



OPEN ACCESS

EDITED BY

Sukhjiwan Kaur,
Department of Environment and Primary
Industries (DEPI), Australia

REVIEWED BY

Ajoy Kumar Roy,
Indian Council of Agricultural Research
(ICAR), India
Aline da Costa Lima Moraes,
State University of Campinas, Brazil

*CORRESPONDENCE

Leidy G. Bobadilla
✉ leidy.bobadilla@untrm.edu.pe
Héctor V. Vásquez
✉ hvasquez@untrm.edu.pe

RECEIVED 27 February 2026

REVISED 26 May 2026

ACCEPTED 26 May 2026

PUBLISHED 10 June 2026

CITATION

Bobadilla LG, Ortiz R, Castillo G,
Matias FI, Gutiérrez L, Valqui L,
Carrasco-Chilón W, Díaz-Valderrama JR,
Venancio TM, Alava EI, Castillo MS and
Vásquez HV (2026) Genomic insights
into *Lolium multiflorum* diversity for
forage breeding in Andean livestock
systems.
Front. Plant Sci. 17:1818688.
doi: 10.3389/fpls.2026.1818688

COPYRIGHT

© 2026 Bobadilla, Ortiz, Castillo, Matias,
Gutiérrez, Valqui, Carrasco-Chilón, Díaz-
Valderrama, Venancio, Alava, Castillo and
Vásquez. 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.

Genomic insights into *Lolium multiflorum* diversity for forage breeding in Andean livestock systems

Leidy G. Bobadilla^{1,2*}, Rodomiro Ortiz^{1,3}, Gianmarco Castillo⁴,
Filipe I. Matias¹, Lucía Gutiérrez³, Leandro Valqui²,
William Carrasco-Chilón⁵, Jorge R. Díaz-Valderrama⁶,
Thiago M. Venancio⁴, E. I. Alava^{1,7}, Miguel S. Castillo^{1,8}
and Héctor V. Vásquez^{1,2*}

¹Escuela de Posgrado, Programa Doctoral en Ciencia para el Desarrollo Sustentable, Facultad de Ingeniería Zootecnista, Biotecnología, Agronegocios y Ciencia de Datos, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Chachapoyas, Peru, ²Laboratorio de Agrostología, Instituto de Investigación en Ganadería y Biotecnología, Facultad de Ingeniería Zootecnista, Biotecnología, Agronegocios y Ciencia de Datos, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Chachapoyas, Peru, ³Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Lomma, Sweden, ⁴Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biotecnologia e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro (LQFPP/UENF), Estado do Rio de Janeiro, Brazil, ⁵Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Estación Experimental de Baños del Inca, Cajamarca, Peru, ⁶Facultad de Ingeniería y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru, ⁷Facultad de Ciencias de la Vida, Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Guayas, Ecuador, ⁸Crop and Soil Sciences Department, North Carolina State University, Raleigh, NC, United States

Understanding the genetic diversity of *Lolium multiflorum* is essential for developing targeted breeding programs that can effectively support pasture-based livestock systems in the Peruvian Andean. This study assessed genomic variability and population structure in 27 *L. multiflorum* accessions from the Cajamarca region and the INIA Amazonas germplasm bank (Peru), using the genotyping-by-sequencing (GBS) technique. DNA extracted from young leaves was sequenced on an Illumina NovaSeq 6000 platform. After bioinformatic processing, 2,070 single nucleotide polymorphisms (SNPs) with heterogeneous distribution were obtained across seven chromosomes. A Principal Coordinate and an Unweighted Pair Group Method with Arithmetic Mean analyses revealed two distinct genetic groups were identified, reflecting a complex structure shaped by gene flow and local selection. The analysis of molecular variance showed that 90% of the genetic variation occurs within populations, whereas the remaining 10% corresponds to interregional differences ($\Phi_{IPT} = 0.099$, $p < 0.006$). The negative Inbreeding Coefficient (F_{IS}) values (Cajamarca = -0.2312; Amazonas = -0.5489) indicate an excess of heterozygotes, a pattern typically associated with predominantly outcrossing species. Additionally, high observed heterozygosity ($H_o > 0.57$) points to potential hybrid vigor and indicates that these populations may maintain stable genetic equilibrium. Collectively, these findings demonstrate that Peruvian

L. multiflorum harbors a broad genetic base, shaped by historical germplasm exchange and local environmental adaptation. This diversity provides critical insights for conservation strategies and can support breeding programs aimed at enhancing forage resilience and productivity in high-Andean ecosystems.

KEYWORDS

Andean grasslands, genotyping-by-sequencing, highland forage germplasm, *Lolium multiflorum*, population structure

1 Introduction

Grasslands cover more than 40% of the Earth's ice-free land surface and provide essential ecosystem services, including livestock products, carbon storage, and nutrient cycling (Millennium Ecosystem Assessment, 2005; Hewins et al., 2018; Erb et al., 2016). Land-use intensification and inadequate management practices negatively affect grassland ecosystems, resulting in biomass loss, biodiversity decline, and reduced ecosystem functioning (Chang et al., 2020; Jiao et al., 2024; Cao et al., 2024). Mountain grasslands are particularly vulnerable to these pressures, as pronounced environmental gradients and climatic variability can accelerate degradation processes and constrain the resilience of pasture-based livestock systems (Rui et al., 2025). In this context, conserving and characterizing genetic diversity of forage species is critical for sustaining productive and resilient livestock systems.

Lolium multiflorum Lam. is a widely cultivated forage species in temperate and high-altitude regions due to its high herbage mass production, nutritional value, and capacity to perform under diverse environmental conditions (Carrasco-Chilón et al., 2024). As a self-incompatible, obligate outcrossing species, *L. multiflorum* typically exhibits high levels of genetic variation within populations, providing a solid foundation for selection and genetic improvement (Fernández-Otero et al., 2024). Studies across the genus *Lolium* have consistently reported that most genetic variation occurs within populations rather than among them, emphasizing the importance of local germplasm for breeding and conservation programs (Nam et al., 2025).

Despite its agronomic relevance in Andean livestock systems, the genomic diversity and population structure of *L. multiflorum* in the Peruvian Andes remain poorly characterized (Vallejos-Cacho et al., 2024). Most genomic research on ryegrass has focused on European and Asian germplasm (Johansen et al., 2025), whereas genomic information for Andean populations remains limited. In the Peruvian Andes, *L. multiflorum* has been cultivated under heterogeneous environmental conditions and diverse management practices, which may contribute to genetic differentiation and local adaptation. In the Peruvian Andes, *L. multiflorum* is cultivated between 2,300 and 3,800 m.a.s.l. under highly heterogeneous environmental conditions associated with altitudinal gradients (Vallejos-Fernández et al., 2020). Traditional management practices, including informal seed exchange, intensive sowing, and recurrent reseeding, combined with an allogamous mating system, are expected to influence patterns of genetic diversity and population structure (Vieira et al., 2004). These conditions make

Andean germplasm particularly well suited for genome-wide diversity analyses.

Single-nucleotide polymorphism (SNP) markers generated through genotyping-by-sequencing (GBS) provide an efficient and reproducible approach for assessing genomic diversity in forage species (Elshire et al., 2011; Jaškunė et al., 2020; Altaf et al., 2025). SNP datasets derived from GBS have been widely applied to infer genetic diversity, population structure, and genetic relationships in *Lolium* species and other forage crops (Blackmore et al., 2016; Tamura et al., 2022; Altaf et al., 2025). The application of SNP- and GBS-based approaches to Andean ryegrass germplasm addresses an important knowledge gap regarding the genomic organization of Peruvian populations and contributes to a more complete understanding of forage biodiversity in high-altitude environments (Kumar et al., 2024; Bunjkar et al., 2024). Hence, the objective of this study was to characterize, at the genomic scale, the genetic diversity and population structure of *L. multiflorum* accessions from the northern Peruvian Andes using SNP-based GBS with the goals to identify a genomic baseline for Peruvian ryegrass germplasm and to offer a scientific foundation to guide forage breeding programs, genetic conservation strategies, and the adaptation of Andean livestock systems to high-altitude environmental variability.

2 Materials and methods

2.1 Plant material

A total of 50 ryegrass accession samples were collected in 2025 from selected accessions maintained in the germplasm bank of the National Institute for Agricultural Innovation (INIA) in Peru, at geographic coordinates 78°27'07" W longitude and 07°09'56" S latitude, at an elevation of 2667 m a.s.l. These samples represented the complete ryegrass germplasm collection conserved by INIA, maintained *in situ* and properly managed to prevent crossing among accessions. Sampling was conducted randomly within each plot; plant material was obtained from multiple non-adjacent individuals, maintaining a minimum distance of 5 m between plants. The collected accessions originated from the Cajamarca Region at the following sites: Cutervo (5), Tacabamba (7), Sendamal (5), Paccha (4), El Agrario (4), Calquis (4), Bambamarca (4), Campiña (3), Santa Cruz (2), San Pablo (1), Cochan (3), Celendín (4), and Baños del Inca (4).

In addition, 12 commercial ryegrass accessions were collected from materials available at the Chachapoyas Agricultural Experiment Station, within the Agrostological Garden (institutional forage collection) of the Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Amazonas Region. These included the following cultivars: Wanca Grass (3), Bison II (3), Inglés (2), Max (2), and the AGP ecotype (2) (Figure 1; Table 1).

2.2 DNA extraction

Ten young leaves were collected from each ryegrass accession and stored in Ziploc bags with silica gel to preserve the plant material prior to DNA extraction (Wilkie et al., 2013). Each ryegrass accession was then cut until a total weight of 650 mg per accession was obtained. The samples were transferred to 1.5-mL microcentrifuge tubes and mechanically disrupted using a tissue homogenizer. Grinding was performed using liquid nitrogen to ensure efficient homogenization. DNA extraction was carried out using the NucleoSpin® Plant II kit (MACHEREY-NAGEL, Dueren, Germany) following the manufacturer's instructions for grasses. Upon completion of the extraction process, the quality of the extracted DNA was assessed using the Quantus Fluorometer (Promega, Madison, USA). DNA integrity was verified through 1% agarose gel electrophoresis using a 1X Tris-Acetate-EDTA

(TAE) buffer, and GelRed (Merck, Rahway, New Jersey, USA) for loading. The DNA samples were further stored at -20 °C for preservation. The material was later sent to Novogene (Sacramento, CA, United States) for high-throughput sequencing.

2.3 Sequencing and SNP genotyping analysis

Sequencing-based genotyping libraries were developed following the protocol of Elshire et al. (2011). Genomic DNA was digested with the ApeKI enzyme, and fragments were ligated to Illumina sequencing adapters and sequence barcodes unique to each sample. After multiplexing, this allowed for the recovery of sample identity for each sequenced DNA fragment. Pooled samples were sequenced on the Illumina NovaSeq 6000 platform, obtaining 100-bp paired-end reads. Raw data quality was examined using FastQC v0.11.726. The reads obtained from the 62 ryegrass samples were filtered into clean data using Fastp v0.22.0 (Chen et al., 2018). The BWA v0.7.18 software (Li and Durbin, 2009) was used to align the reads to the *L. multiflorum* genome (GCA_030979885.1) of the Rabiosa cultivar (Goettelmann et al., 2024). The alignment files were processed using SAMTOOLS v1.22.1 (Li et al., 2009), reorganized and corrected with samtools fixmate, and reweighted with PICARD v3.3.0 (<https://github.com/broadinstitute/picard>). SNPs within the 62

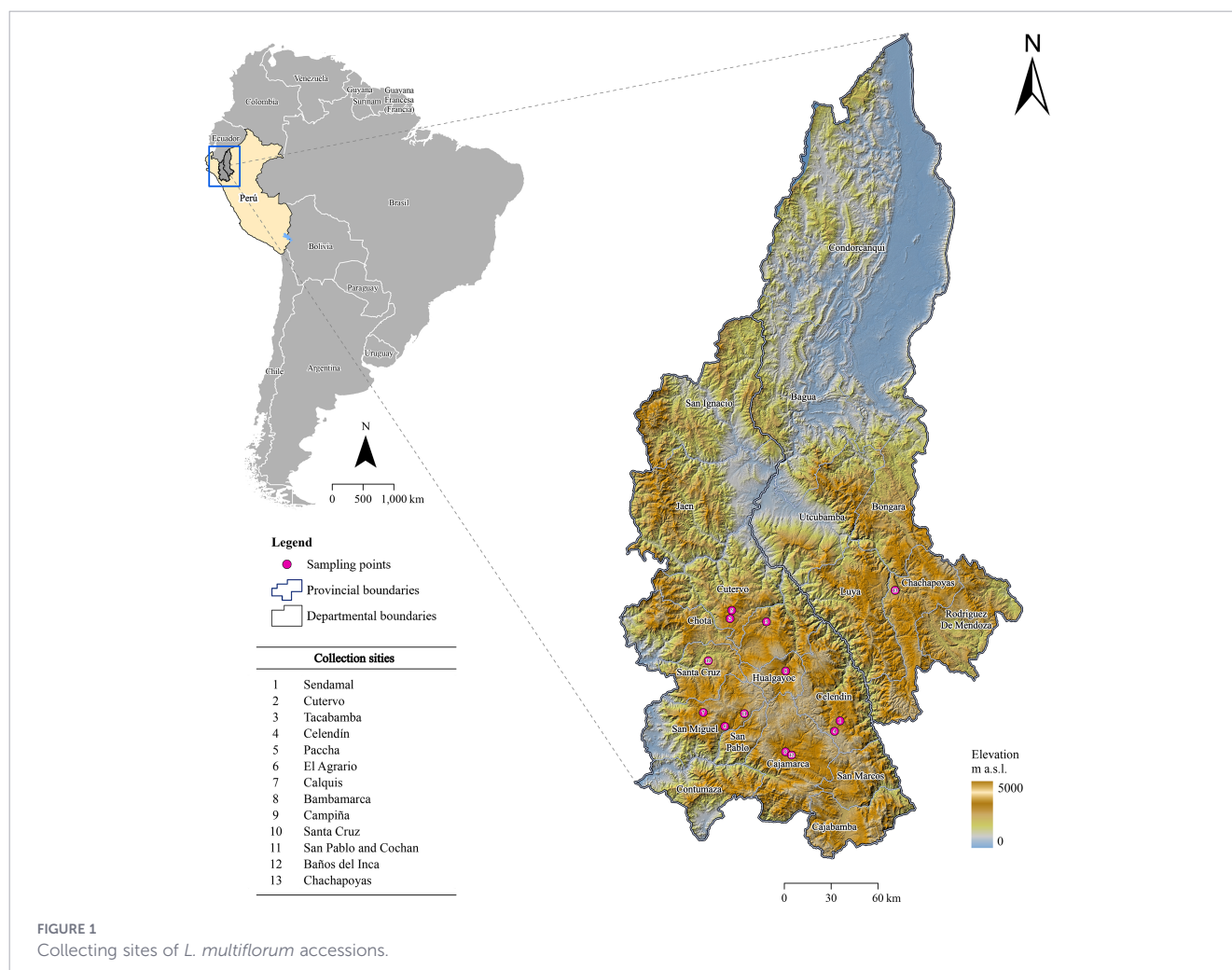


TABLE 1 Identification and origin of *L. multiflorum* accessions.

Sample ID	Accession name	Collection site	Province	Department	Coordinates	
					East	North
LM1 - P1	<i>Lolium multiflorum</i> Paccha - LM1	Paccha	Chota	Cajamarca	78°48'40.85" W	6°19'43.16" S
LM1 - P2	<i>Lolium multiflorum</i> Paccha - LM1	Paccha	Chota	Cajamarca	78°48'40.85" W	6°19'43.16" S
LM1 - P3	<i>Lolium multiflorum</i> Paccha - LM1	Paccha	Chota	Cajamarca	78°48'40.85" W	6°19'43.16" S
LM1 - P4	<i>Lolium multiflorum</i> Paccha - LM1	Paccha	Chota	Cajamarca	78°48'40.85" W	6°19'43.16" S
LM2 - P1	<i>Lolium multiflorum</i> Cutervo - LM2	Cutervo	Cutervo	Cajamarca	78°49'15.76" W	6°22'34.56" S
LM2 - P2	<i>Lolium multiflorum</i> Cutervo - LM2	Cutervo	Cutervo	Cajamarca	78°49'15.76" W	6°22'34.56" S
LM2 - P3	<i>Lolium multiflorum</i> Cutervo - LM2	Cutervo	Cutervo	Cajamarca	78°49'15.76" W	6°22'34.56" S
LM2 - P4	<i>Lolium multiflorum</i> Cutervo - LM2	Cutervo	Cutervo	Cajamarca	78°49'15.76" W	6°22'34.56" S
LM3 - P1	<i>Lolium multiflorum</i> Tacabamba - LM3	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM3 - P2	<i>Lolium multiflorum</i> Tacabamba - LM3	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM3 - P3	<i>Lolium multiflorum</i> Tacabamba - LM3	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM3 - P4	<i>Lolium multiflorum</i> Tacabamba - LM3	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM4 - P1	<i>Lolium multiflorum</i> Tacabamba- LM4	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM4 - P2	<i>Lolium multiflorum</i> Tacabamba- LM4	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM4 - P3	<i>Lolium multiflorum</i> Tacabamba- LM4	Tacabamba	Chota	Cajamarca	78°36'36.29" W	6°23'37.05" S
LM5 - P1	<i>Lolium multiflorum</i> Cutervo- LM5	Cutervo	Cutervo	Cajamarca	78°49'15.76" W	6°22'34.56" S
LM6 - P1	<i>Lolium multiflorum</i> Calquis - LM6	Calquis	San Miguel	Cajamarca	78°58'26.35" W	6°55'15.54" S
LM6 - P2	<i>Lolium multiflorum</i> Calquis - LM6	Calquis	San Miguel	Cajamarca	78°58'26.35" W	6°55'15.54" S
LM6 - P3	<i>Lolium multiflorum</i> Calquis - LM6	Calquis	San Miguel	Cajamarca	78°58'26.35" W	6°55'15.54" S
LM6 - P4	<i>Lolium multiflorum</i> Calquis - LM6	Calquis	San Miguel	Cajamarca	78°58'26.35" W	6°55'15.54" S
LM7 - P1	<i>Lolium multiflorum</i> El Agrario - LM7	El Agrario	San Miguel	Cajamarca	78°50'55.37" W	7°0'2.52" S
LM7 - P2	<i>Lolium multiflorum</i> El Agrario - LM7	El Agrario	San Miguel	Cajamarca	78°50'55.37" W	7°0'2.52" S
LM7 - P3	<i>Lolium multiflorum</i> El Agrario - LM7	El Agrario	San Miguel	Cajamarca	78°50'55.37" W	7°0'2.52" S
LM7 - P4	<i>Lolium multiflorum</i> El Agrario - LM7	El Agrario	San Miguel	Cajamarca	78°50'55.37" W	7°0'2.52" S
LM8 - P1	<i>Lolium multiflorum</i> Bambamarca - LM8	Bambamarca	Hualgayoc	Cajamarca	78°29'52.3" W	6°40'42.1" S
LM8 - P2	<i>Lolium multiflorum</i> Bambamarca - LM8	Bambamarca	Hualgayoc	Cajamarca	78°29'52.3" W	6°40'42.1" S
LM8 - P3	<i>Lolium multiflorum</i> Bambamarca - LM8	Bambamarca	Hualgayoc	Cajamarca	78°29'52.3" W	6°40'42.1" S
LM8 - P4	<i>Lolium multiflorum</i> Bambamarca - LM8	Bambamarca	Hualgayoc	Cajamarca	78°29'52.3" W	6°40'42.1" S
LM9 - P1	<i>Lolium multiflorum</i> Campiña- LM9	Campiña	Cajamarca	Cajamarca	78°29'46.61" W	7°8'46.53" S
LM10 - P1	<i>Lolium multiflorum</i> Campiña- LM10	Campiña	Cajamarca	Cajamarca	78°29'46.61" W	7°8'46.53" S
LM11 - P1	<i>Lolium multiflorum</i> Sendamal- LM11	Sendamal	Celendin	Cajamarca	78°10'52.49" W	6°57'54.96" S
LM11 - P2	<i>Lolium multiflorum</i> Sendamal- LM11	Sendamal	Celendin	Cajamarca	78°10'52.49" W	6°57'54.96" S
LM12 - P1	<i>Lolium multiflorum</i> Sendamal- LM12	Sendamal	Celendin	Cajamarca	78°10'52.49" W	6°57'54.96" S
LM12 - P2	<i>Lolium multiflorum</i> Sendamal- LM12	Sendamal	Celendin	Cajamarca	78°10'52.49" W	6°57'54.96" S
LM12 -P3	<i>Lolium multiflorum</i> Sendamal- LM12	Sendamal	Celendin	Cajamarca	78°10'52.49" W	6°57'54.96" S
LM13 - P1	<i>Lolium multiflorum</i> Celendín- LM13	Celendín	Cajamarca	Cajamarca	78°12'45.07" W	7°1'32.91" S
LM13 - P2	<i>Lolium multiflorum</i> Celendín- LM13	Celendín	Cajamarca	Cajamarca	78°12'45.07" W	7°1'32.91" S
LM13 - P3	<i>Lolium multiflorum</i> Celendín- LM13	Celendín	Cajamarca	Cajamarca	78°12'45.07" W	7°1'32.91" S
LM13 - P4	<i>Lolium multiflorum</i> Celendín- LM13	Celendín	Cajamarca	Cajamarca	78°12'45.07" W	7°1'32.91" S
LM14 - P1	<i>Lolium multiflorum</i> Campiña- LM14	Campiña	Cajamarca	Cajamarca	78°29'46.61" W	7°8'46.53" S
LM15 - P1	<i>Lolium multiflorum</i> Cochan- LM15	Cochan	San Pablo	Cajamarca	78°44'9.65" W	6°55'34.67" S
LM16 - P1	<i>Lolium multiflorum</i> Cochan- LM16	Cochan	San Pablo	Cajamarca	78°44'9.65" W	6°55'34.67" S
LM17 - P1	<i>Lolium multiflorum</i> Cochan- LM17	Cochan	San Pablo	Cajamarca	78°44'9.65" W	6°55'34.67" S

(Continued)

TABLE 1 Continued

Sample ID	Accession name	Collection site	Province	Department	Coordinates	
					East	North
LM18 - P1	<i>Lolium multiflorum</i> Santa Cruz- LM18	Santa Cruz	Santa Cruz	Cajamarca	78°56'45.22" W	6°37'15.11" S
LM19 - P1	<i>Lolium multiflorum</i> Santa Cruz- LM19	Santa Cruz	Santa Cruz	Cajamarca	78°56'45.22" W	6°37'15.11" S
LM20 - P1	<i>Lolium multiflorum</i> San Pablo - LM20	San Pablo	San Pablo	Cajamarca	78°44'9.65" W	6°55'34.67" S
LM21 - P1	INIA 910 – Kumymarca- LM21	Baños del Inca	Cajamarca	Cajamarca	78°27'39.55" W	7°9'53.78" S
LM21 - P2	INIA 910 – Kumymarca- LM21	Baños del Inca	Cajamarca	Cajamarca	78°27'39.55" W	7°9'53.78" S
LM21 - P3	INIA 910 – Kumymarca- LM21	Baños del Inca	Cajamarca	Cajamarca	78°27'39.55" W	7°9'53.78" S
WC - P1	Wanca grass-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
WC - P2	Wanca grass-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
WC- P3	Wanca grass-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
BISII - P1	Bison II-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
BISII - P2	Bison II-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
BISII - P3	Bison II-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
Ingl - P1	Ingles-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
Ingl - P2	Ingles-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
MAX - P1	Max-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
MAX - P3	Max-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
AGP - P1	Ecotipo AGP-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
AGP - P2	Ecotipo AGP-Comercial	Comercial	Chachapoyas	Amazonas	77°52'0.75" W	6°12'27.03" S
LM 22 - P1	Ecotipo Cajamarquino LM-22	Baños del Inca	Cajamarca	Cajamarca	78°27'39.55" W	7°9'53.78" S

A total of 62 samples were collected. Samples corresponding to repeated collections of the same accession were concatenated and analyzed as a single representative accession, resulting in a final dataset of 27 accessions used in the genetic analyses.

sequences of the ryegrass accessions were called using GATK v4.6.2 (Auweru et al., 2013). Filtering was performed using the following parameters: SNPs were selected while excluding indels and other variants; only biallelic SNPs were retained; $QD < 2.0$, $FS > 60.0$, $MQ < 40.0$, $MQRankSum < -12.5$, $ReadPosRankSum < -8.0$, $QUAL < 30.0$, and $10 \leq DP \leq 100$ (Song et al., 2023). Additionally, the VCF files generated with GATK v4.6.2 were concatenated using BCFtools v1.22 (Danecek and McCarthy, 2017).

The initial dataset consisted of 62 ryegrass plant samples. However, some of these samples were biological replicates of the same genetic accession. Therefore, the number of samples was not equivalent to the number of unique accessions. To resolve this, replicate samples were merged by retaining a single representative genotype per accession, thereby removing redundancy. This procedure generated a final non-redundant dataset of 27 unique accessions, which were used for downstream genomic analyses (Table 1). This approach was implemented to avoid redundancy and pseudo-replication, ensuring that downstream population genetic analyses reflected true genetic diversity rather than replicate structure. Subsequently, the data were curated using VCFtools v0.1.17 (Danecek et al., 2011) by applying filters for $MAF > 0.1$, missing data ≤ 0.1 , and $MAC \geq 3$ (Arbizu et al., 2025). Finally, SNPs in linkage disequilibrium (Ragsdale and Gravel, 2020) with ($r^2 = 0.85$) were removed using cyvcf2 v0.31.2 (Pedersen and Quinlan, 2017) and scikit-allel v1.3.13 (Miles et al., 2024) in Python v3.12.3. This computational approach enabled the retrieval of highly curated SNPs and a robust inference of the genetic relationships among the accessions.

2.4 Analysis of population structure and genetic diversity

Population genetic structure was analyzed using ADMIXTURE v1.3.0 (Alexander and Lange, 2011) with the maximum-likelihood algorithm. Values of K from 1 to 10 were evaluated, performing 10 independent replicates for each K, and cross-validation was used to estimate the CV error and determine the optimal number of clusters. For each value of K, the replicate with the highest likelihood reported by the software was selected for subsequent analyses. Population structure was visualized using the 'ggplot2' package v4.0.0 (Wickham, 2016) in R. Additionally, to determine the ancestry of the accessions within each ADMIXTURE cluster, the Weir and Cockerham (1984) F_{ST} was calculated. For this analysis, the packages hierfstat v0.5.11 (Goudet, 2005) and adegenet v2.1.11 (Jombart, 2008) were used. Likewise, an individual-by-individual genetic correlation matrix was generated by calculating the identity-by-state coefficient (PI_HAT). For this calculation, the filtered and robust SNPs obtained and PLINK v1.9.0 (Purcell et al., 2007) software were used to process the data and generate the similarity matrix. Visualization was performed using the pheatmap v1.0.12 (Kolde, 2019) package in R.

A molecular analysis of variance (AMOVA) was also performed using the R 'ape' package (Paradis and Schliep, 2019) to determine the sources of genetic variance among ryegrass lineages. To evaluate statistical significance, a randomization test with 999 permutations was conducted using the permutation test function of the ade4

package v1.7.23 (Dray and Dufour, 2007). Subsequently, genetic diversity parameters were estimated, including sample size (N), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F_{IS}), and the Shannon diversity index, using the Hierfstat v0.5.11 package (Goudet, 2005) in R. Levels of genetic diversity were classified following Kanaka et al. (2023) and Mohammadi and Prasanna (2003), considering expected heterozygosity (H_e) values close to one as indicative of a high number of alleles with similar frequencies, and values close to zero as indicative of no heterozygosity.

Principal Coordinate Analysis (PCoA) was performed in R. Genotypes were extracted from the previously filtered VCF file using the 'vcfR' package v.15.0 (Knaus and Grünwald, 2017). Euclidean distances among individuals were computed using 'ape' v5.8.1 (Paradis and Schliep, 2019), and principal coordinates were obtained through the PCoA function of the same package. Visualization of the results was carried out with 'ggplot2', and marginal distributions were added using ggExtra v0.11.0 (Attali and Baker, 2019) to represent both intra- and interregional genetic dispersion and variability.

Additionally, a principal component analysis (PCA) was performed using productive traits related to dry matter (DM) and fresh forage (FF) to explore phenotypic patterns among the accessions. This analysis was conducted on the final set of 27 accessions and is presented as supplementary material (Supplementary Figure 3).

2.5 Phylogenetic analysis

The phylogenetic tree was constructed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based on the total high-quality SNPs obtained, calculating binary genetic distances through Hamming distance in Python v3.12.3, using the NumPy (Harris et al., 2020), SciPy v1.15.3 (Virtanen et al., 2020), scikit-allel 1.3.13 (Miles et al., 2024), and Bio.Phylo v1.85 (Talevich et al., 2012) packages. Hamming distance was selected because it efficiently measures pairwise dissimilarity among binary SNP genotypes by counting allelic mismatches between individuals (Wang et al., 2015; Agazzi et al., 2023; Sanjeevan and König, 2025). This metric is commonly applied in SNP-based phylogenetic and population analyses, as it preserves marker-level differences and is computationally efficient, making it suitable even when large volumes of genomic data are not available (Hamming, 1950; Di Pasquale et al., 2021). Biallelic genotypes were generated with scikit-allel, and the robustness of the clustering was evaluated using 1,000

bootstrap replicates by resampling loci. The tree was exported in Newick format and visualized in iTOL v7 (Letunic and Bork, 2024). This procedure ensured a reproducible and consistent phylogenetic inference based on high-quality SNP data.

3 Results

3.1 Sequencing and distribution of SNPs

The total raw sequencing reads for the 27 *L. multiflorum* accessions were 320,554,538, with an average of 11.9 million reads per accession. Variant calling performed with GATK identified a total of 7,379,574 SNPs across the 27 accessions, which were then filtered to ensure the quality and reliability of the analyzed variants. After applying the GATK filtering criteria, 614,161 high-quality variants were retained. Subsequently, 22,689 SNPs were removed using VCFtools, and a final linkage disequilibrium (LD) filtering step implemented with the cyvcf2 and scikit-allel packages yielded a curated set of 2,070 high-quality SNPs distributed across the seven ryegrass chromosomes (Table 2). This rigorous filtering strategy ensured a robust and representative marker dataset suitable for downstream genomic analyses. The highest number of SNPs was physically mapped to chromosome four (19.28%, 399 SNPs), and the lowest to chromosome six (10.68%, 221 SNPs). Chromosomes 1 (1.18 Mb) and 7 (0.77 Mb) showed the highest and lowest SNP densities, respectively.

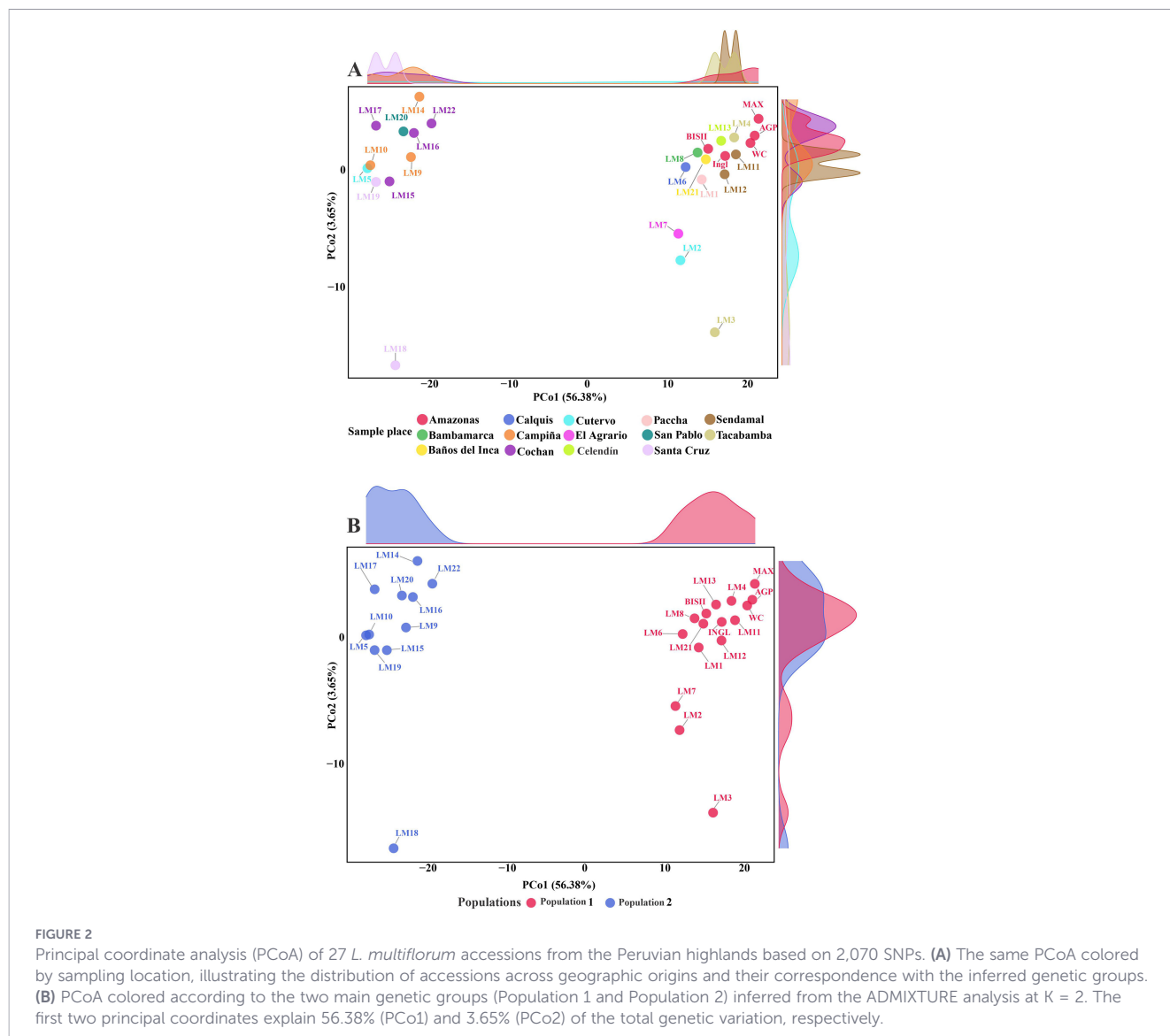
3.2 Population structure

Population structure among the 27 *L. multiflorum* accessions was evaluated using Principal Coordinate Analysis (PCoA), ADMIXTURE, hierarchical clustering, and genetic correlation analyses. The PCoA revealed consistent clustering patterns among accessions, with the first two principal coordinates explaining 56.38% and 3.65% of the total genetic variation, respectively (Figure 2A).

The ADMIXTURE analysis supported the presence of two genetically differentiated groups, with $K = 2$ identified as the most supported number of clusters based on the lowest cross-validation error (Figure 2B). The ancestry assignments showed a clear separation of the Cutervo accessions (LM5 and LM2), which were grouped into distinct clusters. In contrast, accessions from

TABLE 2 Genomic distribution of 2,070 single nucleotide polymorphisms (SNPs) across the seven ryegrass chromosomes.

Chromosome	SNP (#)	SNP (%)	Total length (Mb)	Average total (Pb)	Density (SNPs/Mb)
1	303	14.64	257.70	853314.15	1.18
2	350	16.91	341.23	977724.86	1.03
3	315	15.22	355.29	1131510.51	0.89
4	399	19.28	425.99	1070333.75	0.94
5	223	10.77	242.22	1091079.00	0.92
6	221	10.68	254.32	1156006.58	0.92
7	259	12.51	334.89	1298019.17	0.77



Cochán, Santa Cruz, Campiña, and San Pablo clustered within a single genetic group. Two accessions from Santa Cruz (LM18) and Tacabamba (LM13) displayed divergent positions relative to the main clusters. Accessions from Amazonas, Celendín, Sendamal, Baños del Inca, Paccha, Calquis, Bambamarca, and Tacabamba formed a consolidated genetic group. Additional divergence was observed for one accession from Tacabamba (LM3) and for accessions from El Agrario, which were positioned outside the main cluster.

The phylogenetic tree reconstructed using the UPGMA method, based on binary distances from 2,070 high-quality SNPs (Figure 3), revealed a bifurcated structure composed of two main phylogenetic groups. The first group consists of 11 accessions from Cochán, Santa Cruz, Campiña, Cutervo, and San Pablo. The second group comprises 16 accessions from Amazonas, Tacabamba, Cutervo, El Agrario, Paccha, Sendamal, Baños del Inca, Calquis, Bambamarca, and Celendín. This phylogenetic structure is consistent with the PCoA results. To facilitate the interpretation of the phylogenetic structure, detailed information on intragroup and intra-subgroup genetic distances together with accession

composition is provided in [Supplementary Table 1](#). Additional information describing group and subgroup composition, as well as intra- and inter-genetic distances derived from the UPGMA analysis, is presented in [Supplementary Table 2](#).

The genetic structure analysis performed with ADMIXTURE identified K = 2 as the optimal number of ancestral populations, determined by the lowest cross-validation error (Figure 4A). Under this model, the 27 accessions were clearly grouped into two genetically differentiated populations (Figure 4B). Population 1 (blue) included 16 accessions, while Population 2 (orange) comprised 11 accessions. Within Population 1, the accessions AGP, MAX, and WC formed a differentiated subgroup, whereas in Population 2, the accessions LM10, LM5, LM17, LM19, and LM15 constituted another clearly defined subgroup.

To evaluate the genetic differentiation between accessions with apparently pure ancestry from population 1 (AGP, MAX, and WC) and population 2 (LM10, LM5, LM17, LM15, and LM19) and those showing mixed ancestry within each population, Weir and Cockerham's F_{ST} estimator (Weir and Cockerham, 1984) was calculated. In population 2, the F_{ST} between pure and admixed

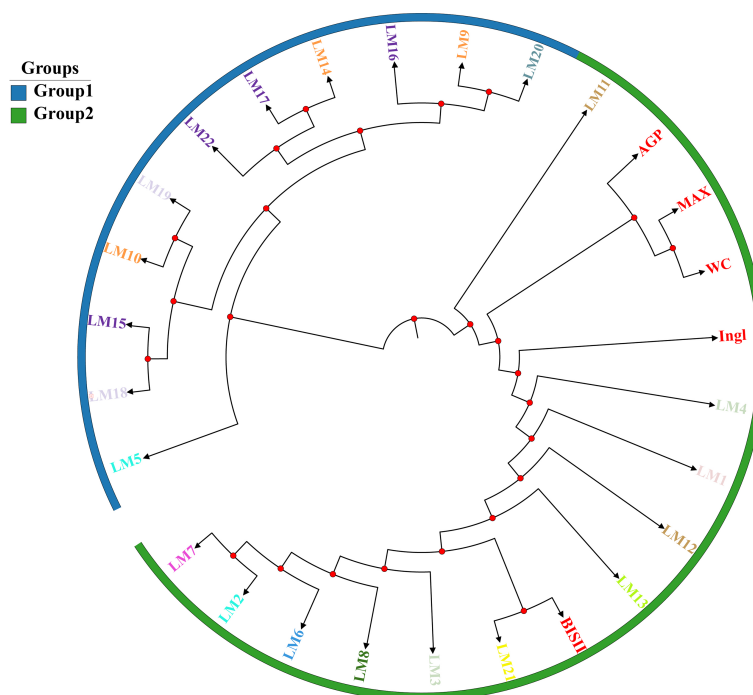


FIGURE 3

Phylogeny based on binary genetic distances and the unweighted pair group method with arithmetic mean (UPGMA) clustering method for the 27 ryegrass accessions using 2,070 single nucleotide polymorphism (SNP) markers. Detailed intragroup and intra-subgroup genetic distances are provided in [Supplementary Table 1](#), whereas group and subgroup composition together with intra- and inter-genetic distances are summarized in [Supplementary Table 2](#).

individuals was 0.0082, whereas in population 1 the value was 0.0216 ([Table 3](#)). Both values indicated low genetic differentiation within populations. In contrast, higher differentiation was observed between populations, with an overall F_{ST} value of 0.057.

The individual-by-individual genetic correlation matrix based on SNPs, together with hierarchical clustering, showed a clear segregation of individuals into two main groups ([Figure 5](#)). The first group was composed of individuals LM18, LM15, LM19, LM10, LM5, LM17, LM14, LM9, LM20, LM16, and LM22, which exhibited high genetic correlation values among themselves, as evidenced by warm color tones in the matrix. The second group included individuals LM2, LM7, LM6, LM8, LM13, LM4, LM11, LM12, AGP, MAX, WC, LM3, BISII, LM1, and LM21, likewise showing high levels of intragroup correlation. Correlations between

individuals belonging to opposite groups were markedly lower, reflected by cool color tones. The main diagonal displayed maximum correlation values, corresponding to the genetic identity of each individual with itself ([Figure 5](#)).

3.3 Population structure and genetic differentiation

The AMOVA revealed significant genetic structure among the populations. A total of 10% of the genetic variation was attributed to differences among populations, whereas most of the diversity (90.01%) was found within populations. This population differentiation was statistically significant ($\Phi_{PT} = 0.0999$, $p = 0.006$), thus indicating that the groups analyzed are not genetically

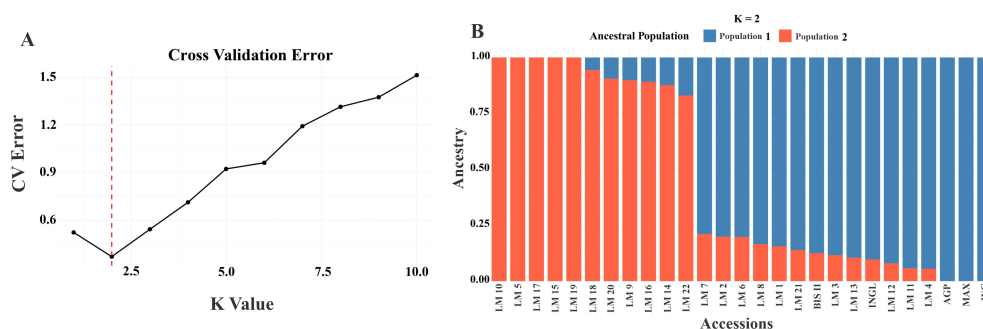


FIGURE 4

(A) Cross-validation error plot identifying $K = 2$ as the optimal number of ancestral populations. (B) Ancestry bar plot at $K = 2$ showing two genetically differentiated groups, with several individuals exhibiting mixed ancestry.

TABLE 3 Weir and Cockerham FST matrix among the groups defined based on the ADMIXTURE results.

	Population 2 – pure	Population 2 – admixed	Population 1 – pure	Population 1 – admixed
Population 2 – Pure	0.0000	0.0082	0.4610	0.3377
Population 2 – Admixed	0.0082	0.0000	0.3850	0.2830
Population 1 – Pure	0.4610	0.3850	0.0000	0.0216
Population 1 – Admixed	0.3377	0.2830	0.0216	0.0000

homogeneous and that processes such as drift, selection, or barriers to gene flow have effectively generated a moderate but meaningful genetic structure (Table 4). Additionally, statistical significance was evaluated using a randomization test with 999 permutations, which indicated that approximately 10% of the genetic variation can be explained by population structure ($F = 2.776$, $p < 0.015$) (Supplementary Figure 2).

3.4 Genetic diversity patterns

The genetic diversity analysis based on 2,070 SNPs revealed moderate to high levels of genetic diversity in both populations. The F_{IS} values were negative, with -0.231 in Cajamarca and -0.549 in

Amazonas. The observed heterozygosity (H_o) showed high values in Cajamarca (0.569) and Amazonas (0.664), while the expected heterozygosity (H_e) presented moderate levels in Cajamarca (0.463) and Amazonas (0.429).

Genetic diversity, assessed using the Shannon index, was higher in Cajamarca (0.857) compared to Amazonas (0.572) (Table 5).

The principal component analysis (PCA) based on productive traits (DM and FF) revealed a clear phenotypic differentiation among the accessions (Supplementary Figure 3). Commercial cultivars such as BISII, MAX, AGP, and INGL clustered in a region characterized by intermediate dry matter levels and low fresh forage production, whereas the remaining accessions were distributed across a broader range of phenotypic combinations.

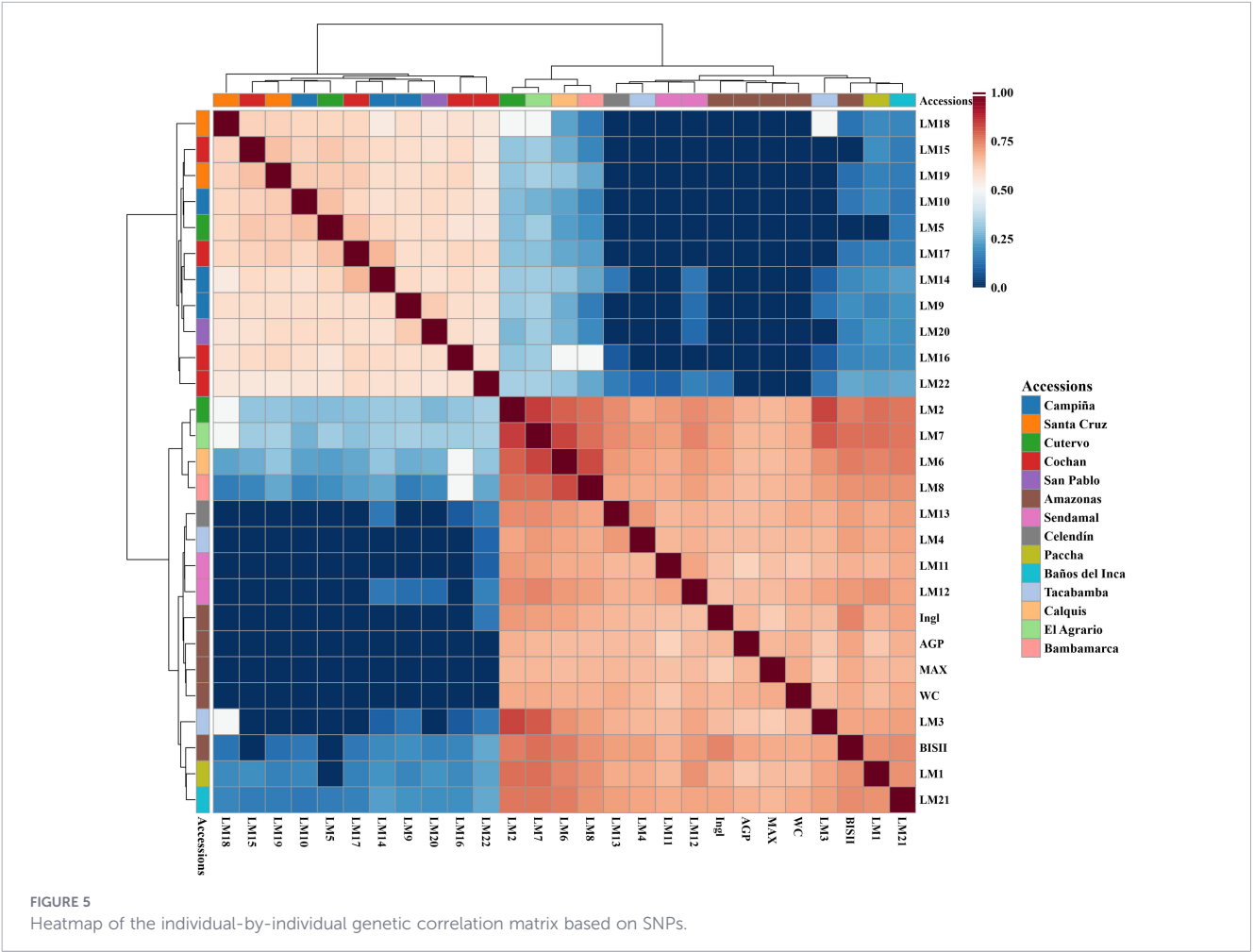


TABLE 4 Analysis of molecular variance (AMOVA) of 27 *L. multiflorum* accessions based on single nucleotide polymorphism (SNP) markers.

Source of variation	DF	SSD	MSD	Variance	%	PhiPT	P-value
Among populations*	1	2,121.1	2,121.10	81.55	9.99	0.0999	0.006
Within populations	25	19,102.8	764.11	734.75	90.01		
Total	26	21,223.9	816.30	816.30	100		

DF: degrees of freedom; SSD: sum of squares deviations; MSD: mean square deviations; PhiPT: genetic differentiation index.

*The population refers to the origin of the samples.

TABLE 5 Genetic diversity between the Cajamarca and Amazonas populations of 27 ryegrass accessions.

Population	N	Ho	He	F _{IS}	Shannon
Cajamarca	22	0.57	0.46	-0.2312	0.8568
Amazonas	5	0.66	0.43	-0.5489	0.5722

N: sample size; Ho: observed Heterozygosity; He: expected Heterozygosity; F_{IS}: inbreeding coefficient (fixation Index).

4 Discussion

Genetic diversity is a key component of plant breeding, as it supports the selection and development of superior genotypes (Bhanu, 2017; Chauhan et al., 2025). In this context, molecular markers, particularly SNPs, have enabled a more precise characterization of genetic structure and variability within germplasm banks (Geethanjali et al., 2024). However, in outcrossing species, variation within cultivars is often high due to extensive pollen dispersal among them, which results in low levels of differentiation between populations (Roldán-Ruiz et al., 2000). These characteristic limits the effectiveness of traditional identification methods, thereby highlighting the need to employ high-throughput markers to optimize selection and accelerate the development of cultivars with superior agronomic traits and greater adaptability (Yu et al., 2022). To date, only limited research using molecular markers has been conducted to determine the genetic diversity of ryegrass in Peru. Here, for the first time, we employed genome-wide SNPs (Figure 1) to infer the genetic diversity and population structure of ryegrass germplasm from the northeastern highlands of Peru (Table 1).

A total of 2,070 high-quality SNPs were identified across the seven chromosomes of *L. multiflorum*, with SNPs detected in both genic and intergenic regions, providing a valuable resource for characterizing genetic diversity and population structure (Figure 2). These results are similar to those reported by Blackmore et al. (2016), who examined the role of genetic diversity in *Lolium perenne* using 2,199 SNPs. According to the results of the analyses of the 27 ryegrass accessions, these can be divided into two main groups (Figure 2). Group I comprised fewer representatives and consisted of local accessions from Cochán, Campiña, Santa Cruz, San Pablo, and one accession from Cutervo, all originating from the Cajamarca Region. In contrast, Group II included a larger set, composed mainly of commercial accessions from the Amazonas Region such as MAX, BIS II, INGLES, and the AGPy Wanca Grass ecotype, which are closely related to materials from Sendamal, Celendín, Tacabamba, Bambamarca, Calqui, Paccha, Baños del Inca, and El Agrario, as well as an additional individual from Cutervo and Tacabamba.

The formation of these groups may be influenced by factors such as agricultural selection or environmental conditions (Xu et al., 2017). The geographic and climatic heterogeneity of Cajamarca could restrict gene flow and promote genetic divergence (Balfourier et al., 2019). Although Cajamarca and Amazonas are not separated by a strict geographical barrier, the variation in topography, altitude, and environmental conditions across the Andean landscape can contribute to genetic structuring while maintaining overall population connectivity. Interestingly, previous research in *L. multiflorum* Lam. reported a similar pattern of genetic structuring (Tamura et al., 2022), where the analysis of 456 SNPs distinguished two main groups: one composed of domesticated cultivars and landraces, and another corresponding to the introduced cultivar ‘Gulfo’. Consistent with this, our results showed that the commercial accessions from Amazonas and the cultivar LM22 (INIA910) share the same genetic group, thus suggesting a possible common ancestor. Notably, INIA910 exhibited a close relationship with the BISII cultivar from Amazonas, forming a monophyletic clade (Figure 3). This cultivar has been reported for its high productivity and nutritional content (Vásquez et al., 2025), which may explain its regional distribution.

The Cajamarca ecotype described by Vallejos-Cacho et al. (2024) and Rojas-Vásquez et al. (2023) was located within Group I, together with the cultivars from Campiñas, Cajamarca, and San Pablo, thereby forming a monophyletic clade (Figure 3). This pattern confirms a regional genetic differentiation, likely influenced by agricultural selection processes and by the geographic and climatic heterogeneity of Cajamarca (Balfourier et al., 2019). The presence of moderate gene flow among populations suggests that genetic connectivity persists, although with marked signals of local adaptation (Cheng et al., 2020). This ecotype is also characterized by its high nutritional content (Vallejos-Cacho et al., 2024). An important breeding application is to continue improving genetic gains within each population, e.g. through recurrent selection, while maintaining genetic distance between the two groups. This strategy enables the later exploitation of hybridization power by crossing progenitors from different groups, similar to approaches currently used in other crops such as maize (Xiao et al., 2021).

The genetic structure analysis using ADMIXTURE clustered the 27 ryegrass accessions from the Peruvian highlands into two well-defined genetic populations (Figure 4), suggesting differentiated groups potentially associated with local adaptation and environmental heterogeneity. This pattern is supported by the overall F_{ST} value of 0.057, indicating moderate but evident genetic differentiation between populations. In contrast, the F_{ST} values estimated between individuals with apparently pure and mixed

ancestry within each population were low (0.0082 for Population 2 and 0.0216 for Population 1; Table 3), remaining below the threshold commonly associated with weak genetic differentiation (Hartl and Clark, 1997; Charlesworth, 1999). These results indicate limited internal differentiation and suggest that the ancestry patterns detected by ADMIXTURE are more likely related to natural variation within populations rather than strong internal subdivision. This interpretation is further supported by kinship relationships based on PL_HAT (Figure 5).

However, the presence of a third cluster ($K = 3$) indicates a finer genetic substructure that may reflect founder effects, genetic drift, or recent adaptation to local microenvironments (Supplementary Figure 1). This type of hierarchical structure is characteristic of allogamous species of the genus *Lolium*, where high gene flow coexists with the formation of genetically coherent groups associated with local selective pressures. Consistently, Kovi et al. (2015) reported $K = 2$ as the optimal number of populations in *L. perenne*, demonstrating that genetic differentiation can arise in response to specific selective pressures even in the absence of geographic isolation. Similarly, Nam et al. (2025) identified two main subpopulations ($K = 2$) in *Lolium* varieties analyzed with SSR markers, distinguishing diploid and tetraploid groups, and reported increased structural resolution when increasing to $K = 3$. In the present study, however, ploidy levels were not assessed; therefore, the finer substructure observed at $K = 3$ is more plausibly attributed to underlying genetic differentiation driven by local adaptation, demographic processes, or management-related factors rather than cytogenetic variation.

On the other hand, the integrated analysis of the productive principal components for dry matter (DM) and green forage (GF) (Supplementary Figure 3) showed that commercial accessions such as BISII, MAX, AGP, and INGL cluster in a region characterized by intermediate levels of dry matter and low green forage production, suggesting a more compact ideotype with lower aerial biomass volume but moderate productivity in terms of DM accumulation.

Genetically, individual-by-individual genetic correlation analyses (Figure 5), together with the population structure inference obtained from ADMIXTURE, indicate that these cultivars are consistently integrated within the same genetic population, a pattern that is reflected in their close proximity in the PCoA, indicating that they share a set of frequent alleles associated with this medium-to-low GF production profile (Supplementary Figure 3), consistent with improved materials selected for stability rather than maximum forage volume (Zhu et al., 2025a). In contrast, LM13, which in the PCA is located at the low end of DM and GF, also falls within this same genetic group in the PCoA but in a more peripheral position (Supplementary Figure 3), suggesting that although it shares ancestry with the commercial cultivars, it carries a less favorable allelic combination for forage productivity—possibly due to the loss of advantageous alleles or retention of variants associated with more conservative growth strategies (Verwimp et al., 2018; Nie et al., 2019). Meanwhile, ryegrass accessions L21 and L22 occupy an intermediate position in the PCA with moderate levels of DM and GF, and in ADMIXTURE they exhibit mixed ancestry with contributions from both ancestral genetic populations, which is consistent with their placement near the transition zone in the PCoA. This mixed genetic background may explain their intermediate phenotype and greater plasticity, acting as bridge genotypes between more

productive lineages and others with lower yield (Faville et al., 2017; Fu and Wang, 2023).

Previous research in ryegrass indicates that its outcrossing nature and high hybridization capacity promote extensive genetic introgression among coexisting populations, which is facilitated by wind-driven pollen dispersal and seed exchange, thereby maintaining high genetic variability within populations (Beckie and Jasieniuk, 2021). Our results align with this pattern. The AMOVA showed that 90.1% of the total genetic variation is concentrated within populations, while only 9.99% is explained among populations (Table 4). This confirms that allelic diversity is predominantly shared at the intrapopulation level, and that differentiation among groups is low. This behavior has been widely documented in ryegrass ecotypes and cultivars (Vieira et al., 2004; Guthridge et al., 2001).

The genetic structure observed suggests high gene flow, likely associated with outcrossing reproduction, efficient pollen dispersal, and management practices that favor germplasm exchange (Guan et al., 2017; Nam et al., 2023). This is reinforced by the high observed heterozygosity values ($H_o = 0.57–0.66$), which consistently exceed the expected heterozygosity ($H_e = 0.46–0.43$) (Table 5). Although this value is lower than the 0.71 reported by Nam et al. (2025) for South Korea, it is similar to the values reported by Tamura et al. (2022) (0.562) and Blackmore et al. (2016) (0.33) for the United Kingdom. The higher values observed in the Peruvian highland populations suggest greater genetic mixing or environmental variability that maintains high levels of diversity (Castro et al., 2024).

Differences among these investigations may be attributed to the type of markers used. While simple sequence repeats (SSRs) are multiallelic and have high discriminatory power, SNPs are biallelic and require a larger number of polymorphic loci to achieve equivalent resolution to that offered by multiallelic SSR markers (Guichoux et al., 2011; Singh et al., 2013). Nevertheless, the consistency observed among investigations that use SNPs reinforces the reliability of the pattern detected in the present study.

Regarding the genetic diversity profile, contrasting characteristics were observed between populations. The notably negative F_{IS} values (-0.231 in Cajamarca and -0.549 in Amazonas) indicate a systematic excess of heterozygotes in both populations. These results align with those reported in European ecotypes (-0.06 to -0.11), thereby indicating an excess of heterozygotes associated with an outcrossing mating system (Blackmore et al., 2016). Similarly, research in Norway with *L. perenne* reported values of -0.039 (Kovi et al., 2015). However, the more negative values observed in the Peruvian highland populations suggest greater gene mixing or gene flow, possibly driven by germplasm exchange among communities, traditional agricultural practices, or the coexistence of multiple genetic materials in nearby areas (Zhu et al., 2025b). The observed heterozygote excess is biologically explained by the predominantly self-incompatible reproductive system of ryegrass, which operates through a gametophytic mechanism controlled by the *S* and *Z* loci, enforcing obligate outcrossing and maintaining high levels of heterozygosity within populations despite occasional breakdowns under specific environmental conditions (Cropano et al., 2021). The excess of heterozygotes is an expected characteristic in species with an obligatory outcrossing mating system (Blackmore et al., 2015). This pattern strongly suggests the presence of heterosis or balancing selection, in which heterozygous genotypes have an adaptive advantage

(Wang et al., 2014). Although interspecific hybridization is a plausible hypothesis that could contribute to this pattern, the history of intensive breeding in ryegrass may have directly favored the retention of beneficial alleles in heterozygous condition (Labroo et al., 2021; Yannicari et al., 2018; Karn and Jasieniuk, 2017; Knorst et al., 2018). The evidence of hybrid vigor and balancing selection indicates that breeding history and management practices have favored the retention of beneficial alleles in heterozygous form, which, combined with the high intrapopulation diversity, positions these germplasms as strategic resources for breeding programs, facilitating the introgression of desirable traits and the development of adapted cultivars. Moreover, this pattern underscores the importance of their conservation in the face of genetic drift and selective pressures arising from human management and intensive cultivation (Ghamkhar et al., 2023; Knorst et al., 2019; Ozkose and Tamkoc, 2014; Yang et al., 2023).

The determination and management of ploidy level constitute a strategic axis within *Lolium* breeding programs, considering that ryegrass naturally occurs in a diploid state ($2n = 14$) and also includes artificially induced tetraploids ($4n = 28$), which have been widely incorporated into modern forage systems due to their distinct agronomic attributes (Faville et al., 2022; Tozer et al., 2017). Diploid genotypes, characterized by smaller cell size and higher sward density, have traditionally been used in older cultivars and turf-specific applications, whereas tetraploids exhibit larger cells, greater palatability and nutritional value for livestock, as well as improved stress tolerance, although they generally show lower sward density and reduced long-term persistence (Guo et al., 2018; Pembleton et al., 2018).

In the present study, although this work provides the first genomic characterization of *L. multiflorum* germplasm from the Peruvian Andes, an important limitation is the lack of ploidy evaluation in the analyzed accessions. Karyotype studies, high-throughput sequencing (HTS), and chromosome conformation capture sequencing (Hi-C) have confirmed that *L. multiflorum* is predominantly diploid ($2n = 14$) (Abbaszade et al., 2017; Pašakinskienė et al., 1997; Knorst et al., 2018; Chen et al., 2025; Brunharo et al., 2025), although induced tetraploid cultivars ($4n = 28$) have also been reported (Ahloowalia, 1967; Rocha et al., 2018). In natural populations of the Peruvian Andes, the occurrence of tetraploids has not been systematically documented; however, the presence of individuals with different ploidy levels cannot be ruled out. Therefore, the population subdivision observed at $K = 3$ should be interpreted with caution, as it may partially reflect undetermined cytogenetic variation. Furthermore, future studies should incorporate broader geographic sampling and explicitly evaluate ploidy using cytogenetic approaches (Meirmans et al., 2018; Perez-Vicente et al., 1992). Such efforts would improve diversity estimates and provide stronger support for breeding programs and conservation strategies focused on Andean germplasm.

To exploit the genetic diversity identified in the present study, ryegrass breeding programs commonly use strategies such as polycross schemes and synthetic populations (Syn1), in which multiple elite progenitors are intercrossed to maintain high levels of heterozygosity associated with forage yield and digestibility (Faville et al., 2018, 2022; Wang et al., 2016). Similarly, half-sib family selection has enabled the identification of superior maternal lines for forage

quality under controlled conditions (Wilkins, 1991; Cornish et al., 1979). In parallel, the use of doubled haploids generated through another culture have facilitated the development of homozygous materials required for breeding purposes (Begheyn et al., 2016).

In this context, the genomic diversity and population structure identified in the Peruvian *L. multiflorum* germplasm may provide useful information for future breeding programs, particularly for parent selection, conservation of locally adapted materials, and the identification of contrasting genetic backgrounds. The differentiation observed between populations, together with the high intrapopulation diversity and excess of heterozygosity, suggests the existence of valuable genetic resources that could be incorporated into improvement strategies. Beyond these conventional approaches, the application of genomic selection approaches based on GBS represents a promising future direction for ryegrass improvement. Previous studies have shown that genomic prediction can improve the selection of traits associated with dry matter yield, forage quality, persistence, disease resistance, and environmental adaptation in both diploid and tetraploid germplasm (Guo et al., 2018; Pembleton et al., 2018). Therefore, future studies integrating broader germplasm collections, phenotypic evaluations, association analyses, and genomic prediction approaches may strengthen breeding programs and accelerate cultivar development for Andean livestock systems.

Within this framework, the strategic management of germplasm with different ploidy levels, with particular emphasis on locally adapted diploid materials, emerges as a fundamental tool to enhance resilience to climatic variability while simultaneously optimizing forage productivity and quality of ryegrass in Andean livestock systems. Collectively, this work represents an initial genomic characterization of Peruvian *L. multiflorum* germplasm, laying the foundation for future studies that integrate broader collections, phenotypic assessments, and genomic prediction approaches to strengthen ryegrass breeding and conservation programs in Andean environments.

5 Conclusion

This study provides the first genomic assessment of genetic diversity of *L. multiflorum* accessions from the northern Peruvian highlands using genotype-by-sequencing. A total of 2,070 high-quality SNPs were identified, revealing the existence of two well-defined genetic populations. Cluster analysis and PCoA separated the accessions into two groups: a smaller group comprising the Cajamarca ecotype and local cultivars from Cochan, San Pablo, Campiña, Cutervo, and Santa Cruz; and a larger group including all commercial cultivars from Amazonas (Chachapoyas), along with local cultivars from Sendamal, Celendin, Bambamarca, Tacabamba, Calquis, Paccha, El Agrario, Cutervo, and the INIA 910 cultivar. These genetic relationships reflect a moderate yet significant population structure, alongside high genetic diversity within populations, particularly in Cajamarca. Collectively, these findings position Peruvian germplasm as a valuable resource for breeding programs aimed at improving forage productivity, nutritional quality, environmental adaptation, persistence, and stress tolerance.

Following agronomic and phenotypic characterization, this genetic resource may support parent selection, genomic-assisted breeding strategies, and conservation efforts, thereby contributing to the sustainability and resilience of mountain livestock systems.

Data availability statement

The raw sequencing data generated in this study have been deposited in the NCBI BioProject database under accession number PRJNA1427078.

Author contributions

LB: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – original draft. RO: Supervision, Validation, Writing – review & editing. GC: Data curation, Formal Analysis, Software, Writing – review & editing. FM: Supervision, Validation, Visualization, Writing – review & editing. LG: Data curation, Visualization, Writing – review & editing. LV: Resources, Software, Visualization, Writing – original draft. WC-C: Investigation, Resources, Writing – original draft. JD-V: Supervision, Validation, Writing – review & editing. TV: Data curation, Formal Analysis, Software, Writing – review & editing. EA: Supervision, Validation, Visualization, Writing – original draft. MC: Supervision, Validation, Visualization, Writing – review & editing. HV: Project administration, Resources, Supervision, Visualization, Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the National Council for Science, Technology and Technological Innovation (CONCYTEC) and the National Program for Scientific Research and Advanced Studies (PROCIENCIA) under call E077-2023-01-BM “Scholarships for Doctoral Programs in Interinstitutional Alliances”, grant, PE501088652-2024 and under call E033-2023-01-BM “Interinstitutional Alliances for Doctoral Programs”, grant PE501084305-2023. Finally, we express our gratitude to the Vicerrectorado de Investigación of the National University Toribio Rodríguez de Mendoza for covering the APC for the article.

References

- Abbaszade, S., Jafari, A. A., Safari, H., and Shirvani, H. (2017). Genetic variability and karyotype analysis for 13 accessions of *Lolium multiflorum*. *J. Ornamental Plants* 7, 1–8. Available online at: http://jornamental.iaurasht.ac.ir/article_528900_0134625dd7bf12d91ab55f67aa7585b0.pdf.
- Agazzi, M. A., Bocalini, F., Francisco, M. R., Reillo, P., and Silveira, L. F. (2023). Genetic variability and kinship analyses of seized red-browed amazon, *Amazona rhodocorytha* (Aves, Psittacidae). *Diversity* 15, 923. doi: 10.3390/d15080923

Acknowledgments

The authors would like to thank the Doctoral Program in Sciences for Sustainable Development of the Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas. Also, to the CONCYTEC and PROCIENCIA within the framework of the Call E033-2023-01-BM “Interinstitutional Alliances for Doctoral Programs”, under grant No. PE501084305-2023.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.2026.1818688/full#supplementary-material>

- Ahloowalia, B. S. (1967). Colchicine induced polyploids in ryegrass. *Euphytica* 16, 49–60. doi: 10.1007/bf00034098

- Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinf.* 12, 246. doi: 10.1186/1471-2105-12-246

- Altat, M., Cavagnaro, P., Kökten, K., Ali, A., Morales, A., Tatar, M., et al. (2025). Genotyping-by-sequencing derived SNP markers reveal genetic diversity and

- population structure of *Dactylis glomerata* germplasm. *Front. Plant Sci.* 16, 1530585. doi: 10.3389/fpls.2025.1530585
- Arbizu, C. I., Bazo-Soto, I., Flores, J., Ortiz, R., Blas, R., García-Mendoza, P. J., et al. (2025). Genotyping by sequencing reveals the genetic diversity and population structure of Peruvian highland maize races. *Frontiers in Plant Science* 16, 1526670. doi: 10.3389/fpls.2025.1526670
- Attali, D., and Baker, C. (2019). “ggExtra: add marginal histograms to “ggplot2”, and more “ggplot2” enhancements,” in *R package version*. Available online at: <https://CRAN.R-project.org/package=ggExtra> (Accessed December 2025).
- Auwer, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Balfourier, F., Bouchet, S., Robert, S., Oliveira, R. D., Rimbart, H., Kitt, J., et al. (2019). Worldwide phylogeography and history of wheat genetic diversity. *Sci. Adv.* 5, eaav0536. doi: 10.1126/sciadv.aav0536
- Beckie, H. J., and Jasieniuk, M. (2021). “*Lolium rigidum* and *Lolium multiflorum*,” in *Weed biology and management* (Amsterdam, The Netherlands: Elsevier), 261–283. doi: 10.1016/B978-0-12-822917-0.00017-3
- Beghey, R., Lübberstedt, T., and Studer, B. (2016). Haploid and doubled haploid techniques in perennial ryegrass (*Lolium perenne* L.) to advance research and breeding. *Agronomy* 6, 60. doi: 10.3390/agronomy6040060
- Bhanu, A. N. (2017). Assessment of genetic diversity in crop plants—an overview. *Adv. Plants Agric. Res.* 7, 255. doi: 10.15406/apar.2017.07.00255
- Blackmore, T., Thomas, I., McMahon, R., Powell, W., and Hegarty, M. (2015). Genetic-geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (*Lolium perenne*) using array-based SNP genotyping. *Theor. Appl. Genet.* 128, 1917–1932. doi: 10.1007/s00122-015-2556-3
- Blackmore, T., Thorogood, D., Skot, L., McMahon, R., Powell, W., and Hegarty, M. (2016). Germplasm dynamics: the role of ecotypic diversity in shaping the patterns of genetic variation in *Lolium perenne*. *Sci. Rep.* 6, 22603. doi: 10.1038/srep22603
- Brunharo, C. A., Short, A. W., Bobadilla, L. K., and Streisfeld, M. A. (2025). The genome of *Lolium multiflorum* reveals the genetic architecture of paraquat resistance. *Mol. Ecol.* 34, 1–15. doi: 10.1111/mec.17775
- Bunjkar, A., Walia, P., and Sandal, S. (2024). Unlocking genetic diversity and germplasm characterization with molecular markers: strategies for crop improvement. *J. Adv. Biol. Bio/Technol.* 27, 6873. doi: 10.9734/jabb/2024/v27i6873
- Cao, F., Li, W., Jiang, Y., Gan, X., Zhao, C., and Liu, J. (2024). Effects of grazing on grassland biomass and biodiversity: a global synthesis. *Field Crops Res.* 300, 109204. doi: 10.1016/j.fcr.2023.109204
- Carrasco-Chilón, W., Cervantes-Peralta, M., Mendoza, L., Muñoz-Vilchez, Y., Quilcate, C., Nuñez-Melgar, D., et al. (2024). Morphological differentiation, yield, and cutting time of *Lolium multiflorum* L. under acid soil conditions in highlands. *Plants* 13, 2331. doi: 10.3390/plants13162331
- Castro, J. C., Vasquez-Guizado, S. J., Vigil, B. E., Ascue, F., Rojas-Villa, P. N., Paredes, J. D., et al. (2024). Development and application of microsatellite markers for genetic diversity assessment and construction of a core collection of *Myrciaria dubia* (Kunth) McVaugh germplasm from the Peruvian Amazon. *Forests* 15, 1873. doi: 10.3390/f15111873
- Chang, J., Ciais, P., Gasser, T., Smith, P., Herrero, M., Havlik, P., et al. (2020). Climate warming from managed grasslands cancels the cooling effect of carbon sinks in sparsely grazed and natural grasslands. *Nat. Commun.* 12, 20406. doi: 10.1038/s41467-020-20406-7
- Charlesworth, B. (1999). Updating population genetics. *BioEssays* 20, 1055. doi: 10.1002/(SICI)1521-1878(199812)20:12<1055::AID-BIES14>3.0.CO;2-X
- Chauhan, R., Prabhakaran, S., Tiwari, A., Joshi, D. C., Chandora, R., Taj, G., et al. (2025). Genetic distances and genome-wide population structure analysis of a grain amaranth (*Amaranthus hypochondriacus*) diversity panel using genotyping by sequencing. *Sci. Rep.* 15, 33816. doi: 10.1038/s41598-025-01626-7
- Chen, Y., Kiesbauer, J., Copetti, D., Frei, D., Frey, J., Grieder, C., et al. (2025). Chromosome-level haplotype-resolved genome assembly provides insights into the highly heterozygous genome of Italian ryegrass (*Lolium multiflorum* Lam.). *Plant Genome* 18, e70079–e70079. doi: 10.1002/tpg2.70079
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Cheng, J., Kao, H., and Dong, S. (2020). Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* Maxim. revealed by reduced representation sequencing. *BMC Plant Biol.* 20, 594. doi: 10.1186/s12870-020-02594-y
- Cornish, M. A., Hayward, M. D., and Lawrence, M. J. (1979). Self-incompatibility in ryegrass. *Heredity* 43, 95–106. doi: 10.1038/hdy.1979.63
- Cropano, C., Manzanares, C., Yates, S., Copetti, D., Do Canto, J., Lübberstedt, T., et al. (2021). Identification of candidate genes for self-compatibility in perennial ryegrass (*Lolium perenne* L.). *Front. Plant Sci.* 12, 707901. doi: 10.3389/fpls.2021.707901
- 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
- Danecek, P., and McCarthy, S. A. (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33, 2037–2039. doi: 10.1093/bioinformatics/btx100
- Di Pasquale, A., Radomski, N., Mangone, I., Calistri, P., Lorusso, A., and Cammà, C. (2021). SARS-CoV-2 surveillance in Italy through phylogenomic inferences based on Hamming distances derived from pan-SNPs, -MNP and -InDels. *BMC Genomics* 22, 782. doi: 10.1186/s12864-021-08112-0
- Dray, S., and Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Software* 22, 1–20. doi: 10.18637/jss.v022.i04
- 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
- Erb, K.-H., Fetzl, T., Kastner, T., Kroisleitner, C., Lauk, C., Mayer, A., and Niedertscheider, M. (2016). Livestock grazing, the neglected land use. In: H. Haberl, M. Fischer-Kowalski, F. Krausmann and V. Winiwarter (eds.) *Social ecology: human-environment interactions*, 5. (Cham, Switzerland: Springer), 295–313. doi: 10.1007/978-3-319-33326-7_13
- Faville, M., Ganesha, S., Cao, M., Jahufer, M. Z. Z., Bilton, T., Easton, H., et al. (2017). Predictive ability of genomic selection models in a multi-population perennial ryegrass training set using genotyping-by-sequencing. *Theor. Appl. Genet.* 131, 703–720. doi: 10.1007/s00122-017-3030-1
- Faville, M. J., Ganesha, S., Cao, M., Jahufer, M. Z. Z., Bilton, T. P., Easton, H. S., et al. (2018). Predictive ability of genomic selection models in a multi-population perennial ryegrass training set using genotyping-by-sequencing. *Theor. Appl. Genet.* 131, 703–720. doi: 10.1007/s00122-017-3030-1
- Faville, M., Schmidt, J., Trollove, M., Moran, P., Hong, W., Cao, M., et al. (2022). Empirical assessment of a genomic breeding strategy in perennial ryegrass. *J. N. Z. Grassl.* 83, 115–122. doi: 10.33584/jnzg.2021.83.3490
- Fernández-Otero, C. I., Ramos-Cabrera, A. M., and Pereira-Lorenzo, S. (2024). Diversity of ecotypes of five species of ryegrass from northwestern Spain by phenotypic traits and microsatellites. *BMC Plant Biol.* 24, 5440. doi: 10.1186/s12870-024-05440-7
- Fu, R., and Wang, X. (2023). Modeling the influence of phenotypic plasticity on maize hybrid performance. *Plant Commun.* 4, 100548. doi: 10.1016/j.xplc.2023.100548
- Geethanjali, S., Kadirvel, P., and Periyannan, S. (2024). Wheat improvement through advances in single nucleotide polymorphism (SNP) detection and genotyping with a special emphasis on rust resistance. *Theor. Appl. Genet.* 137, 224. doi: 10.1007/s00122-024-04730-w
- Ghamkhar, K., Williams, W. M., and Brown, A. H. D. (2023). *Plant genetic resources for the 21st century* (New York: Apple Academic Press). doi: 10.1201/9781003302957
- Goettelman, F., Chen, Y., Knorst, V., Yates, S., Copetti, D., Studer, B., et al. (2023). High-resolution bulked segregant analysis enables candidate gene identification for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Plant J.* 118, 1516–1527. doi: 10.1111/tpj.16693
- 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.1471-8286.2004.00828.x
- Guan, X., Yuyama, N., Stewart, A., Ding, C., Xu, N., Kiyoshi, T., et al. (2017). Genetic diversity and structure of *Lolium* species surveyed on nuclear simple sequence repeat and cytoplasmic markers. *Front. Plant Sci.* 8, 584. doi: 10.3389/fpls.2017.00584
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., et al. (2011). Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* 11, 591–611. doi: 10.1111/j.1755-0998.2011.03014.x
- Guo, X., Cericola, F., Fè, D., Pedersen, M. G., Lenk, I., Jensen, C. S., et al. (2018). Genomic prediction in tetraploid ryegrass using allele frequencies based on genotyping by sequencing. *Front. Plant Sci.* 9, 1165. doi: 10.3389/fpls.2018.01165
- Guthridge, K. M., Dupal, M. P., Kölliker, R., Jones, E. S., Smith, K. F., and Forster, J. W. (2021). AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). *Euphytica* 122, 191–201. doi: 10.1023/A:1012658315290
- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell. Sys. Tech. J.* 29, 147–160. doi: 10.1002/j.1538-7305.1950.tb00463.x
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2
- Hartl, D. L., and Clark, A. G. (1997). *Principles of population genetics*. 3rd ed (Massachusetts, USA: Sinauer Associates).
- Hewins, D. B., Lyseng, M. P., Schoderbek, D. F., Alexander, M., Willms, W. D., Carlyle, C. N., et al. (2018). Grazing and climate effects on soil organic carbon concentration and particle-size association in northern grasslands. *Sci. Rep.* 8, 19785. doi: 10.1038/s41598-018-19785-1
- Jaškūnė, K., Aleliūnas, A., Statkevičiūtė, G., Kemešytė, V., Studer, B., Yates, S., et al. (2020). Genome-wide association study to identify candidate loci for biomass formation under water deficit in perennial ryegrass. *Front. Plant Sci.* 11, 570204. doi: 10.3389/fpls.2020.570204

- Jiao, J., Li, S., Wang, W., Qi, L., Degen, A., Bai, Y., et al. (2024). Land parcellation reduced spatial redundancy resulting in grassland degradation. *People Nat* 6, 2528–2542. doi: 10.1002/pan3.10725
- Johansen, N. H., Bellucci, A., Hansen, P. B., Marum, P., Amdahl, H., Gylström, K. H., et al. (2025). Genomic prediction of agronomic traits in perennial ryegrass (*Lolium perenne* L.) and genotype × environment interactions at the limit of the species distribution. *Theor. Appl. Genet.* 138, 281. doi: 10.1007/s00122-025-05064-x
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Kanaka, K. K., Sukhija, N., Goli, R. C., Singh, S., Ganguly, I., Dixit, S. P., et al. (2023). On the concepts and measures of diversity in the genomics era. *Curr. Plant Biol.* 33, 100278. doi: 10.1016/j.cpb.2023.100278
- Karn, E., and Jasieniuk, M. (2017). Genetic diversity and structure of *Lolium perenne* ssp. multiflorum in California vineyards and orchards indicate potential for spread of herbicide resistance via gene flow. *Evol. Appl.* 10, 616–629. doi: 10.1111/eva.12478
- Knaus, B. J., and Grünwald, N. J. (2017). VcfR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17, 44–53. doi: 10.1111/1755-0998.12549
- Knorst, V., Byrne, S., Yates, S., Asp, T., Widmer, F., Studer, B., et al. (2018). Pooled DNA sequencing to identify SNPs associated with a major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Theor. Appl. Genet.* 132, 947–958. doi: 10.1007/s00122-018-3250-z
- Knorst, V., Yates, S., Byrne, S., Asp, T., Widmer, F., Studer, B., et al. (2019). First assembly of the gene-space of *Lolium multiflorum* and comparison to other Poaceae genomes. *Grassl. Sci.* 65, 125–134. doi: 10.1111/grs.12225
- Kolde, R. (2019). pheatmap: pretty heatmaps. R package version 1.0.12. Available online at: <https://CRAN.R-project.org/package=pheatmap> (Accessed December 2025).
- Kovi, M. R., Fjellheim, S., Sandve, S. R., Larsen, A., Rudi, H., Asp, T., et al. (2015). Population structure, genetic variation, and linkage disequilibrium in perennial ryegrass populations divergently selected for freezing tolerance. *Front. Plant Sci.* 6, 929. doi: 10.3389/fpls.2015.00929
- Kumar, R., Das, S., Choudhury, B., Kumar, A., Prakash, N., Verma, R., et al. (2024). Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing. *Biol. Res.* 57, 562. doi: 10.1186/s40659-024-00562-6
- Labroo, M. R., Studer, A. J., and Rutkoski, J. E. (2021). Heterosis and hybrid crop breeding: a multidisciplinary review. *Front. Genet.* 12, 643761. doi: 10.3389/fgene.2021.643761
- Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. doi: 10.1093/nar/gkac268
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Meirmans, P. G., Liu, S., and Van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *J. Heredity* 109, 283–296. doi: 10.1093/jhered/esy00
- Miles, A., Bot, P., Rodrigues, M. F., Ralph, P., Kelleher, J., Schelker, M., et al. (2024). cggh/scikit-allel: v1.3.8 (Version 1.3.8) [Software]. *Zenodo*. doi: 10.5281/zenodo.10876220
- Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: synthesis* (Washington, DC, USA: Island Press).
- Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43, 1235–1248. doi: 10.2135/cropsci2003.1235
- Nam, D., Baek, E., Hwang, E., Gwak, S., Lee, Y., Cho, S., et al. (2025). Genetic diversity patterns within and among varieties of Korean Italian ryegrass (*Lolium multiflorum*) and perennial ryegrass (*Lolium perenne*) based on simple sequence repetition. *Agriculture* 15, 224. doi: 10.3390/agriculture15030244
- Nam, D.-G., Gwak, S.-C., Baek, E.-S., Lee, Y.-H., Choi, B. R., and Hwang, T.-Y. (2023). The current status of breeding research in *Lolium* genus. *J. Crop Sci. Biotechnol.* 26, 649–659. doi: 10.1007/s12892-023-00226-z
- Nie, G., Huang, T., Xu, X., Huang, L., Peng, Y., Yan, Y., et al. (2019). Genetic variability evaluation and cultivar identification of tetraploid annual ryegrass using SSR markers. *PeerJ* 7, e7742. doi: 10.7717/peerj.7742
- Ozkose, A., and Tamkoc, A. (2014). Morphological and agronomic characteristics of perennial ryegrass (*Lolium perenne* L.) genotypes. *Turk. J. Field Crops* 19, 231. doi: 10.17557/tjfc.15567
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Pašakinskienė, I., Anamthawat-Jónsson, K., Humphreys, M. W., and Jones, R. N. (1997). Novel diploids following chromosome elimination and somatic recombination in *Lolium multiflorum* × *Festuca arundinacea* hybrids. *Heredity* 78, 464–469. doi: 10.1038/hdy.1997.74
- Pedersen, B. S., and Quinlan, A. R. (2017). cyvcf2: fast, flexible variant analysis with Python. *Bioinformatics* 33, 1867–1869. doi: 10.1093/bioinformatics/btx057
- Pembleton, L. W., Inch, C., Baillie, R. C., Drayton, M. C., Thakur, P., Ogaji, Y. O., et al. (2018). Exploitation of data from breeding programs supports rapid implementation of genomic selection for key agronomic traits in perennial ryegrass. *Theor. Appl. Genet.* 131, 1891–1902. doi: 10.1007/s00122-018-3121-7
- Perez-Vicente, R., Petris, L., Osusky, M., Potrykus, I., and Spangenberg, G. (1992). Molecular and cytogenetic characterization of repetitive DNA sequences from *Lolium* and *Festuca*: applications in the analysis of *Festulolium* hybrids. *Theor. Appl. Genet.* 84, 145–154. doi: 10.1007/bf00223994
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Ragsdale, A. P., and Gravel, S. (2020). Unbiased estimation of linkage disequilibrium from unphased data. *Mol. Biol. Evol.* 37, 923–932. doi: 10.1093/molbev/msz265
- Rocha, L. C., Lopes, L. S., Vitoriano, M. B. Z., Mittelman, A., and Techio, V. H. (2018). Meiotic irregularities in diploid and tetraploid plants of *Lolium multiflorum* Lam. *Biologia* 73, 1145–1152. doi: 10.2478/s11756-018-0120-8
- Rojas-Vásquez, Z., Gutiérrez-Arce, F., and Gutiérrez-Arce, W. (2023). Efecto de la altura de pastura sobre el pH ruminal y producción láctea en vacas Holstein en una pradera de rye grass – ecotipo cajamarquino (*Lolium multiflorum*) y trébol blanco (*Trifolium repens*). *Rev. Vet. Zootec. Amaz.* 3, e400. doi: 10.51252/revza.v3i1.400
- Roldán-Ruiz, I., Dendaau, J., Van Bockstaele, E., Depicker, A., and De Loose, M. (2000). AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol. Breed.* 6, 125–134. doi: 10.1023/A:1009680614564
- Rui, H., Luo, B., Wang, Y., Zhu, L., and Zhu, Q. (2025). Quantitative impacts of climate change and human activities on grassland growth in Xinjiang, China. *Front. Plant Sci.* 15, 1497248. doi: 10.3389/fpls.2024.1497248
- Sanjeevan, V., and König, P. (2025). Efficient phylogenetic inference using SNP-based approaches: a comparison with full sequence data. *bioRxiv*. doi: 10.1101/2025.03.10.642383
- Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., et al. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8, e84136. doi: 10.1371/journal.pone.0084136
- Song, S., Wang, S., Li, N., Chang, S., Dai, S., Guo, Y., et al. (2023). Genome-wide association study to identify SNPs and candidate genes associated with body size traits in donkeys. *Front. Genet.* 14, 1112377. doi: 10.3389/fgene.2023.1112377
- Talevich, E., Invergo, B. M., Cock, P. J. A., and Chapman, B. A. (2012). BioPhylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinf.* 13, 209. doi: 10.1186/1471-2105-13-209
- Tamura, K., Arakawa, A., Kiyoshi, T., and Yonemaru, J. (2022). Genetic diversity and structure of diploid Italian ryegrass (*Lolium multiflorum* Lam.) cultivars and breeding materials in Japan based on genome-wide allele frequency. *Grassl. Sci.* 68, 263–276. doi: 10.1111/grs.12363
- Tozer, K. N., Carswell, K., Griffiths, W. M., Crush, J. R., Cameron, C. A., Chapman, D. F., et al. (2017). Growth responses of diploid and tetraploid perennial ryegrass (*Lolium perenne*) to soil-moisture deficit, defoliation and a root-feeding invertebrate. *Crop Pasture Sci.* 68, 632–642. doi: 10.1071/CP17154
- Vallejos-Cacho, R., Vallejos-Fernández, L., Alvarez-García, W., Tapia-Acosta, E., Saldanha-Odrizola, S., and Quilcate-Pairazaman, C. (2024). Sustainability of *Lolium multiflorum* L. ‘Cajamarquino Ecotype’, associated with *Trifolium repens* L. at three cutting frequencies in the northern highlands of Peru. *Sustainability* 16, 6927. doi: 10.3390/su16166927
- Vallejos-Fernández, L. A., Alvarez, W. Y., Paredes-Arana, M. E., Pinares-Patiño, C., Bustíos-Valdivia, J. C., Vásquez, H., et al. (2020). Comportamiento productivo y valor nutricional de 22 genotipos de raigrás (*Lolium* spp.) en tres pisos altoandinos del norte de Perú. *Sci. Agropecu.* 11, 537–545. doi: 10.17268/sci.agropecu.2020.04.09
- Vásquez, H., Valqui, L., Bobadilla, L., Meseth, E., Trigos, M., Zagaceta, L., et al. (2025). Agronomic and nutritional evaluation of INIA 910—Kumymarca ryegrass (*Lolium multiflorum* Lam.): an alternative for sustainable forage production in the Department of Amazonas (NW Peru). *Agronomy* 15, 100. doi: 10.3390/agronomy15010100
- Verwimp, C., Ruttink, T., Muylle, H., Van Glabeke, S., Cnops, G., Quataert, P., et al. (2018). Temporal changes in genetic diversity and forage yield of perennial ryegrass in monoculture and in combination with red clover in swards. *PLoS One* 13, e0206571. doi: 10.1371/journal.pone.0206571
- Vieira, E. A., Castro, C. M., Oliveira, A. C., Carvalho, F. I. F., Zimmer, P. D., and Martins, L. F. (2004). Genetic structure of annual ryegrass (*Lolium multiflorum*) populations estimated by RAPD. *Sci. Agric.* 61, 407–413. doi: 10.1590/S0103-90162004000400009
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

- Wang, C., Kao, W.-H., and Hsiao, C. K. (2015). Using hamming distance as information for SNP-sets clustering and testing in disease association studies. *PLoS One* 10, e0135918. doi: 10.1371/journal.pone.0135918
- Wang, J., Pembleton, L. W., Baillie, R. C., Drayton, M. C., Hand, M. L., Bain, M., et al. (2014). Development and implementation of a multiplexed single nucleotide polymorphism genotyping tool for differentiation of ryegrass species and cultivars. *Mol. Breed.* 33, 435–451. doi: 10.1007/s11032-013-9961-6
- Wang, J., Pembleton, L., Cogan, N., and Forster, J. (2016). Evidence for heterosis in Italian ryegrass (*Lolium multiflorum* Lam.) based on inbreeding depression in F2 generation offspring from biparental crosses. *Agronomy* 6, 49. doi: 10.3390/agronomy6040049
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. (New York, NY, USA: Springer-Verlag). doi: 10.1007/978-3-319-24277-4
- Wilkie, P., Dalberg, A., Harris, D., and Forrest, L. L. (2013). The collection and storage of plant material for DNA extraction: the teabag method. *Gard. Bull. Singap.* 65, 231–234.
- Wilkins, P. W. (1991). Breeding perennial ryegrass for agriculture. *Euphytica* 52, 201–214. doi: 10.1007/BF00029397
- Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* 22, 148. doi: 10.1186/s13059-021-02370-7
- Xu, B., Sun, G., Wang, X., Lu, J., Wang, I. J., and Wang, Z. (2017). Population genetic structure is shaped by historical, geographic, and environmental factors in the leguminous shrub *Caragana microphylla* on the Inner Mongolia Plateau of China. *BMC Plant Biol.* 17, 147. doi: 10.1186/s12870-017-1147-7
- Yang, M., Zhao, H.-Y., Xian, X., Wang, R., Ni, Y., Chen, L., et al. (2023). Assessing risk from invasive alien plants in China: reconstructing invasion history and estimating distribution patterns of *Lolium temulentum* and *Aegilops tauschii*. *Front. Plant Sci.* 14, 1113567. doi: 10.3389/fpls.2023.1113567
- Yanniccari, M., Istilart, C., Giménez, D., and Castro, A. (2018). Dynamics of glyphosate resistance in *Lolium perenne*: pollen-mediated gene flow under field conditions in Argentina. *Agronomy* 8, 286. doi: 10.3390/agronomy8120286
- Yu, Q., Yao, L., Xiong, Y., Zhao, W., Xiong, Y., Dong, Z., et al. (2022). RAD-seq as an effective strategy for heterogeneous variety identification in plants—a case study in Italian ryegrass (*Lolium multiflorum*). *BMC Plant Biol.* 22, 3617. doi: 10.1186/s12870-022-03617-6
- Zhu, J., Giri, K., Lin, Z., Cogan, N., Jacobs, J., and Smith, K. (2025b). Estimation of ryegrass (*Lolium*) dry matter yield using genomic prediction considering genotype by environment interaction across south-eastern Australia. *Front. Plant Sci.* 16, 1579376. doi: 10.3389/fpls.2025.1579376
- Zhu, J., Malmberg, M., Shinozuka, M., Retegan, R., Cogan, N., Jacobs, J., et al. (2025a). Machine learning solutions for integrating partially overlapping genetic datasets and modelling host–endophyte effects in ryegrass (*Lolium*) dry matter yield estimation. *Front. Plant Sci.* 16, 1543956. doi: 10.3389/fpls.2025.1543956