

Genetic architecture of rhizome development in diploid perennial sorghum reveals exploratory genomic signals with heterogeneous complementary support

Pheonah Nabukalu¹ · Wenqian Kong · Stan Cox · Jaxk Reeves · Andrew H. Paterson

Received: 25 May 2026 / Accepted: 7 September 2026

© The Author(s) 2026

Abstract

Background The genetic architecture of perenniality in sorghum remains incompletely understood, particularly following interspecific hybridization and recovery of diploid progeny. Rhizome development and overwinter regrowth are components of perennial persistence but may differ genetically and environmentally. Characterizing their genomic association patterns in diploid *Sorghum bicolor* × *S. halepense* material could identify candidate genomic regions for subsequent validation and breeding.

Results We investigated perenniality in S3011, a diploid-derived interspecific population from *Sorghum bicolor* × *S. halepense*. A structured F₄ panel was phenotyped under field conditions, and accession-level best linear unbiased predictors (BLUPs) were used for genome-wide association analyses. No association exceeded the Bonferroni-adjusted genome-wide significance threshold ($P \leq 1.15 \times 10^{-6}$; 43,317 SNPs tested), and none remained significant after Benjamini–Hochberg false discovery rate correction ($q < 0.05$). A predefined suggestive threshold ($P < 1 \times 10^{-4}$; $-\log_{10}(P) > 4$) identified 13 SNPs, which were consolidated into seven spatially distinct exploratory rhizome-number signals. Complementary characterization using high–low allele-dosage contrasts, local linkage disequilibrium, historical positional correspondence, candidate-gene annotation, and a descriptive marker-based genome scan revealed heterogeneous regional evidence. The strongest high–low allele-dosage differentiation occurred at chromosome 1 cluster 2 (S1_63359080; +0.3364; 100.0th chromosome-relative percentile), followed by chromosome 1 cluster 1 (S1_62967991; +0.3002; 99.3rd percentile). The complementary marker scan showed within-dataset cross-method concordance for selected rhizome-number signals, including the chromosome 3 region. Overwinter regrowth produced three exact marker-level GWAS–marker-scan matches but lacked the localized multi-marker regional structure observed for rhizome number. Because both association approaches used the same S3011 dataset, their agreement represents cross-method concordance rather than independent validation. Given the modest sample size, low trait heritability, and absence of associations surviving multiple-testing correction, all identified signals remain exploratory.

Conclusions Rhizome development and overwinter regrowth showed contrasting genomic association patterns in the S3011 population. Rhizome number was associated with seven exploratory signals that differed in allele-dosage differentiation and complementary regional support, whereas overwinter regrowth showed marker-level concordance without comparable localized multi-marker structure. These results do not establish confirmed QTL

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



or causal loci but prioritize genomic regions and allelic combinations for validation in independent populations and environments before their breeding utility can be established.

Keywords Perennial sorghum · Rhizome number · Overwinter regrowth · Genome-wide association study · Allele dosage · Linkage disequilibrium · *Sorghum halepense*

Introduction

Developing perennial grain sorghum requires integrating the agronomic attributes of cultivated *Sorghum bicolor* with the belowground persistence characteristic of its rhizomatous wild relative, *S. halepense*. The latter forms extensive rhizome networks, maintains belowground vegetative meristems, and exhibits vigorous regrowth following seasonal stress—traits that contribute to its ecological persistence and provide a biological route toward perennializing a major C₄ cereal [1–3]. In agricultural systems, perennial root and rhizome architectures can improve soil structure, enhance water and nutrient capture, increase soil carbon inputs, and contribute to resilience under climatic stress [4–8]. Despite these potential advantages, breeding perennial grain sorghum remains constrained by limited understanding of the genetic architecture underlying rhizome development and post-winter persistence.

Perenniality in sorghum is physiologically complex and encompasses multiple interacting components, including rhizome initiation and development, belowground meristem survival, dormancy, crown survival, resource storage, and overwinter regrowth. These processes are developmentally integrated but are also strongly conditioned by temperature, moisture, seasonal dynamics, and plant developmental state [9–11]. Rhizomes can contribute to perennial persistence by protecting vegetative meristems belowground and supporting renewed growth after seasonal stress, but rhizome development alone need not determine whether a plant survives winter and resumes growth [2, 12]. Rhizome development and overwinter regrowth may therefore represent related but genetically and environmentally distinct components of perenniality.

Early genetic studies involving perennial sorghum relatives identified genomic regions associated with rhizomatousness and related perenniality traits [2, 13]. More recent transmission-genetic and mapping studies likewise support the involvement of multiple genomic regions in the inheritance of rhizomatous growth and perennial persistence [14, 15]. However, many interspecific sorghum populations have involved polyploid, aneuploid, or mixed-ploidy material in which irregular meiosis, allele-dosage complexity, segregation distortion, and reduced fertility complicate genetic analysis [3, 14–16]. These limitations make it difficult to determine how perenniality-associated variation is organized after recombination and recovery of stable diploid progeny.

Comparative studies in other crop–wild relative systems provide additional evidence that perennial and adaptive traits may be genetically heterogeneous rather than governed by a single genomic architecture. In maize–teosinte populations, traits associated with rhizomatous growth and regrowth have been mapped to multiple genomic regions, indicating that perennial growth can be decomposed into genetically tractable components while still involving several loci and environmentally responsive processes [17–19]. More broadly, introgressed and domestication-related variation can occur within genomic regions that differ in local linkage disequilibrium, recombination history, and allelic composition [20–24]. These observations motivate genomic analyses that compare multiple forms of regional evidence without assuming that all biologically relevant loci will exhibit the same local genetic signature.

Diploid extraction provides a useful framework for examining these questions in sorghum. Recovery of diploid derivatives from interspecific material stabilizes chromosome number at $2n=20$ and can simplify segregation and allele-dosage interpretation relative to polyploid or cytogenetically unstable backgrounds. In the *KS105* × Gypsum 9 N pedigree, triploid intermediates expressed strong rhizomatous phenotypes but retained the meiotic instability typical of interspecific hybrids. Subsequent recovery and selfing of fertile diploid derivatives produced

the S3011 population, providing a diploid-extraction panel in which perenniality-associated traits can be evaluated within a genetically stabilized background.

Association-mapping frameworks that account for population structure provide one approach for characterizing marker–trait relationships in such populations [24–26]. In a modestly sized, structured diploid-extraction panel, however, suggestive marker associations should be interpreted cautiously because statistical power, allele frequency, relatedness, and local marker structure can influence the apparent strength and spatial organization of association signals. Accordingly, complementary analyses were used to compare regional patterns of linkage disequilibrium, allele-dosage differentiation, marker-scan correspondence, historical QTL context, and positional gene content without treating those secondary analyses as confirmation of the suggestive GWAS signals.

Under this framework, localized genomic structure need not imply a single major-locus model, and dispersed associations need not imply an absence of genetic control. Rhizome development may produce localized association signals that differ in the degree of support provided by complementary analyses, whereas overwinter regrowth may exhibit a more dispersed association profile because it integrates multiple biological processes and is particularly sensitive to environmental conditions. Distinguishing between these patterns is important for identifying genomic hypotheses that warrant validation while avoiding overinterpretation of suggestive associations as established QTL.

Here, we investigate the genomic architecture of rhizome number and overwinter regrowth in S3011, a diploid-extraction population derived from *S. bicolor* × *S. halepense*. We combine accession-level best linear unbiased predictors (BLUPs) with population-structure-adjusted genome-wide association analysis and complementary marker-based genome scans. For rhizome number, we further characterize the seven spatially distinct suggestive GWAS signals using lead-to-neighbor local linkage disequilibrium, GWAS-effect-oriented high–low allele-dosage contrasts, historical QTL correspondence, positional candidate-gene annotation, and selected region-specific sensitivity analyses. We use these complementary analyses to (i) identify and describe exploratory genomic signals associated with rhizome number, (ii) compare the degree and type of regional characterization among those signals, (iii) evaluate whether allele-dosage differentiation between high- and low-rhizome accessions is exceptional relative to chromosome and genome-wide backgrounds, (iv) assess positional correspondence with previously reported rhizomatousness loci, and (v) contrast the genomic association profile of rhizome development with that of overwinter regrowth. This framework is intended to prioritize testable genomic hypotheses for subsequent validation rather than to define confirmed QTL or causal loci.

Materials and methods

Plant materials and population development

The S3011 population was developed as a diploid-extraction system to recover perenniality-associated alleles from *Sorghum halepense* within a diploid *S. bicolor* background. In 2017, the diploid inbred KS105A (2×) was crossed with the tetraploid perennial accession Gypsum 9 N (4×). Interspecific hybridization produced F₁ progeny spanning diploid, triploid, and tetraploid cytotypes, with triploid derivatives exhibiting strong rhizomatous phenotypes but limited fertility.

Selfing of triploid individuals generated segregating F₂ progeny with mixed ploidy. Fertile diploid individuals exhibiting rhizome formation were selected and advanced through successive selfing to generate diploid F₄ families. These lines retain introgressed *S. halepense* genomic segments while exhibiting stabilized diploid inheritance.

Because the population was derived through a partially fertile triploid intermediate, seed production varied substantially among F₄ families. Consequently, the realized field population was constrained by seed availability, and neither the number of plots nor the number of plants established per accession was uniform across the experiment. Some accessions produced sufficient seed to be represented in multiple plots, whereas others could only be

established in a single plot or with fewer individual plants. As a result, the realized experiment did not constitute a perfectly balanced evaluation of all accessions.

A total of 129 diploid F₄ accessions comprising 855 individual plants were established and phenotyped in the field. Leaf tissue was collected from every surviving plant for DNA extraction. Following the selective-genotyping strategy originally proposed for this population, only a subset of these samples was selected for genotyping in order to improve the efficiency of detecting marker–trait associations while substantially reducing genotyping costs [27]. After genotype quality filtering, 606 individuals representing 124 accessions were retained for downstream genetic analyses.

Phenotypic observations were collected at the individual-plant level, whereas all downstream genetic analyses were conducted at the accession level. Because multiple plants were evaluated for each accession whenever available, accession effects were estimated using linear mixed models fitted by restricted maximum likelihood (REML), with plot included as a random effect to account for environmental variation among field plots and accession included as a random effect to estimate accession-level genetic effects. Accession-level best linear unbiased predictors (BLUPs) were extracted from these models and used as the phenotypic inputs for all downstream genome-wide analyses. Genotype data from selectively genotyped individuals were summarized as mean allele dosage per accession, resulting in one phenotypic value and one genotype vector for each accession included in the association analyses.

Source and identification of plant materials

The perennial accession Gypsum 9 N (4×), derived from *Sorghum halepense*, originated from a natural population located at the northern city limits of Gypsum, Kansas, USA, on the east side of Kansas Highway 4. The material was collected and subsequently developed within a breeding context by Dr. Stan Cox (The Land Institute, Salina, KS, USA).

Collection was conducted on publicly accessible land, and no permits or licenses were required under local regulations. No voucher specimen was deposited, as the material was incorporated directly into an ongoing breeding program.

Taxonomic identification of *Sorghum halepense* was performed by Dr. Stan Cox (The Land Institute, Salina, KS, USA) based on established morphological descriptors and breeding records maintained within the perennial sorghum breeding program.

The annual inbred KS105A (*Sorghum bicolor*) was obtained from the USDA Germplasm Resources Information Network (GRIN; accession PI 612408) and is maintained as part of publicly available sorghum germplasm resources.

All experimental research involving plant materials complied with institutional, national, and international guidelines. Field experiments were conducted in accordance with local regulations at the University of Georgia Coastal Plain Experiment Station, Tifton, GA, USA.

DNA extraction, library preparation, and sequencing

Leaf tissue samples from individual plants were flash-frozen at −80 °C and lyophilized for 48 h prior to DNA extraction. Genomic DNA was extracted from lyophilized tissue using a modified salt-extraction protocol following the rapid polysaccharide- and polyphenol-free method described by [28]. DNA concentrations were normalized to 20 ng μL^{−1} prior to library preparation.

Genotyping-by-sequencing (GBS) libraries were prepared using restriction enzyme-based complexity reduction. Libraries were constructed following the ApeKI-based GBS protocol originally described by [29], consistent with reduced-representation sequencing approaches widely used in sorghum and related cereal species. Fragment libraries were purified and size-selected prior to sequencing.

Sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA), generating paired-end FASTQ reads for downstream genotyping analyses.

Read processing, alignment, and SNP discovery

Sequence reads were processed using the TASSEL-GBS v5 pipeline [30]. Reads were aligned to the *Sorghum bicolor* reference genome BTx623 v3.1.1 [31] using BWA-MEM [32] implemented in BWA [33]. Variant discovery and initial quality filtering were performed using the Genome Analysis Toolkit (GATK; RRID: SCR_001876) [34].

SNP filtering, imputation, and data aggregation

Quality control was applied at both individual and marker levels. Individuals with >50% missing genotype data were excluded. SNPs were retained if they satisfied the following criteria: call rate $\geq 80\%$, minor allele frequency (MAF) ≥ 0.05 , and genotype quality score $Q \geq 20$. Monomorphic loci were removed.

Following plant-level genotype quality control, 43,318 SNPs across 606 individuals were retained. Genotype data from retained individuals were then summarized at the accession level as mean allele dosage across available genotyped plants within each accession. Remaining missing accession-level dosage values were imputed using the SNP-wise mean allele dosage.

After accession-level aggregation and imputation, one of the 43,318 QC-retained markers was invariant across accession-level dosage values and was therefore excluded before association testing because a zero-variance predictor has no estimable marker effect. The final accession-level genotype matrix used for GWAS consequently contained 43,317 variable SNPs. Thus, 43,318 refers to markers retained after plant-level genotype quality control, whereas 43,317 refers to the variable markers actually tested in the accession-level GWAS.

Reference genome and annotation provenance

Sequence alignment, variant discovery, and the physical coordinates used for GWAS were based on the *Sorghum bicolor* BTx623 v3.1.1 reference genome. To maintain a single coordinate framework throughout the final analysis, positional candidate-gene annotation was also performed using the BTx623 v3.1.1 structural gene annotation available through Phytozome v12 (Sbicolor_454_v3.1.1). No cross-version coordinate conversion was used in the final GWAS, signal-window definition, historical-QTL distance calculations, or positional candidate-gene catalogue. BTx623 v5.1 gene models considered during earlier exploratory work were not used to define the final candidate-gene catalogue because direct gene-model correspondence across annotation releases was not established.

Field site and experimental design

Field trials were conducted at the University of Georgia Coastal Plain Experiment Station in Tifton, Georgia, USA (31.5070° N, 83.5085° W). The experimental site is characterized by Tifton loamy sand soils.

The field experiment was established on 5 May 2022 using the available seed from each diploid F₄ accession. Individual accessions were planted in 3.7 m single-row plots with 0.90 m row spacing and approximately 0.30 m within-row plant spacing. Because the S3011 population originated through a partially fertile triploid intermediate, seed production varied substantially among F₄ families. Consequently, the realized field experiment was unbalanced, with variation among accessions in both the number of field plots established and the number of plants per plot. Some accessions produced sufficient seed to be represented in multiple plots, whereas others were represented in a single plot with fewer established plants. Plot identifiers were unique across the experiment and served as the experimental units in subsequent mixed-model analyses.

Weed control consisted of pre-emergence application of S-metolachlor followed by manual weeding as required. No supplemental irrigation was applied during the growing season. During the 2022–2023 overwintering period, minimum air temperatures reached -9.2°C , with multiple freeze events below 0°C . Leaf tissue was collected from every surviving plant for DNA extraction prior to selective genotyping.

Phenotypic measurements

Rhizome development

Rhizome expression was scored in late fall 2022 using an ordinal scale (0, 0.5, 1, 2, 3) reflecting increasing levels of rhizome-derived tiller emergence.

Although ordinal, this trait was analyzed as continuous to approximate an underlying biological gradient (liability scale). This approach is commonly applied in quantitative genetic analyses of ordered traits.

Overwinter regrowth

Regrowth was recorded in spring 2023 as a binary trait (0=no regrowth; 1=regrowth observed).

Although binary, regrowth was analyzed using linear mixed models to enable BLUP estimation and compatibility with downstream association analysis. This approach approximates a liability-threshold model and is widely used in quantitative genetics [35]. As a sensitivity consideration, results should be interpreted with awareness of potential deviations from logistic models.

Mixed models and BLUP estimation

Phenotypic observations were analyzed at the individual-plant level using linear mixed models fitted by restricted maximum likelihood (REML). For each trait, the following model was fitted:

$$y_{ijk} = \mu + P_i + A_j + \varepsilon_{ijk},$$

where y_{ijk} is the phenotype of an individual plant, μ is the overall mean, P_i is the random effect of plot, A_j is the random effect of accession, and ε_{ijk} is the residual error. Plot identifiers were unique across the experiment, with each plot representing a single accession planting. Accordingly, plot was treated as the experimental unit and modeled as a random effect to account for environmental heterogeneity among field plots.

Although the original experimental design followed a randomized complete block design with two nominal replicates, variation in seed availability, field establishment, and plant survival resulted in an unbalanced realized experiment with unequal representation of accessions across plots. Consequently, the mixed-effects model was specified to reflect the realized experimental design by treating plot as the experimental unit rather than fitting a nominal replicate effect. The accession random effect was used to estimate accession-level genetic effects across all available plant observations.

Accession-level best linear unbiased predictors (BLUPs) were extracted from the accession random effects and used as the phenotypic inputs for all downstream genome-wide analyses. Thus, both the exploratory genome-wide association analysis (GWAS) and the complementary marker-based genome scan were conducted at the accession level rather than by treating individual plants as independent genetic units. Genotype data were likewise summarized as mean allele dosage per accession prior to association analysis. Consequently, individual plants contributed to the estimation of accession-level genetic effects through the mixed model but were not treated as independent genetic observations in downstream association analyses, thereby avoiding pseudo-replication.

Population structure and kinship assessment

Population structure was evaluated using principal component analysis (PCA) [36, 37]. The first three principal components (PC1–PC3) were included as fixed covariates in all association models.

Because the S3011 panel contains family structure and relatedness, association models with and without a kinship term were evaluated diagnostically. Inclusion of kinship broadly attenuated marker-association statistics and produced only modest changes in genomic-inflation diagnostics relative to models containing PC1–PC3 alone. The PCA-adjusted model was therefore retained as the primary exploratory association framework, while the kinship-adjusted analysis was treated as a sensitivity assessment rather than as independent evidence. Genomic inflation factors for the primary PCA-adjusted analyses ranged from $\lambda = 1.17$ to 1.26, indicating modest residual inflation. Accordingly, all marker associations are interpreted conservatively as exploratory, and residual relatedness is retained as a limitation of the present study.

Genome-wide association analysis

Genome-wide association analysis (GWAS) was conducted at the accession level using accession-level BLUPs as phenotypic values and mean allele dosage as the corresponding genotype representation. Association testing followed the linear model.

$$y = \mu + \beta_{\text{snp}}X + \sum_{k=1}^3 \beta_k PC_k + \varepsilon,$$

where y denotes the accession-level phenotype, μ is the intercept, x is the accession-level mean allele dosage at the SNP being tested, β_{snp} is the estimated additive allele-substitution effect, PC_k denotes the first three principal-component covariates, and ε is the residual error. This framework corresponds to a population-structure-adjusted linear association model [25, 26, 38].

Genome-wide significance was evaluated using a Bonferroni-adjusted threshold based on the 43,317 variable SNPs that entered the final accession-level association scan:

$$\text{The Bonferroni – adjusted threshold was } 0.05/43,317 = 1.15 \times 10^{-6},$$

corresponding to $-\log_{10}(P) \approx 5.94$. Benjamini–Hochberg false discovery rate (BH-FDR) correction was additionally applied at $q < 0.05$. A predefined suggestive threshold of $P < 1 \times 10^{-4}$ ($-\log_{10}(P) > 4$) was used to identify exploratory associations for subsequent characterization. Associations meeting only the suggestive threshold were not considered genome-wide significant and were interpreted as exploratory signals rather than confirmed QTL.

To distinguish localized multi-marker signals from isolated suggestive associations, SNPs exceeding the suggestive threshold were grouped descriptively according to physical proximity. Adjacent suggestive SNPs separated by ≤ 250 kb were assigned to the same localized cluster, whereas suggestive SNPs separated by greater distances were treated as spatially distinct signals. For each resulting signal or cluster, the SNP with the smallest P-value was designated as the representative lead marker. This clustering procedure was used to summarize the spatial organization of suggestive associations and did not constitute linkage-based QTL interval mapping.

For each representative SNP, the proportion of phenotypic variance explained (PVE) was calculated as.

$$PVE(\%) = 100 \times \beta^2 \text{Var}(X) / \text{Var}(y),$$

where β is the estimated SNP effect from the population-structure-adjusted GWAS, $\text{Var}(X)$ is the variance of the accession-level marker-dosage vector, and $\text{Var}(y)$ is the variance of the corresponding accession-level phenotype. These PVE values describe the marker-level variance contribution under the GWAS effect estimate and should

not be interpreted as independently estimated QTL variance components. Given the modest sample size, low trait heritability, and exploratory significance of the associations, PVE estimates were interpreted cautiously.

Definition and downstream evaluation of suggestive genomic signals

All spatially distinct signals defined from SNPs exceeding the predefined suggestive GWAS threshold were retained in the primary summary of exploratory associations. Subsequent analyses were used to evaluate the degree of complementary support available for these signals rather than to redefine suggestive GWAS associations as confirmed QTL.

Some complementary analyses had been developed within the original analytical workflow around a subset of candidate regions on chromosomes 1, 3, and 4. Accordingly, results from these analyses were retained as additional characterization of selected exploratory signals and were not interpreted as evidence that chromosomes 1, 3, and 4 represented the complete set of suggestive GWAS associations. Differences in the amount or type of downstream evidence available among signals were therefore treated as differences in analytical characterization rather than as a biological criterion for excluding other suggestive signals.

Descriptive marker-scan concordance analysis

A second marker-level scan was used to describe cross-method positional concordance with the primary population-structure-adjusted GWAS. At each retained SNP, accession-level BLUPs were regressed on accession-level mean allele dosage by ordinary least squares. This analysis was not treated as classical linkage-based QTL mapping because a complete linkage map was not available, nor was it treated as an independent association experiment.

For each trait, Manhattan and Q–Q plots summarized the marker scan. Chromosome-wise marker-scan peaks were compared with the seven representative rhizome-number GWAS markers. Exact SNP matches were recorded directly, and physical distance to the chromosome-wise marker-scan peak was calculated for each GWAS lead marker. A ± 250 -kb window defined proximal concordance.

Because both analyses used the same S3011 accessions and genotype dataset, exact or proximal agreement was interpreted strictly as within-dataset cross-method concordance. It was used to describe whether an exploratory GWAS position was also prioritized by an alternative marker-level model, not to validate, replicate, or exclude a GWAS signal.

All seven exploratory rhizome-number GWAS signals were characterized using the same core secondary analyses: high–low allele-dosage contrasts, local lead-to-neighbor linkage disequilibrium, historical positional correspondence, positional candidate-gene annotation, and descriptive marker-scan concordance. Additional dosage-response and regional quality-control or sensitivity analyses remained restricted to selected regions from the earlier analytical workflow.

GWAS-effect-oriented high–low allele-dosage contrast analysis

Phenotype-associated allele-dosage differentiation was evaluated after orienting genotype dosage to the GWAS effect direction, such that increasing dosage corresponded to the allele associated with increasing rhizome number. This orientation standardizes the direction of the contrast but does not assign alleles to either parental species. Accordingly, we use the term high–low allele-dosage contrast throughout and do not interpret this analysis as ancestry inference.

The analysis included 124 accessions with aligned genotype data and rhizome-number BLUPs. Accessions were ranked by BLUP and divided into lower and upper quartiles (31 low-rhizome and 31 high-rhizome accessions). For each marker, the contrast was the mean GWAS-effect-oriented dosage in the high-rhizome group

minus the corresponding mean in the low-rhizome group. Positive values therefore indicate higher dosage of the rhizome-increasing allele in the high-rhizome group.

Marker-level contrasts were summarized in non-overlapping 500-kb windows, yielding 1,248 informative windows genome-wide. Each window-level contrast was ranked within its chromosome and across the genome by absolute magnitude. For each of the seven representative rhizome-number GWAS signals, we recorded the contrast for the 500-kb window containing the lead SNP, its chromosome-relative percentile, and its genome-wide absolute-contrast rank.

These quantities are phenotype-oriented dosage contrasts, not local-ancestry estimates. Phased haplotypes and sufficiently informative parental reference panels were unavailable, and GWAS-effect orientation does not identify parental origin. Accordingly, no contrast is interpreted as direct evidence of *S. halepense* ancestry or segment-level introgression.

Historical QTL correspondence analysis

The spatial correspondence between the seven exploratory rhizome-number GWAS signals identified in the S3011 population and previously reported rhizomatousness loci was evaluated using QTL identified in independent interspecific *Sorghum bicolor* × *S. halepense* mapping populations [39]. The comparison focused on published rhizomatousness loci reported for the rhizome-groups (RG) phenotype because these provided the most directly relevant historical genomic reference for comparison with the S3011 rhizome-number associations.

For each of the seven spatially distinct suggestive GWAS signals, the position of the representative S3011 lead SNP was compared with the corresponding published historical rhizomatousness locus on the same chromosome. For historical QTL identified by interval mapping, physical correspondence was evaluated relative to the published QTL interval boundaries reported by [39]. When an S3011 lead SNP fell within a published historical QTL interval, the minimum physical distance was defined as zero. When the lead SNP fell outside the interval, the minimum distance was calculated to the nearest published interval boundary.

The chromosome 3 comparison was treated separately because the historical locus *qRG.3.H6.1* was detected for the rhizome-groups (RG) phenotype by the complementary single-marker analysis of [39], rather than by their interval-mapping analysis. In that study, RG denotes categorical rhizomatousness derived from rhizome occurrence and is distinct from winter survival, which was analyzed separately as ISur. Accordingly, *qRG.3.H6.1* is treated here as a historical rhizomatousness locus and not as an overwinter-regrowth or survival QTL. Accordingly, the S3011 chromosome 3 lead SNP was compared with the published peak marker (*S3_58273458*) and the reported left–right flanking-marker span associated with *qRG.3.H6.1*. Because this span was derived from the single-marker analysis, it was not interpreted as equivalent to a QTL confidence interval obtained by interval mapping. Distance for the chromosome 3 comparison was therefore reported relative to the published peak marker, together with a descriptive assessment of whether the S3011 signal fell within the broader reported marker span.

Historical correspondence was summarized descriptively according to the physical relationship between each S3011 lead SNP and the relevant published locus. For interval-mapped QTL, relationships were classified as positional correspondence when the S3011 lead SNP occurred within the published interval, close regional proximity when it occurred near an interval boundary, or broader/little positional correspondence at increasing physical distances. The chromosome 3 relationship was reported separately as correspondence with a single-marker-defined historical locus. These categories were descriptive and were not treated as statistical tests of replication.

Because the S3011 population and the historical mapping populations differ in population structure, marker density, analytical framework, and mapping resolution, physical correspondence was interpreted as contextual evidence rather than evidence that an S3011 association and a historical QTL represent the same underlying causal locus. This qualification is particularly important for broad historical intervals, for which physical inclusion of an S3011 signal provides limited resolution regarding locus identity. Historical QTL correspondence was therefore used as a complementary line of evidence for biological interpretation and prioritization of exploratory signals and not as independent validation, replication, or confirmation of the S3011 GWAS associations.

Local linkage disequilibrium analysis

Local linkage disequilibrium (LD) was evaluated for each of the seven spatially distinct exploratory rhizome-number GWAS signals. For each signal, the representative SNP was used as the focal marker, and neighboring SNPs located within ± 250 kb of the representative SNP were identified. The representative SNP itself was excluded from the set of neighboring markers.

Pairwise LD between each representative SNP and each neighboring marker was quantified as the squared Pearson correlation (r^2) between their genotype-dosage vectors across accessions with estimable dosage values [40, 41]. Thus, the analysis characterized lead-to-neighbor LD within each 500-kb signal-centered window rather than calculating all pairwise marker comparisons or defining formal haplotype blocks.

For each of the seven signals, local LD was summarized by the number of neighboring markers in the ± 250 -kb window, the median and maximum lead-to-neighbor r^2 , and the numbers of neighboring markers with $r^2 \geq 0.20$ and $r^2 \geq 0.50$. The neighboring marker showing the maximum r^2 with each representative SNP was also identified. These thresholds were used descriptively to compare the extent of local marker correlation among the seven exploratory signals and were not used to define haplotype blocks, determine GWAS signal retention, or infer causal variants.

Because none of the rhizome-number associations satisfied genome-wide multiple-testing-adjusted significance criteria, the LD analysis was interpreted as complementary regional characterization of the seven exploratory signals rather than evidence of confirmed QTL or causal loci.

Candidate gene identification

Positional genes were catalogued uniformly for all seven exploratory rhizome-number signals by extracting BTx623 v3.1.1 gene models (Phytozome v12; Sbicolor_454_v3.1.1) overlapping the ± 250 -kb interval around each representative SNP. Signal-level gene counts, the nearest annotated gene, and a deduplicated seven-signal catalogue were recorded; genes shared by the overlapping chromosome 1 intervals were counted once in the deduplicated total.

The catalogue is positional rather than functional. No gene was prioritized on proximity alone, and earlier BTx623 v5.1 candidates were not carried forward because direct gene-model identity across annotation releases was not established. The resulting list therefore defines hypotheses for subsequent functional work rather than causal candidates.

Statistical software and reproducibility

All statistical and genomic analyses were conducted in R using the packages lme4, data.table, ggplot2, emmeans, and additional packages documented within the reproducibility scripts. The analytical workflow included phenotype processing, mixed-model estimation, genome-wide association analysis (GWAS), complementary marker-based genome scans, linkage disequilibrium (LD) analysis, GWAS-effect-oriented high–low allele-dosage contrast analysis, signal-centered and candidate-gene analyses, historical QTL correspondence analyses, quality-control diagnostics, and generation of supplementary outputs.

The final curated datasets, analysis scripts, supplementary materials, and reproducibility documentation supporting this study have been permanently archived in Zenodo Version V4 and are publicly available at:

<https://doi.org/10.5281/zenodo.22074040>.

The archived repository includes processed phenotype and genotype data, analysis scripts, locked analytical outputs, supplementary tables and figures, audit manifests, provenance records, and a README describing the repository structure and reproducibility workflow. The deposited analysis package was verified for file integrity, R-script syntax, dependency resolution, portability, and cold-start reproducibility.

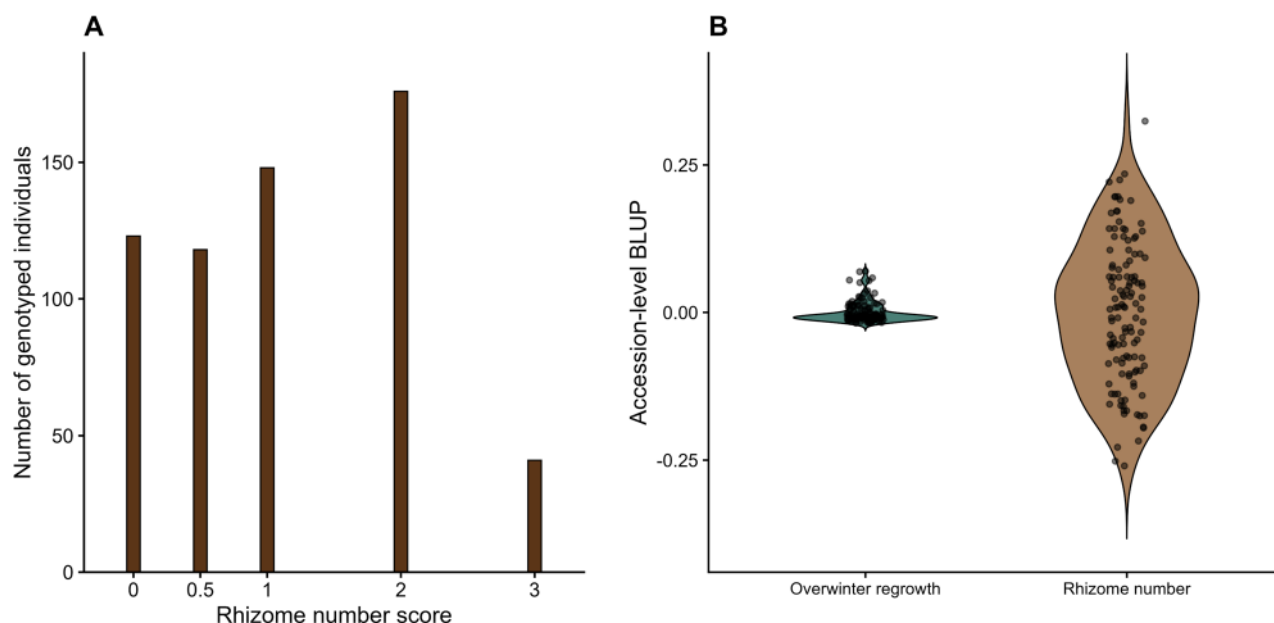


Fig. 1 Phenotypic distributions in the final S3011 analytical panel. **A** Distribution of rhizome number scores among the 606 genotyped individuals retained after genotype quality control and contributing to the 124-accession analytical panel used for genome-wide association analysis. **B** Accession-level best linear unbiased predictors (BLUPs) for overwinter regrowth and rhizome number across the 124 genotyped accessions used for genome-wide association analysis. Points represent individual accessions and violin plots summarize the distribution of BLUP values

Table 1 Phenotypic summary statistics and broad-sense heritability in the S3011 diploid-extraction panel

Trait	Mean	Range	SD	H ² (±SE)
Rhizome number	1.10	0.0–3.0	0.88	0.07±0.02
Overwinter regrowth	0.08	0–1	0.27	0.04±0.01

Rhizome number and overwinter regrowth are reported with broad-sense heritability (H²) estimates

Means, ranges, and standard deviations summarize the observed phenotypic distributions. Broad-sense heritability estimates are reported with their corresponding standard errors

Results

Phenotypic variation for rhizome development and overwinter regrowth

The S3011-derived diploid F₄ panel exhibited substantial phenotypic variation for key components of perennality (Fig. 1). Phenotypic observations were analyzed using mixed models to derive accession-level best linear unbiased predictors (BLUPs).

Rhizome number ranged from 0 to 3 (mean=1.10) and displayed a right-skewed distribution. Accession-level BLUPs spanned a broad continuum, consistent with segregation of introgressed *S. halepense* genomic segments within a diploid *S. bicolor* background.

Overwinter regrowth was observed in 7.8% of plants under winter conditions reaching a minimum temperature of −9.2 °C. The low frequency of regrowth limited statistical power for association detection.

Rhizome number and overwinter regrowth were weakly correlated ($r=0.19$), indicating that rhizome initiation alone is not sufficient to confer overwinter survival.

Broad-sense heritability was low for rhizome number ($H^2 = 0.07 \pm 0.02$ SE) and very low for overwinter regrowth ($H^2 = 0.04 \pm 0.01$ SE) (Table 1; Supplementary Table S1). These low estimates are consistent with

strong environmental influence and evaluation in a single environment. Although heritability was low, estimates remained non-zero, indicating detectable genetic variance.

Population structure of the diploid mapping panel

Principal component analysis (PCA) revealed modest genomic structure (Supplementary Fig. S3). The first three principal components each explained 5–7% of total genetic variation.

No discrete clustering was observed, indicating continuous genomic variation rather than sharply stratified subpopulations. Genomic inflation factors for the primary PCA-adjusted analyses were modest ($\lambda_{\text{Rhiz}}=1.26$; $\lambda_{\text{Regrowth}}=1.17$), consistent with modest residual inflation. Diagnostic comparison of models with and without kinship showed that the kinship term broadly attenuated marker-association statistics while producing only modest changes in genomic-inflation diagnostics. The PCA-adjusted model was retained as the primary exploratory framework, the kinship-adjusted analysis was treated as a sensitivity assessment, and residual relatedness was retained as a limitation.

Genome-wide association analysis identifies seven spatially distinct suggestive signals for rhizome development

Genome-wide association analyses were conducted using accession-level BLUPs and mean allele dosage. For rhizome number, 13 SNPs exceeded the predefined suggestive threshold ($P < 1 \times 10^{-4}$), distributed across chromosomes 1, 2, 3, 4, and 10 (Fig. 2; Supplementary Fig. S1).

Using an adjacent-marker gap of ≤ 250 kb to describe localized clustering, these 13 SNPs resolved into seven spatially distinct suggestive signals (Table 2). Chromosome 1 contained eight suggestive SNPs distributed across two localized multi-marker clusters, whereas the chromosome 2, chromosome 3, and chromosome 10 signals were represented by single suggestive SNPs. Chromosome 4 contained two spatially separated singleton signals (Supplementary Table S2).

The seven representative lead markers were S1_62967991 for chromosome 1 cluster 1, S1_63359080 for chromosome 1 cluster 2, S2_23212567 for the chromosome 2 singleton, S3_62120700 for the chromosome 3 singleton, S4_490163 for the chromosome 4 0.49-Mb singleton, S4_8603936 for the chromosome 4 8.60-Mb singleton, and S10_52748375 for the chromosome 10 singleton (Table 2).

No SNP exceeded the Bonferroni-adjusted genome-wide significance threshold ($P \leq 1.15 \times 10^{-6}$, based on 43,317 tested SNPs), and no association remained significant after Benjamini–Hochberg false discovery rate (BH-FDR) correction ($q < 0.05$). Accordingly, the seven signals are interpreted as exploratory associations rather than confirmed QTL.

The representative lead markers had estimated PVE values ranging from 13.32% to 20.35% under the marker-level GWAS calculation (Table 2). Given the modest sample size, low heritability of rhizome number, and absence of associations surviving multiple-testing correction, these values are interpreted as descriptive measures of marker-level effect magnitude rather than precise estimates of QTL variance.

All seven exploratory rhizome-number GWAS signals were characterized using a common set of complementary analyses, including GWAS-effect-oriented high–low allele-dosage contrasts, local lead-to-neighbor linkage disequilibrium, historical-QTL correspondence, positional candidate-gene annotation, and comparison with the complementary marker-based genome scan. Additional downstream analyses were not applied uniformly across all seven signals. In particular, signal-specific allele dosage–response analyses and regional quality-control or sensitivity assessments were restricted to selected regions examined in the earlier analytical workflow, including the chromosome 4 leave-one-accession-out analysis. These differences in analytical depth do not alter the primary GWAS result of seven spatially distinct suggestive signals across chromosomes 1, 2, 3, 4, and 10.

Additional quality-control analyses conducted for the selected chromosome 4 region included evaluation of local linkage disequilibrium, allele-dosage distributions, and leave-one-accession-out sensitivity analyses. In the

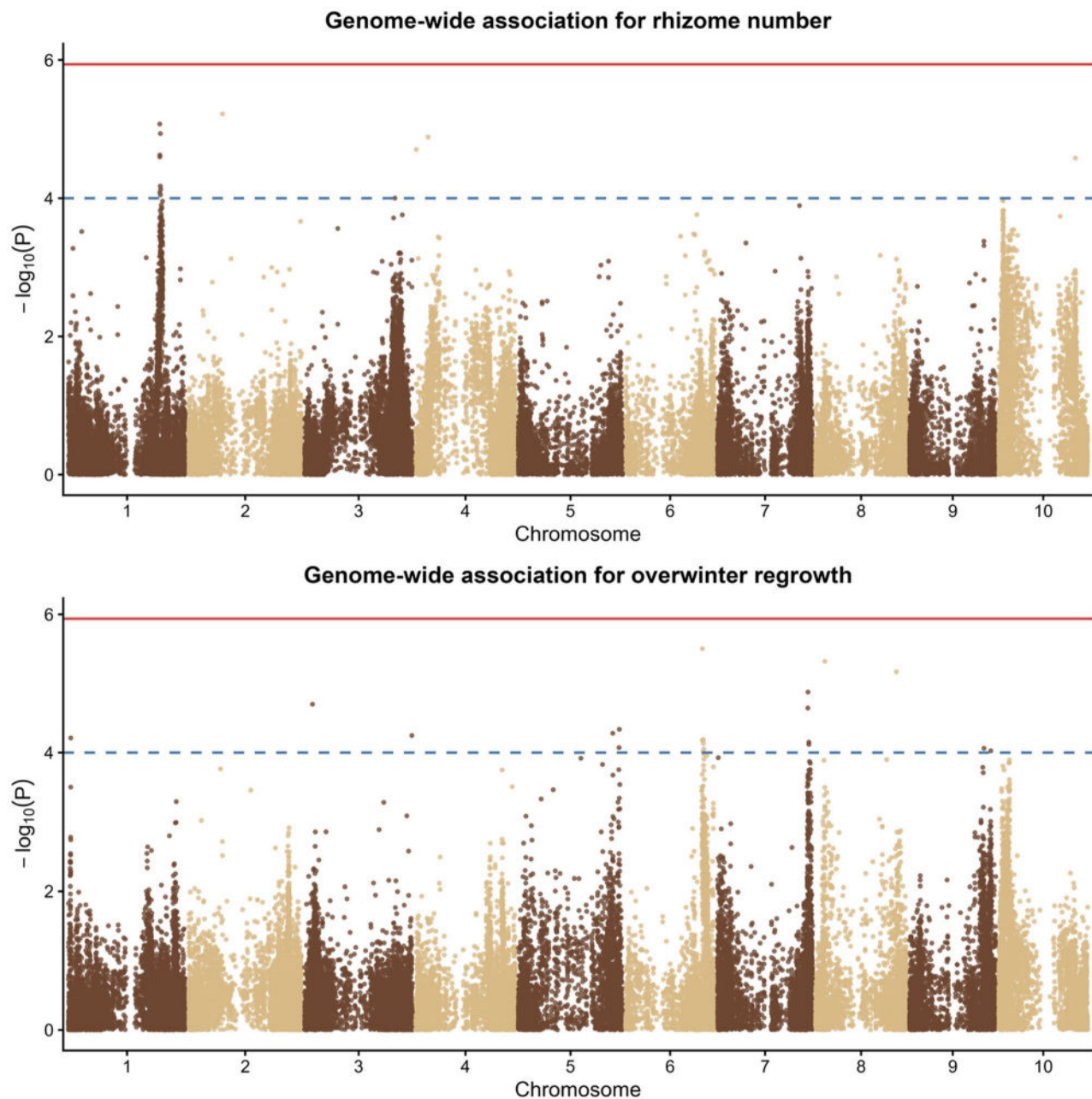


Fig. 2 Genome-wide association profiles for rhizome number and overwinter regrowth in the S3011 diploid-extraction panel. Manhattan plots show association results for rhizome number (top) and overwinter regrowth (bottom) across chromosomes 1–10. The dashed blue line denotes the predefined suggestive threshold ($P < 1 \times 10^{-4}$; $-\log_{10}(P) > 4$), and the solid red line denotes the Bonferroni-adjusted genome-wide significance threshold ($P \leq 1.15 \times 10^{-6}$; $-\log_{10}(P) \approx 5.94$), based on 43,317 tested SNPs. For rhizome number, 13 SNPs exceeded the suggestive threshold on chromosomes 1, 2, 3, 4, and 10 and resolved into seven spatially distinct suggestive signals: two localized multi-marker clusters on chromosome 1, singleton signals on chromosomes 2, 3, and 10, and two spatially separated singleton signals on chromosome 4. Representative SNPs for these seven signals are reported in Table 2. Overwinter-regrowth associations were more dispersed and did not show the localized multi-marker structure observed for rhizome number. No association for either trait exceeded the Bonferroni-adjusted genome-wide significance threshold. Rhizome-number associations also did not remain significant after Benjamini–Hochberg false discovery rate correction. Accordingly, the associations are interpreted as exploratory signals rather than confirmed QTL

Table 2 Representative lead markers for seven exploratory rhizome-number GWAS signals in the S3011 panel

Marker	Chr	Position (bp)	β	<i>P</i> -value	PVE (%)	Signal type
S1_62967991	1	62,967,991	0.101	8.482505×10^{-6}	18.13	Multi-SNP cluster
S1_63359080	1	63,359,080	0.079	1.163571×10^{-5}	15.77	Multi-SNP cluster
S2_23212567	2	23,212,567	0.137	6.066368×10^{-6}	16.77	Singleton
S3_62120700	3	62,120,700	0.146	9.990040×10^{-5}	13.32	Singleton
S4_490163	4	490,163	0.124	1.977008×10^{-5}	17.54	Singleton
S4_8603936	4	8,603,936	0.143	1.307686×10^{-5}	20.35	Singleton
S10_52748375	10	52,748,375	0.127	2.622930×10^{-5}	17.16	Singleton

The signals were spatially distinct and identified in the S3011 diploid-extraction panel. Representative lead markers for the seven spatially distinct exploratory rhizome-number GWAS signals in the S3011 diploid-extraction panel. Thirteen SNPs exceeded the predefined suggestive threshold ($P < 1 \times 10^{-4}$) and resolved into two multi-SNP clusters on chromosome 1 and singleton signals on chromosomes 2, 3, 4, and 10. The two chromosome 4 singleton signals were spatially distinct. Note: Spatial clustering was defined descriptively using an adjacent-marker gap of ≤ 250 kb. The chromosome 1 signals represent the most strongly associated SNP within each multi-SNP cluster; the remaining signals are singletons. Beta (β) denotes the estimated additive allele-substitution effect on rhizome-number BLUPs. PVE was calculated as $100 \times \beta^2 \text{Var}(X) / \text{Var}(y)$ using the PC-adjusted GWAS effect estimate, accession-level marker-dosage variance, and variance of accession-level rhizome-number BLUPs. No SNP exceeded the Bonferroni-adjusted genome-wide significance threshold ($P \leq 1.15 \times 10^{-6}$; 43,317 SNPs tested), and no association remained significant after Benjamini–Hochberg false discovery rate correction ($q < 0.05$). Accordingly, all seven signals are interpreted as exploratory rather than confirmed QTL. High–low allele-dosage contrasts, local LD, historical-QTL correspondence, positional candidate-gene annotation, and descriptive marker-scan concordance were evaluated across all seven signals. Additional dosage-response and regional quality-control or sensitivity analyses were restricted to selected regions

leave-one-accession-out analysis, the GWAS model was sequentially refitted after removal of each accession, and both chromosome 4 lead-SNP associations remained below the predefined suggestive threshold ($P < 1 \times 10^{-4}$) after removal of the most influential accession. These results indicate that the evaluated chromosome 4 associations were not attributable to a single influential genotype (Supplementary Fig. S2; Supplementary Table S10).

Overwinter regrowth exhibits a dispersed association profile

The genome-wide association profile for overwinter regrowth contrasted with that observed for rhizome number (Fig. 2; Supplementary Fig. S1). No overwinter-regrowth association exceeded the Bonferroni-adjusted genome-wide significance threshold or remained significant after BH-FDR correction.

Although individual SNPs exceeded the predefined suggestive threshold, these associations were comparatively dispersed across the genome and did not exhibit the localized multi-marker structure observed for rhizome number. Consequently, the current analysis did not resolve a robust discrete genomic region for overwinter regrowth.

The dispersed association profile is consistent with the low frequency of overwinter regrowth and the limited statistical resolution available for this binary trait in the current population. Thus, whereas rhizome number yielded seven spatially distinct exploratory signals, overwinter regrowth did not show comparable localized genomic architecture under the present experimental conditions.

Complementary marker-based genome scan supports selected rhizome-associated signals

To evaluate whether positions identified by the exploratory GWAS were also prioritized using an alternative analytical framework, we compared the seven representative rhizome-number GWAS signals with results from the complementary marker-based genome scan (Supplementary Figs. S7–S10; Supplementary Table S9).

Exact SNP concordance between the two analytical approaches was observed for chromosome 1 cluster 2 at S1_63359080 and for the chromosome 3 signal at S3_62120700. In both cases, the representative GWAS lead SNP was also the corresponding marker-scan peak, giving a physical distance of 0 bp between the two results.

The remaining representative GWAS signals did not show equivalent colocalization with their corresponding chromosome-wise marker-scan peaks under the predefined ± 250 -kb criterion. For chromosome 1 cluster 1, the

GWAS lead marker S1_62967991 was 391,089 bp from the chromosome 1 marker-scan peak S1_63359080 and therefore fell outside the defined colocalization window. The chromosome 2, chromosome 4, and chromosome 10 representative signals likewise did not satisfy the defined colocalization criterion.

These results demonstrate heterogeneous cross-method correspondence among the exploratory GWAS signals. Exact concordance for chromosome 1 cluster 2 and chromosome 3 provides within-dataset positional concordance for those positions, whereas absence of concordance for the other signals does not invalidate their suggestive GWAS associations. Because the GWAS and marker-based genome scan were conducted using the same population and genotype dataset, this concordance is interpreted as cross-method concordance rather than independent replication or validation of QTL.

For overwinter regrowth, the complementary marker-based scan showed three exact marker-level matches with suggestive GWAS markers—S3_5435691, S6_52895314, and S9_55930253—but the broader association profile remained dispersed and lacked the localized multi-marker regional structure observed for rhizome number (Supplementary Figs. S9–S10; Supplementary Table S9). Because both analyses used the same S3011 dataset, these matches represent within-dataset cross-method concordance rather than independent validation. Thus, neither analytical framework resolved a robust discrete overwinter-regrowth region under the current experimental conditions.

High–low allele-dosage contrasts differ among the seven exploratory rhizome-associated signals

The GWAS-effect-oriented high–low allele-dosage analysis revealed substantial heterogeneity among the seven spatially distinct rhizome-number GWAS signals. Using beta-oriented allele dosage, the high-rhizome minus low-rhizome contrast was evaluated for the 500-kb window containing the representative SNP of each signal and compared with chromosome-specific and genome-wide background distributions (Fig. 3; Supplementary Fig. S5; Supplementary Tables S4–S5).

The strongest high–low allele-dosage differentiation occurred at the two chromosome 1 signals. The window containing chromosome 1 cluster 2, represented by S1_63359080, showed a high-rhizome minus low-rhizome dosage contrast of +0.3364 and ranked at the 100.0th percentile of chromosome 1 windows. This window had the second-highest absolute contrast among the 1,248 informative 500-kb windows evaluated genome-wide. Chromosome 1 cluster 1, represented by S1_62967991, showed a similarly strong positive contrast of +0.3002, corresponding to the 99.3rd chromosome-relative percentile and the fifth-highest absolute contrast genome-wide. Thus, both chromosome 1 signals occurred within windows showing exceptional allele-dosage differentiation between the high- and low-rhizome groups (Supplementary Tables S4–S5).

Intermediate high–low allele-dosage differentiation was observed for the chromosome 10, chromosome 4 0.49-Mb, and chromosome 2 signals. The chromosome 10 window containing S10_52748375 showed a contrast of +0.1416 and was at the 77.5th chromosome-relative percentile, ranking 169th genome-wide by absolute contrast. The chromosome 4 0.49-Mb signal represented by S4_490163 showed a contrast of +0.1266, corresponding to the 86.3rd chromosome-relative percentile and a genome-wide rank of 241. The chromosome 2 signal represented by S2_23212567 showed a contrast of +0.1134 and was at the 90.6th chromosome-relative percentile, with a genome-wide rank of 334. These signals therefore showed positive allele-dosage differentiation, although their contrasts were substantially less exceptional genome-wide than those observed for the two chromosome 1 signals (Supplementary Table S5).

In contrast, the chromosome 3 and chromosome 4 8.60-Mb signals were not exceptional relative to their respective chromosome backgrounds. The window containing S3_62120700 showed a positive contrast of +0.1123 but was at only the 46.4th chromosome-relative percentile and ranked 349th genome-wide. Similarly, the chromosome 4 8.60-Mb signal represented by S4_8603936 showed the smallest contrast among the seven signals (+0.0815), corresponding to the 45.2nd chromosome-relative percentile and a genome-wide rank of 642. Thus, although both windows showed positive high-rhizome minus low-rhizome dosage contrasts, neither was unusual relative to the genomic background on its chromosome (Supplementary Table S5).

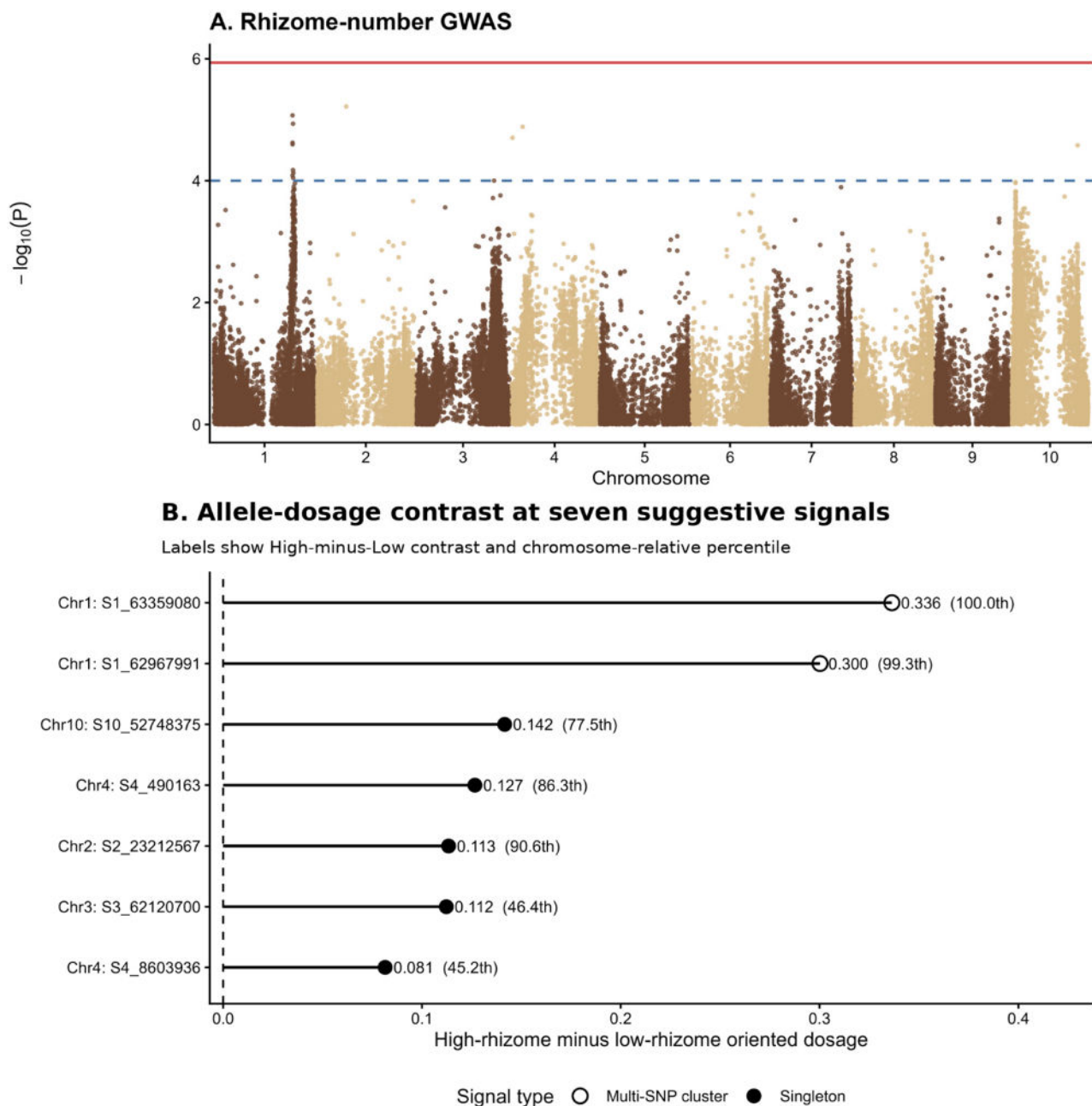


Fig. 3 Integration of rhizome-number GWAS signals with GWAS-effect-oriented high–low allele-dosage contrasts. **A** Genome-wide association profile for rhizome number across chromosomes 1–10. The dashed line denotes the predefined suggestive threshold ($P < 1 \times 10^{-4}$; $-\log_{10}(P) > 4$), and the solid line denotes the Bonferroni-adjusted genome-wide significance threshold ($P \leq 1.15 \times 10^{-6}$; $-\log_{10}(P) \approx 5.94$). No SNP exceeded the Bonferroni-adjusted threshold or remained significant after Benjamini–Hochberg false discovery rate correction. **B** Allele-dosage-derived contrasts for the seven spatially distinct suggestive rhizome-number signals reported in Table 2. Accessions in the upper and lower quartiles of rhizome-number BLUPs comprised 31 high-rhizome and 31 low-rhizome accessions, respectively. SNP dosages were oriented according to GWAS effect direction such that increasing dosage corresponded to the rhizome-increasing allele. High-minus-low dosage contrasts were summarized in non-overlapping 500-kb windows. For each signal, the plotted value represents the mean contrast for the window containing the representative SNP; labels report the contrast and its chromosome-relative percentile based on absolute contrasts among informative windows on the corresponding chromosome. These contrasts quantify phenotype-associated allele-dosage differentiation and should not be interpreted as direct evidence of *S. halepense* ancestry, phased local-ancestry assignments, or independent tests of statistical significance

Genome-wide allele-dosage-derived differences between the phenotypic classes were generally modest (Supplementary Fig. S4; Supplementary Table S4). Against this background, the exceptional contrasts observed for the two chromosome 1 signals indicate localized allele-dosage differentiation rather than a broad genome-wide shift between the high- and low-rhizome groups. The remaining five signals exhibited a continuum of chromosome-relative support, ranging from comparatively elevated contrasts on chromosomes 2, 4, and 10 to chromosome-unexceptional contrasts at the chromosome 3 and chromosome 4 8.60-Mb signals.

Overall, the seven-signal analysis demonstrates that phenotype-associated allele-dosage differentiation is heterogeneous among the exploratory GWAS signals. The two chromosome 1 signals showed the strongest localized differentiation, whereas the other signals showed moderate to comparatively unexceptional contrasts depending on their chromosomal background. Because phased haplotypes and parental reference panels sufficient for formal ancestry assignment were unavailable, and because dosage orientation was based on GWAS effect direction, these results should be interpreted as phenotype-associated allele-dosage contrasts rather than direct evidence of *S. halepense* ancestry or segment-level local ancestry assignment.

Historical QTL correspondence provides heterogeneous positional context for the exploratory GWAS signals

The seven spatially distinct rhizome-number GWAS signals were compared with previously reported rhizomatousness loci identified in an independent interspecific *S. bicolor* × *S. halepense* mapping population [39]. The degree of positional correspondence varied substantially among the seven S3011 signals (Table 3; Supplementary Tables S6 and S8), ranging from interval-level correspondence to close or broad regional proximity and little positional correspondence.

The strongest interval-level correspondence was observed for the chromosome 10 singleton signal. The S3011 lead marker S10_52748375, located at 52.748 Mb, fell within the published 5.7–56.7 Mb interval for *qRG.10B.H6.1*, giving a minimum distance of 0 Mb to the historical QTL interval. Because this published interval spans a large portion of chromosome 10, however, the physical overlap provides broad positional correspondence rather than evidence that the S3011 association and the historical QTL represent the same underlying locus.

The chromosome 3 singleton, S3_62120700 at 62.121 Mb, also showed regional correspondence with the previously reported rhizome-groups (RG) signal, *qRG.3.H6.1*. In [39], RG denotes a categorical rhizomatousness phenotype derived from the proportion of plants exhibiting rhizomes; it is distinct from the winter-survival phenotype analyzed separately as ISur. This historical rhizomatousness locus was identified by complementary single-marker analysis rather than interval mapping [39]. The S3011 marker fell within the broad genomic range represented by the historical single-marker analysis but was approximately 3.847 Mb from the reported historical peak marker. Accordingly, the chromosome 3 relationship is interpreted as broad single-marker regional correspondence and is not directly equivalent to overlap with an interval-mapped QTL.

Among the remaining signals, the chromosome 4 8.60-Mb singleton (S4_8603936) showed the closest proximity to an interval-mapped historical QTL. This marker was approximately 0.296 Mb below the published interval for *qRG.4B.H6.1*. The two chromosome 1 signals showed broader regional proximity to *qRG.1 C.H6.1*: S1_62967991 and S1_63359080 were approximately 3.932 and 3.541 Mb, respectively, below the published historical interval.

In contrast, the chromosome 4 0.49-Mb signal (S4_490163) was approximately 8.410 Mb from the represented *qRG.4B.H6.1* interval, indicating little positional correspondence. The chromosome 2 signal (S2_23212567) was still more distant from its represented historical chromosome 2 locus, with a minimum separation of approximately 43.187 Mb.

Collectively, these comparisons demonstrate heterogeneous positional correspondence between the exploratory S3011 rhizome-number signals and previously reported perenniality-associated loci. The chromosome 10 signal showed interval-level correspondence, the chromosome 3 signal showed broad correspondence with a historical single-marker rhizome-groups (RG) signal, and the chromosome 4 8.60-Mb and chromosome 1 signals

Table 3 Correspondence between S3011 exploratory rhizome-number signals and published rhizomatousness loci

S3011 signal	Lead SNP	Position (Mb)	Historical locus	Mapping method	Published position	Distance	Positional correspondence
Chr1 cluster 1	S1_62967991	62.968	qRG.1 C.H6.1	Interval mapping	Peak 66.947 Mb; interval 66.9–69.1 Mb	3.932 Mb	Broad regional proximity
Chr1 cluster 2	S1_63359080	63.359	qRG.1 C.H6.1	Interval mapping	Peak 66.947 Mb; interval 66.9–69.1 Mb	3.541 Mb	Broad regional proximity
Chr2 singleton	S2_23212567	23.213	qRG.2 C.H6.1	Interval mapping	Peak 66.432 Mb; interval 66.4–72.7 Mb	43.187 Mb	Little positional correspondence
Chr3 singleton	S3_62120700	62.121	qRG.3.H6.1	Single-marker analysis	Peak 58.273 Mb; flanking markers 3.357–69.917 Mb	3.847 Mb to peak	Within broad marker span ^a
Chr4 0.49-Mb singleton	S4_490163	0.490	qRG.4B.H6.1	Interval mapping	Peak 8.893 Mb; interval 8.9–53.5 Mb	8.410 Mb	Little positional correspondence
Chr4 8.60-Mb singleton	S4_8603936	8.604	qRG.4B.H6.1	Interval mapping	Peak 8.893 Mb; interval 8.9–53.5 Mb	0.296 Mb	Close regional proximity
Chr10 singleton	S10_52748375	52.748	qRG.10B.H6.1	Interval mapping	Peak 52.116 Mb; interval 5.7–56.7 Mb	0 Mb	Within published QTL interval ^b

Physical correspondence between the seven exploratory rhizome-number GWAS signals identified in the S3011 diploid panel and previously reported rhizomatousness loci from an independent interspecific Sorghum bicolor × S. halepense mapping population. Historical loci and physical positions are from [39]. For interval-mapped QTL, distance denotes the minimum physical distance from the representative S3011 lead SNP to the published QTL interval and is zero when the SNP falls within that interval. The chromosome 3 locus qRG.3.H6.1 was identified by single-marker analysis and is therefore compared with its published peak and flanking-marker span rather than an interval-mapping boundary. Note: RG denotes “rhizome groups,” the categorical rhizomatousness phenotype analyzed by [39]. Winter survival was analyzed separately as ISur; accordingly, qRG.3.H6.1 is treated here as a rhizomatousness locus rather than a regrowth or survival locus. The historical study also evaluated rhizome distance (RD), but did not report a rhizome-number (RN) phenotype. Previously attributed chromosome 3 loci qRD.3.H6.1 and qRN.3.H6.1 are therefore not included because they are not reported as such by [39]. ^a The flanking-marker span reported for qRG.3.H6.1 derives from single-marker analysis and is not equivalent to a QTL confidence interval from interval mapping. ^b The published interval for qRG.10B.H6.1 is exceptionally broad (5.7–56.7 Mb); physical inclusion of the S3011 chromosome 10 signal therefore indicates positional correspondence but provides limited evidence that the same underlying locus is involved. All correspondence classifications are descriptive and do not establish QTL replication, causal-locus identity, or functional validation

showed varying degrees of regional proximity. The chromosome 2 and chromosome 4 0.49-Mb signals showed comparatively little positional correspondence. These relationships provide historical genomic context for the exploratory GWAS results but do not constitute independent validation or establish identity of the underlying causal loci.

Local linkage disequilibrium differs among the seven exploratory rhizome-associated signals

Local linkage disequilibrium (LD) was evaluated within ± 250 kb of the representative SNP for each of the seven spatially distinct exploratory rhizome-number GWAS signals (Supplementary Fig. S6; Supplementary Table S3). Pairwise LD was quantified as squared Pearson correlation (r^2) between the representative SNP and neighboring markers within each signal-centered window. Local LD patterns differed substantially among the seven signals.

The most extensive lead-to-neighbor LD was observed for chromosome 1 cluster 2 and the chromosome 10 signal. For chromosome 1 cluster 2, represented by S1_63359080, 20 of 35 neighboring markers had $r^2 \geq 0.50$, with a median r^2 of 0.5408 and a maximum of 0.6621. For the chromosome 10 signal represented by S10_52748375, 16 of 61 neighboring markers had $r^2 \geq 0.50$, with a median r^2 of 0.4023 and the highest maximum r^2 observed among the seven signals (0.7064).

The chromosome 3 signal, represented by S3_62120700, was surrounded by a larger number of neighboring markers but showed less extensive lead-to-neighbor LD. Among 104 neighboring markers, 30 had $r^2 \geq 0.20$ and only one had $r^2 \geq 0.50$; the median and maximum r^2 values were 0.1475 and 0.6534, respectively. Chromosome 1 cluster 1 showed intermediate local LD, with 33 of 54 neighboring markers having $r^2 \geq 0.20$, although none reached $r^2 \geq 0.50$.

The remaining signals showed comparatively weak local LD with their neighboring markers. No neighboring marker reached $r^2 \geq 0.50$ for the chromosome 2, chromosome 4 0.49-Mb, or chromosome 4 8.60-Mb signals. Median r^2 values for these signals were 0.0009, 0.0007, and 0.0369, respectively. The chromosome 4 8.60-Mb signal had one neighboring marker with $r^2 \geq 0.20$, whereas none reached this threshold for the chromosome 2 or chromosome 4 0.49-Mb signals.

Overall, these results demonstrate heterogeneous local LD structure among the seven exploratory GWAS signals. Chromosome 1 cluster 2 and the chromosome 10 signal showed the clearest evidence of extended lead-to-neighbor LD, whereas several other signals occurred in regions with comparatively weak LD to surrounding markers. No signal contained a neighboring marker with $r^2 \geq 0.80$. These LD patterns provide regional characterization of the exploratory associations but do not define haplotype blocks, identify causal variants, or establish confirmed QTL.

Positional candidate-gene catalogue

The seven ± 250 -kb signal-centered intervals contained 15–62 annotated BTx623 v3.1.1 genes each and 286 unique positional genes after deduplication (Supplementary Table S7). The two chromosome 1 intervals overlapped by 108,912 bp and shared eight genes. Nearest annotated genes were 0.103–13.4 kb from the corresponding representative SNP.

Because inclusion was based solely on physical overlap, this catalogue is not a functional ranking. Gene count, proximity to the lead SNP, and annotation category were not treated as evidence of causality or as additional statistical support for a GWAS signal; Table S7 is intended as a searchable positional resource for subsequent functional investigation.

Multilocus allele-dosage indices are more strongly associated with rhizome number than with overwinter regrowth

Multilocus allele-dosage indices constructed from the exploratory rhizome-number GWAS signals showed substantially stronger relationships with rhizome number than with overwinter regrowth (Supplementary Table S11). For the index based on the seven representative SNPs, the Spearman correlation with rhizome number was $\rho=0.486$ ($N=123$, $P=1.17 \times 10^{-8}$), compared with $\rho=0.183$ ($N=123$, $P=0.0428$) for overwinter regrowth. When the two chromosome 1 signals were combined before constructing a six-region index, the corresponding correlations were $\rho=0.467$ ($N=124$, $P=4.48 \times 10^{-8}$) for rhizome number and $\rho=0.176$ ($N=124$, $P=0.0509$) for overwinter regrowth.

The previously defined historical chromosome 3/chromosome 4 dosage index was retained as a comparator only. Its correlation with rhizome number was $\rho=0.370$ ($N=124$, $P=2.29 \times 10^{-5}$), whereas its correlation with overwinter regrowth was $\rho=0.104$ ($N=124$, $P=0.252$) (Supplementary Table S11).

Because the seven representative SNPs and six spatial signal units were derived from rhizome-number associations in this same S3011 panel, these analyses quantify within-sample multilocus allelic accumulation. They therefore provide a descriptive comparison between perenniality traits and are not an independent validation of the rhizome-number GWAS signals.

Identification of families carrying prioritized allelic combinations

Integration of allele-dosage patterns and accession-level BLUPs within the selected candidate regions identified several F₄ families carrying elevated allele dosage at multiple evaluated markers.

Four families—A1D220-S2, A1D179-S4, A1D1-AA6-R1-S4, and A1D119-T6—showed elevated allele dosage across multiple evaluated candidate regions together with rhizome BLUPs above the population mean.

These families represent priorities for follow-up evaluation of whether the observed allele-dosage combinations and associated rhizome phenotypes are reproducible in subsequent breeding populations and environments. Their identification is intended as a breeding-oriented prioritization rather than evidence that the evaluated marker combinations are causal.

Summary of genetic architecture

The primary exploratory GWAS for rhizome number identified 13 SNPs exceeding the predefined suggestive threshold, which resolved into seven spatially distinct signals across chromosomes 1, 2, 3, 4, and 10. No association exceeded the Bonferroni-adjusted genome-wide significance threshold or remained significant after BH-FDR correction; consequently, all seven signals are interpreted as exploratory associations rather than confirmed QTL.

Secondary genomic analyses revealed heterogeneous regional patterns among rhizome-associated signals. The marker-based genome scan showed exact SNP concordance with chromosome 1 cluster 2 and the chromosome 3 signal, whereas equivalent colocalization was not observed for the remaining representative rhizome-number GWAS signals under the predefined criterion. The high–low allele-dosage analysis showed the strongest localized differentiation at chromosome 1 cluster 2 (S1_63359080; +0.3364, 100.0th chromosome-relative percentile), with chromosome 1 cluster 1 (S1_62967991; +0.3002, 99.3rd percentile) providing a second strong chromosome 1 contrast. High–low allele-dosage contrasts, local lead-to-neighbor LD, historical-QTL correspondence, positional candidate-gene annotation, and marker-scan comparison were evaluated across all seven signals; additional allele dosage–response and regional quality-control or sensitivity analyses were restricted to selected regions. Multilocus indices based on the seven representative SNPs or six spatial signal units were more strongly associated with rhizome number ($\rho=0.486$ and 0.467 , respectively) than with overwinter regrowth ($\rho=0.183$ and 0.176 , respectively), providing a within-sample descriptive contrast rather than independent validation.

Overwinter regrowth exhibited a more dispersed suggestive association profile without the localized multi-marker structure observed for rhizome number. Nevertheless, comparison of the complementary marker-based genome scan with the exploratory GWAS identified exact cross-method SNP concordance for three of the ten chromosome-wise marker-scan peaks (30%): S3_5435691, S6_52895314, and S9_55930253 (Supplementary Table S9). These concordances represent agreement between analytical approaches applied to the same S3011 dataset rather than independent validation. Together with the low heritability and low frequency of overwinter regrowth in the present experiment, the overall dispersed association pattern indicates limited genomic resolution for this trait under the current design despite localized cross-method agreement at these three markers.

Collectively, the results support a contrasting pattern of genomic resolution between the two perenniality-associated traits. Rhizome development showed detectable but exploratory localized genomic associations, including multi-marker signals and heterogeneous support from complementary analyses. Overwinter regrowth showed localized cross-method marker concordance on chromosomes 3, 6, and 9 but did not resolve into the comparable multi-marker genomic regions observed for rhizome number. These complementary analyses refine the relative support for the exploratory signals but do not establish confirmed QTL, causal loci, or independent validation.

Discussion

Rhizome development and overwinter regrowth showed different levels of genomic resolution in the S3011 diploid-extraction panel. The primary rhizome-number GWAS identified 13 suggestive SNPs resolving into seven spatial signals on chromosomes 1, 2, 3, 4, and 10; none survived Bonferroni or BH-FDR correction. The seven signals therefore remain hypotheses for validation. Secondary analyses did not converge on a single regional hierarchy: chromosome 1 showed the strongest phenotype-oriented allele-dosage differentiation, chromosome 1 cluster 2 and chromosome 3 showed exact within-dataset marker-scan concordance, and local LD and historical correspondence varied among signals. Overwinter regrowth, by contrast, remained genomically dispersed under the present design. This heterogeneous pattern is more informative than a simple claim of “supported” versus “unsupported” loci because each analysis captures a different property of the same structured population.

Heterogeneous regional patterns among exploratory rhizome-associated signals

S3011 is an advanced diploid-extraction panel with family structure rather than a classical linkage-mapping population. A population-structure-adjusted association framework was therefore used to identify marker-trait relationships, with residual relatedness retained as an explicit limitation. The primary result is seven exploratory signals across five chromosomes, not three candidate regions on chromosomes 1, 3, and 4. Chromosome 1 contains two multi-marker clusters; chromosomes 2, 3, and 10 contribute singleton signals; and chromosome 4 contains two separated singleton signals. None met multiple-testing-adjusted significance criteria.

The secondary analyses were designed to characterize, not validate, these signals. High-low allele-dosage contrasts, local LD, historical positional correspondence, positional gene content, and marker-scan concordance were available for all seven signals. Region-specific dosage-response and quality-control analyses were available only for selected regions. Thus, differences in analytical depth should not be interpreted as biological differences in credibility.

For the two chromosome 4 signals, leave-one-accession-out sensitivity analyses provided an additional assessment of whether the suggestive associations were disproportionately influenced by individual accessions. The GWAS model was sequentially refitted after removal of each accession, and both chromosome 4 lead-SNP associations remained below the predefined suggestive threshold ($P < 1 \times 10^{-4}$) after removal of the most influential accession (Supplementary Table S10). These results indicate that the observed chromosome 4 associations were not attributable to a single influential genotype, although this within-sample sensitivity analysis does not provide independent validation of either signal.

For rhizome number, the complementary marker-based genome scan (Supplementary Figs. S7–S8) provided exact SNP concordance with two of the seven spatially distinct suggestive GWAS signals: chromosome 1 cluster 2 at S1_63359080 and the chromosome 3 singleton at S3_62120700. Viewed from the marker-scan perspective, these corresponded to two of the ten chromosome-wise peaks (20%). The remaining representative rhizome-number GWAS signals did not show equivalent colocalization with chromosome-wise marker-scan peaks under the predefined ± 250 -kb criterion (Supplementary Table S9). Because both approaches were applied to the same S3011 population and genotype dataset, these exact correspondences represent cross-method concordance rather than independent replication or validation. Nevertheless, they provide an additional basis for prioritizing these positions for subsequent investigation.

The corrected comparison also identified cross-method concordance for overwinter regrowth (Supplementary Figs. S9–S10), although its overall association profile remained more dispersed than that of rhizome number. Three of the ten chromosome-wise regrowth marker-scan peaks (30%) were exact matches to suggestive GWAS markers: S3_5435691, S6_52895314, and S9_55930253 (Supplementary Table S9). These exact matches demonstrate localized agreement between the two analytical approaches but do not constitute independent validation, because the analyses used the same underlying S3011 dataset. Importantly, the presence of these cross-method matches does not alter the broader observation that overwinter regrowth lacked the localized multi-marker GWAS structure observed for rhizome number under the present experimental design.

Local LD structure also differed substantially among the seven rhizome-number signals (Supplementary Table S3). Chromosome 1 cluster 2 and the chromosome 10 signal showed the clearest evidence of extended lead-to-neighbor LD, with 20 and 16 neighboring markers, respectively, having $r^2 \geq 0.50$. In contrast, the chromosome 3 signal had only one neighboring marker with $r^2 \geq 0.50$, despite having 104 neighboring markers within its ± 250 -kb window, and the remaining signals showed weaker lead-to-neighbor LD. The high–low allele-dosage analysis provided a different regional pattern, identifying particularly strong localized allele-dosage differentiation at the two chromosome 1 signals while the remaining signals showed a continuum of chromosome-relative contrasts. Thus, LD, allele-dosage differentiation, and marker-scan concordance described distinct regional properties and did not uniformly prioritize the same rhizome-number signals.

Positional candidate-gene annotation provided a further common description of the seven signal-centered intervals (Supplementary Table S7). However, the presence, number, or proximity of annotated genes within a signal-centered interval does not provide independent evidence for the association itself or establish functional involvement of any particular gene. Candidate-gene information is therefore most appropriately used to define hypotheses for subsequent functional investigation rather than to rank the statistical credibility of the exploratory signals.

Each secondary analysis measures a different regional property. None overcomes the absence of multiple-testing-adjusted GWAS significance, and none provides independent validation because the analyses were conducted within the same S3011 panel. The seven signals therefore remain explicitly exploratory.

These results are broadly consistent with previous work demonstrating that rhizomatousness in interspecific sorghum populations can be associated with localized genomic regions [39]. However, the historical correspondence observed for the seven S3011 rhizome-number signals was heterogeneous rather than uniform. As described below, the comparisons ranged from interval-level positional correspondence to close or broader regional proximity and little positional correspondence. These relationships provide useful historical context for prioritizing exploratory signals, but they do not establish replication of previously reported QTL or identity of the underlying causal loci.

Historical QTL comparisons reveal heterogeneous positional correspondence among exploratory signals

Comparison with previously reported rhizomatousness loci provides an additional line of evidence for interpreting the seven exploratory S3011 rhizome-number signals (Supplementary Tables S6 and S8). Rhizomatousness

loci for the rhizome-groups (RG) phenotype have previously been reported in an independent interspecific *S. bicolor* × *S. halepense* population [39]. The present comparison indicates that correspondence between those historical loci and the S3011 signals is heterogeneous, ranging from interval-level positional correspondence to close or broad regional proximity and little positional correspondence (Table 3).

The chromosome 10 signal provides the clearest example of interval-level positional correspondence. The S3011 representative marker S10_52748375 at 52.748 Mb falls within the published 5.7–56.7 Mb interval of *qRG.10B.H6.1*. This overlap is notable because it places the exploratory S3011 association within a genomic region previously associated with rhizomatousness, specifically the rhizome-groups (RG) phenotype. However, the historical interval spans approximately 51 Mb and therefore encompasses a substantial portion of chromosome 10. Physical inclusion within such a broad interval provides positional correspondence but relatively limited resolution for inferring that the same underlying locus is responsible for the two observations. Accordingly, the chromosome 10 result should be considered a candidate for further investigation rather than a replicated historical QTL.

The chromosome 3 comparison provides a different form of historical correspondence. The S3011 representative marker S3_62120700 at 62.121 Mb lies within the broad marker span associated with the historical chromosome 3 rhizome-groups (RG) signal *qRG.3.H6.1* and is approximately 3.847 Mb from its reported peak marker, S3_58273458 [39]. Importantly, *qRG.3.H6.1* was identified through the complementary single-marker analysis rather than the interval-mapping analysis used to define several of the other historical QTL. The chromosome 3 comparison therefore cannot be interpreted in the same way as physical inclusion within an interval-mapped QTL. Instead, it represents broad regional correspondence between two marker-based signals. This distinction is particularly important because the S3011 chromosome 3 signal also received complementary support from exact GWAS–marker-scan concordance and additional regional analyses; those additional lines of characterization may justify prioritization of the region, but they do not convert the historical correspondence into QTL replication.

The chromosome 4 8.60-Mb signal showed the closest correspondence among S3011 markers lying outside a historical interval. S4_8603936 was approximately 0.296 Mb below the published interval for *qRG.4B.H6.1*, representing close regional proximity. The two chromosome 1 signals showed broader regional correspondence: S1_62967991 and S1_63359080 occurred approximately 3.932 and 3.541 Mb, respectively, below the published interval for *qRG.1 C.H6.1*. These observations indicate that several S3011 signals occur near regions previously implicated in rhizomatousness, but the physical separations and resolution of the historical intervals preclude treating them as direct overlaps.

Other S3011 signals showed substantially less correspondence with the historical loci represented in the comparison. The chromosome 4 0.49-Mb signal was approximately 8.410 Mb below the published *qRG.4B.H6.1* interval, whereas the chromosome 2 signal was approximately 43.187 Mb below the represented chromosome 2 historical locus, *qRG.2 C.H6.1*. These signals therefore provide little evidence of positional correspondence with the particular historical QTL considered here. Their lack of correspondence does not, however, argue against a biological role for these regions, because the S3011 GWAS and the historical mapping study sampled different genetic backgrounds and used different experimental and analytical designs.

Taken together, the historical comparison therefore does not support a simple interpretation in which the S3011 GWAS either reproduces or fails to reproduce previously mapped rhizomatousness loci. Instead, it reveals a continuum of positional relationships among the seven exploratory signals. Chromosome 10 shows interval-level positional correspondence, chromosome 3 shows broad correspondence with a historical single-marker rhizomatousness signal, chromosome 4 at 8.60 Mb shows close regional proximity, the chromosome 1 signals show broader regional proximity, and the chromosome 2 and chromosome 4 0.49-Mb signals show comparatively little positional correspondence.

These distinctions are important because positional correspondence alone cannot establish replication or causal identity. Differences in mapping population, recombination history, marker density, analytical framework, and mapping resolution can substantially affect the location and width of detected genomic signals. Moreover, broad historical QTL intervals may contain many genes and linked variants. Consequently, the historical comparisons

reported here are best viewed as contextual evidence that can contribute to prioritization rather than as independent validation of the S3011 associations. The seven S3011 signals remain exploratory hypotheses requiring validation in larger populations, additional environments, and independent genetic backgrounds before their utility for perennial sorghum breeding can be established.

High–low allele-dosage contrasts distinguish the chromosome 1 signals from genomic background

The GWAS-effect-oriented high–low allele-dosage analysis provides a common phenotype-associated contrast across all seven exploratory rhizome-number signals. The analysis asks whether a signal occurs in a window with unusual dosage differentiation between high- and low-rhizome accessions; it does not assign parental ancestry. Formal local-ancestry inference would require sufficiently informative parental reference panels and haplotypic information that were not available here.

The strongest localized differentiation occurred at the two chromosome 1 signals. Chromosome 1 cluster 2, represented by S1_63359080, showed a high-rhizome minus low-rhizome dosage contrast of +0.3364 and ranked at the 100.0th percentile of chromosome 1 windows. Its absolute contrast was the second highest among the 1,248 informative 500-kb windows evaluated genome-wide. Chromosome 1 cluster 1, represented by S1_62967991, showed a similarly strong contrast of +0.3002, corresponding to the 99.3rd chromosome-relative percentile and the fifth-highest absolute contrast genome-wide. The occurrence of two spatially distinct chromosome 1 signals within windows showing exceptional dosage differentiation strengthens the rationale for prioritizing these regions for further investigation (Fig. 3; Supplementary Fig. S5; Supplementary Tables S4–S5).

The other five signals showed more moderate or chromosome-unexceptional allele-dosage contrasts. The chromosome 10 signal represented by S10_52748375 showed a contrast of +0.1416 and was at the 77.5th chromosome-relative percentile. The chromosome 4 0.49-Mb signal represented by S4_490163 showed a contrast of +0.1266 and was at the 86.3rd percentile, whereas the chromosome 2 signal represented by S2_23212567 showed a contrast of +0.1134 and was at the 90.6th percentile. These windows therefore exhibited positive differentiation and were relatively elevated within their respective chromosomes, although their absolute contrasts were less exceptional genome-wide than those of the chromosome 1 signals (Supplementary Table S5).

In contrast, the chromosome 3 signal represented by S3_62120700 and the chromosome 4 8.60-Mb signal represented by S4_8603936 showed positive contrasts of +0.1123 and +0.0815, respectively, but ranked at only the 46.4th and 45.2nd percentiles of their corresponding chromosome-wide distributions. Their dosage contrasts therefore provide little evidence of unusual localized differentiation relative to their chromosomal backgrounds. This is particularly informative for the chromosome 3 signal because that region shows exact GWAS–marker-scan concordance and broad historical regional correspondence despite weak lead-to-neighbor LD and an unexceptional allele-dosage contrast. Different analytical approaches therefore provide distinct and non-equivalent forms of evidence for prioritizing exploratory signals.

Genome-wide allele-dosage differences between the high- and low-rhizome groups were generally modest (Supplementary Fig. S4; Supplementary Table S4), indicating that the exceptional chromosome 1 contrasts reflect localized differentiation rather than a broad genome-wide shift between phenotypic classes. More generally, the continuum of contrasts among the seven signals demonstrates that phenotype-associated allele-dosage differentiation is heterogeneous rather than uniformly associated with suggestive GWAS evidence.

The biological interpretation of these contrasts nevertheless remains limited by the absence of phased haplotypes and sufficiently informative parental reference panels. Moreover, orienting dosage according to GWAS effect direction standardizes the direction of phenotype-associated allele-dosage differentiation but does not establish the parental origin of the rhizome-increasing allele. Consequently, even the exceptional chromosome 1 contrasts cannot by themselves demonstrate introgression of specific *S. halepense* segments. Parental reference panels, phased haplotypes, and formal local-ancestry reconstruction will be required to determine whether the localized dosage patterns identified here correspond to specific introgressed genomic segments.

Rhizome development and overwinter regrowth show contrasting levels of genomic resolution

The genomic association profiles of rhizome number and overwinter regrowth differed substantially under the present experimental conditions. Rhizome number produced seven spatially distinct exploratory signals, including two localized multi-marker clusters on chromosome 1. Overwinter regrowth, in contrast, exhibited suggestive associations that were comparatively dispersed across the genome and did not show the same localized multi-marker structure. Nevertheless, comparison with the complementary marker-based genome scan identified exact cross-method concordance for three of the ten chromosome-wise regrowth peaks (30%): S3_5435691, S6_52895314, and S9_55930253 (Supplementary Table S9). These matches demonstrate localized agreement between analytical approaches applied to the same S3011 dataset, but they do not constitute independent replication or validation and do not establish discrete regrowth QTL.

This contrast should therefore not be interpreted as evidence that overwinter regrowth lacks detectable genetic associations. Rather, the primary distinction observed here concerns the degree and pattern of genomic resolution. The ability to resolve genomic associations for regrowth was constrained by the low frequency of successful overwintering, low estimated heritability, binary phenotypic encoding, modest sample size, and evaluation within a single environment. Statistical power to identify genotype–phenotype associations depends on sample size, allele frequency, trait architecture, effect magnitude, and phenotype quality, and limited power can prevent detection of genuine small- to moderate-effect variants [42].

The multilocus allele-dosage sensitivity analyses provided an additional within-sample comparison between the two perennality traits. Indices constructed from the seven representative rhizome-associated SNPs, or from six spatial signal units after combining the two chromosome 1 signals, showed substantially stronger relationships with rhizome number than with overwinter regrowth (Supplementary Table S11). The seven-SNP index was associated with rhizome number at $\rho=0.486$ ($N=123$, $P=1.17 \times 10^{-8}$), compared with $\rho=0.183$ ($N=123$, $P=0.0428$) for overwinter regrowth, while the corresponding correlations for the six-region index were $\rho=0.467$ ($N=124$, $P=4.48 \times 10^{-8}$) and $\rho=0.176$ ($N=124$, $P=0.0509$), respectively. The historical chromosome 3/chromosome 4 comparator index showed a weaker relationship with rhizome number ($\rho=0.370$; $N=124$; $P=2.29 \times 10^{-5}$) and little relationship with overwinter regrowth ($\rho=0.104$; $N=124$; $P=0.252$). Because the component markers were selected from rhizome-number associations in the same S3011 panel, however, these contrasts represent within-sample multilocus allelic-accumulation patterns and should not be interpreted as independent validation of different genetic architectures.

Biologically, overwinter regrowth is also likely to integrate processes beyond rhizome initiation or abundance alone. Successful perennial persistence can depend on coordinated belowground storage, meristem and bud survival, dormancy, resource acquisition, stress tolerance, and responses to environmental conditions [6]. In perennial sorghum specifically, rhizomatous growth and post-winter regrowth are related components of perennality but need not have identical genetic architectures [14, 43]. Rhizome development may therefore contribute to perennial persistence without being sufficient to determine whether a plant survives and resumes growth after winter.

Taken together, the present results support a distinction between stronger localized genomic resolution for rhizome development and a more dispersed association profile for overwinter regrowth under the current design, rather than a distinction between a genetically controlled rhizome trait and a non-genetic regrowth trait. The three exact GWAS–marker-scan concordances for regrowth demonstrate localized within-dataset cross-method concordance, while the multilocus dosage-index analyses indicate that allelic accumulation across the rhizome-associated signals was substantially more strongly related to rhizome number than to overwinter regrowth. Neither result constitutes independent validation. Larger populations, independent genetic backgrounds, and multi-environment measurements of survival and regrowth will therefore be necessary to determine whether the observed regrowth associations are reproducible and whether they can ultimately be resolved into stable genomic regions.

Candidate genes define testable biological hypotheses

The seven signal-centered intervals contain genes annotated for meristem activity, hormone signaling, carbohydrate metabolism, vegetative development, and stress responses—processes plausibly relevant to rhizome initiation, belowground resource storage, and seasonal persistence [6]. These annotations are useful for hypothesis generation but do not identify causal genes. The 286-gene catalogue was defined by physical position alone, and neither proximity to a lead SNP nor functional annotation was treated as evidence of causality. Fine mapping, haplotype analysis, expression profiling, and experimental validation will be required to move from positional association to mechanism.

Modeling considerations and statistical limitations

Several features of the experimental population and analytical framework constrain interpretation of the genomic results. Rhizome expression was analyzed quantitatively despite its ordinal scoring scale. Although this approach facilitates quantitative-genetic analysis, it assumes approximately meaningful differences among phenotypic categories. Overwinter regrowth was binary and was analyzed within a linear-model framework; generalized linear mixed models and larger numbers of regrowing individuals may provide improved inference for this trait in future experiments.

Population structure was modeled with PC1–PC3, and kinship-adjusted models were evaluated as a sensitivity analysis [25, 26, 38]. Kinship broadly attenuated association statistics while changing genomic-inflation diagnostics only modestly. That attenuation is itself informative: the primary PCA-adjusted signals are not robust to every form of relatedness correction. Residual relatedness is therefore treated as a substantive limitation rather than as a problem resolved by the observed inflation factors.

Quantile–quantile diagnostics indicated modest genomic inflation ($\lambda=1.17$ – 1.26). Given the family structure and relatedness among S3011 diploid-extraction accessions, residual structure may contribute to this inflation. Population stratification and cryptic relatedness are recognized sources of inflation in genome-wide association analyses [25, 42]. Consequently, the suggestive associations should not be interpreted as definitive locus discoveries. The absence of associations surviving Bonferroni or BH-FDR correction provides an additional reason for treating the seven signals as hypothesis-generating.

The complementary marker-based genome scan also represents a marker-level analytical framework rather than classical linkage-based interval mapping. Exact concordance with selected GWAS signals describes within-dataset cross-method agreement; it does not constitute independent validation because both analyses used the same population and genotype dataset.

Finally, the principal region-comparison analyses were applied across the complete seven-signal set: allele-dosage contrasts, local LD, historical-QTL correspondence, positional candidate-gene annotation, and comparison with the complementary marker-based genome scan. The remaining asymmetry in analytical depth arises from selected signal-specific allele dosage–response analyses and regional quality-control or sensitivity assessments, which were not applied uniformly across all seven signals. Signals lacking these additional region-specific analyses should therefore not be interpreted as biologically weaker on that basis alone.

Implications for perennial sorghum breeding

The seven exploratory rhizome-associated signals are targets for validation, not markers ready for deployment (Fig. 4). Their predictive value across genetic backgrounds and environments is unknown, and none met multiple-testing-adjusted significance criteria. Near-term breeding value therefore lies in designing validation populations and haplotype tests rather than marker-assisted selection.

Prioritization among the seven signals can nevertheless draw on the heterogeneous complementary evidence generated in this study. Chromosome 1 cluster 2, represented by S1_63359080, and the chromosome 3 signal,

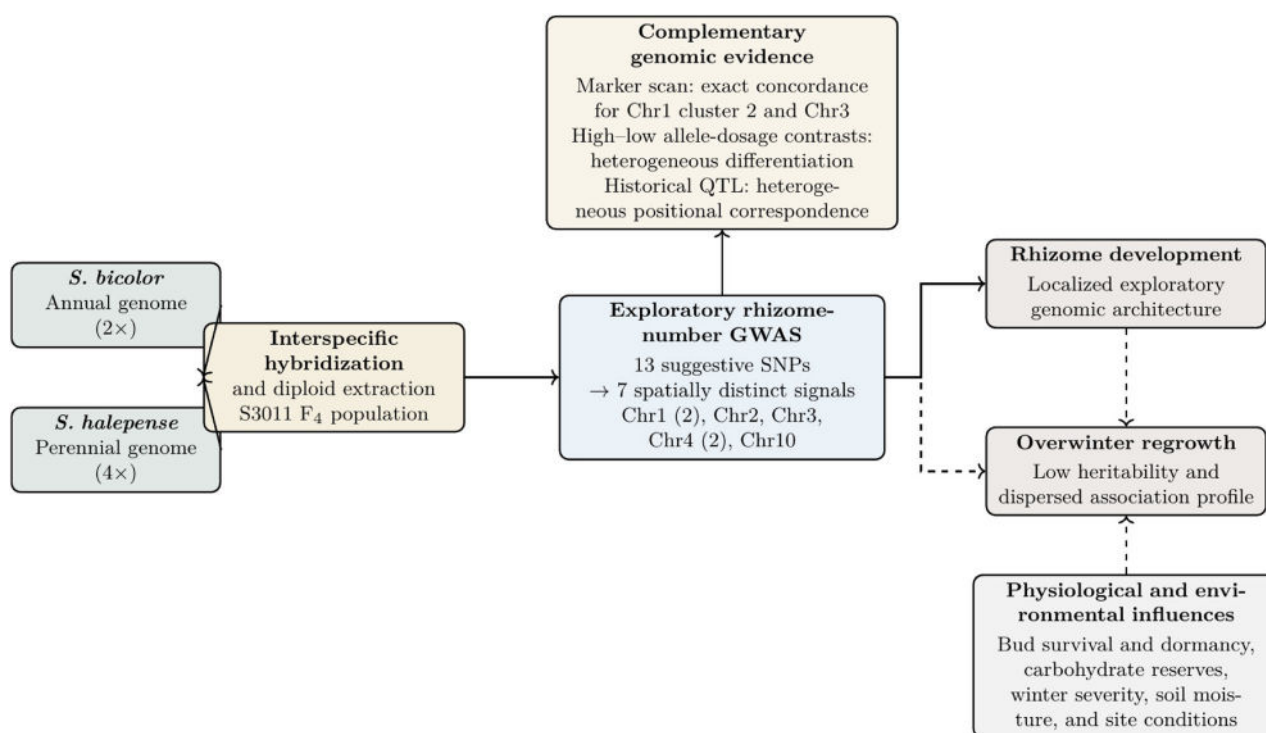


Fig. 4 Conceptual model of perenniality components in the S3011 diploid-extraction population. The S3011 population combines genomic contributions from *S. bicolor* and *S. halepense* following interspecific hybridization and recovery of fertile diploid derivatives. Exploratory mapping of rhizome number identified 13 suggestive SNPs that resolved into seven spatially distinct signals across chromosomes 1, 2, 3, 4, and 10. Complementary analyses provided heterogeneous regional characterization: the marker-based genome scan showed exact SNP concordance for chromosome 1 cluster 2 and chromosome 3, high–low allele-dosage contrasts were strongest for the chromosome 1 signals, and historical-QTL correspondence varied among regions [39]. These analyses provide context for prioritizing exploratory signals but do not establish QTL replication, causal-locus identity, parental ancestry, or functional validation. Overwinter regrowth showed low heritability and a more dispersed association profile. The dashed connection between rhizome development and overwinter regrowth represents a biological hypothesis that rhizome development may contribute to perennial persistence but is not sufficient to determine overwinter survival

represented by S3_62120700, showed exact SNP concordance between the exploratory GWAS and complementary marker-based genome scan. The high–low allele-dosage analysis identified the two chromosome 1 signals as the most exceptional for localized phenotype-associated dosage differentiation. Chromosome 1 cluster 2 and cluster 1 were at the 100.0th and 99.3rd chromosome-relative percentiles, respectively, and ranked second and fifth among 1,248 informative 500-kb windows genome-wide by absolute contrast (Supplementary Tables S4–S5). The other five signals showed more moderate or chromosome-unexceptional allele-dosage contrasts. These patterns provide additional evidence for prioritizing the chromosome 1 regions, but they represent phenotype-associated allele-dosage differentiation rather than direct evidence of *S. halepense* ancestry.

Local LD characterization provided a second common comparison across all seven representative signals (Supplementary Table S3). The most extensive lead-to-neighbor LD was observed for chromosome 1 cluster 2 and the chromosome 10 signal, for which 20 of 35 and 16 of 61 neighboring markers, respectively, had $r^2 \geq 0.50$. The chromosome 3 signal, despite exact GWAS–marker-scan concordance, had only one of 104 neighboring markers with $r^2 \geq 0.50$. These differences illustrate that the complementary analyses capture different properties of the genomic signals and need not identify the same regions as having the strongest support.

Positional candidate-gene annotation was likewise conducted across all seven representative signals. These analyses define genes physically located within the predefined signal-centered intervals and provide targets for subsequent functional investigation, but they do not establish causal relationships or constitute independent

validation of the corresponding GWAS signals. Candidate-gene content should therefore be considered together with, but distinguished from, the statistical, LD, allele-dosage differentiation, and historical-correspondence evidence.

Taken together, allele-dosage differentiation, LD, historical correspondence, positional gene content, and marker-scan concordance provide complementary regional descriptors rather than independent validation. Their rankings are not expected to coincide, and weaker evidence from any single descriptor is not sufficient to discard a suggestive GWAS signal.

Historical-QTL correspondence provides an additional, but distinct, basis for prioritization. The chromosome 10 signal represented by S10_52748375 fell within the broad published interval of the rhizome-groups (RG) QTL *qRG.10B.H6.1*, whereas the chromosome 3 signal showed broad regional correspondence with the historical single-marker rhizome-groups signal *qRG.3.H6.1* [39]. The chromosome 4 8.60-Mb signal showed close regional proximity to the rhizome-groups QTL *qRG.4B.H6.1*, and the two chromosome 1 signals showed broader regional proximity to *qRG.1 C.H6.1*. These relationships increase the biological interest of particular regions but should not be interpreted as replication of the historical QTL. In particular, the very broad historical chromosome 10 interval provides limited positional resolution despite containing the S3011 representative marker.

Signals receiving weaker support from one line of complementary evidence should not be discarded. For example, the chromosome 2 signal and chromosome 4 0.49-Mb signal showed comparatively little positional correspondence with the historical rhizomatousness loci evaluated here, yet their allele-dosage windows were relatively elevated within their respective chromosomes. Conversely, the chromosome 3 signal showed exact GWAS–marker-scan concordance but comparatively limited lead-to-neighbor LD and an allele-dosage contrast that was unexceptional relative to its chromosome background. The chromosome 4 8.60-Mb signal showed close historical-QTL proximity but similarly unexceptional allele-dosage differentiation and weak local LD. These contrasting patterns demonstrate that the complementary analyses provide distinct and non-equivalent forms of evidence. Lack of support from any single analysis therefore cannot be interpreted as evidence that a signal lacks biological relevance.

A logical next step is targeted validation of the complete set of seven signals in larger populations, independent genetic backgrounds, and multiple environments. Dedicated linkage populations, fine-mapping populations, or segregating families carrying alternative haplotypes at prioritized regions could help determine whether the observed marker–trait relationships are reproducible and refine their genomic positions. Haplotype characterization and targeted genotyping could also establish whether prioritized allelic combinations occur at useful frequencies within breeding germplasm and whether their effects are consistent across genetic backgrounds. For regions showing strong allele-dosage-derived differentiation, phased haplotypes and informative parental reference panels will additionally be required to determine whether those patterns correspond to specific introgressed *S. halepense* segments. Functional evaluation of positional candidate genes will likewise be required before individual genes can be implicated in rhizome development. Such validation will be necessary before individual markers or haplotypes can be considered for marker-assisted selection.

More broadly, rhizome development and perennial persistence are unlikely to be adequately represented by a small number of individual markers alone. The contrast observed here between localized exploratory signals for rhizome number and the comparatively dispersed association profile for overwinter regrowth further suggests that breeding for perennality will require integration of multiple component traits and evaluation across environments. Genome-wide prediction may therefore provide a complementary strategy for capturing distributed genetic effects while targeted validation clarifies the contribution of individual candidate regions [44]. In the near term, the principal breeding value of the seven S3011 signals lies in guiding validation, haplotype characterization, parental selection, and the design of populations capable of determining which genomic regions contribute reproducibly to rhizome development and perennial persistence.

Study limitations and future directions

The principal limitations of the present study include modest sample size, single-environment phenotyping, low frequency of overwinter regrowth, residual population structure, and the use of phenotype-oriented allele-dosage contrasts without formal ancestry inference. These factors can constrain the statistical power and reproducibility of association mapping, particularly for complex traits controlled by multiple loci of modest effect [42]. In addition, the absence of a complete linkage map prevented classical interval mapping, and the depth of complementary genomic characterization differed among the seven exploratory rhizome-associated signals.

Future work should first determine whether the seven signals are reproducible in larger populations and independent environments. Dedicated mapping populations and denser haplotype information would help distinguish stable genomic effects from population-specific associations and improve localization of potentially causal variants. Multi-environment phenotyping will be particularly important for overwinter regrowth because perennial persistence reflects both genetic and environmental components and their interaction [6].

Explicit local-ancestry reconstruction using informative parental reference panels would permit direct evaluation of *S. halepense*-derived introgression, which the present phenotype-oriented dosage contrasts cannot provide. Functional characterization of prioritized genes and genomic regions will ultimately be required to establish causality.

For breeding applications, validation should precede deployment of individual markers. If reproducible effects are established, validated regions could subsequently be evaluated within marker-assisted or genomic-selection frameworks. For complex persistence traits such as overwinter regrowth, genome-wide prediction may ultimately complement locus-specific approaches by capturing distributed small-effect variation [44].

Conclusions

In the structured S3011 diploid-extraction panel, rhizome number showed 13 suggestive SNP associations that resolved into seven spatially distinct signals on chromosomes 1, 2, 3, 4, and 10. None survived Bonferroni or BH-FDR correction. Secondary analyses revealed heterogeneous regional patterns rather than uniform confirmation: the strongest phenotype-oriented allele-dosage differentiation occurred at the two chromosome 1 signals, exact within-dataset marker-scan concordance occurred at chromosome 1 cluster 2 and chromosome 3, and local LD and historical positional correspondence varied among regions.

Overwinter regrowth was less resolved genomically, with low heritability, a dispersed association profile, and three exact within-dataset marker-level GWAS–marker-scan matches but no comparable localized multi-marker regions. Together, these results identify testable genomic hypotheses for rhizome development while emphasizing that the present study does not establish QTL, causal loci, parental ancestry, or marker utility for breeding. Validation in larger, independently evaluated populations and environments is required before these regions can be used for genetic inference or selection.

Abbreviations

BLUP	Best linear unbiased predictor
FDR	False discovery rate
GBS	Genotyping-by-sequencing
GWAS	Genome-wide association study
LD	Linkage disequilibrium
MAF	Minor allele frequency
PCA	Principal component analysis
PVE	Phenotypic variance explained

QTL Quantitative trait locus
REML Restricted maximum likelihood
SNP Single-nucleotide polymorphism

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s12870-026-09930-8>.

Acknowledgements We thank the field and laboratory teams at The Land Institute and the University of Georgia for their essential contributions to planting, phenotyping, sample processing, and data collection across multiple seasons. We also acknowledge staff who supported greenhouse management, field logistics, and sample handling. We further thank colleagues whose discussions strengthened the design, analysis, and interpretation of this study.

Declaration of Generative AI Use During manuscript preparation, the authors used a large language model (OpenAI ChatGPT) to assist with language editing and refinement of selected text. All content was critically reviewed and approved by the authors, who take full responsibility for the manuscript

Authors' contributions Author Contributions Statement A.H.P. and S.C. conceived and designed the study. W.K., P.N., S.C., J.R., and A.H.P. contributed to methodology development and study implementation. W.K. and P.N. curated the data and conducted formal analyses. P.N. performed data visualization and prepared the original manuscript draft. W.K., S.C., J.R., and A.H.P. reviewed and edited the manuscript. A.H.P. and S.C. supervised the research and acquired funding. All authors read and approved the final manuscript.

Funding This work was supported by the Malone Family Land Preservation Foundation and the Perennial Agriculture Project, a joint initiative of The Land Institute and the Malone Family Land Preservation Foundation. Additional institutional support was provided by The Land Institute and the University of Georgia. The funding bodies had no role in the design of the study; collection, analysis, or interpretation of data; or in writing the manuscript.

Data availability The final curated datasets, analysis code, supplementary materials, and reproducibility documentation supporting the conclusions of this study are publicly available through Zenodo Version V4 at DOI: 10.5281/zenodo.22074040 [45]. The repository contains processed phenotype and genotype data, genome-wide association analyses, complementary marker-based genome scans, allele-dosage contrast analyses, signal-centered and candidate-gene analyses, historical QTL correspondence analyses, supplementary tables and figures, associated R scripts, audit manifests, and provenance records. Raw genotyping-by-sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession PRJNA1464112; individual SRA accession numbers will become available following completion of NCBI processing and public release.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

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

References

- Warwick SI, Black LD. The biology of Canadian weeds. 61. *Sorghum halepense* (L.) Pers. Can J Plant Sci. 1982;62(4):997–1014.
- Hu FY, Tao DY, et al. Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci USA. 2003;100:4050–4.
- Nabukalu P, Cox TS. Response to selection in the initial stages of a perennial sorghum breeding program. Euphytica. 2016;209:103–11.
- Cox S, Glover J, Van Tassel D, Cox C, DeHaan L. Prospects for developing perennial grain crops. Bioscience. 2006;56(8):649–59.
- Wade LJ. Perennial crops: Needs, perceptions, essentials. Perennial crops for food security. Rome: FAO; 2014.
- Glover JD, Reganold JP, Bell LW, Borevitz J, Brummer EC, Buckler ES, et al. Increased food and ecosystem security via perennial grains. Science. 2010;328(5986):1638–9.
- Crews TE, Rumsey BE. What agriculture can learn from native ecosystems in building soil organic matter: A review. Sustainability. 2017;9(4):578.
- Crews TE, DeHaan LR. The strong perennial vision: A response. Agroecology Sustainable Food Syst. 2015;39(5):500–15.
- Sarath G, Baird LM, Vogel KP, Mitchell RB. Senescence, dormancy and tillering in perennial C4 grasses. Plant Sci. 2014;217–218:140–51.
- Gillespie LM, Volaire F. Are winter and summer dormancy symmetrical seasonal adaptations? Ann Botany. 2017;119:311–23.
- Palmer NA, Saathoff AJ, Scully ED, Tobias CM, Twigg P, Madhavan S, et al. Seasonal below-ground metabolism in switchgrass. Plant J. 2017;92(6):1059–75.
- Ryder N, Dorn KM, Huitsing M, Adams M, Ploegstra J, DeHaan L, et al. Transcriptome assembly and annotation of johnsongrass (*Sorghum halepense*) rhizomes identify candidate rhizome-specific genes. Plant Direct. 2018;2(6):e00065.
- Washburn JD, Murray SC, Burson BL, Klein RR, Jessup RW. Targeted mapping of quantitative trait locus regions for rhizomatousness in chromosome SBI-01 and analysis of overwintering in a *Sorghum bicolor* × *S. propinquum* population. Mol Breeding. 2013;31:153–62. <https://doi.org/10.1007/s11032-012-9778-8>.
- Kong W, Nabukalu P, Cox TS, Goff VH, Pierce GJ, Lemke C, Robertson JS, Compton R, Tang H, Paterson AH. Transmission Genetics of a *Sorghum bicolor* × *S. halepense* Backcross Populations. Front Plant Sci. 2020;11:467. <https://doi.org/10.3389/fpls.2020.00467>.
- Paterson AH, Kong W, Johnston RM, Nabukalu P, Wu G, Poehlman WL, et al. The evolution of an invasive plant, *Sorghum halepense* L. ('Johnsongrass'). Front Genet. 2020;11:317. <https://doi.org/10.3389/fgene.2020.00317>.
- Hodnett GL, Ohadi S, Pugh NA, Bagavathiannan MV, Rooney WL. *Sorghum bicolor* × *S. halepense* interspecific hybridization is influenced by the frequency of 2n gametes in *S. bicolor*. Sci Rep. 2019;9(1):17901.
- Westerbergh A, Doebley J. Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. Theor Appl Genet. 2004;109(7):1544–53.
- Ma A, Qiu Y, Raihan T, Paudel B, Dahal S, Zhuang Y, et al. The genetics and genome-wide screening of regrowth loci, a key component of perennialism in *Zea diploperennis*. G3: Genes, Genomes, Genetics. 2019;9(5):1393–403.
- Swentowsky KW, Bell HS, Wills DM, Dawe RK. QTL map of early- and late-stage perennial regrowth in *Zea diploperennis*. Front Plant Sci. 2021;12:707839. <https://doi.org/10.3389/fpls.2021.707839>.
- McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ et al. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proceedings of the National Academy of Sciences. 2009;106(30):12273–8.
- Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J. The genomic signature of crop-wild introgression in maize. PLoS Genet. 2013;9(5):e1003477. <https://doi.org/10.1371/journal.pgen.1003477>.
- Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. Proc Natl Acad Sci USA. 2016;113(22):E3177–84. <https://doi.org/10.1073/pnas.1525244113>.
- Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: Advances and applications. Nat Rev Genet. 2012;13:85–96.
- Morris GP, Ramu P, Deshpande SP, Hash C, Shah T, Upadhyaya HD, et al. Population genomics and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci. 2013;110(2):453–8.
- Yu J, Pressoir G, Briggs WH, Vroh Bi IG, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38(2):203–8.
- Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60.
- Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;121(1):185–99.
- Aljanabi SM, Martinez I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 1997;25(22):4692–3.

29. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*. 2011;6(5):e19379.
30. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*. 2014;9(2):e90346.
31. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J*. 2018;93(2):338–54. <https://doi.org/10.1111/tpj.13781>.
32. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013. arXiv:1303.3997.
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
34. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
35. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Longman; 1996.
36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–9.
37. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
38. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. 2012;28(18):2397–9.
39. Kong W, Nabukalu P, Cox S, Johnston R, Scanlon MJ, Robertson JS, et al. Unraveling the genetic components of perenniality: Toward breeding for perennial grains. *Plants People Planet*. 2022;4(4):367–81. <https://doi.org/10.1002/ppp.3.10253>.
40. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968;38:226–31.
41. Gaut BS, Long AD. The lowdown on linkage disequilibrium. *Plant Cell*. 2003;15(7):1502–6.
42. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: Biology, function, and translation. *Am J Hum Genet*. 2017;101(1):5–22.
43. Paterson AH, Schertz KF, Lin YR, Liu SCS, Chang YL. The weediness of johnsongrass: A genomic resource for the study of evolution and domestication of Sorghum. *Field Crops Res*. 1995;43(1):13–26.
44. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819–29.
45. Nabukalu P, Kong W, Cox S, Reeves J, Paterson AH. Data and analysis code for Genetic Architecture of Rhizome Development in Diploid Perennial Sorghum Reveals Candidate Regions Supported by Local Ancestry. *Zenodo*. 2026. <https://doi.org/10.5281/zenodo.22074040>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Pheonah Nabukalu¹  · Wenqian Kong² · Stan Cox³ · Jaxk Reeves³ · Andrew H. Paterson^{2,4}

✉ Pheonah Nabukalu
nabukalu@landinstitute.org

¹ The Land Institute, Salina, KS, USA

² Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA

³ Department of Statistics, University of Georgia, Athens, GA, USA

⁴ Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL 36849, USA