Developing First Microsatellites and Analysing Genetic Diversity in Six Chia Cultivars

GH Yue (genhua@tll.org.sg)
Temasek Life Sciences Laboratory https://orcid.org/0000-0002-3537-2248
CC Lai
Temasek Life Sciences Laboratory
M Lee
Temasek Life Sciences Laboratory
L Wang
Temasek Life Sciences Laboratory
ZJ Song
Temasek Life Sciences Laboratory

Research Article

Keywords: Plant, protein, oil, variation, breeding, microsatellite

Posted Date: June 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-601089/v1

Abstract

Chia (*Salvia hispanica* L.), originated in central and southern Mexico and Guatemala, is an emerging industry crop due to its high content of omega-3 fatty acids and dietary fiber in its seeds. The seeds also have a high concentration of proteins and essential amino acids, and are becoming a promising source of bioactive peptides. Polymorphic DNA markers are essential tools to analyse genetic diversity and to accelerate genetic improvement. However, in Chia, polymorphic and codominant DNA markers are still lacking. In this study, fourteen polymorphic microsatellites were identified from DNA sequences and were characterized. The average allele number was 4.8 while the expected and observed heterozygosity was 0.24 and 0.34, respectively. The average PI was 0.50 while the combined PI was 9×10^{-6}. These first 14 microsatellites in Chia are useful in genetic analysis and traceability. These 14 polymorphic microsatellites were used in analysing genetic diversity and population relationships in six cultivars originating in Mexico, Australia and Bolivia. Results showed that allelic diversity and gene diversity were low and ranged from 2.79 to 3.64 and 0.27 to 0.38, respectively. The Mexico black cultivar showed the highest allelic (3.64) and gene diversity (0.38). The six cultivars were closely related with high identity ($\geq$ 0.893). Taken together, these Chia cultivars contain low genetic variation. Therefore, to initiate a breeding program for improving traits, it is essential to use seeds from multiple cultivars to enlarge genetic variation in the founder population.

1. Introduction

Chia (*Salvia hispanica* L.), which belongs to the *Lamiaceae* family, is an emerging industrial crop (Ayerza and Coates, 2011). The seeds of *S. hispanica* have been used as food since 3500 BC in Mexico (Sosa et al., 2016). This species was a cash crop in the centre of Mexico from 1500 to 900 BC. Chia was cultivated in the Central Valley of Mexico between 2000 and 2600 and was an important component of the Aztec diet (de Falco et al., 2017; Dincoglu and Yesildemir, 2019). Chia is a hardy annual plant and grows up to two meters high. This plant adapts to a wide range of soils, climates and minimal rainfall (Dincoglu and Yesildemir, 2019). It starts flowering at the age of 3 months. Flowers are blue in colour. The spikes grow to 10 cm long, set on terminal stems, and fill out to a seed head (Caruso et al., 2018). The seeds are pin-head sized, and shiny. Two months after flowering, the seeds can be harvested. Chia seeds contain soluble and insoluble fibre, high omega-3 content, proteins, minerals, vitamins, and phytochemicals, that have great potential as nutraceutical compounds, which benefit human health (Dincoglu and Yesildemir, 2019; Grancieri et al., 2019; Reyes-Caudillo et al., 2008). Chia seeds supply excellent ingredients to consider in the formulation of functional foods, such as bakery products, cereal bars (Munoz et al., 2013; Sargi et al., 2013). The United States Department of Agriculture (USDA) has promoted the plantation of Chia as an industrial crop (Valdivia-López and Tecante, 2015). In USA, Mexico, Argentina, Australia, Bolivia, India and other countries, Chia is grown as a commercial crop and its seeds are available in supermarkets and health food shops (Dincoglu and Yesildemir, 2019). The annual yield of Chia seeds is around 100–1700 kg/ha/round (Ayerza, 2016; Grimes et al., 2018; Mack et al., 2018). Chia seeds are increasingly popular and are an important ingredient among consumers and manufacturers (Iglesias-Puig

and Haros, 2013; Kuznetcova et al., 2020; Pizarro et al., 2013; Ribes et al., 2021; Zettel and Hitzmann, 2018). The production of Chia seeds has increased in recent years (Grancieri et al., 2019; Jamshidi et al., 2019). Through the centuries, humans have modified Chia, like all crops, by selective breeding. Early Mesoamerican breeders produced lines with well-developed agronomic characteristics including good, uniform seed yield and retention (Jamboonsri et al., 2012; Miranda-Ramos et al., 2020; Ullah et al., 2016). While most commercially available Chia is a mixture of different seeds, some companies offer seeds derived from a single cultivar, which they claim boosts the nutritional value and ensures consistency. However, the yield of Chia is low although conventional breeding has been conducted (Grimes et al., 2018). Further genetic improvement through molecular breeding is essential to make the production of Chia profitable and sustainable.

Polymorphic DNA markers are an essential tool in molecular breeding programs and genetic diversity studies (Nadeem et al., 2018; Tanksley, 1983). RAPD (Random Amplified Polymorphic DNA) (Welsh and McClelland, 1990) has been used in analysing genetic diversity in Chia populations. The study using RAPD found a near lack of diversity in modern commercial Chia varieties (Cahill, 2004). However, RAPD is laboratory dependent and is not highly reproducible (Wilkie et al., 1993). Thus, the RAPD results are difficult to interpret. Among all available molecular markers, microsatellites or simple sequence repeats (SSRs) have played an important role in plant breeding (Nadeem et al., 2018). Microsatellite sequences consist of tandemly repeated short motif with a length of two to six base pairs (bp) (Litt and Luty, 1989). Microsatellites are highly polymorphic, co-dominant and easy to genotype. They are randomly distributed across a genome of interest. They can be easily amplified with PCR using primers flanking the repeat DNA sequences (Guichoux et al., 2011; Yue and Xia, 2014). Genotyping of microsatellites can be conducted with automated DNA sequencers, which enables high-throughput analysis (Guichoux et al., 2011; Yue and Xia, 2014). Therefore, microsatellites are one of the most powerful DNA marker systems in analysing genetic diversity, population relationships in plant germplasm, and in genetic traceability of plant products (Kalia et al., 2011). However, there is currently no polymorphic and codominant DNA marker in Chia.

The purposes of this study were (1) to identify some microsatellites from publicly available DNA sequences in the Chia crop (*Salvia hispanica*), (2) to optimize PCR conditions of selected microsatellites, (3) to characterize 14 microsatellites using automated DNA sequencers and (4) to apply the developed microsatellites to analyse genetic diversity and population relationships in six Chia cultivars. The 14 microsatellites characterized are polymorphic and easily genotyped with PCR and an automated DNA sequencer. The novel information on genetic diversity and population relationships would facilitate the initiation of a molecular breeding program to increase the production of *Salvia hispanica*.

2. Materials And Methods

2.1. *Plant materials*

Seeds of Chia (*Salvia hispanica* L.) from six different sources were purchased from a local supermarket (NTUC, Singapore). The six sources are Mexico black chia seeds, Bolivia black Chia seeds-1, Bolivia black Chia seeds-2, Australia black Chia seeds-1, Australia black Chia seeds-2, and Australia white Chia seeds. Since black seeds from Australia and Bolivia were from four different companies, we regarded them as different cultivars. These seeds from different sources were germinated at room temperature (~ 22-24^{O}C) as described (Geneve et al., 2017). Over 100 seedlings from each source were planted in pots with soil and grown in a greenhouse at the Temasek Life Sciences Laboratory, Singapore.

2.2. Sample collection and DNA extraction

At the age of 30 days post germination, leaf samples from 48 seedlings from each source were collected and stored in a -80^{O}C freezer. DNA from each plant was extracted using the method described previously (Bai et al., 2017). DNA quality and quantity were examined using Nanodrop 2000 (Thermo Scientific, SG, Singapore) and electrophoresis on 1% agarose gels.

2.3. Identifying microsatellites and optimizing PCR for microsatellites

All DNA and cDNA sequences (Pelaez et al., 2019; Wimberley et al., 2020) deposited in Genbank were downloaded. Microsatellite sequences with trinucleotide repeats were extracted using SciRoKo software (Kofler et al., 2007). Primers for 20 selected microsatellites (Supplementary Table S1) were designed using PrimerSelect from the DNAstar software package (DNAstar, CA, USA). Both normal primers and labelled primers (Table 1) were ordered from IDT (IDT, SG, Singapore). For optimizing PCR for each microsatellite, unlabelled primers were used. Each 25 µl PCR reaction contained 40 ng of genomic DNA, 0.5 units of Taq polymerase (Roche, SG, Singapore), 1x Roche Taq PCR buffer containing 1.5 mM $MgCl_2$, 0.2 µM dNTPs, and 200 nM of each primer. PCR was conducted under the following conditions: 2 min denaturation at 94^{O}C; followed by 38 cycles of 30 s at 94^{O}C, 30 s at annealing temperature and 30 s at 72^{O}C, with a final extension at 72^{O}C for 5 min. For each microsatellite, three annealing temperatures (52, 55 and 60^{O}C) were tested. After PCR with the optimal annealing temperature for each microsatellite, the PCR products were checked using electrophoreses on 2–3% agarose gels.

Table 1
Fourteen newly identified and characterized microsatellites in Chia (*Salvia hispanica* L.)

Name	Repeat	Primer sequence (5' to 3')	Size range (bp)	Ta (°C)	Accession no.
Chia01	(ACA)11	CACATAAATATAGGAGAGGGTTG	276–286	55	MZ285863
		AAGGCCAGAAGGGAATG			
Chia03	(TCC)10	TCGTCCTTGTGCAAAACAGTTCAA	216–218	55	MZ285864
		AACAGCCCCAGGAAGTGAGAGC			
Chia04*	(CTT)13	CAAATAGCCAATATTTTACGC	130–157	55	MZ285865
		GTCAGGCAGAGTAGAGTGTGT			
Chia05	(TAT)8	TCCTTATTTTGGACAAGTGG	311–377	55	MZ285866
		TTCATAAGGCATCAAGGAATA			
Chia06	(GTG)10	TATTTTGGACAAGTGGCAGCATA	366–383	55	MZ285867
		TTGCGATTCATAAGGCATCAAG			
Chia08*	(GTG)8	CTGTTTATGGACACAGAGGAG	145–165	55	MZ285868
		CCAAAGAGCATACAAACATTC			
Chia09	(GTG)8	GGTATGGCGGTGGCTATGA	215–222	55	MZ285869
		ACATTCCCATTCCACCAAAGA			
Chia10*	(GAG)8	ACCTAGAATCGGATCTTAACG	150–175	55	MZ285870
		TGAACTTCCAAATACTCCTCA			
Chia12*	(CAG)8	ACTACTGGGATGAAGACGAAT	165–178	52	MZ285871
		ACGGTAAACGAGGTAATCTTC			
Chia13	(ATG)10	ATGGGCTCCACTTCTTCTTCTC	226–251	55	MZ285872
		CACCTCATTTCCATCTTCATCTAA			
Chia15	(GCG)8	GTTGTAATAGTTGGAGAAGAAGGA	327–379	52	MZ285873
		AAAAGCTTTTGTTGAGCGATG			
Chia18	(ACT)8	AAAACCGTCCGACTAAAACCTAAT	197–208	55	MZ285874
		TATTTGGGCATTCAAGTGGTAGTT			

*. Primer sequences were adapted from the paper of Pelaez et al. (Pelaez et al. 2019), however, these markers have not been characterized by Pelaez et al.

Name	Repeat	Primer sequence (5’ to 3’)	Size range (bp)	Ta (°C)	Accession no.
Chia19*	(ATC)8	TTTCTGTCTTGGTCTTGGTTA CTGAGCCAACTCTCTGTGTAA	126–134	55	MZ285875
Chia20*	(CAC)8	CAGCAGCTTTGTCTGACTC CTTAACCCCTTGCAGAATAAT	246–256	55	MZ285876

*. Primer sequences were adapted from the paper of Pelaez et al. (Pelaez et al. 2019), however, these markers have not been characterized by Pelaez et al.

2.4. Genotyping microsatellites in six cultured populations of Chia

To genotype the 14 selected microsatellites in 288 plants from six different sources, labelled primers (Table 1) were used. The PCR reactions and PCR programs are the same as described above. The optimized annealing temperatures of PCR for each microsatellite are shown in Table 1. After PCR, 0.6-2 µl PCR products were genotyped using an automated DNA sequencer ABI3730xl (Applied Biosystems, CA, USA) as described previously (Bai et al., 2017). After the fragment analysis on the ABI3730xl sequencer, genotypes were then analysed with the Rox-500 size standard (Applied Biosystems, CA, USA) for each marker in each sample using software GeneMapper (Applied Biosystems, CA, USA). Genotypes were exported in Excel files for data analysis.

2.5. Data analysis of genetic diversity, potential bottleneck and population relationships

The software GDA (Lewis and Zaykin, 2000) was used to calculate allele number (A), observed heterozygosity (Ho), expected (He) heterozygosity, inbreeding index (F), pairwise linkage and Hardy-Weinberg equilibrium (HWE). The probability of identity (PI) was calculated using GenAlEx 6.5 (Peakall and Smouse, 2006). Polymorphic Information Content (PIC) was calculated with the program Cervus (Kalinowski et al., 2007).

The calculation of all genetic distances (i.e. Nei's standard genetic distance D, 1972 and 1978, as well as co-ancestry identity) were carried out using the software GDA (Lewis and Zaykin, 2000). Phylogenetic trees were constructed using matrices of these genetic distances based on un-weighted pair group methods with arithmetic (UPGMA) averages and neighbour-joining (NJ) methods using software MEGAX (Kumar et al., 2018).

Population structure was analysed using the software Structure 2.3.4 (Pritchard et al., 2010) based on the 14 polymorphic microsatellites. To identify the number of populations (K) making up the structure of the data, the burn-in period was set at 5000 with the Markov Chain Monte Carlo iterations and the run length set at 5000 in an admixture model. Ten runs were performed for each simulated value of *K*, which ranged

from 1 to 6. Each individual was then assigned to a cluster (Q) based on the probability determined by the software that the genotype belonged in the cluster. Based on $K = 6$, a Bar plot with “Sort by Q” was obtained to show the visual of the population structure among the 288 individuals.

Since cultured populations are often subjected to founder effects and bottleneck, which leads to lower genetic diversity (Piry et al., 1999), the bottleneck hypothesis was tested under a two-phased model of mutation (TPM) using the program Bottleneck (Cornuet and Luikart, 1996) because most microsatellites mutated in a two-phased model (Schlötterer, 2000). Wilcoxon’s signed rank and standard difference tests were used to examine the statistical significance of the potential bottleneck in each cultivar.

3. Results

3.1. Microsatellite sequences and polymorphisms in Chia

Twenty DNA sequences (see supplementary Table S1) from Chia containing trinucleotide microsatellites were retrieved from DNA sequences in NCBI Genbank. Twenty pairs of primers flanking the repeat sequences were designed (see supplementary Table S1). After PCR optimization, 14 primer pairs (Table 1) amplified clear PCR products. These 14 microsatellites were genotyped using 288 individuals with the automated ABI3730xl DNA sequencer. They were all polymorphic (Table 2). The allele numbers averaged at 4.8, and ranged from two for Chia12 to six for Chia03 Chia04, Chai13 and Chia15. The average expected heterozygosity was 0.34 while the observed heterozygosity was 0.26. PIC (Polymorphic information content) was averaged at 0.30 with a range from 0.08 to 0.59. Nine of the 14 markers showed deviation from HWE. Possibility of identity (PI) ranged from 0.19 for Chia05 to 0.84 for Chia12. The combined PI of all the 14 markers was 9×10^{-6}. Pairwise analysis of linkage showed that these markers were unlikely to be linked. Detailed data of the characteristics of the 14 microsatellites are presented in Table 2.

Table 2
Characterization of 14 polymorphic microsatellites in 288 Chia (*Salvia hispanica* L.) individuals

Marker	N	Ne	Ho	He	PIC	PI	*P*	F(null)
Chia01	5	1.33	0.15	0.25	0.22	0.59	***	0.25
Chia03	6	1.25	0.05	0.20	0.19	0.66	ND	0.61
Chia04	6	1.71	0.22	0.42	0.38	0.38	***	0.30
Chia05	4	2.97	0.83	0.66	0.59	0.19	***	-0.14
Chia06	5	1.24	0.12	0.19	0.19	0.66	ND	0.25
Chia08	4	1.58	0.23	0.37	0.33	0.44	***	0.24
Chia09	3	1.12	0.08	0.11	0.10	0.80	ND	0.12
Chia10	5	1.93	0.38	0.48	0.37	0.38	**	0.12
Chia12	2	1.09	0.02	0.08	0.08	0.84	ND	0.45
Chia13	6	1.87	0.23	0.47	0.38	0.37	***	0.34
Chia15	6	2.11	0.70	0.53	0.47	0.28	***	-0.17
Chia18	5	1.29	0.12	0.22	0.22	0.61	ND	0.36
Chia19	5	1.55	0.27	0.36	0.31	0.46	**	0.13
Chia20	5	1.72	0.20	0.42	0.36	0.39	***	0.36
Average	4.8	1.62	0.26	0.34	0.30	0.50	-	-

N: number of alleles, Ne: effective number of alleles, Ho: observed heterozygosity, He: expected heterozygosity, PIC: polymorphic information content, PI: probability of identity, *P*: probability for deviation from Hardy-Weinberg equilibrium (HWE), and F(null): estimated frequency of null alleles, ND: No deviation from HWE, **: $P < 0.01$, ***: $P < 0.001$

3.2. Genetic diversity, inbreeding and bottleneck in the six cultivars

Fourteen microsatellites were used in analysing genetic diversity and population relationships in six cultivars that originated from Mexico, Bolivia, and Australia (Table 3). The Mexico black (MB01) showed the highest allelic (A = 3.64) and gene diversity (He = 0.38). The allelic diversity was highest in Mexico black (MB), followed by AUW, AUB02, BB02, AUB01 and BB01. The gene diversity (He) was highest in MB, followed by BB02, AUB02, BB01, AUB01 and AUW. The observed heterozygosity ranged from 0.27 for BB02 and AUB02 to 0.24 for AUB01. All populations showed a high degree of inbreeding with an inbreeding index (F) ranging from 0.11 for BB01 to 0.31 for MB.

Table 3
Genetic diversity and possible bottleneck in six Chia (*Salvia hispanica* L.) cultivars

Cultivar	N	NA	He	Ho	F	*P* (sign-TPM)	*p* (stdv_TPM)	*p* (W_2t_TPM)
BB01	48	2.79	0.28	0.25	0.11	0.08	0.09	0.19
AUW	48	3.21	0.27	0.26	0.03	0.04	0.00	0.02
MB	48	3.64	0.38	0.26	0.31	0.10	0.02	0.19
AUB01	48	2.86	0.27	0.24	0.12	0.07	0.04	0.17
AUB02	48	3.07	0.32	0.27	0.17	0.45	0.08	0.42
BB02	48	2.93	0.36	0.27	0.26	0.37	0.47	0.89

N: Number of individuals, NA: Average number of alleles at 14 microsatellites; He: expected heterozygosity, Ho: observed heterozygosity, F: Inbreeding index; *P*: possibility of bottleneck, sign-TPM: Sign test under two-phased model of mutation, stdv_TPM: Standardized differences test under two-phased model of mutation, W_2t_TPM: Wilcoxon test under two-phased model of mutation, BB01: Bolivia black Chia population-1, AUW: Australia white, MB: Mexico black, AUB01: Australia black population-1, AUB02: Australia black population-2 and BB02: Bolivia black Chia population-2.

Bottleneck analysis showed that some populations might have experienced a bottleneck (Table 3). Based on the three different statistics, it was highly possible that the three cultivars (AUW, MB, AUB01) might have experienced a bottleneck while in other three cultivars, including BB01, AUB02 and BB02, the possibility of bottleneck was low although they showed low genetic diversity.

3.3. Genetic relationships between the six cultivars

Based on genetic distances estimated with the genotypes of the 14 microsatellites, the genetic relationships between the six cultivars were analysed using several methods. The obtained phylogenetic trees using different methods were not uniform. The most common tree is shown in Fig. 1. The two Bolivia black populations (BB01 and BB02) were clustered into a group, while the Mexico black (MB) was grouped with Australian populations (AUB01, AUSB02, AUW) into a big cluster. The genetic distances between pairwise populations were rather small (Supplementary Table S2).

Analysis of population structure showed that with $K = 6$, the individuals from different cultivars were scattered in different positions in the plot of estimates of Q (Fig. 2), suggesting low differentiation of the six cultivars.

4. Discussion

4.1. Chia as an industrial crop

Although Chia is a promising industrial crop due to the many interesting characteristics of its seeds (e.g. high omega-3 contents, high fibers, and high proteins) (Ayerza and Coates, 2011), its seed yield in

commercial production is still low (ranging from 100 to 1700 kg/ha/round) (Ayerza, 2016; Grimes et al., 2018; Mack et al., 2018). Therefore, to make Chia a profitable and sustainable industrial crop, it is essential to increase its yield and capability to grow under different culture conditions and climate conditions through breeding. To initiate a breeding program, a founder population with high genetic variation is essential (Cahill, 2004; Zhao et al., 2009) while polymorphic DNA markers is an essential tool to access genetic diversity in wild and cultured populations (Nadeem et al., 2018). To the best of our knowledge, although some DNA sequences containing microsatellites have been reported (Pelaez et al., 2019; Wimberley et al., 2020), no microsatellite has been characterized and applied in analysing genetic variations in Chia.

4.1. DNA markers in Chia

Microsatellites are one of the most preferred DNA markers in genetic studies in plants (Nadeem et al., 2018). Taking advantage of DNA sequences, including cDNA sequences, available in public databases (e.g. Genbank), we have derived 20 sequences containing trinucleotide microsatellites. After PCR optimization, 14 microsatellites could be easily amplified, indicating that publicly available sequences are a good source for finding DNA markers, especial for the emerging industrial crop species. In fact, this strategy has been used for identifying DNA markers in other agronomic species, livestock and aquaculture species (Arbeiter et al., 2021; Schlötterer, 2000; Zane et al., 2002). Therefore, for novel species, where researchers do not have much initial funding for identifying and characterizing DNA markers, identifying DNA markers in public databases for characterization is a cost-effective and rapid way to start a study on genetic diversity to initiate a breeding program.

Fourteen microsatellites were characterized in 288 individuals from six different sources in Mexico, Australia and Bolivia. All the 14 microsatellites were polymorphic. The allele number is relatively lower in comparison to that in rice (Zhao et al., 2009). Nine of the 14 microsatellites deviated from HWE. The reason for the lower allele number and deviation from HWE may be because the samples used in characterizing the microsatellites are from cultured populations and Chia is a selfing species (Grimes et al., 2018). Using samples from wild populations, more alleles may be found. All these parameters of the 14 newly identified microsatellites suggest that these microsatellites could be used in initial studies on genetic diversity and population structure to initiate a breeding program. Since the combined PI of the 14 microsatellites was 9×10^{-6}, these microsatellites could also be used for genetic traceability of Chia seeds. It is of note that only 14 microsatellites were characterized in this study. These microsatellites are not enough for linkage mapping, QTL (quantitative trait locus) mapping and GWAS (genome wide association studies), which are essential for molecular breeding to accelerate genetic improvement. Therefore, it is essential to identify more DNA markers for future molecular breeding. GBS (genotyping by sequencing) using next-generation sequencing is a cost-effective and rapid way to identify a large number of polymorphic DNA markers (Peterson et al., 2014; Wang et al., 2017), including SNPs and microsatellites.

4.3. Genetic diversity and population structure in Chia

Genetic diversity is the foundation for any breeding program for genetic improvement (Wilkie et al., 1993). Although conventional breeding has been conducted in the Chia crop (Pelaez et al., 2019), not much is known about the genetic variation in Chia. We analysed the genetic variations in six cultivars from Mexico, Bolivia and Australia. In general, the allele number and gene diversity are lower in comparison to most of cultured agronomic plant species, including rice (Zhao et al., 2009), but quite similar to corn (Ferreira et al., 2018). All these data indicate that the current cultivars analysed in this study contain very low genetic variation. Therefore, to further improve the yield, quality and resilience traits of Chia, it is essential to enlarge genetic variations of cultured populations by introducing new breeding materials from wild populations

The analysis of the genetic distances showed that there was genetic differentiation among the six cultured cultivars although the genetic distance was quite small. This low genetic differentiation between cultivars is in agreement with a previous analysis of genetic diversity in Chia with RAPD analysis (Wilkie et al., 1993). The close relationship between the two cultivars (i.e. BB01 and BB02) in Bolivia is logical as they are from the same country. The close relationship between the Mexico population with the cultivars in Australia might be because Australia imported seeds from Mexico for commercial plantation recently (McAloon, 2016). Taken together, our data indicates that the genetic variation and population differentiation in the six Chia cultivars are very small. Certainly, as an initial study on genetic diversity and population structure in Chia, our study only covers a few cultivars and may not reflect the whole picture of genetic diversity of Chia. In future studies, it is essential to examine more cultured and wild populations with more DNA markers, including microsatellites and SNPs (Single nucleotide polymorphisms).

5. Conclusion

Chia is an emerging industrial crop. In this study, we identified 14 polymorphic trinucleotide microsatellites in Chia. This first set of polymorphic microsatellites enables us to start genetic diversity and genetic traceability in Chia. Analysing genetic diversity and population structure in six Chia cultivars revealed that the genetic variation and genetic differentiation were very low. Therefore, to initiate breeding programs to accelerate genetic improvements of economic traits, it is essential to develop more DNA makers to examine genetic variations in cultured and wild population globally.

Declarations

CRediT authorship contribution statement

Conceptualization: GHY; Data curation: GHY, CCL, ML, LW, ZJS; Formal analysis: GHY; Funding acquisition: GHY; Investigation: GHY, CCL, ML, LW, ZJS; Methodology: GHY, CCL, ML, LW, ZJS; Project administration: GHY; Resources: GHY; Software: GHY; Supervision: GHY; Validation: GHY, CCL, ML, LW, ZJS; Visualization: GHY; Roles/Writing - original draft: GHY; Writing - review & editing: GHY, CCL, ML, LW, ZJS.

Declaration of competing interest

The authors declare that they have no conflict of interest

Acknowledgements

This research was financially supported by the Internal Funding of the Temasek Life Sciences Lab, Singapore

References

1. Arbeiter AB,Hladnik M,Jakše J,Bandelj D 2021. First set of microsatellite markers for immortelle (*Helichrysum italicum* (Roth) G. Don): A step towards the selection of the most promising genotypes for cultivation. Indus Crop Prod, 162, 113298.
2. Ayerza R 2016. Crop year effects on seed yields, growing cycle length, and chemical composition of chia (*Salvia hispanica* L) growing in Ecuador and Bolivia. Emirates J Food Agricul, 28, 196-200.
3. Ayerza R,Coates W 2011. Protein content, oil content and fatty acid profiles as potential criteria to determine the origin of commercially grown chia (*Salvia hispanica* L.). Indus Crop Prod, 34, 1366-1371.
4. Bai B,Wang L,Lee M,Zhang Y,Alfiko Y,Ye BQ et al. 2017. Genome-wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol, 17, 93.
5. Cahill JP 2004. Genetic diversity among varieties of Chia *(Salvia hispanica* L.). Genet Resour Crop Evol, 51, 773-781.
6. Caruso MC,Favati F,Di Cairano M,Galgano F,Labella R,Scarpa T et al. 2018. Shelf-life evaluation and nutraceutical properties of chia seeds from a recent long-day flowering genotype cultivated in Mediterranean area. Lwt-Food Sci Technol, 87, 400-405.
7. Cornuet JM,Luikart G 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics, 144, 2001-2014.
8. de Falco B,Amato M,Lanzotti V 2017. Chia seeds products: an overview. Phytochem Rev, 16, 745-760.
9. Dincoglu AH,Yesildemir O 2019. A renewable source as a functional food: Chia seed. Cur Nutr Food Sci, 15, 327-337.
10. Ferreira F,Scapim CA,Maldonado C,Mora F 2018. SSR-based genetic analysis of sweet corn inbred lines using artificial neural networks. Crop Breed Appl Biotechnol, 18, 309-313.
11. Geneve RL,Hildebrand DF,Phillips TD,Al-Amery M,Kester ST 2017. Stress influences seed germination in mucilage-producing Chia. Crop Sci, 57, 2160-2169.
12. Grancieri M,Martino HSD,de Mejia EG 2019. Chia seed *(Salvia hispanica* L.) as a source of proteins and bioactive peptides with health benefits: A review. Comprehen Rev Food Sci Food Safety, 18, 480-499.
13. Grimes SJ,Phillips TD,Hahn V,Capezzone F,Graeff-Hoenninger S 2018. Growth, Yield performance and quality parameters of three early flowering chia (*Salvia hispanica* L.) genotypes cultivated in southwestern Germany. Agri-Basel, 8, 154.

14. Guichoux E,Lagache L,Wagner S,Chaumeil P,Léger P,Lepais O et al. 2011. Current trends in microsatellite genotyping. Mol Ecol Res, 11, 591-611.
15. Iglesias-Puig E,Haros M 2013. Evaluation of performance of dough and bread incorporating chia (*Salvia hispanica* L.). Eur Food Res Technol, 237, 865-874.
16. Jamboonsri W,Phillips TD,Geneve RL,Cahill JP,Hildebrand DF 2012. Extending the range of an ancient crop, *Salvia hispanica* L.-a new omega 3 source. Genet Resour Crop Evol, 59, 171-178.
17. Jamshidi AM,Amato M,Ahmadi A,Bochicchio R,Rossi R 2019. Chia (*Salvia hispanica* L.) as a novel forage and feed source: A review. Ital J Agron, 14, 1-18.
18. Kalia RK,Rai MK,Kalia S,Singh R,Dhawan A 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica, 177, 309-334.
19. Kalinowski ST,Taper ML,Marshall TC 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol, 16, 1099-1106.
20. Kofler R,Schlötterer C,Lelley T 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics, 23, 1683-1685.
21. Kumar S,Stecher G,Li M,Knyaz C,Tamura K 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol, 35, 1547-1549.
22. Kuznetcova DV,Linder M,Jeandel C,Paris C,Desor F,Baranenko DA et al. 2020. Nanoliposomes and nanoemulsions based on chia seed lipids: Preparation and characterization. Intern J Mol Sci, 21, 9079.
23. Lewis P,Zaykin D 2000. Free program distributed by the authors over the internet from the GDA Home Page.
24. Litt M,Luty JA 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet, 44, 397.
25. Mack L,Munz S,Capezzone F,Hofmann A,Piepho HP,Claupein W et al. 2018. Sowing date in Egypt affects chia seed yield and quality. Agron J, 110, 2310-2321.
26. McAloon K 2016. What you need to know about Australian grown Chia Seeds (https://www.kissreadylips.com/blogs/news/what-you-need-to-know-about-australian-grown-chia-seeds).
27. Miranda-Ramos K,Millan-Linares MC,Haros CM 2020. Effect of Chia as breadmaking ingredient on nutritional quality, mineral availability, and glycemic index of bread. Foods, 9, 663.
28. Munoz LA,Cobos A,Diaz O,Aguilera JM 2013. Chia seed (*Salvia hispanica*): An ancient grain and a new functional food. Food Rev Int, 29, 394-408.
29. Nadeem MA,Nawaz MA,Shahid MQ,Doğan Y,Comertpay G,Yıldız M et al. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnolog Equip, 32, 261-285.

30. Peakall R,Smouse PE 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes, 6, 288-295.
31. Pelaez P,Orona-Tamayo D,Montes-Hernandez S,Valverde ME,Paredes-Lopez O,Cibrian-Jaramillo A 2019. Comparative transcriptome analysis of cultivated and wild seeds of *Salvia hispanica* (Chia). Sci Rep, 9, 9761.
32. Peterson GW,Dong Y,Horbach C,Fu Y-B 2014. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity, 6, 665-680.
33. Piry S,Luikart G,Cornuet J-M 1999. BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. J Hered, 90, 502-503.
34. Pizarro PL,Almeida EL,Samman NC,Chang YK 2013. Evaluation of whole chia (*Salvia hispanica* L.) flour and hydrogenated vegetable fat in pound cake. Lwt-Food Sci Technol, 54, 73-79.
35. Pritchard JK,Wen W,Falush D 2010. Documentation for STRUCTURE software: Version 2. University of Chicago, Chicago, IL.
36. Reyes-Caudillo E,Tecante A,Valdivia-Lopez MA 2008. Dietary fibre content and antioxidant activity of phenolic compounds present in Mexican chia *(Salvia hispanica* L.) seeds. Food Chem, 107, 656-663.
37. Ribes S,Pena N,Fuentes A,Talens P,Barat JM 2021. Chia (*Salvia hispanica* L.) seed mucilage as a fat replacer in yogurts: Effect on their nutritional, technological, and sensory properties. J Dairy Sci, 104, 2822-2833.
38. Sargi SC,Silva BC,Santos HMC,Montanher PF,Boeing JS,Santos OO et al. 2013. Antioxidant capacity and chemical composition in seeds rich in omega-3: chia, flax, and perilla. Food Sci Technol, 33, 541-548.
39. Schlötterer C 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma, 109, 365-371.
40. Sosa A,Ruiz G,Rana J,Gordillo G,West H,Sharma M et al. 2016. Chia crop (*Salvia hispanica* L.): its history and importance as a source of polyunsaturated fatty acids omega-3 around the world: a review. J. Crop Res. Fert., 1, 1-9.
41. Tanksley SD 1983. Molecular markers in plant breeding. Plant Mol. Biol. Rep., 1, 3-8.
42. Ullah R,Nadeem M,Khalique A,Imran M,Mehmood S,Javid A et al. 2016. Nutritional and therapeutic perspectives of Chia (*Salvia hispanica* L.): a review. J Food Sci Technol, 53, 1750-1758.
43. Valdivia-López MÁ,Tecante A 2015. Chia (*Salvia hispanica*): a review of native Mexican seed and its nutritional and functional properties. Adv Food Nutri Res, 75, 53-75.
44. Wang L,Bai B,Huang S,Liu P,Wan ZY,Ye B et al. 2017. QTL mapping for resistance to iridovirus in Asian seabass using genotyping-by-sequencing. Mar Biotechnol, 19, 517-527.
45. Welsh J,McClelland M 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res., 18, 7213-7218.
46. Wilkie SE,Isaac PG,Slater RJ 1993. Random amplified polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. Theor Appl Genet, 86, 497-504.

47. Wimberley J,Cahill J,Atamian HS 2020. De novo sequencing and analysis of *Salvia hispanica* tissue-specific transcriptome and identification of genes involved in terpenoid biosynthesis. Plants, 9, 405.
48. Yue GH,Xia JH 2014. Practical considerations of molecular parentage analysis in fish. J World Aqua Soc, 45, 89-103.
49. Zane L,Bargelloni L,Patarnello T 2002. Strategies for microsatellite isolation: a review. Mol Ecol, 11, 1-16.
50. Zettel V,Hitzmann B 2018. Applications of chia (*Salvia hispanica* L.) in food products. Trends Food Sci Technol, 80, 43-50.
51. Zhao W,Chung J-W,Ma K-H,Kim T-S,Kim S-M,Shin D-I et al. 2009. Analysis of genetic diversity and population structure of rice cultivars from Korea, China and Japan using SSR markers. Genes & Genomics, 31, 283-292.

Figures

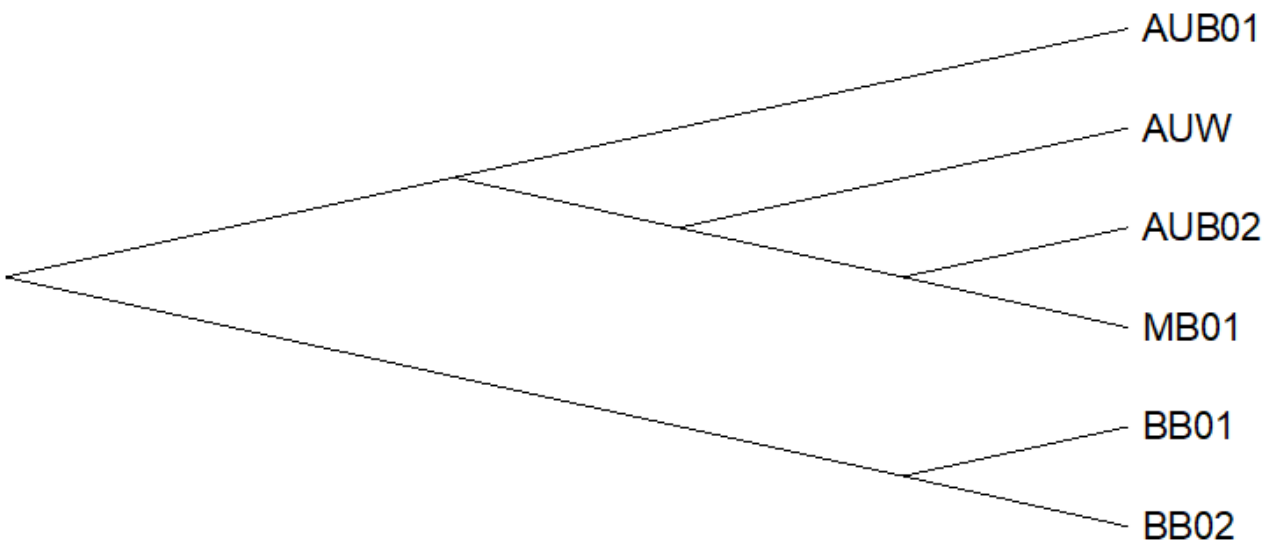

Figure 1

A Neighbor-Joining phylogenetic tree showing relationships between six cultured populations based on the genotypes at 14 microsatellites.

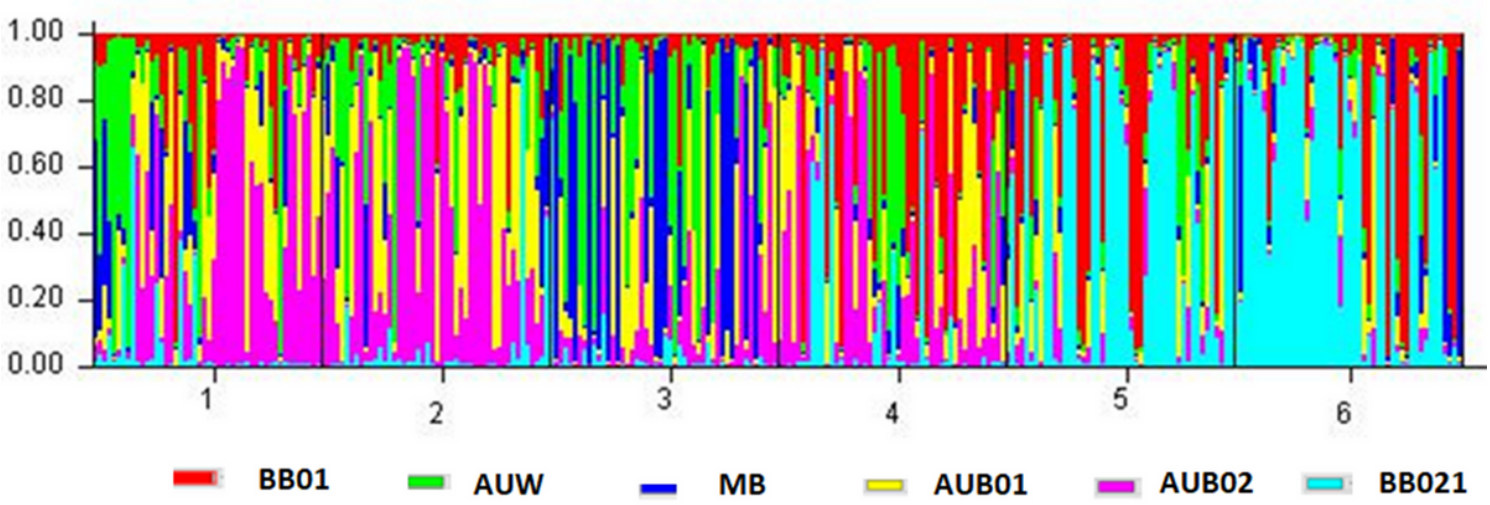

Figure 2

Summary plot of estimates of Q from the structure analysis of six chia populations originated in Mexico, Bolivia and Australia. Each individual is represented by a single vertical line broken into K colored segments, with lengths proportional to each of the K inferred clusters. The numbers (1, 2, 3, 4, 5 and 6) correspond to the predefined populations. Red colour represents

Supplementary Files

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

- SupplementaryTableS1.xlsx
- SupplementaryTableS2.docx