Genome-wide association mapping for agronomic and quality traits in foxtail millet

Yanhong Lou

Shandong agricultural University

Yun Chen

Hubei Normal University

Zhihao Liu

Hubei Normal University

Mingjie Sun

Shandong Agricultural University

Fei Han

Shandong Agricultural University

Tingting Wang

Shandong Agricultural University

Hui Wang

Shandong Agricultural University

Yuping Zhuge (✉ zhugeyp@sdau.edu.cn)

Shandong Agricultural University <https://orcid.org/0000-0002-8298-196X>

Liang Chen

Wuhan Botanical Garden, CAS

Research article

Keywords: Gene pyramiding, Marker-assisted breeding, Phenotypic value, Poaceae, Quantitative trait loci, *Setaria italica*

Posted Date: June 6th, 2019

DOI: <https://doi.org/10.21203/rs.2.10042/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Foxtail millet [*Setaria italica* (L.) P. Beauv.] is a particularly important cereal and fodder crop in arid and semi-arid regions. The genomic variation and alleles underpinning agronomic and quality traits are important for foxtail millet improvement. To better understand the diversity of foxtail millet and facilitate the genetic dissection of its agronomic and quality traits, we used high-quality single nucleotide polymorphisms (SNPs) to perform a genome-wide association study (GWAS).

Results: Using genotyping-by-sequencing, 107 foxtail millet accessions were sequenced, and further analysis revealed 72,181 high-quality SNPs, of which 53 were significantly associated with 15 agronomic and quality traits. These SNPs were distributed across the nine chromosomes of foxtail millet; 44 were located in intergenic regions, whereas one and eight SNPs were located in exon and intron regions, respectively. The GWAS revealed that 28 SNPs were associated with a single trait.

Conclusions: For some of the significant SNPs, favourable genotypes showed pyramiding effects for several traits. The 53 loci identified in this study will therefore be useful for breeding programs aimed at foxtail millet improvement.

Background

Foxtail millet [*Setaria italica* (L.) P. Beauv.] is a perennial C4 plant belonging to the family Poaceae (subfamily Panicoideae, tribe Paniceae) and it is recognised as a particularly important cereal and fodder crop in arid and semi-arid regions [1]. This ancient crop, originating from China, is grown worldwide (especially in East and South Asia, Africa, and Europe) owing to its high tolerance to drought and salinity [2](Zhang et al., 2014). Foxtail millet is known for both its nutritional value and medicinal value, as its grains are rich in proteins, dietary fibres, phenolics, flavonoids, lysin, and minerals. It has also been recognised for its potential as a health-promoting functional food that helps reduce the risk of disease [3], reportedly by reducing blood glucose levels and controlling cholesterol in both healthy subjects and patients with diabetes [4]. Despite its health benefits, foxtail millet's cultivation has decreased significantly compared to other crops due to its lower yield and decreased economic potential [5]. However, increasing awareness of the importance of dietary diversification has led to increased market demand and new opportunities for this crop [6]. Thus, new cultivars with improved nutritional value and higher yield, especially salt-resistant cultivars, are in demand.

Foxtail millet germplasm has a high level of genetic diversity owing to the wide distribution of this plant and its adaption to various agro-climatic conditions [7]. Agronomic traits, such as grain morphology, are the targets of breeding programmes designed to improve both yield and quality as they largely affect crop productivity and responses to environmental stressors [8]. Understanding the genetic basis of phenotypic variation within the germplasm, especially for agronomic traits that are qualitatively important (such as salinity tolerance), has been a major focus in foxtail millet genetic studies. Notably,

most agronomic traits are quantitative, making it difficult to unravel the genetic basis for phenotypes of interest.

Linkage mapping, which identifies quantitative trait loci (QTL) that are closely related to complex traits, has been successfully applied and widely utilised in plants. A complementary approach is linkage disequilibrium (LD) mapping or association mapping. Linkage disequilibrium mapping is based on two strategies: i) re-sequencing of selected candidate genes and ii) genome-wide association, which exploits marker polymorphisms across all chromosomes [9]. Genome-wide association studies (GWAS) are widely utilised to identify valuable natural variations in trait-associated loci, as well as allelic variations in candidate genes underlying quantitative and complex traits, including those related to growth, development, stress tolerance, and nutritional quality [10]. Owing to its high resolution and cost-effectiveness, which are beneficial for gene discovery and molecular marker identification, GWAS have successfully identified numerous loci for complex traits in various plants, including wheat [11], maize [12], sugarcane [13], and alfalfa [14]. However, analyses of foxtail millet by GWAS are limited.

To better understand the diversity of foxtail millet, facilitate the genetic dissection of its agronomic and quality traits, and accelerate its marker-assisted breeding, high-quality single nucleotide polymorphisms (SNPs) were first detected in the present study using genotyping-by-sequencing of 107 foxtail millet accessions, and then used for GWAS.

Results

Double-digest restriction-associated DNA sequencing and variation detection

The double-digest restriction-associated DNA (ddRAD) sequencing carried out to genotype the 107 millet accessions generated 187 Gb of data with reads of 150 bp in length, on average. After quality control, 169 Gb of high-quality sequences were obtained and then mapped to the reference genome sequence of foxtail millet (GCF_000263155.2_v2.0, https://www.ncbi.nlm.nih.gov/assembly/GCF_000263155.2/). The mapping rates of the 107 accessions varied between 78.75% and 100% (Additional file 1: Table S1), and these data were then used for SNP calling by GATK. Initially, 506,788 SNPs were called for the 107 accessions, and 72,181 high-quality SNPs [coverage depth ≥ 3 , mapping quality ≥ 20 , missing ratio of samples within population $\leq 20\%$, and multiple allele frequency (MAF) ≥ 0.05] were retained for subsequent analyses. Of these, 71,947 SNPs were spread across nine chromosomes, with NC_028457.1 containing the most SNPs (13,010) and NC_028450.1 containing the least SNPs (4,854) (Fig. 1), while the remaining 234 SNPs were scattered across 31 scaffolds.

Population structure and linkage disequilibrium

To correct the GWAS model for population structure, we performed three analyses to determine the relationships between the 107 accessions. These were classified into four groups (G1, G2, G3, and G4) according to the results of the principal components analysis (PCA), neighbour-joining (NJ)-tree analysis, and ancestry structure analysis (Fig. 2). Groups G1, G2, G3, and G4 included 36, 14, 27, and 30

accessions, respectively. Whereas G1 primarily included accessions from northern China, G2 and G3 mainly contained accessions from northern and eastern China, respectively, and G4 included accessions distributed from western China to northeast China (Fig. 2A).

To determine the mapping resolution of the GWAS, the LD of the 107 accessions was analysed (Additional file 1: Fig. S1), and the LD decay rate was estimated as 100 kb ($r^2 = 0.2$).

Genome-wide association analysis

Based on the 72,181 high-quality SNPs obtained, we performed an association analysis for 16 traits, most of which presenting unimodal distributions (Fig. 3 and Additional file 1: Table S2). In total, 76 SNPs were significantly associated with these traits after Bonferroni correction (Fig. 4), and 53 significant SNPs were retained after filtering (Additional file 1: Table S3). These SNPs were distributed across eight chromosomes (no significant SNP was detected in NC_028454.1). Of these SNPs, 44 were located in intergenic regions, whereas one and eight SNPs were located in exon and intron regions, respectively. Up to 28 SNPs were associated with a single trait (for leaf width and time_of_peak_value, respectively). The 417 genes in the LD decay regions around each significant SNP were considered associated to the 15 traits. These genes included 23 genes encoding long non-coding RNAs and 10 genes encoding transfer RNAs. The number of genes associated with the different traits ranged from 2 to 64.

A favourable genotype is defined as one in which a significant SNP leads to an increase in phenotypic value (Additional file 1: Table S5). Eleven traits were associated with more than one significant SNP. To assess the potential pyramiding effects between favourable genotypes for traits, the mean phenotypic values of the accessions that contained multiple favourable genotypes were analysed. As shown in Fig. 5, the correlation between the number of favourable genotypes and phenotypic value was >0.4 for three traits (average, ear_length, and minium_of_viscosity_index). These findings suggested a certain degree of pyramiding effect between the favourable genotypes and these three traits.

Frequencies of favourable genotypes in accession groups

The frequencies of favourable genotypes at different significant SNPs were calculated for G1, G2, G3, and G4. The favourable genotypes of SNP locus NC_028452.1::13,868,887 (the 13,868,887th base of chromosome NC_028452.1) were all distributed in G2 (Fig. 6 and Additional file 1: Table S6). Favourable genotypes involving 17, 10, and 25 SNP loci were obtained from two, three, and four different accession groups, respectively. These results suggested that accessions in the different groups might have had different evolutionary or domestication directions.

Discussion

Genome-wide association analysis is an effective method for identifying trait-associated loci in natural populations. To date, only a few SNP-based association analyses in foxtail millet have been reported, with one SNP-based association analysis performed using low-depth (0.8×) re-sequencing [15]. In the

current study, we sequenced 107 millet accessions through ddRAD sequencing with an average coverage of 3.58×, and identified 53 SNP loci associated with 15 traits in the subsequent GWAS.

The accessions used in this study were collected primarily from China, with two accessions from the United States. Population structure analysis revealed that these accessions could be classified into four groups, although the obtained division did not correlate with the geographic origin of the accessions. Similar findings have been reported for other plant species, such as *Brassica rapa* L.[16] and *Matricaria chamomilla* L. [17]. This discrepancy might be explained by the transport routes and exchange of plant materials between regions.

Moderate LD decay is important for association analysis, ranging over several hundred kilobases between plants such as rice, soybean [18], and cotton [19]. The LD decay in the present study was estimated as 100 kb when r^2 decreased to 0.2, which is in agreement with previous findings [20](Zhang et al., 2014). This LD decay will be useful for identifying unknown genes that are linked to significant SNPs.

We employed a mixed linear model in the present association analysis where the significance threshold was estimated as $P = 10^{-4.86}$ after the Bonferroni correction (1/72,181). The *t*-tests used to analyse differences in phenotypic values of different genotypes at each significant SNP locus identified 53 loci associated with 15 traits and established the favourable genotypes for each locus. Marker-based gene pyramiding strategies have been demonstrated in several studies [20,21,22], and the favourable genotypes obtained in the present study also showed dosage pyramiding effects for several traits. The favourable genotypes identified here have substantial potential for future breeding programs, while the different frequencies of favourable genotypes in the different accession groups may imply that these accessions have had different evolutionary or domestication directions.

Conclusions

We identified a substantial number of SNP markers in foxtail millet and performed a GWAS to identify trait-associated loci. For a portion of the significant SNPs, favourable genotypes showed pyramiding effects for several traits. These favourable genotypes are expected to be useful for breeding programs aimed at foxtail millet improvement.

Methods

Plant materials and growth conditions

This study was conducted at the Mazhuang experimental station of Shandong Agricultural University, Taian, China, located at 36.02 °N, 117.00 °E, and 85 m above sea level. This site has a sub-humid, warm-temperate, continental, monsoon climate. The average annual temperature is 13 °C, the annual average precipitation is 688.3 mm, the sunshine duration is 2536.2 h, and the average frost-free period is 172.9 days. Basic physicochemical properties of the topsoil (0–20 cm) are as follows: pH 7.24 (soil:water ratio of 1:2.5); soil organic matter, 11.6 g kg⁻¹; total nitrogen, 0.75 g kg⁻¹; alkali-hydrolysable nitrogen, 72.3 mg

kg⁻¹; available phosphorus, 21.3 mg kg⁻¹; and available potassium, 78.6 mg kg⁻¹. Field experiments were conducted between June and September 2017, employing 107 foxtail millet accessions (collected from main cultivation areas of China or provided by the Chinese Crop Germplasm Resources Center of the Chinese Academy of Agricultural Sciences) as plant materials.

Compound fertiliser (N:P₂O₅:K₂O = 15:15:15) was applied at 1200 kg ha⁻¹ as a base fertiliser and the soil was tilled before sowing. Experimental plots of 40 m² were randomly arranged and seeds were sown at 7.5 kg ha⁻¹ on 28 June 2017. A prophylactic programme of herbicides was applied to control weed infestation. No significant incidences of disease, pests, or weeds affected foxtail millet throughout its growth stage. The spacing between rows was 40 cm, and three replicates were employed for each accession. All plots were thinned on 10 July 2017. Harvesting was carried out on 28 September 2017.

Measurements

Agronomic traits

Data on agronomic traits (plant height, stem diameter, leaf length, leaf width, chlorophyll SPD value, spike length, and spike diameter) were collected in the maturation period (27 September 2017). Ten random plants from each treatment plot were selected and the distances from the ground to the tips of the plants were measured and averaged per replication. Stem diameters were measured with a Vernier calliper and averaged per replication. The lengths and widths of flag leaves were measured and averaged as the leaf length per replication. Ten labelled leaves in each plot were selected to assess the chlorophyll SPD value by using a portable chlorophyll meter (Konica Minolta, Tokyo, Japan); five measurements per leaf were averaged for each replication. Ten random plants from each plot were hand-harvested to determine spike length, diameter, and weight; all spikes were harvested by hand to assess crop yield. Measurements were averaged per replication for each accession.

Quality traits

At maturity, all panicles were harvested from each replication, dried naturally, and dehulled. Quality traits, including grain length, width, and length/width ratio, were determined according to China National Standards (GB/T 17891-1999). The millet was ground in a stainless-steel grinder for 3 min and the resulting powder was used for chemical analysis. Amylose content was determined according to the Chinese National standard method GB7648-87, with minor modifications: 10 mg of millet powder was transferred into a 14-mL capped tube, dispersed in 0.1 mL 95% ethanol, and treated with 0.9 mL 1 M NaOH for 16 h at room temperature. Each sample was thoroughly mixed and 10 µL of the formed supernatant was pipetted into 96-well plates; after adding 190 µL of freshly prepared I₂-KI solution (3% iodine solution diluted 100 times in 0.01 M HCl before use) to each well, the plates were incubated for 10 min. Potato amylose (Sigma-Aldrich, St. Louis, MO, USA) treated in the same way was used to construct a standard curve. Amylose content was measured according to the absorbance of each sample at 620 nm (OD₆₂₀) and it was normalised based on sample weight.

Viscosity profiles were measured using a Rapid Visco Analyser (Super 3; Newport Scientific, Warriewood, Australia) following the procedure of the American Association of Cereal Chemists [23]. Data were recorded using the RVA-3D model Thermocline Windows Control 1.2 software (New Port Scientific, Sydney, Australia), and included peak viscosity, hot viscosity, cool viscosity, peak time, gelatinisation temperature, breakdown viscosity, setback viscosity, and consistency viscosity expressed in centipoises (cp).

DNA isolation and sequencing

Genomic DNA was extracted from leaves using the cetyltrimethylammonium bromide method [24]. Double-digest restriction-associated DNA libraries were prepared using 500 ng of DNA per sample, following the protocol described by Peterson et al. [25], with some modifications. Briefly, genomic DNA was digested with the restriction enzymes HindIII and BfaI at 37 °C for 5 h, followed by a ligation step whereby each sample was assigned one of 24 unique adaptors. Pooled digests of 24 individuals were run on an agarose gel; fragments ranging in size from 220 bp to 450 bp were manually excised and purified using a Zymoclean Gel DNA recovery kit (Zymo Research, Irvine, CA, USA). Each pool was amplified using 14 PCR cycles in 25-µL reactions containing 5 µL 5× reaction buffer, 5 µL 5× high GC enhancer, 0.25 µL Q5 polymerase, 4 µL library DNA, and a unique indexing primer for each pool that corresponds with the standard Illumina (Illumina Inc., San Diego, CA, USA) multiplexed sequencing protocol. The PCRs were carried out in a Veriti 96-well thermal cycler (Life Technologies, Carlsbad, CA, USA) using the following profile: initial denaturation at 98 °C for 30 s, 14 cycles at 98 °C for 15 s, 65 °C for 30 s, and 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. The DNA libraries were quantified using a high-sensitivity DNA analysis kit (No. 5067-4626; Agilent Technologies, Santa Clara, CA, USA) in a 2100 Bioanalyser (Agilent Technologies). Pools were combined in equimolar concentrations to form a single genomic library, and were sequenced in one lane of a HiSeq 2500 Illumina sequencer (pair-end, 2×150 bp).

Sequencing quality check and filtering

Adapter sequences were removed using AdapterRemoval 2 [26]. Reads with Phred scores <20 (average on sliding window), or with incorrect restriction sites, and reads of <50 bp were removed.

Sequence alignment and SNPs detection

Illumina paired-end reads were aligned to the Foxtail millet reference sequence (GCF_000263155.2_v2.0, https://www.ncbi.nlm.nih.gov/assembly/GCF_000263155.2/) using BWA-MEM [27] with default parameters, and SNPs were called using GATK [28](UnifiedGenotyper, stand_call_conf 30, stand_emit_conf 10) with subsequent filtering based on read map quality score (≥ 20), base quality score (≥ 5), and read depth (≥ 3).

Annotation of genetic variants

Variants were annotated using ANNOVAR 2016-02-01 [29] with gene-based annotation to assess whether SNPs or indels caused protein-coding changes and to identify which amino acids were affected.

Population structure and linkage disequilibrium analyses

We conducted both PCA and ancestry analysis to evaluate genetic structure using Plink 1.90 beta [30] and ADMIXTURE 1.3 ([http://www.genetics.ucla.edu /software/admixture/](http://www.genetics.ucla.edu/software/admixture/)), respectively. An individual-based NJ tree was constructed based on p-distance using TreeBest 1.92 (<http://treesoft.sourceforge.net/treebest.shtml>).

Genome-wide association study

In total, 107 accessions were used in GWAS for different traits. Association analysis was conducted with the genome-wide efficient mixed-model association (GEMMA) software package [31]. For mixed-linear-model analysis, the following equation was used:

$$y = Xa + S\beta + K\mu + e,$$

where y represents phenotypes; a and β are fixed effects representing marker effects and non-marker effects, respectively; and μ represents unknown random effects. X , S , and K are the incidence matrices for a , β , and μ , respectively.

For each significant SNP in the GWAS results, accessions with phenotypic value were classified into three groups according to their genotypes (reference homozygosis, heterozygosis, and altered homozygosis). Pairwise comparisons of phenotypic values between genotype groups were performed using t -tests, and significant SNPs were filtered according to the t -test results (at least one p -value ≤ 0.05).

Abbreviations

SNP: single nucleotide polymorphisms; GWS: genome-wide association; QTL: quantitative trait loci; LD: linkage disequilibrium; ddRAD: double-digest restriction-associated DNA; GEMMA: genome-wide efficient mixed-model association

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financially supported by the Modern Agriculture Industrial Technology Systems Project of Shandong Province (SDAIT-15-04), the Postdoctoral Innovation Project of Shandong Province (201603052), and the Postdoctoral funding by Shandong Agricultural University (76560). The funders had no role in study design, data collection, data analysis, data interpretation, the writing of the manuscript or decision to publish.

Author Contributions

YHL and YC contributed equally to this study and therefore are co-first authors. YPZG and LC conceived and designed the study. ZHL provided cDNA libraries and generated Illumina libraries. HW performed statistical analysis. FH, TTW, and MJS conducted the experiment, YHL and YC wrote and corrected the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors thank the Chinese Crop Germplasm Resources Centre of the Chinese Academy of Agricultural Sciences for providing foxtail millet accessions.

References

1. Zhang LZ, Liu RH. Phenolic and carotenoid profile and antiproliferative activity of foxtail millet. *Food Chem.* 2015;174:495-501. <https://doi.org/10.1016/j.foodchem.2014.09.089>
2. Zhang B, Li W, Chang X, Li R, Jing R. Effects of favorable alleles for water-soluble carbohydrates at grain filling on grain weight under drought and heat stresses in wheat. *PLoS One.* 2014;18:e102917. <https://doi.org/10.1371/journal.pone.0102917>
3. Liu KG, Qi SH, Li D, Jin CY, Gao CH, Duan SW, Feng B, Chen M. TRANSPARENT TESTA GLABRA 1 ubiquitously regulates plant growth and development from Arabidopsis to foxtail millet (*Setaria italica*). *Plant Sci.* 2017;254:60-69. <https://doi.org/10.1016/j.plantsci.2016.10.010>
4. Narayanan J, Sanjeevi V, Rohini U, Trueman P, Viswanathan V. Postprandial glycaemic response of foxtail millet dosa in comparison to a rice dosa in patients with type 2 diabetes. *Indian J. Med. Res.* 2017;144:712-717. https://dx.doi.org/10.4103%2Fijmr.IJMR_551_15

5. Tian BH, Luan SR, Zhang LX, Liu YL, Zhang L, Li HJ. Penalties in yield and yield associated traits caused by stem lodging at different developmental stages in summer and spring foxtail millet cultivars. *Field Crops Res.* 2018;217:104-112. <https://doi.org/10.1016/j.fcr.2017.12.013>
6. Goron TL, Raizada MN. Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front. Plant. Sci.* 2015;24:157. <https://doi.org/10.3389/fpls.2015.00157>
7. Han B, Huang XH. Sequencing-based genome-wide association study in rice. *Curr. Opin. Plant Biol.* 2013;16:133-138. <https://doi.org/10.1016/j.pbi.2013.03.006>
8. Biselli C, Bagnaresi P, Cavalluzzo D, Urso S, Desiderio F, Orasen G, Gianinetti A, Righettini F, Gennaro M, Perrini R, Ben-Hassen M, Sacchi GA, Cattivelli L, Valè G. Deep sequencing transcriptional fingerprinting of rice kernels for dissecting grain quality traits. *BMC Genomics.* 2015;16:1091. <https://doi.org/10.1186/s12864-015-2321-7>
9. Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Gaseneyer G, Kilian B, Graner A. Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biol.* 2012;12:16. <https://doi.org/10.1186/1471-2229-12-16>
10. Kumar V, Singh A, Mithra SVA, Krishnamurthy SL, Parida SW, Jain S, Tiwari KK, Kumar P, Rao AR, Sharma SK, Khurana JP, Singh NK, Mohapatra T. Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 2015;22:133-145. <https://doi.org/10.1093/dnares/dsu046>
11. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 2015;128:353-363. <https://doi.org/10.1007/s00122-014-2435-3>
12. Ding JQ, Ali F, Chen GS, Li HH, Mahuku G, Yang N, Narro L, Magorokosho C, Makumbi D, Yan J. Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. *BMC Plant Biol.* 2015;15:206. <https://doi.org/10.1186/s12870-015-0589-z>
13. Gouy M, Rousselle Y, Chane AT, Anglade A, Royaert S, Nibouche S, Costet L. Genome wide association mapping of agromorphological and disease resistance traits in sugarcane. *Euphytica.* 2015;202:269-284. <https://doi.org/10.1007/s10681-014-1294-y>
14. Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, Annicchiarico P. Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS One.* 2017;12:e0169234. <https://doi.org/10.1371/journal.pone.0169234>
15. Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H, Zhu C, Lu Y, Zhou C, Fan D, Weng Q, Guo Y, Huang T, Zhang L, Lu T, Feng Q, Hao H, Liu H, Lu P, Zhang N, Li Y, Guo E, Wang S, Wang S, Liu J, Zhang W, Chen G, Zhang B, Li W, Wang Y, Li H, Zhao B, Li J, Diao X, Han B. A haplotype map of

genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat. Genet. 2013;45:957-961. <https://doi.org/10.1038/ng.2673>

16. Tanhuanpää P, Erkkilä M, Tenhola-Roininen T, Tanskanen J, Manninen O. SNP diversity within and among *Brassica rapa* accessions reveals no geographic differentiation. Genome. 2016;59:11-21. <https://doi.org/10.1139/gen-2015-0118>

17. Pirkhezri M, Hassani ME, Hadian J. Genetic diversity in different populations of *Matricaria chamomilla* L. growing in southwest of Iran, based on morphological and RAPD markers. J. Med. Plants. Res. 2010;4:1-13. <https://doi.org/10.3923/rjmp.2010.1.13>

18. Gupta PK, Rustqi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 2005;57:461-485. <https://doi.org/10.1007/s11103-005-0257-z>

19. Ma ZY, He SP, Wang XF, Sun JL, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, Yan Y, Jia Y, Yang J, Pan Z, Gu Q, Li X, Sun Z, Dai P, Liu Z, Gong W, Wu J, Wang M, Liu H, Feng K, Ke H, Wang J, Lan H, Wang G, Peng J, Wang N, Wang L, Pang B, Peng Z, Li R, Tian S, Du X. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat. Genet. 2018;50:803-813. <https://doi.org/10.1038/s41588-018-0119-7>

20. Zhang S, Tang CJ, Zhao Q, Li J, Yang LF, Qie L, Fan X, Li L, Zhang N, Zhao M, Liu X, Chai Y, Zhang X, Wang H, Li Y, Li W, Zhi H, Jia G, Diao X. Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in Foxtail millet [*Setaria italica* (L.) P. Beauv.]. BMC Genomics. 2014;15:78. <https://doi.org/10.1186/1471-2164-15-78>

21. Werner K, Friedt W, Ordon F. Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol. Breed. 2005;16:45-55. <https://doi.org/10.1007/s11032-005-3445-2>

22. Sacco A, DiMatteo A, Lombardi N, Trotta N, Punzo B, Mari A, Barone A. Quantitative trait loci pyramiding for fruit quality traits in tomato. Mol. Breed. 2013;31:217-222. <https://doi.org/10.1007/s11032-012-9763-2>

23. Sun XJ, Li WH, Hu YY, Zhou XJ, Ji MY, Yu DD. Comparison of pregelatinization methods on physicochemical, functional and structural properties of tartary buckwheat flour and noodle quality. J. Cereal. Sci. 2018;80:63-71. <https://doi.org/10.1016/j.jcs.2018.01.016>

24. Uncu AO, Torlak E, Uncu AT. A cost-efficient and simple plant oil DNA extraction protocol optimized for DNA-based assessment of product authenticity. Food Anal. Methods. 2018;11:939-950. <https://doi.org/10.1007/s12161-017-1070-4>

25. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7:e37135. <https://doi.org/10.1371/journal.pone.0037135>
26. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes. 2016;9:88. <https://doi.org/10.1186/s13104-016-1900-2>
27. Madoui MA, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A, Lemainque A, Wincker P, Aury JM. Genome assembly using Nanopore-guided long and error-free DNA reads. BMC Genomics. 2015;16:327. <https://doi.org/10.1186/s12864-015-1519-z>
28. Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV, Kristinsson KG, Gottfredsson M, Barrett JC, Gudmundsson LJ, Blondal K, Gylfason A, Gudjonsson SA, Helgadóttir HT, Jonasdóttir A, Jonasdóttir A, Karason A, Kardum LB, Knežević J, Kristjánsson H, Kristjánsson M, Love A, Luo Y, Magnusson OT, Sulem P, Kong A, Masson G, Thorsteinsdóttir U, Dembic Z, Nejentsev S, Blondal T, Jonsdóttir I, Stefansson K. *HLA* class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat. Genet. 2016;48:318-322. <https://doi.org/10.1038/ng.3498>
29. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164. <https://doi.org/10.1093/nar/gkq603>
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007;81:559-575. <https://doi.org/10.1086/519795>
31. Zhou X, Stephens M. Genome-wide efficient mixed model analysis for association studies. Nat. Genet. 2012;44:821. <https://doi.org/10.1038/ng.2310>

Figures

Figure 1

Distribution of the SNPs across the nine chromosomes.

Figure 2

Geographic distribution, principal components analysis (PCA), neighbour-joining (NJ) phylogenetic tree, and ancestry structure of the 107 foxtail millet accessions examined here. (A) PCA plot of the first two components. (B) Upper panel, NJ-phylogenetic tree of the 107 accessions used in this study; lower panels, results of ancestry structure analysis with $K = 4$. (C) Geographic distribution of the 105 accessions from China. The pie charts on the map show the proportions of the four groups (G1, G2, G3, and G4) at the geographical locations. Blue, yellow, purple, and red represent G1, G2, G3, and G4, respectively.

Figure 3

Frequency distributions of the phenotypic values of the 16 traits of the 107 foxtail millet accessions.

Figure 4

Genome-wide association study of the 16 traits of the 107 foxtail millet accessions.

Figure 5

Linear regression analyses of the numbers of favourable genotypes on the different phenotypic values.

Figure 6

Distribution of favourable genotypes in G1, G2, G3, and G4 groups.

Supplementary Files

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

- [TableS5.xlsx](#)
- [TableS4.xlsx](#)
- [TableS3.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)

- [FigS1.tif](#)