Canadian Journal of Plant Science

Diversity and relationships of *Miscanthus sinensis* from 20 geographical distributions in China based on SSR markers

Journal:	<i>Canadian Journal of Plant Science</i>
Manuscript ID	CJPS-2021-0024.R1
Manuscript Type:	Article
Date Submitted by the Author:	24-Mar-2021
Complete List of Authors:	Li, Qiang; Northeast Forestry University, Ji, Wenyue; Northeast Forestry University Zhang, Jing; Luohe Municipal Engineering Design Institute Peng, Junhua; China National Seed Group CO LTD He, Miao; Northeast Forestry University, Department of Landscape Architecture
Is the invited manuscript for consideration in a Special Issue?_:	Not applicable (regular submission)
Keywords:	<i>Miscanthus sinensis</i> , Genetic diversity, eSSR, gSSR

SCHOLARONE™
Manuscripts

Diversity and relationships of *Miscanthus sinensis* from 20 geographical distributions in China based on SSR markers

Li Qiang^{1,*}, Wenyue Ji^{1,*}, Jing Zhang², Junhua Peng³ & Miao He¹

1 Department of Landscape Architecture of Northeast Forestry University, 26 Hexing Road, Harbin, Heilongjiang 150040, China

2 Luohe Municipal Engineering Design Institute, 162 Jiaotong Road, Yuanhui District, Luohe, Henan, 462099, China

3 China National Seed Group CO., LTD., A2 Fuxingmenwai Street, Xicheng District, Beijing, 100045, China

Correspondence: Miao He, Department of Landscape Architecture of Northeast Forestry University, 26 Hexing Road, Harbin, Heilongjiang 150040, China .Tel: (86) 136-1364-9547;(86) 188-4610-1120. E-mail: hemiao_xu@126.com

Both authors contributed equally to this work and should be considered co-first authors.

Funded by Special Fund Project of Basic Scientific Research Operating Expenses of Central Universities (257214BA21)

Diversity and relationships of *Miscanthus sinensis* from 20 geographical distributions in China based on SSR markers

Miscanthus is a C₄ herbaceous perennial species and it was chosen as a bioenergy crop due to high biomass yield. *Miscanthus sinensis* has many phenotypes which are adapted to various environments in China. In this study, 421 accessions of *M. sinensis* were collected from 22 provinces, and the genetic polymorphisms amongst these germplasm collections were identified using 20 primer pairs designed against 10 each from EST-SSR(eSSR) and genomic SSR(gSSR) transferable markers from barley. A total of 95 SSR Polymorphic bands were detected producing a 100% polymorphic rate among these *M. sinensis* accessions. The gSSR markers showed a richer genetic polymorphism than eSSR markers. Based on the UPGMA clustering, there was a distinct sub-population separation in *M. sinensis*, which indicates that geographical differences and natural selection are the driving forces for genetic variation and evolution in the species. The 20 pairs of barley markers matched to 26 Polymorphic bands associated with date of heading, plant height, leaf weight, stem weight, leaf/stem ratio, and total biomass yield. Eleven marker Polymorphic bands were associated with the date of heading, 4 with plant height, 10 with leaf weight, 7 with stem weight, 3 with leaf/stem ratio, and 10 with biomass yield.

Keywords: *Miscanthus sinensis*, Genetic diversity, eSSR, gSSR

Subject classification codes: Q948.15

1 Introduction

M. is herbaceous perennial species which is widely grown as ornamental plants. As early as in the 1930s, several *M.* varieties were selected for horticultural uses (Jorgensen, 2000). Up to today, there are 192 varieties in the registry of the Royal Horticultural Society (RHS). Among them, varieties 'Silberfeder', 'Zebrinus' and 'Variegatus' are widely used in landscaping.

M. sinensis is the most widely distributed species in the genus. The C₄ photosynthesis mechanism with a high carbon utilization efficiency makes these plants strong, tall and large-sized (Heaton, 2008). A few European countries have enlisted the species as the bioenergy crop because of the high biomass yield (Heaton, 2004; Heaton; 2008). These

plants are highly tolerant to drought, cold, salinity, disease and insects, and many biotic and abiotic stress factors (Deuter, 2000).

M. sinensis grows widely under tropical and subtropical climates, from extremely hot to cold regions such as the coldest Siberia at 50°N to tropical Polynesia at 22°S, and from Mid-India in the west to Polynesia in the east (Sang, T, 2011). China is the distribution center of *M. sinensis*. (Liu et al., 2001). Plants are found mainly in Anhui, Chongqing, Fujian, Gansu, Guangdong, Guangxi Zhuang Autonomous Region, Guizhou, Heilongjiang, Henan, Hebei, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Shanxi, Shandong, Sichuan, Taiwan, Yunnan, and Zhejiang. These 23 provinces include direct-controlled municipalities and autonomous regions. Under the influences of this broad range of weather conditions, accessions in *M. sinensis* have formed a large number of geographical populations (Liu et al., 2001). These plants have morphological properties that are adapted to the respective weather conditions.

M. sinensis has a highly complex genetic background due to the strong self-incompatibility, and therefore the germplasm contains rich phenotypic variations among populations within the species (Greef & Deuter, 1993). In *M. spp*, phenotypic polymorphisms in leaf, stem, and spikelets and florets have been identified among natural populations, and some of these parameters have extremely significant differences between and within wild accessions (Xiao et al., 2013). Furthermore, *M. sinensis* distributes widely in nature, along volcano belt, in agricultural fields, on river bank, or on mountain slopes from low to high elevations. Adaptation to these ecological environments has led to the formation of different ecotypes within the same species (Chou, 2009). *M. sinensis* plants, when grown under water deficit conditions, develop a flexible water storage strategy to maintain the biological functions of canopy (Clifton-Brown & Lewandowski, 2000). The highly effective water utilization efficiency

mechanism makes these plants extremely tolerant to the drought condition (Jørgensen & Schelde, 2001). In the cold region in Japan, a two-year out-door research experiment confirmed that *M.* can germinate and survive under cold condition (Inoue, 2003). Cold tolerance is a rare trait in C4 plants which was found only in a few C4 species (Sage & Kubien 2007). Thus studies on the *M. sinensis* not only will benefit exploration of this bioenergy plant species, they can also lead to the discoveries of elite genes conferring resistance to these adverse environmental conditions. Understanding the genetic structure of *M. sinensis* will help to establish the relationship between plants and environments, to explore phylogenetic relationships, and to answer questions as how plant adapts to adverse environments. This fundamental information is the bases for the exploration and utilization of elite genes from *M. sinensis*.

SSR (Simple Sequence Repeat) molecular marker technology has the advantages of good polymorphism, good repeatability, and high reliability (Wang et al., 2014). According to the different properties of the sequences on which SSR markers are developed, SSR markers can be divided into genomic SSR (gSSR) and expressed sequence tags SSR (eSSR) (Zhou et al., 2012). Developing SSR primers by library screening is expensive, and there is also little opportunity to develop *Miscanthus* SSR primers by searching public databases to obtain suitable sequences. Therefore, according to the characteristics of SSR flanking region sequence conservatism, it is a realistic choice to select SSR primers for the amplification of *Miscanthus* species from the existing related species (Lu et al., 2012). Barley is a relative species of *Miscanthus*, and the primers of barley have high versatility in *Miscanthus* (Zhou et al., 2012), so the primers of barley are used in this study.

2 Materials and methods

2.1 *Experimental materials and phenotypic characteristics*

In this study, 421 accessions of *M. sinensis* were collected from 20 provinces, and two barley samples were used as references to validate the experimental procedure. Barley samples were provided by Dr. Sun Dongfai in Huazhong Agricultural University, with code number of HV1 (hulless barley, China) and HV10 (HuaDa, China) which are common commercial varieties in China.

All the collected *M.* plants were grown in Wuhan Botanical Garden, Chinese Academy of Sciences (30°33'N, 114°24'E, altitude 19-35m).

Information for collecting the *M.* samples was described in Table 1. During growth season, plants were recorded for the following data: date of heading, plant height, leaf biomass, stems weight, leaf/stem ratio and biomass yield. Three measurements were taken and the average was used for each plant. Data of heading was the days from March 01 when tillers emerged, to the day when spikelets emerged.

2.2 *DNA extraction method*

Fresh leaves were harvested and immediately frozen in liquid nitrogen. DNA was extracted using a modified CTAB method (Doyle, 1987). The quality and integrity of total DNA were evaluated by using 1.0% agarose gel electrophoresis and confirmed by comparing band intensity with a λ DNA standard (Promega Corporation Madison, USA).

2.3 *SSR genotyping*

Primers were designed against the DNA marker sequences from barley. PCR reactions were conducted using these primers and genomic cDNAs from *M. sinensis* and two barley samples. The 10 primer pairs each from eSSR and gSSR were transferrable markers in barley confirmed by Chen Yinyin's group; and these markers were evenly distributed

on all the barley chromosomes. The primer sequence information (Table 2), and their distribution on barley chromosomes (Table 3) were provided. The barley eSSR marker sequences were provided by Heike Ballhausen at IPK (nonpublished data).

PCR amplification was carried out using a BIORAD-My Cycle Thermocycler. The PCR reaction was done in a 20 μ L volume containing 1 $\times$ PCR buffer, $MgCl_2$ 1.8mmol/L, forward and reverse primers each 500 nmol/L, dNTP 200 μ Mmol/L, genomic DNA 60 ng and Taq DNA polymerase (TOYOBO Blend Taq®) 1 unit. PCR program used was: after initial denaturation at 95°C for 3 min, there were 10 cycles of 95 °C denaturation for 30s, annealing at 50~60°C (Table 2) for 30s, and extension at 72°C for 30s, and followed by a final extension at 72 °C for 7 min. PCR products were separated by electrophoresis on a 6% native- polyacrylamide gel (PAGE) (2 L), and silver stained before imaging.

2.4 Data analysis

The association between parameters of population genetics and phylogenetic relationship were analyzed using excel and Popgene32 (Yeh and Boyle 2000). The number of alleles observed in a population (N_a), the effective allele number (N_e), Shannon information index (I), Nei's gene genetic heterogeneity (H_e) and polymorphism information content (PIC) were determined by using Popgene32. Dendrogram was constructed using the functions of cluster analysis in Popgene32, UPMGA (unweight pair method using arithmetic averages) was used to develop the phylogenetic tree among accessions. TASSEL2.1 was used in association analysis between phenotypic traits and marker Polymorphic bands.

3 Results

3.1 SSR polymorphism and genetic diversity

Every primer pair produced several bands, 95 polymorphic Polymorphic bands were

generated from the 20 primer pairs, and all of the amplified products were polymorphic (Table 4). These results indicate that *M. sinensis* accessions contain a large number of allelic variations. Two to eight Polymorphic bands were amplified from a single primer pair, and the average Polymorphic band per primer pair was 4.75. The electropherograms of two pairs of primers were shown in Fig.1 and Fig.2. The 10 eSSR markers produced 46 Polymorphic bands with an average of 4.6 Polymorphic bands, and the marker GBM1012 produced the largest number of Polymorphic bands of 8. The 10 gSSR markers produced 49 Polymorphic bands with an average of 4.9 Polymorphic bands per marker. The Bmag0115 produced the largest number of Polymorphic bands of 7. Using the 20 pairs of barley SSR primers, we were successful in PCR amplification of bands from genomic DNAs of *M. sinensis*, and all the PCR products were polymorphic. These results confirmed that these markers can be used in the analysis of genomic polymorphism of the species. Among the SSR primers, gSSRs markers identified more Polymorphic bands than eSSRs.

On each locus, the observed allele number (N_a) was between 2~3, and the mean Polymorphic bands per marker was 2.0947. For the gSSRs, the average observed allele number (N_a) was 2.1304, which is slightly bigger than the eSSRs of 2.0612. The average effective allele number for all Polymorphic bands (N_e) was 1.5096. The 10 pairs of eSSR primers produced 46 Polymorphic bands, at each locus the effective allele number (N_e) was 1.0047~2.7465 with an average value of 1.4966. For the gSSR primers derived 49 Polymorphic bands, the effective allele number N_e was 1.0047~2.7646 with average of 1.5234. The effective allele number is proportional to the actual allele number. A study on wheat revealed a similar yet slightly stronger correlation between the observed and actual allele number (Peng et al., 2009).

The Shannon information index (I) average value was 0.4287 when the 427 accessions

of *M. sinensis* were included. The index I of the 46 Polymorphic bands derived from eSSR primers was 0.0165 to 1.0482, with an average of 0.4183. Index I of the 49 Polymorphic bands derived from the gSSR primers was 0.0165 to 1.0522, with an average of 0.4399 which is higher than eSSR. The Shannon information index (I) indicated that *M. sinensis* contains a high level of genetic diversity.

The Nei's gene polymorphism index (H_e) is parameter measuring the diversity in population genetics (Nei, 2009). Similar to PIC, H_e can also be used in the analysis of SSR diversity. In this study, the average H_e was 0.2795. The eSSR Polymorphic bands index H_e was from 0.0047 to 0.6359 with an average of 0.2757. There were six Polymorphic bands with $H_e \geq 0.5$, they are e-1011-c, e-1021-a, e-1013-d, e-1040-d, e-1022-a and e-1022-c. The gSSR Polymorphic bands index H_e was from 0.0047 to 0.6383 with an average of 0.2836. The two Polymorphic bands with $H_e \geq 0.5$ were g-0579-c and g-0251-a.

The PIC value was from 0.0048 to 0.5611, with an average of 0.2268. The eSSR locus PIC was between 0.0048~0.5578, with an average of 0.2219. The gSSR locus PIC varied from 0.0048~0.5611 with an average of 0.2319.

The species PIC value is a measure of the capability of markers to differentiate populations based on the polymorphic level (Rong et al., 2018). The value of PIC is determined by the detected allele locus number and the frequency of these Polymorphic bands. When $PIC \geq 0.5$, Polymorphic bands are ranked as highly polymorphic; medium polymorphic Polymorphic bands are in the range of $0.25 \leq PIC < 0.5$, and $PIC < 0.25$ refers to low polymorphic Polymorphic bands.

3.2 Clustering analysis

All the data generated using eSSR and gSSR were used to calculate the Euclidean

distance for clustering analysis. The dendrogram using SSR markers were shown in Fig.3 Barley was placed in a separate group from *M.* of which all the accession were divided into five large groups (Fig. 3).

The first group is comprised accessions from Hubei, Shanxi, Gansu, Henan and Hebei, this group was divided into two smaller subgroups, one is Gansu, Shanxi and Hubei, the other subsection covers Henan and Hebei provinces on the plain region in China (Fig. 3). The first group spreads across the central China region where there are four distinct seasons and plentiful precipitation.

The second group contains accessions from Guizhou, Yunnan, Fujian, Guangxi, Guangdong, Hunan, Anhui, Jiangsu, and Hainan provinces. The second group was divided into the following five subgroups (Fig. 3): Guizhou and Yunnan; Jiangxi, Fujian and Guangdong; Guangxi and Hunan; Anhui and Jiangsu; and Hainan. The second group is the Southern Region of China. This area rich in heat resources, and has warm climate and plentiful precipitation, and constantly under strong influences of ocean.

The third group is comprised of Chongqing, Sichuan and Zhejiang (Fig 3), these two provinces are humid and hot rich in heat energy. They all belong to warm and humid climate.

Shandong in fourth group (Fig 3), this province is located at the north point of China where air temperature is lower than the rest of the regions. Due to the close proximity to the Bohai Sea, weather here is influenced by the ocean.

Liaoning in the fifth group (Fig 3), Shandong and Liaoning both are located along the Bohai Sea Coast, and the local weather is influenced by the ocean. Compared to Liaoning, Shandong has a lower elevation with warmer temperate monsoon climate.

3.3 Association analysis between phenotypic traits and SSR makers

The STRUCTURE function was used to classify the 421 accessions into groups based on mathematical models. The 95 SSR Polymorphic bands were used for the construction of population structure. It was assumed that among all the accessions there are K specimens that would fit into the Hardy-Weinberger balance subgroup. The population number K was set at 1-20, each K value was derived from 5 replicates; the Delta K was used to calculate and determine accurately the number of subpopulations. Fig.4 shows that at K=4, the Delta K value was the highest, which indicates that the whole sample sets should be divided into 4 subpopulations (Fig.5). Subpopulation I contains 128 accessions, which accounts for 30.0% of total sample number. Subpopulation II contains 123 samples, accounting for 28.8% of total sample number. Subpopulation III contains 42 samples, accounting for 9.8% of whole population. Subpopulation IV has 132 samples, accounting for 31.0% of total sample number.

The association between the SSR marker Polymorphic bands and phenotypic traits, such as heading date, plant height, leaf weight, stem weight, leaf/stem ratio and biomass are described in Table 5. In total, 26 Polymorphic bands were associated with phenotypic traits. Each of 12 Polymorphic bands including e-1042-c, g-0021-b, g-0115-f, g-0251-a, g-0115-a, g-021-f, g-0115-b, e-1013-d, g-0115-g, g-0788-c, e-1022-d, g-0023-d, was simultaneously associated with two or more traits.

The locus e-1042-c was associated with the Date of Heading and Leaf/stem ratio; locus g-0021-b associated with heading date, leaf weight and biomass weight; locus g-0115-f simultaneously associated with heading date and stem weight, locus g-0251-a associated with plant height, leaf weight and biomass yield; locus g-0115-a associated with leaf weight and biomass yield; locus g-021-f associated with leaf weight and biomass yield, locus g-0115-b associated with leaf weight, stem weight and biomass yield; locus e-1013-d associated with leaf weight, stem weight, leaf/stem ratio and biomass yield; locus g-

0115-g associated with leaf weight, stem weight and biomass yield; locus g-0788-c associated with leaf weight, stem weight and biomass yield, locus e-1022-d associated with leaf weight, stem weight and biomass yield, locus g-0023-d associated with stem weight and biomass yield.

There are 11 Polymorphic bands associated with heading date, which accounts for 42.31% of total association Polymorphic bands. These Polymorphic bands are g-0223-f, g-0579-c, e-1042-c, g-0021-b, g-0115-f, g-0807-a, g-0223-e, g-0579-d, e-1040-c, e-1040-d, e-1022-a, the explanation rate ranges in 0.1079-0.1590, and the locus giving the maximum explanation rate is e-1022-a.

There are four Polymorphic bands associated with stem height which accounted for 15.38% of total Polymorphic bands. These Polymorphic bands are e-1041-d, g-0251-a, g-0788-e, e-1041-a, the explanation rate is 0.0218-0.0424, and the maximum explanation rate locus is g-0251-a. There are ten Polymorphic bands associated with leaf weight, which accounts for 38.46% of total Polymorphic bands. These Polymorphic bands are g-0115-a, g-021-f, g-0115-b, e-1013-d, g-0251-a, g-0115-g, e-1041-b, g-0021-b, g-0788-c, and e-1022-d. The explanation rate is 0.0689-0.1069, and the locus giving the maximum explanation rate is e-1022-d.

There are seven Polymorphic bands associated with stem weight, accounting for 26.92% of total Polymorphic bands. These Polymorphic bands are g-0115-g, e-1013-d, g-0023-d, g-0115-b, g-0115-f, g-0788-c, e-1022-d, the explanation rate is 0.0408-0.0642, and the locus giving the maximum explanation rate is e-1013-d.

Three Polymorphic bands are associated with the leaf/stem ratio, which accounts for 11.54% of total association Polymorphic bands. These Polymorphic bands are e-1013-d, e-1042-c, g-0906-a, and the explanation rate is 0.0707-0.0909, the locus giving the maximum explanation rate is e-1013-d. Ten Polymorphic bands are associated with

biomass, which accounts for 38.46% of total associated Polymorphic bands. These Polymorphic bands are g-021-f, g-0115-b, e-1013-d, g-0115-g, g-0115-a, g-0023-d, g-0251-a, g-0021-b, g-0788-c, e-1022-d, and the explanation rate is 0.0620-0.0907; the locus giving the maximum explanation rate is e-1022-d.

4 Discussion

4.1 The use of barley SSR markers in genetic studies of *M. sinensis*

Several studies have shown that microsatellite markers from one species can be utilized in the studies of related species (Peakall et al., 1998). But in the *Poaceae*, the transferability of SSRs is very low. In a study by Zhong (2009) to test maize SSRs on *M. sinensis*, the transferable rate was 24.61%. Yu et al. (2013). utilized rice SSRs in an analysis of polymorphism of *M. sinensis* and they obtained 46.6% transferable rate. The current study used 20 barley SSR markers to determine the polymorphism of *M. sinensis*. Result showed that primers against all of these markers were able to amplify DNA bands in the genome of *M. sinensis*, and the transferable rate was 100% which is much higher than those from maize and rice. These results suggest a closer phylogenetic relationship between barley and *M. sinensis*, and therefore barley SSR markers can be used in the genetic classification of *M. sinensis*. In the dendrogram analysis, accessions from neighboring locations were placed in the same cluster, which indicates that these SSR markers effectively detected the genetic variation among different genotypes in *M. sinensis*. The transferable rate by using Barley transferable SSR markers in *M. sinensis* is much higher than other cereal species that have been studied. Therefore, in general, barley transferable SSR markers can be used as an effective tool in genetic studies of *M. spp.* for analyses genetic polymorphism, gene discovery and identification, and marker-assisted breeding.

4.2 Genetic diversity among *M. sinensis*

Genetic polymorphism is a measure of genetic variation within a population; it is used as the basic information for studying species diversification under a changing ecosystem. A thorough understanding of genetic complexity in a germplasm collection is very useful for plant breeding. In this study, barley SSR marker sequences were used to detect the polymorphism in *M. sinensis*, and every one of them revealed at least one polymorphic locus (100% polymorphic rate). These results confirmed the existence of a large number of allelic variation in *M. sinensis*. This conclusion is supported by genetic parameters including N_a of 2.0947, N_e of 1.5096, I at 0.4287 and H_e at 0.2795. These results are in agreement with the results obtained using transferable SSR from rice (Yu et al., 2013), maize (Zhong, 2009), and *Brachypodium distachyon* (Zhao et al., 2013).

The higher genetic diversity in a species, the stronger its adaptability to environments, and the more powerful its evolution potential (Reed & Frankham, 2003). We have shown in this study that *M. sinensis* contains rich genetic diversity, which is the molecular bases why this species has such a strong adaptability to various types of environments.

In this study, there were five highly polymorphic Polymorphic bands accounting for 5.265% of total locus number. The 48 medium polymorphic Polymorphic bands accounted form 50.53% of total polymorphism Polymorphic bands. There were 42 low polymorphic Polymorphic bands accounting for 44.21% of total Polymorphic bands. Results in Fig.1 and Fig.2 shows that SSRs markers amplified several smear bands. These smear bands were possibly derived from shifting of the SSR primers on template DNA during PCR reaction which were seen on the highly sensitive silver stained polyacrylamide gels. To eliminate the interference of these smears, only clear bands were counted during data analysis to avoid human introduced errors.

4.3 Differences in transferability between barley gSSR and eSSR markers

In this study, the gSSR marker primer pairs generated a higher polymorphism indices (Na, 2.1304; Ne, 1.5234; I, 0.4399; He, 0.2836; PIC, 0.2319) than eSSR markers (Na, 2.0612; Ne, 1.4966; I, 0.4183; He, 0.2757; PIC, 0.2219) (Table 4). These results indicate that gSSR markers produced a higher polymorphic level than the eSSR markers because the latter sets were derived from coding regions and therefore they are more conserved (Chabane et al., 2005). Cho also found that gSSRs markers were associated with higher polymorphism compared to eSSRs (83.8% vs 54.0%) (Cho et al., 2000). The eSSR-derived PCR products contained fewer heterogeneous bands which led the lower polymorphic level (Fig.1). The gSSR amplified bands had more heterogeneous bands, with lower degree of conservation (Fig.2). The eSSR markers are from the coding regions; they more conserved among the closely related species than gSSR. These eSSR markers are considered as more suitable portable /transferable marker. On the other hand, the conserved eSSRs sequences will produce lower levels of polymorphism compared to gSSR (Cho et al., 2000; Eujayl et al., 2000). These eSSRs have a higher level transferability across species than gSSRs, whereas gSSRs are more polymorphic. These two sets of markers both have valuable roles in germplasm evaluation and selection of new variety in *M. sinensis*.

4.4 Cluster analysis of *M. sinensis*

In this study, we used UPGMA to cluster all the accessions in *M. sinensis*. Accessions in *M. sinensis* were divided into 5 major clusters according to the geographical distance. Results (Fig.3) from the dendrogram clustering analysis indicate that accessions from nearby locations were always clustered together. These accessions are more likely to have closer phylogenetic relationship, therefore they were the first ones to be clustered into the same group. In other words, plants grown at nearby locations experience less ecological

separation, more active genetic exchange may have occurred resulting in the phylogenetic associations. These accessions thus showed smaller genetic distance and a higher level of genetic homogeneity. The larger genetic differences between clusters were caused by reproduction separation due to geographical barriers, and the geographical barriers also affected gene exchange.

The phylogenetic tree, while dividing the accessions according to the original locations, also showed big variation for samples from the South and North China. Such phenomenon indicates that plants have undergone genetic variations during the adaptation process to the environment at different geographical locations. The correlation between accession clusters and their origin locations indicates that there are inseparable ties between the geographical as well as genetic distribution of *M. sinensis* accessions, and these plants have been heavily influenced by the weather condition.

Previous studies have reported that genetic differentiation is somehow correlated to the germplasm origin locations (Xie et al., 2009; Su et al., 2010). Results from this study provided convincing evidence for the association between genetic diversity and geographical locales for *M. sinensis*. One reasonable explanation for this phenomenon is that *M. sinensis* is obligate out-crossers. For cross-pollinated plants, more frequent gene flow and exchange within the same species would occur when no high mountains or water streams as a separation geographical barrier exist (Xiao et al., 2013). Plants from neighbor areas would share some degrees with smaller genetic distance and remarkably a higher level of genetic identity.

Ethical Statement

Title of the article: Diversity and relationships of *Miscanthus sinensis* from 20 geographical distributions in China based on SSR markers.

I certify that this manuscript is original and has not been published and will not be submitted elsewhere for publication while being considered by Genetic Resources and Crop Evolution.

And the study is not split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time.

No data have been fabricated or manipulated (including images) to support our conclusions.

No data, text, or theories by others are presented as if they were our own.

The submission has been received explicitly from all co-authors. And authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results.

Conflict of Interest:

The authors declare that they have no conflict of interest.

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

Informed consent was obtained from all individual participants included in the study.”

References

- Chabane, K., Ablett, G. A., Cordeiro, G. M., Valkoun, J., & Henry, R. J. (2005). EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution*, 52(7), 903-909.
- Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCOUCH, S. R., & Cartinhour, S. (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 100(5), 713-722.
- Chou, C. H. (2009). *Miscanthus* plants used as an alternative biofuel material: the basic studies on ecology and molecular evolution. *Renewable Energy*, 34(8), 1908-1912.
- Clifton-Brown, J. C., & Lewandowski, I. (2000). Water use efficiency and biomass partitioning of three different *Miscanthus* genotypes with limited and unlimited water supply. *Annals of Botany*, 86(1), 191-200.
- Deuter, M. (2000). Breeding approaches to improvement of yield and quality in *Miscanthus* grown in Europe. *EMI Project*, Final report, 28-52.
- Doyle, J. J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem bull*, 19, 11-15.
- Eujayl, I., Sorrells, M. E., Baum, M., Wolters, P., & Powell, W. (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics*, 104(2-3), 399-407.
- Greef, J. M., & Deuter, M. (1993). Syntaxonomy of *Miscanthus* × *giganteus* GREEF et DEU. *Angewandte Botanik*, 67(3-4), 87-90.

- Heaton, E. A., Dohleman, F. G., & Long, S. P. (2008). Meeting US biofuel goals with less land: the potential of *Miscanthus*. *Global change biology*, 14(9), 2000-2014.
- Heaton, E. A., Long, S. P., Voigt, T. B., Jones, M. B., & Clifton-Brown, J. (2004). *Miscanthus* for renewable energy generation: European Union experience and projections for Illinois. *Mitigation and Adaptation Strategies for Global Change*, 9(4), 433-451.
- Inoue, T. (2003). Chronosequential change in a butterfly community after clear-cutting of deciduous forests in a cool temperate region of central Japan. *Entomological Science*, 6(3), 151-163.
- Jørgensen, U., & Schelde, K. (2001). Energy crop water and nutrient use efficiency. *Prepared for the International Energy Agency Bioenergy Task*, 17
- Liu, L., Zhu, M., Zhu, T.P. (2001). Exploitation and utilization of *Miscanthus* and *Triarrhena*. *Journal of Natural Resources*, 16(6), 562-563.
- Lu, Y. F., Jiang, J. X., Yi, Z. L., (2012). Study on the transferability of maize SSR and Sugarcane EST-SSR markers to *Miscanthus* (Poaceae). *Acta Prataculture Sinica*, 21(5), 86-96.
- Nei, M. (1972). Genetic distance between populations. *American naturalist*, 106(949), 283-292.
- Peakall, R., Gilmore, S., Keys, W., Morgante, M., & Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Molecular biology and evolution*, 15(10), 1275-1287.
- Peng, J. H., Bai, Y., Haley, S. D., & Lapitan, N. L. V. (2009). Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. *Genetica*, 135(1), 95.
- Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation biology*, 17(1), 230-237.
- Rong, H., Huang, B., Zhou, Q., (2018). The construction of fingerprints and genetic diversity analysis of 61 *Malus crabapple* cultivars based on SSR markers. *Journal of Nanjing Forestry University (Natural Sciences Edition)*, 42(3), 45-50.
- Sage, R. F., & Kubien, D. S. (2007). The temperature response of C3 and C4 photosynthesis. *Plant, Cell & Environment*, 30(9), 1086-1106.
- Sang, T. (2011). Toward the domestication of lignocellulosic energy crops: learning from food crop domestication free access. *Journal of Integrative Plant Biology*, 53(2), 96-104.

- Su, D., Zhou, Y.L., Yu, Q.L., Chang, X.Q.(2010). Genetic Diversity Analysis of *Medicago falcate* in North China Based on SSR Markers. *Chinese Journal of Grassland*, 32 (5): 85-89.
- Xiao, L., Jiang, J.X., Yi, Z.L., Ai, X., Deng, N.D. & Lin, C. (2013). Study on phenotypic diversity of *Miscanthus floridulus*. *Journal of Hunan Agricultural University (Natural Sciences)*, 39(2), 150-154.
- Xiao, L., Xue, D., Jiang, J.X., Ai, X., Qin, J.P., Chen, Z.Y., Liu, S.L., & Yi, Z.L.(2013). Genetic Diversity in *Miscanthus sinensis* of China Revealedby Simple Sequence Repeats DNA Markers, *Journal of Plant Genetic Resources*14(1): 36-41.
- Xie, W. G. ,Zhang, X.Q., &Ma, X. (2009). Genetic variation of *Dactylis glomerata* germplasm from Southwest China detected by SSR markers. *Acta Prataculturae Sinica*,138:146.
- Wang, S. X., Zhou, Y. L., Zhang, H. P., *et al.*(2014). Analysis of Genetic Diversity and Relationship among Wheat Breeding Parents by SSR Markers. *Journal of Triticeae Crops*, 34(5),624-627.
- Yeh, F. C., Yang, R., Boyle, T. J., Ye, Z., & Xiyan, J. M. (2000). PopGene32, Microsoft Windows-based freeware for population genetic analysis, version 1.32. *Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Alberta, Canada*.
- Yu, J., Zhao, H., Zhu, T., Chen, L., & Peng, J. (2013). Transferability of rice SSR markers to *Miscanthus sinensis*, a potential biofuel crop. *Euphytica*,191(3), 455-468.
- Zhao, H., Wang, B., He, J., Yang, J., Pan, L., Sun, D., & Peng, J. (2013). Genetic diversity and population structure of *Miscanthus sinensis* germplasm in China. *PLoS One*, 8(10), e75672.
- Zhong, Z.L. (2009). *Genetic diversity of Miscanthus species in South China revealed by SSR makers*, (Doctoral dissertation, Changsha: Hunan Agricultural University).
- Zhou,H.F. Li,S.S, Ge,S. (2011). Development of microsatellite markers for *Miscanthus sinensis* (Poaceae) and cross-amplification in other related species. *American Journal of Botany*, 98(7),195-197.

Figure 1 Segmental electrophoresis figure of primer GBM1015

Figure 2 Segmental electrophoresis figure of primer Bmag0223

Figure 3 Clustering tree graph

Figure 4 The values of Delta K are plotted against (K= 2-20 based on the structure analysis of 95 SSR Polymorphic bands in the 421 *M. sinensis* accessions)

Figure 5 Population structure of 421 *M. sinensis* accessions.
(The population is partitioned into four subpopulations with different colors)

Table1 Collecting district of materials

District	Number of samples	District	Number of samples
Hubei	45	Anhui	10
Chongqing	15	Jiangsu	6
Sichuan	40	Zhejiang	24
Shanxi	32	Fujian	22
Gansu	11	Guangdong	31
Henan	41	Hainan	7
Hebei	5	Guangxi	14
Guizhou	28	Hunan	25
Yunnan	17	Shandong	7
Jiangxi	19	Liaoning	12

Table2 Sequence of SSR primers

Primers	Sequence(F)	Sequence(F)	Annealing temperature(°C)
GBM1011	CAATCAGAAAGAAATCATCTCTC G	AGATCCATCACCTCCGTCAC	55
GBM1012	TTGCTTCCTGCTTCCATCTT	GCACCAGGTAGACGATGAGG	52
GBM1013	TCCCTGATCCATGTTTTTC	GATGCCAAGTTCTTCTTCGC	50
GBM1015	TTGTTGGAACATACAAACATGC	GTTCCGTGTGAATTAGCGGT	50
GBM1017	AGCTTCCTCGCTACAAACCA	GACCCTCCTCCTTCAACCTG	55
GBM1021	GGTTCTGATCAAGGAACCTTTT	TCAAATATACGCCCTCCGTC	60
GBM1022	CAAGCAGATGCAACTACACCA	TCTGCACAGCTGTTTGGTTC	55
GBM1040	AACAGAGTGTGGCCATGC	TTCAGTATCGACCACTTGC	50
GBM1041	ACTAGGTGTGGGTGCTTTG	CCGATCATAGGTGGAAGTGA	53
GBM1042	CACACAAGATCTCTCACGGC	TCCCTGACCAAACCAAGTTC	60
GMS021	CTATCACACGACGCAACATG	CCTGAGAAAGAAAGCGCAAC	56.5
Bmag0021	ATTTTATCAGAACGTCTCTCTC	CTAACTTCTCTCTCCCTCTCC	57.5
Bmag0023	AACACAGACCTACGGGTC	CATGAGATAGATCCAAGCAC	56.5
Bmag0115	TGATCTGTGAGTAGTTCACCA	AATTGAAGATGCGATGATC	50
Bmag0223	TTAGTCACCCTCAACGGT	CCCCTAACTGCTGTGATG	50
Bmac0251	CAGATCTCAACAAACACACC	ATTCATATAAACATATGGGT CC	50
Bmag0579	CCTAGATAAGGAACATAGCCA	CAAAGACCCTAACTCATGTTC	50
EBmac0788	TAACTTACTTTATATCCATGGCA	ATGATGAGAACTCTTCACCC	56.5
Bmag0807	GGATATAAGGGTCCATAGCA	AATTACATCAAATAGGCTCCA	53
EBmac0906	CAAATCAATCAAGAGGCC	TTTGAAGTGAGACATTCCA	50

Table3 Distribution of primers in chromosomes of barley

	eSSR		gSSR	
	LongArm	ShortArm	LongArm	ShortArm
Chromosome 1	GBM1013	GBM1042	Bmag0579	GMS021
Chromosome 1	GBM1012	GBM1011	Bmag0115	
Chromosome 1		GBM1040		Bmag0023
Chromosome 1	GBM1015		EBmac0788	EBmac0906
Chromosome 1	GBM1041		Bmag0223	
Chromosome 1	GBM1022	GBM1021	Bmac0251	Bmag0807
Chromosome 1		GBM1017		Bmag0021

Table 4 Genetic parameters of eSSR and gSSR

	Polymorphic bands	Na ^a	Ne ^b	I ^c	He ^d	PIC ^e
eSSR	e-1042-a	2.0000	1.4778	0.5043	0.3233	0.2680
	e-1042-b	2.0000	1.7584	0.6228	0.4313	0.3401
	e-1042-c	2.0000	1.9604	0.6830	0.4899	0.3704
	e-1015-a	2.0000	1.0094	0.0297	0.0093	0.0093
	e-1015-b	2.0000	1.0094	0.0297	0.0093	0.0093
	e-1015-c	2.0000	1.1285	0.2286	0.1139	0.1050
	e-1015-d	2.0000	1.1337	0.2350	0.1180	0.1088
	e-1011-a	2.0000	1.0094	0.0297	0.0093	0.0093
	e-1011-b	2.0000	1.0094	0.0297	0.0093	0.0093
	e-1011-c	3.0000	2.0462	0.07484	0.5113	0.3921
	e-1021-a	3.0000	2.1318	0.8963	0.5309	0.4663

e-1021-b	2.0000	1.0047	0.0165	0.0047	0.0048
e-1017-a	2.0000	1.1548	0.2594	0.1341	0.1231
e-1017-b	2.0000	1.0094	0.0297	0.0093	0.0093
e-1017-c	2.0000	1.5865	0.5565	0.3697	0.3021
e-1017-d	2.0000	1.5685	0.5485	0.3625	0.2942
e-1017-e	2.0000	1.5625	0.5458	0.3600	0.2942
e-1041-a	2.0000	1.5263	0.5287	0.3448	0.2859
e-1041-b	2.0000	1.4235	0.4741	0.2975	0.2534
e-1041-c	2.0000	1.6400	0.5790	0.3902	0.3149
e-1041-d	2.0000	1.9312	0.6752	0.4822	0.3652
e-1041-e	2.0000	1.9578	0.6823	0.4892	0.3686
e-1041-f	2.0000	1.8829	0.6617	0.4689	0.3585
e-1041-g	2.0000	1.6623	0.5882	0.3988	0.3149
e-1012-a	2.0000	1.5021	0.5168	0.3343	0.2807
e-1012-b	2.0000	1.0094	0.0297	0.0093	0.0093
e-1012-c	2.0000	1.1926	0.2992	0.1615	0.1469
e-1012-d	2.0000	1.1655	0.2712	0.1420	0.1266
e-1012-e	2.0000	1.3168	0.4047	0.2406	0.2113
e-1012-f	2.0000	1.1655	0.2712	0.1420	0.1301
e-1012-g	2.0000	1.6223	0.5717	0.3836	0.3122
e-1012-h	2.0000	1.5323	0.5316	0.3474	0.2894
e-1013-a	2.0000	1.0094	0.0297	0.0093	0.0093
e-1013-b	2.0000	1.0094	0.0297	0.0093	0.0093
e-1013-c	2.0000	1.9982	0.6927	0.4995	0.3748
e-1013-d	3.0000	2.0677	0.7949	0.5164	0.4081
e-1040-a	2.0000	1.0094	0.0297	0.0093	0.0093
e-1040-b	2.0000	1.0094	0.0297	0.0093	0.0093
e-1040-c	2.0000	1.5685	0.5485	0.3625	0.2975
e-1040-d	3.0000	2.5955	1.0098	0.6147	0.5349
e-1022-a	3.0000	2.7465	1.0482	0.6359	0.5578
e-1022-b	2.0000	1.0094	0.0297	0.0093	0.0093
e-1022-c	3.0000	2.5465	1.0019	0.6073	0.5294
e-1022-d	2.0000	1.9027	0.6674	0.4744	0.3615
e-1022-e	2.0000	1.7203	0.6095	0.4187	0.3308
e-1022-f	2.0000	1.7898	0.6332	0.4413	0.3421
means	2.0612	1.4966	0.4183	0.2757	0.2219
gSSR	g-0021-a	2.0000	1.0094	0.0297	0.0093
	g-0021-b	2.0000	1.6749	0.5927	0.4030
	g-0021-c	2.0000	1.9991	0.6929	0.4998
	g-0115-d	2.0000	1.8787	0.6605	0.4677
	g-0115-e	2.0000	1.7147	0.6075	0.4168
	g-0115-f	2.0000	1.6045	0.5642	0.3767
	g-0115-g	2.0000	1.4657	0.4978	0.3178
	g-0115-a	2.0000	1.2256	0.3307	0.1841
	g-0115-b	2.0000	1.2090	0.3152	0.1729
	g-0115-c	2.0000	1.0094	0.0297	0.0093
	g-0906-a	3.0000	1.5392	0.6540	0.3503
	g-0906-b	2.0000	1.0047	0.0165	0.0047
	g-0906-c	2.0000	1.0047	0.0165	0.0047
	g-0906-d	2.0000	1.9782	0.6876	0.4945
	g-0023-e	2.0000	1.9896	0.6905	0.4974
	g-0023-f	2.0000	1.9801	0.6881	0.4950
	g-0023-a	2.0000	1.0094	0.0297	0.0093
	g-0023-b	2.0000	1.5625	0.5458	0.3600
	g-0023-c	2.0000	1.0094	0.0297	0.0093
	g-0023-d	2.0000	1.1655	0.2712	0.1420
	g-0807-a	2.0000	1.8745	0.6593	0.4665

g-0807-b	2.0000	1.0094	0.0297	0.0093	0.0093
g-0807-c	2.0000	1.7147	0.6075	0.4168	0.3285
g-0807-d	2.0000	1.3520	0.4293	0.2604	0.2263
g-0223-a	2.0000	1.9437	0.6786	0.4855	0.3674
g-0223-b	2.0000	1.9982	0.6927	0.4995	0.3749
g-0223-c	2.0000	1.0047	0.0165	0.0047	0.0048
g-0223-d	2.0000	1.0047	0.0165	0.0047	0.0048
g-0223-e	2.0000	1.3936	0.4561	0.2824	0.2426
g-0223-f	2.0000	1.8989	0.6663	0.4734	0.3603
g-0223-g	2.0000	1.7898	0.6332	0.4413	0.3439
g-0788-a	2.0000	1.3344	0.4172	0.2506	0.2164
g-0788-b	2.0000	1.0047	0.0165	0.0047	0.0048
g-0788-c	2.0000	1.5081	0.5198	0.3369	0.2789
g-0788-d	2.0000	1.0047	0.0165	0.0047	0.0048
g-0788-e	2.0000	1.3285	0.4131	0.2473	0.2138
g-0788-f	2.0000	1.3936	0.4561	0.2824	0.2448
g-0579-a	2.0000	1.9604	0.6830	0.4899	0.3697
g-0579-b	2.0000	1.0094	0.0297	0.0093	0.0093
g-0579-c	3.0000	2.7646	1.0522	0.6383	0.5611
g-0579-d	2.0000	1.8388	0.6487	0.4562	0.3515
g-0251-a	3.0000	2.4337	0.9585	0.5891	0.5022
g-0251-b	2.0000	1.5444	0.5374	0.3525	0.2910
g-021-a	2.0000	1.8481	0.6515	0.4589	0.3530
g-021-b	2.0000	1.0094	0.0297	0.0093	0.0093
g-021-c	2.0000	1.0094	0.0297	0.0093	0.0093
g-021-d	2.0000	1.0094	0.0297	0.0093	0.0093
g-021-e	2.0000	1.6691	0.5905	0.4009	0.3226
g-021-f	2.0000	1.6459	0.5813	0.3924	0.3136
Means	2.1304	1.5234	0.4399	0.2836	0.2319

^aNa: The observed number of polymorphic alleles.

^bNe: The effective number of alleles (Kimura and Crow 1964).

^cI: Shanno's information index (Lewontin 1972).

^dHe: Nei's genetic diversity (Nei 1973).

^ePIC = polymorphism information content.

Table 5 Association of SSR markers with *M. sinensis* traits

Items	locus	p-M	p-P	p-adj	R ²
Date of heading	g-0223-f	0.0015	0.002	0.011	0.1124
	g-0579-c	0.0031	0.005	0.034	0.1213
	e-1042-c	0.0042	0.007	0.049	0.1079
	g-0021-b	<0.0001	0.002	<0.0001	0.1166
	g-0115-f	<0.0001	<0.0001	<0.0001	0.1422
	g-0807-a	<0.0001	0.002	<0.0001	0.1208
	g-0223-e	<0.0001	<0.0001	<0.0001	0.1155
	g-0579-d	<0.0001	<0.0001	<0.0001	0.1148
	e-1040-c	<0.0001	<0.0001	<0.0001	0.1258
	e-1040-d	<0.0001	<0.0001	<0.0001	0.1315
	e-1022-a	<0.0001	<0.0001	<0.0001	0.159
Plant height	e-1041-d	0.0031	0.002	0.007	0.0298
	g-0251-a	0.0035	0.024	0.021	0.0424
	g-0788-e	0.01	0.02	0.036	0.0241

Leaf biomass	e-1041-a	0.0165	0.016	0.047	0.0218
	g-0115-a	0.0015	0.002	0.005	0.0758
	g-021-f	0.0016	0.002	0.006	0.0754
	g-0115-b	0.0013	0.004	0.008	0.0764
	e-1013-d	0.0012	0.01	0.014	0.0911
	g-0251-a	0.0025	0.007	0.02	0.0872
	g-0115-g	0.0031	0.006	0.028	0.0724
	e-1041-b	0.0062	0.005	0.045	0.0689
	g-0021-b	<0.0001	<0.0001	<0.0001	0.0812
	g-0788-c	<0.0001	0.003	<0.0001	0.0795
Stem biomass	e-1022-d	<0.0001	<0.0001	<0.0001	0.1069
	g-0115-g	0.0012	0.004	0.006	0.0493
	e-1013-d	0.0012	0.005	0.007	0.0642
	g-0023-d	0.0027	0.003	0.009	0.0451
	g-0115-b	0.0053	0.006	0.036	0.0418
	g-0115-f	0.0065	0.006	0.049	0.0408
	g-0788-c	<0.0001	<0.0001	<0.0001	0.0539
	e-1022-d	<0.0001	<0.0001	<0.0001	0.0562
Leaf/stem ratio	e-1013-d	0.0019	0.008	0.003	0.0909
	e-1042-c	0.0068	0.009	0.019	0.0707
	g-0906-a	0.0139	0.018	0.05	0.074
Biomass yield	g-021-f	0.0023	<0.0001	0.005	0.0638
	g-0115-b	0.0017	0.003	0.008	0.0653
	e-1013-d	0.001	0.004	0.009	0.0824
	g-0115-g	0.0017	0.005	0.01	0.0653
	g-0115-a	0.002	0.005	0.011	0.0647
	g-0023-d	0.0034	0.002	0.016	0.062
	g-0251-a	0.0036	0.006	0.04	0.0754
	g-0021-b	<0.0001	0.003	<0.0001	0.0687
	g-0788-c	<0.0001	0.002	<0.0001	0.0719
	e-1022-d	<0.0001	<0.0001	<0.0001	0.0907

The default value of 10000 permutations (p-value= 0.0001)

p-M: p value of F test for specific markers.

p-P: p value of permutation test for specific markers.

p-adj: P value of experiment-wise test of specific markers.

R² : The total variance explained by a marker using a normalized model.