



OPEN ACCESS

EDITED BY

Leif Skot,
Aberystwyth University, United Kingdom

REVIEWED BY

Alejandra Moreno-Letelier,
National Autonomous University of Mexico,
Mexico
Joao Paulo Gomes Viana,
University of Illinois at Urbana-Champaign,
United States

*CORRESPONDENCE

Carlos I. Arbizu

✉ carlos.arbizu@untrm.edu.pe

Rodomirol Ortiz

✉ rodomirol.ortiz@slu.se

RECEIVED 12 November 2024

ACCEPTED 30 January 2025

PUBLISHED 25 February 2025

CITATION

Arbizu CI, Bazo-Soto I, Flores J, Ortiz R,
Blas R, García-Mendoza PJ, Sevilla R, Crossa J
and Grobman A (2025) Genotyping by
sequencing reveals the genetic diversity and
population structure of Peruvian highland
maize races.

Front. Plant Sci. 16:1526670.

doi: 10.3389/fpls.2025.1526670

COPYRIGHT

© 2025 Arbizu, Bazo-Soto, Flores, Ortiz, Blas,
García-Mendoza, Sevilla, Crossa and Grobman.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Genotyping by sequencing reveals the genetic diversity and population structure of Peruvian highland maize races

Carlos I. Arbizu^{1,2*}, Isamar Bazo-Soto³, Joel Flores³,
Rodomirol Ortiz^{4*}, Raul Blas³, Pedro J. García-Mendoza⁵,
Ricardo Sevilla³, José Crossa⁶ and Alexander Grobman³

¹Centro de Investigación en Germoplasma Vegetal y Mejoramiento Genético de Plantas (CIGEMP),
Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Chachapoyas, Peru,

²Facultad de Ingenierías y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de
Amazonas (UNTRM), Chachapoyas, Peru, ³Laboratorio de Genómica y Bioinformática, Universidad Nacional
Agraria la Molina (UNALM), Lima, Peru, ⁴Department of Plant Breeding, Swedish University of Agricultural
Sciences (SLU), Alnarp, Sweden, ⁵Facultad de Ingeniería, Universidad Nacional Autónoma de Tayacaja
(UNAT), Huancavelica, Peru, ⁶Biometrics and Statistics Unit, International Maize and Wheat Improvement
Center (CIMMYT), Mexico City, Mexico

Peruvian maize exhibits abundant morphological diversity, with landraces cultivated from sea level (sl) up to 3,500 m above sl. Previous research based on morphological descriptors, defined at least 52 Peruvian maize races, but its genetic diversity and population structure remains largely unknown. Here, we used genotyping-by-sequencing (GBS) to obtain single nucleotide polymorphisms (SNPs) that allow inferring the genetic structure and diversity of 423 maize accessions from the genebank of Universidad Nacional Agraria la Molina (UNALM) and Universidad Nacional Autónoma de Tayacaja (UNAT). These accessions represent nine races and one sub-race, along with 15 open-pollinated lines (purple corn) and two yellow maize hybrids. It was possible to obtain 14,235 high-quality SNPs distributed along the 10 maize chromosomes of maize. Gene diversity ranged from 0.33 (sub-race Pachia) to 0.362 (race Ancashino), with race Cusco showing the lowest inbreeding coefficient (0.205) and Ancashino the highest (0.274) for the landraces. Population divergence (F_{ST}) was very low (mean = 0.017), thus depicting extensive interbreeding among Peruvian maize. A cluster containing maize landraces from Ancash, Apurímac, and Ayacucho exhibited the highest genetic variability. Population structure analysis indicated that these 423 distinct genotypes can be included in 10 groups, with some maize races clustering together. Peruvian maize races failed to be recovered as monophyletic; instead, our phylogenetic tree identified two clades corresponding to the groups of the classification of the races of Peruvian maize based on their chronological origin, that is, anciently derived or primary races and lately derived or secondary races. Additionally, these two clades are also congruent with the geographic origin of these maize races, reflecting their mixed evolutionary backgrounds and constant evolution. Peruvian maize

germplasm needs further investigation with modern technologies to better use them massively in breeding programs that favor agriculture mainly in the South American highlands. We also expect this work will pave a path for establishing more accurate conservation strategies for this precious crop genetic resource.

KEYWORDS

germplasm, Andes, SNP markers, genetic resources, NGS

1 Introduction

Maize is a major global crop that is cultivated on approximately 200 million hectares and is considered a key component of food security (Erenstein et al., 2022). In several regions of the world (western South America, Mesoamerica, sub-Saharan Africa), this crop has an important social and economic value because maize is consumed daily, mainly by the poor in their single daily meal. Flourey maize grows in about one million hectares in the South American Andes, where it is consumed with very little processing; that is, boiled grain (mote), fried grain without oil (cancha), or boiled ear as green corn (choclo). Migration of rural populations to urban areas dropped maize consumption by switching to more expensive products (Sevilla, 1994).

Since the publication of “The Races of Maize in Mexico” (Wellhausen et al., 1952), each country, where maize thrives, released similar books. The South American Andean races were described for Colombia (Roberts et al., 1957), Brazil and other eastern South American countries (Brieger et al., 1958), Bolivia (Ramírez et al., 1960), Perú (Grobman et al., 1961), Chile (Timothy et al., 1961), Ecuador (Timothy et al., 1963), and Venezuela (Grant et al., 1963). These investigations described 132 races in the South American Andes, that is, about 52% of the known 260 maize races (Goodman and Brown, 1988). Due to the methodology of sampling for collecting maize in the farms or rural households, maize landraces were recognized together with farmers according to their assessment for different purposes and under the name of their native language. More than half a century after these maize race books were published, we recognize the importance of the race concept to classify the diversity of this crop. Race, which is important for maize diversity classification, has been the unit of maize diversity for more than 60 years and provides means for its monitoring. For example, to the best of our knowledge, all described races remain available in Perú (Ministerio del Ambiente, 2018). To assess the current maize diversity, a new racial classification of the Peruvian maize is being conducted (Ministerio del Ambiente, 2018).

Race diversity is not an indicator of genetic structure in maize because races were described using only a few morphological characters, mostly from the ear and grain. Likewise, there are ecological and cultural criteria for classification that are not well understood. Maize farming began in Mexico about 9,000 years ago (Piperno et al., 2009). According to Kistler et al. (2018), a wild

teosinte (*Zea mays* ssp. *parviglumis*) was domesticated in Mexico and thereafter spread toward the south and arrived in Peru, where it continued to expand to the Andean region and later to the Amazon region. The southwestern Amazon is, however, considered now a secondary improvement center for the partially domesticated maize (Kistler et al., 2018). Maize was grown in Peru about 6,700 years ago in the Chicama Valley, where a sample of maize with well-conserved cobs, husks, stalks, and tassels was found (Grobman et al., 2012). The macrofossil records indicated that maize included various races at that time. Grobman (1982) indicated that maize diversification began very early in human settlements. There is evidence of massive use of maize as food at Los Gavilanes, an archaeological site in Huarney, north of Lima (Bonavia, 2008).

Andean maize possesses many beneficial (favorable) alleles to address its limitation in the production across the South American highlands (Sevilla Panizo and Holle Ostendorf, 2004; Salazar et al., 2017). The success of plant breeding depends on genetic diversity, but this must be organized, tested, and evaluated. To maintain the total maize diversity, every race of this crop should be improved. This breeding strategy ensures both variety adaptation and further adoption. Therefore, accurate classification is essential; all diversity belonging to a maize race should be understood, and all races properly defined. Precise and objective techniques and tools for these research tasks are urgently needed, as classification based on morphology, adaptation, and cultural criteria is insufficient.

The genetic structure and diversity of the races of Peruvian maize remain largely unknown. On the other hand, Vigouroux et al. (2008) used a set of almost 350 races of maize to assess their genetic and population parameters with 96 microsatellite markers, detecting four main groups: (i) highland Mexican, (ii) northern United States of America, (iii) tropical lowland, and (iv) Andean maize races. They suggested that isolation by distance could be the main factor accounting for the historical maize diversification. Further research with microsatellites suggested that the Andean group of maize displayed little mixing with other races (Bedoya et al., 2017). Genotyping-by-sequencing (GBS) is now a feasible technique for describing highly diverse and large genomes as maize (He et al., 2014). This approach is increasingly important as a cost-effective and unique tool for genomic diversity research and gene discovery in maize and provides more single nucleotide polymorphisms (SNPs) markers than SNP arrays (Wang et al., 2020).

Maize evolved rapidly due to human selection, leading to significant phenotypic changes and adaptation to various environments, such as those in Mesoamerica and the Andes. There are, however, some unanswered questions regarding maize domestication and its further evolution. The main aim of this investigation was to determine the genetic diversity and population structure of nine races and one subrace of maize grown in the highlands of Peru using SNPs covering all its 10 chromosomes generated by GBS, thus providing consistent means for understanding its classification and spread in the Peruvian Andes.

2 Materials and methods

2.1 Plant material

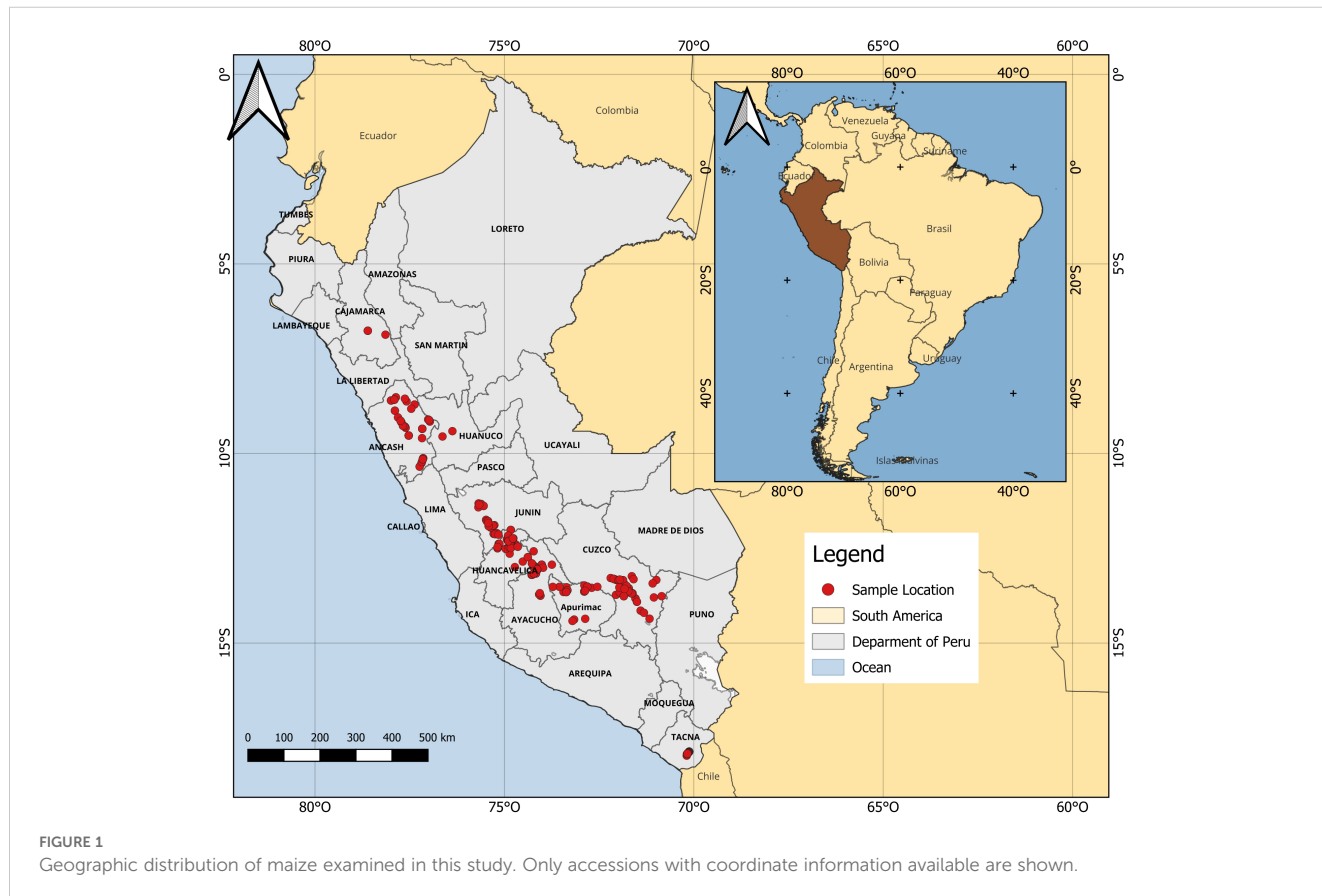
We examined (i) 406 accessions of nine races and one sub-race of Peruvian maize that are currently cultivated in 10 Andean geographic departments of Peru, (ii) 15 open-pollinated (OP) purple maize lines, and (iii) two yellow maize hybrids (423 individuals in total) (Figure 1). Maize accessions were obtained from the Maize Research Program Germplasm Bank maintained at the UNALM in Lima, except for 19 accessions that were obtained from the corn research project of UNAT. One yellow dent hybrid and all purple maize were donated by Associate Prof. Hugo E.

Huanuqueño (Maize Research Unit, UNALM); the other maize hybrid was a local cultivar. All possible accessions available as germplasm were employed. Further details of the maize genotypes examined in this study are available at the [Supplementary Table S1](#).

2.2 Genotyping-by-sequencing

All 423 maize individuals (one seed per accession) were planted at UNALM, and three weeks after germination, leaf samples from one plant of each sample were collected, and total genomic DNA was extracted following Doyle and Doyle (Doyle and Doyle, 1987) protocol with some modifications. The concentration and purity of DNA samples were determined with a NanoDrop 1000 spectrophotometer. DNA samples showing absorbance ratios above 1.8 at 260/280 nm were used for further analysis. For quality determination, DNA samples were electrophoresed on 1% agarose gel, and 50 random samples were digested with *ApeKI* following the manufacturer's protocol. Samples were sent to the University of Minnesota-Biotechnology Center for DNA sequencing.

Genotyping by sequencing libraries were developed following Elshire et al. (2011) protocol. Genomic DNA was digested with the *ApeKI* enzyme, and fragments were ligated to Illumina sequencing adapters and with sequence barcodes that are unique to each sample, which allows the recovery of sample identity for each



sequenced DNA fragment after multiplexing. The pooled samples were sequenced on the Illumina NovaSeq 6000 platform, from which 100 bp single-end sequences reads were obtained. Quality of the raw data was examined with FastQC v0.11.7 software (Andrews, 2010). Thereafter, we employed the TASSEL v5.2.42 bioinformatic pipeline (Bradbury et al., 2007; Glaubitz et al., 2014) for SNPs calling with maize Zm-B73-REFERENCE-NAM-5.0 (Hufford et al., 2021) as the reference genome. Parameters employed in this pipeline were the same as in the study of Huaranga-Joaquin et al. (2023). Data curation was performed using software VCFtools v0.1.16 (Danecek et al., 2011) with the following criteria of retention: (i) minimum minor allele frequency of 0.1, (ii) number of alleles less than or equal to 2, and (iii) maximum missing data of 0.1. Additional filtering was conducted by removing SNPs in linkage disequilibrium (LD) at a threshold of $r^2 = 0.2$ with the function *snpGdsLDpruning* of the SNPrelate package (Zheng et al., 2012) in the R v4.2.2 (R Core Team, 2024) program. Lastly, TASSEL software was employed to convert the .vcf file to PHYLIP format with the argument *-exportType Phylip_Inter*.

2.3 Genetic diversity and population structure

Genetic diversity indices were calculated for each race of maize and for the OP and hybrid cultivars using the *adegenet* v2.1.10 (Jombart, 2008; Jombart and Ahmed, 2011) and the HIERFSTAT v0.5-11 (Goudet, 2005) R packages. Bootstrapped 95% confidence intervals were estimated by running 10,000 bootstraps (BS) using HIERFSTAT package to determine if F_{IS} was significantly different from zero. A maximum likelihood (ML) tree was constructed using the .phy file with the multi-threaded version of the program RAxML v8.2.11 (Stamatakis, 2014), *raxmlHPC-PTHREADS*, with the rapid bootstrapping algorithm and a total of 100 nonparametric BS inferences. Model *ASC_GTRGAMMA* with the ascertainment correction of Lewis (Lewis, 2001) was also considered, and the resulting tree was plotted with the *ggtree* (Yu et al., 2017) R package. Moreover, genetic distances based on Provesti's coefficient (Prevesti et al., 1975) were calculated, and then a dendrogram was generated using the neighbor-joining clustering algorithm with 100 BS replicates from the *poppr* v.1.1.4 (Kamvar et al., 2014, 2015) package in R. A principal coordinate analysis (PCoA) was performed with the *dudi.pco* function of the *ade4* v1.7.22 (Dray and Dufour, 2007) package in R. To determine the population structure, first the filtered .vcf file was converted into .str format with VCFtools and PGDSpider v2.1.1.5 (Lischer and Excoffier, 2012) programs.

We then employed the Bayesian clustering program STRUCTURE v2.3.4 (Pritchard et al., 2000) with populations (K) of 1 to 25 and 10 replicates. A burn-in length of 50,000 with 100,000 Monte Carlo iterations was considered, and the optimal K value was estimated by the Evanno method (Evanno et al., 2005). Population structure was visualized with POPHELPER v2.3.1 (Francis, 2017) R package. Genetic diversity indices were also estimated for the clusters determined by STRUCTURE. The Evanno method determined that the best K (number of populations) is two for our data set, and the next two largest peaks are at $K = 4$ and $K = 10$

(Supplementary Figure S1). Previous research in capirona (Saldaña et al., 2021), carrot (Iorizzo et al., 2013; Arbizu et al., 2016), and maize (Vigouroux et al., 2008) found a false highest peak at $K = 2$ in population structure analysis as the null hypothesis of no structure ($K = 1$) was strongly rejected. In addition, Waples and Gaggiotti (2006); Frantz et al. (2009), and Janes et al. (2017) indicated that the Evanno method tends to underestimate the number of genetic clusters. Hence, it is very likely the second highest peak obtained with our dataset of 14,235 SNPs ($K = 4$) is caused by a strong rejection of the hypothesis of three clusters only. Furthermore, the ML value was obtained at $K = 10$ (Supplementary Figure S2), which is concordant also with our ML analyses of the Peruvian maize. Hence, we decided to discuss our results with $K = 10$.

An analysis of molecular variance (AMOVA) with the *poppr* (Kamvar et al., 2014) package in R to determine the sources of genetic variance within and among the races of maize was also conducted. To evaluate statistical significance, a randomization test with 999 permutations was performed using the *randtest* function of the *ade4* (Dray and Dufour, 2007) package. Finally, a pairwise fixation index (F_{ST}) was estimated using the R package HIERFSTAT, according to Weir and Cockerham (1984).

3 Results

3.1 Sequencing analysis and SNPs distribution

After filtering out the raw reads, the total demultiplexed reads for all 423 genotypes were 1,566.7 M with good barcoded reads representing 99.9%, and the average read per accession was 3.7 M. A total of 5,010,502 tags were identified, of which 86.4% uniquely aligned to the maize reference genome. Next, we detected a total of 1,002,078 raw SNPs after using the TASSEL software, and we kept 31,132 SNPs after filtering with VCFtools program. A set of 14,235 SNPs distributed across the 10 chromosomes of maize was selected after LD pruning in R, which was used for subsequent genetic structure and diversity analysis. The highest and lowest numbers of physically mapped SNPs were identified in chromosome 1 (2170, 15.2%) and 10 (954, 6.7%), respectively (Table 1). The 10 maize chromosomes exhibited a very consistent distribution of SNP markers that spanned virtually the whole genome, showing a low SNP density near the centromeres, whereas the telomere region exhibited a high density of SNPs (Supplementary Figure S3). Chromosome 1 possessed the highest density (7.04 SNPs/Mb), and chromosome 10, with 6.26 SNPs/Mb, the lowest (Supplementary Figure S3).

3.2 Population structure

The PCoA showed that the first and second axis explained 2.0% and 1.5% of the variance, respectively. Sub-race Pachia and the two types of improved maize (purple and yellow cultivars) are separated into two well-distinctive groups. On the contrary, there is not a consistent grouping of the other amylaceous maize. Accessions of race Cusco Gigante are closely related to some individuals of Cusco,

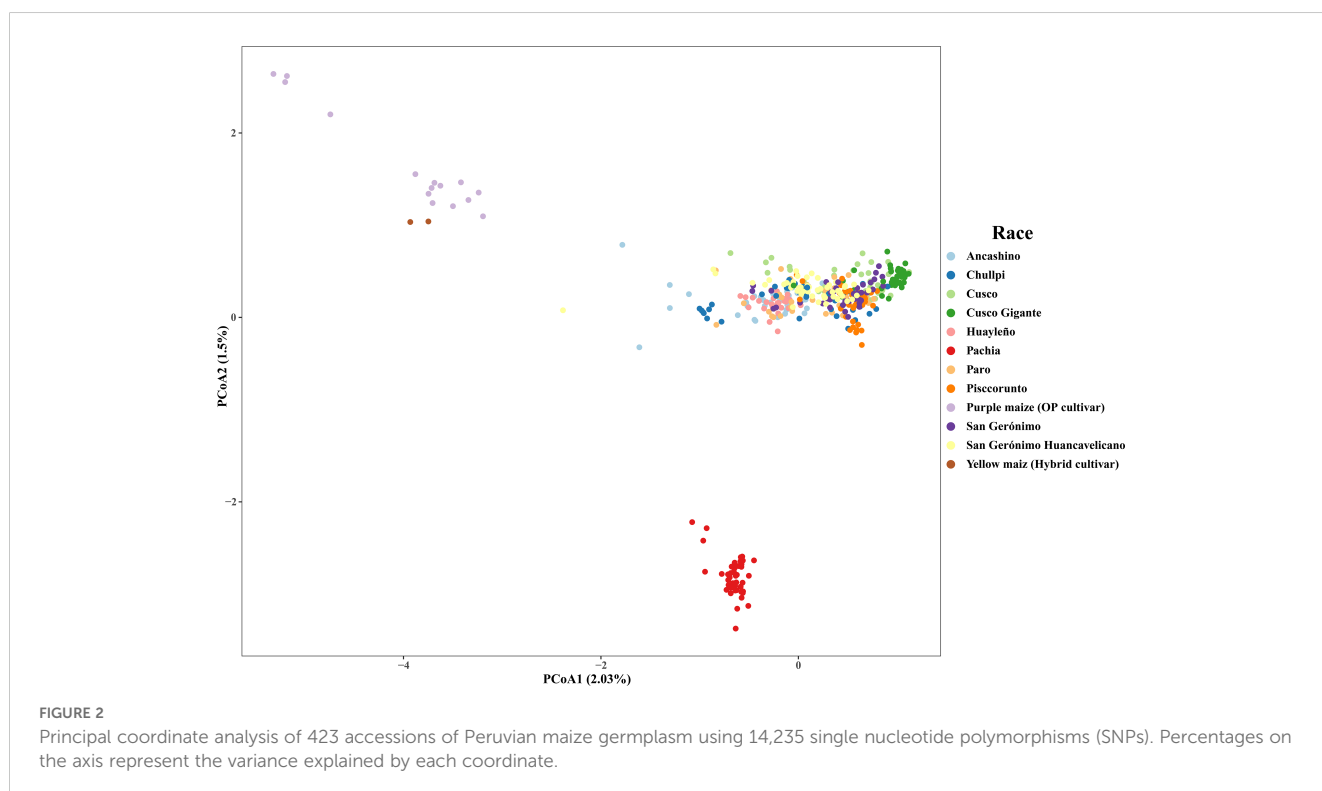
TABLE 1 Genome-wide distribution and density of 14,235 single nucleotide polymorphisms (SNPs) across the 10 chromosomes of maize.

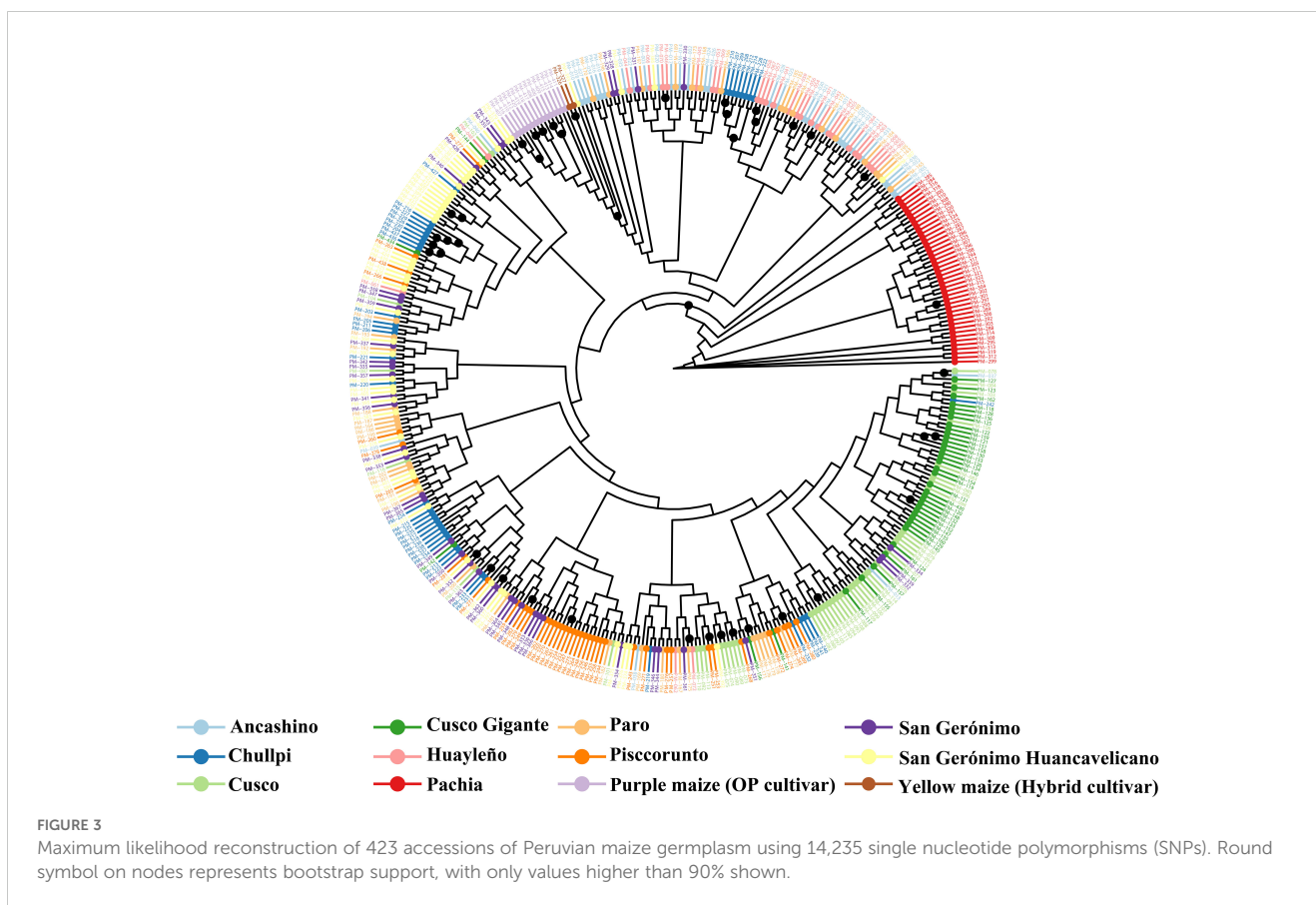
Chromosome	SNP #	SNP %	Total length (Mb)	Density (SNPs/Mb)
1	2170	15.2	308.45	7.04
2	1793	12.6	243.68	7.36
3	1631	11.5	238.02	6.85
5	1615	11.3	226.35	7.13
4	1410	9.9	250.33	5.63
7	1260	8.9	185.81	6.78
8	1229	8.6	182.41	6.74
6	1109	7.8	181.36	6.12
9	1064	7.5	163	6.53
10	954	6.7	152.44	6.26

and races Ancashino, Chullpi, Huayleño, and Paro are grouping together, but without a clear resolution (Figure 2). Deletion of improved maize slightly changed the structure in the PCoA, showing that most accessions of race Chullpi are separated, and race Cusco Gigante and most accessions of Cusco are roughly grouping together (Supplementary Figure S4). When accessions of sub-race Pachia were also deleted, it was possible to detect a better grouping of race Cusco Gigante. Similarly, races Ancashino, Huayleño, and Paro tend to cluster together. In addition, eight accessions of race Chullpi are separated from the other individuals of maize; some accessions of race San Gerónimo Huancavelicano are separated too but mixed with very few individuals of Piscorunto and Cusco Gigante (Supplementary Figure S5).

Accessions of race Ancashino and Huayleño are grouping together, but some other accessions are intermixed in this group. The same feature was shown by another group of some accessions of race Piscorunto. However, most accessions cannot be clearly separated by race criteria. Our ML tree recovered two main clades containing the following maize landraces: (CR1) almost all accessions of Ancashino, Huayleño, and Paro, some accessions of Chullpi, and a few of San Gerónimo and San Gerónimo Huancavelicano, and (CR2) all accessions of Cusco Gigante, Cusco, and Piscorunto, some accessions of Chullpi, San Gerónimo, and San Gerónimo Huancavelicano (Figure 3).

In addition, all individuals of improved maize were placed in a subcluster within CR1. Interestingly, the CR1 mainly comprises the





anciently derived or primary races (ADPR) of maize in Peru, and CR2 consists of the lately derived or secondary races (LDSR), according to the classification described by Grobman et al. (1961) (Figure 3). There are, however, few exceptions: (i) races Cusco and Piscorunto are considered an ADPR, but they are not within the CR1 clade but in CR2 clade, (ii) initially defined as an imperfectly defined race by Grobman et al. (1961), San Gerónimo should be considered as an LDSR. A clade of two accessions of Ancashino (PM-005, PM-012) and one of Paro (PM-165) races is sister to those two clades, and sister to them there is a grade comprising sub-race Pachia. Our ML tree also resolved a subclade within CR2 containing almost all accessions of race Cusco Gigante and Cusco. Moreover, we observed three subclusters of race Chullpi, one of them with above 90% BS within CR1. One grade comprising 10 accessions of race San Gerónimo Huancavelicano, including one Chullpi maize (PM-427), was also detected. For most accessions labeled by their race, a consistent grouping pattern was not noticed. The topology of our neighbor-joining dendrogram differs to some degree to the ML tree. Only clade CR1 was recovered, and a polytomy was present (Supplementary Figure S6).

A clear grouping was not observed when maize accessions were labeled based on their Peruvian geographic department of origin, except for accessions from Tacna, who form a grade with above 90% BS. Maize from Ancash also showed another grade, but accessions from other locations are also intermingled; similarly, another grade formed by maize from Huancavelica and Junín (Supplementary Figure S7). Interestingly, when accessions were labeled according to

their geographic zone of origin (north, center, south) in Peru, our ML tree revealed the following two clades: (CZ1) individuals from the northern Andes (Ancash and Cajamarca) and purple maize OP lines obtained in Lima, (CZ2) individuals from the center (Huancavelica, Huánuco, Junín) and southern (Apurímac, Ayacucho, Cusco, Moquegua) of the Peruvian Andes; both also contained within their corresponding subclades. These two clades possess a BS <90%, and very few accessions from other geographic zones are intermingled. A similar pattern was detected in our PCoA (Supplementary Figure S8).

STRUCTURE analysis showed abundant admixture, except for the accession of sub-race Pachia, whose accessions were placed in cluster (C) 7, thus exhibiting very low admixture (Figure 4). Furthermore, Cusco Gigante and Cusco races are clustering together; that is, most of those accessions are within cluster 4. Like Ancashino and Huayleño, which are forming a group (cluster 5) but with some degree of admixture, improved maize is also clustering together (C2). Chullpi accessions are mainly distributed between clusters 1 and 9, and race Paro was placed in clusters 1, 3, and 5. San Gerónimo Huancavelicano is grouped within clusters 1 and 9 mainly, whereas race San Gerónimo was placed mainly in cluster 1. For $K = 2$, yellow (hybrid) and purple maize (OP) were the only groups (C2) clearly differentiated from the other maize. Admixture was also observed for $K = 4$, and similar to $K = 2$ and $K = 10$, bred maize clearly grouped together (C3). Furthermore, subrace Pachia was also differentiated (C1), and a grouping of most accessions of races Cusco Gigante, Cusco, Piscorunto, Paro, and San Gerónimo was detected

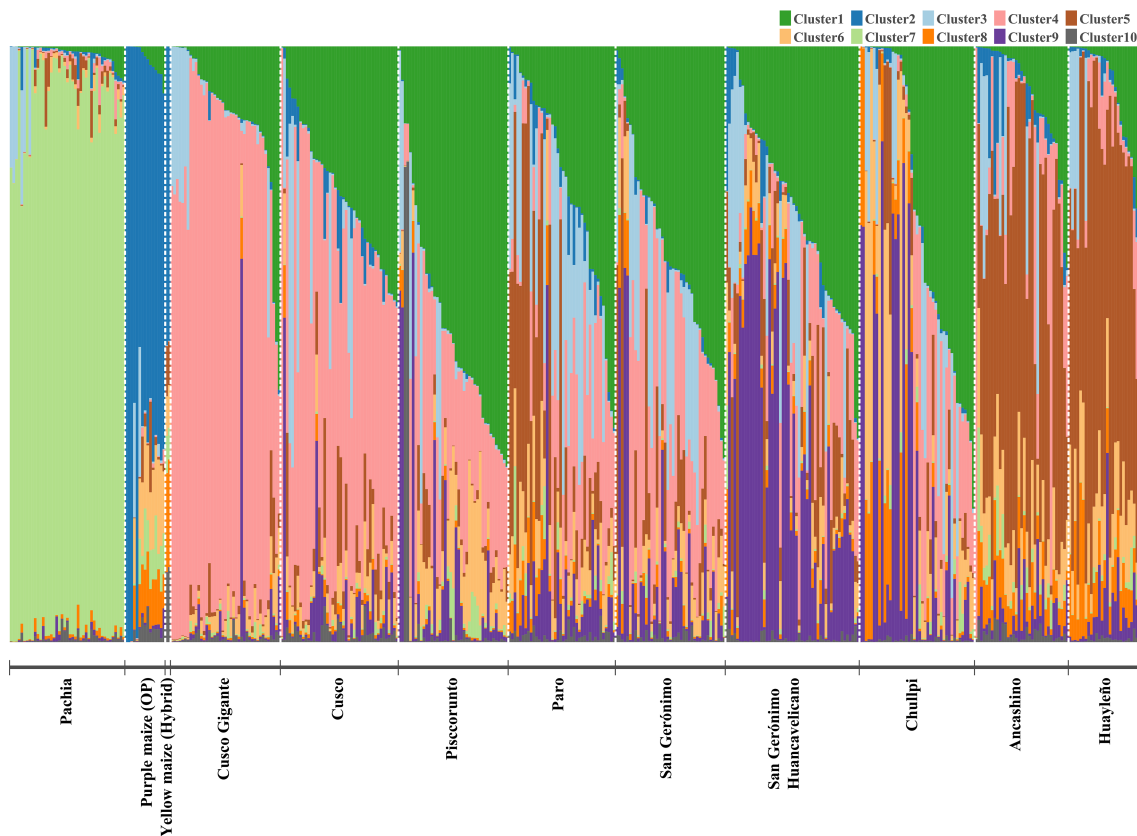


FIGURE 4 Population structure of 423 maize accessions based on 14,235 single nucleotide polymorphisms (SNPs). Each accession is represented by a vertical bar, and each color corresponds to a population (10 in total).

(C4). On the other hand, C2 consisted of most accessions of races Ancashino, Huayleño, Chulpi, and San Gerónimo Huancavelicano (Supplementary Figure S9). An erratic behavior for ML values for K exceeding 11 was noticed (Supplementary Figure S2). This requires further methodological research, as perhaps large SNP data sets require more iterations in burn-in.

There was not a clear cluster assignment when accessions were labeled according to their geographic origin, except for maize from Ancash, Huánuco, Moquegua, Tacna, and Lima (bred maize). The geographic zone criterion of clustering exhibited that most accessions from northern Peru are grouped together (C5), while those from Lima were placed in cluster 2. Maize accessions from the center of Peru were mainly grouped within clusters 1, 4, and 9. Clusters 1, 4, and 7 contained accessions

of maize from southern Perú (Supplementary Table S2, Supplementary Figure S10). Clusters 8 and 10 possessed one and two individuals only, respectively; therefore, results of these clusters will not be described.

Fixation indices (F_{ST}) were very low in general. Population divergence between yellow maize (improved maize) and Cusco Gigante revealed the highest genetic difference (0.18), while races Huayleño and Ancashino exhibited the lowest (0.001), which might not be significantly different from zero (Supplementary Table S3). Furthermore, the greatest genetic variation was observed within races of Peruvian maize (95.48%), while 4.52% was reported for between races, according to our AMOVA. A significant PhiPT value of 0.045 (p -value < 0.001) was obtained (Table 2).

TABLE 2 Analysis of molecular variance of the genetic variation for 423 accessions of Peruvian maize germplasm using 14,235 SNPs.

Source of variation	df	SS	MS	Est. var.	%	PhiPT	p -value
Between races	11	179,018.3	16,274.39	290.99	4.52	0.045	0.001
Within races	411	2,526,998.7	6148.42	6148.42	95.48		
Total	422	2,706,017.0	6412.36	6439.41	100		

df, degree of freedom; SS, sum of squares; MS, mean squares; Est. Var., estimated variance; %, percentage of genetic variation; PhiPT, differentiation statistics; p -value, probability of accepting the null hypothesis.

3.3 Genetic diversity

The nine races and one subrace of Peruvian maize examined in this work showed a very similar number of different alleles; the allelic richness ranged from 1.33 ± 0.15 (subrace Pachia) to 1.36 ± 0.12 (race Ancashino). In addition, the sub-race Pachia possessed the lowest observed heterozygosity (H_O) (0.25 ± 0.17), whereas the race Cusco had the highest (0.28 ± 0.16), and the expected heterozygosity (H_E , genetic diversity) ranged from 0.33 ± 0.14 (Pachia) to 0.36 ± 0.12 (Ancashino). On the other hand, Cusco (0.21) exhibited the lowest inbreeding coefficient (F_{IS}), while Ancashino the highest (0.27) (Table 2). Cluster 2 (1.33 ± 0.16) and 5 (1.36 ± 0.11) possessed the lowest and highest allelic richness, respectively. The lowest H_O was exhibited by C2 (0.2 ± 0.16), whereas the highest was by C1 (0.28 ± 0.15). In addition, C2 (0.33 ± 0.16) also showed the lowest H_E , while C5 (0.36 ± 0.11) showed the highest. Interestingly, clusters containing mainly maize landraces

from Apurímac (C3) and Ayacucho (C1) possessed very high levels of genetic diversity (0.35 – 0.36). F_{IS} ranged from 0.21 (C4) to 0.4 (C2) (Table 3).

4 Discussion

The foundation for crop improvement lies in genetic diversity (Hoisington et al., 1999; Bhandari et al., 2017), which can be assessed by DNA (molecular) markers like SNPs. Analyzing the molecular genetic variation in germplasm provides valuable insights into allelic richness, population structure, and diversity parameters. This information helps plant breeders utilize genetic resources more effectively, reducing the need for extensive pre-breeding tasks when developing new cultivars (Govindaraj et al., 2015). In recent years, due to the advances in next-generation sequencing (NGS), GBS has emerged as a promising genomic approach for estimating plant

TABLE 3 Genetic diversity indices of Peruvian maize germplasm based on 14,235 single nucleotide polymorphisms (SNPs).

Race (or cultivar type)	Number of accessions	N_A	A_R	H_O	H_E	F_{IS}
Ancashino	35	2.000 ± 0.01	1.360 ± 0.12	0.263 ± 0.16	0.362 ± 0.12	0.274 [0.27;0.28]
Chullpi	43	1.999 ± 0.02	1.353 ± 0.12	0.264 ± 0.16	0.354 ± 0.12	0.254 [0.25;0.26]
Cusco	44	2.000 ± 0.02	1.349 ± 0.12	0.278 ± 0.16	0.350 ± 0.12	0.205 [0.2;0.21]
Cusco Gigante	41	1.996 ± 0.06	1.336 ± 0.13	0.262 ± 0.17	0.337 ± 0.13	0.223 [0.22;0.23]
Huayleño	28	1.998 ± 0.05	1.355 ± 0.13	0.265 ± 0.16	0.357 ± 0.13	0.257 [0.25;0.26]
Pachia	43	1.988 ± 0.11	1.332 ± 0.15	0.249 ± 0.17	0.333 ± 0.14	0.253 [0.25;0.26]
Paro	40	2.000 ± 0.01	1.356 ± 0.12	0.277 ± 0.16	0.357 ± 0.12	0.225 [0.22;0.23]
Piscorunto	41	1.999 ± 0.03	1.346 ± 0.12	0.274 ± 0.17	0.347 ± 0.12	0.211 [0.2;0.22]
Purple corn (OP)	15	1.914 ± 0.28	1.313 ± 0.17	0.190 ± 0.17	0.318 ± 0.17	0.401 [0.39;0.41]
San Gerónimo	41	1.999 ± 0.02	1.350 ± 0.12	0.267 ± 0.16	0.351 ± 0.12	0.238 [0.23;0.24]
San Gerónimo Huancavelicano	50	2.000 ± 0.01	1.354 ± 0.12	0.270 ± 0.16	0.355 ± 0.12	0.239 [0.23;0.24]
Yellow maize (hybrid)	2	1.440 ± 0.52	NA	0.230	0.297	0.227
Cluster (C)						
C1	112	2.000 ± 0	1.349 ± 0.12	0.274 ± 0.15	0.350 ± 0.12	0.217 [0.21;0.22]
C2	18	1.948 ± 0.22	1.325 ± 0.16	0.198 ± 0.16	0.329 ± 0.16	0.399 [0.39;0.4]
C3	31	1.999 ± 0.03	1.353 ± 0.12	0.263 ± 0.17	0.355 ± 0.12	0.258 [0.25;0.26]
C4	91	2.000 ± 0.01	1.342 ± 0.12	0.272 ± 0.16	0.343 ± 0.12	0.206 [0.2;0.21]
C5	76	2.000 ± 0	1.359 ± 0.11	0.268 ± 0.15	0.360 ± 0.11	0.255 [0.25;0.26]
C6	7	1.852 ± 0.35	1.326 ± 0.18	0.247 ± 0.22	0.333 ± 0.19	0.258 [0.25;0.27]
C7	43	1.988 ± 0.11	1.332 ± 0.14	0.249 ± 0.17	0.333 ± 0.14	0.253 [0.25;0.26]
C8	1	0.019 ± 0.14	NA	0.004 ± 0.06	0.004	-0.001
C9	42	1.998 ± 0.04	1.346 ± 0.13	0.262 ± 0.16	0.348 ± 0.13	0.246 [0.24;0.25]
C10	2	1.342 ± 0.49	NA	0.250 ± 0.38	0.190 ± 0.26	-0.311 [-0.43;-0.39]

N_A , number of different alleles; A_R , allelic richness; H_O , observed heterozygosity; H_E , expected heterozygosity; F_{IS} , inbreeding coefficient. Values after \pm represent standard deviation. Number in brackets correspond to 95% confidence intervals. NA, not applicable as number of samples is limited.

genetic diversity and population structure on a genome-wide scale and has been successfully employed, *inter alia*, in *Brassica* (McAlvay et al., 2021), *Daucus* (Arbizu et al., 2016; Mezghani et al., 2018; Martínez-Flores et al., 2020), finger millet (Brhane et al., 2022), maize (Swarts et al., 2017; Shu et al., 2021; Dube et al., 2023), spruce (Korecký et al., 2021), wheat (Alipour et al., 2017), and watermelon (Lee et al., 2019). However, GBS has not been used so far for genotyping Peruvian races of maize, whose morphological diversity seems to be the largest worldwide (Grobman et al., 1961; Ministerio del Ambiente, 2018). Studies on Peruvian maize have primarily focused on morpho-agronomic characteristics (Ortiz et al., 2008b, 2008c; López Alejandría, 2011; Chavarry Gómez, 2014; Macuri Núñez, 2016; Gamarra Sánchez et al., 2020; Chambergo Zúñiga, 2021; García Mendoza et al., 2021; Mamani-Huarcaya et al., 2022; Prieto Rosales and Manayay Sánchez, 2022; García-Mendoza et al., 2023), leaving their molecular composition largely unexplored. Herein, we determined the gene diversity and composition of Peruvian maize races from the Andean highlands by means of SNP markers spanning each chromosome.

Unfortunately, despite the significant diversity within Peruvian maize germplasm, knowledge of its genetic components remains very limited. Catalán et al. (2019) reported a high level of variability using eight microsatellites in 83 accessions of six races of maize from Cusco. The genetic diversity of the nine races and one subrace of maize assessed in this study is very high, which is concordant with its improvement status (i.e., landraces), as reported for other landraces of beans (Özkan et al., 2022), peas (Martín-Sanz et al., 2011), squash (Lorello et al., 2020), tarwi (Huaríngá-Joaquín et al., 2023), or wheat (Tehseen et al., 2022), among others. Our genetic diversity indices align with other studies on maize landraces. For example, Warburton et al. (2008) examined the genetic diversity of 24 maize landraces from Mexico with 25 simple sequence repeats (SSR) and reported a total gene diversity of 0.61 across all populations. Similarly, Herrera-Saucedo et al. (2019) determined the genetic variability of 63 native maize accessions from northern Mexico using 31 SSRs, reporting an expected heterozygosity of 0.68. A study of 30 maize landrace accessions from the southern Andean region of South America using 22 SSRs showed a genetic diversity of 0.72 (Rivas et al., 2022).

In a more comprehensive study (Vigouroux et al., 2008), employing 96 SSRs that encompass most of the described races in the American continent, 136 accessions of 47 races of Peruvian maize were included, and these, together with other maize from Ecuador and Bolivia (a total of 235 plants), possessed a total gene diversity of 0.71. Here, we determined that the genetic diversity of the Peruvian maize (0.35) from more diverse geographic regions is higher than the value reported for 46 Mexican landraces (161 accessions) of maize using SNPs (0.311) (Arteaga et al., 2016), demonstrating that Peru possesses one of the largest genetic diversities of amylaceous maize, pointing to the fact that the central Andean region possesses abundant maize genetic variability. A recent study (Arca et al., 2023) found that gene diversity of landraces from seven countries from South America, assessed with 23,412 SNPs, was slightly lower (0.323 ± 0.007) than landraces from Central America and Mexico (0.328 ± 0.006). The higher gene diversity reported with SSR compared to SNP markers

may be due to the multi-allelic nature and higher level of polymorphism of SSR compared to bi-allelic SNP. However, SNPs are more reliable for inferring genome-wide genetic diversity, as demonstrated by previous work (Fischer et al., 2017; García et al., 2018). Cluster 5, containing mainly landraces of Ancashino, Hualyeño, and Paro from Ancash, Apurímac, and Ayacucho, exhibited the highest gene diversity. In addition, C3 and C1, which are mainly comprised of maize landraces from Apurímac and Ayacucho, respectively, also showed very high levels of genetic variability. This result is in agreement with Salhuana (2004), as he indicated that maize from the geographical departments of Ancash and Ayacucho exhibited a greater degree of variability and a more contrasting pattern of allele frequencies than all other maize from other Peruvian locations. It is very likely maize landraces from Ancash, Apurímac, and Ayacucho were involved in the origin of Andean maize. Furthermore, Ancash is the worldwide center of brown and red coloration of pericarp and cob of maize (Grobman et al., 1961; Salhuana, 2004).

Consistent with previous investigations (Warburton et al., 2008; Salazar et al., 2017), bred maize is clearly separated from Peruvian maize landraces, which is explained by their intensity of selection. The well-defined grouping of accessions of sub-race Pachia may be explained in the light of its cultivation in a restricted area in southern Peru (Valley of Pachia, Tacna). Although Grobman et al. (1961) indicated that this sub-race derived from the race Arequipeño, which is a lately derived race, our molecular data, however, do not support this fact, as Pachia was not placed within the CR2 clade. Instead, it is more likely related to race Coruca, which also grows in Tacna and is similar to a floury maize landrace Choclero from Chile (Grobman et al., 1961). Further research is needed, including maize samples from other southern Peruvian regions (Arequipa, Moquegua, Puno, and Tacna), as landraces of maize cultivated in Tacna show tolerance to high levels of boron (Mamani-Huarcaya et al., 2020, 2022), which is a trait of interest for breeding maize for locations with high levels of this element in soil and irrigation water. Landrace Cabanita, widely grown in Arequipa, also shows potential as a source of phenolic compounds with *in-vitro* antioxidant capacity (Fuentes-Cardenas et al., 2022). Races Cusco Gigante and Cusco tend to group together as they are mainly cultivated in Cusco. Additionally, races Ancashino and Huayleño, both sympatrically distributed in northern Peru (Ancash), group together and possess a very low F_{ST} , suggesting they evolved simultaneously. Hybridization likely plays a role in the grouping of these races, as noted by Grobman et al. (1961).

Even though the other Peruvian maize races are morphologically distinct, our GBS dataset failed to support them as monophyletic, which agrees with other research that evaluated Peruvian germplasm (Matsuoka et al., 2002; Vigouroux et al., 2008; Bracco et al., 2016; Bedoya et al., 2017; Arca et al., 2023). Similarly, Mexican maize races do not form distinct clusters (Arteaga et al., 2016). However, Caldu-Primo et al. (2017) were able to distinguish Mexican maize races based on a high F_{ST} SNP dataset. Population structure analysis clustered maize races from the American continent based on geographic origin, with the Peruvian germplasm contained within a clade named “Andean” (Matsuoka et al., 2002; Vigouroux et al., 2008; Bracco et al., 2016; Bedoya et al.,

2017; Arca et al., 2023). More consistent clustering was observed when Peruvian races of maize were labeled according to their geographical zones of origin, identifying CZ1 and CZ2. This mixing among maize landraces is likely due to extensive gene flow within these zones explained by their proximities and frequent seed exchange, which is a common practice in the Peruvian Andes. However, Peruvian maize farmers in the Andes usually dynamize seed flow between families and rural communities and conduct selection within their populations to maintain the morphological characteristics of their landraces. Our ML tree mostly agrees with the classification of Peruvian maize races based on the chronological origin described by Grobman et al. (1961) as it was possible to reconstruct the ADPR (CR1) and LDSR (CR2) clades. Races Chullpi and Paro possess very low F_{ST} (0.004), reflecting very low genetic differentiation. Moreover, the phylogenetic position of race Chullpi warrants additional research to determine its origin. The similar traits that race Chullpi from different geographical departments of Peru exhibits are very likely due to environmental pressures, as the climatic gradients of the Andes are laboratories of constant plant evolution. Similarly, further research is needed for races Cusco Gigante, Cusco, Piscorunto, San Gerónimo, and San Gerónimo Huancavelicano, as it is very likely their phenotype is a result of evolution in the Andes of Peru of a set of novel or derived traits.

A more detailed morphological evaluation is needed for the races of Peruvian maize to identify morphotypes and determine their phenetic plasticity. It is very likely that two accessions of Ancashino (PM-005, PM-012) and one of Paro (PM-165) contributed to the origin of other maize races evaluated in this work, as these individuals form a sister clade to CR1 and CR2. Both races are considered ADPR, directly derived from the primitive races (PR), as described by Grobman et al. (1961). Therefore, it is very likely these three accessions may still possess genetic signatures of PR. However, further research is necessary, including samples from a wider geographical area, for more conclusive results. The position of purple maize OP cultivars within CR1 is explained by its origin in race Kullpi, classified as ADPR by Grobman et al. (1961). The close relation among purple (OP) and yellow maize (hybrid) is due to the use of the latter at UNALM to enhance the yield performance of OP lines.

The vast diversity exhibited by maize races in Peru is crucial for research in plant genetic resources (Ortiz et al., 2008a). However, the lack of relevant information on the genetic diversity of conserved plant material hinders the use of accessions preserved in germplasm banks (Ortiz and Engels, 2018; McCouch et al., 2020). To address this gap for Peruvian maize, we suggest that genomic tools can facilitate the characterization and utilization of this invaluable plant genetic resource, as emphasized by Mascher et al. (2019). This approach is particularly important considering that amylaceous maize in Peru are landraces, dynamic populations in constant evolution in the Andes. Moreover, Peruvian maize germplasm requires special attention as a ~5500 cal. BP maize cob from northern Peru (Grobman et al., 2012) was the only sample without the *Zea mays* ssp. *mexicana* ancestry in a recent study (Yang et al., 2023), shedding light on the origin of maize in the Central Andean region. Swarts et al. (2017) did find introgression

with *Z. mays* spp. *mexicana* in modern South American landraces. On the other hand, we expect our study will provide useful guides to researchers and decision makers for establishing a strong conservation strategy and dynamic utilization soon for Peruvian maize germplasm.

Data availability statement

The dataset generated and/or analyzed during the current study are available in the Dryad repository, <https://datadryad.org/stash/share/4OI0AXLoEIDsbSSLvoOdXrhzSyMUAow9x0elGjhyPmQ>. Additional information is presented in the Supplementary Data file accompanying this article.

Author contributions

CA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Validation, Writing – original draft, Writing – review & editing. IB-S: Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. JF: Data curation, Investigation, Validation, Writing – original draft. RO: Conceptualization, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing. RB: Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing. PG-M: Funding acquisition, Investigation, Validation, Writing – review & editing. RS: Conceptualization, Investigation, Writing – review & editing. JC: Investigation, Methodology, Writing – review & editing. AG: Investigation, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by project N°03-2017 of STC-CGIAR (Ministry of Agrarian Development and Irrigation of the Peruvian Government): Mejoramiento de nuevas variedades de maíz amiláceo explotando el germoplasma: uso de secuenciamientos de ADN de última generación y fenotipado en campos experimentales.

Acknowledgments

The authors wish to acknowledge Fabiola Catalán, Hugo Huanuqueño, Gilberto Rodríguez, and Fatima Silva for providing germplasm and technical support, and Wilian Salazar for helping prepare Figure 1 and Supplementary Figure S10. We thank the University of Minnesota Genomics Center for providing facilities and services. The authors also thank the Bioinformatics High-performance Computing server of Universidad Nacional Agraria la Molina (BioHPC-UNALM) and the University of Wisconsin–Madison for providing resources to perform the analyses. CA

thanks Vicerrectorado de Investigación of UNTRM and PROCENCIA (Contract N° PE501085118-2023) for providing financial support for a stay at the Department of Plant Breeding, Swedish University of Agricultural Sciences.

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.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1526670/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Plot of Delta K (ΔK). Twenty-five populations were considered in a data set of 14,235 SNPs and 423 accessions of Peruvian maize germplasm.

SUPPLEMENTARY FIGURE 2

Plot of the log likelihood for K ranging from 1 to 25.

SUPPLEMENTARY FIGURE 3

Density and distribution of 14,235 single nucleotide polymorphism markers on the 10 maize chromosomes. Red bars denote the end of chromosome; nt refers to nucleotide.

SUPPLEMENTARY FIGURE 4

Principal coordinate analysis of 406 accessions of Peruvian maize germplasm (improved maize not included) using 14,235 single nucleotide polymorphism markers. Percentages on the axis represent the variance explained by each coordinate.

SUPPLEMENTARY FIGURE 5

Principal coordinate analysis of 363 accessions of Peruvian maize germplasm (improved maize and sub-race *Pachia* not included) using 14,235 single nucleotide polymorphism markers. Percentages on the axis represent the variance explained by each coordinate.

SUPPLEMENTARY FIGURE 6

Dendrogram based on Provesti's genetic distance and the neighbor-joining clustering method of 423 accessions of Peruvian maize germplasm using 14,235 single nucleotide polymorphisms markers. Numbers above the branches represent bootstrap values, with only values higher than 90% shown.

SUPPLEMENTARY FIGURE 7

(A) Principal coordinate analysis of 423 accessions of Peruvian maize germplasm using 14,235 single nucleotide polymorphism markers. Percentages on the axis represent the variance explained by each coordinate. (B) Maximum likelihood reconstruction of 423 accessions of Peruvian maize germplasm using 14,235 single nucleotide polymorphism markers. Round symbol on nodes represents bootstrap support, with only values higher than 90% shown. Accessions were labeled according to their Peruvian geographic department of origin.

SUPPLEMENTARY FIGURE 8

(A) Principal coordinate analysis of 423 accessions of Peruvian maize germplasm using 14,235 single nucleotide polymorphism markers. Percentages on the axis represent the variance explained by each coordinate. (B) Maximum likelihood reconstruction of 423 accessions of Peruvian maize germplasm using 14,235 single nucleotide polymorphism markers. Round symbol on nodes represents bootstrap support, with only values higher than 90% shown. Accessions were labeled according to their geographic zone of origin.

SUPPLEMENTARY FIGURE 9

(A) Population structure of 423 maize accessions based on 14,235 single nucleotide polymorphism markers. Each accession is represented by a vertical bar, and each color corresponds to a population (two in total). (B) Population structure of 423 maize accessions based on 14,235 single nucleotide polymorphism markers. Each accession is represented by a vertical bar, and each color corresponds to a population (four in total).

SUPPLEMENTARY FIGURE 10

Cluster assignment based on STRUCTURE results for Peruvian maize accessions according to their geographic origin.

References

- Alipour, H., Bihanta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., and Zhang, G. (2017). Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01293
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Accessed April 8, 2024).
- Arbizu, C. I., Ellison, S. L., Senalik, D., Simon, P. W., and Spooner, D. M. (2016). Genotyping-by-sequencing provides the discriminating power to investigate the subspecies of *Daucus carota* (Apiaceae). *BMC Evol. Biol.* 16, 1–16. doi: 10.1186/s12862-016-0806-x
- Arca, M., Gouesnard, B., Mary-Huard, T., Le Paslier, M. C., Bauland, C., Combes, V., et al. (2023). Genotyping of DNA pools identifies untapped landraces and genomic regions to develop next-generation varieties. *Plant Biotechnol. J.* 21, 1123–1139. doi: 10.1111/pbi.14022
- Arteaga, M. C., Moreno-Letelier, A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña-Ochoa, A., Moreno-Estrada, A., et al. (2016). Genomic variation in recently collected maize landraces from Mexico. *Genom. Data* 7, 38–45. doi: 10.1016/j.gdata.2015.11.002
- Bedoya, C. A., Dreisigacker, S., Hearne, S., Franco, J., Mir, C., Prasanna, B. M., et al. (2017). Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PLoS One* 12, e0173488. doi: 10.1371/journal.pone.0173488
- Bhandari, H. R., Nishant Bhanu, A., Srivastava, K., Singh, M. N., Shreya, and Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants - an overview. *Adv. Plants Agric. Res.* 7, 255. doi: 10.15406/apar.2017.07.00255

- Bonavia, D. (2008). *El maíz: su origen, su domesticación y el rol que ha cumplido en el desarrollo de la cultura* (Lima: Universidad de San Martín de Porres, Fondo Editorial).
- Bracco, M., Cascales, J., Hernández, J. C., Poggio, L., Gottlieb, A. M., and Lia, V. V. (2016). Dissecting maize diversity in lowland South America: genetic structure and geographic distribution models. *BMC Plant Biol.* 16, 186. doi: 10.1186/s12870-016-0874-5
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Brhane, H., Haileselassie, T., Tesfaye, K., Ortiz, R., Hammenhag, C., Abreha, K. B., et al. (2022). Finger millet RNA-seq reveals differential gene expression associated with tolerance to aluminum toxicity and provides novel genomic resources. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1068383
- Brieger, F. G., Gurgel, J. T. A., Paterniani, E., Blumenschein, A., and Alleoni, M. R. (1958). *Races of maize in Brazil and other eastern South American countries* (Washington D.C: National Academy of Sciences, National Research Council).
- Caldu-Primo, J. L., Mastretta-Yanes, A., Wegier, A., and Piñero, D. (2017). Finding a needle in a haystack: distinguishing Mexican maize landraces using a small number of SNPs. *Front. Genet.* 8. doi: 10.3389/fgene.2017.00045
- Catalán, F., Blas, R., Catalán, W., and Pompeyo, C. (2019). "Diversidad genética de seis razas de maíz (*Zea mays* L.) de la región Cusco," in *Cuarto Congreso Peruano de Mejoramiento Genético y Biotecnología Agrícola: Perspectivas de las ómicas para enfrentar los retos de la agricultura* (Universidad Nacional Agraria la Molina, Lima).
- Chambergó Zúñiga, K. (2021). *Variabilidad fenotípica de maíz amiláceo (*Zea mays* L.) dentro y entre las razas Cuzco Gigante, Cuzco y Chullpi* (Junín: Universidad Nacional del Centro del Perú).
- Chavarry Gómez, B. (2014). *Caracterización morfológica de una muestra de accesiones de maíces Peruanos del Banco de Germoplasma de maíz (*Zea mays* L.) de la UNALM* (Lima: Universidad Nacional Agraria la Molina).
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Dray, S., and Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi: 10.18637/jss.v022.i04
- Dube, S. P., Sibiya, J., and Kutu, F. (2023). Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers. *Sci. Rep.* 13, 17851. doi: 10.1038/s41598-023-44961-3
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379. doi: 10.1371/journal.pone.0019379
- Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., and Prasanna, B. M. (2022). Global maize production, consumption and trade: trends and R&D implications. *Food Secur.* 14, 1295–1319. doi: 10.1007/s12571-022-01288-7
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.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
- Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17 (1), 27–32. doi: 10.1111/1755-0998.12509
- Frantz, A. C., Cellina, S., Krier, A., Schley, L., and Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J. Appl. Ecol.* 46, 493–505. doi: 10.1111/j.1365-2664.2008.01606.x
- Fuentes-Cardenas, I. S., Cuba-Puma, R., Marcilla-Truyenque, S., Begazo-Gutiérrez, H., Zolla, G., Fuentealba, C., et al. (2022). Diversity of the Peruvian Andean maize (*Zea mays* L.) race Cabanita: Polyphenols, carotenoids, *in vitro* antioxidant capacity, and physical characteristics. *Front. Nutr.* 9. doi: 10.3389/fnut.2022.983208
- Gamarra Sánchez, G., Munive Cerrón, R., Munive Yachachi, Y., Azabache Leytón, A., Sevilla Panizo, R., and Lapa Chanca, A. (2020). Behavior of populations of floury maize of the Blanco del Cusco variety in the Mantaro valley, Peru. *Agroind. Sci.* 10, 279–286. doi: 10.17268/agroind.sci.2020.03.09
- García, C., Guichoux, E., and Hampe, A. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. *turbinata*). *Tree Genet. Genomes* 14, 87. doi: 10.1007/s11295-018-1301-x
- García Mendoza, P. J., Medina Castro, D. E., Prieto Rosales, G. P., Manayay Sánchez, D., and Ortecho Llanos, R. (2021). Comportamiento agronómico de variedades de maíz amiláceo tradicionales y mejoradas evaluadas en diferentes ambientes de Tayacaja. *Llamkasun* 2, 121–143. doi: 10.47797/llamkasun.v2i1.36
- García-Mendoza, P. J., Pérez-Almeida, I. B., Prieto-Rosales, G. P., Medina-Castro, D. E., Manayay-Sánchez, D., Marin-Rodríguez, C. A., et al. (2023). Genetic diversity and productive potential of starchy corn varieties evaluated in Peruvian highland environments. *J. Saudi Soc. Agric. Sci.* 23, 168–176. doi: 10.1016/j.jssas.2023.10.007
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9, e90346. doi: 10.1371/journal.pone.0090346
- Goodman, M. M., and Brown, W. L. (1988). "Races of Corn," in *Corn and corn improvement*. Eds. G. F. Sprague and J. W. Dudley (American Society of Agronomy, Wisconsin), 33–79. doi: 10.2134/agronmonogr18.3ed.c2
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8278
- Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015, 1–14. doi: 10.1155/2015/431487
- Grant, U. J., Hatheway, W. H., Timothy, D. H., Cassalet, C., and Roberts, L. M. (1963). *Races of maize in Venezuela* (Washington D.C: National Academy of Sciences, National Research Council).
- Grobman, A. (1982). "Maíz (*Zea mays*)," in *Los Gavilanes: mar, desierto y oásis en la historia del hombre: precerámico Peruano*. Ed. D. Bonavia (Corporación Financiera de Desarrollo, Oficina de Asuntos Culturales, Lima), 157–180.
- Grobman, A., Bonavia, D., Dillehay, T. D., Piperno, D. R., Iriarte, J., and Holst, I. (2012). Preceramic maize from paredones and Huaca Prieta, Peru. *Proc. Natl. Acad. Sci.* 109, 1755–1759. doi: 10.1073/pnas.1120270109
- Grobman, A., Salhuana, W., Sevilla, R., and Mangelsdorf, P. C. (1961). *Races of maize in Peru* (Washington D.C: National Academies of Sciences, NRC).
- He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H. K., and Li, Z. (2014). Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00484
- Herrera-Saucedo, V., Santacruz-Varela, A., Rocandio-Rodríguez, M., Córdova-Téllez, L., Moreno-Ramírez, Y. R., and Hernández-Galeno, C. A. (2019). Genetic diversity of maize landraces of northern Mexico analyzed through microsatellites. *Agrociencia* 53, 535–548.
- Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J.-M., Skovmand, B., Taba, S., et al. (1999). Plant genetic resources: What can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci.* 96, 5937–5943. doi: 10.1073/pnas.96.11.5937
- Huaringa-Joaquin, A., Saldaña, C. L., Saravia, D., García-Bendez, S., Rodríguez-Grados, P., Salazar, W., et al. (2023). Assessment of the genetic diversity and population structure of the Peruvian Andean legume, Tarwi (*Lupinus mutabilis*), with high quality SNPs. *Diversity (Basel)* 15, 15. doi: 10.3390/d15030437
- Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., et al. (2021). *De novo* assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Sci.* (1979) 373, 655–662. doi: 10.1126/science.abg5289
- Iorizzo, M., Senalik, D. A., Ellison, S. L., Grzebelus, D., Cavagnaro, P. F., Allender, C., et al. (2013). Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *Am. J. Bot.* 100, 930–938. doi: 10.3732/ajb.1300055
- Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., et al. (2017). The $K = 2$ conundrum. *Mol. Ecol.* 26, 3594–3602. doi: 10.1111/mec.14187
- Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart, T., and Ahmed, I. (2011). *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6. doi: 10.3389/fgene.2015.00208
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2014, 1–14. doi: 10.7717/peerj.281
- Kistler, L., Maezumi, S. Y., Gregorio de Souza, J., Przelomska, N. A. S., Malaquias Costa, F., Smith, O., et al. (2018). Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Sci.* (1979) 362, 1309–1313. doi: 10.1126/science.aav0207
- Korecký, J., Čepel, J., Stejskal, J., Faltinová, Z., Dvořák, J., Lstibůrek, M., et al. (2021). Genetic diversity of Norway spruce ecotypes assessed by GBS-derived SNPs. *Sci. Rep.* 11, 23119. doi: 10.1038/s41598-021-02545-z
- Lee, K. J., Lee, J.-R., Sebastin, R., Shin, M.-J., Kim, S.-H., Cho, G.-T., et al. (2019). Genetic diversity assessed by genotyping by sequencing (GBS) in watermelon germplasm. *Genes (Basel)* 10, 822. doi: 10.3390/genes10100822
- Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925. doi: 10.1080/106351501753462876
- Lischer, H. E. L., and Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28, 298–299. doi: 10.1093/bioinformatics/btr642
- López Alejandria, J. (2011). *Evaluación de siete variedades de maíz choclero (*Zea mays* L. ssp. *amilacea*), bajo condiciones de Cutervo* (Lambayeque: Universidad Nacional Pedro Ruiz Gallo).

- Lorello, I. M., García Lampasona, S. C., and Peralta, I. E. (2020). Genetic diversity of squash landraces (*Cucurbita maxima*) collected in Andean Valleys of Argentina. *Rev. Fac. Cienc. Agrar.* 52, 293–313.
- Macuri Núñez, E. R. (2016). *Estudio de la diversidad fenotípica del maíz (Zea mays L.) en la sierra baja y media del Perú* (Lima: Universidad Nacional Agraria La Molina). Available online at: <https://repositorio.lamolina.edu.pe/bitstream/handle/20.500.12996/1981/F30-M32-T.pdf?sequence=1> (Accessed April 22, 2024).
- Mamani-Huarcaya, B. M., González-Fontes, A., Navarro-Gochicoa, M. T., Camacho-Cristóbal, J. J., Ceacero, C. J., Herrera-Rodríguez, M. B., et al. (2022). Characterization of two Peruvian maize landraces differing in boron toxicity tolerance. *Plant Physiol. Biochem.* 185, 167–177. doi: 10.1016/j.plaphy.2022.06.003
- Mamani-Huarcaya, B. M., Luque, B., Ceacero, C. J., Fernández Cutire, O., and Rexach, J. (2020). Caracterización fisiológica de tres razas de maíz Peruano cultivadas con altos contenidos de boro. *Ciencia Desarrollo* 32–40, 167–177. doi: 10.33326/26176033.2020.26.930
- Martínez-Flores, F., Crespo, M. B., Simon, P. W., Ruess, H., Reitsma, K., Geoffriau, E., et al. (2020). Subspecies variation of *daucus carota* coastal (“Gummifer”) morphotypes (Apiaceae) using genotyping-by-sequencing. *Syst. Bot.* 45, 688–702. doi: 10.1600/036364420X15935294613527
- Martin-Sanz, A., Caminero, C., Jing, R., Flavell, A. J., and Perez de la Vega, M. (2011). Genetic diversity among Spanish pea (*Pisum sativum* L.) landraces, pea cultivars and the World Pisum sp. core collection assessed by retrotransposon-based insertion polymorphisms (RBIPs). *Spanish J. Agric. Res.* 9, 166. doi: 10.5424/sjar/20110901-214-10
- Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081. doi: 10.1038/s41588-019-0443-6
- Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez G., J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci.* 99, 6080–6084. doi: 10.1073/pnas.052125199
- McAlvay, A. C., Ragsdale, A. P., Mabry, M. E., Qi, X., Bird, K. A., Velasco, P., et al. (2021). *Brassica rapa* domestication: untangling wild and feral forms and convergence of crop morphotypes. *Mol. Biol. Evol.* 38, 3358–3372. doi: 10.1093/molbev/msab108
- McCouch, S., Navabi, Z. K., Abberton, M., Anglin, N. L., Barbieri, R. L., Baum, M., et al. (2020). Mobilizing crop biodiversity. *Mol. Plant* 13, 1341–1344. doi: 10.1016/j.molp.2020.08.011
- Mezghani, N., Ruess, H., Tarchoun, N., Ben Amor, J., Simon, P. W., and Spooner, D. M. (2018). Genotyping-by-sequencing reveals the origin of the Tunisian relatives of cultivated carrot (*Daucus carota*). *Genet. Resour. Crop Evol.* 65, 1359–1368. doi: 10.1007/s10722-018-0619-4
- Ministerio del Ambiente (2018). *Línea de base de la diversidad genética del maíz Peruano con fines de bioseguridad*. Available online at: <https://bioseguridad.minam.gob.pe/wp-content/uploads/2019/01/Línea-de-base-ma%C3%ADz-LowRes.pdf> (Accessed April 21, 2024).
- Ortiz, R., Crossa, J., Franco, J., Sevilla, R., and Burgueño, J. (2008a). Classification of Peruvian highland maize races using plant traits. *Genet. Resour. Crop Evol.* 55, 151–162. doi: 10.1007/s10722-007-9224-7
- Ortiz, R., Crossa, J., and Sevilla, R. (2008b). Minimum resources for phenotyping morphological traits of maize (*Zea mays* L.) genetic resources. *Plant Genet. Res.: Character. Util.* 6, 195–200. doi: 10.1017/S1479262108994168
- Ortiz, R., and Engels, J. (2018). “Genebank management and the potential role of molecular genetics to improve the use of conserved genetic diversity,” in *The evolving role of genebanks in the fast-developing field of molecular genetics*. Ed. M. C. de Vicente (International Plant Genetic Resources Institute (IPGRI), Rome, Italy), 19–25.
- Ortiz, R., Sevilla, R., Alvarado, G., and Crossa, J. (2008c). Numerical classification of related Peruvian highland maize races using internal ear traits. *Genet. Resour. Crop Evol.* 55, 1055–1064. doi: 10.1007/s10722-008-9312-3
- Özkan, G., Haliloğlu, K., Türkoğlu, A., Öztürk, H. I., Elkoca, E., and Pocza, P. (2022). Determining genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from Türkiye using SSR markers. *Genes (Basel)* 13, 1410. doi: 10.3390/genes13081410
- Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., and Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci.* 106, 5019–5024. doi: 10.1073/pnas.0812525106
- Prevosti, A., Ocaña, J., and Alonso, G. (1975). Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theor. Appl. Genet.* 45, 231–241. doi: 10.1007/BF00831894
- Prieto Rosales, G. P., and Manayay Sánchez, D. (2022). Caracterización morfológica de 25 variedades de maíz amiláceo evaluadas en dos localidades de la provincia de Tayacaja, Huancavelica. *Llamkasun* 3, 15–29. doi: 10.47797/llamkasun.v3i2.103
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- Ramírez, R. E., Timothy, D. H., Díaz, E., and Grant, U. J. (1960). *Races of maize in Bolivia* (Washington D.C: National Academy of Sciences, National Research Council).
- R Core Team (2024). *R: A language and environment for statistical computing*. Available online at: <https://www.R-project.org/> (Accessed January 8, 2024).
- Rivas, J. G., Gutierrez, A. V., Defacio, R. A., Schimpf, J., Vicario, A. L., Hopp, H. E., et al. (2022). Morphological and genetic diversity of maize landraces along an altitudinal gradient in the Southern Andes. *PLoS One* 17, e0271424. doi: 10.1371/journal.pone.0271424
- Roberts, L. M., Grant, U. J., Ramirez, R., Hatheway, W. H., Smith, D. L., and Mangelsdorf, P. C. (1957). *Races of maize in Colombia* (Washington D.C: National Academy of Sciences, National Research Council).
- Salazar, E., González, M., Araya, C., Mejía, N., and Carrasco, B. (2017). Genetic diversity and intra-racial structure of Chilean Choclero corn (*Zea mays* L.) germplasm revealed by simple sequence repeat markers (SSRs). *Sci. Hortic.* 225, 620–629. doi: 10.1016/j.scienta.2017.08.006
- Saldaña, C. L., Cancan, J. D., Cruz, W., Correa, M. Y., Ramos, M., Cuellar, E., et al. (2021). Genetic diversity and population structure of capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon revealed by RAPD markers. *Forests* 12, 1125. doi: 10.3390/f12081125
- Salhuana, W. S. (2004). Diversidad descripción de las razas de maíz en el Perú. In: W. En, W. Eds, A. Salhuana, F. Valdez and J. Scheuch Davelouis *Programa Cooperativo de Investigaciones en Maíz (PCIM), Logros y Perspectivas, 50º Aniversario*. (Universidad Nacional Agraria La Molina), pp. 204–251.
- Sevilla, R. (1994). “Variation in Modern Andean Maize and Its Implications for Prehistoric Patterns,” in *Corn and culture in the prehistoric New World*. Eds. S. Johannessen and C. A. Hastorf (Westview Press, Colorado), 219–244.
- Sevilla Panizo, R., and Holle Ostendorf, M. (2004). *Recursos genéticos vegetales* (Boulder, Colorado, USA: Luis León Asociados S.R.L.).
- Shu, G., Cao, G., Li, N., Wang, A., Wei, F., Li, T., et al. (2021). Genetic variation and population structure in China summer maize germplasm. *Sci. Rep.* 11, 8012. doi: 10.1038/s41598-021-84732-6
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., et al. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* 357, 512–515. doi: 10.1126/science.aam9425
- Tehseen, M. M., Tonk, F. A., Tosun, M., Istiqliler, D., Amri, A., Sansaloni, C. P., et al. (2022). Exploring the genetic diversity and population structure of wheat landrace population conserved at ICARDA genebank. *Front. Genet.* 13. doi: 10.3389/fgene.2022.900572
- Timothy, D. H., Hatheway, W. H., Grant, U. J., Torregroza, M., Sarria, D., and Varela, D. (1963). *Races of maize in Ecuador* (Washington D.C: National Academy of Sciences, National Research Council).
- Timothy, D. H., Peña, B., Ramirez, R., Brown, W. L., and Anderson, E. (1961). *Races of maize in Chile* (Washington D.C: National Academy of Sciences, National Research Council).
- Vigouroux, Y., Glaubitz, J. C., Matsuoka, Y., Goodman, M. M., Sánchez G., J., and Doebley, J. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* 95, 1240–1253. doi: 10.3732/ajb.0800097
- Wang, N., Yuan, Y., Wang, H., Yu, D., Liu, Y., Zhang, A., et al. (2020). Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Sci. Rep.* 10, 16308. doi: 10.1038/s41598-020-73321-8
- Waples, R. S., and Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15, 1419–1439. doi: 10.1111/j.1365-294X.2006.02890.x
- Warburton, M. L., Reif, J. C., Frisch, M., Bohn, M., Bedoya, C., Xia, X. C., et al. (2008). Genetic diversity in CIMMYT nontemperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. *Crop Sci.* 48, 617–624. doi: 10.2135/cropsci2007.02.0103
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evol. (N Y)* 38, 1358. doi: 10.2307/2408641
- Wellhausen, E. J., Roberts, L. M., Hernández Xolocotzi, E., and Mangelsdorf, P. C. (1952). *Races of maize in Mexico: their origin, characteristics and distribution* (Massachusetts: The Bussey Institution of Harvard University).
- Yang, N., Wang, Y., Liu, X., Jin, M., Vallebuena-Estrada, M., Calfee, E., et al. (2023). Two teosintes made modern maize. *Sci. (1979)* 382, eadg8940. doi: 10.1126/science.adg8940
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. (2017). ggtree : an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolkit for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606