Genetic relatedness among Ethiopian *Oryza longistaminata* populations and other AA genome *Oryza* species

Melaku Getachew^{1,2} · Liyu Huang¹ · Shilai Zhang¹ · Guangfu Huang¹ · Jing Zhang¹ · Tesfaye Kassahun³ · Haileselassie Teklehaimanot³ · Fengyi Hu¹

Received: 28 September 2019 / Accepted: 27 February 2020 / Published online: 7 March 2020
© The Author(s) 2020

Abstract

Oryza longistaminata is the only AA-genome *Oryza* species that is perennial via rhizome production. This undomesticated rice species, which is native to Africa, is hypothesized to be a good candidate for expanding the cultivated rice gene pool. However, its phylogenetic relationships with other *Oryza* members are still unresolved, and it is underutilized as a genetic resource in the breeding of cultivated rice (*Oryza sativa* L.). This study therefore genotyped 361 *O. longistaminata*, 35 cultivated rice, 1 *Japonica* weedy-type, 25 AA genome and 8 CC genome wild rice accessions by using 67 SSR markers. Genotypic grouping confirmed the distinctness of *O. longistaminata* from other rice accessions and the sub-differentiation of this population influenced by eco-geographical conditions. The higher genetic diversity within the *O. longistaminata* population also implies its candidacy as a donor of diverse traits of interest.

Keywords *Oryza longistaminata* · Rice germplasm · Genetic diversity · Phylogenetic relationship · SSR markers

Melaku Getachew and Liyu Huang have contributed equally to this paper.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10725-020-00597-1>) contains supplementary material, which is available to authorized users.

✉ Fengyi Hu
hfengyi@ynu.edu.cn

Melaku Getachew
getachewmelaku68@yahoo.com

Liyu Huang
lyhuang@ynu.edu.cn

Shilai Zhang
shilaizhang@ynu.edu.cn

Guangfu Huang
gfhuang@ynu.edu.cn

Jing Zhang
zhangjing603608@163.com

Tesfaye Kassahun
kassahuntesfaye@yahoo.com

Introduction

The genus *Oryza* is considered an economically important plant group in the grass family Poaceae (Khush 1997). Rice (*Oryza sativa* L.) feeds over half of the world's human population (Nachimuthu et al. 2015). During domestication, the majority of the alleles of the ancestral progenitors of cultivated rice were lost (Sun et al. 2001). According to Ali et al. (2010), rice productivity is affected by its narrow gene pool. Hence, diverse sources of genes must be explored

Haileselassie Teklehaimanot
tekle1961@yahoo.com

- ¹ State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Research Center for Perennial Rice Engineering and Technology of Yunnan, School of Agriculture, Yunnan University, Kunming 650091, Yunnan, China
- ² Department of Biotechnology, Biological and Chemical Engineering College, Addis Ababa Science and Technology University, 16417 Addis Ababa, Ethiopia
- ³ College of Natural and Computational Sciences, Institute of Biotechnology, Addis Ababa University, 1176 Addis Ababa, Ethiopia

(Joshi et al. 2000; Huang et al. 2019). Wild *Oryza* species can be important resources in rice improvement programmes (Brar and Khush 1997); however, it is often difficult to transfer agronomically important traits to cultivated rice (Ren et al. 2003). Thus, precise genetic manipulation of desirable traits requires the prior assessment of kinship relatedness (Nachimuthu et al. 2015).

Due to the relative ease of interspecific hybridization, species that share the AA genome of *O. sativa* are the most accessible genetic resources (Lu et al. 2000; Ren et al. 2003; Chen et al. 2017; Xu et al. 2020). The AA genome clade includes *Oryza rufipogon*, *Oryza nivara*, *Oryza barthii*, *Oryza glumaepatulata*, *Oryza meridionalis*, *Oryza glaberrima*, *O. longistaminata* and *O. sativa*. The African wild rice variety *O. longistaminata* Chev. et. Rohr is the only member of the AA gene pool that is perennial via rhizome production (Vaughan et al. 2005).

O. longistaminata is strongly rhizomatous (Ghesquiere 1985; Vaughan 1994). In addition to its high rate of self-incompatibility, morphological features such as high pollen production along with long and exerted stigma presumably facilitate outcrossing (Ghesquiere 1987; Jones et al. 1996). This African wild rice is resistant to bacterial blight, drought, and the nematode *Meloidogyne graminicola* (Second et al. 1977; Taillebois 1983; Khush et al. 1990; Warda 1997; Jones et al. 1996). For this reason, *O. longistaminata* was suggested for further utilization and application in conservation schemes (Melaku et al. 2013; Wambugu et al. 2013).

Most DNA marker-based genetic relationship studies of the genus *Oryza* or taxa in the AA gene pool have supported the taxonomic classification within the genus *Oryza* based on morphological traits (Duan et al. 2007). For instance, its self-incompatibility, rhizomatous character, and unique ligular characteristics make the African *O. longistaminata* unique in its AA genome group (Ghesquiere 1985; Vaughan 1994). However, the different marker types used in evolutionary studies of the AA genome *Oryza* are still limited and provide ambiguous and controversial relationships (Ren et al. 2003; Ge et al. 2004). This is likely due to marker types, sampling methods and polytomy in the clade (Duan et al. 2007). Thus, an adequate number of accessions and SSR primer pairs must be utilized to achieve a reliable analysis of genetic structure and an improved resolution over related species.

In the present study, 67 polymorphic SSR markers were used to investigate a population structure of 360 *O. longistaminata* accessions from Ethiopia and their genetic relationships with worldwide collections of 17 *Japonica* varieties, 18 varieties of *Indica*, and 1 *O. longistaminata* from Niger, 10 *O. rufipogon*, 6 *O. nivara*, 6 *O. glumaepatula*, and 3 *O. barthii* accessions along with a CC genome outgroup consisting of 2 *Oryza rhizomatis*, 2 *Oryza eichengeri*, 2 *Oryza*

australiensis, 1 *Oryza officinalis*, 1 *Oryza latifolia* and 1 *Japonica* weedy-type accession.

Materials and methods

Plant material

A total of 430 accessions comprising 35 cultivated rice accessions from the Yunnan University Genebank (Table S1), 25 accessions of 5 AA genomes, 9 accessions from 5 species of the *O. officinalis* complex, one *Japonica* weedy-type accession from the International Rice Research Institute collections (Table S2) and 360 accessions from 12 Ethiopian populations of African wild rice (*O. longistaminata*) (Table S3) were included in this study. Based on their geographic location, mode of pollination and species type, the accessions were complexed and analysed as twenty-six independent groups.

Genomic DNA extraction and polymerase chain reaction (PCR)

Total genomic DNA was extracted from fresh leaves by using the CTAB protocol as described by Doyle and Doyle (1987). The quality of the extracted DNA was determined by electrophoresis in a 1% agarose gel, and quantification was carried out using a spectrophotometer. The extracted DNA samples were diluted to 20 ng/μl using Tris-EDTA (TE) buffer. A total of 67 SSR markers (Table S4) covering 12 rice chromosomes were selected for the analysis based on the reports from Panaud et al. (1996), Chen et al. (1997) and Temnykh et al. (2000).

PCR was performed in a 10 μl reaction mixture containing 4 μl of genomic DNA, 0.5 μl of each of two primers (at a concentration of 10 μM), 1.75 μl of a 10×Taq buffer (with MgCl₂ at a final concentration of 2 mM), 0.5 μl of a 2.5 mM dNTP mixture, 0.1 μl of 2.5 U/μl Taq DNA polymerase and 3 μl of double-distilled water. The PCR amplification protocol consisted of 5 minutes of preheating at 94 °C followed by 34 cycles of denaturation at 94 °C for 45 s, annealing at 55 °C to 67 °C (depending on the primer type) for 30 s, extension at 72 °C for 30 s, and a final extension at 72 °C for 7 min.

Polyacrylamide gel electrophoresis and SSR allele scoring

The amplified PCR products were subjected to electrophoresis in an 8.0% polyacrylamide gel and detected through silver staining as described by Panaud et al. (1996), and their size was determined using a standard 50 bp DNA ladder (M1800, Solarbio). Based on the expected PCR product

size indicated in the GRAMENE marker database (<https://www.gramene.org>), the sizes of the most intensely amplified bands were identified as the alleles of the SSR loci. Differently sized amplified bands were scored as genotypes. The bands were recorded as (11, 22, 33. .) to represent homozygous genotypes or (12, 13, 23.. .) to indicate heterozygous genotypes, and '?' was used to denote missing data (Fig. S1).

SSR data analysis

The number of alleles per locus, expected and observed heterozygosity and F statistics such as genetic differentiation (F_{st}), Wright's fixation index (F_{is}) and the total inbreeding coefficient (F_{it}) were calculated using GenAlEx 6.502 (Peakall and Smouse 2012). The major allelic frequency and polymorphic information content (PIC) of each marker were computed using Power Marker Version 3.25 (Liu and Muse 2005). Analysis of molecular variance (AMOVA) was performed for 426 accessions, excluding the *Japonica* weedy-type accession, *O. latifolia*, *O. officinalis* and *O. longistaminata* from Niger (each of which was represented by only one accession). Both AMOVA and principal component analysis (PCA) were also conducted in GenAlEx 6.502.

Based on their computed pairwise genetic distance values, a dendrogram was drawn for the 360 *O. longistaminata* accessions by using DARwin V6 software (Perrier and Jacquemoud-Collet 2006). The unweighted pair group method with arithmetic mean (UPGMA) was applied to the 12 *O. longistaminata* populations using Poptree2 software (<https://www.ualberta.ca/~fyeh/fyeh>).

The population structure was inferred via the STRUCTURE HARVESTER v. 2.3 model (Pritchard et al. 2000). Three replications were run with a burn-in period of 150,000 and 150,000 repeats of the MCMC method for K values of 1 to 26. The optimal number of subpopulations (K) was determined via the ΔK approach as described by Evanno et al. (2005) using STRUCTURE HARVESTER v0.6.8. (Earl and von Holdt 2011).

Results

SSR genotyping

From the 430 wild, cultivated and weedy rice accessions, a total of 440 alleles, with a range of 2 (RM22 and RM171) to 19 (RM225) alleles per locus (average 6.57), were amplified (Table S5). The major allele frequencies over the 67 SSR loci varied from 0.337 (RM1) to 0.953 (RM60), with an overall mean of 0.656. Generally, from all SSR markers used in this study, 663 genotypes were inferred, with an average of 9.89 per locus.

As shown in Table 1, 42 heterozygous loci of the assayed SSR markers presented observed heterozygosity (H_o) and expected heterozygosity (H_e) values ranging from 0.04 to 0.52 and from 0.32 to 0.76, respectively. Based on their polymorphic information content (PIC), 26 (38.9%) of the SSR markers used in this study were highly informative ($PIC > 0.5$).

Extent of genetic diversity

At the group level, genetic parameters such as the observed number of alleles (N_a), effective number of alleles (N_e), mean observed and expected proportions of heterozygous loci and Wright's fixation index (F_{is}) were calculated (Table 1). In the 26 assessed groups, the actual number of alleles (N_a) ranged from the lowest value of 0.72 (*O. longistaminata* accession from Niger) to the highest of 2.485 in both the *Indica* and *Japonica* subspecies of cultivated rice. The effective number of alleles (N_e) ranged from 0.72 (*O. longistaminata* from Niger) to 1.991 (*O. glumaepatula*). Expected heterozygosity (H_e) ranged from 0.015 for *O. officinalis* to 0.421 for *O. glumaepatula*. When all groups were pooled together, 1.864 observed alleles, 1.448 effective alleles, an observed heterozygosity of 0.126 and an expected heterozygosity of 0.239 were obtained. With the exception of the outgroups, all populations showed lower observed heterozygosity than expected heterozygosity. The relatively high Wright's fixation index (F_{is}) in the domesticated versus wild AA genome *Oryza* species implied significant deviation from Hardy-Weinberg expectations.

Genetic differentiation and variance analysis

For all 67 SSR loci, $F_{st} = 0.347$, $F_{is} = 0.603$ and $F_{it} = 0.741$ (Table 2). All three F statistic values were highly significant ($P > 0.001$). In particular, the highly significant F_{st} value showed higher divergence or differentiation among the assessed groups. The indirect estimation of gene flow using F_{st} showed $N_m = 0.471$ (Table 2). Based on the twenty-two study groups, AMOVA showed 35% variance among groups, 39% variance among accessions and 26% variance within accessions (Table 3).

Principal component analysis

A scatter plot of the first two principal components of all 430 accessions showed five clusters (Fig. 1). The first and second axes accounted for 8.96% and 7.26% of the molecular variance, respectively. When only the 12 groups of *O. longistaminata* from Ethiopia were considered, a clear geographic pattern emerged. A total of 180 *O. longistaminata* accessions of the six Gambella groups were located in the right half of Quadrant I and left half of Quadrant II. However,

Table 1 Group-based genetic diversity parameters

Populations	Na	Ne	Ho	He	F
Gambella 1 (GAM1)	2.088	1.521	0.139	0.267	0.457
Gambella 2 (GAM2)	2.309	1.575	0.160	0.279	0.447
Gambella 3 (GAM3)	2.088	1.518	0.150	0.251	0.424
Gambella 4 (GAM4)	2.162	1.621	0.119	0.271	0.545
Gambella 5 (GAM5)	1.897	1.53	0.132	0.199	0.309
Gambella 6 (GAM6)	2.059	1.505	0.161	0.255	0.357
Amhara 1 (AMH1)	2.206	1.525	0.150	0.274	0.381
Amhara 2 (AMH2)	2.294	1.518	0.145	0.271	0.427
Amhara 3 (AMH3)	2.132	1.465	0.167	0.259	0.402
Amhara 4 (AMH4)	2.221	1.438	0.171	0.236	0.371
Amhara 5 (AMH5)	2.029	1.383	0.132	0.216	0.395
Amhara 6 (AMH6)	1.574	1.282	0.135	0.150	0.185
<i>O. barthi</i> (BART)	1.809	1.659	0.108	0.305	0.601
<i>O. glumaepatula</i> (GLUM)	2.426	1.991	0.095	0.421	0.753
<i>O. officinalis</i> (OFFI)	0.750	0.750	0.029	0.015	-1.000
<i>O. nivara</i> (NIVA)	2.397	1.869	0.107	0.402	0.720
<i>O. rufipogon</i> (RUF)	2.368	1.892	0.110	0.400	0.698
Open-pollinated <i>O. rufipogon</i> (O-RUF)	2.059	1.771	0.119	0.378	0.688
<i>Japonica</i>	2.485	1.659	0.045	0.329	0.873
<i>Indica</i>	2.485	1.801	0.054	0.385	0.845
<i>Japonica</i> weedy type (<i>Japonica</i> -w)	1.044	1.044	0.074	0.037	-1.000
<i>O. rhizomatis</i> (RHYZ)	1.426	1.374	0.103	0.217	0.526
<i>O. eichengeri</i> (EICH)	1.147	1.135	0.162	0.107	-0.511
<i>O. latifolia</i> (LATI)	1.162	1.162	0.279	0.140	-1.000
<i>O. australiansis</i> (AUST)	1.132	1.116	0.154	0.108	-0.431
<i>O. longistaminata</i> (LONG)	0.721	0.721	0.088	0.044	-1.000
Mean	1.86	1.45	0.13	0.24	0.47

Na actual number of alleles, Ne effective number of alleles, Ho observed heterozygosity, He expected heterozygosity, F fixation index

Table 2 F statistic and probability P (rand $\geq$ data)-based genetic differentiation of the *Oryza* gene pool (where F_{st} =genetic differentiation, F_{is} =Wright's fixation index, F_{it} =total inbreeding coefficient and Nm =gene flow)

F-Statistics	Value	P
F_{st}	0.347	0.001
F_{is}	0.603	0.001
F_{it}	0.741	0.001
Nm	0.471	

Table 3 Analysis of molecular variance (AMOVA) of 22 wild and cultivated *Oryza* groups at three levels

Source	Df	SS	MS	Est. Var.	%
Among groups	21	5040.827	240.039	5.853	35
Among accession	404	7145.324	17.686	6.655	39
Within accession	425	1864	4.376	4.376	26
Total	850	14,050.15		16.884	100

df degrees of freedom, SS sum of squares, MS mean squares, Est. Var. estimated variance and % = proportion of genetic variability

the *O. longistaminata* accessions of the six Amhara groups resided in Quadrant III. Although very few Amhara1 and Amhara2 accessions were positioned at the very bottom of Quadrant II, all 30 accessions from Amhara6 formed their own distinct cluster near the middle of Quadrant III. PCA also illustrated the distinction between the two rice subspecies. Almost all *Indica* varieties along with both self- and open-pollinated *O. rufipogon* varieties were located between Quadrant I and Quadrant IV. However, all of the *Japonica* varieties were found in the centre of Quadrant IV, and the space between the *Indica* and *Japonica* varieties was occupied by the relatively overlapping outgroups and some other wild AA genome *Oryza* species.

Cluster analysis

Clustering analysis based on the UPGMA and 1000 bootstrap replicates separated the 26 *Oryza* groups into three main clusters, consisting of the *Japonica* weedy-type

Fig. 1 Two-dimensional scaling of principal component analysis of 26 rice groups inferred from 67 SSR marker genotypic data

accession in Cluster I, AA genome *Oryza* species in Cluster II and the *O. officinalis* complex members in Cluster III (Fig. 2). The members of the *O. sativa* species complex in Cluster II were further partitioned as subcluster A and subcluster B. In the *O. longistaminata* group (subcluster A), the accession from Niger was separated from the 12 Ethiopian *O. longistaminata* groups. However, the Ethiopian groups were subjected to further partitioning on the basis of their geographic provenance. Subcluster-B (AA genome) separated both the *Indica* and *Japonica* subspecies of cultivated rice from the other wild *Oryza* types. In this group, the other African wild species, *O. barthi*, appeared distant from the

Asian and American wild AA genome species. In Cluster III, *O. officinalis* was distinct from the rest of the CC genome *Oryza* species. Surprisingly, the EE genome of *O. australiensis* presented the highest similarity to *O. eichengeri*.

Unlike the dendrogram generated by UPGMA analysis, the hierarchical clustering of the 430 accessions showed 2 main clusters categorizing all *O. longistaminata* accessions in one and the remaining *Oryza* members in the other group (Fig. 3). As subcluster A of Fig. 2, in the main cluster of *O. longistaminata* accessions, the Amhara groups were subsequently partitioned from their Gambella counterparts. In the other main cluster, the outgroups widely deviated from

Fig. 2 UPGMA tree of 26 *Oryza* groups based on Dst-corrected genetic distance

Fig. 3 Dendrogram based on Ward's minimum variance for 430 *Oryza* accessions

the other accession. Further splitting within this group indicated the distinction of *Japonica* types and a higher association of *Indica* with the wild AA genome group. Regarding population admixture, almost all accessions resided in their specific populations. However, admixture was observed for some accessions of Gambella6 with Gambella5, Gambella2 with Gambella4 and Amhara2 with Amhara1. Most interestingly, no *O. longistaminata* accession were found to exhibit admixture in any other cluster (Fig. 3).

Population structure

STRUCTURE analysis suggested that the optimal number of subpopulations for all individuals in the study was three (Fig. 4A). At $K=3$, all Gambella *O. longistaminata* groups comprised one cluster, all Amhara *O. longistaminata* groups constituted the second, and all the wild and cultivated AA genome *Oryza* species along with the outgroups were clustered into a third group (Fig. 4B). In the third cluster, no distinction was detected between the CC genome outgroup and the AA genome *Oryza* species.

Discussion

SSR genotyping and genetic diversity

In this study, a total of 440 alleles, showing an average of 6.6 alleles per locus, were generated. Using an equivalent number of SSR markers and 192 cultivated rice germplasm lines, Nachimuthu et al. (2015) reported 3 alleles per locus.

According to Brar and Khush (2003), domestication practices have caused significant genetic diversity loss of approximately 60%, while *O. longistaminata* collections from a wide agro-ecological area of Ethiopia show high allelic diversity (Melaku et al. 2013). Therefore, the higher number of alleles identified in this study could be attributed to the higher number of studied accessions of *O. longistaminata*. On average, a gene diversity value of 0.46 and observed heterozygosity of 0.14 were obtained from the assessed 67 SSR primers (Table 1). The differences from other studies may be caused by differences in the sampling schemes, the number of SSR markers used, the repeat size of the SSR markers and their positions across the genome, which complicates the comparison of such genetic diversity indices (Deu et al. 2008; Chen et al. 2018).

Genetic differentiation and partitioning of genetic variation

For all 67 SSR loci, highly significant F statistic values ($F_{st} = 0.347$, $F_{is} = 0.603$ and $F_{it} = 0.741$) are reported (Table 2). A higher F_{st} value or broader differentiation among the rice groups was expected, as genetic differentiation is not evenly distributed across the sampled area (Cao et al. 2006), and the assessed rice accessions represented different genome groups distributed all around the world (Tables 1, 2, 3). In contrast, AMOVA implicated a higher component of variance (39%) among the accessions (Table 3). Perhaps this is due to the representation of most outgroups by very few accessions. With the exception of these outgroups, all AA genome groups showed lower

Fig. 4 Delta K analysis and STRUCTURE bar graph of 26 *Oryza* groups. **A** The modal value (Delta K). **B** STRUCTURE bar graph of 26 *Oryza* groups at $k=3$ (the highest peak)

observed than expected heterozygosity (Table 1). However, the computed Wright's fixation index ($F_{is} = 0.603$) was lower than the $F_{is} = 0.75$ obtained for 360 *O. longistaminata* accessions from Ethiopia (Melaku et al., 2013) and the $F_{is}=0.958$ obtained for 900 cultivated, wild and weedy rice individuals (Cao et al. 2006). Such a low fixation index for both wild and cultivated AA genome *Oryza* species indicated relatively less significant deviation from Hardy-Weinberg expectations.

Population structure and genetic relationships

The two-dimensional PCA explained 8.96% and 7.26% of the total genetic variation. A similar pattern of molecular variance for the first two principal coordinates was explained by Zhang et al. (2011). To the right of coordinate 1, a wide-open space separated all *O. longistaminata* accessions from the rest of the rice species (Fig. 1). Furthermore, almost all of the Ethiopian *O. longistaminata* populations were clustered on the basis of their geographic locations. Near the left corner of Quadrant IV, only 3 accessions from Gambella showed relative proximity to the *O. longistaminata* accession from Niger. This distribution highlights the genetic distinction of the Ethiopian *O. longistaminata* accessions.

The UPGMA clustering analysis merged *O. longistaminata* populations with their AA genome wild relatives in one main cluster and separated the CC genome *Oryza* species and the weedy rice in two independent clusters (Fig. 2). However, the hierarchical clustering of all accessions

showed 2 main groups, with all *O. longistaminata* accessions sorted into one group and the remaining *Oryza* members in the other group (Fig. 3). The slight clustering difference between Figs. 2 and 3 could be associated with the generation of more balanced and fewer clusters by using Ward's minimum variance compared with UPGMA (Odong et al. 2011). However, all main clusters, including that of *O. longistaminata*, were subjected to further division on the basis of their respective geographic locations and genome types. As shown in Fig. 3, very few *O. longistaminata* accessions were admixed.

All distance-based clustering methods, with minor deviations, showed a generally similar trend. The model-based structure analysis was also in accordance with the principal coordinate analysis, the clustering pattern of WARD minimum variance and the UPGMA analysis. In this study, population structure results showed three main clusters resulting from the sorting of weedy rice, AA genome and CC genome *Oryza* members in one cluster and the Amhara and Gambella *O. longistaminata* groups from Ethiopia in two independent clusters (Fig. 4B). Interestingly, the CC genome *Oryza* species and the AA genome wild and cultivated rice populations were structured together. The lack of distinction in this cluster was, however, not obvious because of the genetic similarity of its members. Rather, the inclusion of fewer accessions in the outgroups made their separation less distinct.

From the STRUCTURE analysis, it was very clear that the Amhara and Gambella *O. longistaminata* groups of

Ethiopia represent two different gene pools (Fig. 4B). The distinct population structuring between the two regions may be a function of both geographical and climatic factors. The impact of climatic differences can be illustrated by the 1990–2007 G.C. climatic data, which indicate a 25.7 °C mean temperature and 896 mm annual rainfall for the Gambella region and a 19.2 °C mean temperature and 1468 mm annual rainfall for the Amhara region (Melaku 2011). Differences between the populations of the two ecogeographic regions may also be associated with diverse modes of adaptation to variable altitudinal levels, soil types and ecological conditions (Nachimuthu et al. 2015).

Conclusion

Sixty-seven microsatellite markers were used to assess the genetic relatedness among Ethiopian *O. longistaminata* accessions and other AA genome wild and cultivated rice species. The genotypic grouping through structure analysis, distance-based clustering and principal component analysis equally confirmed the distinctness of *O. longistaminata* from other AA genome *Oryza* species. The further subgrouping of *O. longistaminata* is in accord with geographical proximity.

Acknowledgements The research was supported financially by the National Natural Science Foundation of China (U1302264 to FYH), the Department of Sciences and Technology of Yunnan Province (2016BB001 to FYH), the National Key Research and Development Plan of China (2016YFD0300508 to FYH), and the Yunling Scholar Talent Project (to FYH).

Authors contributions FYH, TK and HT designed the experiment. MG and LYH prepared the draft of the manuscript. MG, SLZ, GFH and JZ performed the experiments and analysed the data. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ali ML, Sanchez PL, Yu S, Lorieux M, Eizenga CG (2010) Chromosome segment substitution lines: a powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). *Rice* 3:218–234. <https://doi.org/10.1007/s12284-010-9058-3>
- Brar DS, Khush GS (1997) Alien gene introgression in rice. *Plant Mol Biol* 35:35–47. <https://doi.org/10.1023/a:1005825519998>
- Brar DS, Khush GS (2003) Utilization of wild species of genus *Oryza* in rice improvement. In: Nanda JS, Sharma SD (eds) Monograph on genus *Oryza*. Science Enfield, London, pp 283–309
- Cao Q, Lu B, Xia H, Rong J, Sala F, Spada A, Grassi F (2006) Genetic Diversity and Origin of Weedy Rice (*Oryza sativa* f. *spontanea*) Populations Found in North-eastern China Revealed by Simple Sequence Repeat (SSR) Markers. *Ann Bot* 98:1241–1252. <https://doi.org/10.1093/aob/mcl210>
- Chen H, Wu Y, He J, Guan C, Li A, Fang N, He W, Wang R, Wang J, Bao Y, Zhang H (2017) Population structure analysis and association mapping of bacterial blight resistance in indica rice (*Oryza sativa* L.) accessions. *Plant Growth Regul* 82:21–35. <https://doi.org/10.1007/s10725-016-0234-3>
- Chen P, Hu H, Zhang Y, Wang Z, Dong G, Cui Y, Qian Q, Ren D, Guo LB (2018) Genetic analysis and fine-mapping of a new rice mutant, *white and lesion mimic leaf1*. *Plant Growth Regul* 85:425–435. <https://doi.org/10.1007/s10725-018-0403-7>
- Chen X, Temnykh S, Xu Y, Cho GY, McCouch RS (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor Appl Genet* 95:553–567. <https://doi.org/10.1007/s001220050596>
- Deu M, Sagnard F, Chanterreau J, Calatayud C, He´rault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore´ PS, Mamadou A, Ge´rard B, Ndjeunga J, Bezanc, on G (2008) Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. *Theor Appl Genet* 116:903–913. <https://doi.org/10.1007/s00122-008-0721-7>
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Duan S, Lu B, Li Z, Tong J, Kong J, Yao W, Li S, Zhu Y (2007) Phylogenetic analysis of AA-genome *Oryza* Species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochem Genet* 45:113–129. <https://doi.org/10.1007/s10528-006-9062-x>
- Earl DA, von Holdt BM (2011) STRUCTURE HARVESTER: a web-site and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Ge S, Guo Y, Zhu Q (2004) Molecular phylogeny and divergence of the rice tribe *Oryzaeae*, with special reference to the origin of the genus *Oryza* In: Session 1 The genus *Oryza*, its diversity, and its evolution. pp 40–44
- Ghesquiere A (1985) Evolution of *Oryza longistaminata*. In: International Rice Research Institute (IRRI) Rice Genetics Los Banos, Philippines pp 15–27
- Ghesquiere A (1987) Evolution of *Oryza longistaminata*. ORSTOM, Paris
- Huang X, Feng J, Wang R, Zhang H, Huang J (2019) Comparative analysis of microRNAs and their targets in the roots of two cultivars with contrasting salt tolerance in rice (*Oryza sativa* L.). *Plant Growth Regul* 87:139–148. <https://doi.org/10.1007/s10725-018-0459-4>

- Jones MP, Dingkuhn M, Johnson DE, Fagade SO (1996) Inter-specific hybridization: progress and prospects. In: Proceedings of Workshop on: Africa/Asia Joint Research on Inter-specific Hybridization between the African and Asian Rice Species (*O. glaberrima* and *O. sativa*), West Africa Rice Development Association, Bouake, Cote D'Ivoire, pp 44–49
- Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, Brar DS (2000) Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theor Appl Genet 100:1311–1320. <https://doi.org/10.1007/s001220051440>
- Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34. <https://doi.org/10.1023/a:1005810616885>
- Khush GS, Balacangco E, Ogawa T (1990) A new gene for resistance to bacterial blight from *Oryza longistaminata*. Rice Genet Newsl 7:121–122
- Liu K, Muse SV (2005) PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129
- Lu BR, Naredo ME, Juliano AB, Jackson MT (2000) Preliminary studies on taxonomy and biosystematics of the AA-genome *Oryza* species (Poaceae). In: Jacobs SWL, Enerett J (eds) Grasses Syst Evol. CSIRO, Melbourne, pp 51–58
- Melaku G (2011) Genetic diversity of an African wild rice (*Oryza longistaminata* Chev. et Roehr) from Amhara and Gambella regions as revealed by microsatellite markers. MSc. Thesis, Addis Ababa University, Ethiopia
- Melaku G, Haileselassie T, Feyissa T, Kiboi S (2013) Genetic diversity of the African wild rice (*Oryza longistaminata* Chev. et Roehr) from Ethiopia as revealed by SSR markers. Genet Resour Crop Evol 60:1047–1056. <https://doi.org/10.1007/s10722-012-9900-0>
- Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian AB, Ponniah G, Gunasekaran K, Swaminathan M, Suji KK, Sabariappan R (2015) Analysis of population structure and genetic diversity in rice germplasm using SSR Markers: an initiative towards association mapping of agronomic traits in *Oryza Sativa*. Rice 8:1–24. <https://doi.org/10.1186/s12284-015-0062-5>
- Odong TL, Van Heerwaarden J, Jansen J, Van Hintum TL, Van Eeuwijk FA (2011) Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? Theor Appl Genet 123:195–205. <https://doi.org/10.1007/s00122-011-1576-x>
- Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607. <https://doi.org/10.1007/BF02172406>
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier X, Jacquemoud-Collet JP (2006) Darwin software. <https://www.darwin.cirad.fr/darwin>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
- Ren F, Lu B, Li S, Huang J, Zhu Y (2003) A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. Theor Appl Genet 108:113–120. <https://doi.org/10.1007/s00122-003-1414-x>
- Second G, Bezançon G, Bozza J (1977) Variability of *Oryza longistaminata* and the *Oryza sativa* complex in Africa. Ecological and evolutionary aspects. In: Meeting on African rice species. January 25–26, ORSTOM, Paris, pp 47–55
- Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N (2001) Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. Theor Appl Genet 102:157–162. <https://doi.org/10.1007/s001220051631>
- Taillebois J (1983) A new perspective for the production of hybrid seeds F1: the transfer of the allogamy characters of the species *O. longistaminata* A. Chev. At the species *O. sativa* L. Trop Agron 38:303–307
- Temnykh S, Park WD, Ayers N, Cartinhouer S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712. <https://doi.org/10.1007/s001220051342>
- Vaughan DA (1994) The wild relatives of Rice. In: International Rice Research Institute (IRRI), A genetic resources handbook. Manila Philippines 46–47. <https://doi.org/10.2307/4255858>
- Vaughan DA, Morishima H, Kadowaki K (2005) Diversity in the genus *Oryza*. Curr Opin Plant Biol 6:139–146. [https://doi.org/10.1016/S1369-5266\(03\)00009-8](https://doi.org/10.1016/S1369-5266(03)00009-8)
- Wambugu PW, Furtado A, Waters DE, Nyamongo DO, Henry RJ (2013) Conservation and utilization of African *Oryza* genetic resources. Rice 6:1–13. <https://doi.org/10.1186/1939-8433-6-29>
- Warda (1997) Inter-specific hybridization: progress and prospects. In: West Africa Rice Development Association, Proceedings of Workshop on: Africa/Asia Joint Research on Inter-specific Hybridization between the African and Asian Rice Species (*O. glaberrima* and *O. sativa*), Bouake, Cote D'Ivoire
- Xu X, Zhang M, Xu Q, Feng Y, Yuan X, Yu H, Wang Y, Wei X, Yang Y (2020) Quantitative trait loci identification and genetic diversity analysis of panicle structure and grain shape in rice. Plant Growth Regul 90:89–100. <https://doi.org/10.1007/s10725-019-00549-4>
- Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y (2011) Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. PLoS ONE 6(12):e27565. <https://doi.org/10.1371/journal.pone.0027565>