoretical and Applied Genetics 106 (2) 239-250

Parida SK, Kalia SK, Kaul S, Dalal V, Hemaprabha G, Selvi A, Pandit A, Singh A, Gaikwad K, Sharma TR, Srivastava PS (2009). Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theoretical and Applied Genetics 118 (2) 327-338

Park YJ, Lee JK, Kim NS (2009). Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules 14 (11) 4546-4569

Patil RR, Pawar KD, Rane MR, Yadav SR, Bapat VA, Jadhav JP (2016). Assessment of genetic diversity in *Mucuna* species of India using randomly amplified polymorphic DNA and inter simple sequence repeat markers. Physiology and Molecular Biology Of Plants 22 (2) 207-217

Pavlicek A, Hrda S, Flegr J (1999). Short Communication 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 Genus *Frenkelia*. Folia Biologica (Praba) 45 (3) 97-99

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

Peng L, Zhou Y, Kong DY and Zhang WD (2010). Antitumor activities of dammarane triterpene saponins from *Bacopa monniera*. Phytotherapy Research 24 (6) 864-868

Pengelly BC, Maass BL (2001). *Lablab purpureus* (L.) Sweet–diversity, potential use and determination of a core collection of this multi-purpose tropical legume. Genetic Resources and Crop Evolution 48 (3) 261-272

Pérez-Rodríguez P, Riano-Pachon DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B (2009). PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Research (38(Database issue)) D822-D827

Phillips RD (1993). Starchy legumes in human nutrition, health and culture. Plant Foods for Human Nutrition 44 (3) 195-211

Phumichai C, Phumichai T, Wongkaew A (2015). Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild Hevea rubber. Plant Molecular Biology Reporter (33) 1486-1498

Pichersky E, Gang DR (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Science 5 (10) 439-445

Pirman T, Stibilj V, Stekar JM, Combe E (2001). Amino acid composition of beans and lentil. Zb Bioteh Fak Univ Ljubl, Kmet Zooteh 78 (1) 57-68

Poornachandra MN, Khanam S, Shivananda BGTN, Shivananda TN, Dris R (2005). *Mucuna prunensi* (LDDC)- A novel drug for learning and memory retrieval. Journal of Food, Agriculture and Environment 3 (3&4) 13-15

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1 (7) 215-222

Prevost A, Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theoretical and applied Genetics 98 (1) 107-112

Pugalenthi M, Vadivel V (2007a). Agrobiodiversity of eleven accessions of *Mucuna pruriens* (L.) DC. var. *utilis* (Wall. ex Wight) Baker ex Burck (velvet bean) collected

from four districts of South India. Genetic Resources and Crop Evolution 54 (5) 1117-1124

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

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

Purseglove JW (1987). Tropical crop: dicotyledons, Logman, New York, USA, p. 719.

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

Rahi JS, Sripathi S, Gilbert CE, Foster A (1995). Childhood blindness due to vitamin A deficiency in India: regional variations. Archives of Disease in Childhood 72 (4) 330-333

Rai N, Asati BS, Singh AK and Yadav DS (2006). Genetic variability, character association correlation and path coefficient analysis in between seed morphology and seeding growth in French bean. Journal of Assam Science Society (42) 40-43

Rai N, Kumar S, Singh RK, Rai KK, Tiwari GA, Kashyap SP, Singh MA, Rai AB (2016). Genetic diversity in Indian bean (*Lablab purpureus*) accessions as revealed by quantitative traits and cross-species transferable SSR markers. Indian Journal of Agricultural Sciences 86 (9) 654-660

Raina AP, Khatri R (2011). Quantitative determination of L-DOPA in seeds of *Mucuna pruriens* germplasm by high performance thin layer chromatography. Indian Journal of Pharmaceutical Sciences 73 (4) 459-462

Raj A, Stephens M, Pritchard JK (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197 (2) 573-589

Rajaram N, Janardhanan K (1991). The biochemical composition and nutritional potential of the tribal pulse, *Mucuna gigantea* (Willd) DC. Plant Foods for Human Nutrition 41 (1) 45-51

Rakoczy-Trojanowska M, Bolibok H (2004). Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. Cellular and Molecular Biology Letters 9 (2) 221-238

Rana JC, Sood S, Gupta A, Negi KS, Lal H (2014). Rice bean Variety VRB-3 (Him Shakti). Indian Journal of Genetics and Plant Breeding (74) 268

Ravindran V, Ravindran G (1988). Nutritional and anti-nutritional characteristics of *Mucuna* (*Mucuna utilis*) bean seeds. Journal of the Science of Food and Agriculture 46 (1) 71-79

Reddy V, Lerner E (2010). Plant cysteine proteases that evoke itch activate protease-activated receptors. British Journal of Dermatology 163 (3) 532-535

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M (2010). *De novo* assembly and analysis of RNA-seq data. Nature Methods 7 (11) 909-912

Robic A, Dubois C, Milan D, Gellin J (1994). A rapid method to isolate microsatellite markers from cosmid clones. Mammalian Genome 5 (3) 177-179

Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1) 139-140

Rodino AP, Lema M, Pérez-Barbeito M, Santalla M, De Ron AM (2007). Assessment of runner bean (*Phaseolus coccineus* L.) germplasm for tolerance to low temperature during early seedling growth. Euphytica 155 (1-2) 63-70

Roldàn-Ruiz I, Dendauw J, Van Bockstaele E, Depicker A, De Loose MA (2000). AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Molecular Breeding 6 (2) 125-134

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J (2003). Comparative analyses of potato expressed sequence tag libraries. Plant Physiology 131 (2) 419–429

Saldanha CJ (1996). Flora of Karnataka. ISBN 81-204-1040-8. (2) 1-282

Santalla M, Monteagudo AB, Gonzalez AM, De Ron AM (2004). Agronomical and quality traits of runner bean germplasm and implications for breeding. Euphytica (135) 205–215

Sasidharan N (2004). Biodiversity documentation for Kerala, part 6: flowering plants. KFRI hand book 17

Sathiyanarayanan L, Arulmozhi S (2007). *Mucuna pruriens* Linn.-A comprehensive review. Pharmacognosy Reviews 1 (1) 157

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

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

Sathyanarayana N, Pittala RK, Tripathi PK, Chopra R, Singh HR, Belamkar V, Bhardwaj PK, Doyle JJ, Egan AN (2017). Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *de novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers. BMC Genomics 18 (1) 409

Sathyanarayana N, Vikas PB, Rajesha R (2008). In vitro clonal propagation of *Mucuna pruriens* var. *utilis* and its evaluation of genetic stability through RAPD markers. African Journal of Biotechnology 7 (8) 973-980

Sato S, Isobe S, Asamizu E, Ohmido N, Kataoka R, Nakamura Y, Kaneko T, Sakurai N, Okumura K, Klimenko I, Sasamoto S (2005). Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). DNA Research 12 (5) 301-364

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T (2008). Genome structure of the legume, *Lotus japonicus*. DNA research 15 (4) 227-239

Schmidhuber J, Tubiello FN (2007). Global food security under climate change. Proceedings of the National Academy of Sciences 104 (50) 19703-19708

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D (2010). Genome sequence of the palaeopolyploid soybean. Nature 463 (7278) 178-183

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

Schulz MH, Zerbino DR, Vingron M, Birney E (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28 (8) 1086-1092

Schwember AR, Carrasco B, Gepts P (2017). Unraveling agronomic and genetic aspects of runner bean (*Phaseolus coccineus* L.). Field Crops Research (206) 86-94

Scott JM (1919). The velvet bean, Florida Agric. Expt. Stat, Bulletin 102, Gainsmeville, Florida, USA, pp. 43-58

Senan S, Kizhakayil D, Sasikumar B, Sheeja TE (2014). Methods for development of microsatellite markers: an overview. Notulae Scientia Biologicae (6) 1–13

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA (2010). RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biology 10 (1) 160

Sharma S (2014). Nutritional Quality, Functional Properties and Value Addition of Rice Bean (*Vigna umbellata*) (Doctoral dissertation, CSKHPKV Palampur).

Shelef O, Weisberg PJ, Provenza FD (2017). The value of native plants and local production in an era of global agriculture. Frontiers in Plant Science (8) 2069

Shendure J, Ji H (2008). Next-generation DNA sequencing. Nature Biotechnology (26) 1135–1145

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

Shin JH, Kwon SJ, Lee JK, Min HK, Kim NS (2006). Genetic diversity of maize kernel starch-synthesis genes with SNAPs. Genome (49) 1287–1296

Siddhuraju P, Becker K (2001). Effect of various domestic processing methods on antinutrients and in vitro protein and starch digestibility of two indigenous varieties of Indian tribal pulse, *Mucuna pruriens* var. *utilis*. Journal of Agricultural and Food Chemistry 49 (6) 3058-3067

Siddhuraju P, Vijayakumari K, Janardhanan K (1996). Chemical composition and protein quality of the little-known legume, velvet bean (*Mucuna pruriens* (L.) DC.). Journal of Agricultural and Food Chemistry 44 (9) 2636-2641

Singh AP, Sarkar S, Tripathi M, Rajender S (2013). *Mucuna 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

Siriwardhane D, Egawa Y, Tomooka N (1991). Cross-Compatibility of Cultivated Adzuki Bean (*Vigna angularis*) and Rice Bean (*V. umbellata*) with Their Wild Relatives. Plant Breeding 107 (4) 320-325

Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E (2004). A high-density admixture map for disease gene discovery in African Americans. The American Journal of Human Genetics 74 (5) 1001-1013

Somta P, Chankaew S, Rungnoi O, Srinives P (2011). Genetic diversity of the Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as assessed by SSR markers. Genome 54 (11) 898-910

Soowora A (1982). Medicinal plants in traditional medicine in West Africa. John Wiley and Sons 1st Edn. London, pp. 1-256

Spataro G, Tiranti B, Arcaleni P, Bellucci E, Attene G, Papa R, Zeuli PS, Negri V (2011). Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L. Theoretical and Applied Genetics 122 (7) 1281-1291

Sperling CR, King SR (1990). Andean tuber crops: worldwide potential. In Advances in new crops. Proceedings of the first national symposium'New crops: research, development, economics', Indianapolis, Indiana, USA, 23-26 October 1988. Timber Press, pp. 428-435

Srivastav PP, Das H, Prasad S (1990). Effect of roasting process variables on in-vitro protein digestibility of Bengalgram, maize and soybean. Food chemistry 35 (1) 31-37

Srivastava RP, Ali M (2004). Nutritional quality of common pulses. Indian Institute of Pulses Research.

Stamp P, Messmer R, Walter A (2012). Competitive underutilized crops will depend on the state funding of breeding programmes: an opinion on the example of Europe. Plant Breeding 131 (4) 461-464

Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J (2010). Adaptation genomics: The next generation. Trends in Ecology & Evolution (25) 705–712

Staub JE, Serquen FC, Gupta M (1996). Genetic markers, map construction, and their application in plant breeding. HortScience 31 (5) 729-741

Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ (2009). Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. Systematic Botany 34 (1) 115-128

Stein C (2000). Global burden of protein–energy malnutrition in the year 2000. Draft 1-17-2002, World Health Organisation. Global Burden of Disease

Stoddard FL, Balko C, Erskine W, Khan HR, Link W, Sarker A (2006). Screening techniques and sources of resistance to abiotic stresses in cool-season food legumes. Euphytica 147 (1-2) 167-186

Surget-Groba Y, Montoya-Burgos JI (2010). Optimization of *de novo* transcriptome assembly from next-generation sequencing data. Genome Research 20 (10) 1432-1440

Szabo NJ, Tebbett IR (2002). The chemistry and toxicity of *Mucuna* species. In Proceedings of international workshop on food and feed from *Mucuna*: current uses and the way forward'. (Eds FB Milton, M Eilitta, R Myhrman, LB Carew, RJ Carsky), pp. 120-141

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005). Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theoretical and Applied Genetics (110) 1324–1333

Taheri S, Lee Abdullah T, Yusop M, Hanafi M, Sahebi M, Azizi P, Shamshiri R (2018). Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules 23 (2) 399

Tanzi AS, Eagleton GE, Ho WK, Wong QN, Mayes S, Massawe F (2019). Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for food and nutritional security: synthesis of past research and future direction. Planta 1-21

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 (1-3) 93–122

Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, 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 (4) 505-513

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17 (16) 6463–6471

Tautz D, Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12 (10) 4127-4138

Tekalign A, Derera J, Sibiya J, Mumm RH (2018). Molecular analysis for genetic diversity and population structure of Ethiopian faba bean (*Vicia faba* L) accessions. Journal of Plant Biology and Crop Research (1) 1010

Thiel T, Michalek W, Varshney R, 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 (3) 411-422

Tian J, Isemura T, Kaga A, Vaughan DA, Tomooka N (2013). Genetic diversity of the rice bean (*Vigna umbellata*) genepool as assessed by SSR markers. Genome 56 (12) 717–727

Tilman D, Balzer C, Hill J, Befort BL (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences 108 (50) 20260-20264

Tomooka N, Vaughan D, Moss H, Maxted N (2002). The Asian *Vigna*: genus *Vigna* subgenus ceratotropis genetic resources. Springer Science & Business Media; 2012 Dec 6.

Tóth G, Gáspári Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research 10 (7) 967-981

Tracy SM, Coe HS (1918). Velvet beans. United States Department of Agriculture, Washington (DC, É.-U.). Farmers Bulletin (962)

Tresina PS, Daffodil ED, Lincy P, Mohan VR (2014). Assessment of biochemical composition and nutritional potential of three varieties of *Vigna radiata* (L.) Wilezek. Biolife 2 (2) 655-667

Tresina PS, Mohan VR (2011). Effect of gamma irradiation on physicochemical properties, proximate composition, vitamins and antinutritional factors of the tribal pulse *Vigna unguiculata* subsp. *unguiculata*. International Journal of Food Science and Technology 46 (8) 1739-1746

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

Tun YT, Yamaguchi H (2007). Phylogenetic relationship of wild and cultivated *Vigna* (Subgenus *Ceratotropis*, Fabaceae) from Myanmar based on sequence variations in non-coding regions of trnT-F. Breeding Science 57 (4) 271-280

UNICEF (2009). Tracking progress on child and maternal nutrition: a survival and development priority, November 2009.

United Nations (2017). United Nations World Population Prospects: the 2017 Revision (2017) (New York)

Urquiaga S, Boddey RM, Neves MCP (1999). A necessidade de uma revolução mais verde. In: Siqueira JO, Moreira FMS, Lopes AS, Guilherme LRG, Faquin V, Furtini Neto AE, Carvalho JG (Ed.). Inter-relação fertilidade, biologia do solo e nutrição de plantas. Lavras: SBCS. 175-181

Valverde BE (2003). Progress on Rottboellia cochinchinensis management. FAO Plant Production and Protection Paper 120 (Add. 1), pp. 67-79

Van Ooijen JW (2006). JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands (33)

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

Varshney RK, Graner A, Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23 (1) 48-55

Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Nigam SN, Saxena K, Vadez V, Sethy NK, Bhatia S, Aruna R, Gowda MC, Singh NK (2007). Molecular genetics and breeding of grain legume crops for the semi-arid tropics. In: Genomics Assisted Crop Improvement: Genomics Applications in Crops (2) 207-241

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T (2013). "Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement." Nature Biotechnology 31 (3) 240

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

Vom Endt D, Kijne JW, Memelink J (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry 61 (2) 107-114

Wang L, Do Kim K, Gao D, Chen H, Wang S, Lee S, Jackson SA, Cheng X (2016). Analysis of simple sequence repeats in rice bean (*Vigna umbellata*) using an SSR-enriched library. The Crop Journal 4 (1) 40-47

Wang LX, Chen HL, Bai P, Wu JX, Wang SH, Blair MW, Cheng XZ (2015). The transferability and polymorphism of mung bean SSR markers in rice bean germplasm. Molecular Breeding 35 (2) 77

Wang ML, Morris JB, Barkley NA, Dean RE, Jenkins TM, Pederson GA (2007). Evaluation of genetic diversity of the USDA *Lablab purpureus* germplasm collection using simple sequence repeat markers. The Journal of Horticultural Science and Biotechnology 82 (4) 571–578

Wang Z, Gerstein M, Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics (10) 57–63

Wanjekeche E, Wakasa V, Mureithi JG (2003). Effect of germination, alkaline and acid soaking and boiling on the nutritional value of mature and immature *Mucuna* (*Mucuna pruriens*) beans. Tropical and Subtropical Agroecosystems 1 (2/3) 183-192

Warrier PK, Nambiar VPK, Ramankutty C (1996). Indian medicinal plants, a compendium of 500 species, Orient Longman Limited, Madras, India, Vol 4.

Waterlow JC (1997). Protein-energy malnutrition: the nature and extent of theproblem. Clinical nutrition (16) 3-9

Watson JR (1922). Bunch velvet beans to control rot-knot. University of Florida Agricultural Experiment Station, Bulletin 163, Gainsmeville, Florida, USA

Weber JL (1990). Informativeness of human (dC-dA) n·(dG-dT) n polymorphisms. Genomics 7 (4) 524-530

West Jr KP (2002). Extent of vitamin A deficiency among preschool children and women of reproductive age. The Journal of nutrition. 132 (9) 2857S-2866S.

Whyte RO, Nilsson-Leissner G, Trumble HC (1955). Leguminosas en la agricultura, FAO, Estudios Agropecuarios Núm. 21, Roma, Italia, p. 344.

Wiersema JH, Leon B (2016). World economic plants: a standard reference. New York: CRC press; 2016.

Wilmot-Dear C (1987). A revision of *Mucuna* (Leguminosae: Phaseoleae) in the Indian subcontinent and Burma. Kew Bulletin 42 (1) 23-46

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

Wong QN, Massawe F, Mayes K, Blythe M, Mayes S (2014) Molecular genetic tools to support genetic improvement of winged bean (*Psophocarpus tetragonolobus*) for food and nutrition security. ISHS Acta Horticulturae1110.

World Health Organization. The State of Food Security and Nutrition in the World 2018: Building climate resilience for food security and nutrition. Food & Agriculture Org.; 2018 Sep 12.

Yang H, Tao Y, Zheng Z, Zhang Q, Zhou G, Sweetingham MW, Howieson JG, Li C (2013). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. PloS One 8 (5) e64799

Yang T, Fang L, Zhang X, Hu J, Bao S, Hao J, et al. (2015). High-Throughput Development of SSR Markers from Pea (*Pisum sativum* L.) Based on Next Generation Sequencing of a Purified Chinese Commercial Variety. PloS One 10 (10) e0139775

Yates SA, Swain MT, Hegarty MJ, Chernukin I, Lowe M, Allison GG, Ruttink T, Abberton MT, Jenkins G, Skøt L (2014). *De novo* assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics 15 (1) 453

Yeh FC, Yang R, Boyle T, Ye Z, Mao JX (1999). POPGENE, version 1.32: the user friendly software for population genetic analysis. In. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, AB, Canada.

You FM, Huo N, Gu YQ, Luo M-c, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9 (1) 1

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480 (7378) 520-524

Yundaeng C, Somta P, Amkul K, Kongjaimun A, Kaga A, Tomooka N (2018). Construction of genetic linkage map and genome dissection of domestication-related traits of moth bean (*Vigna aconitifolia*), a legume crop of arid areas. Molecular Genetics and Genomics 1-5

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, Bellin D, Pezzotti M, Delledonne M (2010). Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. Plant physiology 152 (4) 1787-1795

Zhan S, Huang J, Guo Q, Zhao Y, Li W, Miao X, Goldsmith MR, Li M, Huang Y (2009). An integrated genetic linkage map for silkworms with three parental combinations and its application to the mapping of single genes and QTL. BMC Genomics 10 (1) 389

Zhang G, Xu S, Mao W, Gong Y, Hu Q (2013). Development of EST-SSR markers to study genetic diversity in hyacinth bean (*Lablab purpureus* L.). Plant Omics 6 (4) 295

Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K (2004). Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. Bioinformatics 20 (7) 1081-1086

Nucleus
https://doi.org/10.1007/s13237-019-00276-1

ORIGINAL ARTICLE

Microsatellite analysis reveals low interpopulation differentiation in velvet bean (*Mucuna pruriens* var. *utilis*) of India

Pittala Ranjith Kumar[1] · Sai Sundeep[1] · N. Sathyanarayana[1]

Received: 1 March 2019 / Accepted: 17 May 2019

Abstract

Velvet bean (*Mucuna pruriens* var. *utilis*) offers significant economic benefits to tropical agriculture as food, fodder and medicinal crop; but suffers from poor breeding efforts. The genetic diversity and taxonomic confusion prevailing in this plant are not well investigated so far. In this study, we used a set of 43 transcriptome derived genic-microsatellite markers on a panel of 18 pan-India velvet bean accessions to determine the genomic composition and population structure. The results revealed low to moderate genetic diversity (average Ht = 0.25), low population differentiation (Fst: 0.05) and a high estimate of gene flow (Nm: 4.71). Extensive admixture and absence of clear-cut population structure characterize this set of collection. The results underscore the need for comprehensive pre-breeding efforts to broaden the genetic diversity of this important medicinal legume crop.

Keywords Microsatellite · Velvet bean · Population differentiation · Genetic diversity · Admixture

Introduction

Mucuna pruriens (L.) DC. is a self-pollinated tropical legume characterized by high N_2 fixing ability, biomass production and protein content [15]. It is also the key natural source of L-Dopa (3, 4-dihydroxy-L-phenylalanine)—a nonprotein amino acid used in treatment of Parkinson's disease [1, 4]. The plant is a native to China and eastern India where it was once cultivated as green manure cover crop [9, 16, 30]. It is now widely distributed all over the Indian sub-continent and beyond—both in wild and cultivated forms. In India, it is recognized by three distinct botanical varieties classified based on the nature of pod pubescence. The wild varieties viz., var. *pruriens* and var. *hirsuta* possesses abundant, long stinging pod hairs and are known by the name "itching bean"—the name evocative of itchy dermatitis they induce due to an allergic reaction caused by proteinaceous substance mucunain in the trichome hair [5]. The cultivated var. *utilis* possesses non-stinging, silky hairs and is commonly referred to as velvet bean. The economic benefits of the latter are dealt in greater detail in more recent review papers [7, 30]. From the taxonomic view point, dozen or so cultivated *Mucuna* species are reported so far in the genus *Mucuna* [28]. These include *M. deeringiana, M. utilis, M. pruriens, M. cochichinensis, M. nivea, M. capitata, M. hassjoo, M. diabolica,* and *M. aterrima* [19, 21, 30]. However, a more recent monograph on this genus [22] determined these as merely synonyms of *M. pruriens* var. *utilis*. Nonetheless, wide-ranging morphological variability coupled with numerous inter-varietal hybrids [1, 3] as well as genotypes introduced from earlier seed exchange programs pose significant challenge for the taxonomic identification and parental selection in velvet bean breeding programs even today. Thus, DNA profiling using molecular markers could help to determine the genomic composition in the germplasm collection and provide way forward for future genetic improvement studies in this crop [32]. Thus, we applied a set of 50 polymorphic genic-microsatellite markers derived from our recent transcriptome sequencing efforts to analyze genetic variation in velvet bean germplasm of India. Genic microsatellites were particularly chosen as they represent the coding region of the genome and thus provide a reliable estimate of genetic diversity.

✉ N. Sathyanarayana
nsathyanarayana@cus.ac.in

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

Published online: 25 May 2019

1

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

Patrush Lepcha, Pittala Ranjith Kumar and N. Sathyanarayana*

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

Abstract

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

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

1.1 Introduction

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

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

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

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

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