

RESEARCH ARTICLE

Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing

Wenjin Su^{1,2}, Lianjun Wang¹, Jian Lei¹, Shasha Chai¹, Yi Liu³, Yuanyuan Yang¹, Xinsun Yang^{1*}, Chunhai Jiao^{4*}

1 Food Crops Institute, Hubei Academy of Agricultural Sciences, Wuhan, China, **2** Beijing Key Laboratory of Crop Genetic Improvement/Laboratory of Crop Heterosis and Utilization, Ministry of Education, China Agricultural University, Beijing, China, **3** College of Agriculture, Yangtze University, Jingzhou, China, **4** Hubei Academy of Agricultural Sciences, Wuhan, China

* yangxins013@163.com (XY); jiaoch@hotmail.com (CJ)



OPEN ACCESS

Citation: Su W, Wang L, Lei J, Chai S, Liu Y, Yang Y, et al. (2017) Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. PLoS ONE 12(2): e0172066. doi:10.1371/journal.pone.0172066

Editor: Genlou Sun, Saint Mary's University, CANADA

Received: October 18, 2016

Accepted: January 30, 2017

Published: February 10, 2017

Copyright: © 2017 Su et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Hubei Engineering and Technology Research Center of Sweet Potato, Introduction and Utilization of Sweet Potato Germplasm Resources from America (2013-Z61), Science and Technology Innovation Center of Hubei Academy of Agricultural Sciences (2007-620-001-03), China Agriculture Research

Abstract

Sweet potato, *Ipomoea batatas* (L.) Lam., is an important food crop that is cultivated worldwide. However, no genome-wide assessment of the genetic diversity of sweet potato has been reported to date. In the present study, the population structure and genetic diversity of 197 sweet potato accessions most of which were from China were assessed using 62,363 SNPs. A model-based structure analysis divided the accessions into three groups: group 1, group 2 and group 3. The genetic relationships among the accessions were evaluated using a phylogenetic tree, which clustered all the accessions into three major groups. A principal component analysis (PCA) showed that the accessions were distributed according to their population structure. The mean genetic distance among accessions ranged from 0.290 for group 1 to 0.311 for group 3, and the mean polymorphic information content (PIC) ranged from 0.232 for group 1 to 0.251 for group 3. The mean minor allele frequency (MAF) ranged from 0.207 for group 1 to 0.222 for group 3. Analysis of molecular variance (AMOVA) showed that the maximum diversity was within accessions (89.569%). Using CoreHunter software, a core set of 39 accessions was obtained, which accounted for approximately 19.8% of the total collection. The core germplasm set of sweet potato developed will be a valuable resource for future sweet potato improvement strategies.

Introduction

Sweet potato (*Ipomoea batatas* (L.) Lam.) is a crop of considerable economic and social importance in developing countries [1]. Because of its high productivity and abundant protein, calorie and vitamin contents, it plays a key role in alleviating hunger and malnutrition in

System (CARS-11-C-15), Characteristic Discipline of Hubei Academy of Agricultural Sciences (2015TSXK06), National Natural Science Foundation of China Grant (31301377), Innovation and Application of Major Crops Germplasm Resources Introduced from Africa and Southeast Asia (2016ABA087). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

impoverished areas [2]. In China, sweet potato has been ranked fourth in importance among staple crops; in 2010, it was grown on approximately 3.65 million ha, with a total annual yield of 21.26 t.ha⁻¹, occupying approximately 45.1% of the worldwide sweet potato planted area and accounting for approximately 75.3% of the worldwide total annual sweet potato production [3].

Genetic resources are of paramount importance for crop improvement. Studies from archaeology, linguistics, history and biotechnology indicate that Central and South America is the primary center of sweet potato diversity, with East Africa, Asia and Oceania suggested as the secondary centers of diversity [4,5]. Over time, natural hybridization and selection have resulted in the evolution of different kinds of native sweet potato cultivars and a magnitude gene pool is reserved in China [6]. In the 1970s, China made significant progress in sweet potato breeding, which led to the release of many cultivars, including the excellent cultivar Xushu 18, which is the offspring of the most important exotic cultivars: Okinawa 100 from Japan and Nancy Hall from the United States [7]. However, the recurrent use of a few elite lines as parental stocks has decreased genetic diversity and narrowed the genetic background for sweet potato improvement [8]. Additionally, a large number of germplasm resources have been conserved, but their use is limited due to an unmanageable number of accessions. As a result, genetic diversity analyses and core germplasm development have been proposed to better manage these collections [9].

Genomic tools, such as molecular markers, can help elucidate the genetic background of the accessions, which could support breeding strategies. Random amplified polymorphic DNA (RAPD) [10,11], amplified fragment length polymorphisms (AFLPs) [12,13] and inter-simple sequence repeats (ISSRs) [14,15] are frequently used to fingerprint and characterize sweet potato varieties. However, these types of markers are far from saturated due to an insufficient number of markers. Simple sequence repeats (SSRs) are co-dominant markers that are more saturated [16]. A previous study of SSRs in 380 accessions provided a foundation upon which to study the genetic diversity of the germplasm at a fine scale [17]. SNPs are more useful than conventional markers because they are the most abundant and stable type of genetic marker in most genomes [18]. In recent years, deep sequencing technology has been rapidly developed to exploit these advantages and has enabled the high-throughput identification of SNPs [19–23], albeit with the disadvantage of becoming cost-prohibitive when the population is large. A new strategy for *de novo* SNP discovery and genotyping of large populations, referred to as specific length amplified fragment (SLAF) sequencing (SLAF-seq) [24], was recently reported. This high-resolution method has been tested on many organisms, including soybean, sesame, cucumber and the common carp (*Cyprinus carpio* L.), whose genome sequence was not reported at the time of publication [24–27].

This study presents a comprehensive view of the genome-wide variation among 197 sweet potato accessions, most of which are from China, and provides a core germplasm set representing the maximum diversity of the total collection. The described core set can be more efficiently used for breeding than the whole collection.

Materials and methods

Plant varieties and DNA extraction

A set of 197 sweet potato accessions was evaluated in the present study. These accessions were generated from different agro-climatic zones and were cultivated on the Experimental Farm of Hubei Academy of Agricultural Sciences in Jiangxia District, Wuhan, China, in 2015. The experimental site (29°18'N latitude and 113°42'E longitude with an altitude of 20–40 meters above the sea level) had a humid climate.

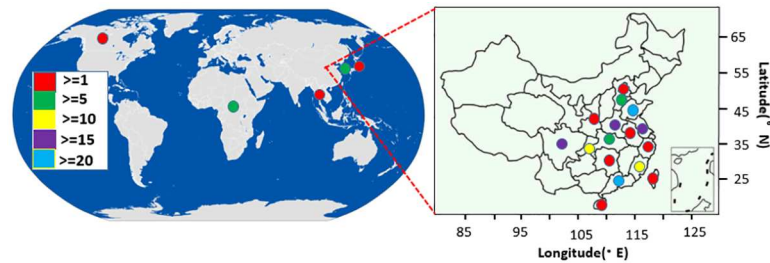


Fig 1. Location of the sweet potato accessions from around the world, highlighting China.

doi:10.1371/journal.pone.0172066.g001

The samples included 50 landraces and 147 modern cultivars. Among these accessions, 178 came from China, 6 from Africa, 2 from Japan, 8 from South Korea, 1 from Thailand, and 2 from the USA (Fig 1). Detailed information on the regional distribution of the 197 accessions is provided in S1 Table.

Bulked young healthy leaves from each accession were collected, frozen in liquid nitrogen and used for DNA extraction. DNA was isolated via the CTAB protocol [28]. The DNA concentration was quantified using a NanoDrop-2000 spectrophotometer, and DNA samples were diluted to 50 ng μL^{-1} .

High-throughput sequencing and data processing

Genomic DNA was analyzed according to the SLAF-seq method. To obtain evenly distributed SLAF tags and to avoid repetitive SLAF tags for maximum SLAF-seq efficiency, simulated restriction enzyme digestion was carried out *in silico*. The genomic DNA of the materials was digested with the *RsaI* restriction enzyme, and *Arabidopsis thaliana* DNA was used as a control to assess the normal rate of enzyme digestion [29]. The SLAF library was constructed according to procedures described by Sun with a few modifications [24]. DNA fragments of 264–314 bp were selected as SLAFs and prepared for paired-end sequencing on the Illumina High-Seq 2500 sequencing platform (Illumina, Inc.; San Diego, CA, US) at Beijing Biomarker Technologies Corporation.

The raw reads were clustered based on similarity above 90%. The SLAF tags were defined as the group with the most samples. The samples with the most tags were used as references, and GATK [30] and SAMTOOLS [31] were employed for SNP calling. Only SNPs called by both GATK and SAMTOOLS were considered to be of high quality.

Statistical analyses

The raw data were qualified [29], and these data can be further used for data mining and additional analyses. A total of 62,363 SNPs from 197 accessions were developed to calculate genetic structure and relationships (S2 Table). ADMIXTURE was employed to investigate population structure based on the maximum-likelihood method; five independent simulations were carried out for each K (number of groups) ranging from 1 to 10, and accessions were assigned to a corresponding population based on their maximum membership probabilities [32]. A phylogenetic tree based on the neighbor-joining method and a UPGMA dendrogram based on Nei's distance were constructed in MEGA 5 [33,34] using the developed SNPs. A principal component analysis (PCA) was performed with Cluster software [35,36].

Genetic diversity, polymorphic information content (PIC), and the minor allele frequency (MAF) were calculated using calculation scripts developed by Biomarker Technologies Corporation. The presence of molecular variance among groups, among accessions within groups

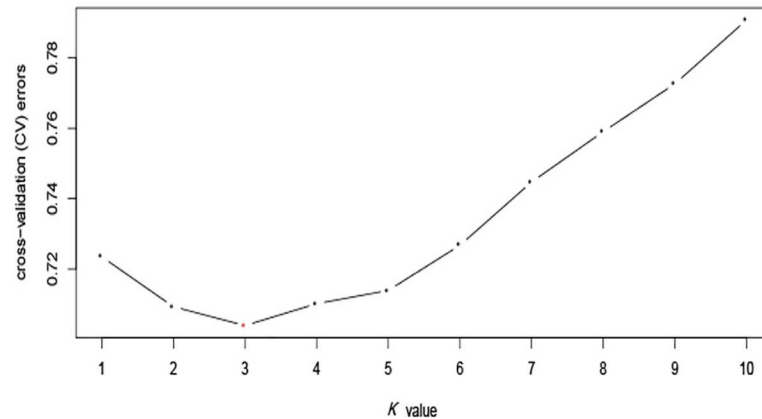


Fig 2. ADMIXTURE estimation of the number of groups for K values ranging from 1 to 10.

doi:10.1371/journal.pone.0172066.g002

and within accessions was assessed via analysis of molecular variance (AMOVA) using Arlequin [37]. Furthermore, pairwise levels of differentiation were estimated using the PopGen package in BioPerl. Finally, CoreHunter software was used to develop a core germplasm set [38,39].

Results

Population structure and phylogenetic relationships

The estimated membership fractions of the 197 accessions for different values of K ranged from 1 to 10, and the maximum likelihood revealed by the population structure showed an optimum value of 3 ($K = 3$) (Figs 2 and 3), which indicated that the entire population could be categorized into three groups: group 1, group 2 and group 3. Group 1 contained 54 accessions, one of which was from Africa, while the remainder were from different provinces in China. Group 2 contained 63 accessions from China. Group 3 contained 80 accessions, 18 of which were from Africa, Japan, South Korea, and Thailand, while the remainder were from different provinces in China (Table 1). Among the modern cultivars, 54 accessions were classified into group 1, 13 into group 2, and 80 into group 3. Among the landraces, 50 accessions were assigned to group 2 (Table 2). For a more thorough classification of the model-based structure relative to prior grouping information, the following analyses were based on the results of model-based population structure.

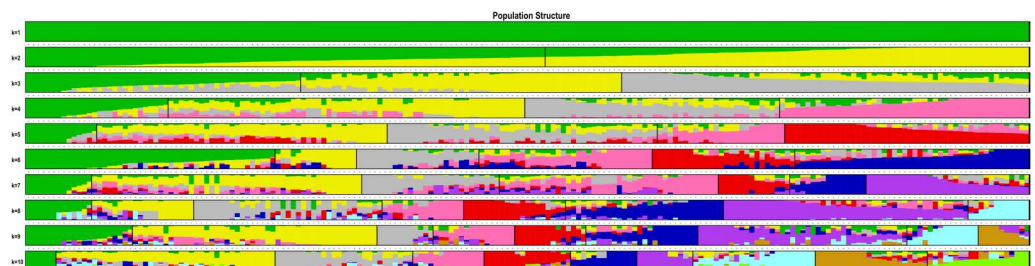


Fig 3. Pattern of variation among the 197 accessions based on 62,363 SNPs. The x-axis showed different accessions. The y-axis quantified the membership probability of accessions belonging to different groups. Colors in each row represented structural components.

doi:10.1371/journal.pone.0172066.g003

Table 1. Regional distribution of the total collection and the core set in this study.

Collection region	Group			Number of core sets
	Group 1	Group 2	Group 3	
Africa	1	0	5	1
Anhui	3	0	0	0
Beijing	1	0	2	0
Chongqing	7	0	4	2
Fujian	0	3	8	2
Guangdong	0	47	5	8
Hainan	0	3	0	2
Hebei	3	0	4	2
Henan	7	2	6	2
Hubei	1	3	4	3
Hunan	2	0	0	0
Japan	0	0	2	0
Jiangsu	11	1	4	4
South Korea	0	0	8	1
Shanxi	1	0	2	1
Shandong	9	0	11	1
Sichuan	7	2	10	8
Taiwan	0	1	0	0
Thailand	0	0	1	1
USA	0	0	2	0
Zhejiang	1	1	2	1
Total	54	63	80	39

doi:10.1371/journal.pone.0172066.t001

Neighbor-joining cluster analysis clearly divided the 197 accessions into three groups (Fig 4A); this result was consistent with the assignments made using ADMIXTURE. The UPGMA dendrogram of the 197 accessions revealed that group 1 was genetically more similar to group 3 than to group 2 (Fig 4B). Group 1 and group 3 clustered together with a genetic distance of 0.271, while group 2 stood alone, exhibiting a relatively large genetic distance from the other groups (i.e., 0.284 for group 1 and 0.288 for group 3) (Table 3). The PCA also separated the 197 accessions into three major groups (Fig 4C). Nucleotide diversity (P_i) indicated that the accessions in group 2 exhibited a higher genetic diversity than those in group 1 and group 3 (Table 2). The three groups were intermixed.

Genetic diversity

Genetic parameters, including genetic distance, PIC and MAF, were estimated separately to evaluate the genetic diversity of the three groups (Table 4; S1 Fig). The highest mean genetic

Table 2. Comparison between the model-based groups.

Group	Landraces	Modern cultivars	P_i	Number of Core sets
Group 1	0	54	0.333	9
Group 2	50	13	0.338	14
Group 3	0	80	0.319	16

P_i : Nucleotide diversity

doi:10.1371/journal.pone.0172066.t002

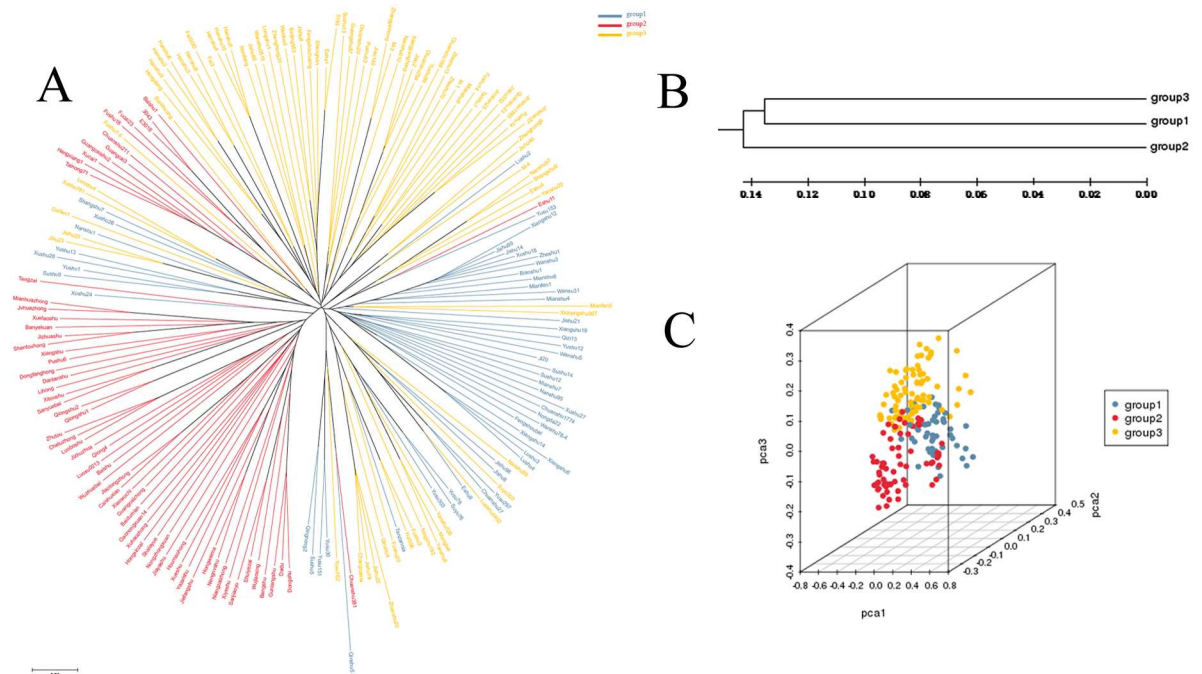


Fig 4. Characterization of the genetic structure of the 197 sweet potato accessions. (A) Phylogenetic tree of the 197 accessions based on the analysis of 62,363 SNPs. (B) UPGMA dendrogram based on Nei's distance among three inferred groups. (C) PCA plot of the 197 accessions based on the analysis of 62,363 SNPs.

doi:10.1371/journal.pone.0172066.g004

distance was present in group 3 (0.311), and the lowest was found in group 1 (0.290). Comparison of the mean PIC revealed that group 3 was highly polymorphic, whereas group 1 exhibited the lowest PIC. The mean MAF across the three groups ranged from 0.207 to 0.222.

Population differentiation

A population differentiation analysis was performed to analyze the genetic variations among and within groups, as revealed by the population structure. AMOVA revealed that the maximum diversity of 89.569% occurred within accessions, while the minimum diversity of 3.152% was attributed to genetic differentiation among groups (Table 5). The pairwise F_{st} analysis among the three inferred groups indicated that group 1 and group 2 showed the highest differentiation, with an F_{st} of 0.069; group 1 and group 3 were the most closely related, with an F_{st} of 0.045 (Table 3), this corroborated the results of UPGMA.

Development of a core set

Among the 197 sweet potato accessions studied, a core set of 39 accessions was selected using CoreHunter software (S3 Table), three of which were from Africa, South Korea and Thailand,

Table 3. Genetic distance (Downward Diagonal) and pairwise F_{st} (Upward Diagonal) among the three groups inferred through structure analysis.

	Group 1	Group 2	Group 3
Group 1		0.069	0.045
Group 2	0.284		0.046
Group 3	0.271	0.288	

doi:10.1371/journal.pone.0172066.t003

Table 4. Diversity statistics for the 197 sweet potato accessions collected in this study.

Group	Genetic distance		PIC		MAF	
	Mean	Range	Mean	Range	Mean	Range
Group 1	0.290	0~0.524	0.232	0~0.375	0.207	0~0.500
Group 2	0.307	0~0.527	0.246	0~0.375	0.219	0~0.500
Group 3	0.311	0~0.512	0.251	0~0.375	0.222	0~0.500

PIC: Polymorphic information content

MAF: Minor allele frequency

doi:10.1371/journal.pone.0172066.t004

Table 5. AMOVA of the 197 sweet potato accessions collected in this study.

Source of variation	Df	Sum of squares	Mean square	Sigma	Components of covariance (%)
Among groups	2	29,001.660	14,500.830	89.178	3.152
Among accessions within groups	194	571,656.800	2946.684	205.996	7.279
Within accessions	197	499,334.400	2534.692	2534.692	89.569
Total	393	1,099,992.860	2798.964	2829.866	100.00

doi:10.1371/journal.pone.0172066.t005

while the remainder were from different provinces of China. Among these core set accessions, nine accessions were from group 1, 14 from group 2, 16 from group 3 (Tables 1 and 2). Ten allele types were produced, of which 4 alleles were homozygous and 6 were heterozygous. Allele frequencies were calculated for the total collection and the core set, and the results showed that there were no loss of alleles in the resulting core set (Table 6; Fig 5). Genetic diversity parameters and population structure were analyzed for the core set.

Genetic diversity of the core set

The genetic diversity of the core set was estimated to determine the extent of diversity captured from the total collection. Comparisons of all genetic parameters revealed that the values for the core set were greater than those for the total collection (Table 7). For example, the mean genetic distance of the total collection was 0.303, but this value increased to 0.319 in the core set. Similarly, the mean PIC and the mean MAF of the total collection were 0.243 and 0.216, while those of the core set were 0.255 and 0.226, respectively. PCA clustering showed that the accessions of the core set were distributed into two clusters, with the landraces distanced from the modern cultivars (Fig 6).

AMOVA of the core set

AMOVA of the core set showed 3.876% of the variance among groups (Table 8); in contrast, 11.954% of the variance was present among accessions within groups, and the maximum diversity occurred within accessions. The partitioning of molecular variance was similar to that in the total collection (Table 5). A plot of the percentage allele frequency for the core set

Table 6. Comparison of percent of alleles generated in the total collection versus the core set.

	Alleles									
	AA	CC	GG	TT	GT/TG	AC/CA	AG/GA	CG/GC	AT/TA	CT/TC
Total collection (197 acc.)	19.29	21.47	21.60	19.21	1.95	1.97	5.42	1.28	2.40	5.39
Core set (39 acc.)	18.61	20.92	21.02	18.54	2.22	2.23	6.15	1.46	2.70	6.15

doi:10.1371/journal.pone.0172066.t006

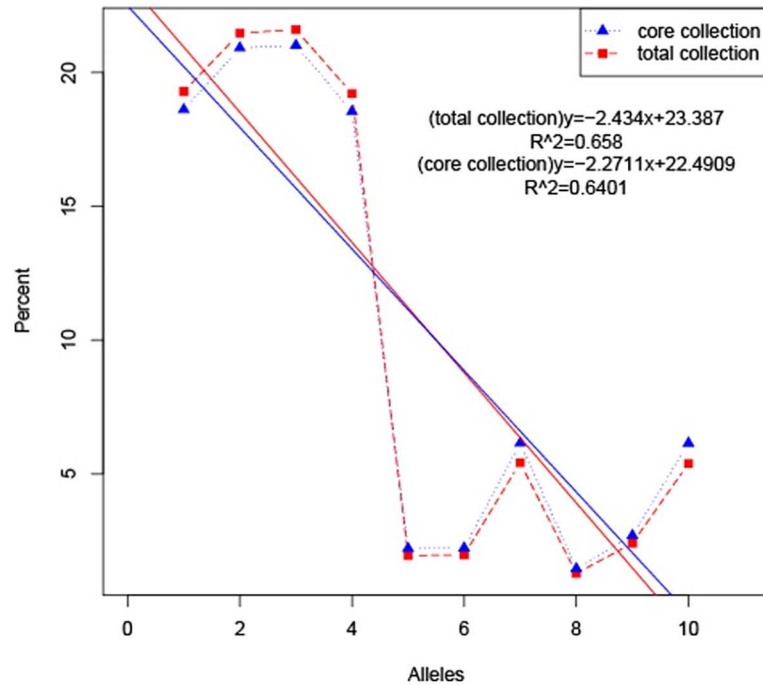


Fig 5. Comparison of the frequency of occurrence of alleles in the total collection versus the core set. The x-axis indicated ten allele types, (1) AA; (2) CC; (3) GG; (4) TT; (5) GT/TG; (6) AC/CA; (7) AG/GA; (8) CG/GC; (9) AT/TA; (10) CT/TC. The y-axis showed the allele frequencies.

doi:10.1371/journal.pone.0172066.g005

versus the total collection showed that nearly all alleles in the total collection were represented in the core set with similar frequencies (Fig 5).

Discussion

SNP-based assessment of population structure

The genetic architecture of diverse sweet potato accessions was precisely estimated using 62,363 SNPs. In this study, the concordance of the model-based structure analysis revealing three groups in the population (Fig 2) with the phylogenetic tree and PCA clustering agreed with the results of Yang et al. [17], which were generated using SSRs. Furthermore, for landraces and modern cultivars, the assignment to groups was basically in accordance with the previous study using SSRs, which also clustered landraces separately from modern cultivars (Table 2) [17]. This indicates, not surprisingly, that the landraces are genetically distant from modern cultivars. The calculation of genetic distance between groups shown in Table 3

Table 7. Comparison of the genetic diversity of the total collection versus the core set.

Genetic distance				PIC				MAF			
Total collection (197 acc.)		Core set (39 acc.)		Total collection (197 acc.)		Core set (39 acc.)		Total collection (197 acc.)		Core set (39 acc.)	
Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
0.303	0~0.527	0.319	0~0.520	0.243	0~0.375	0.255	0~0.375	0.216	0~0.500	0.226	0~0.500

PIC: Polymorphic information content

MAF: Minor allele frequency

doi:10.1371/journal.pone.0172066.t007

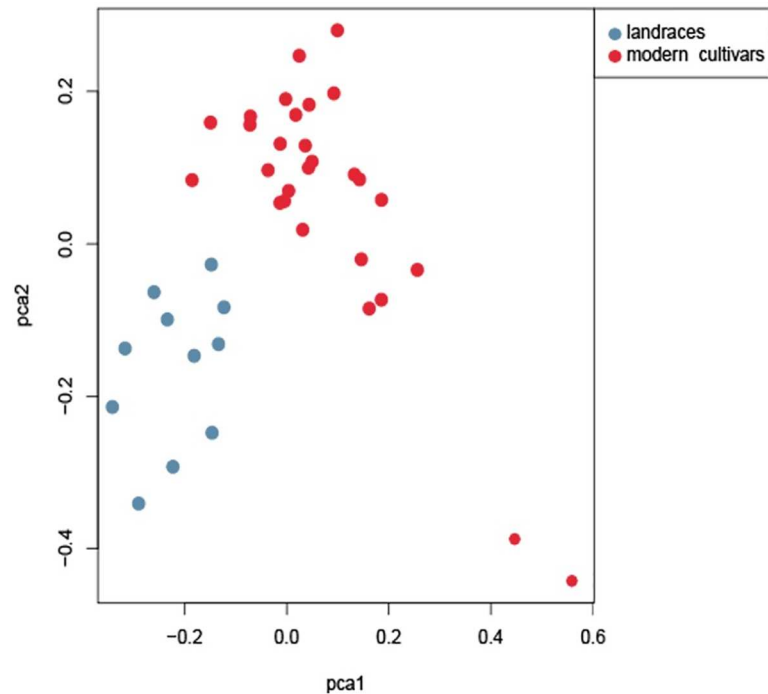


Fig 6. PCA of the sweet potato core set.

doi:10.1371/journal.pone.0172066.g006

confirms this result. Nevertheless, the clustering of the 197 accessions according to the phylogenetic tree or PCA and their assignment into groups did not agree with the information on their geographic origin. This discrepancy can be explained by the acceleration of germplasm resource exchange between regions; many previous studies also support these results [7,11,12,14].

The UPGMA dendrogram based on Nei’s distance among the three inferred groups (Fig 4B) indicated that group 2, which was mostly composed of landraces, was genetically distant from group 1 and group 3. This distance may be due to the breeding history of sweet potato; it is possible that the landraces of group 2 originated overseas and dispersed across a broad region of China, resulting in group 1 and group 3. This pattern also agrees with the results obtained by Yang et al. [17]. Thus, we inferred that the sweet potato population genetic structures estimated based on SNPs and SSRs were similar.

Model-based AMOVA showed that the maximum diversity of 89.569% occurred within accessions, while the minimum diversity of 3.152% was attributed to genetic differentiation among groups (Table 5). Yang et al. [17] performed AMOVA on 380 sweet potato accessions and revealed 16.47% variation among groups and 83.53% variation within accessions. The variation within accessions observed in our study was higher than that reported previously, possibly due to the large number of markers developed in the present study.

Table 8. AMOVA of the sweet potato core set.

Source of variation	Df	Sum of squares	Mean square	Sigma	Components of covariance (%)
Among groups	3	30,834.410	10,278.136	221.948	3.876
Among accessions within groups	35	216,587.900	6188.226	684.459	11.954
Within accessions	39	187,953.000	4819.308	4819.308	84.170
Total	77	435,375.310	5654.225	5725.715	100.00

doi:10.1371/journal.pone.0172066.t008

Genetic diversity assessed based on SNPs

Long-term selection gain requires genetic variability; thus, it is important to examine not only population structure but also genetic diversity [40]. Across the 197 sweet potato accessions examined in this study, we observed mean genetic distances of 0.290 in group 1, 0.307 in group 2, and 0.311 in group 3; mean PICs of 0.232 in group 1, 0.246 in group 2, and 0.251 in group 3; and mean MAFs of 0.207 in group 1, 0.219 in group 2, and 0.222 in group 3 (Table 4). Yang et al. [17] reported that the average genetic distances of 380 sweet potato accessions (as determined by SSRs) ranged from 0.220 to 0.254, while the average PIC ranged from 0.181 to 0.204. The higher genetic diversity observed in this study might be explained by the number of SNPs evaluated and the combination of SNP alleles at different loci.

Development of the core germplasm set

Recently, various types of molecular markers have been used to develop core germplasm sets for different types of crops. These markers include RAPD for common bean [41] and Spanish melon [42]; AFLPs for barley [43]; SSRs for rice, wheat, common bean and olive [44–47]; and SNPs for olive and *Arabidopsis* [47,48]. In this study, genomic characterization revealed high genetic diversity within the 197 sweet potato accessions; therefore, we decided to identify a core germplasm set to aid sweet potato breeders in effectively using the accessions in their crop improvement programs. This was the first time that SNPs have been used to identify major variances and select a fully representative germplasm set from a large sweet potato collection.

CoreHunter software was employed in this study to develop a core germplasm set of 39 sweet potato accessions, accounting for 19.8% of the total collection. However, the sampling percentage of the core germplasm set to fit all crops has long been under debate, possibly due to the large extent of germplasm resources and the complex types of data involved [9]. A sampling percentage of 20~30% was once suggested by Yonezawa et al. [49], and mini core sets representing approximately 1% of total collections have been used to characterize very large collections [50,51]. A perfect ratio and fixed size for all core germplasm sets do not exist because different crops and different goals require different sampling percentages [9]. In the current study, nearly all alleles were represented in the core set (Fig 5); thus, this was the perfect sampling percentage for this study.

Genetic parameters and cluster analysis were used to evaluate the efficiency of the development of the core germplasm set; these methods have been described in many other reports [9,52–55]. In the present study, the mean genetic distance, PIC, and MAF values of the core germplasm set were higher than those of the total collection (Table 7), which was expected because diversity increases after the elimination of genetically similar accessions during core germplasm set development [56]. Cluster analysis clearly separated the core set into two groups according to their types; these groups were distributed separately into a two-dimensional plot for PCA (Fig 6). AMOVA based on this model showed that the partitioning of molecular variance was similar to that for the total collection. The germplasm core set developed in this study was statistically supported by the above genetic analyses.

Conclusion

To the best of our knowledge, the SNPs reported in this study are the most saturated markers yet obtained for sweet potato. The SNP-based molecular characterization of the sweet potato collection in this study revealed large variations within accessions. The pattern of population structure and genetic diversity varied across model-based groups. Group 2, which consisted mostly of landraces, was genetically distant from the other two groups. A core germplasm set

of 39 sweet potato accessions, accounting for 19.8% of the total collection, was developed. The genome-level profiling of 197 sweet potato accessions and the development of the core set in this study will provide a foundation for genomic studies and for the identification of potential parents for sweet potato improvement.

Supporting information

S1 Fig. Boxplots of genetic diversity among the three groups.

(TIF)

S1 Table. Regional distribution of the 197 sweet potato accessions collected in this study.

(DOC)

S2 Table. List of SNPs developed in this study.

(SNPLIST)

S3 Table. List of sweet potato accessions identified for the core collection based on SNPs.

(XLS)

Acknowledgments

We are grateful to the Xuzhou Sweet Potato Research Center, Guangdong Academy of Agricultural Sciences and Shandong Academy of Agricultural Sciences for providing the plant materials.

Author Contributions

Conceptualization: WS XY CJ.

Data curation: WS YL.

Formal analysis: WS YL YY.

Investigation: WS LW JL SC YY.

Methodology: WS XY CJ.

Project administration: XY CJ.

Resources: LW JL SC.

Supervision: XY CJ.

Validation: WS XY CJ.

Visualization: WS XY CJ.

Writing – original draft: WS.

Writing – review & editing: XY CJ LW JL SC YY.

References

1. He G, Prakash CS, Jarret RL. Analysis of genetic diversity in a sweet potato (*Ipomoea batatas*) germplasm collection using DNA amplification fingerprinting. *Genome*. 1995; 38: 938–945. PMID: [8537002](#)
2. Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, et al. A sweet potato gene index established by de novo assembly of pyrosequencing and sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics*. 2010; 11: 604. doi: [10.1186/1471-2164-11-604](#) PMID: [20977749](#)

3. Department of Science, Technology and Education of Ministry of Agriculture of China, Department of Education, Science and Culture of Ministry of Finance of China. The agriculture industrial development report in China. Beijing: China Agriculture Press; 2010.
4. O'Brien PJ. The sweet potato: its origin and dispersal. *Am Anthropol.* 1972; 74: 342–365.
5. Austin DF. The taxonomy, evolution and genetic diversity of sweet potato and related wild species. In: Gregory P, editor. Report of the first sweet potato planning conference. Lima: International Potato Center; 1987. pp. 27–61.
6. Fang BP, Zhang XJ, Chen JY, An K. The history and status of sweet potato germplasm research in China. *Guangdong Agricultural Science.* 2004; 1: 1.
7. Liu DG, Zhao N, Zhai H, Yu XX, Jie Q, Wang LJ, et al. AFLP fingerprinting and genetic diversity of main sweetpotato varieties in China. *J Integr Agric.* 2012; 11: 1424–1433.
8. Hoisington D, Khairallah M, Reeves T, Ribaut JM, Skovmand B, Taba S, et al. Plant genetic resources: what can they contribute toward increased crop productivity? *Proc Natl Acad Sci.* 1999; 96: 5937–5943. PMID: [10339521](#)
9. Roy Choudhury D, Singh N, Singh AK, Kumar S, Srinivasan K, Tyagi RK, et al. Analysis of genetic diversity and population structure of rice germplasm from North-Eastern region of India and development of a core germplasm set. *PLoS ONE.* 2014; 9: e113094. doi: [10.1371/journal.pone.0113094](#) PMID: [25412256](#)
10. Gichuki ST, Berenyi M, Zhang D, Hermann M, Schmidt J, Glössl J, et al. Genetic diversity in sweetpotato (*Ipomoea batatas* (L.) Lam.) in relationship to geographic sources as assessed with RAPD markers. *Genet Resour Crop Evol.* 2003; 50: 429–437.
11. Jarret RL, Austin DF. Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. *Genet Resour Crop Evol.* 1994; 41: 165–173.
12. Zhang D, Cervantes J, Huamán Z, Carey E, Ghislain M. Assessing genetic diversity of sweetpotato (*Ipomoea batatas* (L.) Lam.) cultivars from tropical America using AFLP. *Genet Resour Crop Evol.* 2000; 47: 659–665.
13. Zhang D, Rossel G, Kriegner A, Hijmans R. AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: Its implications on the dispersal of the crop. *Genet Resour Crop Evol.* 2004; 51: 115–120.
14. Fajardo DS, La Bonte DR, Jarret RL. Identifying and selecting for genetic diversity in Papua New Guinea sweetpotato (*Ipomoea batatas* (L.) Lam.) germplasm collected as botanical seed. *Genet Resour Crop Evol.* 2002; 49: 463–470.
15. Moulin MM, Rodrigues R, Goncalves LSA, Sudré CP, Pereira MG. A comparison of RAPD and ISSR markers reveals genetic diversity among sweet potato landraces (*Ipomoea batatas* (L.) Lam.). *Acta Scientiarum, Agronomy.* 2012; 34: 139–147.
16. Elizabeth Ann V, Aline B, Mariana Silva R, Jurema RQS, Eduardo de Andrade B, Nivaldo P. Genetic diversity in Brazilian sweet potato (*Ipomoea batatas* (L.) Lam., Solanales, Convolvulaceae) landraces assessed with microsatellite markers. *Genet Mol Biol.* 2008; 31: 725–733.
17. Yang XS, Su WJ, Wang LJ, Lei J, Chai SS, Liu QC. Molecular diversity and genetic structure of 380 sweetpotato accessions as revealed by SSR markers. *J Integr Agric.* 2015; 14: 633–641.
18. Liu J, Huang S, Sun M, Liu S, Liu Y, Wang W, et al. An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods.* 2012; 8: 34. doi: [10.1186/1746-4811-8-34](#) PMID: [22920499](#)
19. Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, et al. A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir bacterial artificial chromosome contigs. *Genetics.* 2007; 176: 2637–2650. doi: [10.1534/genetics.106.067462](#) PMID: [17603124](#)
20. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, et al. Whole-Genome patterns of common DNA variation in three human populations. *Science.* 2005; 307: 1072–1079. doi: [10.1126/science.1105436](#) PMID: [15718463](#)
21. Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature.* 2007; 448: 1050–1053. doi: [10.1038/nature06067](#) PMID: [17660834](#)
22. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet.* 2007; 39: 1151–1155. doi: [10.1038/ng2115](#) PMID: [17676040](#)

23. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 2011; 12: 499–510. doi: [10.1038/nrg3012](https://doi.org/10.1038/nrg3012) PMID: [21681211](https://pubmed.ncbi.nlm.nih.gov/21681211/)
24. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, et al. SLAF-seq: an efficient method of large-scale De Novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE*. 2013; 8: e58700. doi: [10.1371/journal.pone.0058700](https://doi.org/10.1371/journal.pone.0058700) PMID: [23527008](https://pubmed.ncbi.nlm.nih.gov/23527008/)
25. Li B, Tian L, Zhang J, Huang L, Han F, Yan S, et al. Construction of a high-density genetic map based on large-scale markers developed by Specific Length Amplified Fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in glycine max. *BMC Genomics*. 2014; 15: 1086. doi: [10.1186/1471-2164-15-1086](https://doi.org/10.1186/1471-2164-15-1086) PMID: [25494922](https://pubmed.ncbi.nlm.nih.gov/25494922/)
26. Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, et al. Construction of a high-density genetic map for sesame based on large scale marker development by Specific Length Amplified Fragment sequencing (SLAF-seq). *BMC Plant Biol*. 2013; 13: 141. doi: [10.1186/1471-2229-13-141](https://doi.org/10.1186/1471-2229-13-141) PMID: [24060091](https://pubmed.ncbi.nlm.nih.gov/24060091/)
27. Xu X, Xu R, Zhu B, Yu T, Qu W, Lu L, et al. A high-density genetic map of cucumber derived from Specific Length Amplified Fragment sequencing (SLAF-seq). *Front Plant Sci*. 2015; 5: 768. doi: [10.3389/fpls.2014.00768](https://doi.org/10.3389/fpls.2014.00768) PMID: [25610449](https://pubmed.ncbi.nlm.nih.gov/25610449/)
28. Zhao N, Yu X, Jie Q, Li H, Li H, Hu J, et al. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. *Mol Breed*. 2013; 32: 807–820.
29. Su WJ, Zhao N, Lei J, Wang LJ, Chai SS, Yang XS. SNP sites developed by Specific Length Amplification Fragment sequencing (SLAF-seq) in sweetpotato. *Scientia Agricultura Sinica*. 2016; 49: 27–34.
30. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20: 1297–1303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110) PMID: [20644199](https://pubmed.ncbi.nlm.nih.gov/20644199/)
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
32. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; 19: 1655–1664. doi: [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109) PMID: [19648217](https://pubmed.ncbi.nlm.nih.gov/19648217/)
33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
34. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4: 406–425. PMID: [3447015](https://pubmed.ncbi.nlm.nih.gov/3447015/)
35. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20: 1453–1454. doi: [10.1093/bioinformatics/bth078](https://doi.org/10.1093/bioinformatics/bth078) PMID: [14871861](https://pubmed.ncbi.nlm.nih.gov/14871861/)
36. Duntelman GH. Principal components analysis. California: SAGE Publications; 1989.
37. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinformatics Online*. 2005; 1: 47–50.
38. De Beukelaer H, Smýkal P, Davenport GF, Fack V. Core Hunter II: fast core subset selection based on multiple genetic diversity measures using Mixed Replica search. *BMC Bioinformatics*. 2012; 13: 312. doi: [10.1186/1471-2105-13-312](https://doi.org/10.1186/1471-2105-13-312) PMID: [23174036](https://pubmed.ncbi.nlm.nih.gov/23174036/)
39. Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinformatics*. 2009; 10: 243. doi: [10.1186/1471-2105-10-243](https://doi.org/10.1186/1471-2105-10-243) PMID: [19660135](https://pubmed.ncbi.nlm.nih.gov/19660135/)
40. Hamblin MT, Warburton ML, Buckler ES. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS ONE*. 2007; 2: e1367. doi: [10.1371/journal.pone.0001367](https://doi.org/10.1371/journal.pone.0001367) PMID: [18159250](https://pubmed.ncbi.nlm.nih.gov/18159250/)
41. Skroch PW, Nienhuis J, Beebe S, Tohme J, Pedraza F. Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. *Crop Sci*. 1998; 38: 488–496.
42. López-Sesé AI, Staub J, Katzir N, Gómez-Guillamón ML. Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD and SSR markers to assess strategies for large collection evaluation. *Euphytica*. 2002; 127: 41–51.
43. van Treuren R, Tchoudinova I, van Soest LJM, van Hintum TJJ. Marker-assisted acquisition and core collection formation: a case study in barley using AFLPs and pedigree data. *Genet Resour Crop Evol*. 2006; 53: 43–52.
44. Yan W, Agrama H, Jia M, Fjellstrom R, McClung A. Geographic description of genetic diversity and relationships in the USDA rice world collection. *Crop Sci*. 2010; 50: 2406–2417.

45. Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet.* 2007; 114: 1265–1275. doi: [10.1007/s00122-007-0517-1](https://doi.org/10.1007/s00122-007-0517-1) PMID: [17318494](https://pubmed.ncbi.nlm.nih.gov/17318494/)
46. Blair MW, Díaz LM, Buendía HF, Duque MC. Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet.* 2009; 119: 955–972. doi: [10.1007/s00122-009-1064-8](https://doi.org/10.1007/s00122-009-1064-8) PMID: [19688198](https://pubmed.ncbi.nlm.nih.gov/19688198/)
47. Belaj A, Dominguez-García MC, Atienza SG, Martín Urdíroz N, De la Rosa R, Satovic Z, et al. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genet Genomes.* 2012; 8: 365–378.
48. McKhann HI, Camilleri C, Berard A, Bataillon T, David JL, Reboud X, et al. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 2004; 38: 193–202. doi: [10.1111/j.1365-313X.2004.02034.x](https://doi.org/10.1111/j.1365-313X.2004.02034.x) PMID: [15053772](https://pubmed.ncbi.nlm.nih.gov/15053772/)
49. Yonezawa K, Nomura T, Morishima H. Sampling strategies for use in stratified germplasm collections. In: Hodgkin T, Brown AHD, van Hintum Th JL, Morales EAV, editors. *Core collections of plant genetic resources.* Chichester, UK: John Wiley and sons; 1995. pp. 35–53.
50. Li XL, Lu YG, Li JQ, Xu HM, Shahid MQ. Strategies on sample size determination and qualitative and quantitative traits integration to construct core collection of rice (*Oryza sativa*). *Rice Sci.* 2011; 18: 46–55.
51. Wen W, Franco J, Chavez-Tovar VH, Yan J, Taba S. Genetic characterization of a core set of a tropical maize race tuxpeno for further use in maize improvement. *PLoS ONE.* 2012; 7: e32626. doi: [10.1371/journal.pone.0032626](https://doi.org/10.1371/journal.pone.0032626) PMID: [22412898](https://pubmed.ncbi.nlm.nih.gov/22412898/)
52. Kuroda Y, Tomooka N, Kaga A, Wanigadeva SMSW, Vaughan DA. Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genet Resour Crop Evol.* 2009; 56: 1045–1055.
53. Jansen J, van Hintum T. Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theor Appl Genet.* 2007; 114: 421–428. doi: [10.1007/s00122-006-0433-9](https://doi.org/10.1007/s00122-006-0433-9) PMID: [17180377](https://pubmed.ncbi.nlm.nih.gov/17180377/)
54. Upadhyaya HD, Yadav D, Dronavalli N, Gowda C, Singh S. Mini core germplasm collections for infusing genetic diversity in plant breeding programs. *Electronic J Plant Breed.* 2010; 1: 1294–1309.
55. Upadhyaya H, Reddy K, Sharma S, Varshney R, Bhattacharjee R, Singh S, et al. Pigeonpea composite collection and identification of germplasm for use in crop improvement programmes. *Plant Genet Resour.* 2011; 9: 97–108.
56. Agrama H, Yan W, Lee F, Fjellstrom R, Chen MH, Jia M, et al. Genetic assessment of a mini-core subset developed from the USDA Rice Genebank. *Crop Sci.* 2009; 49: 1336–1346.