



# Genome-Wide SNPs Provide Insights on the Cryptic Genetic Structure and Signatures of Climate Adaption in *Amorphophallus albus* Germplasms

Yong Gao<sup>†</sup>, Si Yin<sup>†</sup>, Honglong Chu, Yanan Zhang, Haibo Wang, Huanhuan Chen, Chao Liu, Dongqin Dai\* and Lizhou Tang\*

College of Biological Resource and Food Engineering, Center for Yunnan Plateau Biological Resources Protection and Utilization, Qujing Normal University, Qujing, China

## OPEN ACCESS

### Edited by:

Antonio Gonzalez-Rodriguez,  
Universidad Nacional Autónoma  
de México, Mexico

### Reviewed by:

Daniel Blaine Marchant,  
Stanford University, United States  
Alejandro Casas,  
Universidad Nacional Autónoma  
de México, Mexico

### \*Correspondence:

Lizhou Tang  
124472623@qq.com  
Dongqin Dai  
cicidaidongqin@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 21 March 2021

**Accepted:** 05 July 2021

**Published:** 23 July 2021

### Citation:

Gao Y, Yin S, Chu H, Zhang Y,  
Wang H, Chen H, Liu C, Dai D and  
Tang L (2021) Genome-Wide SNPs  
Provide Insights on the Cryptic  
Genetic Structure and Signatures  
of Climate Adaption  
in *Amorphophallus albus*  
Germplasms.  
*Front. Plant Sci.* 12:683422.  
doi: 10.3389/fpls.2021.683422

Domesticated species represent unique systems in which the evolutionary genomic consequences of intensive selective breeding and adaptation can be thoroughly investigated. *Amorphophallus albus* occurs naturally and is in cultivation throughout the downstream region of the Jinshajiang River in Southwest China. This species is characterised by high konjac glucomannan content, and has been cultivated in China for nearly 2,000 years. To study genetic differentiation and local adaptation of *A. albus*, we sampled 13 distinct local cultivated populations of this species. Restriction site-associated DNA sequencing was conducted with 87 samples, resulting in 24,225 SNPs. The population structure analyses suggest two main genetic groups: one in the relatively upstream region, and one downstream. We found evidence of additional sub-structure within the upstream group, demonstrating the statistical power of genomic SNPs in discovering subtle genetic structure. The environmental and geographic factors were all identified as significant in shaping the genetic differentiation of this species. Notably, the proportion of environmental factors was larger than geographic factors in influencing the population genetic patterns of *A. albus*. We also discovered loci that were associated with local adaptation. These findings will help us understand the genetic differentiation of this newly domesticated species, thereby informing future breeding programs of *A. albus*.

**Keywords:** *Amorphophallus albus*, restriction site-associated DNA sequencing, genetic structure, isolation by distance, isolation by environment, environment adaption

## INTRODUCTION

Understanding the genetic basis of adaptation has been a long standing goal of evolutionary biology (Barton and Keightley, 2002; Grillo et al., 2016). Domesticated species are excellent biological models for studying the genetic basis of rapid evolutionary response to abiotic selective pressure such as climate (Flori et al., 2019). For example, cultivated crops may face strong selective pressure during the domestication process, resulting in population genomics changes. Therefore, the population genomic resources from crops are extremely valuable in understanding the genomic patterns of microevolution under natural and artificial selection.

*Amorphophallus* species are characterised by a single leaf and an underground stem (Figure 1A), and are important sources of konjac glucomannan (KGM) (Chua et al., 2010; Yin et al., 2019). Early investigations show that KGM is a multi-purpose material, with uses as a gelling agent, thickener, film former, and emulsifier (Nishinari et al., 1992; Hu et al., 2006). KGM has also been reported to have the ability to control blood cholesterol and sugar levels, help with weight loss and promote intestinal activity (Zhang et al., 2005; Hu et al., 2006). In the tropical and subtropical regions of Asia, the diverse species of *Amorphophallus* have been historically used as a food source as well as in traditional medicine (Gille et al., 2011). *Amorphophallus* species have been cultivated in China for nearly 2,000 years. *A. albus* is endemic to China, with a native distribution along the upstream region of the Yangtze River (the Jinshajiang River) (Li et al., 2010). *A. albus* has been cultivated by the farmers in this region, where it is harvested for its high KGM content (Zhang et al., 1999). The cultivation of this plant is so recent that no comprehensive breeding programs have been conducted; instead, local farmers simply collected tubers from wild populations and have grown them in their home gardens for many generations (Yin et al., 2019).

In cultivation, *A. albus* is an asexually reproduced crop that can be harvested 2–3 years after planting. *A. albus* can propagate several cormels in one growth cycle, and farmers usually harvest swollen corms and leave cormels in the soil as the “seed” for next year (Tang et al., 2020). Compared to other crops (e.g., rice, wheat) with a domestication history of nearly 10,000 years, the domestication of *A. albus* is relatively short (less than 2,000 years) (Liu, 2003; Zhou et al., 2017). Although *A. albus* has a narrow range along the downstream regions of the Jinshajiang River, heterogeneous micro-environments exist in this area due to the diverse geographic features of the Yunnan–Guizhou Plateau. Considering its short domestication time and limited distribution, *A. albus* offers an excellent model for evaluating genetic differentiation and local adaptations for crops in the early stages of domestication.

The agricultural domestication process, which allows for the crop’s widespread cultivation, often requires substantial adaptation, including for traits related growth, water use, and climate (Aslan et al., 2015). A deep understanding of the genetic response to abiotic selective pressure and of the genes involved in adaptation to climate is essential for broadening the genetic base of the cultivar (Singh, 2001). After years of natural selection and domestication, *A. albus* has formed stable variations in different cultivation areas, leaving local germplasm with specific leaf colors, leaf surface spots, and content of the KGM (Srzednicki and Borompichaichartkul, 2020). Our previous population genetics study, based on 17 simple sequence repeat (SSR) markers, suggested that *A. albus* germplasms should be divided into two genetic groups corresponding to geographic distribution (Yin et al., 2019). However, the study relied on only a few SSRs, which lacked the power to infer fine-scale genetic patterns and functional genomic adaptation in *A. albus*. Besides geographic factors, what role has the heterogeneous climate played in shaping the differentiation of *A. albus* germplasms? Next generation sequencing (NGS) has revolutionised genetic

studies of population differentiation and adaptive selection, making it possible to study these phenomena in a wide range of species (Henry, 2012). In order to further investigate population differentiation and its potential environmental drivers in *A. albus*, we used restriction site-associated DNA sequencing (RAD-seq) to generate genome-wide SNPs and conducted population genetics analyses as well as genome-environment associations (GEAs) of *A. albus*. The goal of our study was to: (1) describe the genome-wide genetic diversity pattern in *A. albus*; (2) evaluate the relative effects of geographic and environmental factors in shaping genetic differentiation of *A. albus* populations; (3) find genomic regions which might be associated with local adaptation in *A. albus*. These results will shed light on the domestication history and patterns of selection during the early stages of domestication in *A. albus*, which could accelerate crop improvement in the future breeding programs.

## MATERIALS AND METHODS

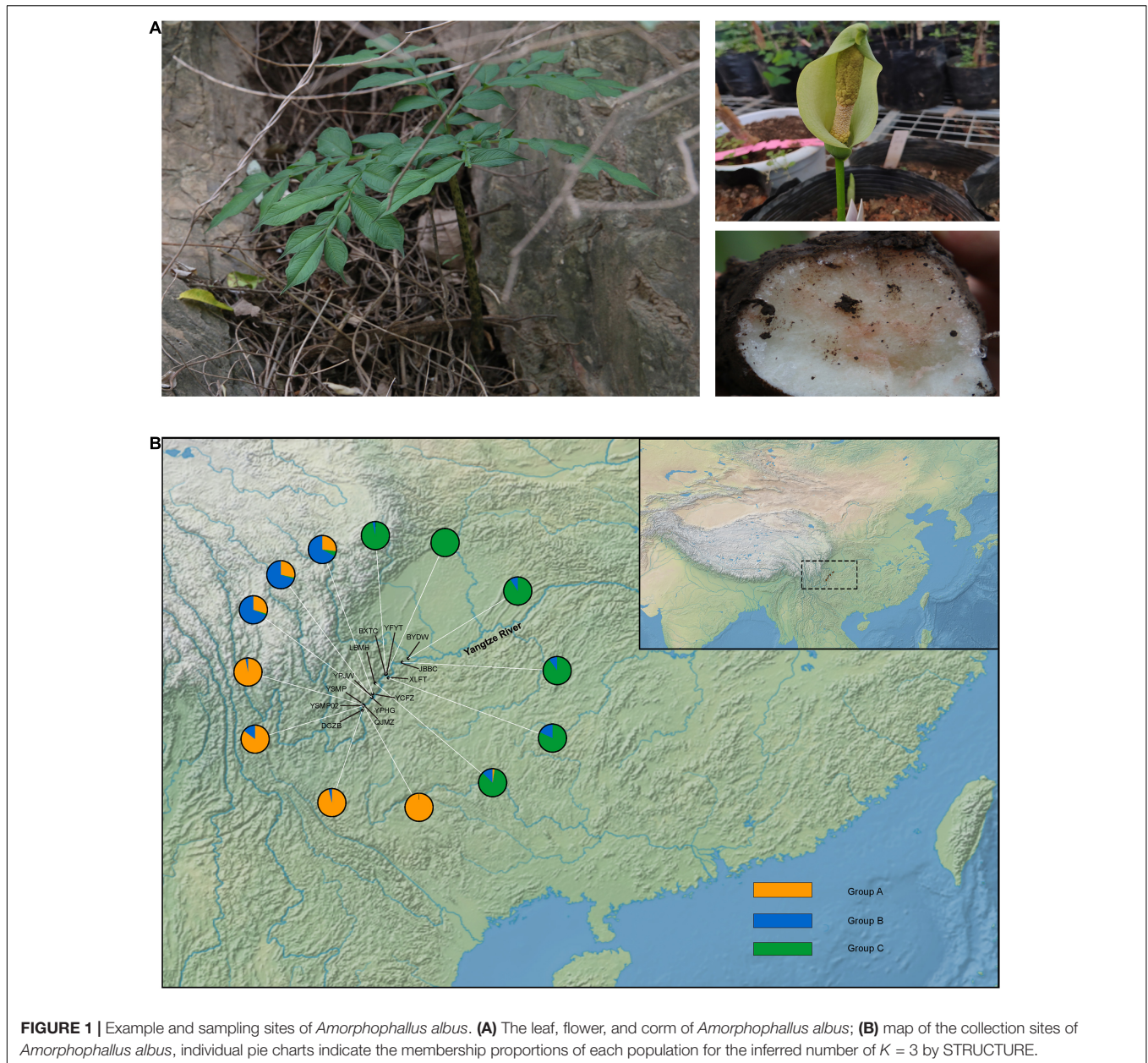
### DNA Sampling and Genotyping

A total of 87 individuals from 13 local cultivated populations of *A. albus* landrace germplasms (2–8 individuals per population) were collected during 2016 and 2017 (Figure 1 and Table 1). Each individual was randomly selected from sampling sites, and the genomic DNA was extracted from the leaves using a commercial DNA isolation kit (Tiangen, Beijing, and China). DNA samples were digested with EcoRI enzyme, randomly fragmented and prepared as multiplexed RAD libraries following the established methods (Baird et al., 2008). Samples were pair-end sequenced with a read length of 150 bp on a NovaSeq platform (Illumina, San Diego, CA, United States).

Prior to assembly, the raw data of each sample underwent quality control by removing adapters and paired reads with alternative reads containing more than 50% low-quality bases ( $Q$ -value  $\leq 5$ ) or 10% unidentified nucleotides (Ns). Stacks v2.3b was then used to assemble pair-ended reads and identify SNPs (Rochette and Catchen, 2017). We set the minimum depth of coverage to create a stack at three sequences and the maximum number of different nucleotides allowed between catalog loci as two nucleotides. The SNP loci were filtered with the “populations” software implemented in Stacks v2.3b (Rochette and Catchen, 2017). To obtain high-quality SNPs, we removed loci with a global minor allele frequency (MAF) less than 0.05 or that were missing in more than 15% of total samples. We retained only loci that were present in at least 80% of the individuals from the population in at least 11 populations. Only the first SNP per locus was retained to avoid linkage bias.

### Analyses of Population Genetic Diversity and Differentiation

Population genetic diversity statistics, including observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ), nucleotide diversity ( $\pi$ ), and inbreeding coefficient ( $F_{IS}$ ) were calculated for variant loci as well as all nucleotide positions using the “populations” program in Stacks v2.3b (Rochette and Catchen, 2017). Population YSMP02, which was comprised of fewer



than three samples, was excluded from all genetic diversity analyses. Population differentiation ( $F_{ST}$ ) among 13 populations of *A. albus* was analysed with the “populations” program and the significance was inferred by Arlequin 3.5 with 1,000 permutations (Excoffier and Lischer, 2010).

### Genetic Structure of *A. albus* Populations

Genetic structure of 13 *A. albus* populations was assessed in multiple ways. First, we used a Bayesian clustering approach implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000). We set the possible cluster numbers ( $K$ ) ranging from 1 to 10, and 10 independent calculations were run for each  $K$  number. The simulations were conducted with 100,000 burn-ins as well as 500,000 Markov Chain Monte Carlo (MCMC)

chains after burn-ins. The most likely cluster number was determined by the  $\Delta K$  value calculated by Structure Harvester 0.6 (Earl and vonHoldt, 2012). In addition to STRUCTURE, principal component analysis (PCA), and fineRADstructure (Milan et al., 2018) were used as complementary assessments of population structure. The PCA plot of *A. albus* germplasms was drawn using the “adeigenet” package in R (Jombart, 2008). The fineRADstructure software is a modified version of the fineSTRUCTURE package (Lawson et al., 2012), which uses haplotype linkage information to calculate a co-ancestry matrix based on the most recent coalescence among sampled individuals (Milan et al., 2018). The input file was generated from the output of Stacks “populations” program using the data conversion script provided by fineRADstructure (Milan et al., 2018). The



fineRADstructure pipeline was run with 100,000 burn-ins and 100,000 MCMC chains, sampling every 1,000 steps. A tree illustrating the relationships among all individuals was also constructed with 10,000 hill-climbing iterations.

To quantify the genetic variance distributed in different hierarchical levels of *A. albus* germplasms, we assigned the 13 populations into genetic groups according to genetic structure analyses. Analysis of molecular variance (AMOVA) implemented in Arlequin 3.5 was used to detect the genetic variance among genetic clusters, among populations within clusters and within populations. The significance was calculated with 9,999 permutations (Excoffier and Lischer, 2010).

## Genetic, Geographic, and Environmental Correlations

To assess the contributions of geography and environment in driving the genetic differentiation of *A. albus*, mantel tests of isolation by distance (IBD) and isolation by environment (IBE) were conducted. Geographic distance between each pair of populations was calculated with the geographical coordinates of each population using GenAlex 6.5 (Peakall and Smouse, 2012). Using GPS coordinates of sampled populations, 19 climate variables with a resolution of 2.5 min were downloaded from the WorldClim<sup>1</sup> (Supplementary Table 1). The 19 climatic variables were first subject to PCA using SPSS Statistics 17.0 to reduce dimensionality. PCA suggested that the first three principal components (PCs) explained 99.19% of the total variation (Supplementary Table 2). We therefore selected the first three PCs of climatic variables to use as data points, and calculated Euclidean distance using “ecodist” package in R to create a pairwise environmental distance matrix for all populations

<sup>1</sup><https://www.worldclim.org/>

(Goslee and Urban, 2007). Finally, the genetic, geographic, and environmental correlations were tested using GenAlex 6.5 with 1,000 permutations (Peakall and Smouse, 2012).

We used redundancy analyses (RDA), implemented in the “vegan” package in R, to simultaneously estimate the effects of geography and environment on genomic variation (Oksanen et al., 2015). The genomic SNPs from each individual were treated as response factors, and geographic (latitude and longitude) and climatic (three climatic PCs) variables were used as explanatory factors. An initial global test was conducted using 999 permutations to determine whether the genome-wide SNP variation of 87 samples could be explained by geography and climatic factors. To determine the role of each variable separately, a partial RDA (with 999 permutations) which the effects of all but the tested explanatory variables were treated as conditioned factors was tested one by one (Wang et al., 2017).

## Genome-Wide Signatures of Adaptive Differentiation

We used BayPass 2.1 to conduct the genome scan for adaptive divergence in *A. albus* germplasms (Gautier, 2015). This software identifies overly differentiated outliers by calculating the  $XtX$  statistics and the matrix  $\Omega$  (Günther and Coop, 2013; Gautier, 2015; Flori et al., 2019). First, the  $XtX$  statistics of *A. albus* germplasms were calculated based on genomic SNPs. Then, a pseudo-observed data set (POD) with 10,000 SNPs was simulated according to parameters estimated from the real data set and used to calibrate the  $XtX$  statistics (Gautier, 2015). The posterior estimate of  $\Omega$  obtained with the POD was compared with that obtained from the real data. Finally, a 1% significance threshold of the  $XtX$  statistics was used to identify genomic footprints of selection.

**TABLE 1** | The statistical values of genetic diversity within populations from variant and all positions data of RAD-seq.

Population	N	A <sub>P</sub>	Percent of polymorphic loci (%)	H <sub>O</sub>		H <sub>E</sub>		π		F <sub>IS</sub>	
				Variant positions	All positions	Variant positions	All positions	Variant positions	All positions	Variant positions	All positions
YPHG	3	1	0.0152	0.1072	0.00007	0.0829	0.00006	0.0994	0.00007	-0.0135	-0.00001
YPJW	5	12	0.0239	0.1456	0.00011	0.1154	0.00008	0.1296	0.00009	-0.0317	-0.00002
YSMP02	2	-	-	-	-	-	-	-	-	-	-
JBBC	8	50	0.0260	0.1515	0.00011	0.1117	0.00008	0.1197	0.00009	-0.0504	-0.00004
XLFT	8	48	0.0265	0.1739	0.00013	0.1202	0.00009	0.1288	0.00009	-0.0820	-0.00006
DGZB	8	36	0.0270	0.1943	0.00014	0.1373	0.0001	0.1471	0.00011	-0.0911	-0.00007
QJMZ	8	81	0.0262	0.1815	0.00013	0.1270	0.00009	0.1361	0.0001	-0.0930	-0.00007
YSMP	8	62	0.0323	0.1483	0.00011	0.1383	0.0001	0.1484	0.00011	-0.0027	0
YCFZ	8	17	0.0288	0.1465	0.00011	0.1258	0.00009	0.1349	0.0001	-0.0247	-0.00002
LBMH	8	245	0.0359	0.1331	0.0001	0.1417	0.0001	0.1517	0.00011	0.0460	0.00003
BXTC	8	35	0.0271	0.1647	0.00012	0.1281	0.00009	0.1374	0.0001	-0.0555	-0.00004
BYDW	8	15	0.0282	0.1526	0.00011	0.1281	0.00009	0.1373	0.0001	-0.0328	-0.00002
YFYT	5	17	0.0136	0.1510	0.00011	0.0825	0.00006	0.0928	0.00007	-0.1046	-0.00008

N, number of individuals; A<sub>P</sub>, private allele number; H<sub>O</sub>, observed heterozygosity; H<sub>E</sub>, expected heterozygosity; π, nucleotide diversity; F<sub>IS</sub>, inbreeding coefficient.

To complement the outlier tests, we also carried out GEA analyses using BayPass 2.1, with the AUX model and parameters estimated above (Gautier, 2015). We treated three PCs of the climate variables as covariables, and the significant association of each SNP with each covariable was judged by a calibrated Bayes Factor (BF) of POD with 1% threshold.

## RESULTS

### RAD-Seq Assembling and Genotyping

Sequencing RAD-Tags from 87 individuals of *A. albus* germplasms produced 453.75 GB of clean data. The genomic data per sample ranged from 2.97 to 10.65 GB, with the average value being 5.22 GB (Supplementary Table 3). The minimum and maximum coverage depth of each sample was 7.26 and 13.92, respectively, and the average depth was 9.12 (Supplementary Table 3). After applying the assembly and filtering criteria described above, a final SNP data set consisting of 24,225 loci was retained for subsequent analyses.

### Genetic Diversity and Differentiation Among Populations

Genetic diversity indices, including  $H_E$  and  $\pi$ , were highest in population LBMH and lowest in YFYT, when calculated from variant loci and from all positions (Table 1). Meanwhile, the observed heterozygosity ( $H_O$ ) within populations was highest in DGZB and lowest in YPHG. The largest number of private alleles was found in LBMH (Table 1). Pairwise  $F_{ST}$  values among populations ranged from 0.038 (YCFZ versus YPJW) to 0.352 (YFYT versus YPHG), and significant genetic differentiations were detected between most population pairs (69 out of 78  $F_{ST}$  values) ( $P < 0.05$ ) (Table 2).

### Analysis of Population Genetic Structure

The STRUCTURE analysis indicated that  $K = 6$  was the best supported number of genetic clusters, with the second largest  $\Delta K$  value detected when  $K = 2$  (Supplementary Figure 1). However,

results from the other two analyses, PCA and fineRADstructure, suggested that the 13 populations should be divided into 3 genetic groups. As the result of  $K = 6$  was not biologically informative, we selected  $K = 3$  as the optimal genetic cluster number (Figure 2B). The three genetic groups were comprised of a geographically upstream group (DGZB, YSMP, YSMP02, and QJMZ; Group A), a midstream group (YPJW, YPHG, and YCFZ; Group B), and a group distributed along the downstream region (LBMH, BXTC, XLFT, JBBC, BYDW, and YFYT; Group C) (Figures 1B, 2B). The 13 populations were essentially grouped by geographic proximity, and substantial genetic introgression was detected in the 3 populations of group B (Figures 1B, 2B). Analysis of fineRADstructure further revealed the hierarchical genetic structure of *A. albus* germplasms. The 13 populations were first divided into 2 groups, with the up-stream group further differentiated into two sub-clusters (Figure 3). The divergence within the upstream group was also supported by an independent STRUCTURE analysis using only the up-stream populations (Supplementary Figure 2).

To quantify the genetic variance distribution in *A. albus* germplasms, scenarios of two and three genetic groups were tested using AMOVA. When assigning populations into two groups, 24.68% of variation was among genetic groups ( $P < 0.0001$ ). When the populations were assigned into three groups, a higher percentage of variation (29.90%) among groups was revealed ( $P < 0.0001$ ) (Table 3).

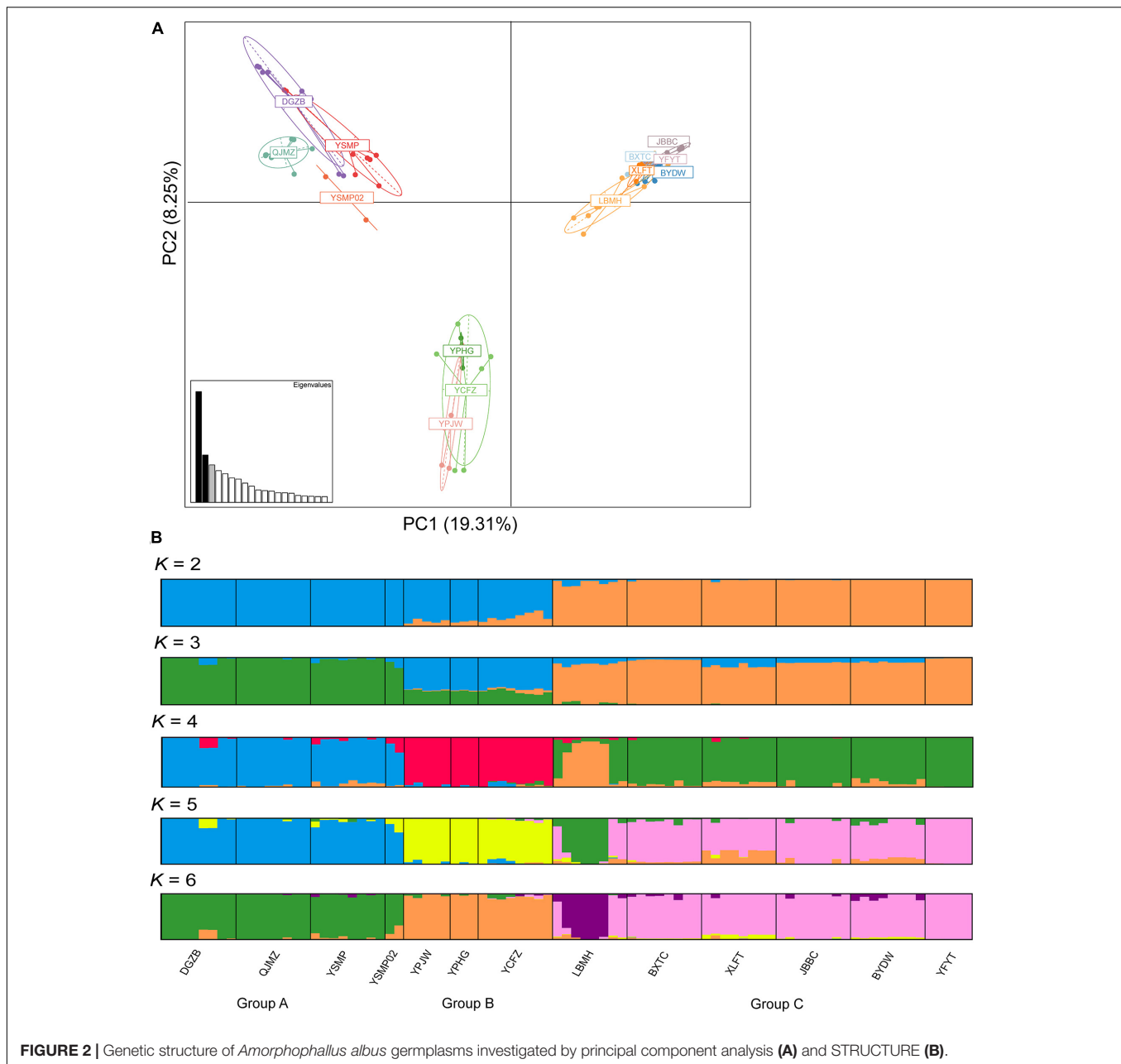
### Genetic, Geographic, and Environmental Correlations

The first three PCs summarised 58.52, 32.22, and 8.45%, respectively, of the total variation in 19 climatic variables downloaded from WorldClim (Supplementary Table 2). Both IBD and IBE analyses all suggested that the genetic distance was significantly correlated with geographic and environmental distances (Figures 4A,B). The proportion of genetic variation that could be explained by geographic and climatic factors was further assessed with RDA. The initial global RDA test was highly significant, and the five constraining variables recovered 33.98%

**TABLE 2** | Pairwise  $F_{ST}$  values among 13 populations of *Amorphophallus albus* based on 24,225 SNPs.

Population	DGZB	QJMZ	YSMP	YSMP02	YPJW	YPHG	YCFZ	LBMH	BXTC	XLFT	JBBC	BYDW
QJMZ	<b>0.149</b>											
YSMP	0.101	<b>0.135</b>										
YSMP02	0.126	0.108	0.098									
YPJW	<b>0.219</b>	<b>0.239</b>	<b>0.202</b>	0.209								
YPHG	<b>0.212</b>	<b>0.245</b>	<b>0.206</b>	0.257	0.091							
YCFZ	<b>0.198</b>	<b>0.215</b>	<b>0.183</b>	<b>0.171</b>	0.038	0.070						
LBMH	<b>0.217</b>	<b>0.233</b>	<b>0.186</b>	<b>0.188</b>	<b>0.175</b>	<b>0.175</b>	<b>0.153</b>					
BXTC	<b>0.251</b>	<b>0.269</b>	<b>0.217</b>	<b>0.225</b>	<b>0.209</b>	<b>0.215</b>	<b>0.179</b>	<b>0.109</b>				
XLFT	<b>0.257</b>	<b>0.274</b>	<b>0.222</b>	<b>0.231</b>	<b>0.218</b>	<b>0.225</b>	<b>0.185</b>	<b>0.128</b>	<b>0.134</b>			
JBBC	<b>0.275</b>	<b>0.292</b>	<b>0.238</b>	<b>0.251</b>	<b>0.233</b>	<b>0.243</b>	<b>0.200</b>	<b>0.126</b>	<b>0.122</b>	<b>0.147</b>		
BYDW	<b>0.247</b>	<b>0.266</b>	<b>0.212</b>	<b>0.223</b>	<b>0.205</b>	<b>0.212</b>	<b>0.175</b>	<b>0.096</b>	<b>0.108</b>	<b>0.112</b>	<b>0.104</b>	
YFYT	<b>0.330</b>	<b>0.349</b>	<b>0.281</b>	<b>0.346</b>	<b>0.303</b>	<b>0.352</b>	<b>0.245</b>	<b>0.163</b>	<b>0.137</b>	<b>0.195</b>	<b>0.175</b>	<b>0.168</b>

Bold values indicate significant levels ( $P < 0.05$ ).



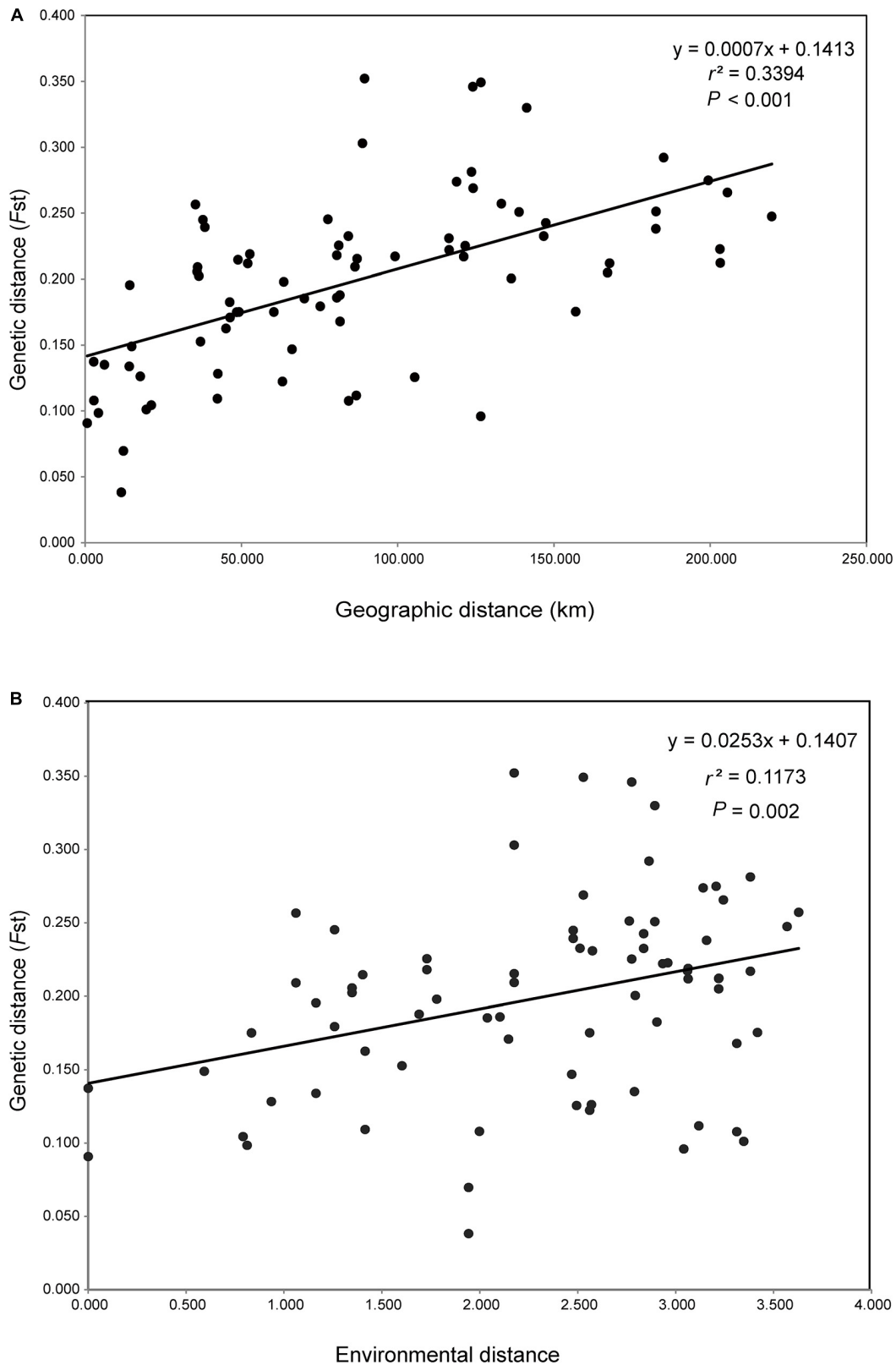
of the genotypic variance (Supplementary Table 4). The first two axes of RDA explained 15.55 and 5.60% of the total variance (Figure 5A). When testing the role of each variable in predicting genotypic variation independently, all partial RDA tests revealed a significant effect for each geographic and climatic factor by conditioning all other variables (Supplementary Table 4). Overall, the climatic variables explained a higher percent of SNP variation (14.29%) than geographic factors (9.63%) (Figure 5B).

## Genome-Wide Footprints of Adaptive Differentiation

For the simulations in BayPass, the Spearman's correlation between the posterior estimate of  $\Omega$  obtained with the POD

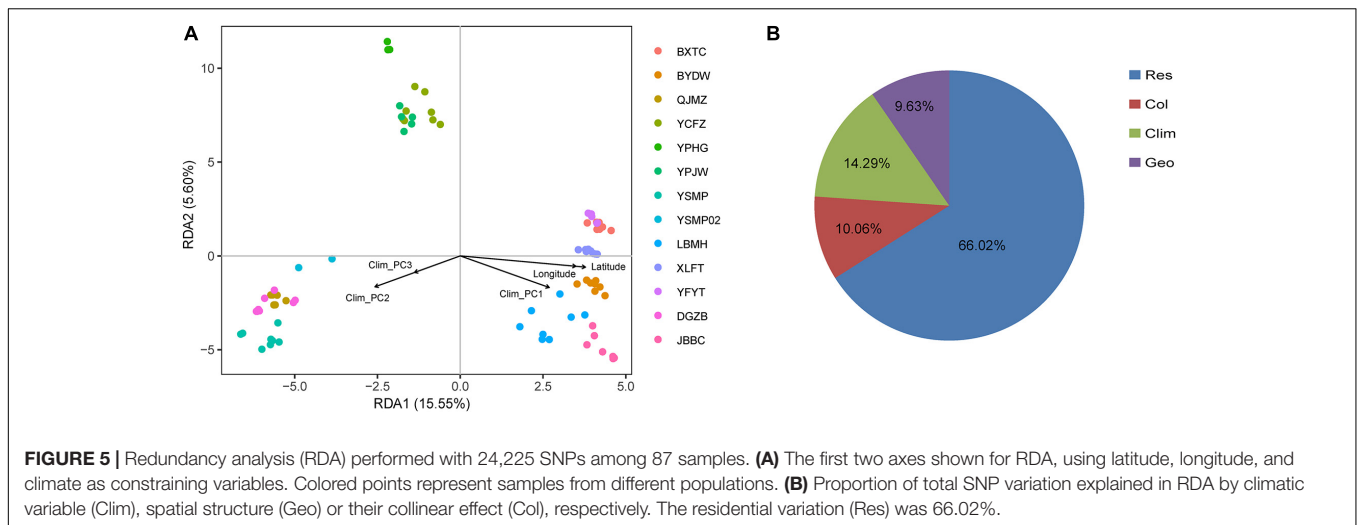
and that acquired from the real data set was nearly one ( $\rho = 0.9988$ ) (Supplementary Figure 3). This indicated that the POD faithfully mimics the real data set, and the 1% significance threshold of the  $XtX$  statistics ( $XtX = 19.62$ ) was robust to identify outliers under genomic selection. Through this genome scan for outlier loci, 20 outliers were detected by the  $XtX$  calibration in BayPass (Supplementary Figure 4A and Supplementary Table 5). GEA analyses revealed that 14 loci were significantly associated with Clim\_PC1, 14 loci with Clim\_PC2, and 22 loci with Clim\_PC3 (Supplementary Figures 4B–D and Supplementary Table 5). In total, 20 loci were detected as significant in two or more tests (Supplementary Table 5).





**FIGURE 4 |** Mantel tests of geographic, environmental, and genetic correlations. **(A)** Correlation of pairwise geographic distance versus pairwise genetic distance ( $F_{ST}$ ); **(B)** correlation of pairwise environmental distance versus pairwise  $F_{ST}$ .





Our results supported the previous inference from SSR makers that *A. albus* genetically differentiated into two geographic groups, one from the upstream region and another (Group C) in the downstream region (**Figure 1B**). A novel result from the current study was the identification of sub-divergence within the upstream group, which can be separated into two genetic clusters (**Figure 2**). Although the plot of  $\Delta K$  value indicate that  $K = 6$  was the most supported genetic cluster number in the STRUCTURE analysis, the result of  $K = 6$  was not biologically informative. The other two analyses (PCA and fineRADstructure) all suggested that populations of *A. albus* germplasms should be divided into three genetic groups (**Figure 2B**). We also conducted an independent STRUCTURE analysis of only the seven populations from the upstream region, which supported the genetic division of the upstream group (**Supplementary Figure 2**). All these tests indicate that the inference about genetic relationships among populations was robust. Our previous study only detected the two genetic clusters. Because of both an increase in the number of independent loci and in the overall number of alleles, the thousands of SNPs outperformed the few SSRs in the genetic assignment analysis (Fischer et al., 2017). Our study clearly demonstrates the power of genome-wide SNPs in discovering cryptic genetic structure of domesticated plant.

## Drivers Shaping the Geographic Differentiation Pattern of *A. albus*

Much previous work has been done to investigate the role of geographic and environmental factors in driving genetic patterns of natural populations (e.g., Orsini et al., 2013; Wang et al., 2017). Our results, based on separate IBD and IBE analyses, indicated that the genetic differentiation of *A. albus* was shaped by both effects of IBD and IBE. The existence of IBD in natural populations is usually interpreted as the equilibrium between random drift and migration, in which gene flow is dependent on geographical distance (Gao et al., 2015). This pattern can be generated by both island and stepping-stone models (Orsini et al., 2013). Because stepping-stone colonisation should result in

lower genetic diversity in populations that were colonised later, we could potentially distinguish these two processes (Prunier and Holsinger, 2010). Given the short domestication time of *A. albus*, the genetic pattern of germplasms should not yet be extensively affected by cultivation of this species. As no clear variation pattern of genetic diversity was observed in 13 populations, we conclude that the geographically isolated landscapes contribute to the genetic differentiation pattern of *A. albus*.

We also observed a significant effect of climatic factor distances on genetic differentiation of *A. albus*. A large body of genetic research has identified IBE as one driver of population differentiation, highlighting the importance of environmental adaptation in shaping genetic differentiation patterns. RDA analysis enables identification and testing of the effect of individual variables influencing genomic variability, while also offering the ability to detect collinearity between them (Lasky et al., 2012; Wang et al., 2017). Our RDA results supported a stronger influence of environmental adaptation over geographic factors in driving population differentiation, with 14.29% of genetic differentiation being explained by the environment (**Figure 5B**). That the environment was more important than geography in predicting genetic differentiation suggests a predominant role of environmental adaptation in *A. albus*. Given the heterogeneous micro-environments in this region, these populations might have adapted to the local conditions of their microhabitats. As reported in other species (Ye et al., 2016), geographic isolation may have induced initial differentiation, which was then reinforced by adaptation in response to local climate and environmental factors.

## Genomic Regions Associated With Local Adaptation

Understanding genetic variation for local adaptation in the gene pool of domesticated landraces is important for the selection of crops with increased yield potential in future environments (Redden, 2013). In this study, BayPass software identified highly differentiated SNPs based on a calibration procedure of the

XtX statistic (Günther and Coop, 2013) by correcting for confounding demographic effects (Gautier, 2015). In total, 20 loci were detected as under positive selection pressure with this method. Additionally, more than 20 loci in total were significantly associated with climate factors as revealed by GEA (Supplementary Figures 4B–D and Supplementary Table 5). These environment-associated SNPs might reflect the impact of local adaptation, which has been implicated in studies of other species (Lotterhos and Whitlock, 2014; Wang et al., 2017). Previous studies of local adaptation have shown a trend toward complex adaptation involving many loci, each under relatively weak selection (Allaby et al., 2015; Zhou et al., 2017). The multiple loci identified in our study support the idea of complex local adaptation in *A. albus* populations. Due to the lack of a reference genome, we could not identify the genomic distribution of the identified loci nor linked functional genes. Despite this knowledge gap, the conservation of these adaptive genomic regions can be included in breeding plans for this species that would lead to robust and locally adapted cultivars (Zhou et al., 2017).

## Breeding Implications

Although *A. albus* has been cultivated and consumed for a long time in China, the crop is cultivated from unimproved original genetic material in most regions (Srzednicki and Borompichaichartkul, 2020). The reproduction ratio of *Amorphophallus* species is much lower compared with most other tuber crops. For example, the annual reproduction coefficient of potatoes, sweet potatoes, and other tuber crops can be up to more than a dozen times, but that of *A. albus* is only one third of these plants (Adams and Stevenson, 1990; Srzednicki and Borompichaichartkul, 2020). The lack of seed corms further hindered the selection on agronomic traits (Yin et al., 2019).

Sexual crossbreeding should be carried out on *A. albus* to offset its demerit of a low yield (Srzednicki and Borompichaichartkul, 2020). The crossbreeding attempts of this plant have just started in last 20 years. The genomic variation pattern of *A. albus* observed in this study can guide breeding programs. Based on the relatively low genetic diversity, more breeding material is needed. When making crossbreeding strategies, the implement of materials from different genetic groups should be considered. Only in this way can the later generations have greater variation and, therefore, provide more selection opportunities.

In conclusion, we described the genomic genetic diversity of *A. albus* germplasm using tens of thousands SNPs developed from RAD-seq. Cryptic genetic structure was also uncovered in this species, which demonstrated the increased power of genomic SNPs over traditional markers. We further analysed the

roles of geographic and environmental factors in shaping genetic patterns of *A. albus* germplasm, and discovered genomic regions that might be associated with local adaptation. Together these findings shed light on the geographic differentiation patterns of *A. albus*, and will inform the future breeding programs of this agriculturally important plant.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA715021.

## AUTHOR CONTRIBUTIONS

YG, LT, and DD conceived and designed the study. YG, HChu, HChen, and SY collected the samples. SY and YZ carried out the laboratory work and performed the analyses. YG drafted the manuscript. All authors contributed to writing and editing and read and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31760103 and 31860620), the Special Basic Cooperative Research Programs of Yunnan Provincial Undergraduate Universities (2018FH001-006 and 202001BA070001-003), the Applied Basic Research Project of Yunnan Province (2018FB050), and the research grant from Qujing Normal University (2018QN001).

## ACKNOWLEDGMENTS

We would like to thank Stephen Gaughran at Yale University for his assistance with English language and grammatical editing of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.683422/full#supplementary-material>

## REFERENCES

- Adams, S. S., and Stevenson, W. R. (1990). Water management, disease development, and potato production. *Am. J. Potato Res.* 67, 3–11. doi: 10.1007/bf02986908
- Allaby, R. G., Kistler, L., Gutaker, R. M., Ware, R., Kitchen, J. L., Smith, O., et al. (2015). Archaeogenomic insights into the adaptation of plants to the human environment: pushing plant–hominin co-evolution back to the Pliocene. *J. Hum. Evol.* 79, 150–157. doi: 10.1016/j.jhevol.2014.10.014
- Aslan, S., Forsberg, N. E. G., Hagenblad, J., and Leino, M. W. (2015). Molecular genotyping of historical barley landraces reveals novel candidate regions for local adaptation. *Crop Sci.* 55, 2766–2776. doi: 10.2135/cropsci2015.02.0119
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Barton, N. H., and Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3, 11–21. doi: 10.1038/nrg700

- Bolger, M. E., Weisshaar, B., Scholz, U., Stein, N., and Mayer, K. F. (2014). Plant genome sequencing - applications for crop improvement. *Curr. Opin. Biotech.* 26C, 31–37. doi: 10.1016/j.copbio.2013.08.019
- Cantelmo, N. F., Pinho, R. G. V., and Balestre, M. (2016). Genomic breeding value prediction for simple maize hybrid yield using total effects of associated markers, under different imbalance levels and environments. *Genet. Mol. Res.* 15:15017232.
- Chua, M., Baldwin, T. C., Hocking, T. J., and Chan, K. (2010). Traditional uses and potential health benefits of *Amorphophallus konjac* K. Koch ex N.E.Br. *J. Ethnopharmacol.* 128, 268–278. doi: 10.1016/j.jep.2010.01.021
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., et al. (2017). Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* 18:69. doi: 10.1186/s12864-016-3459-7
- Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., et al. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol. Ecol.* 28, 1009–1029. doi: 10.1111/mec.15004
- Ganal, M. W., Altmann, T., and Roder, M. S. (2009). SNP identification in crop plants. *Curr. Opin. Plant Biol.* 12, 211–217. doi: 10.1016/j.pbi.2008.12.009
- Gao, Y., Ai, B., Kong, H., Kang, M., and Huang, H. (2015). Geographical pattern of isolation and diversification in karst habitat islands: a case study in the *Primulina eburnea* complex. *J. Biogeogr.* 11, 2131–2144. doi: 10.1111/jbi.12576
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 201, 1555–1579. doi: 10.1534/genetics.115.181453
- Gille, S., Cheng, K., Skinner, M. E., Liepman, A. H., Wilkerson, C. G., and Pauly, M. (2011). Deep sequencing of voodoo lily (*Amorphophallus konjac*): an approach to identify relevant genes involved in the synthesis of the hemicellulose glucomannan. *Planta* 234, 515–526. doi: 10.1007/s00425-011-1422-z
- Goslee, S. C., and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* 22, 1–19.
- Grillo, M. A., De Mita, S., Burke, P. V., Solórzano-Lowell, K. L. S., and Heath, K. D. (2016). Intrapopulation genomics in a model mutualist: population structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution* 70, 2704–2717. doi: 10.1111/evo.13095
- Günther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics* 195, 205–220. doi: 10.1534/genetics.113.152462
- Henry, R. J. (2012). Next-generation sequencing for understanding and accelerating crop domestication. *Brief. Func. Genomics* 11, 51–56. doi: 10.1093/bfgp/eln032
- Hu, J. B., Liu, J., Xie, C. H., and Gao, X. X. (2006). Corm induction and multiplication of *Amorphophallus albus* in vitro. *J. Hortic. Sci. Biotech.* 81, 859–863.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Lasky, J. R., Marais, D. L. D., Mckay, J. K., Richards, J. H., and Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* 21, 5512–5529. doi: 10.1111/j.1365-294x.2012.05709.x
- Lawson, D. J., Simon, G. H., and Falush, M. D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet.* 8:e1002453. doi: 10.1371/journal.pgen.1002453
- Li, H., Zhu, G., Boyce, P. C., Jin, M., and Jacobsen, N. (2010). “Araceae,” in *Flora of China*, eds Z. Wu and H. Li (Beijing: Science Press), 23–33.
- Liu, P. (2003). *Konjac*. Beijing: China Agriculture Press.
- Lotterhos, K. E., and Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol. Ecol.* 23, 2178–2192. doi: 10.1111/mec.12725
- Milan, M., Emiliano, T., John, L. D., and Daniel, F. (2018). RADpainter and fineRADstructure: population inference from RADseq data. *Mol. Biol. Evol.* 35, 1284–1290. doi: 10.1093/molbev/msy023
- Nishinari, K., Williams, P. A., and Phillips, G. O. (1992). Review of the physico-chemical characteristics and properties of konjac mannan. *Food Hydrocolloid* 6, 199–222. doi: 10.1016/s0268-005x(92)80360-3
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2015). *Package Vegan: Community Ecology Package in R*. Available online at: <https://CRAN.R-project.org/package=vegan> (accessed September 10, 2020).
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., and De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol. Ecol.* 22, 5983–5999. doi: 10.1111/mec.12561
- Peakall, R., and Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Pritchard, J. K., Stephens, and Matthew. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- Prunier, R., and Holsinger, K. E. (2010). Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within Protea (*Proteaceae* L.). *Mol. Ecol.* 19, 3968–3980. doi: 10.1111/j.1365-294x.2010.04779.x
- Redden, R. (2013). New approaches for crop genetic adaptation to the abiotic stresses predicted with climate change. *Agronomy* 3, 419–432. doi: 10.3390/agronomy3020419
- Rochette, N. C., and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12, 2640–2659. doi: 10.1038/nprot.2017.123
- Singh, S. P. (2001). Broadening the genetic base of common bean cultivars: a review. *Crop Sci.* 41, 1659–1675. doi: 10.2135/cropsci2001.1659
- Srzednicki, G., and Borompichaichartkul, C. (2020). *Konjac Glucomannan-Production, Processing, and Functional Applications*. Boca Raton: CRC Press.
- Tang, R., Liu, E., Zhang, Y., Schinnerl, J., and Chen, G. (2020). Genetic diversity and population structure of *Amorphophallus albus*, a plant species with extremely small populations (PSESP) endemic to dry-hot valley of Jinsha River. *BMC Genomics* 21:102. doi: 10.1186/s12863-020-00910-x
- Wang, J., Feng, C., Jiao, T., Von Wettberg, E. B., and Kang, M. (2017). Genomic signature of adaptive divergence despite strong nonadaptive forces on edaphic islands: a case study of *Primulina juliae*. *Genome Biol. Evol.* 9, 3495–3508. doi: 10.1093/gbe/evx263
- Ye, Z., Chen, P., and Bu, W. (2016). Terrestrial mountain islands and Pleistocene climate fluctuations as motors for speciation: a case study on the genus *Pseudovelonia* (Hemiptera: Veliidae). *Sci. Rep.* 6:33625.
- Yin, S., Yan, Y., You, L., Chen, Q., and Gao, Y. (2019). Newly developed genomic SSRs reveal genetic diversity in wild and cultivated *Amorphophallus albus* germplasms. *Plant Mol. Biol. Rep.* 37, 365–375. doi: 10.1007/s11105-019-01162-5
- Zhang, S., Liu, P., Zhang, X., Zhang, Y., and Su, C. (1999). Resources of Chinese germplasms of *Amorphophallus* and proposal for its development and utilization. *J. Southwest Agric. Univ.* 21, 515–519.
- Zhang, Y. Q., Xie, B. J., and Gan, X. (2005). Advance in the applications of konjac glucomannan and its derivatives. *Carbohydr. Polym.* 60, 27–31. doi: 10.1016/j.carbpol.2004.11.003
- Zhou, Y., Chen, Z., Cheng, M., Chen, J., Zhu, T., Wang, R., et al. (2017). Uncovering the dispersion history, adaptive evolution and selection of wheat in China. *Plant Biotechnol. J.* 16, 280–291. doi: 10.1111/pbi.12770

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gao, Yin, Chu, Zhang, Wang, Chen, Liu, Dai and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.