Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Leveraging a SNP-based genomic relationship matrix to enhance GWAS for spittlebug resistance and mode of reproduction in tetraploid Urochloa hybrids

Rosangela Maria Simeão
rosangela.simeao@embrapa.br
Brazilian Agricultural Research Corporation
Sanzio Carvalho Lima Barrios
Brazilian Agricultural Research Corporation
Karem Guimarães Xavier Meireles
Brazilian Agricultural Research Corporation
Fabrícia Zimermann Vilela Torres
Brazilian Agricultural Research Corporation
Aline da Costa Lima Moraes
University of Campinas (UNICAMP)
Yasmin Lany Ventura Said
Andrea Raposo
Brazilian Agricultural Research Corporation
Eric Ragalzi
University of Campinas (UNICAMP)
Anete Pereira Souza
University of Campinas (UNICAMP)

Research Article

Keywords: apomixis, Brachiaria, forage breeding, kinship, marker-assisted selection, Notozulia entreriana, polyploidy

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10553865/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Table 1 is available in the Supplementary Files section.

Leveraging a SNP-based genomic relationship matrix to enhance GWAS for spittlebug resistance and mode of reproduction in tetraploid *Urochloa* hybrids

Authors

Rosangela Simeão - Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP Campinas), São Paulo, Brazil, E-mail: rosangela.simeao@embrapa.br (Orcid 0000-0001-8351-846X). Corresponding author

Sanzio Carvalho Lima Barrios - Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: sanzio.barrios@embrapa.br, (Orcid 0000-0002-5490-4959)

Karem Guimarães Xavier Meireles - Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: karem.meireles@embrapa.br, (Orcid 0000-0003-3753-5693)

Fabrícia Zimermann Vilela Torres - Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: fabricia.torres@embrapa.br, (Orcid 0000-0002-1035-3518)

Aline da Costa Lima Moraes - Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: aclima@unicamp.br, (Orcid 0000-0002-6985-4392)

Yasmin Lany Ventura Said-Postgraduate Program in Biotechnology, School of Pharmaceutical Sciences, Food, and Nutrition, Federal University of Mato Grosso do Sul (UFMS), Campo Grande, Mato Grosso do Sul, Brazil, E-mail: yasminvsaid@gmail.com, (Orcid 0000-0003-0819-2011)

Andrea Raposo - Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil and Center for Plant Molecular Breeding (CeM²P), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: andrea.raposo@embrapa.br, (Orcid 0000-0001-9640-4218)

Eric Ragalzi - Postgraduate Program in Genetics and Molecular Biology, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: ericragg@gmail.com, (Orcid 0009-0009-6230-3211)

Anete Pereira de Souza - Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil and Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, E-mail: anete@unicamp.br, (Orcid: 0000-0003-3831-9829)

Abstract

Breeding programs for the tropical forage grass *Urochloa* require the evaluation of interspecific hybrids derived from crosses between apomictic and sexual tetraploid species within the same agamic complex. We evaluated 45,937 genotyping-by-sequencing (GBS)-derived single-nucleotide polymorphism (SNP) markers via dosage-aware calls and a genomic relationship matrix (**G**) to refine the pedigrees of 245 putative hybrids and 14 parents. Genome-wide association studies (GWASs) based on realized kinship coefficients were conducted to dissect labor-intensive and low-throughput traits, specifically spittlebug (*Notozulia entreriana*) resistance and reproduction modes. Classification via the **G**-matrix identified 23.0% of the progeny as true hybrids, 48.5% as *U. brizantha* pollen contaminants, and 19.3% as apomictic clones. Pedigree correction combined with a permissive minor allele frequency threshold (MAF ≥ 0.05) enabled the detection of two dominant loci associated with damage resistance (*qDam-dom1* and *qDam-dom2*, explaining ~48% of the genetic variance) and one additive locus (*qSur-add1*) associated with nymph survival. Additionally, three loci associated with apomixis (*qApom-dom1*, *qApom-dom2*, and *qApom-dom3*) mapped independently of the resistance loci. These findings demonstrate that dosage-aware genomic matrix filtering, combined with polyploid-specific association models and permissive MAF thresholds, is essential for dissecting complex traits and accelerating genomic-assisted selection in polyploid *Urochloa* breeding programs.

Keywords: apomixis, *Brachiaria*, forage breeding, kinship, marker-assisted selection, *Notozulia entreriana*, polyploidy

Introduction

Apomictic, tetraploid cultivars of the tropical grass forage *Urochloa* P. Beauv. (syn. *Brachiaria*) occupy approximately 80% of Brazil′s 100 million hectares of cultivated pastureland. Until 2013, four primary cultivars were released through phenotypic evaluation and selection from the wild, mostly originating from East Africa (Valle et al. 2008; Jank et al. 2014). This selection process was initially successful and simplified by apomixis, an asexual mode of reproduction via seeds, present in *Urochloa brizantha* and *U. humidicola* (Euclides et al. 2010; Jank et al. 2014). However, the reliance on apomictic accessions has created a severe genetic bottleneck, limiting agronomic traits such as productivity, nutritional value and resistance to pests (Miles 2007; Worthington and Miles 2015). This narrow genetic base necessitated interspecific hybridization and the adoption of modern breeding strategies (Jank et al. 2025).

To address these limitations, modern breeding programs in *Urochloa* utilize hybridization schemes that cross artificially chromosome-doubled (tetraploidized) sexual females, primarily *U. ruziziensis* (Swenne et al. 1981; Lutts et al. 1991) and *U. decumbens* (Simioni and Valle 2009), with elite, obligate apomictic paternal accessions (*U. brizantha* and *U. decumbens*), or superior apomictic hybrids (Miles 2007; Barrios et al. 2013; Gouveia et al. 2022; Correa et al. 2025). These three East African *Urochloas* belong to the same agamic complex, share a basic chromosome number of $x = 9$,

and intercross readily (Torres-González and Morton 2005; Risso-Pascotto et al. 2006; Fuzinatto et al. 2012; Pinto de Paula et al. 2017). Although these interspecific crosses yield massive hybrid populations segregating for sexual and apomictic modes of reproduction (Worthington et al. 2019), advancing elite candidates for commercial release remains challenging (Miles 2007). Hybrids must undergo labor-intensive, low-throughput screening for modes of reproduction, spittlebug resistance, forage yield and nutritional quality, and seed yield. Currently, with the exception of early-stage screening for spittlebug damage (Gouveia et al. 2022), most critical agronomic traits and modes of reproduction are evaluated only in advanced breeding stages after candidates with high antibiosis are prioritized.

In this context, high-throughput genotyping via high-density single-nucleotide polymorphism (SNP) datasets from next-generation sequencing (NGS) offers unprecedented opportunities to accelerate selection in long-lived species of *Urochloa* species (Lara et al. 2019; López-Contreras et al. 2025). When conducting genome-wide association studies (GWASs) with high-density markers are conducted, explicitly accounting for population structure, allelic dosage, and kinship significantly increases statistical power (Wang et al. 2020; Tibbs-Cortes et al. 2021; Karunarathne et al. 2025). Genomic relationship matrices provide a robust framework for polyploid breeding programs involving interspecific hybrids. These models incorporate classical parameterization VanRaden (2008) as well as methods tailored for segmental allotetraploids such as *U. brizantha* ($2n = 4x = 36$) (Slater et al. 2014). By utilizing realized kinship coefficients, these matrices accurately account for polysomic inheritance, allelic dosage, and specific scaling parameters. Ultimately, precise genetic characterization optimizes breeding workflows by partitioning populations, effectively separating apomictic hybrids, potential cultivar candidates, from sexual hybrids destined for recurrent selection cycles. Furthermore, capturing genome-wide background variation enhances GWAS detection of loci controlling apospory, providing a powerful, complementary strategy to distinguish facultative apomicts and sexual hybrids early in the breeding cycle (Worthington et al. 2016; Matias et al. 2019; Jank et al. 2025).

Additionally, this unified association mapping framework can be concurrently deployed to dissect complex biotic stress responses within the partitioned population. For example, mapping loci associated with antibiosis to spittlebugs, the most damaging pasture pest in Brazil, requires the integration of comprehensive phenotypic data (combining reduced plant damage and low nymph survival) with genotypic data through GWAS. Identifying these markers has the potential to revolutionize *Urochloa* breeding by enabling high-throughput early-generation selection, accelerating genetic gains, and identifying superior resistant hybrids without prematurely narrowing the population's genetic diversity.

Therefore, the main objective of this study was to develop and validate a genome-assisted selection strategy for tetraploid *Urochloa* by integrating experimental phenotyping, kinship modeling, genome-wide mapping of spittlebug

(*Notozulia entreriana* (Berg) (Hemiptera: Cercopidae)) with antibiosis resistance, and marker-based population partitioning. Specifically, our objectives were as follows:

1. Validate pedigree assignments and model realized kinship: Utilizing genotyping-by-sequencing (GBS)-derived SNP markers in an interspecific hybrid population, we accounted for complex segmental allotetraploid inheritance (mixed disomic–polysomic segregation) and allelic dosage, ranging from simplex to quadruplex, via appropriately parameterized genomic matrices, establishing a robust genetic baseline for quantitative polyploid analyses.
2. Identifying and characterizing loci associated with spittlebug antibiosis: High-density genotypic data were combined with a composite phenotypic index, integrating plant damage scores and nymph survival, in a GWAS framework to map major-effect quantitative trait loci (QTLs). Furthermore, we characterized putative "triple-carrier" hybrids, in which individuals accumulate favorable allele dosages across multiple independent loci, to evaluate cumulative gene action, potential epistatic synergy, and the convergence of maternal and paternal resistance alleles.
3. Characterize the genomic architecture governing mode of reproduction: Using the diagnostic SCAR marker p779/p780 as a proxy for apomixis, we screened the hybrid population to assign presumptive reproductive phenotypes (apomictic vs. sexual) for inclusion in a GWAS targeting mode of reproduction. This enabled cross-validation of the predictive accuracy of the SCAR marker in an interspecific tetraploid background and confirmed whether marker-derived classifications correspond to genomic regions governing apospory.

Materials and methods

Plant material

Controlled and open pollination crosses were performed among five apomictic male parents and ten sexual or facultative apomictic female parents, with each cross designated by a specific code (Table 1). The male parents included three apomictic cultivars derived from natural accessions, *Urochloa brizantha* cv. Marandu and accession B4–2, *U. decumbens* cv. Basilisk, and two apomictic cultivars obtained from interspecific hybridization between three species (*Urochloa* spp. cvs. BRS Ipyporã and Cayman). Henceforth, the male parents are referred to as MAR, B4–2, BAS, IPY, and CAY. The female parents were sexual or facultative apomictic genotypes derived from intra- or interspecific hybridization. For controlled pollination, inflorescences of IPY were collected the day before anthesis and kept in water overnight, and pollen was harvested on the day of hand pollination of accession R81 maintained in pots in a greenhouse. For open pollination, the maternal plant was established in a location extensively surrounded by the designated apomictic male

parent, and seeds were harvested after full maturation. The crosses generated numerous putative hybrids and, following initial screening for spittlebug damage response, subsequently were characterized alongside the parental genotypes through molecular analyses.

Methods

Notozulia entreriana resistance bioassay

The methodology described below was adapted from those proposed by Sotelo et al. (1998) and Cardona et al. (1999). A total of 171 *Urochloa* hybrids were evaluated for two resistance parameters, nymph survival and nymph-induced plant damage, whereas an additional 30 hybrids were evaluated for nymph-induced plant damage only. The following apomictic male parent cultivars were included as standard controls: BAS (susceptible to spittlebugs), MAR, and IPY (both resistant to spittlebugs).

For bioassay establishment, genotypes were initially sown in plug trays and, after approximately 60 days, transplanted into pots. Following an additional 30-day growth period, tillers were detached, individualized, and transferred to a hydroponic system for approximately ten days to promote root development. The rooted tillers were then placed individually into experimental units consisting of two PVC fittings: a 6.5 cm long × 5.3 cm inner diameter coupling, which functions as a miniature pot, and a 3 cm long × 4.8 cm outer diameter reducer, which fits into the coupling and serves as a lid. The lid contained a 2.2 cm diameter central aperture through which the plant culm was threaded and allowed to develop. The base of the coupling was sealed with a disc cut from a polystyrene cup. Each experimental unit was filled with a 4.5 cm layer of commercial potting substrate. Tillers were positioned on top of the substrate layer, and the lid was secured in place by inserting a piece of foam into the central aperture around the culm. After ten days, once the tillers had developed a well-established root system with abundant superficial roots, the plants were infested with *Notozulia entreriana* eggs at the near-eclosion stage (five eggs per plant; two replicates per genotype).

At 25 days, nymph-induced plant damage was assessed via a visual rating scale ranging from 1–5, as follows: 1 - no visible damage; 2 - 25% of the leaf area affected; 3 - 50% of the leaf area affected; 4 - 75% of the leaf area affected; and 5 – the entire leaf area affected and desiccated. Nymph survival was subsequently recorded by counting the number of fifth-instar nymphs and adults present in each experimental unit.

Mode of reproduction determination via a PCR-based marker assay

The mode of reproduction of 42 hybrids from cross CRZ–22 and the male parent hybrid IPY was determined via the SCAR molecular marker p779/p780 (forward primer: 5'-TATGTCACGACAAGAATATG-3'; reverse primer: 5'-TGTAACCATAACTCTCAGCT-3'), which was originally developed for *Pennisetum* (Akiyama et al. 2011) and

validated for *Urochloa* by Worthington et al. (2016). PCR amplification was performed via a touchdown (TD) protocol with an annealing temperature of 57°C, which improves primer specificity by promoting perfect primer-template hybrids and eliminating nonspecific amplification. The amplified products were resolved via electrophoresis on 1.5% agarose gels (1.5 g of agarose in 100 mL of 1× TBE buffer), stained with GelRed® (Biotium), and imaged via a GelDoc XR+ system (Bio-Rad). The presence of an amplicon of approximately 900–950 bp suggests an apomictic individual, whereas the absence of the band indicates sexual reproduction. The marker diagnostic accuracy of markers has been reported to be approximately 85% (Worthington et al. 2016).

Genomic data, quality control, and SNP filtering

The 245 hybrid individuals and 15 parental genotypes, including the 42 hybrids previously evaluated via a PCR marker assay, were genotyped-by-sequencing (GBS) following the methods of Poland et al. (2012). Reads were aligned to the *U. ruziziensis* reference genome (Pessoa-Filho et al. 2019) via *Bowtie2* (Langmead and Salzberg 2012), and SNP calling was performed with the *TASSEL-GBS* pipeline (Glaubitz et al. 2014) modified to obtain allele-specific read counts (Pereira et al. 2018). Biallelic SNPs with a minimum sequencing depth of 20 and genotyped in at least 75% of the individuals were retained via *VCFtools* (Danecek et al. 2011). The allele dosage was inferred with the *Updog* package (Gerard et al. 2018) under the norm model, retaining markers scored in at least 80% of the individuals. The genomic data reflected tetraploid genetics, in which genotypes are encoded as discrete alternative allele dosages ranging from 0 to 4. Thus, dosages classify individual SNPs as nulliplex (e.g., aaaa), simplex (e.g., Aaaa), duplex (e.g., AAaa), triplex (e.g., AAAa) or quadruplex (e.g., AAAA).

The female-parent R81 failed GBS quality control and was excluded from all the genomic analyses, resulting in 259 individuals being retained. For the damage and survival GWAS analyses, two hybrids were further excluded because their extreme phenotypic outlier values were inconsistent with the experimental scoring scale (H300–74 and H358–74), likely reflecting recording errors.

Phenotypic and genotypic data analysis and estimation of genetic effects

Phenotypic evaluations of nymph survival and nymph-induced plant damage were conducted under an unbalanced completely randomized design with two replications per genotype. The genotypic values were predicted as best linear unbiased predictors (BLUPs) for all 201 hybrids and three male parents phenotyped and genotyped. This was achieved by fitting a mixed linear model with genotype as a random effect via restricted maximum likelihood (REML), utilizing the *lme4* (Bates et al. 2015) and *lme4GS* (Caamal-Pat et al. 2021) packages. The linear mixed model (Caamal-Pat et al. 2021) is written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \qquad (1)$$

where, $\boldsymbol{y}$ is a response vector of dimensions $\boldsymbol{n} \times \boldsymbol{1}$, $\boldsymbol{X}$ is a matrix of fixed effects of dimensions $\boldsymbol{n} \times \boldsymbol{r}$, $\boldsymbol{\beta}$ is a vector of fixed effects of dimensions $\boldsymbol{r} \times \boldsymbol{1}$; $\boldsymbol{Z}$ is an incidence matrix of dimensions $\boldsymbol{n} \times \boldsymbol{r}$; and $\boldsymbol{u}$ is a vector of random effects, in which, $\boldsymbol{u} \sim \boldsymbol{MN}\ (\boldsymbol{0}, \boldsymbol{K})$ and $\boldsymbol{e} \sim \boldsymbol{MN}(\boldsymbol{0}, \boldsymbol{\sigma_e^2 I})$, with $\boldsymbol{K}$ being a known variance–covariance matrix; and $\boldsymbol{\sigma_g^2}$ and $\boldsymbol{\sigma_e^2}$ are variance parameters associated with $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$, assuming that $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ are independently distributed.

In the present study, $\boldsymbol{y}$ comprised the observed phenotypic values for damage score or nymph survival across all evaluated genotypes and replicates. The fixed effects matrix $\boldsymbol{X}$ included the overall intercept and the replicate effect, whereas the random effects vector $\boldsymbol{u}$ represented genotype effects assumed to follow a normal distribution with mean zero and variance–covariance structure proportional to the identity matrix ($\boldsymbol{K} = \boldsymbol{I}$), which is consistent with an identity-based (nongenomic) mixed model. The variance components $\boldsymbol{\hat{\sigma}_g^2}$ and $\boldsymbol{\hat{\sigma}_e^2}$ were estimated via restricted maximum likelihood (REML) via the *lme4* package.

Once the mixed model was fitted, the conditional means of the random effects were obtained as best linear unbiased prediction (BLUP) for $\boldsymbol{u}$, as follows:

$$\boldsymbol{\hat{u}^* = \hat{\sigma}_g^2 Z^{*\prime} \hat{V}^{*-1} (y - X\hat{\beta})} \qquad (2)$$

where $\boldsymbol{\hat{V}^* = \hat{\sigma}_e^2 Z^* Z' + \hat{\sigma}_g^2 I}$, with $\boldsymbol{\hat{\sigma}_e^2}$, $\boldsymbol{\hat{\sigma}_g^2}$, and $\hat{\beta}$ denoting the REML estimates of the residual variance, genotypic variance, and vector of fixed effects, respectively. The conditional means of the random effects for the model in equation (2) were obtained as:

$$\boldsymbol{\hat{u} = \Gamma \Lambda^{1/2} \hat{u}^*} \qquad (3)$$

Eigenvalue decomposition was used, where $\boldsymbol{\Gamma}$ contains the eigenvectors and where $\boldsymbol{\Lambda}$ contains the eigenvalues of $\boldsymbol{K}$.

The resulting genotypic effects represent deviations of each genotype from the population mean, adjusted for replicate effects, and were used as the phenotypic input for all subsequent GWAS analyses. This approach eliminates confounding between genotypic and environmental effects while preserving the genetic signal across genotypes, a critical requirement for association mapping in unbalanced experimental designs (Caamal-Pat et al. 2021).

Broad-sense heritability was estimated both by the conventional method $\boldsymbol{H^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}}$ and by the *Cullis* method (Cullis et al. 2006), which accounts for the precision of BLUP estimation:

$$\boldsymbol{H^2_{Cullis} = 1 - \frac{PEV}{2\hat{\sigma}_g^2}} \qquad (4)$$

where $\boldsymbol{\overline{PEV}}$ is the mean prediction error variance of the BLUP, extracted from the conditional variance of the random effects. Heritability was estimated under two frameworks: (i) using the identity covariance matrix $^{**}\boldsymbol{K}^{**} = {}^{**}\boldsymbol{I}^{**}$ with all evaluated genotypes (*lme4*, $n = 201$) and (ii) using the genomic relationship matrix $\boldsymbol{G}$ as the covariance kernel (GBLUP, *sommer* package; Covarrubias-Pazaran 2016) restricted to genotypes with both phenotypic and genotypic data ($n = 201$).

Genomic relationship matrix and parentage verification

The genomic relationship matrix ($\boldsymbol{G}$) was estimated following the VanRaden (2008) methodology adapted for polyploid species via the *AGHmatrix* R package (Amadeu et al. 2016), and cross-validated with the method proposed by Slater et al. (2014). Because the evaluated progeny can exhibit diverse genetic origins driven by the coexistence of sexual reproduction, apospory, pseudogamy, potential self-fertilization, and pollen contamination under open pollination, the G matrix serves as a robust diagnostic tool to dissect these complex reproductive outcomes. Specifically, reproductive variations were partitioned on the basis of realized pairwise genomic relationship coefficients ($\boldsymbol{G_{ij}}$), where $\boldsymbol{i}$ and $\boldsymbol{j}$ denote any two individuals in the population. Under the tetraploid configuration in *AGHmatrix* (using ploidy = 4), which scales the expected diagonal values approximately 1.0, the genomic relationship coefficients for each progeny class behave predictably. The diagonal elements ($\boldsymbol{G_{ii}}$) represent self-kinship or individual homozygosity relative to the base population. Values near 1.0 indicate heterozygosity and homozygosity levels consistent with the population average. Conversely, values substantially exceeding 1.0 (e.g., 1.2 to 1.5) reflect inbreeding, such as selfing, clonal reproduction via apomixis, or mating between close relatives, which elevates the proportion of fixed homozygous loci (allelic dosages of 0 or 4) beyond population expectations.

Consequently, the final 259×259 $\boldsymbol{G}$ matrix was used to classify each individual into one of six categories on the basis of their pairwise relationship coefficients with the declared parental genotypes: confirmed hybrids, apomictic clones, *U. brizantha* pollen contaminants, unconfirmed maternity, under investigation, or unknown parentage. The population structure was further explored and visualized via principal component analysis (PCA) computed from the $\boldsymbol{G}$ matrix. To display genomic relationships and parentage classifications across all individuals simultaneously, a dual-triangle heatmap of the $\boldsymbol{G}$ matrix was constructed. In this heatmap, individuals were organized by cross-family, with parental genotypes repeated within each family block to facilitate direct visual comparison of parent–offspring genomic configurations. The lower triangle displays the continuous pairwise genomic relationship values ($\boldsymbol{G_{ij}}$), whereas the upper triangle depicts the categorical parentage classifications assigned to each individual. Families were ordered by female parents, and offspring within each family were sorted by their classification status in a standardized order: confirmed hybrids, under investigation, apomictic clones, unconfirmed maternity, *U. brizantha* pollen contaminants, and unknown parentage. Categorical color bars identifying the predominant classification of each family were aligned along both the top and left margins of the heatmap. The heatmap was generated via the *ggplot2* package (Wickham et al. 2016), which uses a diverging color scale ranging from dark blue (representing $\boldsymbol{G_{ij} = -1}$ or low relationship) through white (representing $\boldsymbol{G_{ij} = 0}$) to dark red (representing $\boldsymbol{G_{ij} \geq 2}$), with intermediate breakpoints at $\boldsymbol{G_{ij} = 0.5, 1.0, and\ 1.5}$ to optimize visual contrast across the biologically relevant range of relationship values. Finally, genealogical network diagrams were

constructed using both declared and genomically inferred parentage information, explicitly incorporating the discovery of H830–9 as an undeclared F_1 female parent.

Genome-wide association study for *Notozulia entreriana* antibiosis - *GWASpoly*

Genome-wide association analyses were performed via the *GWASpoly* package (Rosyara et al. 2016), which implements mixed linear models specifically designed to explicitly model allelic dosage effects (0-4 copies). For each SNP, five genetic models were tested: additive, 1-dom-alt, 1-dom-ref, 2-dom-alt, and 2-dom-ref, capturing a range of additive and dominance effects characteristic of tetraploid inheritance. The $\boldsymbol{G}$ matrix estimated was incorporated as a kinship matrix ($\boldsymbol{K}$) to control for population structure and genetic relatedness among individuals, following the P3D (population parameters previously determined) approach for computational efficiency. The use of identity-based BLUPs rather than genomic BLUPs (GBLUPs) as phenotypic inputs for GWAS was motivated by the need to avoid double-counting genomic information, which would otherwise inflate test statistics under mixed model association frameworks incorporating a genomic relationship matrix as a covariate.

To assess the robustness of marker–trait associations, three analytical scenarios were investigated: (i) all available hybrids, excluding extreme phenotypic outliers but retaining unphenotyped individuals to maintain the full structure of the genomic relationship matrix ($n = 243$); (ii) hybrids excluding redundant apomictic individuals ($n = 190$); and (iii) genetically confirmed hybrids analyzed via refined BLUPs ($n = 56$). Genome-wide significance thresholds were established via the Bonferroni correction ($\boldsymbol{\alpha = 0.05/m}$, where $\boldsymbol{m}$ is the number of markers tested and a suggestive threshold is defined as $\boldsymbol{1/m}$.

Genomic inflation factors (λ) were calculated for each scenario, trait, and model to assess residual confounding. The proportion of phenotypic variance explained by each candidate SNP ($\boldsymbol{R^2}$) was estimated as follows:

$$\boldsymbol{R^2 = \frac{\hat{a}^2 \cdot 2p(1-p)}{V_P}} \qquad (5)$$

where $\hat{\boldsymbol{a}}$ is the estimated allelic effect, $\boldsymbol{p}$ is the alternative allele frequency, and $\boldsymbol{V_P}$ is the total phenotypic variance. The proportion of genetic variance explained ($\boldsymbol{R^2/V_g}$) was further obtained by dividing $\boldsymbol{R^2}$ by the broad-sense heritability estimate.

BLUP was estimated via a random intercept model without explicit inclusion of replicate effects, as the two replicates were unbalanced across all genotypes and the correlation between BLUP estimated with and without replicates as a fixed effect exceeded 0.999 for both traits, confirming negligible confounding between replicate and genotypic effects in this

design. Manhattan and quantile–quantile (Q–Q) plots, generated via the package *GWASpoly*, were generated for all GWAS scenarios and models.

Genome-wide association study for *Notozulia entreriana* antibiosis - diploid dosage

As a complementary approach, GWAS was also performed via the *GAPIT* package version 4.1.0 (Wang and Zhang 2021), which implements the *FarmCPU* (Fixed and Random Model Circulating Probability Unification; Liu et al. 2016) and *BLINK* (Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway; Huang et al. 2019) algorithms. Since *GAPIT* does not natively model allelic dosage in polyploids, tetraploid genotypic data (dosage scale 0–4) were diploidized by dividing each dosage value by two and rounding to the nearest integer (resulting scale 0–1–2) prior to analysis. The population structure was controlled via the first three principal components derived from the $\boldsymbol{G}$, combined with the VanRaden kinship matrix as a covariate. Analyses were conducted separately for each trait via a scenario with a single apomictic individual per family **(**$n = 198$**)** to enable direct comparison with *GWASpoly* results under equivalent conditions. Method convergence was evaluated by comparing p_values and effect directions of the candidate SNPs identified by *GWASpoly*. All phenotypic and genomic analyses were performed via R version 4.5.3 (R Core Team 2019).

Analysis between PCR markers and the genomic relationship matrix

Cross-validation between the PCR-based mode of reproduction classification and genomic parentage verification was performed by comparing the binary PCR phenotype (apomictic/sexual) with the $\boldsymbol{G}$ matrix-derived classification (apomictic clone, confirmed hybrid, or to investigate) for each individual. Concordance was assessed on the basis of pairwise genomic relationship coefficients, observed heterozygosity, and genomic relationships with the male-parent IPY. The predicted cytological phenotypes, including aposporic-only embryo sacs, polygonum-type sexual embryo sacs, and facultative profiles presenting both sac types simultaneously, were assigned to each individual on the basis of the combination of PCR results, $\boldsymbol{G}$ matrix classification, and allelic dosage at candidate loci. Cytological predictions have not been confirmed experimentally and represent only genomic predictions.

GWAS for mode of reproduction

A GWAS for mode of reproduction was performed using the binary PCR phenotype (0 = sexual, 1 = apomictic) as the response variable ($n = 43$; comprising 42 hybrids and the apomictic male-parent IPY). Prior to analysis, SNPs were filtered for minor allele frequency ($MAF > 0.10$) and missing data rate (< 20%), resulting in 29,910 polymorphic markers. The genomic relationship matrix $\boldsymbol{G}$, estimated as described above, was included to control for population structure within the CRZ–22 family.

A GWAS was conducted via *FarmCPU* and *BLINK* algorithms implemented in *GAPIT3* (Liu et al. 2016; Huang et al. 2019) under an additive model. The first three principal components derived from the $\boldsymbol{G}$ matrix were included as covariates. Genome-wide significance was assessed via a Bonferroni-corrected threshold of $\alpha = 0.05/29{,}910$ ($p \leq 1.67 \times 10 - 6$) and a suggestive threshold of $1/29{,}910$ ($p \leq 3.34 \times 10 - 5$).

Because *GWASpoly* (Rosyara et al. 2016) requires stable variance component estimation that was not achievable with ($n = 43$), dominance effects were evaluated post hoc for the seven Bonferroni-significant candidate SNPs identified by *GAPIT* via logistic regression (GLM, binomial family) implemented in R. Three dominance contrasts were tested for each candidate: additive (linear dosage effect), 2-dom-alt ($\text{dosage} \geq 3 \text{ vs.} \leq 2$), and 2-dom-ref ($\text{dosage} \leq 3 \text{ vs.} = 4$). Model fit was compared via the Akaike information criterion (AIC), and the best-fitting model was selected for each candidate. To assess the relative gain in statistical power from dominance modeling, *p*_values under the additive and 2-dom-ref models were compared for all seven candidates. Pairwise correlations among candidate alleles were calculated to identify redundant loci. The predicted probabilities of apomixis for each individual were obtained from the best–fitting GLM via the $predict(\quad)$ function with $type = \text{“}response\text{”}$.

Results

Phenotypic variation and heritability

Significant variation in both resistance traits was observed across the hybrid population (**Fig. 1**). Initial broad-sense heritability estimates from the full hybrid dataset using an identity covariance structure yielded H^2 = 0.08 for both damage score and nymph survival, values substantially lower than expected given the observed phenotypic variation (**Supplemental Table 1**). Broad-sense Cullis heritability estimates under the identity covariance model (*lme4*) were $H^2_Cullis = 0.570$ for the damage score and $H^2_Cullis = 0.702$ for nymph survival, reflecting moderate-to-high precision of genetic value estimation given the replicated experimental design. Under the genomic BLUP framework incorporating the $\boldsymbol{G}$ matrix as the covariance kernel (GBLUP, *sommer*), Cullis' heritability estimates increased substantially to $H^2_Cullis = 0.793$ for damage and $H^2_Cullis = 0.885$ for survival, demonstrating the additional information contributed by marker-based relationship estimation for both traits. The correlation between identity-based BLUPs and GBLUPs was $r = 0.49$ for damage and $r = 0.77$ for survival, indicating that the two frameworks captured substantially different aspects of genetic variation, particularly for damage scores, where the low conventional heritability ($H^2 = 0.083$) implies that the genomic covariance structure provides considerable additional discriminating power beyond phenotypic means alone. Importantly, these estimates were obtained from the full dataset, which included apomictic clones, pollen contaminants, and individuals with uncertain parentage. Therefore, they should not be directly

compared with the heritability estimates obtained from the purified subset of genetically confirmed hybrids described below.

Following restriction of the analysis to genetically confirmed hybrids (n = 56), heritability estimates increased to H^2 = 0.54 for the damage score ($\sigma_g^2 = 0.186$; $\sigma_e^2 = 0.319$) and $H^2 = 0.43$ for nymph survival ($\sigma_g^2 = 252.25$; $\sigma_e^2 = 669.3$), representing a 2.3-fold increase in nymph survival relative to the uncorrected estimate. Complementary heritability estimates obtained via genomic BLUP (GBLUP) incorporating the $\boldsymbol{G}$ matrix (n = 190, excluding redundant apomictic clones, and only for phenotyped and genotyped hybrids and parents) were $H^2 = 0{,}58$ for the damage score and $H^2 = 0{,}51$ for nymph survival, which is consistent with the *lme4* refined estimates (**Supplemental Table 2**).

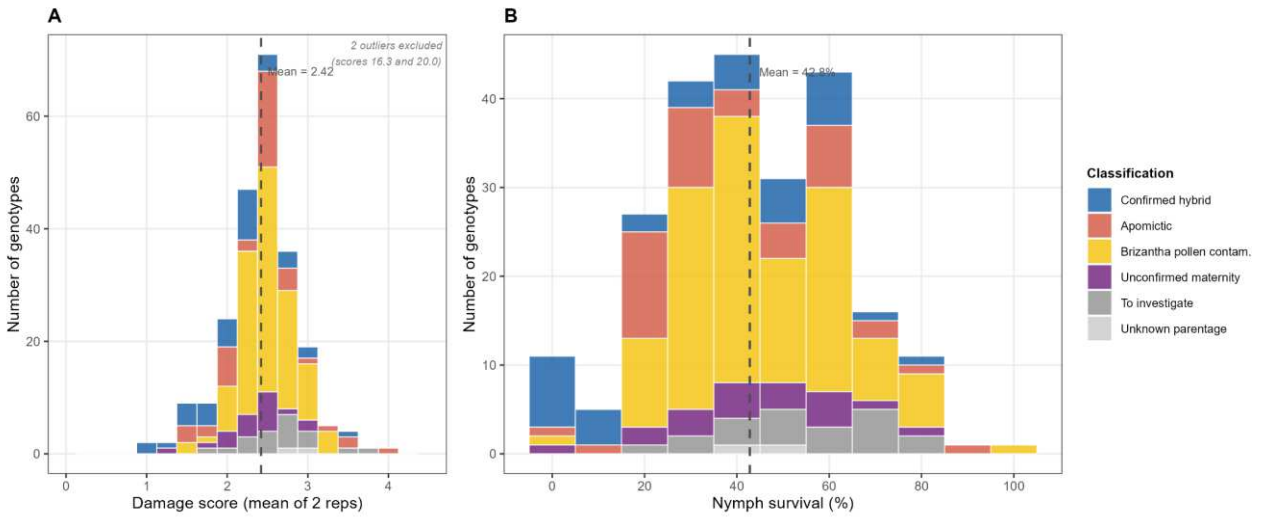


Fig. 1. Distribution of damage scores and nymph survival ratings around the population mean, alongside hybrid classification on the basis of parentage results derived from the genomic relationship matrix ($\boldsymbol{G}$).

Genotyping, quality control, and SNP calling and filtering

Genotyping-by-sequencing of 259 individuals, 243 putative interspecific hybrids and 14 parental genotypes yielded a total of 46,085 biallelic SNPs after initial filtering. Parental genotypes were not included in the GWAS panel, as they were used to establish reference allele frequencies and parentage classification thresholds. The final GWAS panel comprised 243 putative hybrids in the all-hybrid scenario ($n = 243$), 198 hybrids in the one-apomictic-per-family scenario ($n = 198$), and 56 confirmed interspecific hybrids in the confirmed-hybrid scenario ($n = 56$).

A minimum minor allele frequency threshold of $MAF \geq 0.10$ was applied to retain SNPs that were informative across the three parental species. This rationale is consistent with the findings of Krusenbaum and Wissuwa (2025), who demonstrated that strict *MAF* and call-rate filters discard potentially valuable genetic information not linked with common SNPs, leading to reduced detection accuracy for traits associated with rare or dispensable genomic variation.

Following the removal of 148 monomorphic markers, 45,937 polymorphic SNPs were retained for downstream analyses. No SNP exceeded a missing data rate of 20%, and no SNP presented simultaneously high missing data and low minor allele frequency, confirming that missing data were structurally associated with interspecific genomic divergence rather than genotyping error. The parental genotypes differ substantially in their expected genomic composition and missing data rates relative to the diploid *U. ruziziensis* reference genome used for SNP calling. Although the three species involved in the crosses belong to the same agamic complex (Ferreira et al. 2021), *U. decumbens* (BAS) presented the lowest missing data rate among the parental genotypes (4.9%), reflecting its greater genomic proximity to *U. ruziziensis* relative to *U. brizantha*. In contrast, MAR and B4–2, which are pure *U. brizantha* genotypes, presented high missing data rates of 27.4% and 52.7%, respectively, which is consistent with their greater divergence from the reference genome. IPY and CAY, which are three-species interspecific hybrids (*U. brizantha* × *U. decumbens* × *U. ruziziensis*), presented low-to-moderate missing rates of 6.4% and 2.6%, respectively, reflecting their intermediate genomic composition. The female parents H99–9, H100–9, and H317–9 were selected from open pollination and likely carry a relatively high proportion of the *U. decumbens* or *U. ruziziensis* genomic background, resulting in very low missing rates (< 0.1%). In contrast, H176–9, selected from open pollination with a probable higher proportion of *U. brizantha* background, presented a missing rate of approximately 30%. These patterns indicate that the SNP panel is inherently biased toward the *U. decumbens* and *U. ruziziensis* genomic regions, which is practically unavoidable when a diploid *U. ruziziensis* reference is used for interspecific studies involving *U. brizantha* and should be acknowledged as a limitation of the present study.

These patterns collectively indicate that SNPs absent in hybrids and parental genotypes with relatively high *U. brizantha* genomic contributions do not represent genotyping failures but rather reflect genomic regions of *U. brizantha* that are not covered by the diploid *U. ruziziensis* reference sequence. Consequently, the SNP panel used in this study is inherently biased toward the *U. decumbens* and *U. ruziziensis* genomic regions, which is practically unavoidable when a single-species reference genome is used for SNP discovery in highly heterozygous interspecific polyploid populations derived from three divergent parental species. This ascertainment bias should be considered when interpreting association signals, particularly for loci tagged by SNPs with low minor allele frequency, which may disproportionately represent *U. brizantha*-derived variation segregating at low frequency in the panel owing to reference-based calling constraints rather than true biological rarity.

Examination of the 500 SNPs with the highest missing rates revealed a clear interspecific pattern: families from crosses involving female parents with higher *U. brizantha* genomic contributions consistently presented elevated missing rates in this SNP subset. For example, CRZ–4 (H327–9 × BAS) presented 69.4% missing data among the top 500 SNPs, CRZ–6 (H896–9 × BAS) 46.4%, and CRZ–18 (H176–9 × IPY) 43.3%, all of which are consistent with the high *U.*

brizantha genomic proportion of their respective female parents. Conversely, crosses CRZ–26 and CRZ– 24, despite having MAR (*U. brizantha*) as the male parent, presented near-zero missing rates in this SNP subset, a seemingly paradoxical result.

These patterns lead to an unambiguous conclusion: SNPs with elevated missing rates in this dataset represent genomic regions of *U. brizantha* that are absent from the diploid *U. ruziziensis* reference sequence. These markers are not uninformative; they are biologically relevant indicators of interspecific genomic variation, and their removal through aggressive missing data filters would discard valuable information specific to the *U. brizantha* parental contribution. Accordingly, the QC filtering strategy adopted in this study applied no missing data threshold per individual, given the structured interspecific nature of missing data, and no missing data threshold per SNP, as no marker exceeded the 20% threshold. The only filters applied were removal of monomorphic SNPs ($MAF = 0$) and markers with invalid dosage calls (values outside the 0-4 range), resulting in a final panel of 45,937 polymorphic SNPs that retained the full spectrum of interspecific genomic variation present in this three-species hybrid population.

The absence of SNPs with simultaneously high missing rates and low *MAFs* further validated this filtering strategy and had direct practical implications for downstream analyses. A more stringent *MAF* threshold (e.g., > 0.10) was deliberately not applied to preserve species-specific rare alleles potentially underlying resistance and mode of reproduction traits. In interspecific populations derived from three parental species with divergent allele frequency spectra, alleles present at low frequency in the overall panel may represent fixed or near-fixed variants in one parental species—precisely, the alleles most likely to underlie interspecific trait differences—and ones that are well-genotyped in the present dataset as demonstrated by the absence of confounding between missing data and allelic rarity. These low *MAF* values are consistent with resistance alleles derived from *U. brizantha* segregating at low frequency within a predominantly *U. ruziziensis* genetic background, a pattern expected in interspecific populations where species-specific adaptive alleles are rare at the panel level but functionally critical and one that is compounded by the reference-based ascertainment bias described above. The retention of low-frequency alleles ($MAF \geq 0.05$) has proven consequential for candidate locus detection for damage (Dam) and nymph survival (Sur), and this concern is particularly acute in interspecific breeding populations involving multiple parental species, as recently demonstrated for low-call-rate SNPs and presence-absence variation in rice genomic prediction (Krusenbaum and Wissuwa 2025). Future GWAS efforts in *Urochloa* interspecific populations should therefore carefully consider the trade-off between statistical power, which increases with stricter *MAF* filters, and biological completeness, which necessitates the retention of rare, species-specific variants.

Genomic relationship matrix and parentage verification

The diagonal elements of the $\boldsymbol{G}$ matrix (**Supplemental Table 3**) for hybrids, which reflect individual heterozygosity, revealed a bimodal distribution consistent with the mixed reproductive structure of the population. The first quartile and median of the diagonal values were 0.30 and 0.37, respectively, corresponding to interspecific hybrids with high heterozygosity, whereas the third quartile and maximum reached 1.77 and 3.23, respectively, corresponding to apomictic clones and parental genotypes with higher homozygosity relative to the allele frequency base. The off-diagonal elements presented a median of -0.026 and a mean of -0.004, indicating that the majority of pairwise relationships were close to zero, which is consistent with a structured population in which individuals from different crosses share little genomic background. The maximum off-diagonal value of 1.73 identified at least one pair with a very high genomic relationship, indicative of maternal clones or apomictic progeny, whereas the minimum value of -0.855 reflected strong interspecific divergence between certain pairs.

Three distinct patterns of high genomic relationships ($\boldsymbol{G} > 0.8$) were identified. The first and most prevalent pattern involved full-sib pairs within the same cross family, where pairwise $\boldsymbol{G}$ values of 1.5–1.7 were observed among hybrids sharing both parents, for example, H351–74, H365–74, and H368–74 (all from H581–9 × CAY); the K family (all from R81 × IPY); and H340–74, H343–74, and H348–74 (all from H327–9 × CAY). These elevated values are expected in tetraploid full sibs under the VanRaden $\boldsymbol{G}$ matrix scaled by allele frequency variance. The second pattern, indicative of apomixis, involved high $\boldsymbol{G}$ values between a maternal parent and its declared offspring: H327–9 female parents of H348–74 ($\boldsymbol{G} = 1.491$) and H581–9, with several of its putative progeny, presented parent–offspring relationships comparable in magnitude to full-sib relationships, which is consistent with maternal sexual reproduction. The third pattern involved two hybrids from different maternal parents, H274–74 (H176–9 × CAY) and H368–74 (H581–9 × CAY), presenting an unexpectedly high $\boldsymbol{G}$ value of 1.579 despite having different declared mothers, suggesting either that H176–9 and H581–9 are genomically very similar, having been selected from the same open-pollinated field population, or that one of the declared maternities is incorrect.

Conversely, the most negative pairwise $\boldsymbol{G}$ values ($\boldsymbol{G} < -0.6$) identified the poles of maximum interspecific divergence within the panel. The female parent H176–9 and the male parent B4–2 represented the *U. brizantha* extreme, with $\boldsymbol{G}$ values as low as -0.855 and -0.750, respectively, when paired with members of CRZ–22, which represented the *U. ruziziensis*/*U. decumbens* extreme. This pattern indicates that the $\boldsymbol{G}$ matrix successfully captures the full spectrum of interspecific genomic divergence present in the panel. A notable case of divergence between hybrids was also observed: H351–74 (CRZ–12, high *U. brizantha* contribution) and H471–74 (CRZ–27, distinct parental combination) presented $\boldsymbol{G} = -0.653$, approaching interspecific-level divergence between two putative hybrid individuals, an unusual finding that reflects the breadth of genomic variation generated by the three-species crossing design.

Parentage diagnosis on the basis of pairwise $\boldsymbol{G}$ values between putative hybrids and their declared parental genotypes revealed two distinct patterns of maternal relationships. The first group comprised crosses in which the median $\boldsymbol{G}$ between offspring and the declared sexual mother exceeded 1.0, strongly indicative of maternal facultative apomixis: CRZ–11 (H327–9, median $\boldsymbol{G}$_sexual $= 1.38$), CRZ–12 and CRZ–5 (H581–9, median $\boldsymbol{G}$_sexual $= 1.32$), and CRZ–25 (H327–9, median $\boldsymbol{G}$_sexual $= 1.27$). The female parents H581–9 and H327–9 were therefore identified as the principal facultative apomictic candidates among the maternal genotypes, appearing as mothers in multiple crosses with systematically elevated $\boldsymbol{G}$_sexual parent values. CRZ–20 (H581–9 × IPY) presented an intermediate pattern, with a median $\boldsymbol{G}$_sexual of 0.667 but a maximum of 1.37, indicating a mixture of apomictic clones and true sexual hybrids within the same family. The second group comprised crosses with $\boldsymbol{G}$_sexual values of approximately 0.19–0.235, the expected range for a sexual mother–offspring relationship in tetraploids under the VanRaden scaling-confirming sexual maternity law in CRZ–8, CRZ–9, CRZ–16, CRZ–17, CRZ–18, and CRZ–3, among others. Two individual cases of negative $\boldsymbol{G}$_sexual values were identified, H130–74 (CRZ–18) and H274–74 (CRZ–9), indicating incorrectly declared maternity for these genotypes, as negative $\boldsymbol{G}$ values are inconsistent with any degree of genomic relatedness between mothers and offspring.

Paternity diagnosis revealed an equally informative and biologically coherent pattern. Confirmed paternity, indicated by positive and moderate G_apomictic values, was observed in CRZ–22 (R81 × IPY, $\boldsymbol{G}$_$a$pomictic = 0.20 to 0.87), CRZ–15 (H896–9 × CAY, $\boldsymbol{G}$_$a$pomictic = 0.29 to 1.0), and CRZ–20 (H581–9 × IPY, $\boldsymbol{G}$_$a$pomictic= 0.07 to 0.65). In contrast, systematically negative $\boldsymbol{G}$_$a$pomictic values were detected in crosses CRZ–13, CRZ–16, CRZ–8, and CRZ–10 (declared apomictic CAY, median $\boldsymbol{G}$_$a$pomictic= -0.235 to -0.238), CRZ–3 and CRZ–7 (declared apomictic BAS, median $\boldsymbol{G}$_$a$pomictic= -0.150 to -0.151), and CRZ–1 (declared apomictic B4–2, median $\boldsymbol{G}$_$a$pomictic= -0.384). These negative values are biologically interpretable: CAY, BAS, and B4–2 are apomictic genotypes, and apomictic plants produce pollen that is clonally derived from the maternal genome rather than from the declared apomictic parent genotype itself. Consequently, the absence of detectable genomic contributions from these declared apomictic parents in their putative offspring is consistent with either cross-contamination by pollen from an alternative genotype or with the inherent genomic divergence between apomictic pollen donors and the reference allele frequencies used to construct the $\boldsymbol{G}$ matrix.

The most striking individual case of parentage was H471–74 (declared across CRZ–27: H896–9 × MAR), which presented a very low $\boldsymbol{G}$_sexual of 0.10, inconsistent with true maternity, along with a very high $\boldsymbol{G}$_apomictic of 0.875, confirming the paternal contribution of MAR while rejecting the declared maternal parent H896–9. This pattern strongly suggests an error in the declared maternal identity rather than in the paternal assignment, and H471–74 was reclassified accordingly.

The high proportion of *U. brizantha* pollen contaminants was confirmed by consistently negative genomic relationship coefficients with the declared male parents ($\boldsymbol{G}$ = -0.21 to -0.24) and positive coefficients with the MAR ($\boldsymbol{G}$ mean = 0.121), which is fully consistent with paternal half-sib relationships. Pairwise genomic relationships within apomictic families revealed mean $\boldsymbol{G}$ values of 1.36, 1.40, and 1.41 for families derived from H581–9, H327–9, and R81, respectively, confirming their clonal origin, whereas relationships between apomictic families were substantially lower ($\boldsymbol{G}$ mean = 0.629; range: -0.191-1.730), indicating that apomictic families are genomically distinct from one another **(Supplemental Table 4)**.

The results were visualized via principal component analysis (PCA) and a genomic relationship heatmap. Principal component analysis of the genomic relationship matrix ($\boldsymbol{G}$) explained 74.7% and 20.05% of the total genomic variation in PC1 and PC2, respectively, consistently separating individuals according to their reproductive origin and species background (**Fig. 2**). Parentage verification on the basis of pairwise genomic relationship coefficients allowed the classification of all 243 putative hybrids into four categories (**Supplemental Table 4**): confirmed hybrids ($n = 56$; 23.0%); apomictic clones of the maternal parent ($n = 47$; 19.3%); *U. brizantha* pollen contaminants ($n =$118; 48.5%); genotypes with unconfirmed maternity; and genotypes under investigation or of unknown parentage ($n = 22$; 9.0%).

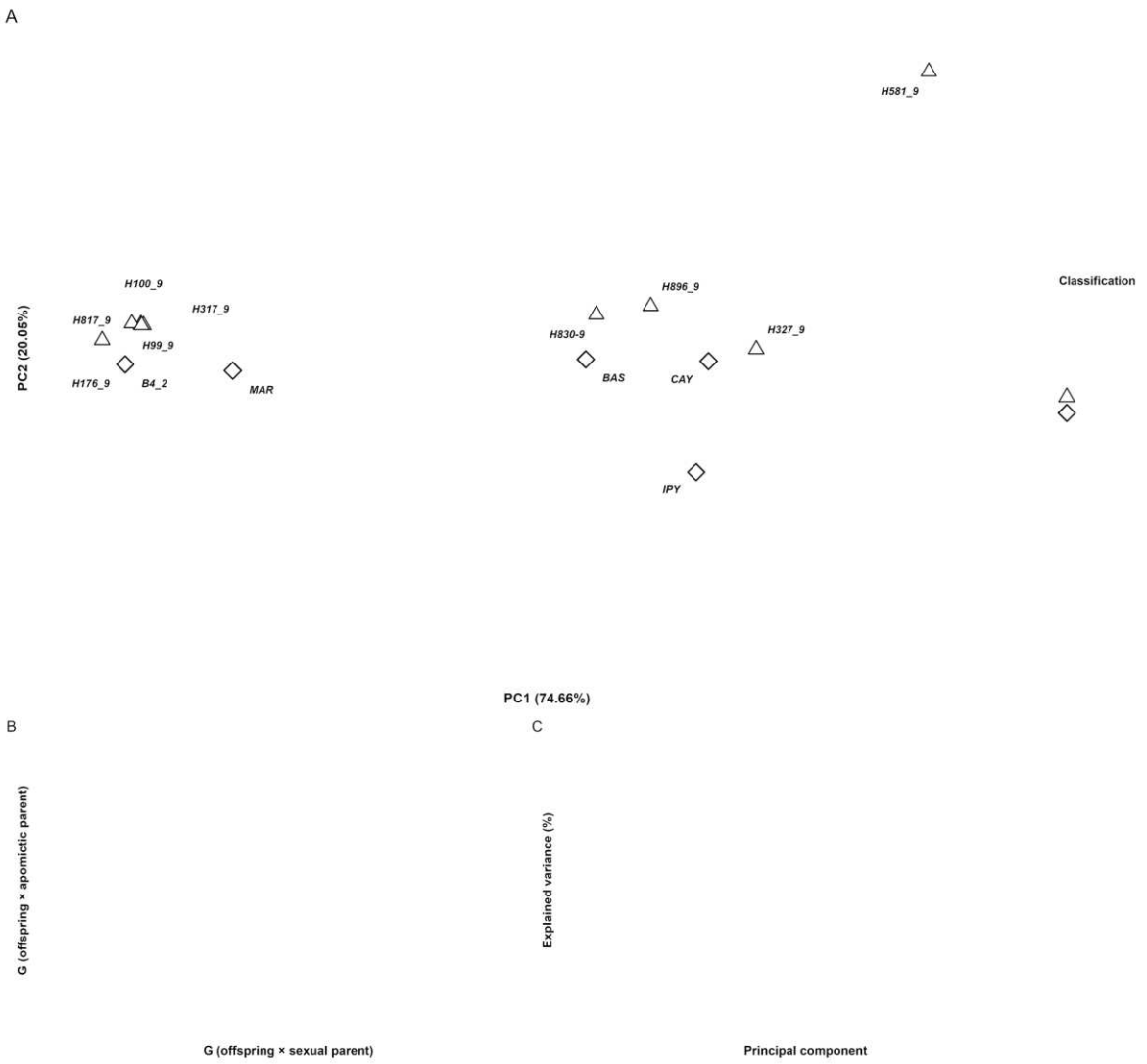


Fig. 2 Principal component analysis (PCA) and population structure based on the genomic relationship coefficients of 243 interspecific *Urochloa* hybrids and 14 parental genotypes with 45,937 GBS-derived SNPs. (A) PCA of the first two principal components, colored by parentage population structure. The female parent R81 failed GBS quality control and was excluded, and the male parents (MAR, B4–2, BAS, IPY, and CAY) are indicated by distinct symbols. (B) Pairwise genomic relationship coefficients between each hybrid and its declared female parent ($\boldsymbol{G}$ offspring × sexual parent) versus its declared male parent ($\boldsymbol{G}$ offspring × apomictic parent). The vertical dotted lines indicate the maternity threshold ($\boldsymbol{G} = 0.15$) and the apomictic clone threshold ($\boldsymbol{G} = 1.0$). The five major groups, confirmed hybrids, apomictic clones, *U. brizantha* pollen contaminants, unconfirmed maternity, and hybrids to investigate, are resolved by their position in the bivariate space. (C) Percentage of total genomic variance explained by each principal component.

A notable finding emerged from the parentage analysis of the female parent H830-9, which presented a genomic relationship coefficient of 0.55 with the female parent H327–9, which was consistent with a mother–offspring relationship in segmental allopolyploids (expected $\boldsymbol{G} \approx 0.50$), and 0.25 with the male parent IPY. These results indicate that H830-9

is an undeclared F_1 hybrid of H327–9 × IPY that is inadvertently used as a founding female parent. Its single descendant, H421–74, represents an unintended F_2 interspecific hybrid of H830-9 × B4–2 ($\boldsymbol{G} = 0.314$), rather than the declared H830–9 × CAY cross ($\boldsymbol{G} = 0.027$), constituting an unplanned three-generation pedigree structure detected exclusively through genomic parentage verification (**Fig. 3**).

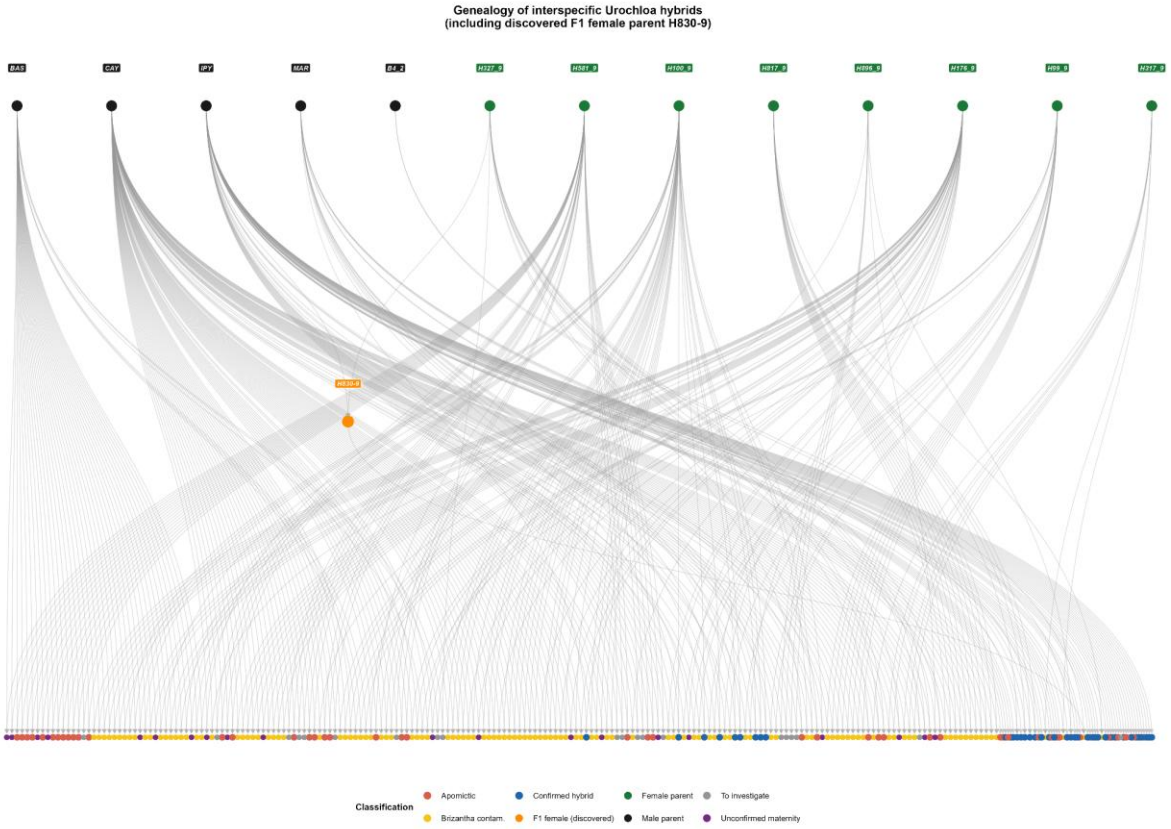


Fig. 3 Genealogy of interspecific *Urochloa* hybrids inferred from the genomic relationship $\boldsymbol{G}$ matrix. The figure illustrates the correction of the pedigree of the female parent H830–9, which was genomically identified as an undeclared F_1 hybrid resulting from the cross H327–9 × IPY ($\boldsymbol{G} = 0.546$), rather than a founding genotype of independent origin as previously assumed. Its single descendant in the present study, H421–74, represents an unintended F_2 interspecific hybrid derived from the cross H830–9 × B4–2 ($\boldsymbol{G} = 0.314$), constituting an unplanned three-generation pedigree structure detected exclusively through genomic parentage verification.

The genomic relationship heatmap (**Fig. 4**) organized by cross family revealed distinct structural patterns consistent with the reproductive and genomic composition of the evaluated population. In the lower triangle, which displays continuous pairwise $\boldsymbol{G}$ values, the CRZ–22 family (R81 × IPY) formed an intense red block in the upper-left corner of the matrix, reflecting the very high genomic relationship values ($\boldsymbol{G} = 1.3 - 1.7$) among members of this family, which

was consistent with the predominance of apomictic clones of the maternal parent R81. Families derived from *U. ruziziensis*-based crosses presented moderate blue tones, reflecting the expected genomic relationship range among full-sibs in tetraploid interspecific populations. In contrast, families with high *U. brizantha* genomic contributions presented light blue to white tones compared with families from other parental backgrounds, indicating low pairwise genomic relationships across interspecific family groups, a pattern consistent with the substantial genomic divergence between *U. brizantha* and *U. ruziziensis*/*U. decumbens* documented in the $\boldsymbol{G}$ matrix diagnostics described above.

In the upper triangle, which displays the parentage classification of each individual, the yellow color corresponding to *U. brizantha* pollen contaminants dominated the matrix, accounting for approximately 48% of all evaluated hybrids and forming the most extensive classification block. The colored stripes corresponding to male (black) and female (green) parental genotypes were clearly visible along the margins of each family block, confirming their correct positioning within the matrix structure. Apomictic clones (red) were visibly clustered within their respective family blocks, most prominently in CRZ–22 and in families derived from H581–9 and H327–9, the two female parents identified as principal apomictic candidates. The dual-triangle structure of the heatmap thus provides simultaneous visualization of genomic relationship magnitude and parentage classification, enabling direct identification of the four major population groups: confirmed hybrids, apomictic clones, pollen contaminants, and unconfirmed parentage genotypes, as well as their distributions across the 27 evaluated cross families.

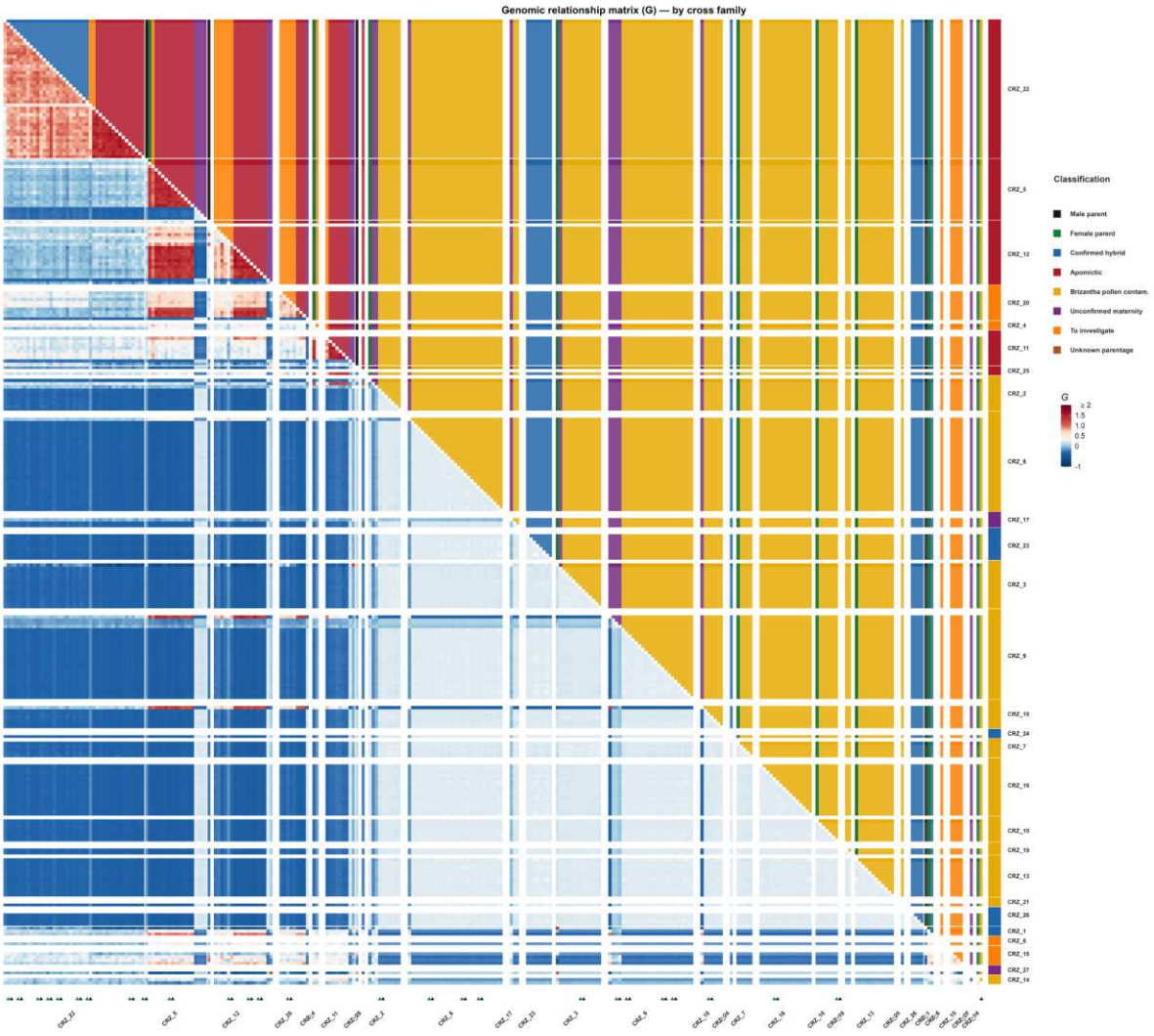


Fig. 4 Dual-triangle heatmap of the genomic relationship matrix ($\boldsymbol{G}$) organized by a cross family. Individuals are grouped by cross, with parental genotypes repeated within each family block to facilitate direct visual comparison of parent–offspring and full-sib genomic relationships. The upper triangle displays the parentage classification assigned to each individual via $\boldsymbol{G}$ matrix analysis: confirmed hybrid (blue), apomictic clone (red), *U. brizantha* pollen contaminant (yellow), unconfirmed maternity (purple), under investigation (orange), and unknown parentage (grey). The lower triangle displays continuous pairwise $\boldsymbol{G}\sim\boldsymbol{ij}\sim$ values on a diverging color scale from dark blue ($\boldsymbol{G}\sim\boldsymbol{ij}\sim = \boldsymbol{-1}$) through white ($\boldsymbol{G}\sim\boldsymbol{ij}\sim = \boldsymbol{0}$) to dark red ($\boldsymbol{G}\sim\boldsymbol{ij}\sim \geq \boldsymbol{2}$). The color bars along the right and bottom margins identify the cross family (CRZ) of each individual block. Triangular markers along the bottom axis indicate the position of parental genotypes within the ordered sequence. Families are ordered by maternal parents, and individuals within each family are sorted by classification status.

Genome-wide association analysis for damage and nymph survival

Using GBLUP-predicted values as GWAS phenotypes resulted in severely inflated genomic inflation factors for dominance models ($\boldsymbol{\lambda} = 2.04\ to\ 2.38$), indicating confounding between GBLUP-adjusted phenotypes and the kinship matrix. Therefore, BLUP genotypic effects predicted under an identity covariance structure (*lme4*) were retained for all GWAS analyses, yielding well-calibrated inflation factors across all scenarios and models (**Supplemental Table 5**).

GWAS for the damage score and nymph survival of *Notozulia entreriana* was conducted under three analytical scenarios differing in population composition: all hybrids ($n = 243$), all hybrids retaining one apomictic per family ($n = 198$), and confirmed interspecific hybrids only ($n = 56$). The genomic inflation factors ($\boldsymbol{\lambda}$) ranged from 0.875 to 1.050 across all scenario × trait × model combinations, indicating adequate control of population stratification by the G matrix kinship covariate. The additive model $\boldsymbol{\lambda}$ values were 0.987 and 1.050 for $n = 243$ (damage and survival, respectively), 1.000 and 1.005 for $n = 198$, and 0.875 and 0.894 for $n = 56$, with slight deflation in the smallest scenario expected given the reduced sample size.

Two candidate loci for damage (Dam) scores were consistently detected across the scenarios. The first candidate, qDam-dom1, located on scaffold 6 (*MAF* = 0.099), reached Bonferroni-corrected significance in the $n = 243$ scenario under the 1-dom-alt model ($score = 5.680$), with suggestive signals in $n = 198$ ($score = 4.493$) and $n = 56$ ($score = 4.432$) under the same model (**Supplemental Table 5**). The positive effect estimate at qDam-dom1 under the 1-dom-alt model suggests that a single copy of the alternative allele is sufficient to confer increased damage susceptibility, which is consistent with a dominant detrimental effect. The second candidate, qDam-dom2, located on scaffold 7 (*MAF* = 0.232), presented the strongest signal in the $n = 198$ scenario under the 1-dom-ref model (*score* = 4.987), with a comparable signal in $n = 56$ ($score = 4.966$) and a weaker signal in $n = 243$ ($score = 2.508$), which is consistent with signal enhancement following the removal of apomictic clones. This pattern is consistent with a spurious signal inflated by apomictic clones (**Fig. 5**), further demonstrating the importance of population purification prior to GWAS in breeding populations with apomictic and sexual modes of reproduction. The negative effect estimated at qDam-dom2 ($\beta = -0.475$) the reference allele confers a protective effect under a dominant model. A third candidate, qDam-add1, which is also located on scaffold 6 ($MAF = 0.085$), presented consistent signals across all three scenarios under the additive model (scores: 2.990, 2.978, and 2.960 for $n = 243$, 198, and 56, respectively), although it did not reach the Bonferroni significance in any scenario individually. The high LD between qDam-dom1 and qDam-add1 ($r = 0.995$) suggests that they tag the same QTL region on scaffold 6, with an estimated allelic substitution effect of -0.475 damage units per reference allele copy, explaining 13.3% of the phenotypic variance and 23.3% of the genetic variance. Together, qDam-dom1 and qDam-dom2 accounted for approximately 48% of the genetic variance in the damage score. The third marker, qDam-add1, which is also on scaffold 6 approximately 636 kb upstream of qDam-dom1, was suggested to have additive

effects in the $n = 198$ ($score = 2.99$) and $n = 56$ ($score = 2.96$) scenarios, explaining 1.0% of the phenotypic variance. Three markers showed consistent evidence of associations with damage (Dam) scores across multiple scenarios and were considered robust candidates (**Supplemental Tables 5 and 6; Fig. 5**; **Supplemental Fig. 1**).

For nymph survival (Sur), a single candidate locus was identified: qSur-add1, located on scaffold 5 ($MAF = 0.053$), which was detectable exclusively in the $n = 56$ confirmed hybrid scenarios ($additive\ score = 4.834$; 1-dom-ref $score = 5.222$), with negligible signals in $n = 243$ ($score = 0.264$) and $n = 198$ ($score = 2.498$). This pattern of exclusive detection in the purified population is consistent with the low *MAF* of this marker, which represents a resistance allele segregating at very low frequency in the overall panel but detectable in the genomically homogeneous confirmed hybrid subset.

Five additional SNPs on scaffold 7, located 74–387 kb from qDam-dom2 and presenting pairwise LD of $r = 0.59 - 0.83$ with this candidate, presented suggestive signals in the $n = 198$ and $n = 56$ damage scenarios and were interpreted as reflecting the same QTL signal as qDam-dom2 rather than independent associations. One SNP on scaffold 6, which reached Bonferroni significance in the $n = 243$ scenario under the 1-dom-alt model, was identified as a false positive inflated by apomictic clones, as its signal disappeared entirely upon population purification, which is consistent with the pattern expected for markers whose allele frequency is distorted by maternal clonal reproduction.

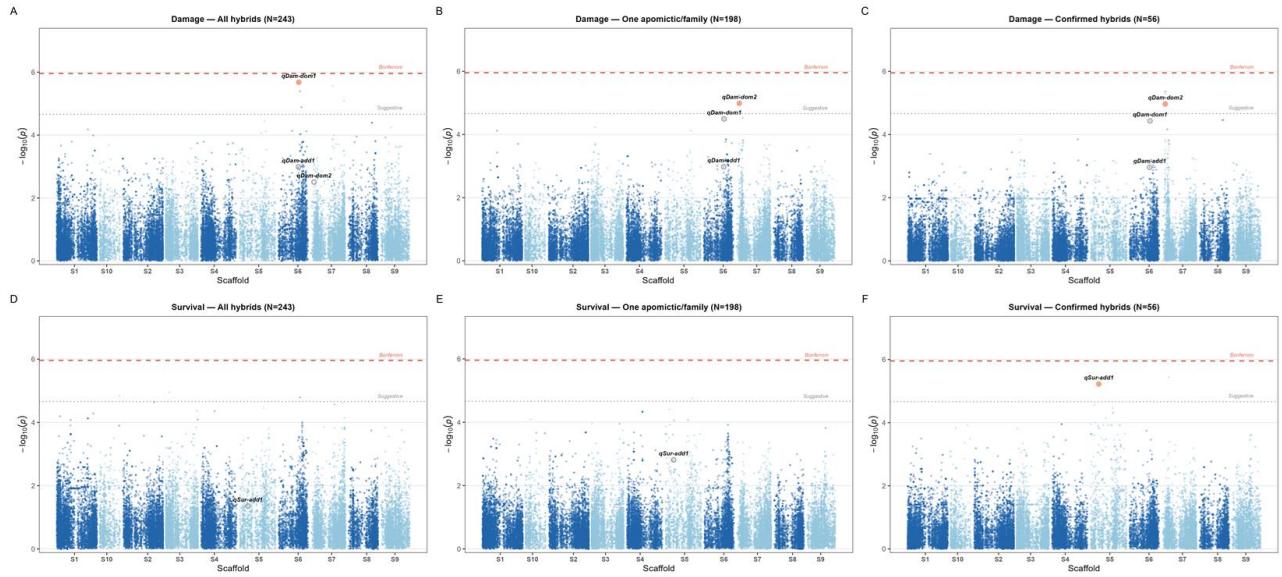


Fig. 5 Manhattan plots of genome-wide association analyses for nymph-induced plant damage (A-C) and nymph survival (D-F) in *Urochloa* spp. hybrids, conducted under three population compositions: all hybrids ($n = 243$; A, D), hybrids excluding apomictic individuals ($n = 198$; B, E), and confirmed hybrids only ($n = 56$; C, F). The $-log{\sim}10{\sim}(p)$ values obtained from GWASpoly under polyploid-adapted dominance and additive models are plotted against the genomic position across ten scaffolds (S1-S10). The red dashed line indicates the Bonferroni-corrected significance threshold

($-log{\sim}10{\sim}(p) \geq 6.0$; $\alpha = 0.05$), and the dotted line indicates the suggestive threshold. Labeled SNPs represent candidate loci identified across multiple population compositions: qDam-dom1 (scaffold S6, dominance model) and qDam-add1 (scaffold S6, additive model) for nymph-induced damage; qDam-dom2 (scaffold S7, dominance model) for damage in the reduced population; and qSur-add1 (scaffold S5, additive model) for nymph survival. The candidate SNPs highlighted in orange exceeded the suggestive threshold in the confirmed hybrid subset, supporting their biological relevance despite the reduced sample size.

As mentioned above, the retention of low-frequency alleles ($MAF \geq 0.05$) is consequential for candidate locus detection for damage and nymph survival: the markers qDam-dom1 ($MAF = 0.099$) and qDam-add1 ($MAF = 0.085$) for spittlebug damage, and qSur-add1 ($MAF = 0.053$) for nymph survival present minor allele frequencies at or near the conventional $MAF \geq 0.05$ threshold. Had a more stringent filter (e.g., $MAF > 0.10$) been applied, qDam-dom1 and qDam-add1 would have been discarded prior to analysis, and qSur-add1, detectable only in the $n = 56$ confirmed hybrid scenarios, would have been lost entirely.

A complementary approach, GWAS was performed via the *FarmCPU* and *BLINK* algorithms in *GAPIT* on diploidized data ($n = 198$) (**Supplemental Fig. 2**). Under this framework, qDam-add1 presented the strongest association signal for damage ($p = 1.85 \times 10^{-4}$), which was convergent with the *GWASpoly* results, whereas qDam-dom1 was nominally significant ($p = 0.016$). In contrast, qDam-dom2 was nonsignificant in *GAPIT* ($p = 0.590$), which is consistent with the loss of dominance information incurred by diploidization of tetraploid allelic dosage data, confirming the superior sensitivity of polyploid-specific models for detecting nonadditive effects (**Supplemental Fig. 2**; **Supplemental Table 6**). The ranking of hybrid genotypes on the basis of genomic breeding values for damage score and nymph survival estimated by GBLUP is presented in **Supplemental Table 7**.

Genomic breeding values and resistance mechanism classification

The scatter plot of phenotypic BLUPs for damage score and nymph survival (**Fig. 6A**) revealed continuous variation across the 204 evaluated genotypes, with a broad distribution across all four resistance quadrants. The scatter plot of GBLUPs estimated under the genomic relationship matrix $\boldsymbol{G}$ (**Fig. 6B**) provided a substantially different ranking of genotypes, particularly for the damage score, where the correlation between phenotypic BLUPs and GBLUPs was $r = 0.49$, indicating that the genomic covariance structure contributed considerable additional discriminating power beyond phenotypic means alone, which was consistent with the low conventional heritability observed for this trait ($H^2 = 0.083$). For nymph survival, the BLUP-GBLUP correlation was greater ($r = 0.77$), reflecting the more heritable nature of this trait ($H^2 = 0.275$). Cullis broad-sense heritability estimates were substantially greater under the GBLUP framework than

under the identity-based model for both traits: *H²_Cullis* increased from 0.570 to 0.793 for the damage score and from 0.702 to 0.885 for nymph survival, confirming the superior precision of genomic value estimation when pedigree information was replaced by marker-derived relationship coefficients.

The GBLUP quadrant analysis (**Fig. 6B**) revealed that 41 genotypes (17.1% of the evaluated population) occupied the antibiosis quadrant (GBLUP_damage < 0, GBLUP_survival < 0), comprising 20 confirmed interspecific hybrids (48.8%), 13 apomictic clones (31.7%), two *U. brizantha* pollen contaminants (4.9%), and six genotypes of other classification (14.6%). In contrast, the tolerance quadrant (GBLUP_damage < 0, GBLUP_survival ≥ 0) was dominated by *U. brizantha* pollen contaminants ($n = 92$, 72.4% tolerance quadrant), reflecting the low damage scores conferred by the *U. brizantha* genomic background in combination with the high nymph survival typical of tolerant genotypes. The susceptibility quadrant (GBLUP_damage ≥ 0, GBLUP_survival ≥ 0) contained 52 genotypes, predominantly *U. brizantha* pollen contaminants ($n = 23$) and apomictic clones ($n = 7$), whereas the antixenosis quadrant (GBLUP_damage ≥ 0, GBLUP_survival < 0) contained 19 genotypes, exclusively apomictic clones ($n = 14$) and other classifications ($n = 5$).

The male parent IPY occupied the antibiosis quadrant with the most extreme GBLUP values among the parental genotypes ($\text{GBLUP_damage} = -0.405$, $\text{GBLUP_survival} = -18.68$), confirming its role as the primary donor of resistance alleles in this population and consistent with its confirmed carrier status for qDam-dom1 and qDam-add1. Analysis of candidate locus dosages among the 41 antibiosis genotypes revealed four triple-carrier genotypes, K15, K6, K25, and H340–74, each simultaneously carrying resistance-associated alleles at qDam-dom1, qDam-dom2, and qDam-add1 (**Supplemental Table 7**), and presented among the most negative GBLUP values for nymph survival in the entire panel (range: -4.76-16.43%). All four were classified as apomictic, carrying the qDam-dom2 reference allele, having inherited qDam-dom1 and qDam-add1 from the IPY male parent, the only male parent confirmed to carry these alleles, suggesting a complementary resistance architecture in which paternal (scaffold 6 and 7 loci from IPY) and maternal (scaffold 7 locus from H327–9 and H581–9) genomic contributions each provide distinct and cumulative resistance components. These triple-carrier genotypes represent priority candidates for use as male parents in subsequent crossing cycles aimed at pyramiding resistance alleles across multiple loci in sexually reproducing interspecific *Urochloa* backgrounds.

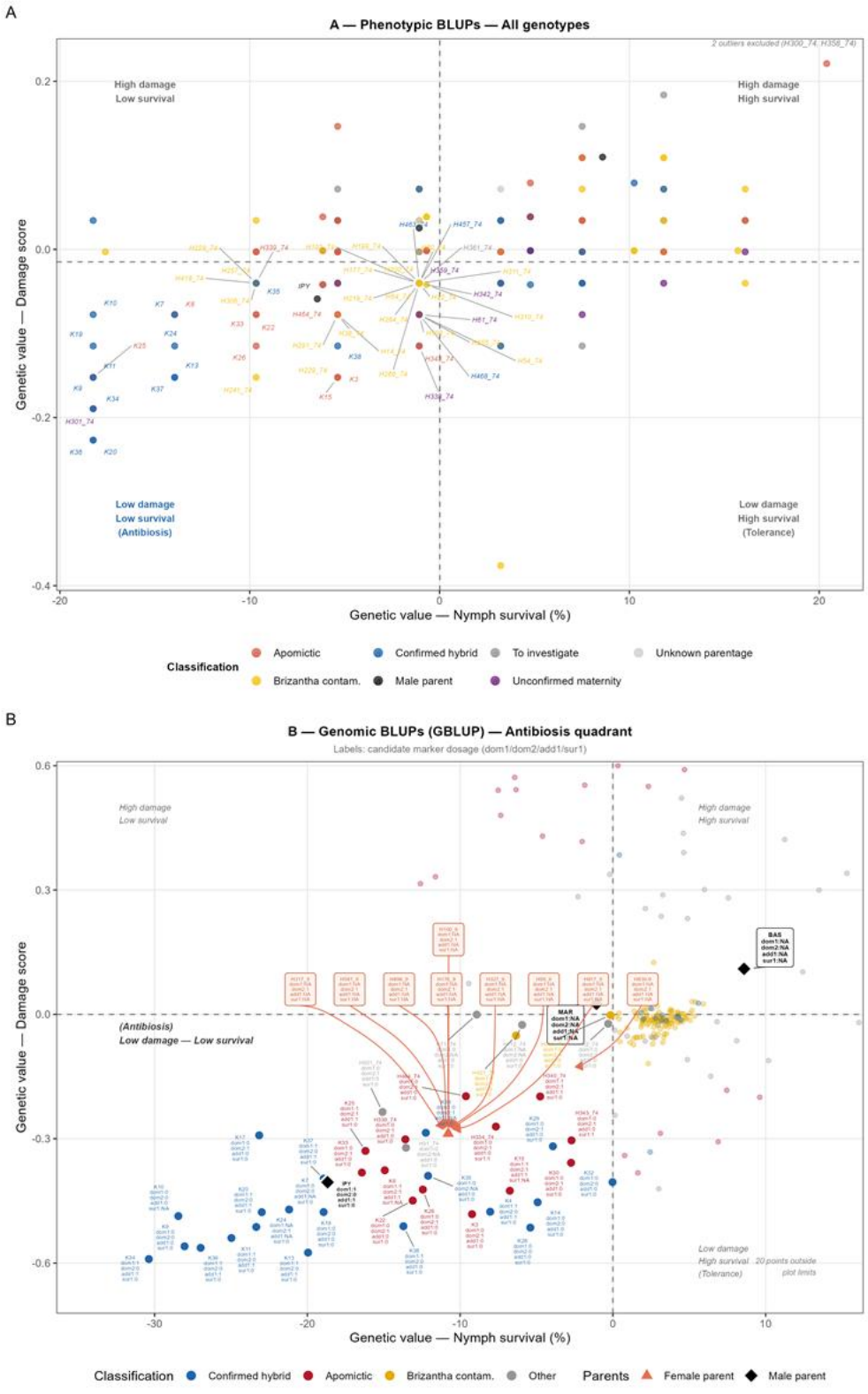


Fig. 6 Scatter plots of genetic values for spittlebug damage score (*y*-axis) and nymph survival percentage (*x*-axis) in interspecific *Urochloa* hybrids and parental genotypes. (A) Phenotypic BLUPs estimated under a mixed model with an identity covariance structure (*lme4*) for all evaluated genotypes ($n = 201$; two extreme outliers, H300–74 and H358–74, were excluded for visualization). The dashed lines indicate the population means for each trait. Individuals in the lower-

left quadrant presented simultaneously low damage scores and low nymph survival, which is consistent with antibiosis resistance. (B) Genomic BLUP (GBLUP) estimation incorporating the genomic relationship matrix G as the covariance kernel (*sommer* package, $n = 201$). Individuals in the antibiosis quadrant (GBLUP_damage < 0, GBLUP_survival < 0) are labeled with their allelic dosage at the four candidate resistance loci: dom1 = qDam-dom1 (scaffold 6, $MAF = 0.099$); dom2 = qDam-dom2 (scaffold 7, $MAF = 0.232$); add1 = qDam-add1 (scaffold 6, $MAF = 0.085$); sur1 = qSur-add1 (scaffold 5, $MAF = 0.053$); NA = missing genotype data at that locus. Male parents (filled diamonds) and female parents (filled triangles) are represented by distinct symbols; female parents are positioned at the mean GBLUP coordinates of dom2, only carrier genotypes, as they were not phenotypically evaluated. The female parent H830-9 is positioned by genomic inference following its reclassification as an undeclared F_1 hybrid. The color coding reflects parentage classification: confirmed interspecific hybrids (blue), apomictic clones (red), *U. brizantha* pollen contaminants (yellow), and others (gray). Thirty-three genotypes with GBLUP values outside the plot limits are not shown. Quadrant annotations indicate putative resistance mechanisms: antibiosis (low damage, low survival), tolerance (low damage, high survival), antixenosis (high damage, low survival), and susceptibility (high damage, high survival).

Apomixis mode of reproduction: phenotypic characterization and marker concordance

The mode of reproduction of 42 hybrids from the R81 × IPY cross (CRZ–22) and the paternal parent IPY was determined via a PCR-based molecular marker with a reported diagnostic accuracy of approximately 85%. Among the 42 hybrids, 25 were classified as apomictic and 17 as sexual by PCR. Cross-validation with genomic relationship coefficients from the $\boldsymbol{G}$ matrix revealed substantial differences between methods (**Supplemental Table 8**).

Examination of the discordant cases provided biological insight into the nature of the disagreement. The 12 individuals classified as sexual by PCR but as apomictic clones by the $\boldsymbol{G}$ matrix presented a mean heterozygosity of 0.507, which was statistically indistinguishable from that of confirmed apomictic clones (0.508) and substantially lower than that of confirmed sexual hybrids (0.613). These results indicate that these individuals are genomically consistent with clonal reproduction and represent likely false negatives of the PCR marker, which was consistent with the known 15% error rate of the assay. Conversely, 22 individuals classified as apomictic by PCR but confirmed hybrids by the $\boldsymbol{G}$ matrix presented a high genomic relationship with IPY ($\boldsymbol{G}$ mean $= 0.789$) and heterozygosity consistent with sexual fertilization (mean $= 0.610$), suggesting that these individuals originated from sexual fertilization events in an otherwise apomictic female parent, a pattern consistent with facultative apomixis in *Urochloa*.

The predicted cytological phenotypes were assigned to each individual on the basis of the combination of PCR results, $\boldsymbol{G}$ matrix classification, and allelic dosage at candidate loci (**Supplemental Table 8**). Three individuals (K6, K18, and K33) were classified as obligate apomicts producing aposporic-only embryonary sacs, presenting apomictic PCR results,

clonal $\boldsymbol{G}$ matrix classification ($\boldsymbol{G} > 1.0$ among pairs), and complete absence of reference alleles at all three candidate loci ($\text{dosage} = 4$). Three individuals (K38, K40, and K43) were classified as high-priority for cytological confirmation because of the discordance between the PCR results (sexual) and marker dosage (presence of a reference allele at all three loci; predicted probability of $\text{apomixis} = 0.88$), suggesting either false-negative PCR results or epigenetic silencing of the apomixis locus. Twenty-two individuals presented profiles consistent with facultative apomixis, combining apomictic PCR results with hybrid genomic composition and the presence of reference alleles, with a predicted probability of $\text{apomixis} = 0.88$.

GWAS for apomictic and sexual modes of reproduction

Genome-wide association analysis for apomixis mode was performed via the *FarmCPU* and *BLINK* algorithms in *GAPIT3* ($n = 43$, including the paternal parent IPY). Owing to the small sample size, variance component estimation in *GWASpoly* was unstable; therefore, dominance effects were evaluated post hoc via logistic regression (GLM) for the seven candidate SNPs identified by *GAPIT3*.

Under the additive model, seven SNPs reached genome-wide Bonferroni significance ($\alpha = 0.05/29{,}910$; **Fig. 7A**; **Supplemental Table 9**). The strongest signal under the additive model was detected at scaffold 7 (qApom-add1; $-\log_{10}\text{p} = 10.5$). However, subsequent dominance testing revealed that qApom-add1 had negligible support under the 2-dom-ref model ($p = 0.636$), and the AIC of the additive-only model for this locus was the highest among the candidates ($\text{AIC} = 59.65$), indicating that its detection reflects a dosage artifact of the additive model rather than a genuine dominance effect. This marker is therefore considered a false positive and is not included among the validated candidates (**Fig. 7B**). The markers qApom-dom1 (scaffold 9, 34.4 Mb) and qApom-dom2 (scaffold 1, 6.7 Mb) presented identical *p*-values of 7.9×10^{-5} under the 2-dom-ref model, a 32-fold improvement over the additive model ($p = 2.5 \times 10^{-3}$). Notably, qApom-dom1 and qApom-dom2 are perfectly correlated ($r = 1.00$), indicating that despite their location on different scaffolds (S9 and S1), they constitute a single informative locus in this population, possibly reflecting incomplete genome assembly of the reference tetraploid *U. ruziziensis*. The marker qApom-dom3 (scaffold 9, 28.0 Mb) showed a similarly pronounced improvement under the dominance model ($p = 1.6 \times 10^{-4}$ vs $p = 4.9 \times 10^{-3}$), located approximately 6.4 Mb from qApom-dom1 on scaffold 9.

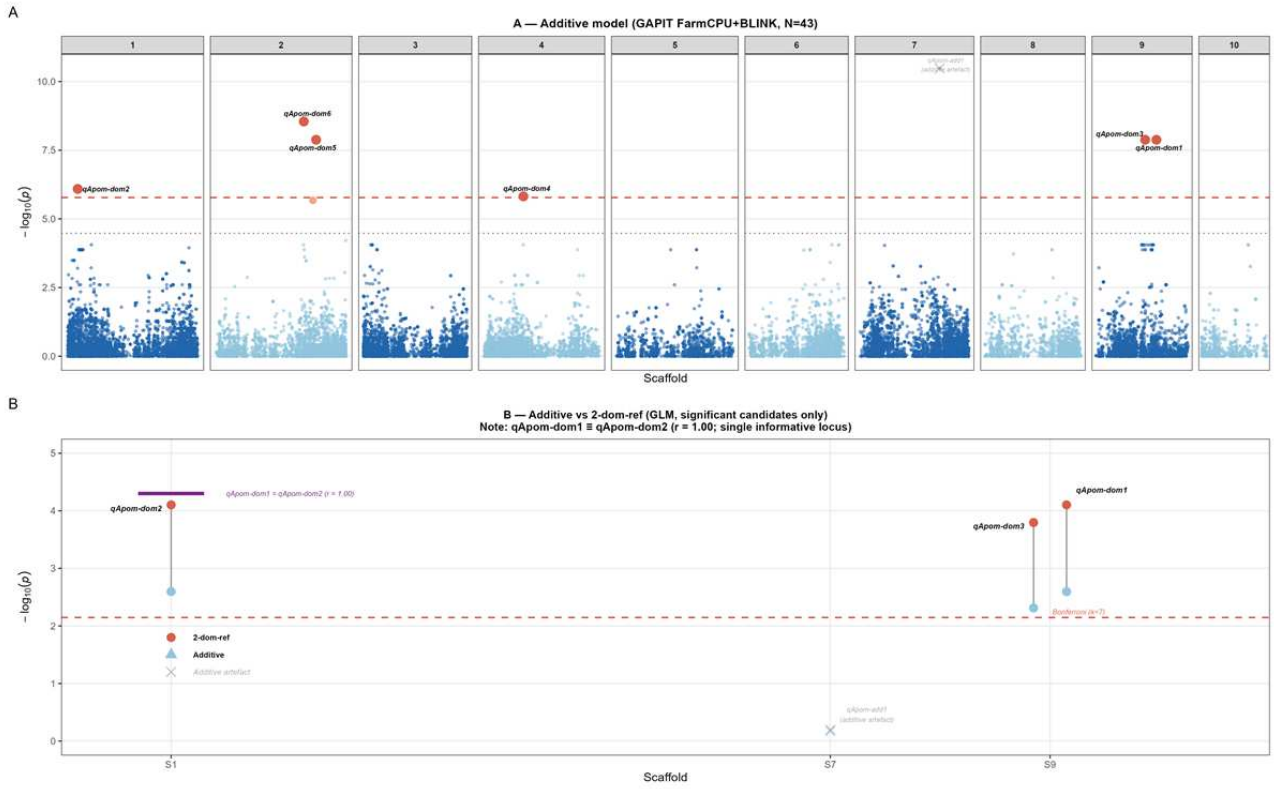


Fig. 7 Genome-wide association analysis for mode of reproduction across CRZ–22 (R81 × IPY; $n = 43$). **(A)** Manhattan plot of $-log_{10}p$ scores under the additive model (*FarmCPU* and *BLINK* algorithms, *GAPIT3*). Seven SNPs reached genome-wide Bonferroni significance ($\alpha = 0.05/29{,}910$ markers tested; dashed red line). The strongest additive signal was detected at qApom-add1 (scaffold 7), which was subsequently identified as a false positive on the basis of dominance testing ($AIC = 59.65$; 2-dom-ref; $p = 0.636$) and excluded from validated candidates. **(B)** Manhattan plot of $-log_{10}(p)$ scores under the 2-dom-ref dominance model (logistic GLM), applied *post hoc* to the seven candidate SNPs identified in panel An owing to instability of variance component estimation in *GWASpoly* at $n = 43$. Three validated candidate loci are labeled: qApom-dom1 (scaffold 9, position 34.4 Mb; $MAF = 0.336$), qApom-dom2 (scaffold 1, position 6.7 Mb; $MAF = 0.460$), and qApom-dom3 (scaffold 9, position 28.0 Mb; $MAF = 0.435$). qApom-dom1 and qApom-dom2 presented identical *p* values under the 2-dom-ref model ($p = 7.9 \times 10^{-5}$), a 32-fold improvement over the additive model ($p = 2.54 \times 10^{-3}$), and were perfectly correlated ($r = 1.00$), indicating that a single informative locus likely reflects incomplete assembly of the diploid *U. ruziziensis* reference genome. qApom-dom3, located approximately 6.4 Mb from qApom-dom1 on scaffold 9, also reached Bonferroni significance under the dominance model ($p = 1.6 \times 10^{-4}$ vs $p = 4.9 \times 10^{-3}$ under the additive model). Model comparison via the AIC confirmed that qApom-dom1 was the most parsimonious single-locus predictor of the mode of reproduction (AIC = 37.79). The dashed red and dotted gray lines indicate the Bonferroni and suggestive significance thresholds, respectively.

Model comparison via the AIC confirmed that qApom-dom1 alone was the most parsimonious predictor of the mode of reproduction ($\mathrm{AIC} = 37.79$), outperforming the combined model with all three candidates ($\mathrm{AIC} = 38.63$) and the model including qApom-dom3 ($\mathrm{AIC} = 39.10$). The 2-dom-ref effect structure of the significant loci indicates that carrying at least one copy of the reference allele (dosage ≤ 3) is sufficient to confer an association with apomixis, which is consistent with a dominant inheritance mechanism. This pattern mirrors the known dominance of the Apospory-Specific Genomic Region (ASGR) in *Urochloa*, where a single dose of the apomictic-associated region is sufficient for apospory expression.

Allelic dosage analysis confirmed the biological coherence of the candidate markers. All confirmed obligate apomicts (K6, K18, and K33) carried a dosage of 4 at all three loci, with the absence of the reference allele, whereas confirmed sexual hybrids (K38, K40, and K43) carried dosages ≤ 3 at all three loci. Three obligate apomictics (K6, K18, and K33) presented a predicted probability of apomixis of 0.19 under the dominant model, which is paradoxically lower than that of facultative apomictics (0.88). This apparent paradox reflects the fact that the candidate markers tag a dominant reference allele associated with facultative apomixis, whereas obligate apomictics (clones of R81) carry the alternative allele in homozygosity (dosage = 4), suggesting that obligate and facultative apomixis in this cross may be governed by distinct allelic configurations at the same locus or by different genetic mechanisms altogether. All candidate loci for apomixis were located on scaffolds distinct from those harboring spittlebug resistance loci (S5, S6, S7), confirming that the genomic architecture of apomixis and resistance are independent in this population.

Discussion

Genomic parentage verification as a prerequisite for GWAS in polyploid forage breeding

The present study demonstrated that genomic parentage verification via the genomic relationship matrix $\boldsymbol{G}$ is an indispensable first step prior to GWAS in interspecific *Urochloa* hybrid populations. Initially, of the 243 hybrids evaluated, only 23% ($n = 56$) were confirmed as true interspecific hybrids of the declared parents, while 48.5% ($n = 118$) were identified as *U. brizantha* pollen contaminants, 19.3% ($n = 47$) were identified as apomictic clones of the maternal parent, and the remaining 9.0% presented ambiguous or unconfirmed parentage profiles. These proportions highlight the magnitude of cryptic genomic misclassification that can occur in forage breeding programs involving apomictic species and semicontrolled crosses. Without prior parentage verification, GWAS analyses would have been conducted on a population in which nearly half of the individuals shared no genomic background with the target cross, and approximately one fifth were genomically identical clones, two conditions that severely compromise the validity of association mapping.

The importance of rigorous population characterization prior to genomic analyses has been emphasized in recent reviews of genetics and genomics in *Urochloa* (Ferreira et al. 2021; Simeão et al. 2021). However, to our knowledge, the present study is the first to systematically quantify the impact of the mode of reproduction misclassification and pollen contamination on GWAS signal validity in this genus. The consequences of ignoring modes of reproduction and parentage are exemplified by the spurious SNP signal detected at scaffold 6 in the $n = 243$ scenario, which reached genome-wide Bonferroni significance under a dominance model but disappeared entirely in purified scenarios phenotyped and genotyped hybrids ($n = 201$) and confirmed interspecific hybrids ($n = 56$), with progressive score deflation consistent with inflation by apomictic clones. This pattern illustrates a fundamental challenge in forage grass genomics: apomictic clones introduce artificial genomic redundancy that inflates association signals regardless of their biological relevance to the trait under study. Although the $\boldsymbol{G}$ matrix incorporated as a kinship covariate in *GWASpoly* partially controls for population stratification (genomic inflation factors $\lambda = 0.833 - 1.097$ across all scenarios), it cannot fully account for the nonrandom clustering of apomictic clones, which represent extreme cases of identity-by-descent that distort allele frequency estimates and shrinkage parameters.

The diverse modes of reproduction, both apomictic and sexual, within intra- and interspecific crosses in *Urochloa*, combined with a limited understanding of chromosomal behavior during meiosis and high heterozygosity, significantly hinder crossing, evaluation, and selection activities (Espitia-Buitrago et al. 2026). The discovery that the female parent H830–9 was itself an undeclared F_1 hybrid of H327–9 × IPY, detected exclusively through $\boldsymbol{G}$ matrix analysis, further illustrates the risks associated with the use of phenotypically or morphologically identified founders in breeding programs. Such cryptic pedigree errors are particularly likely in *Urochloa* programs where apomictic clones and their sexual derivatives are phenotypically similar and where the facultative nature of apomixis in several accessions makes visual identification of the mode of reproduction unreliable. Collectively, these findings support the systematic implementation of genomic parentage panels as a routine quality control step prior to phenotypic evaluation and genomic analysis in forage breeding programs involving *Urochloa* and related apomictic genera. The $\boldsymbol{G}$ matrix framework implemented here, which is based on 45,937 GBS-derived SNPs under the tetraploid VanRaden model, provides unambiguous resolution of parentage, clonality, and mode of reproduction across all 243 hybrids evaluated, at a genotyping cost comparable to that of standard GBS applications (Matias et al. 2019; Simeão et al. 2021).

The interspecific hybrid population evaluated here reflects the complex genomic architecture of the polyploid *Urochloa*, in which the tetraploid apomictic species (*U. brizantha*, *U. decumbens*) evolved predominantly through allopolyploidy but retained autopolyploid segments in specific genomic regions, a process termed segmental allopolyploidy (Worthington et al. 2016; da Costa Lima Moraes et al. 2023). This genomic mosaic has direct

consequences for the interpretation of pairwise $\boldsymbol{G}$ coefficients: partial homeology between subgenomes generates moderate positive $\boldsymbol{G}$ values between individuals of different species, inflating the expected relationship under a purely allopolyploid model and complicating the discrimination between true hybrids and contaminants on the basis of $\boldsymbol{G}$ thresholds alone. The observation that confirmed interspecific hybrids presented mean pairwise $\boldsymbol{G}$ values of 0.898, substantially higher than the value of 0.50 expected for full-sibs under diploid inheritance, is consistent with the amplification of genomic similarity by shared autopolyploid segments and reinforces the need for species-specific calibration of $\boldsymbol{G}$ matrix thresholds in polyploid forage breeding programs. Indeed, da Costa Lima Moraes et al. (2023) demonstrated contrasting allopolyploid and autopolyploid behaviors in parental genotypes of *U. humidicola*, highlighting how the mode of reproduction and ploidy origin jointly determine the genomic architecture of interspecific populations in this genus.

A GWAS for *Notozulia entreriana* resistance identified two robust candidate loci for damage scoring, qDam-dom1 on scaffold 6 (37.3 Mb) and qDam-dom2 on scaffold 7 (6.2 Mb), which were consistently detected across multiple analytical scenarios and explained approximately 23.3% and 25.1% of the genetic variance for damage, respectively, accounting for ~48% of the total genetic variance when considered jointly. These results closely mirror the findings of Espitia-Buitrago et al. (2026), who investigated *Aeneolamia varia* spittlebug resistance in 339 *Urochloa* F_1 interspecific hybrids. By performing a GWAS with 57,051 SNPs aligned to the allotetraploid cv. Basilisk reference genome, the authors identified six robust QTLs for plant damage traits on chromosomes 1, 6, and 7, which accounted for up to 21.5% of the phenotypic variance per locus. The independent convergence of association signals on homologous chromosomal regions across two different populations, spittlebug species, and reference genomes substantially strengthens the candidacy of scaffolds S6 and S7 as harboring functionally relevant resistance loci in interspecific *Urochloa*. Notably, Espitia-Buitrago et al. (2026) similarly reported no QTLs for nymph survival (antibiosis) in their larger panel, which is consistent with our observation that qSur-add1 was detectable only in the $n = 56$ confirmed hybrid scenarios, suggesting that antibiosis-related resistance components may require highly homogeneous populations and precise phenotyping for genomic detection.

Previous QTL mapping for spittlebug resistance in *Urochloa* relied primarily on biparental populations and diploidized marker data. Ferreira et al. (2019) constructed a genetic map of tetraploid *U. decumbens* with allele dosage information and detected QTLs for resistance to *Notozulia entreriana* explaining between 4.66 and 6.24% of the phenotypic variation. The substantially larger proportions of genetic variance explained by qDam-dom1 and qDam-dom2 in the present study (23,3% and 25.1%, respectively) may reflect the advantage of interspecific hybridization in amplifying allelic contrasts between susceptible (*U. ruziziensis*) and resistant (*U. brizantha*) parental species or,

alternatively, may indicate that these loci tag different functional regions than those detected in intraspecific populations. Jank et al. (2025) provided an additional genomic context for candidate gene identification in these regions.

The proportion of genetic variance explained jointly by qDam-dom1 and qDam-dom2 (~48%) suggests that resistance to *N. entreriana* in this interspecific cross may be governed by a relatively small number of loci with large individual effects. This oligogenic architecture contrasts with the highly polygenic patterns typically assumed for complex resistance traits in tropical grasses (Simeão et al. 2021) and may partly reflect the interspecific nature of the cross, in which large genomic blocks from resistant (*U. brizantha*) and susceptible (*U. ruziziensis*) parents segregate with minimal recombination, generating strong marker–trait associations at loci flanking these blocks. A similar oligogenic pattern was suggested by Resende et al. (2024), who estimated moderate-to-high heritability for nymph survival in recurrent selection populations of *U. ruziziensis* ($h_g^2 = 12.2 - 38.4$), and reported that progressive genetic gain across 16 selection cycles, is consistent with the presence of resistance alleles with substantial individual effects that respond predictably to selection pressure.

The retention of low-frequency alleles ($MAF \geq 0.05$) in the SNP panel was a deliberate analytical decision with direct consequences for candidate locus detection. Both qDam-dom1 ($MAF = 0.099$) and qDam-dom2 ($MAF = 0.232$) present moderate minor allele frequencies, which is consistent with alleles derived from the resistant *U. brizantha* parental species segregating in a predominantly *U. ruziziensis* genetic background. Had a more stringent minor allele frequency (*MAF*) filter been applied, these loci, and potentially additional resistance-associated variants, would have been discarded prior to analysis. This concern is particularly acute in interspecific breeding populations, where species-specific alleles conferring resistance, tolerance, or other adaptive traits may be rare at the population level while remaining functionally critical, as recently demonstrated for low-call-rate SNPs and presence-absence variation in rice genomic prediction (Krusenbaum and Wissuwa 2025). Future GWAS efforts in *Urochloa* interspecific populations should therefore carefully consider the trade-off between statistical power, which increases with stricter *MAF* filters, and biological completeness, which necessitates the retention of rare, species-specific variants.

The dominance effect structure of both major loci is noteworthy from a breeding perspective. The 1-dom-alt model at qDam-dom1 indicates that a single copy of the alternative allele is sufficient to confer the susceptibility phenotype, whereas the 1-dom-ref model at qDam-dom2 indicates that a single copy of the reference allele confers a measurable reduction in damage score ($\beta = -0.474$ damage units). This haploinsufficiency-like pattern, where a single allele dose is sufficient for effect expression, is consistent with the dominant resistance mechanisms described in other grass-pest systems and has practical implications for marker-assisted selection in tetraploid *Urochloa*: even heterozygous carriers of the resistance allele at qDam-dom2 (dosages 1 to 3) are expected to express reduced susceptibility, maximizing the

probability of recovering resistant progeny in subsequent crosses. Recent reviews have highlighted the critical importance of incorporating dosage-aware models in polyploid GWASs to avoid the loss of dominance signals caused by diploidization of tetraploid data (Ferreira et al. 2021), a limitation confirmed here by the nondetection of qDam-dom2 in *GAPIT* analyses based on diploidized genotypes ($p = 0.59$) despite its clear signal in *GWASpoly* tetraploid models.

The molecular mechanisms underlying spittlebug resistance in *Urochloa* remain poorly characterized, but recent transcriptomic and genomic studies have begun to provide candidate gene hypotheses. Begnami et al. (2025) demonstrated in *Paspalum regnellii* that resistance to *Mahanarva spectabilis* nymphs involves kinase signaling cascades, glutathione metabolism, and constitutive structural barriers, including lignin deposition, which are pathways that are consistent with the phenylpropanoid-mediated biochemical responses reported for *Urochloa* (Achon and Enciso-Maldonado 2026). The broad-sense *Cullis* heritability estimates obtained here ($H^2_Cullis = 0.570 - 0.793$ for damage; 0.702–0.885 for survival) substantially exceeded conventional broad-sense estimates ($H^2 = 0.083 - 0.343$) (**Supplemental Table 1**), reflecting the high precision of BLUP estimation under the replicated experimental design, and are consistent with the moderate-to-high heritabilities reported in recurrent selection programs for the same trait (Resende et al. 2024). The increase in *Cullis'* heritability from the identity-based model to the GBLUP framework (damage: 0.570 → 0.793; survival: 0.702 → 0.885) indicates that genomic information substantially improves the accuracy of genetic value estimation for both traits, supporting the integration of the genomic selection approaches highlighted by Simeão et al. (2021) into future *Urochloa* resistance breeding cycles.

These genomic relationship patterns provide crucial insights into the complex genetic architecture of the evaluated *Urochloa* population. First, the moderate genomic relationship coefficients ($G \sim 0.1 - 0.4$) observed between parental genotypes of distinct species, such as MAR, BAS, IPY, and CAY, are highly consistent with a segmental allopolyploid model. Because these tetraploid species share conserved autopolyploid genomic regions, their realized kinship remains positive and moderate rather than negative, reflecting significant historical germplasm sharing and ancestral synteny across species boundaries. Furthermore, this segmental architecture explains the exceptionally high genomic relationship coefficient ($\boldsymbol{G} = 0.898$) observed among confirmed hybrids from the R81 × IPY cross. R81 represents an artificially chromosome-duplicated *U. ruziziensis* line, whereas IPY is an interspecific hybrid containing autopolyploid segments. The segregation and inheritance of these shared segmental regions effectively amplify the genomic similarity among full-sib progeny, leading to higher-than-expected relationship values.

Second, the intermediate heterozygosity level of 0.508 observed among the apomictic individuals, which are clonal replicates of the tetraploid maternal parent R81, directly reflects the strict preservation of the maternal segmental structure. Because apomixis bypasses meiotic recombination, the specific combination of autopolyploid and allopolyploid segments

that defines the maternal genome is faithfully transmitted to the offspring, locking in its baseline heterozygosity. This complex polyploid background also shapes the genetic architecture of resistance to the spittlebug. The identification of a 1-dom-alt effect at the qDam-dom1 locus suggests that marker–trait associations are strongly influenced by segmental dosage dynamics. In purely autopolyploid regions of the *Urochloa* genome, the accumulation of alternative alleles typically results in additive phenotypic effects owing to simple dosage scaling. In contrast, within allopolyploid or homoeologous segments, intergenomic interactions often lead to prominent dominance effects, where the presence of a single alternative allele is sufficient to confer robust antibiosis resistance.

The scatter plot of genomic breeding values (GBLUP) for damage score and nymph survival revealed a clear structure in the distribution of resistance-associated genotypes, with 41 individuals occupying the antibiosis quadrant ($\text{GBLUP_damage} < 0; \text{GBLUP_survival} < 0$), corresponding to genotypes that simultaneously reduced plant damage and impaired nymph survival, the hallmark of antibiosis resistance (Cardona et al. 1999). The substantial representation of apomictic clones in the antibiosis quadrant is consistent with their maternal origin from H327–9 and H581–9, which carry the qDam-dom2 reference allele and reflects the genomic adjustment provided by the $\boldsymbol{G}$ matrix covariance structure, GBLUP of apomictic clones is pulled toward the maternal parent genomic value rather than inflated by clonal phenotypic repetition, as would occur under identity-based BLUP.

The correlation between phenotypic BLUPs and GBLUPs was $r = 0.49$ for damage score and $r = 0.77$ for nymph survival, indicating that the genomic framework provided substantially different, and more precisely, ranking of genotypes for the damage trait, where conventional heritability was low ($H^2 = 0.083$) and the $\boldsymbol{G}$ matrix covariance structure contributed the most additional discriminating power. This divergence between phenotypic and genomic rankings underscores the importance of GBLUP-based selection criteria in interspecific *Urochloa* breeding programs, where low-heritability traits such as insect damage are difficult to select for phenotypically but may be amenable to genomic selection given adequate marker density and population structure information (Simeão et al. 2021).

The four triple-carrier genotypes, K15, K6, K25, and H340–74, which are all classified as apomictic and carry qDam-dom1, qDam-dom2, and qDam-add1 simultaneously, presented the most negative GBLUP values for nymph survival (range: -4.75% to -16.20%), which is consistent with a cumulative dosage effect of multiple resistance alleles. These genotypes inherited qDam-dom1 and qDam-add1 from the IPY paternal parent, the only male parent confirmed to carry these alleles, and qDam-dom2 from their maternal parents, suggesting a complementary resistance architecture in which paternal and maternal genomic contributions each contribute distinct resistance components. This pattern supports the breeding strategy of crossing IPY-derived hybrids with maternal parents carrying the qDam-dom2 reference allele to maximize the probability of recovering progeny with cumulative resistance allele configurations in future cycles.

Genomic dissection of the mode of reproduction in the R81 × IPY family

The GWAS for mode of reproduction in the CRZ–22 family identified three Bonferroni-significant candidate loci under a dominant inheritance model, qApom-dom1 (S9, 34.4 Mb), qApom-dom2 (S1, 6.7 Mb), and qApom-dom3 (S9, 28.0 Mb), with substantially stronger signals under the 2-dom-ref model than under the additive model (32-fold improvement in *p value* for qApom-dom1 and qApom-dom2). The dominant inheritance pattern observed here, where a single copy of the reference allele at dosage ≤ 3 is sufficient for association with apomixis, is fully consistent with the well-established dominant inheritance of apospory in *Urochloa* (Worthington et al. 2016) and with the haploinsufficiency model proposed for the ASGR in *Pennisetum* (Akiyama et al. 2011). These results reinforce the conclusions of Ferreira et al. (2021) and Simeão et al. (2021), who emphasized that the dominant and dosage-sensitive nature of apomixis-controlling loci in *Urochloa* demands polyploid-aware statistical frameworks for their genomic detection, a requirement fulfilled here through the use of *GWASpoly* and complementary GLM dominance testing.

The perfect correlation between qApom-dom1 and qApom-dom2 ($r = 1.00$), despite their location on different scaffolds (S9 and S1), most likely reflects incomplete chromosomal assembly in the reference *U. ruziziensis* diploid genome, where syntenic regions of the same chromosome may be distributed across multiple scaffolds. This interpretation is supported by the work of da Costa Lima Moraes et al. (2023), who mapped the ASGR locus in *U. humidicola* to linkage group 3, in a region syntenic with chromosome 3 of *U. ruziziensis*, suggesting that scaffold S9 in the present study may correspond to chromosome 3 of the reference genome and that qApom-dom1 and qApom-dom2 together tag the *U. ruziziensis* ASGR-syntenic region. Model comparison via the AIC confirmed that qApom-dom1 alone was the most parsimonious predictor of the apomixis mode ($\mathrm{AIC} = 37.79$), which is consistent with the single-locus dominant inheritance model for apospory described in *Urochloa* (Worthington et al. 2016) and *Pennisetum* (Akiyama et al. 2011).

The paradoxical observation that confirmed that obligate apomicts (K6, K18, K33) carried dosage = 4 at all three candidate loci, the complete absence of the reference allele, whereas facultative apomicts carried dosage ≤ 3, suggests that obligate and facultative apomixis in this cross may be governed by distinct allelic configurations at the same locus. One interpretation is that the reference allele identified here tags a modifier or suppressor of obligate apomixis, whose presence in one or more copies shifts the reproductive phenotype toward facultatively, whereas its complete absence results in obligate apospory. This model would predict that the paternal parent IPY, which has a dosage = 2 at all three loci and is phenotypically apomictic, contributes the reference allele to the cross, generating a population in which the presence of the IPY-derived reference allele at moderate dosage (1-3 copies) confers facultative apomixis, whereas its absence (dosage = 4, all *U. ruziziensis*-derived alternative alleles) results in obligate clonal reproduction. Functional

characterization of the genomic region tagged by qApom-dom1 in chromosome 3-syntenic regions of *U. ruziziensis* is needed to test this hypothesis.

The three individuals classified as sexually reproducing by PCR but carrying the apomixis-associated reference allele at all three loci (K38, K40, K43; predicted probability of apomixis = 0.88) represent high-priority candidates for cytological confirmation. Given the reported 15% false-negative rate of the SCAR p779/p780 marker (Worthington et al. 2016), these individuals may be facultative apomicts whose aposporic embryo sacs were not detected in the tissue sampled. The cross-validation between PCR-based phenotyping and $\boldsymbol{G}$ matrix classification revealed an overall concordance of only 17%, underscoring the limitations of single-marker PCR assays for mode of reproduction determination in populations with high rates of facultative apomixis. This finding has direct implications for the use of molecular markers in *Urochloa* breeding programs: the combined use of the SCAR p779/p780 marker (Akiyama et al. 2011; Worthington et al. 2016), genomic relationship coefficients, and allele dosages at candidate loci provides substantially more reliable modes of reproduction classification than any single method alone does and should be considered a standard protocol in programs involving interspecific crosses of facultatively apomictic parents (Jank et al. 2025).

All candidate loci for apomixis were located on scaffolds distinct from those harboring spittlebug resistance loci (S5, S6, S7), confirming that the genomic architecture of apomixis and resistance are independent in this population. This independence is an important finding for breeding strategies: marker-assisted introgression of resistance alleles at qDam-dom1 and qDam-dom2 into apomictic backgrounds would not be expected to inadvertently disrupt the apomixis-controlling region, facilitating the development of apomictic cultivars that combine high-level spittlebug resistance with clonal seed production, the ultimate goal of interspecific *Urochloa* improvement programs (Simeão et al. 2021; Ferreira et al. 2021; Espitia-Buitrago et al. 2026).

The interspecific hybrid population evaluated in this study reflects the complex genomic architecture of tetraploid *Urochloa*, in which the tetraploid apomictic species (*U. brizantha* and *U. decumbens*) evolved predominantly through allopolyploidy but retained autopolyploid segments in specific genomic regions, a process termed segmental allopolyploidy (Worthington et al. 2016). This genomic mosaic has direct consequences for the interpretation of pairwise genomic relationship coefficients: the partial homeology between subgenomes generates moderate positive $\boldsymbol{G}$ values between individuals of different species, inflating the expected relationship under a purely allopolyploid model and complicating the discrimination between true hybrids and contaminants on the basis of $\boldsymbol{G}$ thresholds alone. The observation that confirmed interspecific hybrids presented mean pairwise $\boldsymbol{G}$ values of 0.898, substantially higher than the value of 0.50 expected for full-sibs under diploid inheritance, is consistent with the amplification of genomic similarity

by shared autopolyploid segments and reinforces the need for species-specific calibration of $\boldsymbol{G}$ matrix thresholds in polyploid forage breeding programs.

Conclusions

The present study demonstrated that genomic parentage verification via the tetraploid genomic relationship matrix $\boldsymbol{G}$ is an indispensable prerequisite for genome-wide association studies in interspecific *Urochloa* hybrid populations. Among the 243 evaluated hybrids, only 23.0% were confirmed as true interspecific hybrids of the declared parents, whereas 48.0% were identified as *U. brizantha* pollen contaminants and 19.3% as apomictic clones of the maternal parent. Without prior parentage verification, GWAS analyses would have been conducted on a severely misclassified population, as evidenced by the spurious association signal detected at Scaffold 6 in the unfiltered scenario that disappeared entirely upon population purification.

Following parentage verification and population purification, GWAS identified two robust candidate loci for spittlebug (*Notozulia entreriana*) resistance, qDam-dom1 (scaffold 6, 37.3 Mb) and qDam-dom2 (scaffold 7, 6.2 Mb), which were consistently detected across multiple analytical scenarios and jointly explained approximately 48% of the genetic variance in the damage score. Both loci operate under dominant inheritance models, with a single allele dose sufficient for effect expression, which is consistent with resistance mechanisms previously described in *Urochloa* and related grass genera. A third locus, qSur-add1 (scaffold 5), was exclusively detected in the confirmed hybrid scenario under an additive model and was retained only because of the deliberate choice to apply a permissive $MAF \geq 0.05$ filter, a decision that proved critical, as three of four candidate resistance markers would have been discarded under an $MAF > 0.10$ threshold, demonstrating the importance of retaining rare alleles in interspecific breeding populations.

For the apomixis mode, a GWAS of the R81 × IPY family identified three candidate loci, namely, qApom-dom1 (scaffold 9, 34.4 Mb), qApom-dom2 (scaffold 1, 6.7 Mb), and qApom-dom3 (scaffold 9, 28.0 Mb), under a dominant inheritance model consistent with the known dominance of the Apospory-Specific Genomic Region in *Urochloa*. The perfect correlation between qApom-dom1 and qApom-dom2 ($r = 1.00$), despite their location on different scaffolds, likely reflects incomplete assembly of the reference *U. ruziziensis* genome rather than independent loci, and their syntenic correspondence with the ASGR-containing chromosome 3 of *U. ruziziensis* warrants functional validation. Critically, all apomixis candidate loci were located on scaffolds distinct from those harboring spittlebug resistance loci, confirming the genomic independence of these two traits and supporting the feasibility of combining high-level spittlebug resistance with apomictic seed reproduction through marker-assisted introgression in future *Urochloa* breeding cycles.

Taken together, these results establish a genomic framework for resistance and mode of reproduction improvement in interspecific tetraploid *Urochloa* and underscore three methodological lessons of broad applicability to forage grass breeding programs: (i) genomic parentage verification should precede any genomic analysis involving apomictic species or open-pollinated crosses; (ii) permissive *MAF* filters are essential in interspecific populations to avoid discarding rare species-specific resistance alleles; and (iii) dosage-aware and dominant-effect models substantially outperform additive diploidized approaches for detecting nonadditive loci in tetraploid interspecific populations.

Reference list

Achon I, Enciso-Maldonado GA (2026) Integrating molecular insights on stress responses and agronomic traits for breeding in forage grasses. Crop Pasture Sci 77:CP25162. https://doi.org/10.1071/CP25162

Akiyama Y, Goel S, Conner JA, Hanna WW, Yamada-Akiyama H (2011) Evolution of the apomixis transmitting chromosome in *Pennisetum.* BMC Evol Biol 11:289. https://doi.org/10.1186/1471-2148-11-289

Amadeu RR, Cellon C, Olmstead JW, Garcia AAF, Resende MFR Jr, Muñoz PR (2016) *AGHmatrix*: R Package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Gen 9: plantgenome2016.01.0009. https://doi.org/10.3835/plantgenome2016.01.0009

Barrios SCL, Valle CB., Alves GF, Simeão RM, Jank L (2013) Reciprocal recurrent selection in the breeding of *Brachiaria decumbens*. Trop Grass 1:52-54. https://doi.org/10.17138/tgft(1)52-54

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using *lme4*. J Stat Softw 67:1-48. https://doi.org/10.18637/jss.v067.i01

Begnami IdS, Aono AH, Graciano DdS, Cermello-Guerreiro SM, Ferreira RCU, Malagó Jr W, Matta FP, Souza AP, Vigna BBZ (2025) Elucidating molecular responses to spittlebug attack in *Paspalum regnellii*. Plant Mol Biol Rep 43:307-323. https://doi.org/10.1007/s11105-024-01487-w

Caamal-Pat D, Pérez-Rodríguez P, Crossa J, Velasco-Cruz C, Pérez-Elizalde S, Vázquez-Peña M (2021) *lme4GS*: An R-package for genomic selection. Front Genet 12:680569. https://doi.org/10.3389/fgene.2021.680569

Cardona C, Miles JW, Sotelo G (1999) An improved methodology for massive screening of *Brachiaria* spp. genotypes for resistance to *Aeneolamia varia* (Homoptera: Cercopidae). J Econ Entomol 92:490-496. https://doi.org/10.1093/jee/92.2.490

Correa CTR, Vaio M, Barrios SCL, Valle CB, Torres GA, Techio VA (2025) The repetitive DNA landscape in the brizantha agamic complex of *Urochloa* P. Beauv.. Genome 68:1-13. https://doi.org/10.1139/gen-2024-0096

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package *sommer*. PLoS ONE 11: e0156744. https://doi.org/10.1371/journal.pone.0156744

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381-393. https://doi.org/10.1198/108571106X154443

da Costa Lima Moraes A, Mollinari M, Ferreira RCU, Aono A, Lara LAC, Pessoa-Filho M, Barrios SCL, Garcia AAF, Valle CB, Souza AP, Vigna BBZ (2023) Advances in genomic characterization of *Urochloa humidicola*: exploring polyploid inheritance and apomixis. Theor Appl Genet 136:238. https://doi.org/10.1007/s00122-023-04485-w

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and *VCFtools*. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Espitia-Buitrago P, Perea C, Mejia-Medina JC, Hernández LM, Castiblanco V, Ryan C, De Vega JJ, Jauregui RN (2026) Integrating image-based phenotyping and GWAS to map resistance to spittlebug nymphs in interspecific *Urochloa* grasses. G3 Genes Genomes Genet 2026:jkag101. https://doi.org/10.1093/g3journal/jkag101

Euclides VPB, Valle CB, Macedo MCM, Almeida RG, Montagner DB, Barbosa RA (2010) Brazilian scientific progress in pasture research during the first decade of XXI century. R Bras Zootec 39:151-168. https://doi.org/10.1590/S1516-35982010001300018

Ferreira RCU, Moraes ACL, Chiari L, Simeão RM, Vigna BBZ, Souza AP (2021) An overview of the genetics and genomics of the *Urochloa* species most commonly used in pastures. Front Plant Sci 12:770461. https://doi.org/10.3389/fpls.2021.770461

Ferreira RCU, Moraes ACL, do Valle CB, Garcia AAF, Barrios SCL, Souza AP (2019) Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R.D. Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance. Front Plant Sci 10:92. https://doi.org/10.3389/fpls.2019.00092

Fuzinatto VA, Pagliarini MS, Valle CB (2012) Meiotic behavior in apomictic *Brachiaria ruziziensis* × *B. brizantha* (Poaceae) progenies. Sci Agric 69:380-385. https://doi.org/10.1590/S0103-90162012000600006

Gerard D, Ferrão LFV, Garcia AAF, Stephens M (2018) Genotyping polyploids from messy sequencing data. Genetics 210:789–807. https://doi.org/10.1534/genetics.118.301468

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. https://doi.org/10.1371/journal.pone.0090346

Gouveia BT, Mateus RG, Barrios SCL, Valle CB, Bueno Filho JS, Fernando RE, Dias AM, Nunes JAR (2022) Combining ability and selection for agronomic and nutritional traits in *Urochloa* spp. hybrids. Grass Forage Sci 77:33-44. https://doi.org/10.1111/gfs.12555

Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019) *BLINK*: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8:giy154. https://doi.org/10.1093/gigascience/giy154

Jank L, Barrios SC, Valle CB, Simeão RM, Alves GF (2014) The value of improved pastures to Brazilian beef production. Crop Pasture Sci. 65:1132-1137. https://doi.org/10.1071/CP13319

Jank L, Rios E, Santos MF, Jauregui RN, Vigna BBZ, Barrios SCL, Acuña CA, Zilli AL, Valle CB, Reyno R, Raposo A, Altpeter F, Simeão R, Griffa S, Dall'Agnol M, Brunes AP, Tomaszewska P, Vilela M, Gesteira G, De Vega JJ, Colomba EL, Brugnoli EA, Marcón F, Martínez EJ, Do Canto J, Rizza MD, Azevedo ALS, Narancio R, Hojsgaard D (2025) Apomixis in farmers' fields: overview, case studies from forage grasses and considerations for future apomictic crops. Crit Rev Plant Sci 44:345-397. https://doi.org/10.1080/07352689.2025.2578983

Karunarathne P, Reutemann AV, James JE, Zhou Q, Sassone A, Rose LE, Hojsgaard D (2025) Navigating the challenges in apomixis population genetics: insights from past, present, and future perspectives. Crit Rev Plant Sci 44:261-292. https://doi.org/10.1080/07352689.2024.2440296

Krusenbaum L, Wissuwa M (2025) Low-call-rate SNPs and presence-absence variation identified in the rice pangenome can improve genomic prediction of rice gene bank accessions. Theor Appl Genet 138:295. https://doi.org/10.1007/s00122-025-05080-x

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with *Bowtie 2*. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Lara LAC, Santos MF, Jank L, Chiari L, Vilela MM, Amadeu RR, Santos JPR, Pereira GS, Zeng ZB, Garcia AA (2019) Genomic selection with allele dosage in *Panicum maximum* Jacq.. G3 Genes Genomes Genet. 8:2463-2475. https://doi.org/10.1534/g3.118.200986

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12:e1005767. https://doi.org/10.1371/journal.pgen.1005767

López-Contreras LM, Denning-James KE, Carvajal-Tapia J, Arango J, Villegas DM, Jauregui RN, Wenzl P, De Vega JJ, Carvajal-Yepes M (2025) Genome-wide association study of a Guinea grass (*Megathyrsus maximus*) diversity panel reveals the genetic basis of agronomic and nutritional traits. BMC Plant Biol 25:1785. https://doi.org/10.1186/s12870-025-08007-2

Lutts S, Ndikumana J, Louant BP (1991) Fertility of *Brachiaria ruziziensis* in interspecific crosses with *Brachiaria decumbens* and *Brachiaria brizantha*: meiotic behavior, pollen viability and seed set. Euphytica 57:267-274.

Matias FI, Barrios SCL, Meireles KGX, Valle CB, Carazzolle MF, Guimarães CT, Moro GV (2019) Association mapping considering allele dosage: an example of forage traits in an interspecific segmental allotetraploid *Urochloa* spp. panel. Crop Sci 59:2481-2494. https://doi.org/10.2135/cropsci2019.03.0185

Miles JW (2007) Apomixis for cultivar development in tropical forage grasses. Crop Sci 47:S238–S249. https://doi.org/10.2135/cropsci2007.04.0016IPBS

Pereira GS, Garcia AAF, Margarido GRA (2018) A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. BMC Bioinformatics 19:398. https://doi.org/10.1186/s12859-018-2433-6

Pessoa-Filho M, Souza Sobrinho F, Fragoso RR, Silva Junior OB, Ferreira ME (2019) A phased diploid genome assembly for the forage grass *Urochloa ruziziensis* based on single-molecule real-time sequencing. GenBank Assembly GCA_015476505.1. In International Plant and Animal Genome Conference XXVII, 2019, San Diego. https://www.embrapa.br/en/busca-de-publicacoes/-/publicacao/1107378/a-phased-diploid-genome-assembly-for-the-forage-grass-urochloa-ruziziensis-based-on-single-molecule-real-time-sequencing

Pinto de Paula CM, Sobrinho FS, Techio VH (2017) Genomic constitution and relationship in *Urochloa* (Poaceae) species and hybrids. Crop Sci 57: 2605-2616. https://doi.org/10.2135/cropsci2017.05.0307

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. https://doi.org/10.1371/journal.pone.0032253

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Resende TT, Sobrinho FS, Campagnani MO, Veríssimo BA, Calsavara LA, Gonçalves FMA, Nunes JAR, Auad AM (2024) Sixteen years of recurrent selection of ruzi grass for resistance to spittlebugs (Hemiptera: Cercopidae). Agronomy 14:1516. https://doi.org/10.3390/agronomy14071516

Risso-Pascotto C, Pagliarini MS, Valle CB (2006) A new basic chromosome number for the genus *Brachiaria* (Trin.) Griseb. (Poaceae: Panicoideae: Paniceae). Genet Resour Crop Evol 53:7-10. https://doi.org/10.1007/s10722-005-7762-4

Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9:1-10. https://doi.org/10.3835/plantgenome2015.08.0073

Slater AT, Cogan NOI, Hayes BJ, Schultz L, Dale MFB, Bryan GJ, Forster JW (2014) Improving breeding efficiency in potato using molecular and quantitative genetics. Theor Appl Genet 127:2279–2292. https://doi.org/10.1007/s00122-014-2386-8

Simeão RM, Resende MDV, Alves RS, Pessoa-Filho M, Azevedo ALS, Jones CS, Pereira JF, Machado JC (2021) Genomic selection in tropical forage grasses: current status and future applications. Front Plant Sci 12:665195. https://doi.org/10.3389/fpls.2021.665195

Simioni C, Valle CB (2009) Chromosome duplication in *Brachiaria* (A. Rich.) Stapf allows intraspecific crosses. Crop Breed Appl Biotechnol 9:328-334.

Sotelo G, Cardona C, Miles J (1998) An improved methodology to screen genotypes of *Brachiaria* spp for resistance to spittlebug (Homoptera: Cercopidae) in greenhouse conditions. Rev Colombiana Entomol 24 https://doi.org/10.25100/socolen.v24i1.9826

Swenne A, Louant BP, Dujardin M (1981) Induction par la colchicine de formes autotétraploıdes chez *Brachiaria ruziziensis* Germain et Evrard (Gramineae). Agron Trop 36:134-141.

Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14:e20077. https://doi.org/10.1002/tpg2.20077

Torres-González AM, Morton CM (2005) Molecular and morphological phylogenetic analysis of *Brachiaria* and *Urochloa* (Poaceae). Mol Phylogenet Evol 37:36-44. https://doi.org/10.1016/j.ympev.2005.06.003

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414-4423. https://doi.org/10.3168/jds.2007-0980

Wang J, Crossa J, Gai J (2020) Quantitative genetic studies with applications in plant breeding in the omics era. Crop J 8:683-687. https://doi.org/10.1016/j.cj.2020.09.001

Wang J, Zhang Z (2021) *GAPIT* version 3: boosting power and accuracy for genomic association and prediction, Genomics Proteomics Bioinformatics 19:629-640. https://doi.org/10.1016/j.gpb.2021.08.005

Wickham H (2016) *ggplot2*: elegant graphics for data analysis. Springer, New York. https://doi.org/10.1007/978-3-319-24277-4

Worthington ML, Miles JW (2015) Reciprocal full-sib recurrent selection and tools for accelerating genetic gain in apomictic *Brachiaria*. In: Budak H, Spangenberg G (eds) Molecular breeding of forage and turf. Springer, Cham, pp 19–30. https://doi.org/10.1007/978-3-319-08714-6_3

Worthington M, Heffelfinger C, Bernal D, Quintero C, Zapata YP, Perez JG, Vega J, Miles J, Dellaporta S, Tohme J (2016) A parthenogenesis gene candidate and evidence for segmental allopolyploidy in apomictic *Urochloa decumbens*. Genetics 203:1117-1132. https://doi.org/10.1534/genetics.116.190314

Worthington M, Arango J, Jauregui RN, Castiblanco V, Perez JG, Fernandez-Fuentes N, Skot L, Dyer S, Tohme J, Jones CS (2019) Translocation of a parthenogenesis gene candidate to an alternate carrier chromosome in apomictic *Brachiaria humidicola*. BMC Genomics 20:41. https://doi.org/10.1186/s12864-018-5392-4

Statements & declarations

Funding

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Processo 442749/2014-7), Fundação de Apoio à Pesquisa do Estado de Minas Gerais (FAPEMIG - Processo APQ-02645-17), Fundação de Apoio ao Desenvolvimento de Ensino, Ciência e Tecnologia do Mato Grosso do Sul (FUNDECT - Processo 71/044.551/2022), and EMBRAPA (20.23.01.002.00.00, and 10.25.00.031.00.00).

Competing interests

The authors have no relevant financial or nonfinancial interests to disclose.

Declaration of generative AI and AI-assisted technologies

The authors acknowledge the use of generative AI tools, specifically Claude (Anthropic) and Gemini (Google), to assist with English language editing and proofreading, as English is not the authors' native language. Additionally, Claude (Anthropic) was used to optimize and refine R scripts, primarily for generating and formatting graphical figures to ensure a clear presentation of the experimental results. All scientific content, data analysis, interpretations, and final manuscript

edits were conducted, critically reviewed, and approved by the human authors, who take full responsibility for the integrity of the work.

Author contributions

Rosangela Simeão, Sanzio Carvalho Lima Barrios, Karem Guimarães Xavier Meireles and Anete Pereira de Souza conceived the project. Rosangela Simeão analyzed the data, interpreted the results, and wrote the manuscript with major contributions from Aline da Costa Lima Moraes and Sanzio Carvalho Lima Barrios. Aline da Costa Lima Moraes, Andrea Raposo, Yasmin Lany Ventura Said, and Eric Ragalzi conducted the laboratory experiments. Sanzio Carvalho Lima Barrios managed the field experiments and selected the open-pollinated hybrids. Fabrícia Zimermann Vilela Torres conducted the antibiosis experiments. Karem Guimarães Xavier Meireles performed the controlled crosses. All the authors have read and approved the final manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- SupplementalTables.xlsx
- SupplementlaFig.1QQplotscomposite.png
- SupplementalFig2ManhattanGAPIT.png