Assessment and phenotypic identification of millet germplasm (*Setaria italica* [L.]) in Liaoning, China (#99850)

First submission

Guidance from your Editor

Please submit by **24 May 2024** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the materials page.

1 Figure file(s)
1 Table file(s)
1 Raw data file(s)
3 Other file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Assessment and phenotypic identification of millet germplasm (*Setaria italica* [L.]) in Liaoning, China

Xin-tong Li [1], **Weifeng He** [2], **Honghao Wang** [2], **Min Xu** [Corresp. 2]

[1] School of Accounting, GuiZhou University of Finance and Economics, Guiyang, Guizhou, China

[2] Cash Crop Institute, Liaoning Academy of Agricultural Science, Liaoyang, Liaoning, China

Corresponding Author: Min Xu
Email address: shumin690101@163.com

[Aims]This study evaluated millet germplasms in Liaoning Province to support the collection, preservation and innovation of millet germplasm resources. **[Methods]**The study was conducted from 2018 to 2020, involved the selection of 105 millet germplasm resources from the Germplasm Bank of the Liaoning Academy of Agricultural Sciences (LAAS), the observation and recording of 31 traits, and the application of multivariate analysis methods to assess phenotypic diversity. **[Results]** From the diversity analysis and correlation analysis, it was found that the tested traits had abundant diversity and complex correlations among them. principal component analysis (PCA) comprehensively analyzed all quantitative traits and extracted 7 principal components. Grey relational analysis (GRA) highlighted the varied contributions of different traits to yield. Through systematic cluster analysis（SCA）, the resources were categorized into six groups at Euclidean distance of 17.09. K-mean cluster analysis determined the distribution interval and central value of each trait, then identified resources with desirable traits. **[Conclusion]**The results revealed resources that possess characteristics such as upthrow seedling leaves, more tillers and branches, larger and well-formed ears, and lodging resistance prefer to higher grain yield. It was also discovered that the subear internode length (SIL) could be an indicator for maturity selection. Four specific resources, namely, Dungu No. 1, Xiao-li-xiang, Basen Shengu, and Yuhuanggu No. 1, were identified for further breeding and practical applications.

1. Introduction

Millet (*Setaria italica* [L.]) is a diploid ($2n = 2X = 18$) species of foxtail grass (*Gramineae*, *Setaria*). Millet, which originated in China, is one of the world's oldest cultivated crops, with a cultivation history of 5000 to 8000 years (Lu, 2009). Millet possesses attributes such as drought resistance, water efficiency, high light utilization capacity, high storage convenience, and dual utility as both a grain and a grass (Yang, 2012). The millet seed kernel is a reservoir of well-balanced nutrients (Diao, 2011), comprising ample proteins (Liu, 2009) and vitamins (Liu, 2013). It is commonly chosen as the primary dietary option for new mothers or recovering patients.

According to statistical data, there are more than 40,000 millet resources worldwide. Among these, the country with the richest germplasm resources is China, specifically the National Crop Germplasm Bank, which stores over 26,000 resources, accounting for 70% of the global collection (Liu, 2009; Doust, 2009). Previous researchers have made significant progress in researching and utilizing millet germplasm resources. These studies primarily focused on genetic diversity analysis of phenotypic traits. For instance, Wang et al. (2016) comprehensively evaluated 15 phenotypic traits in 878 millet resources globally and identified 8 key indicators, such as leaf sheath color, ear length, seed color, and kernel color, for phenotype identification. Tian et al. (2010) investigated the genetic diversity of 482 millet varieties in Henan and Shandong provinces and discovered that the diversity level of millet breeding cultivars was considerably lower than that of local varieties, suggesting certain traits of greater significance in the breeding process. Li et al. (1996) examined 23,381 Chinese landrace millet samples and conducted a comprehensive analysis of 11 agronomy traits. Of these traits, only seedling leaf color, starch composition, and 1000-grain weight showed significant regional differences in phenotypic diversity indices. In addition, molecular biotechnology has been applied in the study

of millet germplasm resources. For instance, Yang et al. (2003) and Schontz et al. (1998) observed abundant genetic diversity among millet resources from various regions. Jia et al. (2013) conducted genome resequencing to analyse the diversity of 916 core millet germplasms, providing insight into the geographical distribution of millet genetic resources.

According to Liu et al. (2019), the primary millet-producing regions are concentrated in the northern and eastern parts of China. The millet produced in Liaoning, an important millet-producing area, is renowned for its golden-red grain color, pleasant taste, and high quality. In comparison to molecular markers, simpler and more intuitive indicators are required to evaluate resource materials in practical production. Therefore, this study aimed to comprehensively understand the basic situation of millet germplasm resources in Liaoning Province, to screen agronomy traits as selection indicators, and to promote millet scientific research and production in northeastern China.

2. Materials and Methods

2.1 Experimental materials

A total of 105 millet resources preserved in the Germplasm Bank of the Liaoning Academy of Agricultural Sciences (LAAS) were selected for this study. These resources originated from various regions, including Liaoning (52), Beijing (20), Hebei (8), Jilin (9), Inner Mongolia (9), Heilongjiang (2), and Shanxi (5). Approximately half of the materials are from Liaoning, with the remaining resources sourced from provinces that north of the Yellow River.

2.2 Experimental design

The experiment was conducted over three growing seasons, from 2018 to 2020, at the experimental site of the Cash Crops Institute of Liaoning, located in Liaoyang City, Liaoning Province. The soil at the site is sandy loam, with the following composition in the topsoil: 1.97%

organic matter, 0.08% total nitrogen, 73.4 mg/kg alkali-hydrolyzable nitrogen, 23.6 mg/kg available phosphorus, and 247.5 mg/kg available potassium. The resources were arranged sequentially during sowing, with each resource planted in four rows with 3~4 cm plant spacing The length of each row was 5 m, with a spacing of 50 cm, resulting in a plot area of 10 square meters. Sowing took place around May 11th, and harvesting was performed around September 25th. A three-compound fertilizer (N:P:K=15%:15%:15%) was applied as a base fertilizer. Ploughing and weeding were conducted three times during the growth period.

Two points were selected from each plot, and five consecutive plants with uniform growth were chosen from each point as samples. Quality and quantitative traits were investigated following the Descriptors and Data Standard for Millet (*S. italica* [L.]) compiled by Lu (2006). A total of 12 quantitative traits and 19 quality traits were selected for statistical analysis，ignoring traits that show no difference between varieties. The mean value of the ten plants and the interannual mean value were calculated.

2.3 Data processing

Data collection and analysis were performed using EXCEL 2007 and DPS 9.5 (Tang, 2010). Phenotypic diversity analysis and systematic cluster analysis(SCA)were conducted based on all 31 traits, while the Shannon diversity index was analysed using the method described by Wang et al. (2021). Correlation analyses, principal component analyses(PCA), gray relational analysis(GRA), and K-means cluster analysis were performed based on the 12 quantitative traits.

Before conducting SCA and GRA, all the data were standardized using the following formula: $x_{ij}^{'} = (x_{ij} - \overline{x}_j)/S_j$, $\overline{x}_j = \sum_{i=1}^{n} x_{ij}/n$, and $S_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)}$. After standardization, a new data series was obtained with a dimension of 1, a mean value of 0 and a variance of 1.

3. Results and analysis

3.1 Phenotypic diversity analysis of millet germplasm resources

Table 1 shown the Shannon diversity index (DI) and distribution frequency of phenotypic characters of 19 quality traits. By comparing the frequency, it was found that two traits—tiller habit (TH) and branch habit (BH)—showed weak, moderate, and strong characteristics with a uniform distribution, resulting in a greater diversity index (DI). The phenotypic distribution of the other 17 traits showed clear tendencies, with a focus on 1 or 2 characteristics, resulting in a lower DI. It is worth noting that traits related to pigmentation show obvious distribution tendencies, with traits related to leaves mainly being green and traits related to flowers primarily being yellow.

Table 1： Phenotypic diversity of 19 quality traits

The phenotypic diversity analysis of 12 quantitative traits is presented in Table 2. The coefficient of variation (CV), which represents the degree of trait dispersion, ranged from 6.34% to 43.77% for the 12 traits. The highest CV was observed for stem number per plant (SNP), followed by main stem node number (MSN), seed weight per plant (SWP), and ear weight per plant (EWP), indicating a high degree of variation and potential for genetic improvement. On the other hand, the subear internode length (SIL) and growing period (GP) had CVs close to or less than 10%, suggesting lower dispersion and relatively stable performance among the varieties. The Shannon DI, which reflects the distribution of trait performance, ranged from 1.0566 to 2.0428. The main stem length (MSL), EWP, SWP, and grass weight per plant (GWP) had DI values close to or greater than 2.0, indicating these traits performance are concentrated, reflecting a simple genetic basis,. Conversely, SNP had the lowest DI value, close to 1.0,indicating these traits performance were relatively scattered and susceptible to external conditions

Table 2： Phenotypic diversity of 12 quantitative traits

93 *3.3 Correlation analysis of millet germplasm resources*

94 Correlation analysis aims to study the association between two or more random variables of
95 equal status. In this experiment, correlation analysis was performed based on the 12 quantitative
96 traits (Table 3). The EWP showed a significant positive correlation with SWP and both
97 significant positive correlated with the main ear length (MEL), main ear diameter (MED), and
98 SNP. also showed a highly significant negative correlation with the MSN. EWP had a significant
99 positive correlation with MSL and GP, while SWP showed significant positive correlations with
100 these traits. These results indicate that grain yield is positively correlated with ear size (MEL and
101 MED) and ear setting potential (SNP and GP). Additionally, GWP was positively correlated with
102 the MSN and main stem diameter (MSD), which represent vegetative growth. GP, which
103 represents the growth potential of the plant, showed positive correlations with SNP, MSL, MEL,
104 SIL, and MED. Overall, the results of the correlation analysis align with expectations.

105 **Table 3： Correlation analysis of 12 quantitative traits**

106 *3.4 Principal component analysis (PCA) of quantitative traits*

107 Principal component analysis (PCA) is a comprehensive method used for transforming
108 multiple trait indices and reducing dimensionality into several principal components. Each
109 principal component represents a relatively independent indicator system, there is no correlation
110 between each principal component, the numerical value is intuitive and easy to analyze (Shi,
111 2019). The PCA results for 12 quantitative traits from 105 millet germplasms are presented in
112 Table 4. Seven principal components with eigenvalues greater than or close to 1 were selected,
113 which collectively contributed to 81.63% of the variance and encapsulated most of the genetic
114 information of the millet germplasms (Qiao, 2015; Liu, 2020).

The eigenvector value of each trait indicates its contribution to the principal component (PC). By comparing the eigenvector values, the following observations were made: the 1st PC was primarily associated with MEL(-)-, MSL(+)-, seed weight per plant (SWP +), GWP(-) -, and GP(-) -; the 2nd PC was mainly influenced by MED(+), EWP(+), and MSN(+); the 3rd PC was primarily influenced by SNP(+), 1000-seed weight (TSW-), and EWP(+). the 4th PC was mainly influenced by MED(-), GP(+), GWP(+) and EWP(-); the 5th PC was mainly influenced by MSD(+), SIL(-), MSN (+), MEL (+) and SWP(+); the 6th PC was primarily influenced by MSL (-), SIL (+), SWP (+), and EWP (+); and the 7th PC was mainly influenced by MEL (+) and MSN (-). Compared to correlation analysis, PCA provided an interpretation of the correlation between quantitative traits of plants and specifically emphasized the importance of the GP and TSW.

Table 4： Principal components and eigenvalues of 12 quantitative traits

3.5 Gray relational analysis (GRA) of quantitative traits and yield

Gray correlation analysis refers to the quantitative description and comparison of the development of a system to measure the degree of correlation between factors based on the similarities or differences in their development trends (Deng, 2002). After standardized processing, the seed weight per plant (SWP) was taken as the reference column, and the correlation between the main quantitative traits and the SWP was analysed. Among the tested quantitative traits, EWP had the greatest impact on the SWP, followed by SNP, MEL, and MED, these four traits directly represent the ear-bearing capacity of the plant. The coefficients of three traits, namely, GP, MSL, and SIL, were very close, these three traits are mainly related to maturity and main stem growth. The MSD, GWP, and MSN represent the vegetative growth status of the plant, and their contributions to the SWP decrease in turn. Therefore, when selecting

high-yield millet varieties, it is advisable to focus on materials with more tillers, larger ears, and longer growth periods.

Table 5：Gray relational analysis of seed yield and 11 quantitative traits

3.6 *System cluster analysis (*SCA) *of all traits*

Cluster analysis reflects the genetic differences among different varieties, and clustering resources with similar traits into one group can guide the selection of hybridization parents (Bakhsh et al*.,* 2017). The selection of resources from different groups with large genetic distances for hybridization can lead to substantial genetic variation (Santhosha et al., 2011; Zhu et al*.,* 2012).

Based on the performance of all 31 traits (12 quantitative traits + 19 quality traits) of 105 millet resources, SCA analysis was conducted via standardized data transformation - Euclidean distance - deviation square sum method. All materials were divided into six groups at Euclidean distance = 17.09 (Fig. 2).

Fig. 1：System cluster analysis of Setaria italica (L.) germplasm

The quantitative traits of different groups were analysed and compared. The results are presented in Table 6. The 1st group consisted of 11 samples, which exhibited mid-earlier maturity, fewer tillers, taller and slender main stem with multiple nodes, smaller main ears, small seeds with medium grain yield, and higher grass yield. The 2nd group included 30 materials, showing earlier maturation, fewer tillers, lower and slenderer main stem with multiple nodes, smaller main ears, larger seeds but lower grain yield, and higher grass yield. The 3rd group consisted of 33 materials displaying mid-later maturation, medium tiller ability, taller main stem with multiple nodes, medium-sized main ears, larger seeds with higher grain yield, and medium grass yield. The 4th group comprised 6 materials, showing later maturation, multiple tillering, taller and

stronger main stems with multiple nodes, larger main ears, smaller seeds but higher grain yield, and greater grass yield. The 5th group contained 5 materials, exhibiting mid-later maturation, multiple tillering, lower and stronger main stem with fewer nodes, shorter and thicker main ears, medium-sized grains, lower grain yield and lower grass yield. The 6th group consisted of 20 materials that exhibited mid-later maturation, multiple tillering, taller and slenderer main stems, smaller main ears, smaller seeds but higher grain yields, and lower grass yields.

Table 6： System cluster analysis of 12 quantitative traits

In the analysis of 12 quantitative traits in Table 6, the qualities of each group were also compared. The 1st group had unique characteristics in terms of leaf sheath color (LCS) and bristle length (BL) that were not found in the other groups. The 2nd group displayed common characteristics for all traits. The 3rd group exhibited abundant traits characteristics, including red seed color (SC), which was not present in the other groups. The 4th group, with its 6 resources, had relatively simple and concentrated trait phenotypes but also had a rare occurrence of purple bristle color (BC). The 5th group, consisting of only 5 resources, had scattered trait phenotypes, with kernel color (KC), SC, and SH showing unique characteristics. The 6th group showed abundant trait characteristics, with LR and ear shape (ES) displaying characteristics not found in the other groups.

Further examination of the geographical distribution of resources within each group revealed that resources from Liaoning had certain advantages, particularly in Group 4, which exclusively consisted of materials from Liaoning. Resources from Group 1 were also concentrated in Liaoning. Group 2 resources were relatively dispersed and covered almost all geographical origins. Groups 3 and 6 mainly sourced their resources from Liaoning, Beijing, and Inner Mongolia. Group 5 resources were evenly distributed from Liaoning and Jilin.

3.6 K-mean clustering analysis

184 K-means clustering analysis was conducted to analyse the 12 quantitative traits, and the
185 theoretical distribution intervals and central values of these traits are shown in Table 7. The
186 performance of each trait for the studied resources was found to be concentrated near the central
187 value, with more distribution around the minimum value and less around the maximum value.

188 **Table 7： K-mean cluster analysis of 12 quantitative traits**

189 Based on these results, specific resource materials with desirable performance in terms of
190 maturity, plant height, tiller habit, ear size, and grain weight were identified. After a
191 comprehensive evaluation, four specific resources were selected: Dungu No. 1, an early-
192 maturing and draft small-ear and small-seed resource from Taiyuan, Shanxi Province; Xiao-li-
193 xiang, a small-seed and early-maturing resource from Shijiazhuang, Hebei Province; Basen
194 Shengu, a later-maturing and taller large-ear resource from Fuxin, Liaoning Province; and
195 Yuhuanggu No. 1, a later-maturing and low-yield large-seed resource from Chifeng, Inner
196 Mongolia.

197 **4. DISCUSSION**

198 This experiment examined a total of 31 traits, which showed rich diversity in both
199 qualitative and quantitative aspects. Among the 19 qualitative traits, traits related to plant color,
200 such as KC, SC, and LSC, exhibited lower Shannon DI values, suggesting a noticeable tendency
201 toward pigmentation in the local resources. Leaf-related traits are shown in green, while ear-
202 related traits are shown in yellow. Traits representing ear characteristics, such as ear
203 compactness (EC, SD), bristle length (BL), displayed moderate DI values, indicating that the
204 ears of local resources are primarily sparse and loose, making them prone to grain drop. The BL
205 was short, and the peduncle shape was curved, with the ear shape predominantly cylindrical and
206 spindle shaped. Traits representing plant structure, such as TH, BH, and blooming leaf attitude
207 (BLA), showed higher DI values, indicating greater diversity in plant structure types.

For the 12 quantitative traits examined, the DI values ranged from 1.0566 to 2.0428, and the CVs ranged from 6.34% to 45.84%. GP and TSW exhibited small DI values, suggesting strong adaptability through long-term selection and limited potential for genetic improvement. This finding is consistent with the results obtained by Gao et al. (2020) in their study on mung beans. Yield traits, such as EWP, SWP, and GWP, which reflect plant growth capacity, exhibited relatively high CV and DI values. This indicates that these traits are influenced by multiple quantitative genes, are susceptible to external conditions, and have great potential for genetic improvement. This conclusion aligns with the research findings of Wang et al. (2009) and Wang et al. (2021). The results of the correlation analysis indicate that ear size, SNP, and GP are closely correlated with grain yield, while no direct correlation is found between traits representing plant growth status and grain yield. These findings are consistent with those of Jia et al. (2017) but differ from those of Yan et al. (2010), which can be attributed to variations in test sites, sampling methods, and measurement indices. Additionally, the correlation of qualitative traits was also analyzed by assigning values based on their phenotype (result supplied), revealing the correlation between traits associated with pigmentation, and represented BH, TH, Spike density (SD), Shattering habit (SH), Seedling leaf attitude (SC) and Lodging resistance (LR) exhibited correlations with yield, consistent with the results of the diversity analysis.Since the correlation of qualitative traits is completed by assigning values on each phenotype, human factors influence greatly, the analysis results can be provided as a reference in the work only.

PCA revealed the extraction of 7 PCs. The 1st, 2nd, 3rd, 4th, and 6th PCs focus on yield and explain the associations between yield and main stem growth ability, main ear length, and growth period. The 5th and 7th PCs highlight the relationship between main stem growth ability and main ear length. In the breeding process, it is important to consider the contribution rate of

each PC and the breeding goal comprehensively.

GRA based on GWP demonstrated that ear size and ear-bearing capacity are the main factors limiting yield, followed by growth period traits.

Overall, the results of correlation analysis, PCA, and GRA were consistent, suggesting that high-yield lines are expected to have more tillers and branches, larger and well-formed ears, and lodging resistance. It is worth noting that TSW does not directly influence yield, possibly due to millet's seed shatter characteristics and the presence of immature seeds at harvest. These findings are similar to those obtained by Jia et al. (2021) in their study on Tartary buckwheat, where the length of subear internode (SIL) was found to be proportional to the GP, making it a potential indicator of maturity.

A systematic cluster analysis was conducted on all 31 traits, resulting in the division of 105 millet materials into six groups at a Euclidean distance of 17.09. Each group exhibited distinct phenotypic characteristics and showed certain geographical distribution tendencies. According to previous studies on crops such as adzpea (Pu et al., 2003) and cotton (Xu et al., 2017), the characteristics of these groups were found to be correlated with the natural climate conditions of their original source areas, which could be categorized using cluster analysis. The geographical distribution tendencies among groups were not prominently observed in this experiment, due to the frequent introduction of resources between regions and intermixing of bloodlines.

In addition, K-means cluster analysis was performed on 12 quantitative traits. Theoretical distribution intervals and central values were calculated to identify resource materials with specific traits. Fuether more,, four specific resources were identified: one with early maturity and draft, small ear size, and small seeds; one with small seeds and early maturity; one with later maturity, taller and larger ears; and one with later maturity,,lower yields, and larger seeds.

5. Conclusion

In conclusion, this study evaluated the phenotypic diversity of 105 millet germplasms in the Liaoning area by considering 19 qualitative traits and 12 quantitative traits. The results revealed a rich diversity of traits and complex correlations among them. Resources that possess characteristics such as upthrow seedling leaves, more tillers and branches, larger and well-formed ears, and lodging resistance prefer to higher grain yield. It was also discovered that the subear internode length (SIL) could be an indicator for maturity selection. Furthermore, all resource materials were divided into six groups with different phenotypic characteristics, and the distribution interval of each quantitative character was determined. Four specific resources, namely, Dungu No. 1, Xiao-li-xiang, Basen Shengu, and Yuhuanggu No. 1, were identified for further breeding and practical applications.

Author Contribution Statement

LI X. T., HE W.F., WANG H.H. performed experiments.

HE W.F. supplied germplasm resources materials.

WANG H.H. analyzed data.

LI X.T. wrote the paper.

XU M. planned and supervised the research.

Funding:

This study was financially supported by philosophy and Social Sciences Planning Project Guizhou （Grant number:21GZYB62）

Conflict interest:

There are no direct competition and conflict interest of our work.

REFERENCE

Bakhsh A, et al. 2017. Use of RAPD markers in comparison with agro-morphological traits for estimation of diversity among chickpea genotypes. International Journal of Agriculture and Biology. 19: 427-431.

Deng J L. 2002. Fundamentals of Grey Theory (Grey System Theory Series). *Wuhan: Huazhong University of Science and Technology Press*. (in Chinese)

Diao X M. 2011. Chinese Industry and Technical System of Foxtail Millet. *Beijing: China Agricultural Sciences and Technology Press.* 20-30. (in Chinese)

Doust A et al., 2009. Foxtail millet: A sequence-driven grass model system. *Plant Physiology*, 149(1):137-141.

Gao Y Q, et al., 2020. Screening and Evaluation of Germplasm Resources of MungBean．*Seed*. 39(08)：66-69. (in Chinese)

Jia G, et al., 2013. A haplotype map of genomicvariations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). *Nature Genetics,* 45: 957-961.

Jia R L, et al., 2021. Genetic Diversity Analysis and Comprehensive Assessment of Agronomic Traits of 64 Tartary Buckw heat Germplasms．*Crops*，(3)：19-27(in Chinese)

Jia X P, et al., 2017. Grey Incidence Analysis and Comprehensive Evaluation of 71 Miliet Germplasm Resources in China．*Seed*，36(9)：63~66. (in Chinese)

Li Y, et al., 1996. A phenotypic diversity analysis of foxtail millet (Setaria italica (L.) P. Beauv.) landraces of Chinese origin. *Genetic Resources and Crop Evolution,* 43(4): 377-384.

Liu J A, et al., 2019. Spatiotemporal variation and regional advantages of foxtail millet production in recent 30 years in China．*Scientia Agricultura Sinica*，52(11): 1883－1894．(in Chinese).

Liu M X, Lu P. 2013. Distribution of vitamin E content and its correlation with agronomic trait sand carotenoids content in foxtail millet varieties in China. *Acta Agronomica Sinica,* 39(3): 398-408. (in Chinese)

Liu S C, et al.2020. Comprehensive Evaluation of Agronomic Traits and Quality Traits of Foxtail Millet Landrace in Shanxi. *Scientia Agricultura Sinica*. 53(11):2137-2148.

Liu X, et al., 2009. Geographic Distribution and Germplasm-rich Region of Cereals in China. *Journal of Plant Genetic Resources*, 10(1):1-8. (in Chinese)

Lu H Y, et al.,2009. Earliest domestication of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago. *PNAS*，106(18):7367-372.

Lu P. 2006. Descriptors and data standard for foxtial millet [Setaria italica (L.) Beauv.]. Beijing: China Agriculture Press. (in Chinese).

Pu S J, et al., 2003. Study on Adzuki Bean Variety Sources of North China Ⅶ The Genetic Parameters and Correlation Analysis of Agronomic Traits of Adzuki Bean in Beijing Environment. *Journal of Beijing University of Agriculture*, 18(3):174-177. (in Chinese)

Qiao L, et al. 2015. .Genetic Diversity of Foreign Mungbean Germplasm Resources by Agronomic Characters. *Journal of Plant Genetic Resources*, 16(5) : 986-993. (in Chinese)

Santhosha H M,et al. 2011. Genetic diversity in early cauliflower (Brassica oleracea var. botrytis L.) germplasm. *Journal of Horticultural Sciences*, 6(1): 21-24

Schontz D, Rether B. 1998. Genetic variability in foxtail millet, Setaria italica (L.) P. Beauv. - RFLP using a heterologous rDNA probe. *Plant Breeding*, 117(3): 231-234.

Shi J J, et al. 2019. Genetic diversity analysis and comprehensive evaluation of introduced Avena sativa L. germplasm resources in the western area of Heilongjiang. *Journal of Southern Agriculture.* 50(3): 515-523. (in Chinese)

Tang Q Y. 2010. DPS data processing system-Experimental Design, Statistical Analysis and data Mining (Second edition). *Beijing: Science Press,*China. (in Chinese)

Tian B H. 2010. Genetic diversity of landrace and improves cultivars in foxtail millet. *Journal of Plant Genetic Resources*, 11(2):224-228. (in Chinese)

Wang H G, et al., 2016. Phenotypic Diversity Evaluations of Foxtail Millet Core Collections. *Acta Agronomica Sinica*, 42(1): 19-30. (in Chinese)

Wang H G, et al., 2018. Phenotypic Diversity of Local Foxtail Millet Varieties in Shanxi. *Chinese Agricultural Science Bulletin*, 34(32):19-24. (in Chinese)

Wang X J, et al., 2009, Analysis of genetic diversity of local germplasm resources for foxtail millet in Gansu. *Agricultural Research in the Arid Areas*，27(6)：129~133,153. (in Chinese)

Wang Y L, Li J G, Wang L Q, Deng J, Tang R, Yu Y H. (2021). Evaluation on phenotypic variation of rice bean germplasm collected from Hunan Province[J]. *Journal of Plant Genetic Resources*. 22(2): 317-327.

Xu M, et al.,2017. Clustering and Correlation Analysis of Earlier-Mauture Cotton Innovation Germplasm based on Biological Traits. *Crops,* 1:25-31. (in Chinese)

Yan F. 2010. Grey Correlation Degree Analysis and Comprehensive Evaluation of Serotinous Millet Germplasm Resources in Heilongjiang Province. *Heilongjiang Agricultural Sciences*, (7):30~33. (in Chinese)

Yang T Y, et al., 2003. Genetic variation among varieties of foxtail millet from different regions in China based on RAPD markers. *Acta Botanica Boreali-Occidentelia Sinica,* 23(5): 765-770. (in Chinese)

Yang X Y, et al., 2012. Early millet use in northern China. *Proceedings of the National Academy of Sciences of the United States of America*, 109: 3726-3730.

Zhu S Y, et al 2012. Principal component analysis and cluster analysis for main morphological characteristics of cauliflower

371 inbred lines. *Journal of Plant Genetic Resources.* 13(1): 77-82.(in Chinese)

372

373

374

Figure 1

System Cluster Analysis

System Cluster Analysis of all 31 Traits

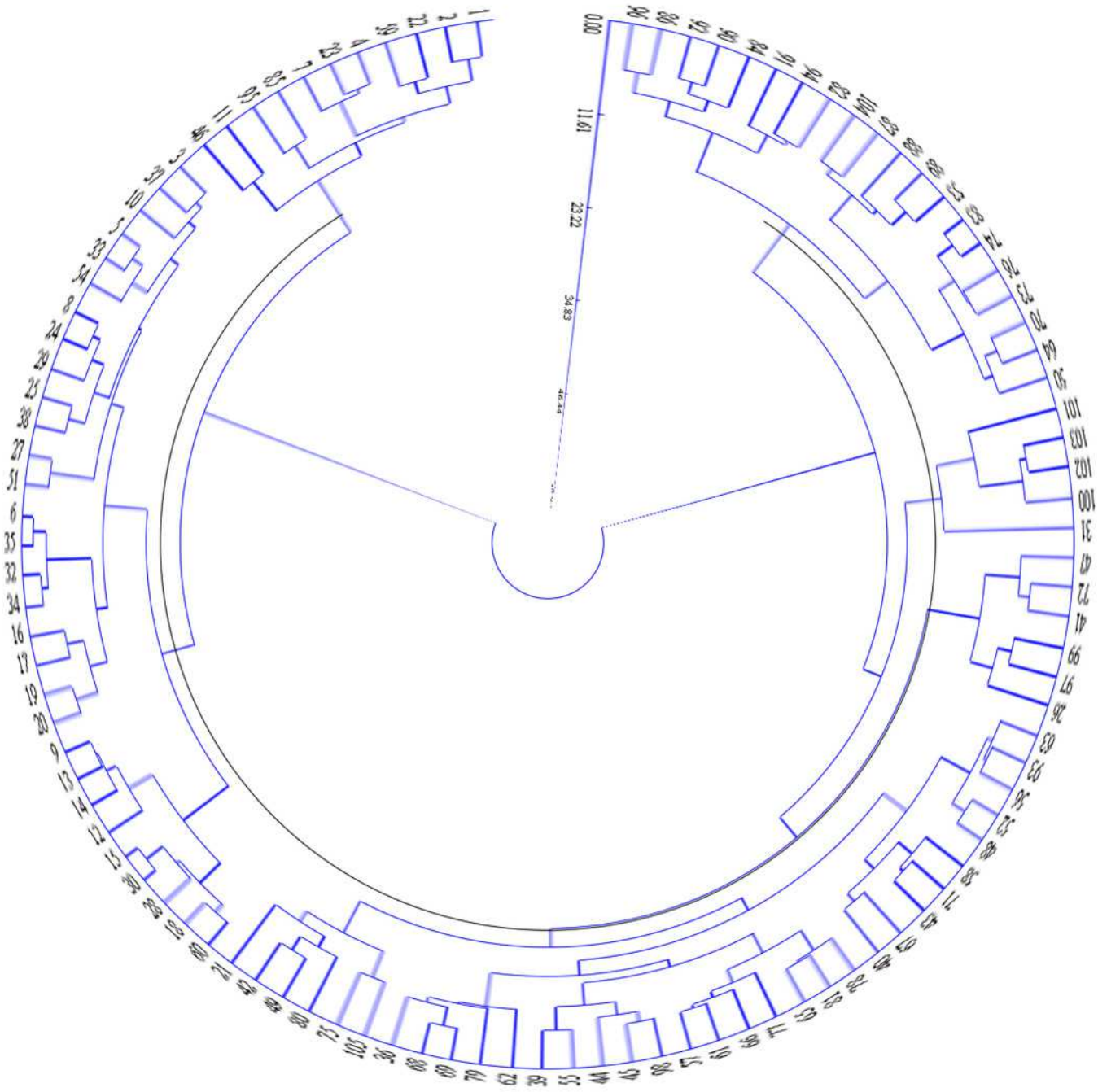

Table 1(on next page)

Tables

7 tables of manuscript

Table 1 Phenotypic diversity of 19 qualitative traits

Trait	Shannon diversity index DI	Distribution Frequency						
		1	2	3	4	5	7	9
Leaf sheath color LSC	0.5257	0.7810	-	0.2190	-	-	-	-
Leaf color of seedling LCS	0.2925	0.9143	0.0857	-	-	-	-	-
Bristle color BC	0.5324	0.0762	0.8476	0.0762	-	-	-	-
Bristle length BL	0.7438	0.0476	0.7524	0.1810	0.0190	-	-	-
Protecting glume color PGC	0.3673	0.0095	0.8952	0.0952	-	-	-	-
Stigma color STC	0.6483	0.2190	0.7524	0.0286	-	-	-	-
Anther color AC	0.6830	0.1524	0.0762	0.7714	-	-	-	-
Seed color SC	0.1075	-	0.9810	-	0.0095	0.0095		
Kernel color KC	0.0943	-	0.9810	-	0.0190	-	-	-
Seedling leaf attitude SLA	0.7825	0.0381	0.6286	0.3333	-	-	-	-
Blooming leaf attitude BLA	0.9655	0.6667	0.1143	0.0476	0.1714	-	-	-
Tiller habit TH	1.0566	0.3524	0.4381	0.2095	-	-	-	-
Branch habit BH	1.0543	0.3429	0.4476	0.2095	-	-	-	-
Peduncle shape PS	0.9367	-	0.2190	0.6095	0.1714	-	-	-
Ear compactness EC	0.5004	0.0286	0.8476	0.1238	0.0000	-	-	-
Spike density SD	0.5799	0.7333	-	0.2667	0.0000	-	-	-
Ear shape ES	1.1206	0.0381	0.4476	0.3810	0.1333	-	-	-
Shattering habit SH	0.7091	-	0.3714	0.6190	0.0095	-	-	-
Lodging resistance LR	1.0316	-	-	0.4762	-	0.4000	0.1048	0.0190
Mean	0.6701	0.3143	0.5554	0.2905	0.0593	0.2048	0.1048	0.0190

Note: Arabic numbers (1–9) in the first line of the table refer to the phenotypic trait according to Descriptors and Data Standard for Millet [*S. italica* (L.)] (Lu, 2006).

16

17

18

19 **Table 2 Phenotypic diversity of 12 quantitative traits**

Traits		Min V.	Max V.	Mean V.	SD	CV %	Shannon diversity index DI
Growing period GP d		101.0	131.0	118.809 5	7.535 9	6.34	1.696 6
Stem	Stem number per plant SNP	1.0	3.0	1.771 4	0.775 3	43.77	1.056 6
	Main Stem Length MSL cm	90.0	170.0	133.847 6	19.025 2	14.21	2.042 8
	Main stem diameter MSD mm	15.0	41.0	25.895 2	4.235 7	16.36	1.850 7
	Main stem nod number MSN	0.3	1.0	0.522 4	0.150 6	28.83	1.517 6
Ear	Peduncle length PL cm	9.0	16.0	13.115 4	1.430 0	10.90	1.707 4
	Main ear length MEL mm	16.0	40.0	22.847 6	4.018 7	17.59	1.874 1
	Main ear diameter MED mm	0.7	4.2	2.635 2	0.503 6	19.11	1.764 9
Yield	Grass weight per plant GWP g	9.2	119.8	47.040 0	21.563 6	45.84	1.951 3
	1000-seed weight TSW g	2.0	3.9	2.688 6	0.350 1	13.02	1.578 4
	Ear weight per plant EWP g	11.5	41.1	22.511 4	5.948 4	26.42	1.996 7
	Seed weight per plant SWP g	9.3	27.3	15.514 3	4.104 3	26.45	1.957 0

20

21

22

23

24

25

26

27 **Table 3 Correlation analysis of 12 quantitative traits**

Correlation Coefficient	SNP	MSL	PL	MSD	MSN	MEL	MED	GP	GWP	TSW	EWP
SNP	1										
MSL cm	0.023 7	1									
PL cm	0.083 4	0.428 4**	1								
MSD cm	0.196 5	-0.150 6	0.082 8	1							
MSN	-0.169 3	0.088 1	0.017 7	-0.075 9	1						
MEL mm	0.223 3*	0.325 5**	0.194 5*	0.246 3*	-0.215 1	1					
MED mm	0.262 2**	-0.008 1	-0.007 3	0.180 9	-0.093 5	0.005 1	1				
GP d	0.410 5**	0.302 8**	0.199 7*	0.097 8	-0.095 6	0.400 0**	0.192 0*	1			
GWP g	-0.010 7	0.081 1	0.075 9	-0.194 5*	0.287 2**	0.042 7	-0.105 8	0.055 8	1		
TSW g	-0.190 4	-0.044 6	-0.027 4	-0.017 0	0.052 5	-0.028 6	-0.045 1	-0.024 2	-0.050 8	1	
EWP g	0.395 5**	0.292 7**	0.182 1	0.093 1	-0.284 6**	0.461 2**	0.279 9**	0.330 3**	-0.160 9	-0.095 5	1
SWP g	0.253 3**	0.214 7 *	0.177 7	0.003 5	-0.267 2**	0.376 2**	0.257 3**	0.193 8*	-0.092 1	-0.083 9	0.826 6**

28 Note: $r_{0.05}$=0.1918, $r_{0.01}$=0.2504

Table 4 Principal components analysis and eigen value of 12 quantitative traits

Traits		1st Principle Component	2nd Principle Component	3rd Principle Component	4th Principle Component	5th Principle Component	6th Principle Component	7th Principle Componen
GP d		-0.503 0	0.381 1	-0.015 9	0.649 1	-0.298 8	-0.070 6	0.270 2
Stem	SNP	-0.055 0	0.180 3	0.842 5	-0.021 7	0.266 3	0.131 7	0.046 3
	MSL cm	0.546 8	0.344 8	0.125 2	0.163 5	-0.177 7	-0.572 9	0.218 9
	MSD cm	0.318 1	0.628 0	-0.149 4	-0.000 8	0.489 3	-0.161 4	-0.106 8
	MSN	-0.377 2	0.409 7	-0.238 3	-0.102 6	0.455 7	-0.035 5	-0.365 3
Ear	SIL cm	-0.268 3	0.383 6	-0.063 3	0.261 7	-0.481 2	0.535 6	-0.319 1
	MEL mm	-0.628 1	-0.235 5	-0.116 6	0.191 7	0.453 1	0.170 3	0.465 7
	MED mm	-0.338 3	-0.332 4	0.092 2	-0.670 9	-0.316 8	-0.221 1	-0.107 7
Yield	GWP g	0.541 5	-0.340 4	0.082 8	0.575 9	-0.212 2	0.110 8	-0.154 0
	TSW g	0.286 7	0.246 8	-0.622 0	-0.388 7	-0.263 2	0.196 7	0.318 4
	EWP g	0.174 0	0.458 1	0.401 2	-0.490 6	-0.187 8	0.422 6	0.239 8
	SWP g	0.616 6	-0.252 7	-0.110 0	0.074 8	0.404 0	0.501 0	0.058 1
Eigen Value		3.114 9	1.663 4	1.209 1	1.114 5	0.970 8	0.919 9	0.803 0
Contribution Rate		25.957 3	13.861 3	10.076 0	9.287 4	8.090 0	7.666 2	6.691 5
Cumulative C.R.		25.957 3	39.818 6	49.894 6	59.182 0	67.272 0	74.938 2	81.629 7
Factor Weight		31.798 8	16.980 8	12.343 5	11.377 5	9.910 6	9.391 4	8.197 4

Table 5 Grey relational analysis of seed yield and 11 quantitative traits

Trait	Interacting coefficient	Correlation sequence	Trait	Interacting coefficient	Correlation sequence
EWP g	0.5987	1	PL cm	0.4168	7
SNP	0.4674	2	MSD cm	0.3915	8
MEL mm	0.4575	3	GWP g	0.3808	9
MED mm	0.4332	4	TSW g	0.3714	10
GP d	0.4169	5	MSN	0.3476	11

MSL cm	0.4168	6	

Table 6 System Cluster Analysis of 12 quantitative traits

Traits		1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Group
	GP d	117.0000	113.4333	121.4848	123.5000	118.4000	122.1500
Stem	SNP	1.1818	1.1333	1.9697	2.3333	2.4000	2.4000
	MSL cm	143.2727	129.2333	136.3333	143.1667	99.8000	137.2000
	MSD cm	0.4909	0.4967	0.4955	0.7500	0.8400	0.4750
	MSN	13.7273	13.7333	12.9091	13.0000	11.6000	12.5500
Ear	PL cm	26.9091	25.0000	25.5758	31.5000	26.0000	25.5000
	MEL mm	21.6364	20.6333	23.8485	33.3333	20.0000	22.7500
	MED mm	49.9636	51.8867	47.6697	58.5167	25.9800	38.9450
Yield	GWP g	49.9636	51.8867	47.6697	58.5167	25.9800	38.9450
	TSW g	2.5455	2.8100	2.7545	2.5833	2.6200	2.5250
	EWP g	18.5909	17.7100	25.3758	27.3667	19.7200	26.3850
	SWP g	14.3455	12.9267	17.1697	17.9333	13.9000	16.9850

Table 7 K-mean cluster analysis of 12 quantitative traits

Traits		Min V.	Freqency	Center V.	Freqency	Max V.	Freqency
	GP d	102.21	0.1333	120.08	0.7143	127.38	0.1524
Stem	SNP	1.00	0.3524	2.00	0.4381	3.00	0.2095
	MSL cm	106.42	0.2286	131.20	0.3810	152.49	0.3905
	MSD cm	0.45	0.7333	0.66	0.2095	0.97	0.0571
	MSN	10.63	0.1524	13.21	0.6952	15.13	0.1524
Ear	PL cm	19.63	0.1524	26.01	0.7429	34.18	0.1048
	MEL cm	19.92	0.4952	24.62	0.4476	34.33	0.0571
	MED cm	0.70	0.0095	2.44	0.7619	3.36	0.2286
Yield	GWP g	29.71	0.5619	50.79	0.3238	82.23	0.2000
	TSW g	2.43	0.5333	2.93	0.4286	3.63	0.0381
	EWP g	16.80	0.4286	24.75	0.4000	31.56	0.1714
	SWP g	11.75	0.4381	16.96	0.4095	22.43	0.1524

66