Investigation of Genetic Relationships within *Miscanthus* using SNP Markers Identified using SLAF-Seq

ZHIYONG Chen (zhiyongchen@hunau.edu.cn)
Hunan Agricultural University https://orcid.org/0000-0001-5604-9514
Yancen He
Hunan Agricultural University
Yasir Iqbal
Hunan Agricultural University
Yanlan Shi
Hunan Agricultural University
Hongmei Huang
Hunan Agricultural University
Zili Yi
Hunan Agricultural University

Research article

Keywords: Miscanthus, SLAF-seq, SNP, SLAF tags, high-throughput sequencing, identification of genetic relationship.

Posted Date: January 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-152687/v1

Abstract

Background: *Miscanthus,* which is a leading dedicated-energy grass in Europe and in parts of Asia, is expected to play a key role in the development of the future bioeconomy. However, due to its complex genetic background, it is difficult to investigate phylogenetic relationships and the evolution of gene function in this genus. Here, we investigated 50 *Miscanthus* germplasms: 1 female parent (*M. lutarioriparius*), 30 candidate male parents (*M. lutarioriparius, M. sinensis,* and *M. sacchariflorus*), and 19 offspring. We used high-throughput Specific-Locus Amplified Fragment sequencing (SLAF-seq) to identify informative single nucleotide polymorphisms (SNPs) in all germplasms.

Results: We identified 800,081 SLAF tags, of which 160,368 were polymorphic. Each tag was 264–364 bp long. The obtained SNPs were used to investigate genetic relationships within *Miscanthus.* We constructed a phylogenetic tree of the 50 germplasms using the obtained SNPs, and found that the germplasms fell into two clades: one clade of *M. sinensis* only and one clade that included the offspring, *M. lutarioriparius,* and *M. sacchariflorus*. Genetic cluster analysis indicated that *M. lutarioriparius* germplasm C3 was the most likely male parent of the offspring.

Conclusions: As a high-throughput sequencing method, SLAF-seq can be used to identify informative SNPs in *Miscanthus* germplasms and to rapidly characterize genetic relationships within this genus. Our results will support the development of breeding programs utilizing *Miscanthus* cultivars with elite biomass- or fiber-production potential.

Background

Miscanthus, which is a high-biomass-yielding perennial C4 grass, has emerged as a candidate second-generation energy crop with great potential utility [1–4]. *Miscanthus* is a heterogeneous gramineous plant that hybridizes interspecifically, generating a complex genetic background. Despite increasing interest, molecular research into *Miscanthus* has been scant, due to its large genome (approximately 2.65 Gb) [5], complex genetic background, and limited available sequence data. Because inter- and intraspecific hybridization is common in *Miscanthus*, this genus is characterized by rich genetic diversity, and hybrid offspring often have high biomass [3]. However, this high degree of genetic diversity also increases the complexity of interspecific relationships in the *Miscanthus*, and, consequently, the difficulty of genetic evolution analyses in this genus. Therefore, it is difficult to mine functional genes in the *Miscanthus*, which seriously affects the utility of *Miscanthus* species for energy production and conversion [6]. Molecular markers would be useful for further investigations of *Miscanthus* plants; such markers have been widely used in studies of genetics, molecular population genetics, species formation, evolutionary and phylogenetic relationships, and molecular taxonomy [7].

First generation molecular markers include restriction fragment length polymorphisms (RFLPs) [8, 9], random amplified polymorphic DNA (RAPD) [10, 11], and amplified fragment length polymorphisms (AFLPs) [12], while second generation molecular markers include simple sequence repeats (SSRs) [13]

and inter-simple sequence repeats (ISSRs) [14]. However, these markers have several limitations: they are low throughput, inaccurate, time-consuming, labor-intensive, and costly [3]. These drawbacks have motivated the development of third-generation molecular markers. Third-generation molecular markers are SNPs. These markers are polymorphic and a generally widely distributed throughout the whole genome [15]. SNP markers are amenable to large-scale automated monitoring and have been instrumental in various techniques associated with crop breeding, such as the construction of genetic maps, the DNA fingerprinting of germplasm resources, the detection of molecular biodiversity, and the analysis of linkage disequilibrium [16]. This continuous development of molecular marker technology has accelerated functional gene identification and characterization in other crops, and has led to the development of varieties with improved functional traits [6]. Thus, these techniques might be useful for molecular genetic research in *Miscanthus*.

Although genotyping-by-sequencing (GBS) and restriction site-associated DNA sequencing (RAD-seq) have been used extensively in *Miscanthus*, there are still some difficulties and challenges associated with the application of these techniques, especially to non-model taxa like *Miscanthus* [17]. One obstacle to the widespread use of GBS is the difficulty of the associated bioinformatics analysis, which is typically hampered by a large number of erroneous SNP interferences that are not easy to diagnose or correct [18].

To overcome these challenges, we aimed to develop and identify SNP markers for *Miscanthus* using SLAF-seq techniques. That is, we aimed to reduce the genomic complexity using specific digestion, develop markers via the high-throughput sequencing of representative libraries, and determine phylogenetic relationships using genotyping.

SLAF-seq uses bioinformatics methods to systematically analyze known genome sequences, genome sequences of related species, bacterial artificial chromosome (BAC) sequences, or Fosmid sequences [19–23]. SLAF-seq techniques differ in several ways from GBS or RAD-seq techniques. First, there are many more SLAF tags, with SLAF-seq identifying one tag about every 10 K; second, SLAF tags are uniformly distributed, so important chromosome segments are not missed; and third, SLAF-seq methods avoid repetitive sequences, thus improving sequencing cost-effectiveness. As such, SLAF-seq utilizes deep sequencing to ensure genotyping accuracy; a reduced representation strategy to reduce sequencing costs; a pre-designed representation scheme to optimize marker efficiency; and a double-barcode system for large populations [24].

SLAF-seq has been widely used for the development of specific molecular markers and genetic maps [25]. For example, Sun et al. used 50,530 SLAFs with 13,291 SNPs to genotype the F1 population of the common carp [24]. Due to its efficient identification of SNP markers, SLAF-seq has been used in a wide variety of crops [26–30]. SLAF-seq was also used to develop the first high-density genetic maps for several economically important species, including sesame [31], cucumber [23], the brown alga *Undaria pinnatifida* (Phaeophyceae) [32], wax gourd [33], watermelon (*Citrullus lanatus* L.) [34], and *Salvia miltiorrhiza* [25]. In addition, an increasing number of studies of the Gramineae have been performed using SLAF-seq [21, 22, 35, 36]. For example, SLAF-seq was used to develop the first 7E-chromosome-

specific molecular markers for *Thinopyrum elongatum* [35], while 5,142 polymorphic SLAFs were analyzed to identify a new maize inflorescence meristem mutant [36]. Zhang et al. used 69,325 high-quality SLAFs, of which 26,248 were polymorphic, to develop sufficient markers for a segregating *Agropyron* F1 population [28]. Furthermore, a high-density genetic linkage map for orchard grass was developed using 2,467 SLAF markers and 43 SSR markers [37], and the semi-dwarf gene in barley was fine-mapped using molecular markers developed with SLAF-seq [21]. The successful application of SLAF-seq in other species provides reference materials for the development of SNPs in this study.

However, SLAF-seq has yet to be used to develop SNP markers in *Miscanthus*. Therefore, we aimed to use SLAF-seq to determine the genetic relationships among *Miscanthus* species, as well as to develop SNP markers. These results provide useful data for molecular marker-assisted *Miscanthus* breeding programs.

Results

Evaluation of the digestive enzymes

Based on the *S. bicolor* reference genome, we predicted that *Eco*RV + *Sca*I was a suitable restriction enzyme combination for the digestion of the *Miscanthus* genome. The pair-end digestion efficiency of *Eco*RV + *Sca*I for the control genome (Nipponbare) was 90.87%, and 97.80% of control reads lacked a complete restriction endonuclease recognition sequence, indicating that this enzyme combination was suitable. After *Eco*RV + *Sca*I digestion, we predicted 52,287 SLAF tags in the sorghum genome.

Analysis of the SLAF-seq data

We generated 57.81 M clean sequence reads after SLAF library construction and high-throughput sequencing (Supplement Table 1). The average GC content across all sequences was 41.39%. Across all sequences, the average number of bases with a quality score ≥ 30 (Q30) was 93.66%, indicating that 93.66% of all bases had a 0.1% chance of error (i.e., 99.9% confidence).

We developed 469,509 SNPs and 800,081 SLAF tags; 160,368 of the SLAF tags were polymorphic (Supplement Table 2). The average sequencing depth per sample was 11.76x for the female parent (sample A12), 15.47x for the male parents (samples A1–A11, B1–B10, and C1–C9), and 7.85x for the progeny (samples D1–D19). Overall, we obtained 469,509 SNPs (Supplement Table 3).

Genetic distance analysis

In total, 11,244 highly consistent population SNPs with integrity > 0.8 and minor allele frequency (MAF) > 0.05 were identified across all samples. The clustering patterns of the 50 *Miscanthus* germplasms were optimal, and cross-validation error was lowest when the number of clusters (ΔK) was four: *M. sinensis*

(A1–A11) fell into group 1; some offspring (D8, D9, D11–D13, D15, and D17–D19) and C3 (*M. lutarioriparius*) fell into group 2; *M. lutarioriparius* (C1, C2, and C4–C9) and *M. sacchariflorus* (B1–B10) fell into group 3; and the remaining offspring (D1–D7, D10, D14, and D16) and the female parent A12 (*M. lutarioriparius*) fell into group 4 (Table 1).

Table 1
The clusters associated with the 50 samples.

Sample ID	Group number	Sample ID	Group number
A2	1	C5	3
A3	1	C6	3
A4	1	C7	3
A5	1	C8	3
A6	1	C9	3
A7	1	D1	4
A8	1	D2	4
A9	1	D3	4
A10	1	D4	4
A11	1	D5	4
A12	4	D6	4
B1	3	D7	4
B3	3	D9	2
B4	3	D10	4
B5	3	D11	2
B6	3	D12	2
B7	3	D13	2
B8	3	D14	4
B9	3	D15	2
B10	3	D16	4
C1	3	D17	2
C2	3	D18	2
C3	2	D19	2

Sample ID: project sample number; Group Number: the group into which the sample was clustered.

PCA analysis indicated that the *M. sinensis* germplasms were distinct from all other groups, while the offspring, the female parent (*M. lutarioriparius)*, and sample C3 (*M. lutarioriparius*) formed a loose cluster (Fig. 1). The remaining *M. lutarioriparius* and most of the *M. sacchariflorus* formed a cluster, as did two *M. sacchariflorus* (Fig. 1).

The phylogenetic analysis indicated that the samples fell into two large clades: one that included all samples for the candidate male parent *M. sinensis*; and one that included all other samples (female *M. lutarioriparius*, male *M. sacchariflorus*, male *M. lutarioriparius*, and all hybrids). In the second large clade, the hybrids were divided into two groups: one included D1–D7, D10, D14, and D16, while the other included D8, D9, D11–D13, D15, and D17–D19 (Fig. 2). Female parent A12 fell into the former group, while candidate male parent C3 fell into the latter. These results were consistent with our clustering analysis (Table 1), suggesting that the recovered genetic distances accurately reflected the genetic relationships among the four clusters.

Our combined phylogenetic analysis also indicated that C3 was the most likely male parent of the crossbred offspring, because this sample had the fewest abnormal offspring (Table 2).

Table 2
Phylogenetic relationships between candidate male parents and offspring.

Candidate male parent	Number of abnormal offspring	Candidate male parent	Number of abnormal offspring
B1	9	C1	11
B2	7	C2	6
B3	9	C3	0
B4	9	C4	11
B5	11	C5	10
B6	9	C6	9
B7	10	C7	11
B8	6	C8	9
B9	9	C9	14
B10	6		

Discussion

Analysis of the SLAF-seq data

Genetic improvements in biomass quality may be substantially accelerated by the development of genetic markers associated with quality traits [38]. SNPs, which are more abundant in the genome than any other molecular markers, are particularly useful for analyses of genetic diversity and population structure [39]. In this study, we used SLAF-seq to efficiently identify SNP markers. Compared with other methods, such as GBS and RAD-seq, SLAF-seq is more accurate, faster, and less expensive; SLAF-seq also reduces genome complexity [40]. Here, we obtained 800,081 SLAF tags and 469,509 SNPs. This is

greater than the number of SNPs previously obtained in the *Miscanthus* genome using RAD-seq [41]. These polymorphic molecular markers are highly discriminatory, and can be used for genetic map construction and gene mapping in *Miscanthus*.

Genetic Relationships Analysis

The heterozygosity and polyploidy that have accumulated in *Miscanthus* genomes over their long evolutionary history challenge efforts to sequence complete genomes in this genus. Therefore, it is difficult to determine genetic relationships within *Miscanthus.*

Some studies have investigated phylogenetic relationships and species evolution using SNP markers [42]. These previous studies uncovered multiple instances of domestication behaviors or gene transfer between cultivated species and wild species [43]. To obtain comprehensive information about genetic variation, SNP markers must be developed for the whole genome. These developed SNP markers may clarify unresolved phylogenetic relationships, especially among closely related species [44]. Specifically, the SNP markers obtained in this study may help to resolve the phylogenetic relationships among similar species of *Miscanthus*, such as *M. lutarioriparius* and *M. sacchariflorus* (Fig. 1, Fig. 2).

Conclusions

We developed 469,509 SNPs using SLAF-seq technology. Among these, 11,244 highly consistent population SNPs, with integrity > 0.8 and MAF > 0.05, were identified across all samples. We successfully identified the paternal parent and obtained an intraspecific hybrid polyploid population of *M. lutarioriparius*. Using the molecular markers obtained for this population in this study, high-density molecular marker linkage maps could be constructed, and QTL mapping analyses could be performed to investigate the genes controlling quantitative traits (such as biomass yield and cellulose content) in detail. These results support genetic improvements in biomass yield and fiber quality in *Miscanthus* species.

Materials And Methods

Plant materials

Fifty *Miscanthus* specimens: the maternal plant *Miscanthus lutarioriparius* (A12; tetraploid, 2n = 4x), the paternal candidate *M. sinensis* (A1–A11; diploid, 2n = 2x), the paternal candidate *M. sacchariflorus* (B1–B10; diploid, 2n = 2x), the paternal candidate *M. lutarioriparius* (C1–C9; diploid, 2n = 2x), and 19 progeny (D1–D19; triploid, 2n = 3x). The 19 progeny were generated via open pollination from A12. A12 was bagged during flowering to test its self-pollination seed-setting rate. The measured self-pollinating seed-setting rate was 0, indicating that A12 was highly self-incompatible. Other *Miscanthus* genotypes that flowered simultaneously with A12 in the Germplasm Nursery were selected as paternal candidates. The maternal and paternal candidate materials were collected from across China (Fig. 3). All the plants were

grown in the *Miscanthus* Germplasm Nursery at Hunan Agricultural University, Changsha, Hunan, China (latitude 28°11', longitude 113°04').

DNA extraction

We used fresh leaf material obtained from 50 *Miscanthus* individuals. Fresh leaves were frozen in liquid nitrogen and ground manually into a fine powder. Total genomic DNA was extracted following a modified cetyltrimethyl ammonium bromide (CTAB) method [45]. Because the leaves of *Miscanthus* are rich in phenols, the CTAB extraction solution was supplemented with 2% poly-N-vinylpyrrolidone (PVP) and 1% β-mercaptoethanol to purify the *Miscanthus* DNA. The concentration and quality of the extracted DNA were detected using 0.8% agarose gel electrophoresis and an ND-1000 spectrophotometer (Nano Drop), respectively.

Enzyme digestion design

To identify the most appropriate enzymes for genomic digestion, we selected the *Sorghum bicolor* genome as the reference genome (http://phytozome.jgi. doe.gov/pz/portal.html#!info?alias = Org_Sbicolor). This genome was selected because a recent study suggested that *S. bicolor* and *Miscanthus* were closely related [46]. In addition, the two genomes were of similar size and had similar GC levels [47, 48]. We then predicted suitable restriction enzyme combinations for the digestion of the *Miscanthus* genome, based on the *S. bicolor* genome, using DNassist [49]. To assess the efficiency of the predicted enzymes, the genome of Japanese rice (*Oryza sativa* L. subsp. *japonica*) was selected as a control.

SLAF library construction and high-throughput sequencing

Using the identified restriction enzyme combination (*Eco*RV + *Sca*I), the total genomic DNA of each sample was digested. After adding an A-tail to the 3' end of each digested fragment and ligating the fragment to the Dual-index sequencing adapter [50], each fragment was PCR amplified. The amplicons were purified, mixed, and cut. The digested fragments were chosen as target segments. The libraries were selected and sequenced using an Illumina HiSeq TM 2500 platform (Biomarker Technologies Corporation), with a read length of 2 ×100 bp. To assess the accuracy of SLAF library construction, the Japanese rice genome was again used as a control.

Development of SLAF tags and SNP markers

The Dual-index tags were used to classify the raw sequencing data by sample. Sequences read from the same locus were grouped using similarity clustering [51]. In general, only high-depth fragments were selected in each cluster group; low-depth segments were removed. Here, we first calculated the SLAF tags

for each sample independently, and then all single-sample SLAF tags were clustered to derive population-wide SLAF tags. The SLAF tags with the greatest sequencing depths across all samples were selected as reference sequences for the development of SNP markers. Each sequence was aligned with the reference sequence using bwa [52]. We used GATK [53] and SAMtools [54] to identify SNPs. The SNPs identified by both methods were considered reliable. Of these reliable SNPs, those with integrity > 0.8 and MAF > 0.05 were considered highly consistent and were used for subsequent analyses.

Genetic Relationships Among Samples

We used admixture software [55] to determine the population structure of the 50 *Miscanthus* germplasms, assuming that these germplasms fell into 1–10 genetic clusters. We also performed principal components analysis (PCA) of the germplasms using cluster software [56]. PCAs perform linear transformations of variables to create orthogonal axes ordered by the proportion of variance explained [57]. An unrooted phylogenetic tree based on our SNP data for these 50 germplasms was constructed using the neighbor-joining (NJ) method [58] in MEGA 5.0 software [59]. SLAF-based paternity tests have two requirements: the sequencing depth of the parents in the SLAF tag must be ≥ 10, while the ratio of sequencing depth between parents and offspring in the SLAF tag must be ≥ 2; and the SLAF tag in the tag sequence must be present in one of the parents [44]. If both conditions are met, the SLAF tag is regarded as abnormal; if there are > 3% abnormal tags between the offspring and a given parent, there is unlikely to be a parental relationship [44].

Abbreviations

AFLPs: amplified fragment length polymorphisms; BAC: bacterial artificial chromosome; CTAB: cetyltrimethyl ammonium bromide; GBS: genotyping by sequencing; ISSRs: inter-simple sequence repeats; MAF: minor allele frequency; NJ: Neighbor-joining; PCA: principal components analysis; PVP: poly-N- vinylpyrrolidone; RAD-seq: restriction site-associated DNA sequencing; RAPD: random amplified polymorphic DNA; RFLPs: restriction fragment length polymorphisms; SLAF-seq: specific-locus amplified fragment sequencing; SNPs: single nucleotide polymorphisms; SSRs: simple sequence repeats

Declarations

Acknowledgements

We would like to thank LetPub (www.letpub.com) for providing linguistic assistance during the preparation of this manuscript.

Author Contributions

This research was designed by ZYC, HHH and ZLY; and was performed by YCH with assistance from YLS. ZYC and YCH wrote the paper with help from the other authors. All authors read and approved the final

manuscript.

Authors' information

At the time of this work, YCH and YLS were postgraduate students in the College of Bioscience and Biotechnology at Hunan Agricultural University, Hunan, China. ZYC, HHH, and YI are Assistant Professors, and ZLY is a Professor in the College of Bioscience and Biotechnology at Hunan Agricultural University. All authors are also members of Hunan Engineering Laboratory of *Miscanthus* Ecological Applications, Hunan, China.

Funding

This work was supported by The National Natural Science Foundation of China [31471557, 31871693].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

a College of Bioscience & Biotechnology, Hunan Agricultural University, Changsha, 410128, PR China

b Hunan Engineering Laboratory of Miscanthus Ecological Applications, Hunan Agricultural University, Changsha, 410128, PR China

References

1. Sang T. Toward the domestication of lignocellulosic energy crops: learning from food crop domestication. J Integr Plant Biol. 2011;53:96–104.
2. Sang T. China's bioenergy potential. Global Change Biology Bioenergy. 2011;3:79–90.
3. Clifton-Brown J, Harfouche A, Casler MD, Dylan Jones H, Macalpine WJ, Murphy-Bokern D, et al. Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow and poplar. Glob Change Biol Bioenergy. 2019;11:118–51.

4. Liu W, Sang T. Potential productivity of the Miscanthus energy crop in the Loess Plateau of China under climate change. Environmental Research Letters. 2013;8.
5. Kirkpatrick J. Construction and analysis of the miscanthus genespace (Doctoral dissertation, University of Illinois at Urbana–Champaign). 2014.
6. Olatoye MO, Clark LV, Labonte NR, Dong H, Dwiyanti MS, Anzoua KG, et al. Training Population Optimization for Genomic Selection in Miscanthus. G3 (Bethesda). 2020;10:2465–2476.
7. Olatoye MO, Clark LV, Wang JP, Yang XP, Yamada T, Sacks EJ, et al. Evaluation of genomic selection and marker-assisted selection in Miscanthus and energycane. Molecular Breeding. 2019;39.
8. Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, et al. Construction of an RFLP map of barley. Theor Appl Genet. 1991;83:250–6.
9. Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature. 1988;335:721–6.
10. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 1990;18:6531–5.
11. John W, Michael M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990;18:7213–8.
12. Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23:4407–14.
13. Litt M, Luty JA. A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. Am J Hum Genet. 1989;44:397–401.
14. Reddy MP, Sarla N, Siddiq EA. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica. 2002;128:9–17.
15. Clark LV, Brummer JE, Glowacka K, Hall MC, Heo K, Peng JH, et al. A footprint of past climate change on the diversity and population structure of Miscanthus sinensis. Ann Bot. 2014;114:97–107.
16. Ge C, Ai X, Jia S, Yang Y, Che L, Yi Z, et al. Interspecific genetic maps in Miscanthus floridulus and M. sacchariflorus accelerate detection of QTLs associated with plant height and inflorescence. Mol Genet Genomics. 2019;294:35–45.
17. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. Plos One. 2011;6:e19379.
18. Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, et al. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet. 2013;126:2699–716.
19. Zeng B, Yan HD, Liu XC, Zang WJ, Zhang AL, Zhou SF, et al. Genome-wide association study of rust traits in orchardgrass using SLAF-seq technology. Hereditas. 2017;154.
20. Liu LQ, Luo QL, Li HW, Li B, Li ZS, Zheng Q. Physical mapping of the blue-grained gene from Thinopyrum ponticum chromosome 4Ag and development of blue-grain-related molecular markers

and a FISH probe based on SLAF-seq technology. Theor Appl Genet. 2018;131:2359–70.

21. Jia Q, Tan C, Wang J, Zhang X-Q, Zhu J, Luo H, et al. Marker development using SLAF-seq and whole-genome shotgun strategy to fine-map the semi-dwarf gene ari-e in barley. Bmc Genomics. 2016;17:911.
22. Zhang Y, Zhang JP, Huang L, Gao AN, Zhang J, Yang XM, et al. A high-density genetic map for P genome of Agropyron Gaertn. based on specific-locus amplified fragment sequencing (SLAF-seq). Planta. 2015;242:1335–47.
23. Wei QZ, Wang YZ, Qin XD, Zhang YX, Zhang ZT, Wang J, et al. An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing. Bmc Genomics. 2014;15:1158.
24. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. Plos One. 2013;8:e58700.
25. Liu T, Guo L, Pan Y, Zhao Q, Wang J, Song Z. Construction of the first high-density genetic linkage map of Salvia miltiorrhiza using specific length amplified fragment (SLAF) sequencing. Sci Rep. 2016;6:24070.
26. Wang WW, Sun Y, Yang P, Cai XY, Yang L, Ma JR, et al. A high density SLAF-seq SNP genetic map and QTL for seed size, oil and protein content in upland cotton. Bmc Genomics. 2019;20.
27. Shen C, Jin X, Zhu D, Lin ZX. Uncovering SNP and indel variations of tetraploid cottons by SLAF-sEq. Bmc Genomics. 2017;18.
28. Gong DP, Huang L, Xu XH, Wang CY, Ren M, Wang CK, et al. Construction of a high-density SNP genetic map in fluecured tobacco based on SLAF-sEq. Molecular Breeding. 2016;36.
29. Dong ZM, Chen L, Li Z, Liu NX, Zhang SC, Liu J, et al. Identification and molecular mapping of the semi-dwarf locus (sdf-1) in soybean by SLAF-seq method. Euphytica. 2020;216.
30. Zhang SZ, Hu XH, Miao HR, Chu Y, Cui FG, Yang WQ, et al. QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (Arachis hypogaea L.). Bmc Plant Biology. 2019;19.
31. Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, et al. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol. 2013;13:141.
32. Shan T, Pang S, Li J, Li X, Su L. Construction of a high-density genetic map and mapping of a sex-linked locus for the brown alga Undaria pinnatifida (Phaeophyceae) based on large scale marker development by specific length amplified fragment (SLAF) sequencing. Bmc Genomics. 2015;16:902.
33. Jiang B, Liu WR, Xie DS, Peng QW, He XM, Lin YE, et al. High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing. Bmc Genomics. 2015;16.
34. Shang JL, Li N, Li NN, Xu YY, Ma SW, Wang JM. Construction of a high-density genetic map for watermelon (Citrullus lanatus L.) based on large-scale SNP discovery by specific length amplified fragment sequencing (SLAF-seq). Sci Hortic. 2016;203:38–46.

35. Chen S, Huang Z, Dai Y, Qin S, Gao Y, Zhang L, et al. The development of 7E chromosome-specific molecular markers for Thinopyrum elongatum based on SLAF-seq technology. Plos One. 2013;8:e65122.
36. Xia C, Chen LL, Rong TZ, Li R, Xiang Y, Wang P, et al. Identification of a new maize inflorescence meristem mutant and association analysis using SLAF-seq method. Euphytica. 2015;202:35–44.
37. Zhao X, Huang L, Zhang X, Wang J, Yan D, Li JM, et al. Construction of high-density genetic linkage map and identification of flowering-time QTLs in orchardgrass using SSRs and SLAF-sEq. Sci Rep. 2016;6:29345.
38. Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, et al. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theor Appl Genet. 2019;132:627–45.
39. Liu S, An Y, Tong W, Qin XJ, Samarina L, Guo R, et al. Characterization of genome-wide genetic variations between two varieties of tea plant (Camellia sinensis) and development of InDel markers for genetic research. BMC Genom. 2019;20:935. https://doi.org/10.1186/s12864-019-6347-0.
40. Yang Y, Xuan L, Yu C, Wang ZY, Xu JH, Fan WC, et al. High-density genetic map construction and quantitative trait loci identification for growth traits in (Taxodium distichum var. distichum × T. mucronatum) × T. mucronatum. BMC Plant Biol. 2018;18:263. https://doi.org/10.1186/s12870-018-1493-0.
41. Clark LV, Stewart JR, Nishiwaki A, Toma Y, Kjeldsen JB, Jørgensen U, et al. Genetic structure of Miscanthus sinensis and Miscanthus sacchariflorus in Japan indicates a gradient of bidirectional but asymmetric introgression, Journal of Experimental Botany, Volume 66, Issue 14, July 2015, Pages 4213–4225, https://doi.org/10.1093/jxb/eru511, 2015.
42. Parkinson J, Blaxter M. Expressed sequence tags: an overview. Methods Mol Biol. 2009;533:1–12.
43. Kanazin V, Talbert H, See D, Decamp P, Nevo E, Blake T. Discovery and assay of single nucleotide polymorphisms in barley (Hordeum vulgare). Plant Mol Biol. 2002;48:529–37.
44. Jones JC, Fan SH, Franchini P, Schartl M, Meyer A. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Mol Ecol. 2013;22:2986–3001.
45. Xia YN, Xu J, Duan JY, Liu QB, Huang HM, Yi ZL, et al. Transgenic Miscanthus lutarioriparius that co-expresses the Cry 2Aa# and Bar genes. Can J Plant Sci. 2019;99:841–51.
46. Sheng J, Zheng X, Wang J, Zeng X, Zhou F, Jin S, et al. Transcriptomics and proteomics reveal genetic and biological basis of superior biomass crop Miscanthus. Sci Rep 7, 13777, doi:10.1038/s41598-017-14151-z (2017).
47. Kim C, Tang H, Paterson AH. Duplication and Divergence of Grass Genomes: Integrating the Chloridoids. Tropical Plant Biology. 2009;2:51–62.
48. Kim C, Zhang D, Auckland SA, Rainville LK, Jakob K, Kronmiller B, et al. SSR-based genetic maps of Miscanthus sinensis and M. sacchariflorus, and their comparison to sorghum. Theor Appl Genet. 2012;124:1325–38.

49. Phillips SM, Dubery IA, Van Heerden H. Molecular characterisation of two homoeologous elicitor-responsive lipin genes in cotton. Mol Genet Genomics. 2013;288:519–33.
50. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013;79:5112–20.
51. Liu T, Zhu J, Zhou J, Zhu YX, Zhu X. Initialization-similarity clustering algorithm. Multimedia Tools and Applications. 2019;78.
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
53. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
55. Wang XQ, Huang GH, Zhao S, Guo JH. An open-source software package for multivariate modeling and clustering: applications to air quality management. Environ Sci Pollut Res. 2015;22:14220–33.
56. Zhao YL, Basak S, Fleener CE, Egnin M, Sacks EJ, Prakash CS, et al. Genetic diversity of Miscanthus sinensis in US naturalized populations. Global Change Biology Bioenergy. 2017;9:965–72.
57. Dunteman GH. Principal components analysis. Newbury Park California Sage Publications. 1989;930:527–47.
58. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
59. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 2011;28:2731–9.

Figures

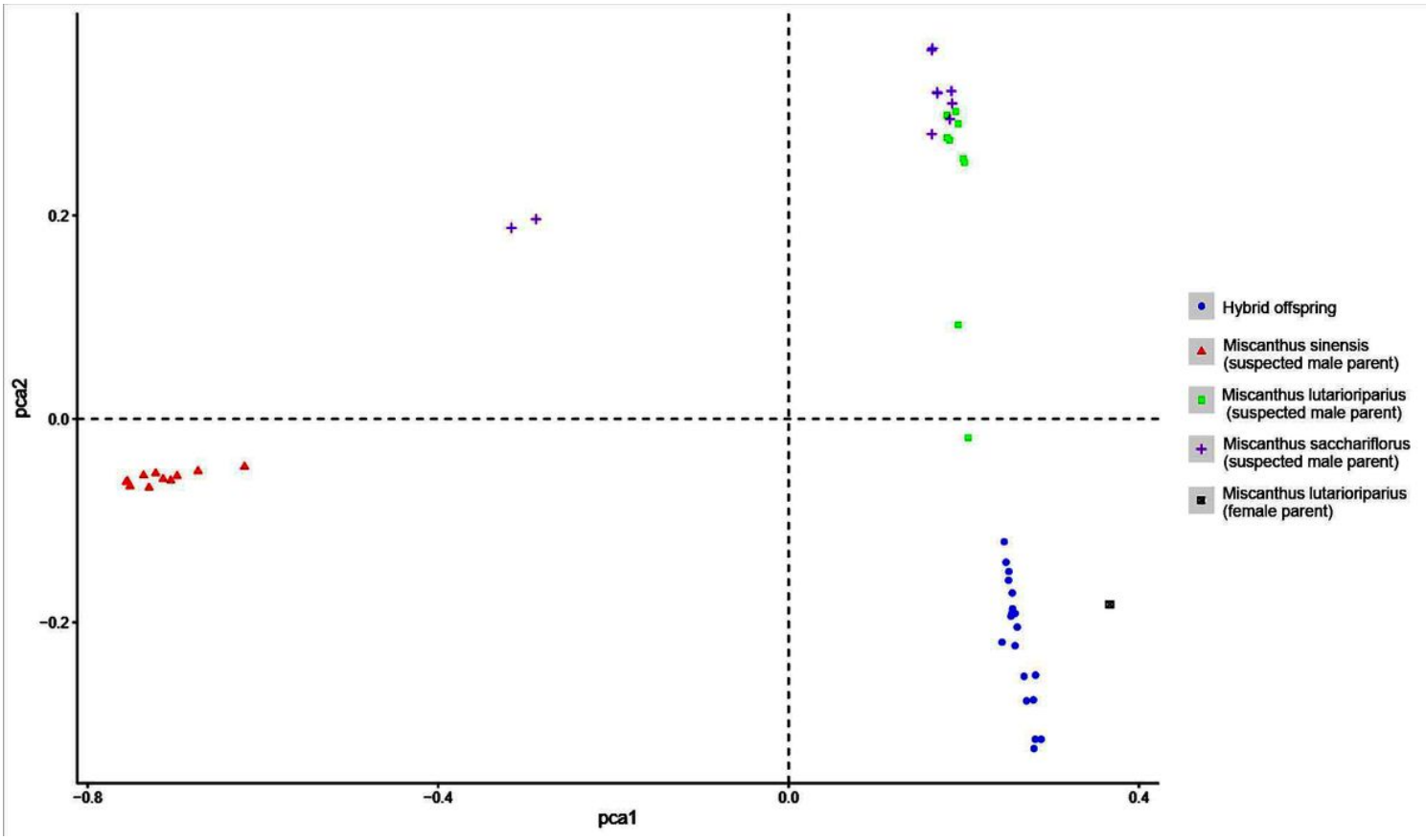

Figure 1

Principal component analysis (PCA) of the 50 Miscanthus germplasms. Each point represents a sample, and points of different color and shape correspond to different sample groups.

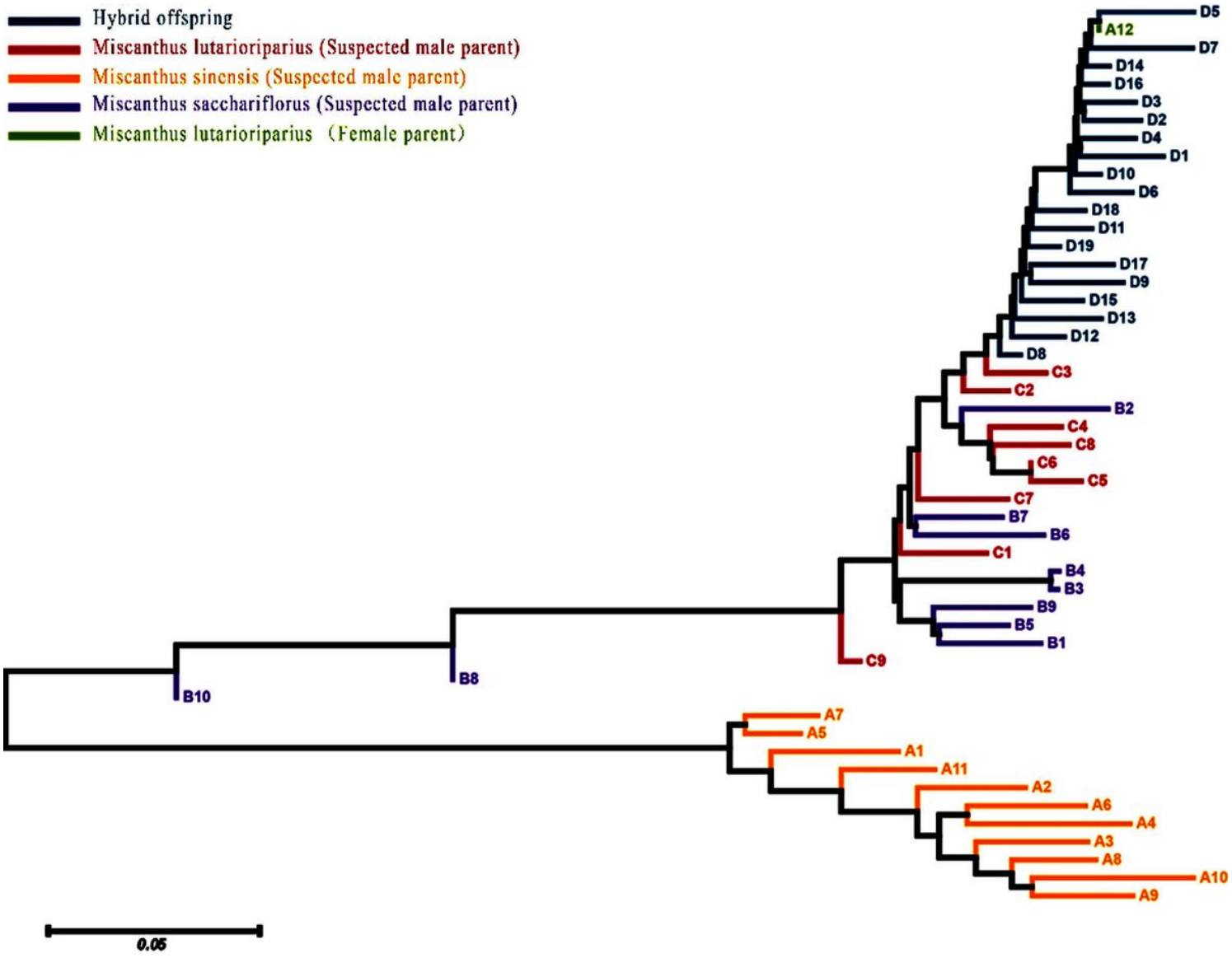

Figure 2

A neighbor-joining phylogenetic analysis of all 50 Miscanthus germplasms. Each node represents a sample. Scale bar represents the genetic distance.

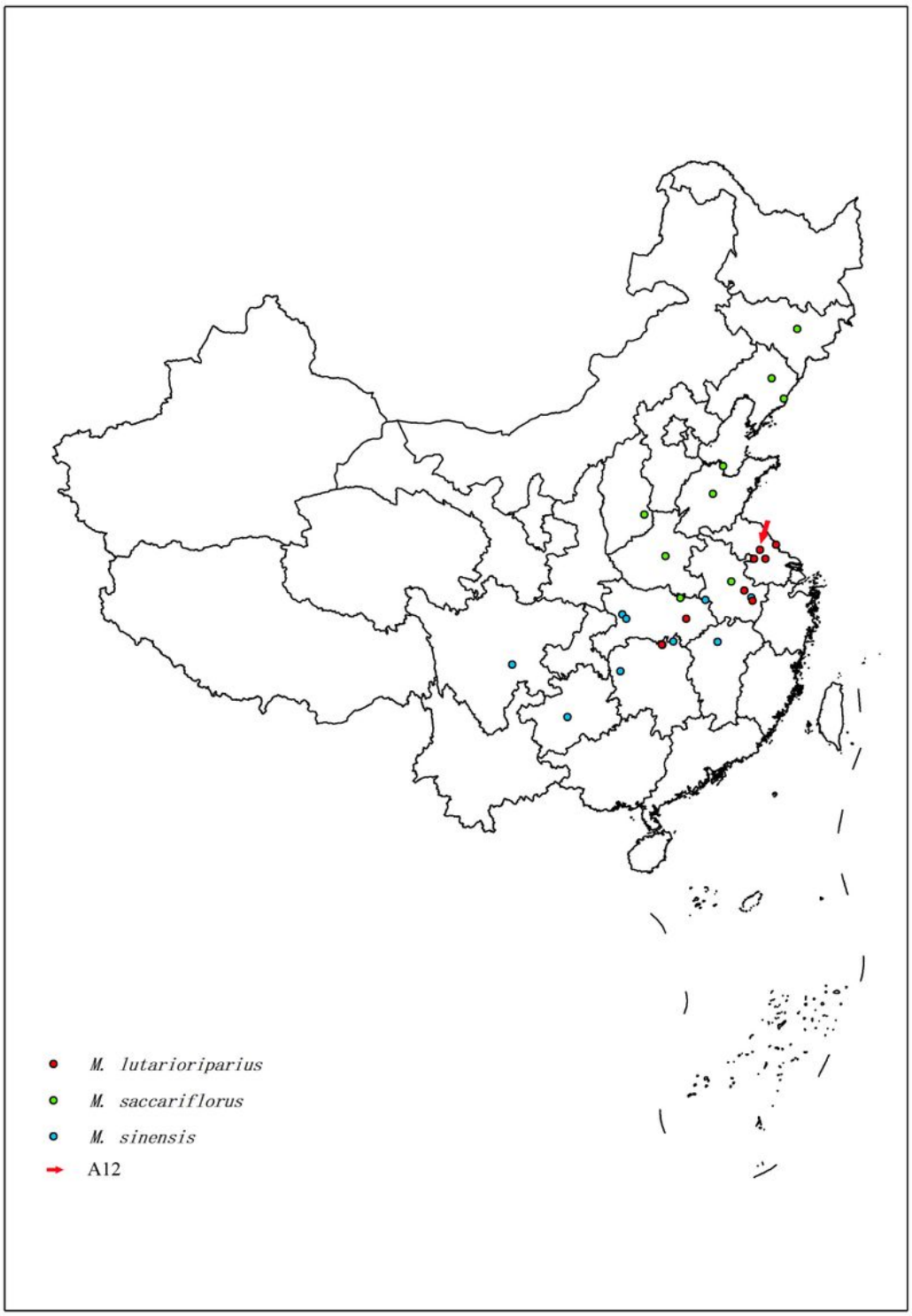

Figure 3

Origins of the 31 pure species Miscanthus germplasms used in this study. Each spot represents a sample; the female parent is sample A12; the remaining samples are the candidate male parents. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country,

territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

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

- renamed4b77e.docx