Genic-SSR-based genetic diversity and population structure analysis in a global germplasm collection highlights the African origin of winged bean (Psophocarpus tetragonolobus L.)

Kishor U. Tribhuvan
ICAR - Indian Institute of Agricultural Biotechnology
Twinkle Mishra
ICAR - Indian Institute of Agricultural Biotechnology
Samptirao Dilip
ICAR - Indian Institute of Agricultural Biotechnology
Avinash Panday
ICAR - Indian Institute of Agricultural Biotechnology
Sudhir Kumar
ICAR - Indian Institute of Agricultural Biotechnology
Jay Lal Mahto
Birsa Agricultural University
Hanuman Lal Raigar
NBPGR: National Bureau of Plant Genetic Resources
A. Pattanayak
ICAR - Indian Institute of Agricultural Biotechnology
Binay Kumar Singh (binaybio@gmail.com)
ICAR - Indian Institute of Agricultural Biotechnology

Research Article

Keywords: Fabaceae, Transcriptome, Molecular markers, Population study

Posted Date: March 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2672155/v1

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on June 12th, 2023. See the published version at https://doi.org/10.1007/s10722-023-01624-6.

Abstract

Winged bean (*Psophocarpus tetragonolobus* L.) is an underutilized legume of the family Fabaceae. We assembled 58,811 unigenes from the publicly available winged bean RNAseq data sets and discovered 4,107 perfect SSRs. Seventy-eight of the 166 SSRs amplified into a single band, of which 22 were polymorphic, in 79 germplasm accessions of winged bean constituting eight populations from India, Thailand, Nigeria, Indonesia, Taiwan, the Philippines, and Papua New Guinea. We found 60 alleles at the 22 polymorphic SSR loci, with a mean value of 2.73 per locus. With a mean of 0.36, the PIC values for the SSR loci ranged from 0.11 to 0.64. We recorded the maximum genetic diversity in advanced breeding lines (I = 0.67, He = 0.41) and the minimum genetic diversity in the germplasm accessions collected from Manipur (I = 0.44, He = 0.28), a north-eastern Indian state. The AMOVA analysis indicated that within-population variation was significantly higher (85%) than between-population variation (15%). The pairwise N_m values between the populations ranged between 0.69–3.41, indicating the varying level of gene flow between them. The analysis of the population structure based on the Bayesian model-based clustering algorithm revealed two distinct groups among the eight populations with different levels of introgression. The fuzzy clustering based on the Manhattan method also identified a similar number of groups, with 72% similarity between the two clustering methods. However, the Neighbour-Joining (NJ)-based clustering grouped all the accessions into four clusters. Nevertheless, all three clustering methods unanimously indicated that most African accessions tended to group, and their leftover members were spread across the hypothetical geographical populations, indirectly supporting the African origin of the winged bean.

1. Introduction

Winged bean [*Psophocarpus tetragonolobus* (L.) DC.] is an underutilized or orphan legume (NAS 1979) of the phaseoloid clade of the family Fabaceae (Lepcha et al. 2017). It is diploid (2n=2x=18) with an estimated 782 Mbp genome size (Bennett and Smith 1976). The winged bean is a short-day self-pollinated species, however, cross-pollination up to 16.0 percent may also occur (Erskine 1979, Sriwichai et al. 2022). It is a tropical legume growing almost exclusively in Southeast Asia and Papua New Guinea as an annual or a perennial twinning herb (Hymowitz and Boyd 1977). The winged bean crop matures in 4-5 months and can yield up to 14 quintals of dry seed and 115 quintal tubers per hectare, even with minimal management (Khan et al. 1977). Besides being high-yielding, it is a perfect cover crop due to its fast-growing and vining nature (Banerjee et al. 1984, Anugroho et al. 2010). It is also highly suitable for crop rotation or intercropping and an ideal crop for low-input self-reliant agricultural systems as it forms symbiotic associations with a broad-spectrum cowpea rhizobial group (Masefield 1961, Ikram and Broughton 1980, Iruthayathas and Vlassak 1982).

The winged bean plant is described as a 'supermarket on a stalk' as every part of the plant, including leaves, shoots, flowers, seeds, fruits, and tubers, are used in different cuisine in different parts of the world (Khan 1976, Haq 1982). It is nutritionally almost equivalent to soybean and is often regarded as the 'soybean of the tropics' (Kadam et al. 1984). Moreover, the winged bean has been used in traditional medicine and for cosmetic purposes (Stopp 1963, Perry and Metzger 1980, Shibata et al. 2011). Winged bean plants contain an exceptionally high amount of carbohydrates, vitamins, minerals, and fiber (Gillespie and Blagrove 1978, NAS 1981, Kortt 1983, Kadam et al. 1984, Henry et al. 1985, Mnembuka and Eggum 1985, Kantha and Erdman 1986). The crude protein content of winged bean seed (~34.0 %) is higher than cowpea (~22.5 %), pigeonpea (~22.0 %), and lima bean (~23.0 %) and similar to soybean (~35.0 %) (Makeri et al. 2017). The carbohydrate content of winged bean seed ranges from 23-40 percent and lipids from 15-20 percent, constituted by 30-40 percent saturated and 60-70 percent unsaturated fatty acids (Jaffe and Korte 1976, Ekpenyong and Borchers 1980, Varriano et al. 1983, Mohanty et al. 2015). Because of its high oxidative stability, solid fat content, and good thermal conductivity, the winged bean seed oil is considered superior to soybean oil (Makeri et al. 2016). Winged bean seed powder is a valuable flour that can be brewed to make a coffee-like drink (Oso and Ashafa 2021). Minerals such as calcium, iron, phosphorus, potassium, sulfur, sodium, magnesium, zinc, manganese, boron, copper, chromium, and strontium are also found abundantly in the winged bean. The winged bean tubers have ivory flesh and contain 12-19 percent protein and 1–4 percent fat (Lepcha et al. 2017).

The origin of the winged bean is still not established (Tanzi et al. 2019). Two alternate hypotheses exist. The first hypothesis proposes an African origin of the winged bean followed by its *in-situ* domestication or human-aided migration to the east or trans-domestication of an African progenitor along the borders of the Indian Ocean (Khan 1976, Harder 1992). This hypothesis is based on the chromosome number and karyotype similarity of the winged bean with four other African species of the genus (Tanzi et al. 2019, Pickersgill 1980). The alternate hypotheses suggest an eastern origin and consider the winged bean, which arose through allopatric speciation, as a species distinctly different from other African species of the genus (Verdcourt and Halliday 1978), although no progenitor was proposed. Yang et al. (2018), from their chloroplast genome and ITS studies, supported the view of Verdcourt and Halliday (1978) and suggested that the progenitor is either not yet discovered or lost. However, their study also established crossability between *P. tetragonolobus* and *P. scamdens* (African), indicating that they may be related.

The winged bean is a relatively less explored pulse crop. Masefield (1973) foresaw its agricultural potential, and the National Academy of Sciences (NAS) of the United States singled out its remarkable worth for special attention by agricultural scientists in 1974. Since then, winged bean research has received a significant impetus (NAS 1975). However, most research efforts were focused on assessing its suitability under different climatic and soil conditions and nutritional and chemical properties. Only minimal efforts were put into genome-based research, including developing molecular markers. Consequently, marker-based trait dissection and breeding of cultivars with high harvest index and other desirable traits like photoperiod insensitivity, determinate growth habits, stress tolerance, and improved nutritional quality could not be taken up (Lepcha et al. 2017). Furthermore, the absolute quantification of genetic diversity and population parameters, critical for developing effective conservation and sustainability strategies, could not be realized.

The allelic count is an important measure of a population's genetic diversity and evolutionary potential. However, it is less commonly used than observed (Ho) and expected (He) heterozygosity measures because of its dependence on sample number. Both Ho and He are sensitive to allele frequencies rather than allele numbers. A decrease in allelic richness (variety of alleles or allelic count rather than frequency of alleles) reduces the fitness of a population to adapt to a variety of environments (Fisher 1930). It has been shown in many studies that founder events and subsequent demographic expansions affect allelic richness more than heterozygosity (Leberg 1992, Luikart et al. 1998, Nei et al. 1975, Templeton 1980). Thus, some alleles lost in a founder event may not reduce heterozygosity but will undoubtedly reduce allelic richness.

Various clustering methods commonly used in genetic diversity studies follow a hard clustering method and group genotypes specifically to one cluster or another. On the other hand, fuzzy clustering allows each individual to be a member of all statistically appropriate clusters rather than forcing it to any specific cluster. The method helps identify an individual's similarities (or dissimilarities) with multiple clusters. Thus, the partitioning of an individual may be 0 (completely fuzzy) to 1 (firm membership to a cluster). Dunn's partitioning coefficient (Fc(U), Dunn 1974) and Kaufman's partitioning coefficient (DC(U), Kaufman and Rousseeuw 1990) together help to identify an appropriate number of clusters. It is also possible to decide the number of clusters from a combination of silhouette value (highest) and Dc(U) value (lowest). In a genetic diversity study, it is crucial that a grouping done by one method is confirmed by other methods to avoid errors and identify unique germplasm, if any (Pattanayak et al. 2018).

In the last 15 years, several genome and transcriptome sequences have been published for underutilized legumes, including the winged bean, and these sequences are increasingly used for developing molecular markers (Varshney et al. 2012, Chapman 2015, Vatanparast et al. 2016). We developed a large set of genic-SSR markers from the publicly available winged bean RNAseq data sets. Moreover, we analyzed the diversity and population structure of a winged bean germplasm collection of 79 accessions from seven countries and shed some light on the origin of the winged bean using various clustering methods.

2. Material And Methods

2.1. Plant material

We used 79 germplasm accessions of winged bean provided by the Birsa Agricultural University (BAU), Ranchi (Jharkhand), India. The seeds of the germplasm accessions were originally collected/procured by the ICAR - National Bureau of Plant Genetic Resources, New Delhi, from seven different countries, including India, Thailand, Nigeria, Indonesia, Taiwan, the Philippines, and Papua New Guinea. The details of the germplasm accessions of the winged bean used in the study are summarized in Supplementary Table S1 and Fig. 1. The different plant parts of the winged bean are shown in Fig. 2.

2.2. De novo transcriptome assembly and mining of genic-SSRs

We acquired the publicly accessible raw RNAseq data sets (SRR5252646, SRR5252647, SRR5252648, and SRR5252649) from the NCBI database and employed FastQC software to assess the quality of the data. Trimmomatic v.0.32 software (Bolger et al. 2014) was used to remove the adaptor contamination and low-quality reads (QV < 20). The ambiguous reads (N > 5%) and reads smaller than 36 nucleotides were also removed. Trinity v.2.11.0 software (Grabherr et al. 2011) was used to assemble the clean reads *de novo* with the default k-mer, k = 25. The CD-HIT-EST v.4.6 software (Li and Godzik 2006) was used to eliminate the redundancies among the assembled transcripts with a global sequence identity cut-off value set at 90 percent. The unigenes were identified using Perl script "get_longest_isoform_seq_per_trinity_gene.pl". Krait v.1.33 software (Du et al. 2018) was employed to detect perfect SSRs and design primers from the unigene sequences. Only genic-SSR loci with simple sequences of 2–6 nucleotides repeated at least four times were chosen. We applied the following criteria to design the primers: primer length = 18–27 bases (optimal of 20 bases), GC content = 30–80 percent with primer GC clamp = 2, annealing temperature = 58–65°C (optimal 60°C), and the product size = 100–300 bp.

2.3. DNA extraction and PCR amplification of genic-SSRs

The young leaves of winged bean were used to extract the high molecular weight genomic DNA using the Cetyl Trimethyl Ammonium Bromide (CTAB) method following Murray and Thompson (1980). The RNA contamination of the DNA was removed by adding 2.0µl of RNase A (10 mg/ml, HiMedia) to 20 µl of DNA dissolved in TE buffer (Tris–EDTA, pH = 8.0), followed by incubation at 37ºC for 3–4 h. The quality of the purified DNA samples was analyzed on 1.0 percent agarose gel and quantified based on the absorbance at 260 nm in a NanoDrop™ One^C microvolume UV-Vis spectrophotometer (Thermo Scientific).The genic-SSRs were amplified in a C1000 Touch Thermal Cycler (Bio-Rad) with the cycling conditions 94ºC for 5 min followed by 35 cycles of 94ºC for 45 s, 60ºC for 30 s, 72ºC for 30 s, and a final extension at 72ºC for 7 min in 10µl PCR reaction mixtures containing 20 ng DNA, 1× PCR buffer, 1.5 µM $MgCl_2$, 200 µM dNTP mix, 250 nM forward and reverse primers, and 0.25 U Taq DNA polymerase. We procured all the PCR reagents from Invitrogen. The amplified products of genic-SSRs were resolved on 3.5 percent MetaPhor (FMC BioProducts, Rockland, ME, USA) agarose gel containing 0.5 ng/ml ethidium bromide (HiMedia). The gel images were captured on a Gel Doc™XR + Imaging System (Bio-Rad). The size of the amplicons was determined against a 50bp DNA ladder (Invitrogen).

2.4. Genetic diversity analysis

The summary statistics for all 22 polymorphic out of 78 single locus genic-SSR markers were calculated using the PowerMarker v.3.25 software (Liu and Muse 2005). The population genetic parameters and the pairwise PhiPT values [for gene flow estimation; $N_m = 0.25[(1/PhiPT)–1]$ were calculated using GenAlEx 6.5 software (Peakall and Smouse 2012). The MEGA 7 software (Kumar et al., 2016) was used to create the neighbor-joining dendrogram. The Ewens-Watterson test was conducted by Popgene v.1.31 software (Watterson 1977).

The entire germplasm set was divided into eight geographical distribution-based hypothetical populations. Allelic counts were calculated for all hypothetical populations based on their collection locations. To confirm the sample size neutrality of allelic counts, the permutation method, as described by Fu et al. (2003) and implemented in FP Test R software, was used. The permutation of alleles was repeated 10,000 times. The SSR data set was also used for fuzzy clustering using Manhattan distance (Gan et al. 2007). The optimum number of clusters was decided based on the lowest Dc(u) value and highest average silhouette values among clusters. Individual silhouette amount was considered for assessing the membership strength of a genotype to a cluster.

2.5. Population structure analysis

The Bayesian model-based clustering algorithm of the STRUCTURE v.2.3.4 software (Pritchard et al. 2000) was employed to analyze the population structure. The Markov Chain Monte Carlo (MCMC) runs were performed for each value of K from 2 to 10 to determine the optimal number of groups (K-value) in the germplasm collection. Each run included 10,000 burn-in periods and 100,000 MCMC replicates. The admixture model option was used for the analyses. The

germplasm accessions with a membership fraction of ≥ 0.8 were deemed pure, whereas the remainder were admixtures. We calculated the ΔK value using Evanno's method (Evanno et al. 2005), which was implemented in the STRUCTURE HARVESTER software.

3. Results

3.1. De novo transcriptome assembly and its evaluation

We obtained 21,794,816 clean reads after pooling the raw RNAseq data and their subsequent cleaning with Trimmomatic v.0.32 software to remove adaptors and low-quality reads. The *de novo* assembly of the clean reads using Trinity v.2.11.0 software generated 87,160 contigs (transcripts henceforth) containing 94,323,676 nucleotides. Using the Perl script "get_longest_isoform_seq_ per_trinity_gene.pl" in Trinity, we retrieved 58,811 unigenes totaling 53,467,440 nucleotides from these transcripts. The N50 values of transcripts and unigenes were 1,697 bp and 1,574 bp, and GC contents were 41.47 and 41,80 percent. The average lengths of transcripts and unigenes were 1,082.19 and 909.14 bp; their median lengths were 718 and 503 bp, and lengths were between 201–15,137 bp. Table 1 shows a summary of the transcriptome *de novo* assembly. A total of 61.3 percent of transcripts were in the range of 200–999 bp, 23.9 percent in 1,000–1,999 bp, 9.9 percent in 2,000–2,999 bp, 3.32 percent in 3,000–3,999 bp, 1.02 percent in 4,000–4,999 bp and 0.7 percent > 5,000 bp, while these counts for the unigenes were 71.89, 16.48, 7.45, 2.7, 0.9, and 0.58 percent respectively (Fig. 3). The quantitative assessment of the completeness of the transcriptome assembly using the Eukaryota BUSCO database yielded the following results: complete: 92.55% [complete and single copy: 88.63%, complete and duplicated: 3.92%], fragmented: 3.53%, missing: 3.92% and the total number of BUSCOs identified: 255, as detailed in Table 2.

Table 1
Summary of transcriptome *de novo* assembly in winged bean.

S. No.	Particulars	Transcripts	Unigenes
1.	Number of sequences	87,160	58,811
2.	Average length (bp)	1,082.19	909.14
3.	N50 (bp)	1,697	1,574
4.	Minimal length (bp)	201	201
5.	Maximal length (bp)	15,137	15,137
6.	Median length (bp)	718	503
7.	Total assembled bases	94,323,676	53,467,440
8.	GC content	41.47%	41.80%

Table 2
BUSCO analysis for assessing transcriptome assembly completeness with the Eukaryote lineage database (eukaryota_orthoDB9).

	Number of BUSCO units found
	Eukaryota_orthoDB9
Complete BUSCOs	236 (92.55%)
Complete and single-copy BUSCOs	226 (88.63%)
Complete and duplicated BUSCOs	10 (3.92%)
Fragmented BUSCOs	9 (3.53%)
Missing BUSCOs	10 (3.92%)
Total BUSCOs searched	**255 (100%)**

3.2. *Identification, analysis, and PCR validation of genic-SSRs*

Over 58,811 unigenes, we identified 4,107 perfect SSRs, averaging 0.07 SSR per unigene. Taking the overall length of the unigenes analyzed to find SSRs as 53,467,440 bp, each 13.02 kbp of the unigenes sequence had one SSR locus. We found 762 different types of repeat motifs among the 4,107 genic-SSR loci (Table 3). The most prevalent SSRs were trinucleotide SSRs (44.05%), followed by di-, hexa-, tetra-, and pentanucleotide SSRs (22.86%, 19.80%, 8.08%, and 5.21%) (Fig. 4a). The most prevalent motifs among the tri- and dinucleotide repeats, with frequencies of 4.7 and 8.7 percent respectively, were GAA/TTC and AG/CT. The simple sequences were found to reiterate 4–31 times, with frequencies ranging from 23.33 percent (n = 4) to < 1 percent (n > 12) (Table 3). The 15 bp SSRs were the most common (24.91%), followed by 24 bp (18.97%), 18 bp (12.61%), and 16 bp (10.32%) (Fig. 4b). The SSRs had a maximum length of 90 bp. We successfully designed PCR primers for all 4,107 genic-SSRs (Supplementary Table S2). We chose 166 loci at random (n ≥ 18) to evaluate their efficacy in assessing the genetic diversity and population structure of 79 germplasm accessions of winged bean, constituting eight populations from seven countries.

Seventy-eight SSR loci amplified to yield a single band in a panel of four genotypes of winged bean, namely AKWB1, IWB1, WB11, and WB13. Of these, 71 loci produced amplicons of the expected size, five produced amplicons larger than the expected size, and two produced amplicons less than the expected size (Supplementary Table S3).

Table 3 Frequency distribution of the ten most abundant genic-SSR repeat motifs in winged bean.

S. No.	Repeat motif	Number of reiterations of the motif																					
		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	2
1	AG/CT	-	-	-	117	57	31	32	30	24	5	12	11	7	7	6	4	2	3	6	-	1	-
2	GA/TC	-	-	-	77	56	36	28	22	31	9	5	5	6	5	2	4	1	1	1	1	1	1
3	GAA/TTC	-	97	38	25	18	2	2	2	1	2	1	1	2	1	-	-	-	-	1	-	-	-
4	AAG/CTT	-	88	34	23	16	3	-	2	1	1	-	-	-	1	-	-	-	-	-	-	1	-
5	AGA/TCT	-	65	35	12	9	6	5	2	1	3	1	2	-	1	-	-	-	-	-	-	-	-
6	AT	-	-	-	33	20	18	12	7	12	-	-	-	-	-	-	-	-	-	-	-	-	-
7	TA	-	-	-	32	26	17	7	4	8	-	-	-	-	-	-	-	-	-	-	-	-	-
8	AAC/GTT	-	39	18	13	1	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	AAT/ATT	-	36	20	6	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	CTC/GAG	-	31	22	10	5	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Other motifs*	958	901	326	185	110	26	17	9	2	5	-	1	-	1	-	-	0	-	-	-	-	-
	Total	958	1257	493	533	326	140	104	80	80	25	19	20	15	16	8	8	3	4	**8**	**1**	**3**	1

*Additional 752 types of repeat motifs in winged bean transcriptome

3.3. Polymorphism and frequency spectrum of genic-SSRs

All 78 genic-SSR loci producing a single band were used to analyze the genetic diversity among the 79 germplasm accessions of winged bean. However, only 22 of the 78 genic-SSRs were polymorphic, and 56 were monomorphic. The profile generated by the primer IIABWB278 in different winged bean germplasm accessions is shown in Fig. 5. In 79 germplasm accessions of winged bean, we found 60 alleles at 22 polymorphic genic-SSR loci. The number of alleles per locus ranged between 2 and 5, with a mean of 2.73. The effective number of alleles (Ne) per locus ranged between 0.14 to 3.28, with a mean of 1.89. The Shannon information index (I) of the genic-SSR loci ranged from 0.24 to 1.28, and the PIC values ranged from 0.11 to 0.64, with an average of 0.36. Among the 60 alleles, five alleles were rare (frequency < 0.05), 37 were common (frequency 0.05–0.5), and 18 were frequent (frequency > 0.5). Major allele frequencies per genic-SSR locus ranged from 0.42 to 0.94, with a mean of 0.68 (Table 4). The distribution analysis of the 60 alleles among the 79 germplasm accessions of winged bean constituting eight populations from seven different countries revealed that (1) 27 alleles were present in each of the eight populations, (2) the maximum number of alleles amplified in Population-I (advanced breeding lines) followed by Population-VIII (Papua New Guinea collection), (3) 56 alleles were common to both Indian (Population-I, II, III, and IV) and exotic (Population- V, VI, VII, and VIII) collections, (4) Two alleles (IIABWB2531: 160bp; IIABWB2090: 180bp) were specific to Indian (Population-I, II, III, and IV) collections while two alleles (IIABWB1767: 135bp; IIABWB2090: 170bp) were specific to exotic (Population- V, VI, VII, and VIII) collections,(5) One allele was specific to Population- III (Maharashtra collection) whereas one allele was specific to Population- II (Jharkhand collection) and Population- I (advanced breeding lines), and (6) Four alleles were rare in both the Indian (Population-I, II, III, and IV) and exotic (Population- V, VI, VII, and VIII) collections. The Ewens-Watterson test for the neutrality of markers performed on overall allele frequencies (Supplementary Table S4) showed that the F-value (sum of the square of allelic frequency) for three genic-SSR loci, namely IIABWB127, IIABWB163, and IIABWB410, was beyond the upper/lower limits of expected F values at 95 percent confidence level.

Table 4 Summary statistics of 22 polymorphic genic-SSR loci from 79 germplasm accessions of winged bean.

S. No.	SSR marker ID	Forward Primer	Reverse Primer	Expected amplicon size (bp)	Observed amplicon size (bp)	Number of alleles	Effective number of alleles (Ne)	Sh inf inc
1	IIABWB2	AATGTGAAGACTCCGCTGCC	GGAGATGATGACGCTGGAGG	156	156	3	1.33	0.4
2	IIABWB77	AGGAGAAAGAGAAACTGAGGAGC	CTTCACAGTTCGCAGTTCGC	125	125	2	1.61	0.5
3	IIABWB118	AGGAGAAATGGCATTAGCATCG	GATGGTAGAATTTGACGCGGG	134	134	4	1.75	0.7
4	IIABWB278	TCTGCTCAGAAATGGGTGCC	ACTAATGGCCCAGCAACTCC	100	290	3	1.84	0.8
5	IIABWB409	TGCTGTTTGAACCATCTTGGG	GACCAAACTCACCATGATGGC	145	145	2	1.95	0.6
6	IIABWB410	AGAAATGTGGAAGTTCAAATTAAAGCC	TGCGGTGTATAGAAAGTTTCACC	186	186	2	2.00	0.6
7	IIABWB970	AACAGCACCCACAATTTTTGG	AGAGGGGATATCTGAAACGGG	127	127	2	1.82	0.6
8	IIABWB1271	TCCTTGTTCAGTGGGTGTGG	GTGAGTTGAATTTGCCCCTGG	207	207	4	3.28	1.2
9	IIABWB1343	ACCAATGATAGACCAACTAGAAACG	TCACATGAAAAGGAGCCACC	107	107	2	1.89	0.6
10	IIABWB1403	AATGCAACAAAAACTATTGAAAGCC	CTCTGTTGCTCGTGCTTTCC	113	113	2	1.58	0.5
11	IIABWB1630	CCATCCCCACAATTGTCTCC	TTTGCTGTCTCCCTTGTTCC	141	141	2	2.00	0.6
12	IIABWB1767	CTGGTCGACAGCAGCAGG	AGAGCTGCATCGTCGTATCC	147	147	2	1.14	0.2
13	IIABWB2090	TGCCTACCACCCTTCAAAGC	CAAATCCTGGAGCTTGGAGC	197	197	5	2.22	1.0
14	IIABWB2422	GGCGAGTAGGCGTCATCC	CGAAACTCGGTGCATTCAGC	179	179	2	1.35	0.4
15	IIABWB2531	CTTGCTTCAAAGGCTCCTCG	AGCGAGCGGAAGATCTAACG	138	138	4	2.72	1.0
16	IIABWB2571	TCTATACGACCTTAAACCCTAGCC	GTCTAGCACCCTGTCTCAGC	123	215	2	1.79	0.6
17	IIABWB2944	CACCAACAGCGTCTTTCTTCC	GGGATCCGAAAGTAGAGCAGC	119	119	3	2.07	0.8
18	IIABWB3180	TCGTAATCTACAAACTGCAGACG	AAAAAGGGTGTAGCCGTTGC	195	195	3	2.42	0.9
19	IIABWB3213	GGAGGAGGAGGTTGTGTTCC	ATCTTGCCGCCTCAACATCC	122	122	4	2.89	1.2
20	IIABWB3849	GGCACTTTAATTTCTCACCCTAGC	AATTTCGGCAGCCCATTTGC	200	200	3	1.33	0.5
21	IIABWB4067	TCGTTTCCATCGACATCCCC	CGAGCTAACGAACTTGCACG	148	210	2	1.40	0.4
22	IIABWB4093	GTCACATTCATTGGTTTGTCTTTCG	TGAATTCAAACAGTGCAATAGGAGG	144	144	2	1.14	0.2
			Mean			**2.73**	**1.89**	**0.7**

3.4. Genetic diversity, differentiation, gene flow, and population structure analysis

Table 5 displays the various diversity indices calculated for each of the eight winged bean populations. The observed number of alleles (n) and the effective number of alleles (Ne) for the populations ranged from 1.86 to 2.55 and 1.59 to 1.90. The expected heterozygosity (He) varied from 0.28 to 0.41. The Population-I (advanced breeding lines)(I = 0.67, He = 0.41) displayed the maximum, while Population-IV (Manipur collection) (I = 0.44, He = 0.28) displayed the minimum genetic diversity. The extent of polymorphism of the genic-SSRs varied across the eight populations, ranging from 100 percent in Population-V (Thailand collection) to 59.09 percent in Population-IV (Manipur collection). The AMOVA (Table 6) revealed a minimal genetic variation (15%) across the eight populations. Nevertheless, the genetic variation between and within the populations was significant ($P < 0.001$). The results showed that genetic variation within populations was the primary source of the overall variation. The within-population variation was 85 percent, while the between-population variation was 15 percent. We observed a significant level of gene flow between Population-III (Maharashtra collection) and Population-VI (Nigeria collection) (0.28), followed by Population-IV (Manipur collection) and Population-VII (Indonesia collection) (0.26). The pairwise N_m values between the populations ranged from 0.69–3.41 (Supplementary Table S5). The STRUCTURE v.2.3.4 software was used to analyze the genotypic data. It revealed the maximum ΔK at K = 2, indicating that at least two groups existed among the eight populations (Fig. 6a). We have designated Groups I and II with green and red bars on the bar plot, with one bar for each accession. The inferred ancestry of individuals for different groups ranged from 0.039 to 0.99, with an average value of 0.49 in Group-I (red) and 0.51 in Group-II (green). Forty-one individuals had inferred ancestry (> 0.5) to Group-I and the rest to Group-II. However, considering a minimum of 80% membership proportion criterion to qualify for non-admixed, 39 individuals were non-admixed while 40 were admixed. The number of non-admixed individuals in Group-I and Group-II was 19 and 20, respectively (Fig. 6b).

Table 5 Summary statistics of genetic diversity parameters on the populations of winged bean calculated based on 22 polymorphic genic-SSR markers.

Population	n	Ne	I	He	P	P%	N	U
Advanced breeding lines	2.55 ± 0.17	1.90 ± 0.16	0.67 ± 0.07	0.41 ± 0.04	21	95.45	56	1
Jharkhand	2.46 ± 0.18	1.75 ± 0.11	0.63± 0.07	0.38 ± 0.04	20	90.91	54	1
Maharashtra	2.18 ± 0.19	1.62 ± 0.10	0.53 ± 0.07	0.33 ± 0.04	17	77.27	48	0
Manipur	1.86 ± 0.18	1.59 ± 0.13	0.44 ± 0.09	0.28 ± 0.05	13	59.09	41	0
Thailand	2.36 ± 0.12	1.69 ± 0.11	0.59 ± 0.06	0.36 ± 0.04	22	100.00	49	1
Nigeria	2.23 ± 0.13	1.60 ± 0.09	0.53 ± 0.06	0.33 ± 0.04	20	90.91	52	1
Indonesia	2.18 ± 0.14	1.64 ± 0.11	0.55 ± 0.06	0.34 ± 0.04	19	86.36	48	0
Papua New Guinea	2.50 ± 0.17	1.66 ± 0.12	0.57 ± 0.07	0.33 ± 0.04	21	95.45%	55	0

n = Observed number of alleles; Ne = Effective number of alleles; I = Shannon's information index; He = Expected heterozygosity; P = Number of polymorphic loci; P% = Polymorphism percentage; N = Total number of alleles; U = Number of unique alleles

Table 6
Summary of analysis of molecular variance (AMOVA).

Source	df	SS	MS	Estimated variance	%variance
Among populations	7	240.34	34.33	2.23	15
Within populations	71	910.39	12.82	12.82	85
Total	**78**	**1150.73**		**15.05**	**100**

3.5. Genetic relationships

The NJ - based clustering (Nei et al. 1983) grouped the 79 germplasm accessions of the winged bean into four main clusters designated Cluster -I, II, III, and IV (Fig. 7). We did observe significant clustering of geographically proximal accessions. Cluster -I and II mainly comprised the Indian and only a few exotic accessions, while Cluster-III and IV consisted only of exotic collections. Out of 32 germplasm accessions showing affiliation to Cluster-I, a majority (22) were Indian, while 10 were exotic accessions. Among the 22 Indian germplasm accessions affiliated to Cluster-I, Populations- II (Jharkhand collection), Population-III (Maharashtra collection), and Population-IV (Manipur collection), each contributed five accessions, and Population-I (advanced breeding lines) contributed seven. Among the ten exotic accessions affiliated with Cluster-I, nine were from Population-V (Thailand collection), while one was from the Philippines (Population-VII). Twenty-three accessions showed the affiliation to Cluster-II. The population-wise distribution of the accessions was Population-I (advanced breeding lines) (11), Population-II (Jharkhand collection) (7), Population- III (Maharashtra collection) (1), Population-V (Thailand collection) (2), and Population- VIII (Papua New Guinea) (2). Cluster- III comprised seven accessions from three populations, namely Population-VII (Indonesia collection) (4), Population- V (Thailand collection) (2), and Population-VIII (Papua New Guinea collection) (1). Cluster- IV contained 17 accessions from only two populations, namely Population- VI (Nigeria collection) (9) and Population- VIII (Papua New Guinea) (8). The NJ clustering, based on pairwise *Fst* values, grouped all eight populations in two distinct clusters. All exotic populations (Population - VI, VII, and VIII) showed their affiliation to a single cluster except the Thailand collection (Population -V), which clustered with the Indian populations (Population - I, II, III, and IV) (Fig. 8).

Allelic counts presented in Fig. 9 indicated that it was higher than expected in Population- VII, almost equal to expected in Population- I and II, less than expected in all others, and lowest in Population- IV. The test of significance (Fig. 10) showed that the difference in allelic counts (allelic diversity) between Population- IV and Population- VIII was significant. Fuzzy clustering showed two distinct clusters as decided by the highest average silhouette value and lowest Dc(U) value. The number of genotypes in each cluster was almost equal (39 and 40), although 24 of the 39 genotypes in Cluster- I showed very weak membership (Supplementary Table S6). Almost all the genotypes of Population- II (Jharkhand, India) belonged to this cluster, although their memberships were very weak. Population- I, which is a cross-derived population, showed the next higher membership to Cluster- I, which was not very unexpected. On the contrary, members of Cluster- II showed much better silhouette values. All genotypes of Population- VIII (except one) showed membership in this group, while there was only one member from Population- II (Supplementary Table S6). Membership was spread equally among other clusters.

A comparison of all three methods of genetic diversity assessment showed that there was 72 percent similarity between fuzzy clustering (Manhattan method) and Structure (Bayesian) analysis. Both methods identified two major groups in which the African accessions (Population- VI and VIII) were grouped. Similarly, in the neighbor-joining cluster analysis where four clusters were delineated, the African and Asian genotypes showed very distinct grouping (Cluster- I & IV) with very little cross presence. In all three analyses, the Taiwan genotype EC0142600 appeared to be distinct from the Indonesian germplasm and showed exactly opposite grouping from the Philippines germplasm.

4. Discussion

Winged bean (*P. tetragonolobus* L.) is a high-protein, high-oil orphan tropical legume ideal for low-input agricultural systems. It can potentially augment the rural economy and nutritional security and sustain the genetic resources to address present and future environmental challenges. However, appropriate policy and scientific interventions are required to mainstream this orphan crop into commercial agricultural systems. Using the publicly available RNAseq data sets, we developed genic-SSR markers to expedite breeding for modern varieties and devise effective conservation strategies. We also assessed the effectiveness of a set of randomly selected genic-SSR markers to discriminate the germplasm accessions of a global collection.

We assembled the publicly available RNAseq data sets into 87,160 transcripts and extracted 58,811 unigenes from the transcripts. The N50 values and the average lengths of the transcripts were 1,697 bp and 1,082.19 bp, while these values for the unigenes were 1,574 bp and 909.14 bp, respectively. The N50 values and the average lengths obtained for the transcripts and the unigenes were in congruence or better than several other reports (Garg et al. 2011; Zhang et al. 2012; Chen et al. 2015; Liu et al. 2016; Chakrabarti et al. 2016). We found 4,107 perfect SSRs with a frequency of one SSR for every 13.02 kbp of the unigenes, higher than reported earlier in winged bean (Wong et al. 2017) but considerably lower than in other legumes, namely pigeon pea (Dutta et al. 2011) and black gram (Souframanien and Reddy 2015). These variations may be due to the genome size differences of legumes, redundancy in the assembled unigenes, and the tools and parameters used to identify the SSRs. We identified 762 distinct repeat motifs among the 4,107 genic-SSRs. The trinucleotide repeat motifs were the most abundant (44.05%). Our results corroborate the earlier reports on SSR mining in winged bean transcriptome (Chapman 2015; Vatanparast et al. 2016; Singh et al. 2017; Wong et al. 2017). The abundance of the trinucleotide repeat motifs in the genic-SSRs is highly evident, as the numeral variation in these repeat motifs does not alter the reading frame of the SSR-containing transcripts. The most frequent trimeric and the dimeric repeat motifs GAA/TTC and AG/CT, respectively obtained in the study, agree with the studies by Vatanparast et al. (2016) and Wong et al. (2017) on the winged bean as well as studies made on other legumes including lotus, soybean, and medicago (Jayashree et al. 2006).

We successfully designed PCR primers for all the 4,107 genic-SSRs identified in the study using the Krait v.1.33 software and tested 166 primer pairs for randomly selected genic-SSRs for PCR amplification in a panel of four winged wean germplasm accessions. All the primer pairs yielded amplification; however, only 78 yielded a single amplicon. Of these, 71 were of the expected size, while five were more and two were less than the expected size. The introns present in the SSR-containing unigenes could cause the amplicon to be larger than expected. Errors in the unigene assembly also could lead to variations in the size of the amplicons (Singh et al. 2016). The allelic richness of a population is a key determinant of its long-term potential for adaptability and resilience to the changing environment. The mean value of genic-SSR alleles per locus (2.73) and the mean PIC value of the genic-SSR markers (0.36) recorded in our study were lower than the respective values of 3.9 and 0.50 obtained by Laosatit et al. (2022) in a study on 457 accessions of winged bean collected from Thailand using 14 polymorphic genic-SSR markers. However, the mean PIC value observed in this study was comparable to the earlier study by Wong et al. (2017) (0.38) using genic-SSR markers. We obtained a moderate number of alleles in the winged bean populations, with a prevalence of common alleles (71.7%) and only a few rare alleles, indicating a restricted genetic base. The rare alleles discovered in the study could be used to identify the genetic identity of the germplasm accession. The Ewens-Watterson test conducted for ascertaining the neutrality of markers showed that except for IIABWB127, IIABWB163, and IIABWB410, none of the tested loci was under selection.

The degree of genetic variation existing in a population forms the basis for devising its breeding and conservation strategies. We observed the maximum genetic diversity in Population-I (advanced breeding lines) (I = 0.67, He = 0.41) and the minimum in Population-IV (Manipur collection) (I = 0.44, He = 0.28). The ability of the genic-SSR markers to resolve genetic variability within and between populations confirms their applicability as effective markers for future research. The AMOVA revealed little genetic difference across the eight populations. It showed that within-population variation was more substantial than that between-population, and genetic variation within populations had the maximum contribution to the overall variation. The resolution of the genetic diversity among the populations by the newly developed genic-SSR markers substantiates their efficiency as effective markers for diversity studies. Although the winged bean is largely self-pollinated, outcrossing to 16.0 percent is also reported, which may be the reason for higher variation levels within populations. The magnitude of gene flow (pairwise N_m values) recorded in the study ranged between 0.69–3.41, indicating a wide variation in gene flow among the populations. The high pairwise values between Population - I (advanced breeding lines), Population - II (Jharkhand collection) (3.41), Population - V (Thailand collection) (2.99), and Population - VII (Indonesia collection) (2.52) indicated that the germplasm accessions might have served as parents for developing the advanced breeding lines. Gene flow across geographically distant populations can be attributed to multiple or diffuse origins, considerable crossbreeding, and/or seed transit between places.

Diversity assessment of a gene pool is more informative when overall allelic diversity (and not frequency of alleles) is targeted and evaluated using appropriate markers like SSRs, which are co-dominant (Fu et al. 2003; William et al. 1990). As suggested in various research reports (Fu et al. 2003 in oat; Russel et al. 2000 in barley), average genetic diversity (allelic counts in our case) is less sensitive to breeding/directional selection than allelic diversity at particular loci. Allelic counts in different populations studied here indicate that the average genetic diversity of the populations is still maintained (except for Population- IV) across continents. This is expected in a cultivated species with fewer breeding and selection efforts. Maintenance of expected allelic count in Population- I (crossbreed) indicates a low selection pressure. Overall, the allelic counts suggest that all the populations are still fit for adaptation to wider environmental variations.

The population structure analysis revealed that many accessions in each of the two groups inferred by the analysis had a high level of genetic admixing (possessing less than 80% of the ancestral genome), similar to the report by Sriwichai et al. (2022) based on the molecular characterization of Thailand and imported accessions of winged bean. They proposed that the high outcrossing rate (16.0%) in winged bean may be the primary source of genetic admixing. The other possible reasons for obtaining a large number of admixtures through structure analysis may be (1) the likelihood of crossing between the germplasm of diverse origins coming through international trade (Khan, 1976) or sharing seeds for international trials organized by the International Council for the Development of Underutilized Plants (ICDUP), (2) incorrect labeling of the accessions at the source, and (3) symbiotic association of winged bean to a broad range of nitrogen-fixing bacteria aiding its spread to non-conventional regions, where other legumes performed poorly. A more refined picture of genetic diversity among the germplasm accessions for planning further explorations and mining additional diversity could be obtained by high-throughput genotyping-by-sequencing. It may also help in identifying regions where additional diversity could be mined. The NJ-based clustering grouped the accessions into four clusters: I, II, III, and IV. Although genetic diversity is not expected to follow political boundaries, we did observe a limited population-wise clustering in the study. Cluster - I and II mainly comprised Indian and only a few exotic accessions, while Clusters III and IV consisted only of exotic collections. The NJ clustering, based on pairwise *Fst* values, grouped all eight populations in two distinct clusters. All exotic populations (Population - VI, VII, and VIII) showed their affiliation to a single cluster except the Thailand population (Population -V), which clustered with the Indian populations (Population - I, II, III, and IV). This is in

congruence with the earlier study by Wong et al. (2016), wherein the accessions from Indonesia, Nigeria, and Papua New Guinea showed close affiliations, and the study by Mohanty et al. (2013), wherein the accessions from India and Thailand showed significant closeness.

The fuzzy clustering showed that all the accessions from Population- VIII were associated with Cluster- II and members of this Cluster were spread across all populations except Population- II. When compared with Structure and Neighbour Joining-based clustering/grouping, we saw that most African accessions (in Population- VI and VIII) tended to group, and their leftover members were spread across different hypothetical geographical populations. The other two analysis methods did not confirm the unique clustering (in Cluster- I) of a part of Asian germplasm in the NJ method. These observations indicate that the African genotypes, although unique, are spread across other hypothetical geographic populations, indirectly supporting the hypothesis of the African origin of the winged bean (Khan 1976; Harder 1992). Population- II showed a very weak membership with Cluster- I (as observed from the silhouette amount). This Population requires further study as it appears to be different from other populations.

Conclusion

We assembled 58,811 unigenes from the publicly available winged bean transcriptomes and discovered 4,107 perfect SSRs. Moreover, we examined the genetic diversity and population structure of 79 germplasm accessions of the winged bean, representing eight populations from seven countries, including the likely centers of species diversity, using 22 polymorphic markers. We found a modest allelic richness and the predominance of common alleles in the populations, pointing to the conserved evolution of the species. The study indicated the need for more extensive exploration in Papua New Guinea, Nigeria, and Indonesia to augment the diversity in the collection. The population structure analysis showed a high level of genetic admixing, indicating extensive outcrossing among the accessions. It reiterates the need to take all appropriate measures to preserve the genetic integrity of the accessions. We observed a high within-population genetic diversity highlighting individual-level selection for genetically improving the species. Overall, our study augmented the available list of informative markers, provided a deeper insight into the genetic makeup of the species, and highlighted its possible African origin.

Declarations

Authors' contributions

All authors contributed to the study conception and design. BKS conceived the project and wrote the manuscript. BKS and KUT designed the experiments. BKS, KUT, TM, and SD performed the experiments. AP, KUT, TM, and SK analyzed the data. JLM and HLR provided the seed materials. AP coordinated the project and edited the manuscript. All the authors read and approved the manuscript.

Declaration of Competing Interest

The authors of this manuscript have no competing interests.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgments

We sincerely acknowledge the Director, ICAR - Indian Institute of Agricultural Biotechnology, Ranchi 834 003, Jharkhand, India, for providing financial support and research facilities. We are also grateful to the Vice Chancellor, Birsa Agricultural University (BAU), Ranchi 834 006, Jharkhand, India, and the Director, ICAR – National Bureau of Plant Genetic Resources, New Delhi 110 012, India, for providing the seed materials of winged bean.

References

1. Anugroho, F., Kitou, M., Kinjo, K., Kobashigawa, N., 2010. Growth and nutrient accumulation of winged bean and velvet bean as cover crops in a subtropical region. Plant Prod. Sci. 13, 360–366. https://doi.org/10.1626/pps.13.360.
2. Banerjee, A., Bagchi, D., Si, LK, 1984. Studies on the potential of winged bean as a multipurpose legume cover crop in tropical regions. Exp. Agric. 20(4), 297-301. https://doi.org/10.1017/S001447970001797X.
3. Bennett, M.D., Smith, J.B., 1976. Nuclear dna amounts in angiosperms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 274(933), 227–274. http://dx.doi.org/10.1098/rstb.1976.0044.
4. Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30 (15), 2114–2120. https://doi.org/10.1093/ bioinformatics/btu170.
5. Chakrabarti, M., Dinkins, R.D., Hunt, A.G., 2016. De novo transcriptome assembly and dynamic spatial gene expression analysis in red clover. Plant Genome. 9 (2). https://doi.org/10.3835/plantgenome2015.06.0048.
6. Chapman, M.A., 2015. Transcriptome sequencing and marker development for four underutilized legumes. Appl. Plant Sci. 3(2), apps.1400111. https://doi.org/10.3732/apps.1400111.
7. Chen, H., Wang, L., Wang, S., Liu, C., Blair, M.W., Cheng, X., 2015. Transcriptome sequencing of mung bean (*Vigna radiata* L.) genes and the identification of EST-SSR markers. PLoS ONE. 10, e0120273. https://doi.org/10.1371/journal.pone.0120273.
8. **Du, L., Zhang, C., Liu, Q., Zhang, X., Yue, B., 2018. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 34(4):681-683.** https://doi. org/10.1093/bioinformatics/btx665.

9. Dunn, J.C., 1974. Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems. J. Cybernet. 4, 1-15.
10. Dutta, S., Kumawat, G., Singh, B.P., Gupta, D.K., Singh, S., Dogra, V., Gaikwad, K., Sharma, T.R., Raje, R.S., Bandhopadhya, T.K., Datta, S., Singh, M.N., Bashasab, F., Kulwal, P., Wanjari, K.B., Varshney, R.K., Cook, D.R., Singh, N.K., 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]. BMC Plant Biology. 11, 17. http://www.biomedcentral.com/1471-2229/11/17.
11. Ekpenyong, T.E., Borchers, R.L., 1980. The fatty acid composition of the oil of the winged bean (*Psophocarpus tetragonolobus* L.) seeds. J. Am. Oil Chem. Soc. 57, 147–149. https://doi.org/10.1007/BF02673105.
12. Erskine, W., 1979. The exploitation of genetic diversity in the winged bean, *Psophocarpus tetragonolobus* (L.) DC. for grain yield. Dissertation, Cambridge University.
13. Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14 (8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.
14. Fisher, R., 1930. The genetical theory of natural selection. Oxford University Press.
15. Fu, Y.B., Peterson, G.W., Scoles, G., Rossnagel, B., Schoen, D.J., Richards, K.W., 2003. Allelic diversity changes in 96 Canadian oat cultivars released from 1886 – 2001. Crop Sci. 43, 1989-1995.
16. Gan, G., Ma, C., Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. SIAM, Society for Industrial and Applied Mathematics.
17. Garg, R., Patel, R., Tyagi, A., Jain, M., 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res. 18, 53–63. https://doi.org/10.1093/dnares/dsq028.
18. Gillespie, J., Blagrove, R., 1978. Isolation and composition of the seed globulins of winged bean, *Psophocarpus tetragonolobus* (L) DC. Aust. J. Plant Physiol. 5(3), 357-369. https://doi.org/10.1071/pp978 0357.
19. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., de Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29(7), 644–652. https://doi.org/10.1038/nbt.1883.
20. Haq, N., 1982. Germplasm resources, breeding and genetics of the winged bean. *Zeitschrift fur Pflanzenzuchtung - Journal of plant breeding.* 88 (1), 1-12.
21. Harder, D.K., Smartt, J., 1992. Further evidence on the origin of the cultivated winged bean, *Psophocarpus tetragonolobus* (L.) DC. (Fabaceae): chromosome numbers and the presence of a host-specific fungus. Econ. Bot. 46, 187–191. https: //doi.org/10.1007/BF029 30637.
22. Henry, C.J.K., Donachie, P.A., Rivers, J.P.W., 1985. The winged bean. Will the wonder crop be another crop? Ecol. Food Nutr. 16, 331–338. https://doi.org/10.1080/03670244.1985.9990872.
23. Hymowitz, T., Boyd, J., 1977. Origin, ethnobotany and agricultural potential of the winged bean- *Psophocarpus tetragonolobus*. Econ. Bot. 31(2), 180–188. https://doi.org/10.1007/BF02866589.
24. Ikram, A., Broughton, W.J., 1980. Rhizobia in tropical legumesVII. Effectiveness of different isolates on *Psophocarpus tetragonolobus* (L.) DC. Soil Biol. Biochem. 12, 7–82. https:// doi.org/10.1016/0038-0717(80)90105-4.
25. Iruthayathas, E.E., Vlassak, K., 1982. Symbiotic specificity in nodulation and nitrogen fixation between winged bean and Rhizobium. Sci. Hortic. (Amsterdam). 16, 313–322. https://doi.org/10.1016/0304- 4238(82)90030-9.
26. Jaffe, W.G., Korte, R., 1976. Nutritional characteristics of the winged bean in rats. Nutr. Rep. Int. 4(4), 449–455.
27. Jayashree, B., Punna, R., Prasad, P., Bantte, K., Hash, C.T., Chandra, S., Hoisington, D.A., Varshney, R.K., 2006. A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. In Silico Biol. 6(6), 607-620.
28. Kadam, S.S., Salunkhe, D.K., Luh, B.S., 1984. Winged bean in human nutrition. Crit. Rev. Food Sci. Nutr. 21, 1–40. https://doi.org/10.1080/10408398409527395.
29. Kantha, S.S., Erdman, J.W., 1986. Winged bean as a source of protein: recent advances, 206–219.
30. Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data—An Introduction to Cluster Analysis. John Wiley & Sons Inc., New York.
31. Khan, T.N., 1976. Papua New Guinea: A centre of genetic diversity in winged bean (*Psophocarpustetragonologus* (L.) DC.). Euphytica. 25, 693–705. https://doi.org/10.1007/BF00041608.
32. Khan, T.N., Bohn, J.C., Stevenson, R.A., 1977. Winged bean: cultivation in Papua New Guinea. World Crops. 29(5), 208–214.
33. Kortt, A.A., 1983. Comparative studies on the storage proteins and antinutritional factors from seeds of *Psophocarpus tetragonolobus* (L.) DC. from five Southeast Asian countries. Plant Foods Hum. Nutr. 33, 29–40. https://doi.org/10.1007/BF01093735.
34. Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7), 1870-4. https://doi.org/10.1093/molbev/msw054.
35. Laosatit, K., Amkul, K., Chankaew, S., Somta, P., 2022. Molecular genetic diversity of winged bean gene pool in Thailand assessed by SSR markers. Hortic. Plant J. 8(1), 81-88. https://doi.org/10.1016/j.hpj.2021.05.001.
36. Lepcha, P., Egan, A.N., Doyle, J.J., Sathyanarayana, N., 2017. A review on current status and future prospects of winged bean (*Psophocarpus tetragonolobus*) in tropical agriculture. Plant Foods Hum. Nutr. 72(3), 225-235. https://doi.org/10.1007/s11130-017-0627-0.
37. Leberg, P.L., 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution (NY). 46, 477–494.
38. Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
39. Liu, C., Fan, B., Cao, Z., Su, Q., Wang, Y., Zhang, Z., Wu, J., Tian, J., 2016. A deep sequencing analysis of transcriptomes and the development of EST-SSR markers in mungbean (*Vigna radiata*). J. Genet. 95, 527–535. https://doi.org/10.1007/s12041-016-0663-9.

40. Liu, K., Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21(9), 2128–2129. https://doi.org/10.1093/ bioinformatics/bti282.
41. Luikart, G., Allendorf, F., Cornuet, J., Sherwin, W.B., 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89, 238–247.
42. Makeri, M.U., Abdulmannan, F., Ilowefah, M.A., Chiemela, C., Bala, S.M., Muhammad, K., 2017. Comparative physiochemical, functional and structural characteristics of winged bean [*Psophocarpus tetragonolobus* DC] and soybean [*Glycine max*] protein isolates. J. Food Meas. Charact. 11, 835-846. https://doi.org/10.1007/s11694-016-9455-4.
43. Makeri, M.U., Karim, R., Adbulkarim, M.S., Ghazali, H.M., Miskandar, M.S., Muhammad, K., 2016. Comparative analysis of the physiochemical, thermal, and oxidative properties of winged bean and soybean oils. Int. J. Food Prop. 19(12), 2769–2787. https://doi.org/10.1080/10942912.2015.1031246.
44. Masefeld, G.B., 1961. Root nodulation and agricultural potential of the leguminous genus Psophocarpus. Trop. Agric. (Trinidad Tobago). 38, 225–229.
45. Masefield, G.B., 1973. *Psophocarpus tetragonolobus* - a crop with a future. Field Crop Abstracts. 26, 157-160.
46. Mnembuka, B.V., Eggum, B.O., 1995. Comparative nutritive value of winged bean (*Psophocarpus tetragonolobus* (L) DC) and other legumes grown in Tanzania. Plant Foods Hum. Nutr. 47, 333–339. https://doi.org/10.1007/BF01088271.
47. Mohanty, C.S., Pradhan, R.C., Singh, V., Singh, N., Pattanayak, R., Prakash, O., Chanotiya, C.S., Rout, PK, 2015. Physicochemical analysis of *Psophocarpus tetragonolobus* (L.) DC seeds with fatty acids and total lipids compositions. J. Food Sci. Technol. 52(6), 3660– 3670. https://doi.org/10.1007/s13197-014-1436-1.
48. Mohanty, C.S., Verma, S., Singh, V., Khan, S., Gaur, P., Gupta, P., Nizar, M.A., Dikshit, N., Pattanayak, R., Shukla, A., Niranjan, A., Sahu, N., Behra, S.K., Rana, T.S., 2013. Characterization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) based on molecular, chemical and physiological parameters. Am. J. Mol. Biol. 3, 187–197. https://doi.org/10.4236/ajmb.2013.34025.
49. Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8(19), 4321–4326. https://doi.org/10.1093/nar/ 8.19.4321.
50. NAS 1975. The winged bean: a high-protein crop for the humid tropics. National Academy of Sciences, Washington, DC.
51. NAS 1979. Tropical Legumes: Resources for the Future. Report of an Ad Hoc Committee on Technology Innovation, Board on Science and Technology for International Development, Commission on International Relations, and National Research Council. National Academy of Sciences, Washington, DC.
52. NAS 1981. The winged bean: a high-protein crop for the humid tropics, 2nd Edition, National Academy of Sciences, Washington, DC.
53. Nei, M., Maruyama, T., Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. Evolution (NY). 29, 1–10.
54. Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19(2), 153–170. https://doi.org/10.1007/BF02300753.
55. Oso, A.A., Ashafa, A.O., 2021. Nutritional composition of grain and seed proteins, in: Jimenez-Lopez, J.C., (Eds.), Grain and Seed Proteins Functionality. IntechOpen. https://doi.org/10.5772/intechopen.97878.
56. Pattanayak, A., Roy, S., Salej, S., Banshanlang, I., Banerjee, A., Gupta, S., Joshi, D., 2019. Rice bean: a lesser known pulse with well-recognized potential. Planta. 250. https://doi.org/10.1007/s00425-019-03196-1.
57. Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics. 28(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.
58. Perry, L.M., Metzger, J., 1980. Medicinal plants of east and southeast Asia: attributed properties and uses. MIT Press, Cambridge.
59. Pickersgill, B., 1980. Cytology of two species of winged bean, *Psophocarpus tetragonolobus* (L.) DC. and *P. scandens* (Endl.) Verdc. (Leguminosae). Bot. J. Linn. Soc. 80, 279–291. https ://doi. org/10.1111/j.1095-8339.1980.tb016 65.x
60. Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2), 945–959.
61. Russell, J.R., Ellis, R.P., Thomas, W.T.B., Waugh, R., Provan, J, Booth, A., Fuller, J., Lawrence, P., Young, G., Powell, W., 2000. A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol. Breed. 6, 553–568.
62. Shibata, M., Ishimatsu-Tsuji, Y., Yokoo, M., Nakai, Y., Abe, K., Muta, K., 2011. Effect of oral intake of winged bean extract on a skin lichenification model: evaluation by microarray analysis. Biofactors. 37(6), 421-428. http://dx. doi.org/10.1002/biof.182.
63. Singh, B.K., Mishra, D.C., Yadav, S., Ambawat, S., Vaidya, E., Tribhuvan, K.U., Kumar, A., Kumar, S., Kumar, S., Chaturvedi, K.K., Rani, R., Yadav, P., Rai, A., Rai, P. K., Singh, V.V., Singh, D., 2016. Identification, characterization, validation and cross-species amplification of genic-SSRs in Indian mustard (*Brassica juncea*). J. Plant Biochem. Biotechnol. 25 (4), 410–420. https://doi.org/10.1007/s13562- 016-0353-y.
64. Singh, V., Goel, R., Pande, V.. Asif, M.H., Mohanty, C.S., 2017. *De novo* sequencing and comparative analysis of leaf transcriptomes of diverse condensed tannin-containing lines of underutilized *Psophocarpus tetragonolobus* (L.) DC. Sci. Rep. 7, 44733. https://doi.org/10.1038/srep44733.
65. Souframanien, J., Reddy, K.S., 2015. De novo assembly, characterization of immature seed transcriptome and development of genic-SSR markers in black gram [*Vigna mungo* (L.) Hepper]. PLoS ONE. 10(6), e0128748. https://doi.org/10.1371/journal. pone.0128748.
66. Sriwichai, S., Laosatit, K., Monkham, T., Sanitchon, J., Jogloy, S., Chankaew, S., 2022. Genetic diversity of domestic (Thai) and imported winged bean [*Psophocarpus tetragonolobus* (L.) DC.] cultivars assessed by morphological traits and microsatellite markers. Ann. Agric. Sci. 67 (1), 34-41. https://doi.org/10.1016/j.aoas.2022.04.002.
67. Stopp, K., 1963. Medicinal plants of the Mt. Hagen people (Mbowamb) in New Guinea. Econ. Bot. 17, 16–22. https://doi.org/10.1007/BF02985349.
68. Tanzi, A.S., Eagleton, G.E., Ho, W.K., Wong, Q.N., Mayes, S., Massawe, F., 2019. Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for food and nutritional security: synthesis of past research and future direction. Planta. 250, 911-931.

69. Varriano-Marston, E., Beleia, A., Lai, C.C., 1983. Structural characteristics and fatty acid composition of *Psophocarpus tetragonolobus* seeds. Ann. Bot. 51(5), 631–640.
70. Varshney, R., Chen, W., Li, Y. et al., 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. 30, 83–89. https://doi.org/10.1038/nbt.2022.
71. Vatanparast, M., Shetty, P., Chopra, R., Doyle, J.J., Sathyanarayana, N., Egan, A.N., 2016. Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*, Leguminosae). Sci. Rep. 6, 29070. https://doi.org/10.1038/srep29070.
72. Verdcourt, B., Halliday, P., 1978. A revision of *Psophocarpus* (*Leguminosae-Papilionoideae -Phaseoleae*). Kew Bull. 33, 191. https://doi.org/10.2307/4109575.
73. Watterson, G.A., 1977. Heterosis or neutrality? Genetics. 85, 789-814.
74. Wong, Q.N., Massawe, F., Mayes, K., Blythe, M., Mayes, S., 2016. Molecular genetic tools to support genetic improvement of winged bean (*Psophocarpus tetragonolobus*) for food and nutrition security. Acta Hortic. 1110. https://doi.org/10.17660/ActaHortic.2016.1110.1.
75. Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18, 6531–6535.
76. Wong, Q.N., Tanzi, A.S., Ho, W.K., 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. https://doi.org/10.3390/genes8030100.
77. Yang, S., Grall, A., Chapman, M.A., 2018. Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.), a multipurpose underutilized legume. Am. J. Bot. 105, 888–897. https :// doi.org/10.1002/ajb2.1093.
78. Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Zhang, Q., Liang, X., Li, Y., 2012. De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.) BMC Genomics. 13, 90. https://doi.org/10.1186/1471-2164-13-90.

Figures

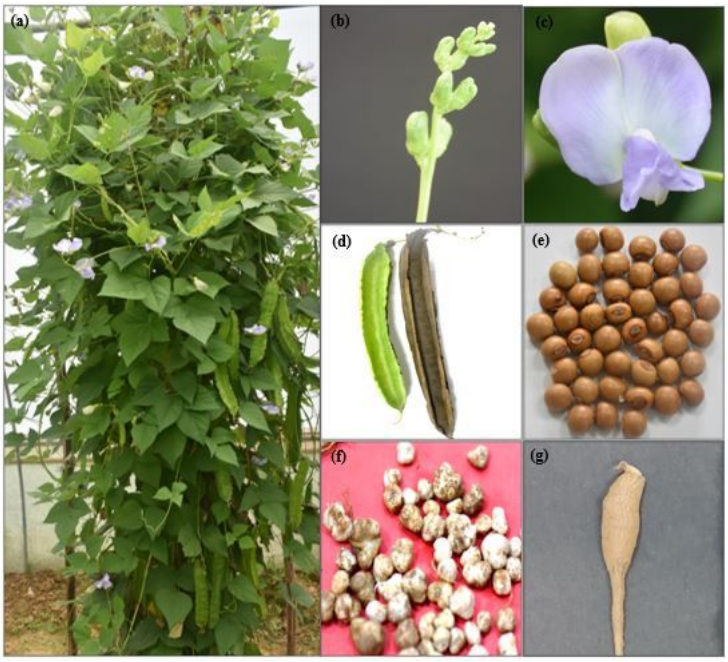

Figure 2

Photographs of winged been showing (a) whole plant at the fruiting stage, (b) inflorescence, (c) opened flower, (d) pods at near-maturity and matured stages, (e) matured seeds, (f) root nodules, and (g) root tuber.

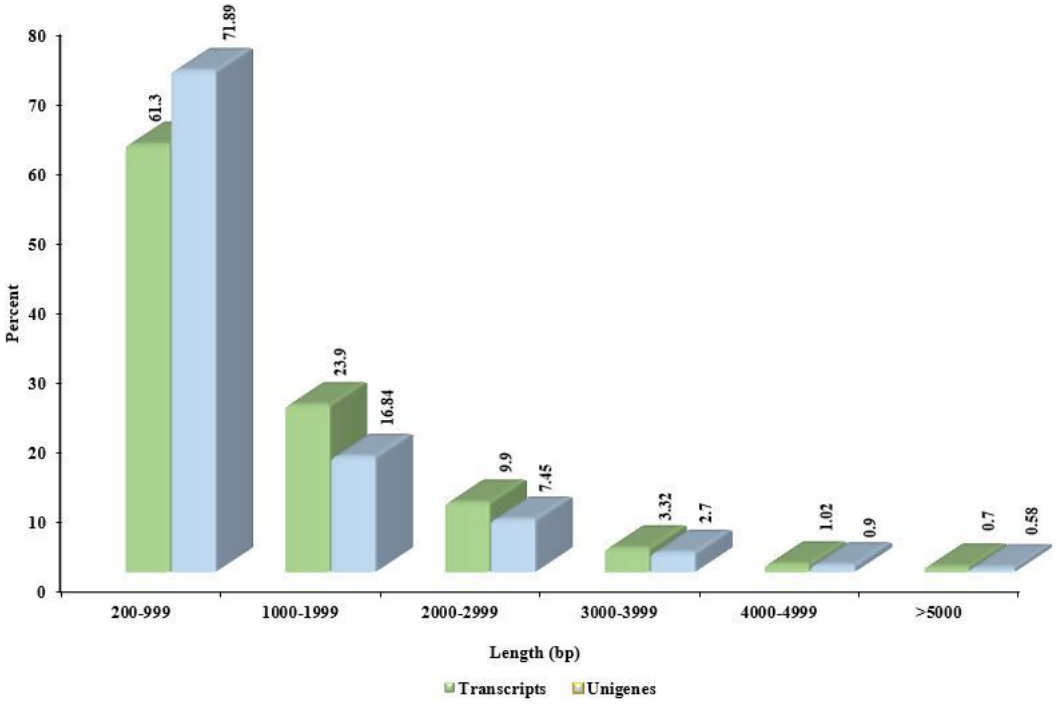

Figure 3

Length-wise distribution frequency of assembled transcripts and unigenes.

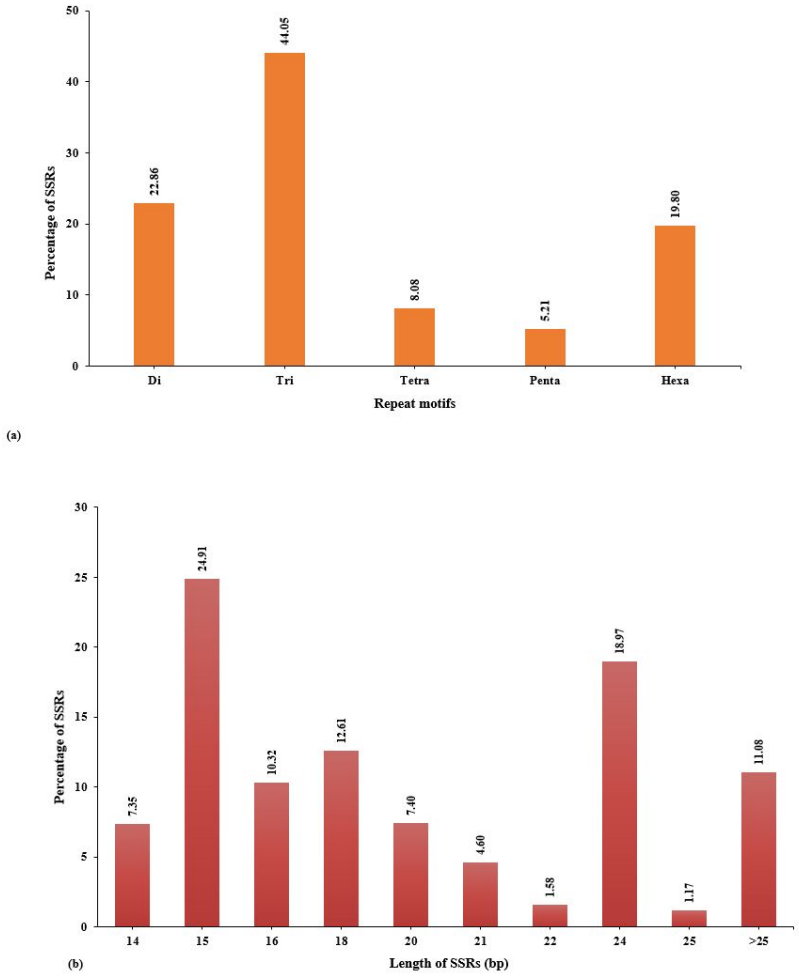

Figure 4

Occurrence and distribution of genic-SSRs of winged bean (a) Length of repeat motifs, (b) Length of SSRs.

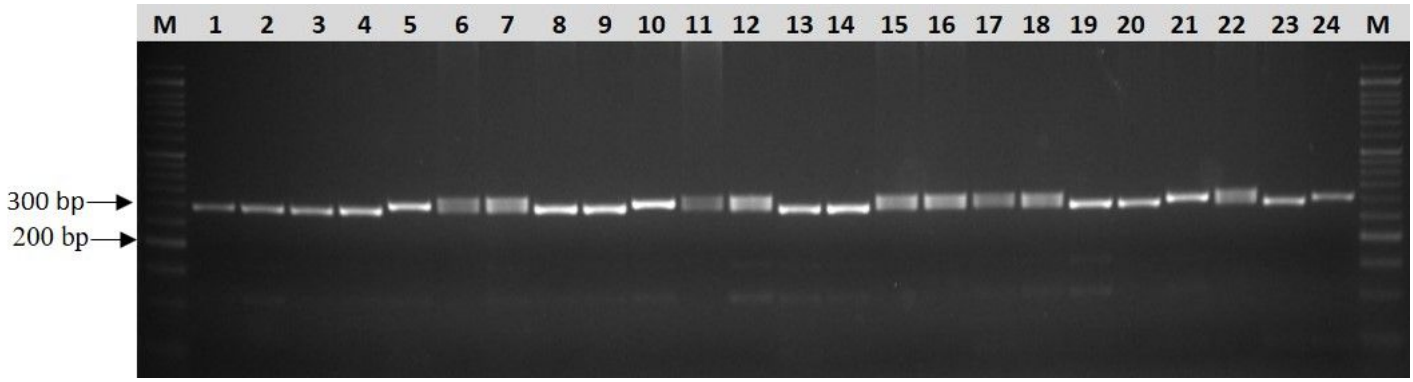

Figure 5

Polymorphism analysis of IIABWB278 in different winged bean germplasm accessions. (1) EC0142666, (2) EC0178276, (3) RWBGP51, (4) EC0178271, (5) EC0251035, (6) IC095222, (7) PURPLE2, (8) RWBGP64, (9) RWBGP48, (10) RWBGP54, (11) RWBGP62, (12) EC0118345, (13) RWBGP75, (14) RWBGP71, (15) EC0142661, (16) RWBGP70, (17) EC0178269, (18) RMBWB1, (19) RWBGP53, (20) RWBGP68, (21) EC0038823, (22) RWBGP57, (23) EC038957, (24) EC0142654. M = 50 bp DNA marker.

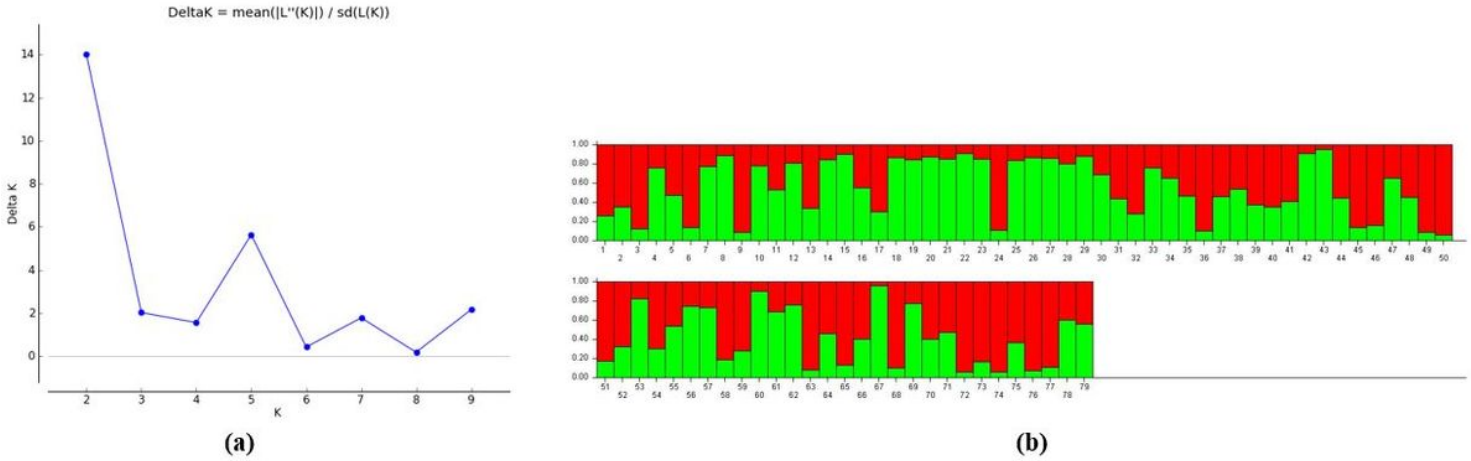

Figure 6

Population genetic structure based on 22 polymorphic genic-SSR markers in 79 germplasm accessions of winged bean. (a) ΔK graph showing peak value at K = 2, (b) population structure at ΔK = 2.

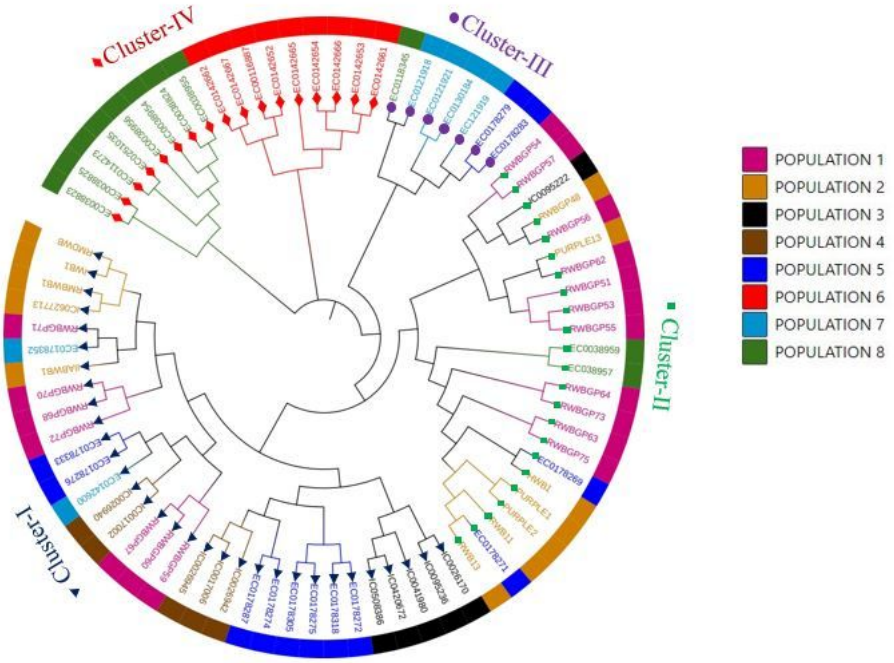

Figure 7

NJ tree showing the genetic relationships among the 79 germplasm accessions of winged bean.

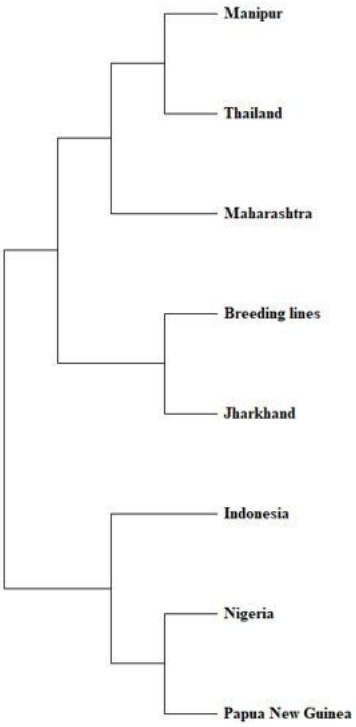

Figure 8

Pairwise Fst value-based NJ tree showing the genetic relationships among the eight populations of winged bean.

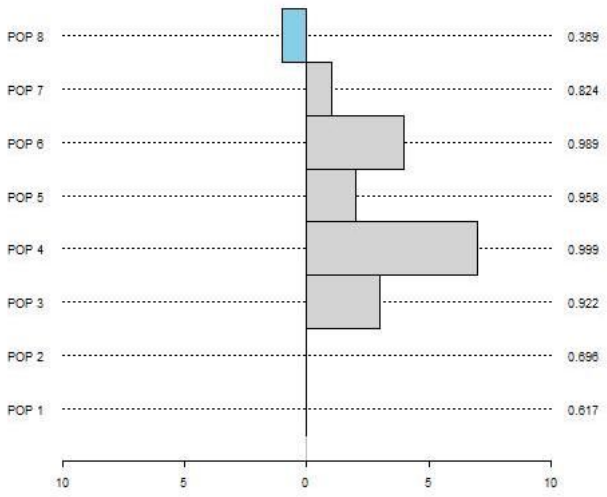

Figure 9

The difference between the observed (left) and expected (right) allelic counts in each population.

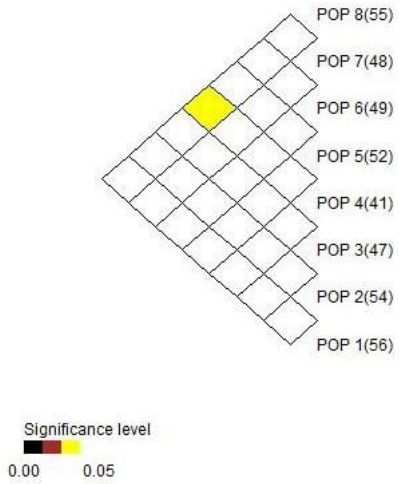

Figure 10

Test of allelic differences among populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS2.xlsx
- SupplementaryTableS4.docx
- SupplementaryTableS5.docx
- SupplementaryTableS6.docx