

# Genetic structure and designing a preliminary core collection of *Zizania latifolia* in China based on 12 microsatellites markers

Xiangliang Lei<sup>1,2,\*</sup>, Xiaona Su<sup>3,\*</sup>, Chengchuan Zhou<sup>4</sup>, Shaolin Jiang<sup>1</sup>, Xiaoquan Yuan<sup>2</sup>, Yao Zhao<sup>5</sup> and Shaomei Jiang<sup>6</sup>

<sup>1</sup> Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Jiangxi Agricultural University, Nanchang, Jiangxi, China

<sup>2</sup> Fuzhou Institute of Agricultural Science, Fuzhou, Jiangxi, China

<sup>3</sup> Nanchang Business College, Jiangxi Agricultural University, Jiujiang, Jiangxi, China

<sup>4</sup> Jiangxi Academy of Forestry, Nanchang, Jiangxi, China

<sup>5</sup> Key Laboratory of Poyang Lake Environment and Resource Utilization, Ministry of Education, Center for Watershed Ecology, School of Life Sciences, Nanchang University, Nanchang, Jiangxi, China

<sup>6</sup> School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China

\* These authors contributed equally to this work.

## ABSTRACT

The genetic diversity and structure of wild crop relatives are crucial for their conservation and utilization in breeding programs. This study presents a comprehensive survey and collection of *Zizania latifolia* across its natural distribution range in China. Using 12 microsatellite markers, the genetic diversity of 357 wild *Z. latifolia* accessions from 25 populations was evaluated, revealing a high genetic diversity ( $H_e = 0.439$ ). The genetic structure analysis indicated significant genetic differentiation among populations, with evidence of isolation by distance. CoreHunter3 and PowerMarker software were employed to design a preliminary core collection, and the final core collection comprised 92 wild accessions. The core collection was found to be representative of the original germplasm, ensuring the effective conservation of *Z. latifolia*'s genetic resources. This study would provide valuable insights for the development of conservation strategies and the utilization of *Z. latifolia*.

**Subjects** Agricultural Science, Biodiversity, Genetics, Plant Science, Population Biology

**Keywords** *Zizania latifolia*, Microsatellites markers, Genetic diversity, Natural distribution, Germplasm

## INTRODUCTION

The emergence and advancement of agricultural civilization has laid the fundamental material foundation for the development and progress of human society. Despite the substantial growth in cultivated plant production over the past century, the diversity of cultivated plants has declined significantly, particularly among major food crops (Meyer, DuVal & Jensen, 2012; Isbell et al., 2017). The three primary staple crops—wheat, maize,

Submitted 26 August 2024

Accepted 6 January 2025

Published 21 February 2025

Corresponding authors

Yao Zhao, yaozhao@ncu.edu.cn

Shaomei Jiang,

13870621151@163.com

Academic editor

Nikolaos Nikoloudakis

Additional Information and  
Declarations can be found on  
page 15

DOI 10.7717/peerj.18909

© Copyright

2025 Lei et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

and rice—currently occupy approximately 80% of the global cereal acreage (*Food and Agriculture Organization (FAO), 2022*). Crop wild relatives play an important role in crop breeding and food security, as they possess far greater genetic diversity and superior agronomic and resistance traits that are often lacking in cultivated varieties (*Zhang et al., 2017*). Exploring and utilizing the genetic resources of wild crop relatives to improve existing crops, breed new high-yielding and high-resistance varieties, or even domesticate novel crops from wild plants, is considered one of the most effective approaches to address future challenges and achieve sustainable societal development (*Choudhary et al., 2017; Kashyap et al., 2022*).

The survey, collection, and evaluation of crop wild relatives form the foundation for their conservation and utilization. However, the conservation of crop wild relatives in large quantities, while providing a rich genetic base for plant genetic improvement and variety selection, also presents challenges in terms of physical space and maintenance costs (*Brush & Meng, 1998; Ford Lloyd et al., 2011*). To address this problem, *Frankel (1984)* first proposed the concept of core germplasm, which was further developed and refined by *Brown (1989)*. The identification and conservation of core germplasm have been widely adopted in cultivated plants, such as rice, sweet potato and olive (*Belaj et al., 2012; Su et al., 2017; Kumar et al., 2020*). However, relatively few efforts have been made to create core germplasm for wild plants, especially wild relatives of crops (*Miyamoto, Ono & Watanabe, 2015; Bai et al., 2019; Wang et al., 2023*), which evidently limits our ability to further explore and utilize their genetic resources.

The genus *Zizania* taxonomically belongs to the family Gramineae, subfamily Rice, and tribe Rice (*Chen & Xu, 1994; Terrell et al., 1997*). Globally, there are four recognized *Zizania* species: *Z. latifolia*, *Z. palustris*, *Z. aquatica*, and *Z. texana*. While *Z. latifolia* is found in East Asia, the latter three species are native to North America. *Z. palustris*, commonly known as wild rice in North America, was traditionally consumed as a food by indigenous communities and subsequently domesticated as a specialty food crop in the region. Similarly, *Z. latifolia* was once utilized as a food crop in China, but its ancestors were eventually domesticated to produce a distinct aquatic vegetable called “Jiaobai” in the form of a plant-fungus complex (*Guo et al., 2007; Guo et al., 2015; Zhao et al., 2019*).

Today, cultivated *Z. latifolia* is a highly popular aquatic vegetable in southern China, second only to lotus root in terms of cultivation area (*Guo et al., 2007*). As a tertiary gene pool of Asian cultivated rice, *Z. latifolia* represents an important genetic resource for rice improvement, as its life history and phenotypic traits are highly analogous to those of rice, while possessing numerous unique and desirable characteristics that allow for hybridization with rice (*Liu, Liu & Li, 1999; Guo et al., 2015*). Inspired by the successful domestication of *Z. palustris* in North America, coupled with the development of advanced genomic breeding technologies and the acknowledged high nutritional value of *Z. latifolia*, some researchers have proposed the possibility of *de novo* domestication of *Z. latifolia* as a grain crop (*Zhao et al., 2018; Yan et al., 2022; Xie et al., 2023*).

The natural distribution range of *Z. latifolia* in China spans a wide latitudinal expanse, with wild populations extending from the southwestern to the northeastern regions, where it is a common aquatic plant in freshwater wetland habitats. However, in recent decades,

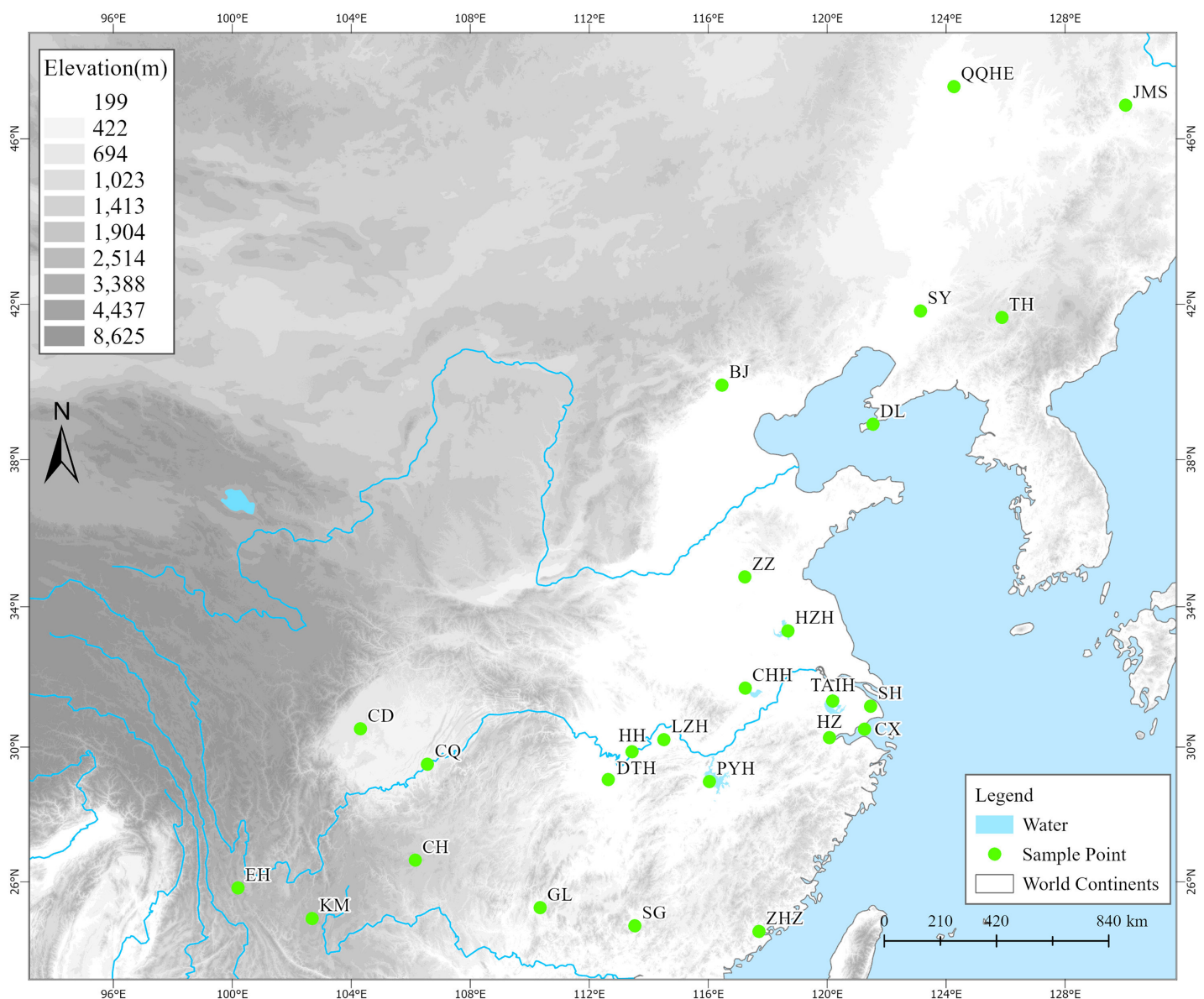
the natural habitat of *Z. latifolia* has been extensively lost due to rapid population growth, intensified human economic activities, and accelerated urbanization, leading to the decline or even local extinction of some populations. Consequently, a comprehensive survey, collection, and evaluation of wild *Z. latifolia* populations, as well as the establishment of a dedicated germplasm resource nursery, are not only necessary for the conservation of its genetic resources but also form the foundation for the exploration and utilization of this species' potential. Several studies have been conducted to investigate the genetic diversity and population genetic structure of *Z. latifolia*. For instance, [Xu et al. \(2008\)](#) analyzed 21 wild *Z. latifolia* populations using a single nuclear gene (*Adh1*) sequence and found low levels of nucleotide diversity, with no evident geographic population structure detected. In a subsequent study, [Chen et al. \(2012, 2017\)](#) evaluated the genetic diversity of *Z. latifolia* populations in the middle reaches of the Yangtze River and the northeastern region of China using SSR markers, while [Zhao et al. \(2018\)](#) also examined *Z. latifolia* populations in the eastern region of China using the same molecular approach and reported a clinal genetic structure from north to south. [Zou et al. \(2024\)](#) also found two genetic groups corresponding to the north and south of *Z. latifolia*'s distribution range using whole genome resequencing. More recently, [Wagutu et al. \(2022\)](#) analyzed 28 *Z. latifolia* populations from five major watersheds using 46 SSR loci and concluded that the species exhibits relatively lower overall genetic diversity and significant genetic differentiation among watersheds. Although the findings of these studies may vary depending on the sampling strategies or methodologies employed, they collectively reveal patterns of genetic differentiation within the existing *Z. latifolia* populations and provide a solid foundation for the collection and identification of genetic resources.

Building upon the findings from previous studies, we conducted a comprehensive survey and collection of *Z. latifolia* samples across its potential natural distribution range in China, and subsequently established an *ex situ* germplasm bank for this species. We then proceeded to evaluate the genetic diversity of the *Z. latifolia* germplasm resources using SSR molecular markers and attempted to construct a preliminary core germplasm subset. This multifaceted investigation can serve as a valuable reference for the conservation and utilization of *Z. latifolia*'s genetic resources. Furthermore, the study has helped to elucidate the endangered status of wild *Z. latifolia* populations, thereby informing the development of appropriate conservation strategies for this important species.

## MATERIALS AND METHODS

### Sample collection strategy

From 2019 to 2021, we conducted surveys and collected samples based on existing digital herbarium records (Chinese Virtual Herbarium, <https://www.cvh.ac.cn>) at 40 potential distribution areas of *Z. latifolia* (including habitats such as lakes, rivers, and swamps) across 28 provinces. Among these areas, wild *Z. latifolia* was found in only 25 survey areas, predominantly concentrated east of the Heihe-Tengchong line ([Fig. 1](#)). In each survey area, we selected a number of evenly distributed sampling sites according to the actual size of the area, collecting three live plants from each site, with a spacing of not less than 1,000 m between individual plants. These plants were then transplanted to the aquatic plant



**Figure 1** The geographical locations of 25 wild *Zizania latifolia* populations in China. A total of 25 survey localities from different areas in China. (1) QQHE, Qiqihar; (2) JMS, Jiamusi; (3) TH, Tonghua; (4) SY, Shenyang; (5) DL, Dalian; (6) BJ, Beijing; (7) ZZ, Zaozhuang; (8) HZH, Hongze; (9) TAIH, Wuxi; (10) SH, Shanghai; (11) CX, Cixi; (12) HZ, Hangzhou; (13) CHH, Chaohu; (14) PYH, Poyang; (15) HH, Honghu; (16) LZH, Wuhan; (17) DTH, Yueyang; (18) CD, Chengdu; (19) CQ, Chongqing; (20) ZHZ, Zhangzhou; (21) SG, Shaoguan; (22) GL, Guilin; (23) CH, Weining; (24) KM, Kunming; (25) EH, Dali. Green filled circles represent survey localities. The base map was obtained from the National Catalogue Service for Geographic Information (<http://www.webmap.cn>). [Full-size DOI: 10.7717/peerj.18909/fig-1](https://doi.org/10.7717/peerj.18909/fig-1)

resource nursery at the Lushan Botanical Garden in Jiujiang, Jiangxi Province, for *ex situ* conservation. Once the collected plants were established and had survived, 3 to 5 healthy, young leaves were selected, placed in a sealed bag with an appropriate amount of color-changing silica gel, and stored in a refrigerator at  $-80^{\circ}\text{C}$  after complete drying for DNA extraction. Ultimately, we obtained a total of 357 wild *Z. latifolia* samples from 132 sampling sites across 25 populations (survey areas).



## DNA extraction and PCR assays

Total genomic DNA was extracted using a plant genomic DNA kit (Tiangen Biotech, Beijing, China). DNA quality and concentration were then examined using NanoDrop ND2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Twelve microsatellites with clear bands and relatively high polymorphism were selected from 24 *Z. latifolia*-specific and cross-specific microsatellites after the pre-experiment, and the specifically developed markers were generally better than the cross-specific microsatellites (Quan *et al.*, 2009; Richards *et al.*, 2004, 2007). PCR amplification was performed following the protocol by Quan *et al.* (2009), and the PCR products were separated on a 6% denaturing polyacrylamide gel (Fig. S1). The fragments were then visualized by silver staining, and alleles were scored in reference to a DNA marker (100 bp ladder; Tiangen Biotech, Beijing, China). To ensure the high quality of the codominant SSR dataset, several precautions were taken, as suggested by Bonin *et al.* (2004). Then we manually converted the raw data into the input format of the software GENALEX 6.5 (Peakall & Smouse, 2012).

## Genetic data analyses

The parameters of population genetic variation were estimated using mean allele number ( $N_a$ ), mean effective allele number ( $A_e$ ), Shannon information index ( $I$ ), polymorphism information content ( $PIC$ ), expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), and Wright's fixation index ( $FI$ ) with GENALEX 6.5 and PowerMarker v3.25 (Liu & Muse, 2005; Peakall & Smouse, 2012).  $F$ -statistics were calculated to estimate the extent of genetic differentiation among populations. Analysis of Molecular Variation (AMOVA) was conducted to examine genetic variation both among and within populations. A Principal Coordinates Analysis (PCoA) was then performed to illustrate the pattern of genetic divergence.  $F$ -statistics, AMOVA, and PCoA analyses were all carried out using GENALEX 6.5 (Peakall & Smouse, 2012). The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrograms of populations and individuals based on Nei's distance were respectively constructed and visualized in MEGA X (Kumar *et al.*, 2018). The genetic structure was further explored using the Bayesian clustering algorithm implemented in STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly, 2000). The program was provided with no prior information on ancestral populations and was run 10 times for each value of  $K$  ( $K$  ranges from 1 to 25), under the admixture model with independent allele frequencies, using 1,000,000 Markov chain Monte Carlo iterations and a burn-in of 100,000 iterations. We inferred  $K$  using the *ad hoc* statistic  $\Delta K$  (Evanno, Regnaut & Goudet, 2005). The resulting matrices of estimated cluster membership coefficients ( $Q$ ) were permuted with CLUMPP (Jakobsson & Rosenberg, 2007). The final matrix for each  $K$  value was visualized with DISTRUCT (Rosenberg, 2004). Isolation by Distance (IBD) among populations was tested using the Mantel test implemented in GENALEX 6.5. The significance of IBD values was assessed using 9,999 permutations.

## Design the core collection

The core collection of *Z. latifolia* germplasm was constructed using the following methods: First, the program CoreHunter3 (De Beukelaer, Davenport & Fack, 2018) was employed to develop independent core collections based on genotypic data. Various cutoff values, including 10%, 15%, 20%, 25%, 30%, 40%, and 50% of the initial collection, were used to design the core collections in CoreHunter3, based on Modified Rogers distance (MR) with default parameters. Next, PowerMarker v3.25 software was utilized, and core collections were constructed using the simulated annealing algorithm (SA), specifically with maximizing allelic richness (SANA) and maximizing genetic diversity (SAGD). The cutoff values were set identically to those used in CoreHunter3. Finally, the core collections extracted by each method with the 10% cutoff value were combined to form the final core germplasm of *Z. latifolia*.

## RESULTS

### Genetic diversity

The genetic variation among 357 wild *Z. latifolia* accessions was estimated using 12 SSR markers. Generally, all 12 SSR loci were polymorphic, each generating two to 12 alleles, indicating relatively low SSR variation. The mean number of alleles ( $N_a$ ), mean number of effective alleles ( $A_e$ ), polymorphism information content (PIC), Shannon information index ( $I$ ), observed heterozygosity ( $H_o$ ), and expected heterozygosity ( $H_e$ ) for each SSR locus are listed in Table S1. The values for heterozygosity varied among these loci ( $H_o = 0-0.101$ ,  $H_e = 0.028-0.373$ ). The *Z. latifolia* populations exhibited a relatively high level of mean genetic diversity, with  $N_a = 1.823$ ,  $A_e = 1.443$ ,  $I = 0.345$ ,  $H_o = 0.012$ , and  $H_e = 0.207$ .

The sampling information, population size, and genetic diversity parameters for each population are listed in Table 1. The overall genetic diversity of wild *Z. latifolia* accessions used in this study was high, with  $N_a = 5.000$ ,  $A_e = 1.963$ ,  $H_o = 0.013$ , and  $H_e = 0.439$ . However, three populations (SH, HZ, and CD) showed no SSR variability, and all samples within these populations were genetically uniform. The unexpectedly high mean fixation index ( $FI = 0.976$ , Table 1) indicated significant heterozygosity deficiency, although these values varied among populations.

### Genetic structure

The  $F$ -statistics indicated high genetic differentiation among populations, with an overall genetic differentiation coefficient  $F_{st} = 0.452$  (Table S1). This was further confirmed by AMOVA, which revealed that 60.35% of the diversity occurred within individuals, while 37.11% was attributed to genetic differentiation among groups (Table S2).

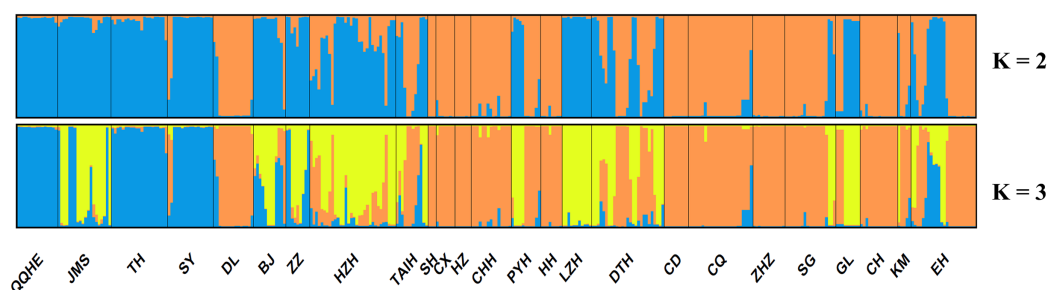
A model-based Bayesian approach was employed to infer the genetic structure of individuals under  $K = 2$ , which resulted in the best fit to the model as estimated by Evanno's  $\Delta K$  statistics (Fig. S2). According to the genetic assignment analysis of STRUCTURE, when  $K = 2$ , the populations of wild *Z. latifolia* could be divided into two genetic groups. Populations from high latitudes (QQHE, JMS, TH, SY, BJ, ZZ, and HZH) were clustered into Group I (North group), while the remaining populations were

**Table 1** Parameters of genetic diversity of 25 wild *Z. latifolia* populations based on 12 microsatellites.

Population code	Location	Latitude	Longitude	Population size	Sampling site number	Accession number	$N_a$	$A_e$	$H_o$	$H_e$	$FI$
QQHE	Qiqihar	47.2	124.25	S	4	16	1.750	1.383	0.024	0.201	0.894
JMS	Jiamusi	46.75	130.08	S	8	23	2.667	2.080	0.037	0.381	0.903
TH	Tonghua	41.65	125.87	S	5	23	1.833	1.486	0.048	0.261	0.890
SY	Shenyang	41.81	123.13	S	6	21	2.083	1.661	0.000	0.337	<b>1.000</b>
DL	Dalian	38.91	121.46	S	3	15	1.750	1.283	0.000	0.172	<b>1.000</b>
BJ	Beijing	39.95	116.45	S	4	12	1.917	1.396	0.000	0.225	<b>1.000</b>
ZZ	Zaozhuang	34.81	117.23	S	3	9	1.833	1.661	0.037	0.289	0.863
HZH	Hongze	33.31	118.66	L	17	32	3.333	2.037	0.016	0.431	0.970
TAIH	Wuxi	31.31	120.19	L	4	12	2.750	1.936	0.064	0.464	0.869
SH	Shanghai	31.19	121.45	S	1	3	1.000	1.000	0.000	0.000	–
CX	Cixi	30.5	121.25	S	3	9	1.167	1.054	0.000	0.044	<b>1.000</b>
HZ	Hangzhou	30.26	120.06	S	3	9	1.000	1.000	0.000	0.000	–
CHH	Chaohu	31.67	117.23	L	5	15	1.333	1.196	0.000	0.094	<b>1.000</b>
PYH	Poyang	28.98	116.03	L	4	11	2.583	1.860	0.015	0.420	0.953
HH	Honghu	29.86	113.43	L	3	9	1.250	1.091	0.000	0.060	<b>1.000</b>
LZH	Wuhan	30.21	114.5	L	6	16	2.500	1.952	0.000	0.428	<b>1.000</b>
DTH	Yueyang	29.03	112.63	L	9	27	3.250	2.169	0.015	0.454	0.900
CD	Chengdu	30.51	104.3	S	3	9	1.000	1.000	0.000	0.000	–
CQ	Chongqing	29.5	106.55	S	8	24	1.583	1.120	0.000	0.090	<b>1.000</b>
ZHZ	Zhangzhou	24.45	117.7	S	4	12	1.250	1.147	0.000	0.068	<b>1.000</b>
SG	Shaoguan	24.65	113.53	S	7	20	1.583	1.259	0.004	0.177	0.975
GL	Guilin	25.13	110.66	S	3	9	1.500	1.365	0.028	0.221	0.800
CH	Weining	26.65	106.13	S	5	14	1.250	1.071	0.000	0.051	<b>1.000</b>
KM	Kunming	24.9	102.68	S	2	5	1.583	1.350	0.000	0.237	<b>1.000</b>
EH	Dali	25.8	100.2	L	12	25	1.833	1.525	0.003	0.288	0.988
Overall					132	357	5.000	1.965	0.013	0.439	0.976

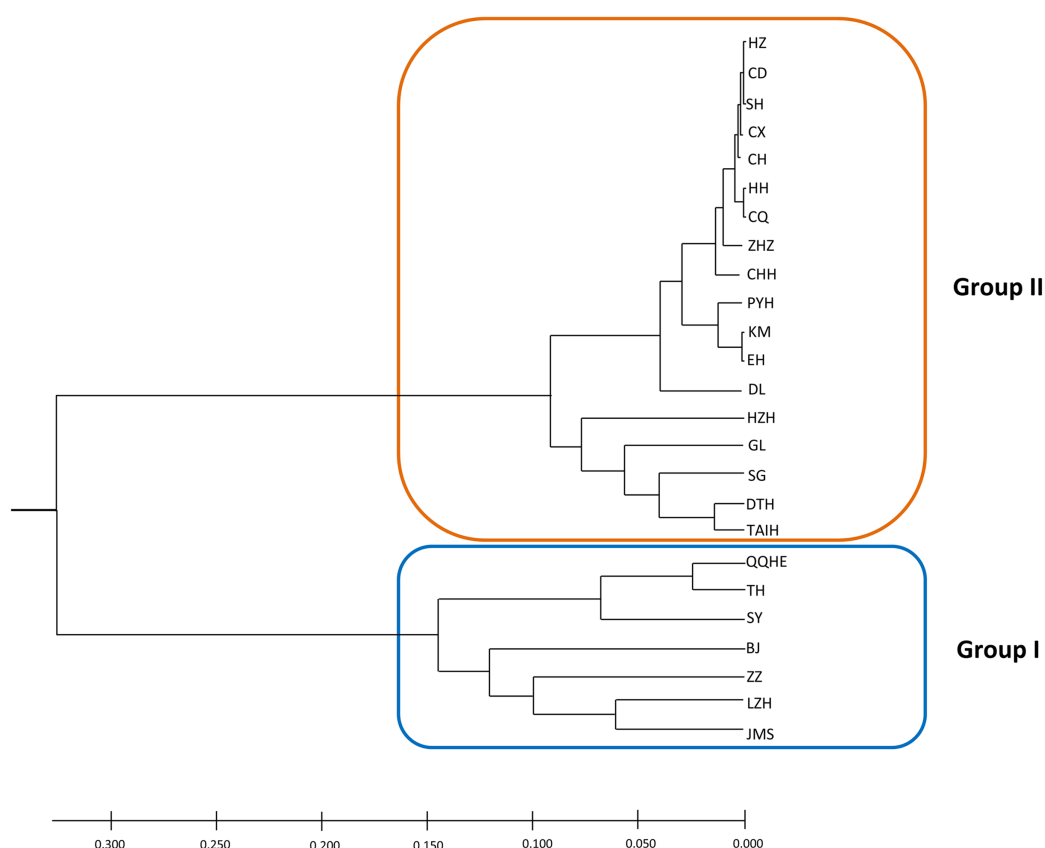
**Note:** The population size was labeled by S (Small, <500) or L (Large, >1,000);  $N_a$ , mean number of alleles;  $A_e$ , number of effective alleles;  $H_o$ , observed heterozygosity;  $H_e$ , expected heterozygosity;  $FI$ , fixation index. The  $FI$  values equal to 1 are shown in bold font.

clustered into Group II (South group) (Fig. 2). This pattern was also supported by the phylogenetic trees (Fig. 3; Fig. S3). We also demonstrated the genetic divergence pattern illustrated by STRUCTURE when  $K = 3$ , where intermediate populations between the north and south groups were further clustered, showing genetic subdivisions within genetic groups (Fig. 2). The PCoA generated a similar pattern of population genetic structure. The first two components of PCoA explained a total variation of 39.52% (Fig. 4). The first principal coordinate accounted for 31.73% of the total variation and allowed the discrimination of the north and south genetic groups. The second coordinate, which accounted for 7.79% of the total variation, showed no clear differentiation pattern. The Mantel test provided significant evidence for isolation by distance (IBD) in *Z. latifolia* ( $R^2 = 0.14$ ,  $p < 0.01$ ) (Fig. 5).



**Figure 2** Proportions of ancestry of 25 *Z. latifolia* populations based on  $K = 2$  and  $K = 3$  subdivisions. Each bar represents the proportional membership assignments of an individual.

Full-size [DOI: 10.7717/peerj.18909/fig-2](https://doi.org/10.7717/peerj.18909/fig-2)



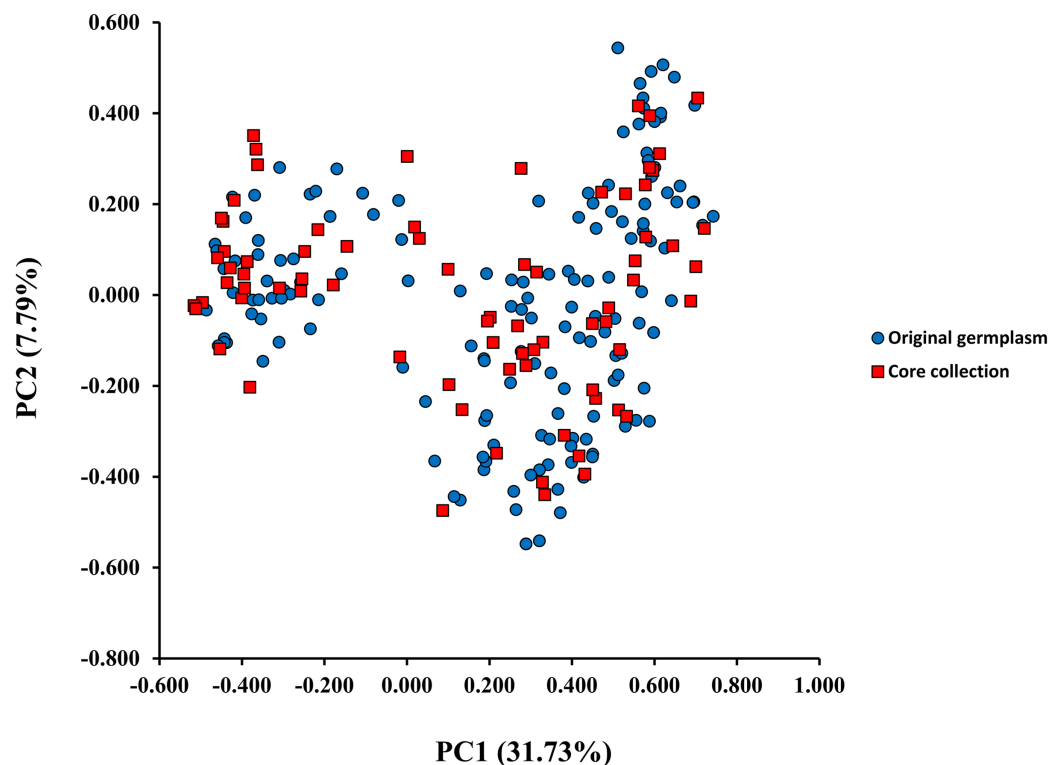
**Figure 3** The phylogeny tree (UPGMA) for the 25 wild *Z. latifolia* populations. Two genetic groups (Group I and Group II) were illustrated.

Full-size [DOI: 10.7717/peerj.18909/fig-3](https://doi.org/10.7717/peerj.18909/fig-3)

## Development of a preliminary core collection

The CoreHunter3 software was utilized by setting the weights of MR to 0.5 each and applying sampling proportions of 10%, 15%, 20%, 25%, 30%, 40%, and 50%. This resulted in core germplasm samples of 36, 54, 71, 89, 107, 143, and 179, respectively. The sampling proportions in the SA method, which includes the SANA and SAGD methods, were kept consistent with those in CoreHunter3.





**Figure 4** The scatterplots of first two PC axis of principal-coordinate-analysis (PCoA) for original germplasm (blue circles) and core collection (red squares) of wild *Z. latifolia*.

Full-size [DOI: 10.7717/peerj.18909/fig-4](https://doi.org/10.7717/peerj.18909/fig-4)

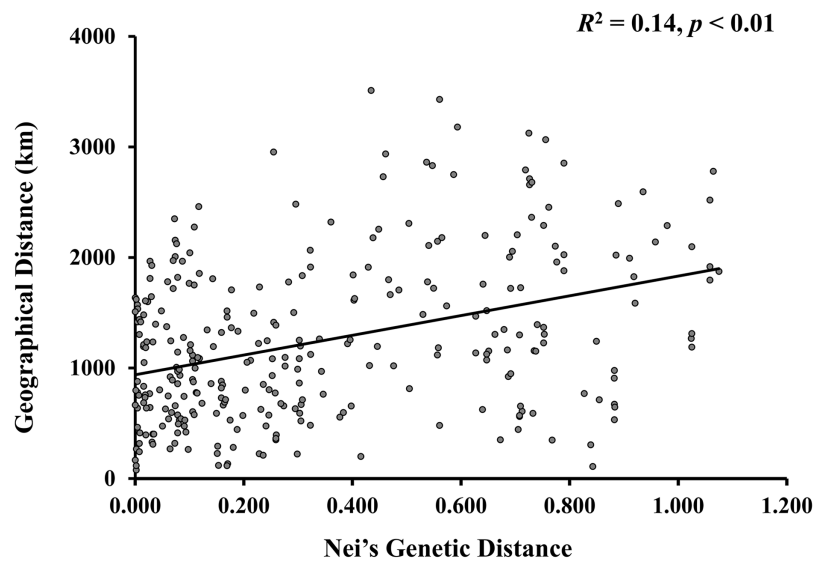
The diversity parameters for each sampled core collection are presented in Table 2. For SANA, the  $H_e$  peaked at 0.444 when the core germplasm number reached 89 (25%), while the genetic diversity under SAGD increased linearly, peaking at 0.426 when the sample size was at its maximum. In contrast, the highest genetic diversity (0.582) was observed when the sampling proportion was set to 10% in CoreHunter3.

The core germplasms generated by different methods showed minimal overlap (Fig. 6). After removing redundant individuals, we obtained a merged core collection containing 92 accessions. Table 3 illustrates that there were no significant differences in genetic diversity parameters between the 92 core accessions and the original 357 accessions. Figure 3 also provides insight into the distribution of individuals as determined by principal coordinates analysis, showing that the 92 core accessions were evenly distributed across the entire set of 357 samples, indicating effective preservation of genetic diversity.

## DISCUSSION

### The distribution and habitat changes of *Z. latifolia* in China

According to the results of our nationwide survey of wild *Z. latifolia* populations during the 3-year period from 2019 to 2021, no wild *Z. latifolia* populations were found in 15 of the 40 survey areas that were pre-set based on the herbarium's historical distribution data. These areas are mainly located in northern and northwestern China. In contrast, wild



**Figure 5** The scatterplots between genetic distance and geographic distance for *Z. latifolia* populations. Genetic distance is represented by pairwise Nei's distance among populations, which is correlated with geographic distance between populations. The regression line overlays the scatterplot (Mantel test,  $R^2 = 0.14$ ,  $p < 0.01$ ). [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02\_img.jpg\) DOI: 10.7717/peerj.18909/fig-5](https://doi.org/10.7717/peerj.18909/fig-5)

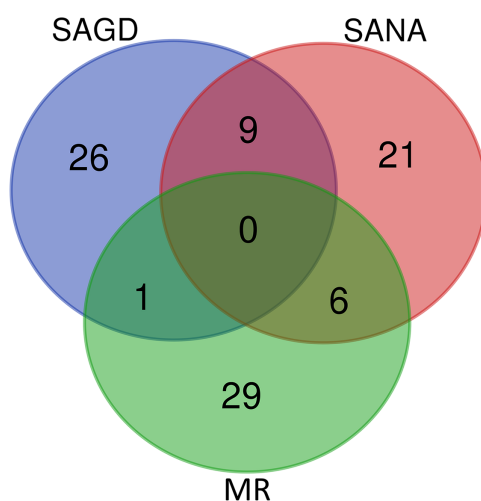
*Z. latifolia* populations were found in the remaining 25 survey areas. Currently, wild *Z. latifolia* is concentrated in the Yangtze River Basin, and some populations also exist in the Haihe River Basin and Northeast China. However, wild *Z. latifolia* populations in areas south of the Yangtze River Basin are difficult to find, which is consistent with the findings of Zhao et al. (2018). Moreover, we found large habitat loss in less than 5 years when we revisited some populations, mainly due to the destruction of wetlands during urbanization. In fact, in recent decades, a large number of freshwater wetlands have been reclaimed and their areas drastically reduced due to factors such as population growth and a lack of understanding of the ecological functions of wetlands. As a result, suitable habitats for wild *Z. latifolia* have been lost. According to incomplete statistics, since the end of the 1940s, more than one-third of the lakes in the middle and lower reaches of the Yangtze River have been reclaimed, with a total area of more than 13,000 km<sup>2</sup>, and over 1,000 lakes have disappeared due to reclamation. About two-thirds of the lakes in the middle reaches of the Yangtze River have vanished due to reclamation, with the area of Dongting Lake shrinking from 4,350 to 2,625 km<sup>2</sup>, and the area of Poyang Lake decreasing from 5,200 to 2,933 km<sup>2</sup> (Yang et al., 2010). Additionally, the blockage of the water system by water conservancy projects, such as the construction of artificial gates and embankments, may further limit the natural expansion of wild *Z. latifolia* populations. For example, the middle and lower reaches of the Yangtze River, which harbor the largest and most concentrated wild populations of *Z. latifolia*, have only three lakes—Dongting Lake, Poyang Lake, and Shijiu Lake—that remain naturally connected to the river.

**Table 2** Genetic diversity parameters of wild *Z. latifolia* germplasms for different core subsets.

Method	Sampling proportion	Accession number	Population number	A	$H_e$
Overall	100%	357	25	60	0.439
SANA	10%	36	17	42	0.422
	15%	54	22	47	0.417
	20%	71	22	48	0.412
	25%	89	23	55	0.444
	30%	107	24	49	0.405
	40%	143	25	56	0.429
	50%	179	24	56	0.437
SAGD	10%	36	15	38	0.333
	15%	54	18	44	0.382
	20%	71	22	48	0.389
	25%	89	21	50	0.402
	30%	107	23	55	0.403
	40%	143	23	55	0.415
	50%	179	24	58	0.426
MR	10%	36	11	54	0.582
	15%	54	14	56	0.569
	20%	71	15	59	0.546
	25%	89	19	60	0.546
	30%	107	20	60	0.534
	40%	143	21	60	0.515
	50%	179	25	60	0.497

**Note:**

A, allele number;  $H_e$ , expected heterozygosity.



**Figure 6** Venn diagram of core germplasms identified by three different methods (SAGD, SANA and MR). Full-size [DOI: 10.7717/peerj.18909/fig-6](https://doi.org/10.7717/peerj.18909/fig-6)

**Table 3** Comparison of genetic parameters between core collection and original germplasm.

Parameters	Original germplasm	Core germplasm
<i>N</i>	357	92
<i>A</i>	60	57
<i>A<sub>e</sub></i>	1.965	2.177
<i>I</i>	0.345	0.917
<i>H<sub>o</sub></i>	0.013	0.016
<i>H<sub>e</sub></i>	0.439	0.490
<i>FI</i>	0.976	0.971

**Note:**

*N*, sample size; *A*, allele number; *A<sub>e</sub>*, number of effective alleles; *I*, Shannon Index; *H<sub>o</sub>*, observed heterozygosity; *H<sub>e</sub>*, expected heterozygosity; *FI*, fixation index.

### Genetic diversity of wild *Z. latifolia* populations

Genetic diversity is an important component of biological diversity, as richer genetic diversity within a species enhances its potential ability to adapt to environmental changes (Hughes *et al.*, 2008). As a perennial aquatic plant, *Z. latifolia* exhibits strong clonal growth (Xu *et al.*, 2008; Chen *et al.*, 2012). We observed this characteristic during the sampling process and took necessary measures to avoid collecting too many clones. The phylogeny of the genus *Zizania* suggests that the ancestors of *Z. latifolia* originated in the Americas and migrated to the Far East and East Asia through the Bering Strait during the Tertiary period (Xu *et al.*, 2010). Xu *et al.* (2008) investigated the phylogeny and genetic structure of *Z. latifolia* in China based on the nuclear gene sequence *Adh1a*, finding that populations at high latitudes (Northeast China) possessed a richer range of haplotypes. Xu *et al.* (2015) also used three pairs of universal microsatellite markers to compare the genetic diversity of four species within the genus *Zizania*. They found that the genetic diversity of *Z. latifolia* was relatively low ( $H_e = 0.374$ ) compared to *Z. palustris* ( $H_e = 0.630$ ). However, this work may have underestimated genetic diversity due to the use of too few molecular markers. Subsequently, Zhao *et al.* (2019) and Wagutu *et al.* (2022) examined the genetic diversity and genetic structure of wild populations of *Z. latifolia* in China using SSR molecular markers. Both studies found high genetic diversity (mean diversity:  $H_e = 0.426$  in Zhao *et al.*, 2019,  $H_e = 0.173$  in Wagutu *et al.*, 2022; overall diversity:  $H_e = 0.432$  in Zhao *et al.*, 2019,  $H_e = 0.241$  in Wagutu *et al.*, 2022). In comparison, the present study found a similar level of overall genetic diversity ( $H_e = 0.439$ ) as Zhao *et al.* (2019), but higher than Wagutu *et al.* (2022); and the mean diversity ( $H_e = 0.207$ ) was at the same level as Wagutu. These findings may be due to sampling differences between studies. Although Wagutu *et al.* (2022) surveyed more populations, their sampling was limited to watersheds on five latitudinal gradients. Populations within the watersheds may trend to be genetically identical due to the presence of gene flow, thus making relatively low overall genetic diversity. In contrast, both Zhao *et al.* (2019) and our work sampled populations extensively across the country. Unfortunately, previous studies (Xu *et al.*, 2015; Zhao *et al.*, 2019; Wagutu *et al.*, 2022) did not calculate the genetic parameters of *I* and *PIC*, preventing us from directly assessing differences in genotypic diversity across studies. However, the very low observed

heterozygosity ( $H_o$ ) coupled with an unusually high fixation index ( $FI$ ) suggests a heterozygote deficiency in most populations. To collect more representative genotypes for constructing a germplasm repository, we sampled an area (population) with multiple sample sites spread far apart to cover as much of the area as possible, spacing samples approximately 1,000 m apart within each sample point. This sampling strategy may have two potential consequences: First, multiple populations might have been mistakenly sampled as one, leading to genetic structure subdivision and high overall genetic diversity (Wahlund effect). Second, if some wild populations of *Z. latifolia* are already experiencing heterozygote deficiency (Zhao et al., 2019), the sign of heterozygote deficiency could be further amplified because of the higher likelihood of collecting homozygous individuals. The increased probability of biparental inbreeding due to extensive clonal growth in *Z. latifolia* may also raise the proportion of homozygous genotypes in the population. Additionally, heterozygote deficiency may result from the insufficient polymorphism and number of molecular markers used in this study. For example, Wagutu et al. (2022) used 46 pairs of SSR markers and found heterozygote excess ( $FI = -0.245$ ), whereas only 12 pairs of SSR markers were used in our study. Although conventional gel electrophoresis may underestimate the actual number of alleles, the relatively low PIC (mean PIC = 0.383) also suggests that this earlier batch of developed SSR markers used in our study may have insufficient polymorphism and that new high-polymorphic SSR markers need to be developed using the genome or transcriptome data (Quan et al., 2009; Wagutu et al., 2022; Wang et al., 2023). Now that a high-quality chromosome-level genome of *Z. latifolia* has been released (Yan et al., 2022), subsequent studies can employ whole-genome resequencing to more accurately evaluate the genetic diversity of *Z. latifolia* (Zou et al., 2024).

### Genetic population structure of wild *Z. latifolia*

The results of this study also showed high genetic differentiation ( $F_{st} = 0.452$ ) among *Z. latifolia* populations in China, suggesting restricted gene flow among these populations. Specifically, we found significant isolation by distance (IBD) through the Mantel test, and STRUCTURE, PCoA, and UPGMA clustering analyses indicated that the wild *Z. latifolia* populations can be divided into north and south genetic groups. The natural habitat of *Z. latifolia* is generally fragmented, forming isolated habitats that limit gene exchange between populations. This aligns with the observation that aquatic plants typically exhibit high genetic differentiation (Barrett, Eckert & Husband, 1993). Even though these habitats may sometimes achieve intermittent connectivity through water regimes, this connectivity does not necessarily coincide with the seed maturity period required for effective gene flow via waterborne seed dispersal. Additionally, whether birds are significant mediators of seed dispersal in *Z. latifolia* remains unclear, and pollen flow is only effective in localized areas. Overall, *Z. latifolia* lacks the ability to disperse gene flow over long distances, resulting in isolation by distance on a large scale (Willson, 1983). Furthermore, the genetic structure of aquatic plants is influenced not only by geographic isolation but also by the connectivity of water systems (Wang et al., 2008; Honnay et al., 2010; Davis et al., 2018). Both Zhao et al. (2019) and Wagutu et al. (2022) have previously reported a north-south genetic differentiation in wild populations of *Z. latifolia*, and our results also support this



conclusion. However, [Wagutu et al. \(2022\)](#) concluded that populations across different latitudinal gradients were more genetically similar within watersheds, whereas genetic subdivision occurred between watersheds in a latitudinal direction. In contrast, [Zhao et al. \(2019\)](#) did not observe latitudinal genetic subdivision, instead pointing out genetic subdivision in an east-west direction within the Yangtze River Basin. To further explore these patterns, we analyzed the genetic structure of *Z. latifolia* populations using STRUCTURE software by applying  $K = 3$  ([Fig. 2](#)). However, we did not find any clear genetic subdivisions associated with the geographic distribution of the populations. Interestingly, in both STRUCTURE and UPGMA analyses, the DL populations located in high northern latitudes were clustered with populations in the south, raising questions about the origin of this population that should be addressed in subsequent work. [Zhao et al. \(2019\)](#) has reported the existence of *Z. latifolia* populations formed by the escape of cultivars into natural habitats. If feral *Z. latifolia* samples were incorrectly collected, this could lead to misclassification of the genetic structure of the populations.

### The designing of a preliminary core collection for wild *Z. latifolia*

Core germplasm is a key component in plant breeding that can increase the effectiveness of utilizing plant genetic resources for genetic improvement and new variety development. The purpose of core germplasm is to maximize the representation of the genetic diversity of the entire species with a minimum number of genetic resources. To accomplishing this goal, the sampling strategy is important, and it determined the results of core germplasm construction. A suitable sampling plan can optimize the preservation of genetic diversity while minimizing genetic redundancy. Of these, stratified sampling is often considered superior to random sampling in many aspects, as demonstrated by numerous studies ([Brown, 1989](#); [Casler, 1995](#); [Escribano, Viruel & Hormaza, 2008](#); [Kumar et al., 2020](#)). Stratified sampling can be implemented using various methods, including proportional, logit, square root, and genetic diversity-based sampling, each with its own advantages and disadvantages. When selecting a sampling strategy, it is important to consider factors such as the amount of germplasm material, grouping, and genetic diversity.

Core germplasm construction involves extracting a portion of the total samples from existing germplasm resources that meet specific criteria. Achieving this goal hinges on selecting appropriate sampling methods and determining suitable sampling ratios. Currently, the commonly used software for constructing core germplasm includes PowerMarker, PowerCore, and CoreHunter. The proportion of core germplasm relative to the original materials in various plants, both domestically and internationally, ranges from 5% to 40% ([Wang et al., 2007](#); [Escribano, Viruel & Hormaza, 2008](#); [Liang et al., 2015](#); [Dervishi et al., 2021](#)). However, none of these studies provided a uniform sampling rate, primarily due to differences in sampling completeness, levels of genetic diversity, and the biological characteristics specific to each plant. In our research, we used PowerMarker v3.25 (SANA and SAGD) and CoreHunter3 (MR) to construct a preliminary core collection of wild *Z. latifolia*. Our results revealed very limited intersection between the core germplasm constructed by different methods using a 10% sampling ratio ([Fig. 6](#)), likely due to differences in the algorithms underlying each method. As a result, we

combined the core germplasm constructed by each of the three methods. Evaluating the representativeness of core samples in relation to the diversity of the entire germplasm resource is also a key focus in core germplasm studies. Referring to previous studies, we first created a principal component map based on the final core germplasm and compared it with the principal component distribution maps of all materials, finding a high degree of agreement (Fig. 4). Additionally, no significant differences were observed between the genetic diversity indices of the final core germplasm and the original germplasm materials (Table 3). The results demonstrated that the final core germplasm was well-matched with the original germplasm, indicating the rationality of the final core germplasm screening results. This core collection will facilitate the utilization of *Z. latifolia* in developing new crop varieties with enhanced yield and resistance traits.

## CONCLUSIONS

The present study addresses the importance of genetic diversity assessment and core collection development for the conservation of *Z. latifolia*, a significant genetic resource for rice improvement. Despite the limited number of molecular markers and relatively insufficient sample size, our findings highlight the impact of habitat loss and fragmentation on the genetic structure of wild populations, with a clear north-south genetic differentiation pattern. The successful design of a preliminary core collection, using various sampling strategies, ensures the efficient representation of genetic diversity and provides a foundation for future breeding and genetic research. Moreover, the insights gained from this study call for immediate conservation actions to protect the remaining wild populations of *Z. latifolia* from further decline due to anthropogenic pressures. Future research should leverage whole-genome resequencing to refine the assessment of genetic diversity, considering the recent availability of a high-quality chromosome-level genome of *Z. latifolia*.

## ACKNOWLEDGEMENTS

We express thanks to Jianli Wu and Huimei Wang providing for critical reading and helpful suggestions on improving our manuscript.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31460378, 31600293, 32260091), Natural Science Foundation of Jiangxi Province (Grant No. 20212BAB205029) and China Agricultural Research System (No. CARS-01-60). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: 31460378, 31600293, 32260091.

Natural Science Foundation of Jiangxi Province: 20212BAB205029.  
China Agricultural Research System: CARS-01-60.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Xiangliang Lei performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xiaona Su performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chengchuan Zhou performed the experiments, prepared figures and/or tables, and approved the final draft.
- Shaolin Jiang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Xiaoquan Yuan performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yao Zhao conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Shaomei Jiang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplemental Files](#) and at Figshare: Zhao, Yao; Lei, Xiangliang (2024). Lei et al. 2024 SSR raw data of *Zizania latifolia*. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.26384773.v1>.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.18909#supplemental-information>.

## REFERENCES

- Bai Q, Cai Y, He B, Liu W, Pan Q, Zhang Q. 2019. Core set construction and association analysis of *Pinus massoniana* from Guangdong province in southern China using SLAF-seq. *Scientific Reports* 9:13157 DOI 10.1038/s41598-019-49737-2.
- Barrett SCH, Eckert CG, Husband BC. 1993. Evolutionary processes in aquatic plant populations. *Aquatic Botany* 44(2–3):105–145 DOI 10.1016/0304-3770(93)90068-8.
- Belaj A, Dominguez-García MD, Atienza SG, Martín Urdíroz N, De la Rosa R, Satovic Z, Martín A, Kilian A, Trujillo I, Valpuesta V, Del Río C. 2012. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genetics and Genomes* 8(2):365–378 DOI 10.1007/s11295-011-0447-6.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13(11):3261–3273 DOI 10.1111/j.1365-294X.2004.02346.x.

- Brown AHD. 1989.** Core collections: a practical approach to genetic resources management. *Genome* **31**(2):818–824 DOI [10.1139/g89-144](https://doi.org/10.1139/g89-144).
- Brush SB, Meng E. 1998.** Farmers' valuation and conservation of crop genetic resources. *Genetic Resources and Crop Evolution* **45**(2):139–150 DOI [10.1023/A:1008650819946](https://doi.org/10.1023/A:1008650819946).
- Casler MD. 1995.** Patterns of variation in a collection of perennial ryegrass accessions. *Crop Science* **35**(4):1169–1177 DOI [10.2135/cropsci1995.0011183x003500040043x](https://doi.org/10.2135/cropsci1995.0011183x003500040043x).
- Chen YY, Chu HJ, Liu H, Liu YL. 2012.** Abundant genetic diversity of the wild rice *Zizania latifolia* in central China revealed by microsatellites. *Annals of Applied Biology* **161**(2):192–201 DOI [10.1111/j.1744-7348.2012.00564.x](https://doi.org/10.1111/j.1744-7348.2012.00564.x).
- Chen YY, Liu Y, Fan X, Li W, Liu Y. 2017.** Landscape-scale genetic structure of wild rice *Zizania latifolia*: the roles of rivers, mountains and fragmentation. *Frontiers in Ecology and Evolution* **5**:17 DOI [10.3389/fevo.2017.00017](https://doi.org/10.3389/fevo.2017.00017).
- Chen SL, Xu KX. 1994.** A study on cladistic taxonomy of *Zizania* L. *Bulletin of Botanical Research* **14**:385–394.
- Choudhary M, Singh V, Muthusamy V, Wani S. 2017.** Harnessing crop wild relatives for crop improvement. *International Journal of Life Sciences* **6**:73 DOI [10.5958/2319-1198.2017.00009.4](https://doi.org/10.5958/2319-1198.2017.00009.4).
- Davis CD, Epps CW, Flitcroft RL, Banks MA. 2018.** Refining and defining riverscape genetics: how rivers influence population genetic structure. *Wiley Interdisciplinary Reviews: Water* **5**(2):e1269 DOI [10.1002/wat2.1269](https://doi.org/10.1002/wat2.1269).
- De Beukelaer H, Davenport GF, Fack V. 2018.** Core Hunter 3: flexible core subset selection. *BMC Bioinformatics* **19**(1):203 DOI [10.1186/s12859-018-2209-z](https://doi.org/10.1186/s12859-018-2209-z).
- Dervishi A, Jakše J, Ismaili H, Javornik B, Štajner N. 2021.** Genetic structure and core collection of olive germplasm from Albania revealed by microsatellite markers. *Genes* **12**(2):256 DOI [10.3390/genes12020256](https://doi.org/10.3390/genes12020256).
- Escribano P, Viruel MA, Hormaza JI. 2008.** Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilised subtropical fruit tree species. *Annals of Applied Biology* **153**(1):25–32 DOI [10.1111/j.1744-7348.2008.00232.x](https://doi.org/10.1111/j.1744-7348.2008.00232.x).
- Evanno G, Regnaut S, Goudet J. 2005.** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**(8):2611–2620 DOI [10.1111/j.1365-294X.2005.02553.x](https://doi.org/10.1111/j.1365-294X.2005.02553.x).
- Food and Agriculture Organization (FAO). 2022.** *World Food and Agriculture–Statistical Yearbook*. Rome: FAO DOI [10.4060/cc2211en](https://doi.org/10.4060/cc2211en).
- Ford Lloyd BV, Schmidt M, Armstrong SJ, Barazani OZ, Engels J, Hadas R, Hammer K, Kell SP, Kang D, Khoshbakht K, Li Y, Long C, Lu B-R, Ma K, Nguyen VT, Qiu L, Ge S, Wei W, Zhang Z, Maxted N. 2011.** Crop wild relatives—undervalued, underutilized and under threat? *BioScience* **61**(7):559–565 DOI [10.1525/bio.2011.61.7.10](https://doi.org/10.1525/bio.2011.61.7.10).
- Frankel OH. 1984.** Plant genetic resources today: a critical appraisal. In: Holden JHW, Williams JT, eds. *Crop Genetic Resources: Conservation and Evaluation*. London: George Allen & Unwin, 249–257.
- Guo HB, Li SM, Peng J, Ke WD. 2007.** *Zizania latifolia* Turcz. cultivated in China. *Genetic Resources and Crop Evolution* **54**(6):1211–1217 DOI [10.1007/s10722-006-9102-8](https://doi.org/10.1007/s10722-006-9102-8).
- Guo L, Qiu J, Han Z, Ye Z, Chen C, Liu C, Xin X, Ye CY, Wang YY, Xie H, Wang Y, Bao J, Tang S, Xu J, Gui Y, Fu F, Wang W, Zhang X, Zhu Q, Guang X, Wang C, Cui H, Cai D, Ge S, Tuskan GA, Yang X, Qian Q, He SY, Wang J, Zhou X-P, Fan L. 2015.** A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. *The Plant Journal* **83**:600–609 DOI [10.1111/tpj.12912](https://doi.org/10.1111/tpj.12912).

- Honnay O, Jacquemyn H, Nackaerts K, Breyne P, Van Looy K. 2010. Patterns of population genetic diversity in riparian and aquatic plant species along rivers. *Journal of Biogeography* 37(9):1730–1739 DOI 10.1111/j.1365-2699.2010.02331.x.
- Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend M. 2008. Ecological consequences of genetic diversity. *Ecology Letters* 11(6):609–623 DOI 10.1111/j.1461-0248.2008.01179.x.
- Isbell F, Adler PR, Eisenhauer N, Fornara D, Kimmel K, Kremen C, Letourneau DK, Liebman M, Polley HW, Quijas S, Scherer-Lorenzen M. 2017. Benefits of increasing plant diversity in sustainable agroecosystems. *Journal of Ecology* 105(4):871–879 DOI 10.1111/1365-2745.12789.
- Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806 DOI 10.1093/bioinformatics/btm233.
- Kashyap A, Garg P, Tanwar K, Sharma J, Gupta NC, Ha PT, Bhattacharya RC, Mason AS, Rao M. 2022. Strategies for utilization of crop wild relatives in plant breeding programs. *Theoretical and Applied Genetics* 135(12):4151–4167 DOI 10.1007/s00122-022-04220-x.
- Kumar A, Kumar S, Singh KB, Prasad M, Thakur JK. 2020. Designing a mini-core collection effectively representing 3004 diverse rice accessions. *Plant Communications* 1(5):100049 DOI 10.1016/j.xplc.2020.100049.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6):1547–1549 DOI 10.1093/molbev/msy096.
- Liang W, Dondini L, De Franceschi P, Paris R, Sansavini S, Tartarini S. 2015. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Molecular Biology Reporter* 33(3):458–473 DOI 10.1007/s11105-014-0754-9.
- Liu B, Liu ZL, Li XW. 1999. Production of a highly asymmetric somatic hybrid between rice and *Zizania latifolia* (Griseb): evidence for inter-genomic exchange. *Theoretical and Applied Genetics* 98(6–7):1099–1103 DOI 10.1007/s001220051173.
- Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129 DOI 10.1093/bioinformatics/bti282.
- Meyer RS, DuVal AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* 196(1):29–48 DOI 10.1111/j.1469-8137.2012.04253.x.
- Miyamoto N, Ono M, Watanabe A. 2015. Construction of a core collection and evaluation of genetic resources for *Cryptomeria japonica* (Japanese cedar). *Journal of Forest Research* 20(1):1–11 DOI 10.1007/s10310-014-0460-3.
- Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537–2539 DOI 10.1093/bioinformatics/bts460.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(3):945–959 DOI 10.1093/genetics/155.2.945.
- Quan Z, Pan L, Ke W, Liu Y, Ding Y. 2009. Sixteen polymorphic microsatellite markers from *Zizania latifolia* Turcz. (Poaceae). *Molecular Ecology Resources* 9(3):887–889 DOI 10.1111/j.1755-0998.2008.02357.x.
- Richards CM, Antolin MF, Reilley A, Poole J, Walters C. 2007. Capturing genetic diversity of wild populations for ex situ conservation: Texas wild rice (*Zizania texana*) as a model. *Genetic Resources and Crop Evolution* 54(4):837–848 DOI 10.1007/s10722-006-9167-4.



- Richards CM, Reilley A, Touchell D, Antolin MF, Walters C. 2004. Microsatellite primers for Texas wild rice (*Zizania texana*), and a preliminary test of the impact of cryogenic storage on allele frequency at these loci. *Conservation Genetics* 5(6):853–859 DOI 10.1007/s10592-004-1977-3.
- Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4(1):137–138 DOI 10.1046/j.1471-8286.2003.00566.x.
- Su W, Wang L, Lei J, Chai S, Liu Y, Yang Y, Yang X, Jiao C. 2017. Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. *PLOS ONE* 12(2):e0172066 DOI 10.1371/journal.pone.0172066.
- Terrell EE, Peterson PM, Reveal JL, Duvall MR. 1997. Taxonomy of north American species of *Zizania* (Poaceae). *SIDA, Contributions to Botany* 1:533–549.
- Wagutu GK, Fan X, Fu W, Tengwer MC, Li W, Chen Y. 2022. Genetic structure of wild rice *Zizania latifolia* in an expansive heterogeneous landscape along a latitudinal gradient. *Frontiers in Ecology and Evolution* 10:929944 DOI 10.3389/fevo.2022.929944.
- Wang JC, Hu J, Xu HM, Zhang S. 2007. A strategy on constructing core collections by least distance stepwise sampling. *Theoretical and Applied Genetics* 115(1):1–8 DOI 10.1007/s00122-007-0533-1.
- Wang Y, Yang T, Wang X, Sun X, Liu H, Wang D, Wang H, Zhang G, Li Y, Wang X, Wei Z. 2023. Develop a preliminary core germplasm with the novel polymorphism EST-SSRs derived from three transcriptomes of colored calla lily (*Zantedeschia hybrida*). *Frontiers in Plant Science* 14:1055881 DOI 10.3389/fpls.2023.1055881.
- Wang MX, Zhang HL, Zhang DL, Qi YW, Fan ZL, Li DY, Pan DJ, Cao YS, Qiu ZE, Yu P, Yang QW, Wang XK, Li ZC. 2008. Genetic structure of *Oryza rufipogon* Griff. in China. *Heredity* 101(6):527–535 DOI 10.1038/hdy.2008.61.
- Willson MF. 1983. *Plant reproductive ecology*. New York, USA: Wiley.
- Xie YN, Qi QQ, Li WH, Li YL, Zhang Y, Wang HM, Zhang YF, Ye ZH, Guo DP, Qian Q, Zhang ZF, Yan N. 2023. Domestication, breeding, omics research, and important genes of *Zizania latifolia* and *Zizania palustris*. *Frontiers in Plant Science* 14:1183739 DOI 10.3389/fpls.2023.1183739.
- Xu XW, Ke WD, Yu XP, Wen J, Ge S. 2008. A preliminary study on population genetic structure and phylogeography of the wild and cultivated *Zizania latifolia* (Poaceae) based on Adh1a sequences. *Theoretical and Applied Genetics* 116(6):835–843 DOI 10.1007/s00122-008-0717-3.
- Xu XW, Walters C, Antolin MF, Alexander ML, Lutz S, Ge S, Wen J. 2010. Phylogeny and biogeography of the eastern Asian-North American disjunct wild-rice genus (*Zizania* L., Poaceae). *Molecular Phylogenetics and Evolution* 55(3):1008–1017 DOI 10.1016/j.ympev.2009.11.018.
- Xu XW, Wu JW, Qi MX, Lu QX, Lee PF, Lutz S, Ge S, Wen J. 2015. Comparative phylogeography of the wild-rice genus *Zizania* (Poaceae) in eastern Asia and North America. *American Journal of Botany* 102(2):239–247 DOI 10.3732/ajb.1400323.
- Yan N, Yang T, Yu XT, Shang LG, Guo DP, Zhang Y, Meng L, Qi QQ, Li YL, Du YM, Liu XM. 2022. Chromosome-level genome assembly of *Zizania latifolia* provides insights into its seed shattering and phytocassane biosynthesis. *Communications Biology* 5(1):36 DOI 10.1038/s42003-021-02993-3.
- Yang G, Ma R, Zhang L, Jiang J, Yao S, Zhang M, Zeng HD. 2010. Lake status, major problems and protection strategy in China. *Journal of Lake Sciences* 22:799–810.

- Zhang H, Mittal N, Leamy LJ, Barazani O, Song BH. 2017.** Back into the wild—apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications* **10**(1):5–24 DOI [10.1111/eva.12434](https://doi.org/10.1111/eva.12434).
- Zhao Y, Song Z, Zhong L, Li Q, Chen J, Rong J. 2019.** Inferring the origin of cultivated *Zizania latifolia*, an aquatic vegetable of a plant-fungus complex in the Yangtze River Basin. *Frontiers in Plant Science* **10**:1406 DOI [10.3389/fpls.2019.01406](https://doi.org/10.3389/fpls.2019.01406).
- Zhao Y, Zhong L, Zhou K, Song Z, Chen J, Rong J. 2018.** Seed characteristic variations and genetic structure of wild *Zizania latifolia* along a latitudinal gradient in China: implications for neo-domestication as a grain crop. *AoB Plants* **10**(6):ply072 DOI [10.1093/aobpla/ply072](https://doi.org/10.1093/aobpla/ply072).
- Zou Y, Yang W, Zhang R, Xu XW. 2024.** Signatures of local adaptation and maladaptation to future climate in wild *Zizania latifolia*. *Communications Biology* **7**(1):1313 DOI [10.1038/s42003-024-07036-1](https://doi.org/10.1038/s42003-024-07036-1).