Received: 8 August 2025 | Accepted: 1 October 2025

DOI: 10.1002/tpg2.70144

The Plant Genome

ORIGINAL ARTICLE

# A whole-genome assembly of St. Augustinegrass and visualizing diversity within the species

**Ashley N. Schoonmaker**[1,2] | **Ashley G. Yow**[3] | **Xingwang Yu**[2] | **Rocio van der Laat**[2] | **Jeffrey C. Glaubitz**[4] | **Kim Thorsted**[5] | **Matthew D. Robbins**[5] | **B. Shaun Bushman**[5] | **Sheron A. Simpson**[6] | **Brian E. Scheffler**[6] | **Nathan P. Lynch**[7] | **Thomas G. Ranney**[7] | **Susana Milla-Lewis**[2] | **Amanda M. Hulse-Kemp**[3,1,2]

[1]Bioinformatics Graduate Program, North Carolina State University, Raleigh, North Carolina, USA

[2]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, North Carolina, USA

[3]USDA-ARS, Genomics and Bioinformatics Research Unit, Raleigh, North Carolina, USA

[4]Bioinformatics Facility, Biotechnology Resource Center, Cornell University, Ithaca, New York, USA

[5]Forage and Range Research Unit, USDA-ARS, Logan, Utah, USA

[6]Genomics and Bioinformatics Research Unit, USDA-ARS, Stoneville, Mississippi, USA

[7]Horticultural Science, North Carolina State University, Mills River, North Carolina, USA

**Correspondence**
Susana Milla-Lewis, Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695, USA.
Email: srmilla@ncsu.edu
Amanda M. Hulse-Kemp, USDA-ARS, Genomics and Bioinformatics Research Unit, Raleigh, North Carolina 27606, USA.
Email: amanda.hulse-kemp@usda.gov

Assigned to Associate Editor David Edwards.

Ashley N. Schoonmaker and Ashley G. Yow are co-first authors.

**Funding information**
USDA Specialty Crops Research Initiative, Grant/Award Number: 2019-51181-30472; Agricultural Research Service, Grant/Award Numbers: 2080-21000-018-000-D, 6066-21310-005-00D

## Abstract

St. Augustinegrass [*Stenotaphrum secundatum* (Walt.) Kuntze] is a warm-season turfgrass species in the family Poaceae. This species is a popular choice for lawns in the Southern United States, due to its higher tolerance to shade, heat and humidity. However, there is little genomic information available to researchers and breeders, limiting knowledge on the genetic basis for favorable traits. We present a reference-grade chromosome-scale genome assembly for the popular freeze-tolerant diploid cultivar Raleigh. The reference genome has been resolved into two haplotype assemblies (465.41 and 401.52 Mb), accounting for 95.2% and 82.1% of the expected haplotype genome size respectively, each anchored on the nine chromosomes and a total of 62,454 genes. Analysis of the assembly revealed 50.70% of the genome contained repeats. Analysis of the diversity within the species was investigated across 79 genotypes including commercial cultivars, breeding lines, and plant introductions by low-coverage sequencing identifying 605,038 single nucleotide polymorphisms (SNPs). The SNPs were used to investigate genetic diversity across the panel and the effectiveness of low-coverage sequencing on the high GC content species. SNPs

**Abbreviations:** CTAB, cetyltrimethylammonium bromide; GLS, gray leaf spot; LAI, long terminal repeat assembly index; LTR, long terminal repeat; MAF, minor allele frequency; MDS, multidimensional scaling; PI, plant introduction; QTL, quantitative trait loci; SAD, St. Augustinegrass decline; SNP, single nucleotide polymorphism; SSR, simple sequence repeats; TE, transposable element.

classified genotypes into groups matching their phylogenetic and breeding history, with the plant introductions clustering into two groups on either side of the plot. Breeding lines for those whose parents existed in the panel clustered in between the two parents. These results showed that the cheaper, low-coverage option can be used for this type of analysis. Together, all of the resources produced in this study allow the start of the genomics-enabled genetic improvement for St. Augustinegrass.

**Plain Language Summary**

The first published genome sequence of the popular turfgrass, St. Augustinegrass, has been completed. St. Augustinegrass is popular for its ability to thrive in sandy soils, while having a higher tolerance to heat, humidity, and shade than most other southern grown grasses. By using some of the newest sequencing technologies, the genome is currently of a higher quality than any other currently published warm-season turfgrass genome. This reference genome provides a high-quality closely related genome for future genomic work with warm-season turfgrasses. This will allow for the expansion of the molecular toolkits that are available to turfgrass breeders to allow building tools to enable more efficient selection. We also found that a cheaper way of sequencing was sufficient for investigating differences among breeding materials to save money in future studies.

## 1 | INTRODUCTION

St. Augustinegrass [*Stenotaphrum secundatum* (Walt.) Kuntze] is a popular warm-season turfgrass grown throughout humid subtropical regions in the United States (south of Virginia and east of mid-Texas), especially along the coast (Busey et al., 1982; Hodges et al., 1994). St. Augustinegrass thrives in many soil types, but it performs well in sandy soils and has a higher tolerance to heat and humidity, making it a popular choice for lawns in the Southern United States and accounts for as much as 70% of lawns in some states in this region (Casler & Duncan, 2003). The species is set apart from other turfgrasses by wider leaves, which create a tight canopy preventing weed growth, making it a more economical and environmentally friendly choice in regions it grows best. St. Augustinegrass is noted for being the most shade tolerant of the warm-season grasses with some cultivars being more tolerant than others (Winstead & Ward, 1974). While most other warm-season turfgrasses shift to a more upright and open growth when exposed to lower light conditions, St. Augustinegrass retains its density and ground cover, which are critical factors in the quality score of a turfgrass stand (Busey et al., 1982; McBee & Holt, 1966; Winstead & Ward, 1974). This turfgrass species exhibits aggressive lateral growth, an ideal trait for sod production, despite only producing stolons and not rhizomes. However, St. Augustinegrass lacks wear tolerance making it not as ideal for areas under heavy foot traffic. Due to being well adapted to warmer, more tropical climates, it is generally not cold hardy enough in more temperate regions. In addition, the species is susceptible to a number of diseases and pests including the southern chinch bug (*Blissus insularis* Barber) (Reinert & Kerr, 1973), the St. Augustinegrass decline (SAD) virus caused by Panicum mosaic virus (McCoy et al., 1970), gray leaf spot (GLS) caused by the fungal pathogen *Pyricularia grisea* (Carbajal Melgar, 2017), and brown patch (*Rhizoctonia* species) (Piper, 1919).

Raleigh (Figure 1) is the commercial standard for cold tolerance in the species (Kimball et al., 2018) and a prominent diploid cultivar in the North Carolina State University breeding program as it allows the species to expand into the transitional climatic zone of the United States. Raleigh was collected from a private residence in Raleigh, NC, in 1964 and was released by North Carolina State University in the early 1980s (Bateman, 1980) following evaluation trials for cold tolerance and turf quality in multiple locations (Bateman, 1980). The cultivar exhibits a moderately long, narrow leaf of a medium green color and shows tolerance to moderate shade though it is not the most shade-tolerant cultivar (Bateman, 1980; Busey et al., 1982). Although this genotype is less adapted to (sub)tropical locations, it is highly resistant to SAD (McCoy et al., 1970). It is more drought tolerant than other St. Augustinegrass cultivars, but has a lower quality (Graham et al., 2022) as it lacks the fine leaf texture and shorter internode length desirable in the industry (Kimball et al., 2016) and has a lower establishment rate (percent coverage at the end of a

planting year) compared to other prominent cultivars (Graham et al., 2022).

*Stenotaphrum* is a tropical genus of the family Poaceae. Species in this genus are mostly found along eastern Africa, the coast of the Indian Ocean, and southern China, with the exception of *S. secundatum*, which is found in all continents, only excluding Antarctica. St. Augustinegrass's haploid chromosome number is $n = 9$ (Long & Bashaw, 1961). While some cultivars are polyploids, including sterile triploids (3n = 3x = 27) and tetraploids (4n = 4x = 36), most are diploid (2n = 2x = 18) (Long & Bashaw, 1961). Thus, breeding programs generally improve the species through use of the diploid genotypes (Genovesi et al., 2009). Most released polyploids are of unknown provenance and are often characterized by wider, coarser leaves (Horn et al., 1973). Despite its popularity and the vast knowledge of morphological and performance aspects, there is little known about the genetic makeup of the species beyond the ploidy level of breeding material.

Historically, genetic classification of genotypes by breeding programs has relied on visual assessment. Plant morphology in controlled environments has been the main driver of race classification in the species (Busey, 1986; Busey et al., 1982). Lately, however, much progress has been made in developing genetic tools for breeding and genetic research in the species. Due to differences in environments promoting different gene expression patterns, mutations, or genetic drift in geographically separated populations, which then affect morphological characteristics, DNA markers have been used in the past decade to further classify genetic variation across this species. Amplified fragment length polymorphism markers were initially used to assess levels of molecular variation across cultivars and plant introductions (PIs, which are more wild-type or unknown materials, a term utilized by the U.S. Department of Agriculture-National Plant Germplasm System) representing the different ploidy levels present in the species (Milla-Lewis et al., 2013). In a later study, simple sequence repeat (SSR) markers were developed for St. Augustinegrass and used to evaluate genetic diversity (Mulkey et al., 2014). Individuals clustered mainly by ploidy level with diploids breaking into two races: Breviflorous and Longicaudatus (Milla-Lewis et al., 2013; Mulkey et al., 2014). Cultivars Texas Common, Raleigh, Palmetto, Seville, and Jade clustered in the same groups within Breviflorous,

**Core Ideas**

- The first chromosome-scale assemblies of both haplotypes of the *Stenotaphrum secundatum* genome were assembled.
- A total of 62,454 protein-coding genes were annotated across both haplotypes.
- A uniform annotation filtering pipeline was developed for turfgrasses to facilitate future comparisons.
- A panel of 79 diploid representatives of the germplasm for S. *secundatum* were investigated for genetic diversity.
- A low-coverage sequencing platform was tested for effectiveness in variant calling on high GC content species.

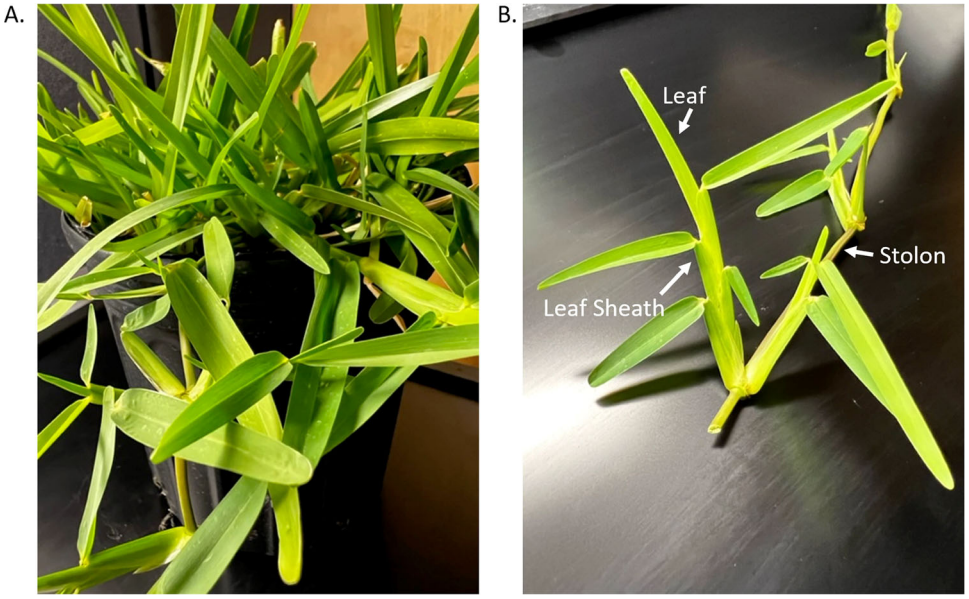



**FIGURE 1** Plant morphology of St. Augustinegrass cultivar Raleigh. (A) Phenotype of the sequenced Raleigh plant. (B) Labeled morphology of the aboveground structures of the plant.

which differed from morphological classifications that separated Seville and Jade from the others (Milla-Lewis et al., 2013). PIs clustered away from cultivars, with those collected from Africa forming a distinct cluster. The SSR study identified slightly different clusters among the diploids, showing results related to the pedigree of the cultivars. In 2014, the first linkage map consisting of 160 SSR markers was developed (Mulkey et al., 2014). More recently, a high-density linkage map of 2871 single nucleotide polymorphism (SNP) and 81 SSR markers, spanning 1241.7 cM, was constructed (Yu et al., 2018). Linkage maps provide a wealth of information in identifying genomic regions associated with traits of interest.

While molecular markers can identify genomic regions linked to traits of interest, they tell little of the composition of the genes themselves. Reference genomes, made possible by long read sequencing, allow researchers to study the chromosomal makeup of the organism and the different haplotypes associated with genes of interest, including resistance genes (Vendelbo et al., 2022; Xia et al., 2019). The reference genome, sometimes referred to as the ultimate genomic map, can be used to determine the true location of the DNA markers used for quantitative trait loci (QTL) mapping and their relationship to important genes. Given St. Augustinegrass is a more heterozygous plant due to outcrossing in the species (Yu et al., 2022), identifying the available haplotypes for each of the genes gives a more complete understanding of the possible combination of traits available for breeders. As no reference genome currently exists for St. Augustinegrass, creating one will greatly assist turfgrass researchers in studying characteristics across grass species, act as a guide for future assembled genomes, and assist breeders in identifying the genes responsible for various traits of interest.

Therefore, the objectives of this study were to (1) use the Raleigh St. Augustinegrass cultivar to develop a high-quality reference genome using long-read sequencing, (2) annotate the phased haplotype assemblies, and (3) utilize this reference genome to identify SNPs in a panel of publicly available diploid cultivars, germplasm, and breeding lines of St. Augustinegrass.

## 2 | METHODS

### 2.1 | DNA isolation, sequencing, and read processing

Young leaves of the St. Augustinegrass cultivar Raleigh were collected for DNA extraction directly into liquid nitrogen. Extraction with typical kits proved difficult to obtain high-quality DNA, therefore, high molecular weight DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) extraction method. Tissue was disrupted with ceramic beads and DNA was obtained following a "slow phenol" adaptation to the method described in Saghai-Maroof et al. (1984). DNA was sequenced with a target read size of 10,000 bp using the circular consensus sequencing (CCS) technology on PacBio's Sequel IIE. The PacBio CCS subreads were run through the PacBio SMRTTOOLS v.7.0.0 software with the default parameters to consolidate and generate consensus reads (Pacific Biosciences).

The leaves were also used to extract DNA for short read sequencing using the Qiagen Plant DNeasy kit. Short reads were sequenced on an Illumina NovaSeq 6000 with a paired-end (PE) 150 bp layout and a target depth of 30x. Illumina reads were evaluated using FastQC and the plot-bamstats feature in Samtools (Andrews, 2010; H. Li et al., 2009).

Purified DNA fragments were obtained from leaves for Hi-C library construction, and libraries were created according to default protocols in the Phase Genomics Kit. Libraries were sequenced on the Illumina NovaSeq platform using the 150 paired-end (PE) layout.

### 2.2 | RNA isolation, sequencing, and read processing

Leaves were collected from five clonal plugs of Raleigh, which were propagated for RNA-seq collection and grown under standard greenhouse conditions. These will be called replicates going forward. Leaves from each of the five replicates were taken at four collection points and placed directly into liquid nitrogen. The collection points were 0, 1, 2, and 6 days after inoculation with *Magnaporthe grisea*, the causal agent of GLS disease. RNA was extracted from the leaves following the standard Trizol extraction protocol (Invitrogen), checked for quality on an Agilent Bioanalyzer using the RNA Integrity Score. RNA was sequenced with a standard RNA-seq protocol using TruSeq adapters on the Illumina HiSeq platform to obtain 2x150 reads for a target of 20–30 million reads per sample. Illumina RNA-seq reads were evaluated for quality using FastQC (Andrews, 2010). Reads were trimmed using Trim Galore, and the quality of the resulting reads was again visualized (Martin, 2011). Due to a high rate of duplication and overrepresented sequences, the reads were trimmed again using fastp v.0.23.4-2 (S. Chen et al., 2018) with –dedup and –dup_calc_accuracy 6. A final quality check was performed with FastQC and the reporting feature of fastp, which confirmed the high quality of the trimmed reads.

### 2.3 | Genome size estimation

Genome size was evaluated with two methods: lab-based and sequence-based estimation. Flow cytometry was performed on Raleigh independently at two separate research locations

with two subsamples each. Propidium iodide was used as a fluorochrome, and *Pisum sativum* 'Ctirad' (8.76 pg) and chicken erythrocyte nuclei were used as internal standards. Read-based genome sizes were estimated using both the CCS and Illumina cleaned reads with GenomeScope v2.0 (Ranallo-Benavidez et al., 2020; Vurture et al., 2017) based on *k*-mer counts from Jellyfish v2.2.9 (Marçais & Kingsford, 2011).

## 2.4 | Assembly and scaffolding

De novo assembly was performed on the CCS reads using both Hifiasm v.0.16.1 and HiCanu v.2 (H. Cheng & al, 2020; Nurk et al., 2020). Using the results of the flow cytometry genome size estimate, the expected haplotype genome size parameter in the assembly process was set to 500 Mb. The haplotype-level assemblies from HiCanu, separated with PurgeDups (Guan et al., 2020), and Hifiasm were evaluated with BBTools v.38.79 (Bushnell, 2014), benchmarking universal single-copy orthologs (BUSCO) v.5.7.1 (Simão et al., 2015), and Merqury v.1.0 (Rhie et al., 2020). Genome size estimates from flow cytometry and GenomeScope, and results of alignment to linkage maps were used to evaluate the assembly statistics and select the best de novo assembly.

Two previously published linkage maps that have cultivar Raleigh as one of the parents (Yu et al., 2018, 2020) were used to scaffold and further assess the assemblies. Marker sequences from the two linkage maps were aligned to the primary assembly with Bowtie2 v.2.5.1 (Langmead & Salzberg, 2012) using default parameters except with flags for exact matches. Positions of the markers on the assemblies were analyzed for order and grouping relative to the linkage maps. Raw reads CCS and Illumina reads were aligned to the contigs of the assembly with pbmm2 v.1.10.0 (https://github.com/PacificBiosciences/pbmm2), a customized wrapper for minimap2 v.2.15 (H. Li, 2018), sorted with Samtools v.1.9 (H. Li et al., 2009), and visualized with the Integrative Genomics Viewer (IGV) v.2.12.3 (Thorvaldsdottir et al., 2013). Contigs were visualized in the locations where a single contig showed regions of both SNP polymorphs (*k*-mers containing flanking sequence plus each version of the SNP base). Contigs were cut based on evidence of mis-assembly, evidence of gaps in the read coverage surrounding the duplicated SNP region, or evidence of a section connected to the rest of the contig by a very short overlap of reads containing multiple mismatches within the overlap. Pieces were then removed to the secondary haplotype assembly. The primary haplotype assembly was then realigned to the linkage maps and scaffolded with ALLMAPS v.1.1.11 (Tang et al., 2015). The primary haplotype assembly was gap filled using Dentist v.1.0 (Ludwig et al., 2022) and the raw CCS reads. The secondary haplotype was scaffolded to the primary haplotype using Ragtag v.1.0.1 (Alonge et al., 2022) and then gap filled using Dentist. Illumina reads were trimmed for adapters and quality (reads less than 55 bp were removed and 12 bp trimmed off the front of the reads) with fastp, aligned with Bowtie2 with sensitivity flagged, and used to polish the combined haplotypes with Pilon v.1.23 (Garrison et al., 2010). Further scaffolding was completed using the Hi-C reads. The first 15 bp of the Hi-C reads were trimmed with fastp. A Hi-C map was generated using the Juicer pipeline (Durand et al., 2016) and scaffolding completed with Juicer-3dna pipeline and Juicebox to orient and scaffold based on the read data. Orientation was verified and chromosome naming was determined by alignment of the chromosomes from each haplotype to *Setaria italica* (Doust et al., 2009) using Mummer v.4.0 (Kurtz et al., 2004).

Markers from the linkage maps were again aligned to both haplotypes to verify the final assembly. BBtools and BUSCO genes from the poales_odb10 database were used to assess the final assembly. Two regions (Chr1 and Chr5) in the assembly containing *Pseudomonas aeruginosa* were identified and masked for downstream analysis. As a final step in the assembly quality control (QC) process, telomeric regions were identified using the Telomere Identification toolKit (tidk) (Brown et al., 2025). The find option was first utilized with the clade parameter set to "Poales." Then, the explore option was utilized to identify potential telomeric repeats between 5 and 13 bp long, and subsequently, the search option was used for all the repeat monomers that were output from explore. The only sequence that appeared to indicate the location of telomeres was the canonical plant telomeric repeat (TTTAGGG).

## 2.5 | Annotation

Both the primary and secondary haplotypes of the genome were concatenated together for annotation using a custom pipeline developed as a part of this project (detailed parameters can be found at: https://github.com/USDA-ARS-GBRU/Grass_annotation_pipeline). Repetitive and transposable elements (TEs) in the genome were identified and soft-masked using the Extensive de novo TE Annotator pipeline (Ou et al., 2019).

As the first step in the gene model prediction process, the quality-controlled Illumina RNA-seq data mentioned previously (see Section 2.2) was aligned to the genome using HISAT2 v.2.2.1 (Kim et al., 2019). The HISAT2 alignments were used in two ways for gene prediction: (1) as supporting RNA-seq evidence for homology-based gene prediction with GeMoMa v.1.9 (Keilwagen et al., 2018), and (2) as input for *de novo* transcript assembly using StringTie v.2.2.0 (Pertea et al., 2015). GeMoMa was used to annotate the genome with protein data obtained from National Center for Biotechnology Information (NCBI) GenBank for six species, including *Eleusine coracana* Genbank Assembly (GCA_032690845.1),

*Oryza sativa* (GCA_009797565.1), *Paspalum notatum* (GCA_036689595.1), *Setaria viridis* (GCA_005286985.2), *Sorghum bicolor* (GCA_000003195.3), and *Urochloa decumbens* (GCA_964030465.3). Then, the RNA-seq alignments were included as evidence for validating the annotations. Simultaneously, de novo transcripts were assembled from the same RNA-seq alignments. Predicted gene models from GeMoMa were combined with the assembled transcripts from StringTie and used as input for Mikado (Mapleson et al., 2018; Venturini et al., 2018), which selects the best transcript for each gene locus based on multiple criteria.

Genes overlapping with TEs were filtered from the results by running TEsorter (Zhang et al., 2022) and subsequently removing the TE-overlapping genes from the dataset. Then, EggNOG-mapper v.2.1.12 (Huerta-Cepas et al., 2019) was used to filter for gene models with orthology assignments to the EggNOG database v.5.0.2. Finally, gFACs (Caballero & Wegrzyn, 2019) was run to filter out any gene models with unrealistic structures, ensuring the final set of annotated genes were high quality. Completeness of the gene set was assessed after each step in the annotation process using BUSCO in transcriptome mode against poales_odb10 conserved orthologs. The final predicted set of proteins was used as input for OMArk (Nevers et al., 2025), which utilized a comparison-based approach against other members of the Panicoideae subfamily to assess the proteome quality of St. Augustinegrass.

Functional annotation was performed on the gene models with EnTAP v.2.3.0 (Hart et al., 2020) to assign putative gene functions based on homology, protein domains, gene ontology (GO) terms, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. As part of the base-level pipeline, EnTAP utilizes its own preformatted database containing the entire set of NCBI taxonomy information and GO terms from the GO Consortium (The Gene Ontology Consortium et al., 2023). EnTAP also includes EggNOG-mapper as part of its default pipeline. In addition to the default sources of gene function assignment, optional databases included in the EnTAP run were InterProScan Pfam (Blum et al., 2025; Paysan-Lafosse et al., 2025), NCBI nonredundant (nr), NCBI RefSeq plant, UniProt TrEMBL, and UniProt Swiss-Prot.

As part of the annotation process, two independent nucleotide-binding leucine-rich repeat (NLR) prediction software packages were used to identify potential genes that play a role in disease occurrence in St. Augustinegrass. NLR-Annotator (Steuernagel et al., 2020) was used to identify potential NLR regions in the genome, while NLRtracker (Kourelis et al., 2021) was used to filter for NLR genes in the predicted gene models. The overlap between genomic regions from NLR-Annotator and the loci of genes identified by NLRtracker was used to determine the overall agreement between the two NLR prediction methods and identify genes with the highest likelihood of being true NLR genes. NLR loci identified by both prediction software were compared to overall gene density by plotting them as tracks on the genome assembly in IGV.

## 2.6 | Orthologous analysis

Orthologous analysis was performed using the OrthoVenn3 web tool (Sun et al., 2023) with the OrthoFinder algorithm and default parameters. This was done for the purpose of comparing the genome annotations of multiple species against the full annotation (including both haplotypes) of St. Augustinegrass Raleigh. The other genomes included in the comparison were derived from the same six species whose proteomes were used as protein evidence during gene annotation (see previous section) and the *Arabidopsis thaliana* genome from the OrthoVenn3 database. As part of the process, OrthoVenn included identification of orthologous clusters and cluster networks, as well as GO term enrichment analysis of the orthologous clusters.

## 2.7 | Synteny analysis

To visualize the similarities and differences between the two haplotype assemblies for St. Augustinegrass Raleigh, synteny analysis was conducted using Synteny and Rearrangement Identifier (SyRI) v.1.7.1 (Goel et al., 2019). The primary and secondary haplotypes were aligned using mummer, and the output was filtered for alignments with similarity ≥90% and length ≥1 kb. Mummer was also used to calculate overall heterozygosity in a standardized method following https://github.com/USDA-ARS-GBRU/StandardizedHeterozygosityEvaluation as indicated in Delorean et al. (2023). Locations of SNPs (see Section 2.8), predicted protein-coding genes, and repeats classified into known repeat families were included as additional tracks, while genomic NLR regions predicted by NLR-Annotator were included as marker points in the haplotype synteny plot.

## 2.8 | Diversity analysis—DNA isolation, sequencing, and analysis

A panel of 79 St. Augustinegrass diploid individuals was selected for inclusion in diversity analysis. The selected samples represented historical cultivars, PIs, as well as breeding lines and a small set of known segregating cross progeny (Table 1). Plants were all grown under standard greenhouse conditions in Raleigh, NC. Young leaf tissue from each individual plant was obtained and DNA was extracted using Qiagen Plant DNAeasy kits. DNA quality was checked using PicoGreen's standard protocol. Library

**TABLE 1** Diversity panel of St. Augustinegrass individuals included for genotyping by sequencing with RipTide.

| Genotype | Type | Female parent | Male parent | Average sequencing depth (x) | NCBI SRA # |
|---|---|---|---|---|---|
| PI 212293 | Plant introduction | N/A | N/A | 2.7 | SRR26199918 |
| PI 410353 | Plant introduction | N/A | N/A | 2.2 | SRR26199917 |
| PI 410355 | Plant introduction | N/A | N/A | 4.0 | SRR26199890 |
| PI 410357 | Plant introduction | N/A | N/A | 8.7 | SRR26199866 |
| PI 410360 | Plant introduction | N/A | N/A | 4.8 | SRR26199855 |
| PI 410361 | Plant introduction | N/A | N/A | 2.9 | SRR26199844 |
| PI 410363 | Plant introduction | N/A | N/A | 14.5 | SRR26199910 |
| PI 410364 | Plant introduction | N/A | N/A | 4.0 | SRR26199883 |
| PI 414079 | Plant introduction | N/A | N/A | 13.7 | SRR26199873 |
| PI 509038 | Plant introduction | N/A | N/A | 9.4 | SRR26199872 |
| PI 509039 | Plant introduction | N/A | N/A | 5.2 | SRR26199900 |
| PI 600734 | Plant introduction | N/A | N/A | 8.6 | SRR26199899 |
| PI 647924 | Plant introduction | N/A | N/A | 0.9 | SRR26199898 |
| PI 647925 | Plant introduction | N/A | N/A | 4.7 | SRR26199897 |
| Amerishade | Cultivar | Unknown | Unknown | 6.6 | SRR26199896 |
| Captiva (NUF-76) | Cultivar | Unknown | Unknown | 10.4 | SRR26199895 |
| Classic | Cultivar | Unknown | Unknown | 7.1 | SRR26199894 |
| Delmar | Cultivar | Unknown | Unknown | 3.4 | SRR26199893 |
| Eclipse | Cultivar | Unknown | Unknown | 11.9 | SRR26199892 |
| FloraVerde (1997-6) | Cultivar | Unknown | Unknown | 7.6 | SRR26199891 |
| Jade | Cultivar | Unknown | Unknown | 4.5 | SRR26199889 |
| Palmetto | Cultivar | Unknown | Unknown | 7.1 | SRR26199888 |
| Raleigh | Cultivar | Unknown | Unknown | 9.1 | SRR26199887 |
| Seville | Cultivar | Unknown | Unknown | 2.8 | SRR26199886 |
| Sunclipse | Cultivar | Unknown | Unknown | 3.6 | SRR26199885 |
| TamStar (DALSA 0605) | Cultivar | Unknown | Unknown | 3.7 | SRR26199871 |
| TX-common | Cultivar | Unknown | Unknown | 4.3 | SRR26199870 |
| Sola (XSA 11377) | Cultivar | Raleigh | PI 410353 | 7.3 | SRR26199869 |
| Silk | Cultivar | Unknown | Unknown | 7.3 | SRR26199868 |
| XSA 10403 | Breeding line | Classic | PI 410364 | 5.8 | SRR26199867 |
| XSA 10137 | Breeding line | Raleigh | Seville | 11.0 | SRR26199865 |
| XSA 10865 | Breeding line | Raleigh | Seville | 8.6 | SRR26199864 |
| XSA 11168 | Breeding line | Raleigh | PI 410353 | 4.8 | SRR26199863 |
| XSA 11237 | Breeding line | TX Common | PI 410353 | 4.6 | SRR26199862 |
| XSA 12166 | Breeding line | Raleigh | PI 410363 | 4.4 | SRR26199861 |
| XSA 14220 | Breeding line | XSA 10034 | XSA 11135 | 13.05 | SRR26199860 |
| XSA 11026 | Breeding line | TX Common | PI 410357 | 11.9 | SRR26199859 |
| XSA 11027 | Breeding line | TX Common | PI 410357 | 9.8 | SRR26199858 |
| XSA 11513 | Breeding line | TX Common | PI 410357 | 5.0 | SRR26199857 |
| XSA 12155 | Breeding line | Raleigh | PI 365032 | 11.9 | SRR26199856 |
| XSA 12341 | Breeding line | Raleigh | PI 410361 | 7.9 | SRR26199854 |
| XSA 12354 | Breeding line | Raleigh | PI 365032 | 1.3 | SRR26199853 |
| XSA 14271 | Breeding line | NUF-77 | XSA 10075 | 7.8 | SRR26199852 |
| XSA 10098 | Breeding line | Raleigh | Seville | 4.8 | SRR26199851 |
| XSA 10127 | Breeding line | Raleigh | Seville | 4.6 | SRR26199850 |

(Continues)

**TABLE 1** (Continued)

| Genotype | Type | Female parent | Male parent | Average sequencing depth (x) | NCBI SRA # |
|---|---|---|---|---|---|
| XSA 192750 | Breeding line | XSA 10098 | XSA 10127 | 3.6 | SRR26199849 |
| XSA 192748 | Breeding line | XSA 10098 | XSA 10127 | 5.8 | SRR26199848 |
| XSA 192664 | Breeding line | XSA 10098 | XSA 10127 | 3.2 | SRR26199847 |
| XSA 192791 | Breeding line | XSA 10098 | XSA 10127 | 3.7 | SRR26199846 |
| XSA 192774 | Breeding line | XSA 10098 | XSA 10127 | 3.3 | SRR26199845 |
| XSA 192789 | Breeding line | XSA 10098 | XSA 10127 | 4.1 | SRR26199843 |
| XSA 10084 | Breeding line | Raleigh | Seville | 7.0 | SRR26199842 |
| XSA 10053 | Breeding line | Raleigh | Seville | 7.4 | SRR26199841 |
| XSA 10056 | Breeding line | Raleigh | Seville | 12.2 | SRR26199840 |
| XSA 10944 | Breeding line | Raleigh | Seville | 13.0 | SRR26199916 |
| XSA 10075 | Breeding line | Raleigh | Seville | 11.2 | SRR26199915 |
| XSA 10106 | Breeding line | Raleigh | Seville | 7.7 | SRR26199914 |
| XSA 16832 | Breeding line | Raleigh | PI 410353 | 10.7 | SRR26199913 |
| XSA 16808 | Breeding line | Raleigh | PI 410353 | 4.7 | SRR26199912 |
| XSA 10011 | Breeding line | Raleigh | Seville | 9.1 | SRR26199911 |
| XSA 10022 | Breeding line | Raleigh | Seville | 12.0 | SRR26199909 |
| XSA 10034 | Breeding line | Raleigh | Seville | 4.5 | SRR26199908 |
| XSA 10074 | Breeding line | Raleigh | Seville | 12.8 | SRR26199907 |
| XSA 10078 | Breeding line | Raleigh | Seville | 5.8 | SRR26199906 |
| XSA 10175 | Breeding line | GF2 | Seville | 7.3 | SRR26199905 |
| Cobalt (DALSA 1618) | Cultivar | PI 300130 | Amerishade | 8.5 | SRR26199904 |
| XSA 11001 | Breeding line | Raleigh | PI 410353 | 3.6 | SRR26199903 |
| XSA 11072 | Breeding line | Raleigh | PI 410353 | 6.1 | SRR26199902 |
| XSA 11187 | Breeding line | Raleigh | PI 410353 | 11.3 | SRR26199901 |
| XSA 11135 | Breeding line | Raleigh | PI 410353 | 8.2 | SRR26199884 |
| DSA 13005 | Breeding line | Raleigh | N/A | 14.6 | SRR26199882 |
| DSA 16001 | Breeding line | Raleigh | N/A | 10.9 | SRR26199881 |
| DSA 16016 | Breeding line | Raleigh | N/A | 9.5 | SRR26199880 |
| GF2 | Breeding line | Raleigh | N/A | 6.5 | SRR26199879 |
| 106SVT3 | Breeding line | Raleigh | N/A | 7.9 | SRR26199878 |
| XSA 11194 | Breeding line | Raleigh | PI 410353 | 3.3 | SRR26199877 |
| XSA 11167 | Breeding line | Raleigh | PI 410353 | 8.1 | SRR26199876 |
| XSA 11272 | Breeding line | Raleigh | PI 410353 | 7.6 | SRR26199875 |
| XSA 16732 | Breeding line | Raleigh | PI 410353 | 7.1 | SRR26199874 |

Abbreviations: N/A, not available as occured in the wild; NCBI, National Center for Biotechnology Information; PI, Plant introduction; SRA, short read archive.

construction for the genotyping by sequencing was conducted using the iGenomX RipTide high-throughput rapid library prep kit (iGenomX, Carlsbad; now Twist 96-Plex Library Prep Kit) following the manufacturer's instructions (https://www.twistbioscience.com/products/ngs/library-preparation/twist-96-plex-library-preparation-kit) with high-GC primers. The library was sequenced on the NextSeq 2K for paired-end reads of the 150 bp target size to obtain an average expected depth per sample of 3.5 Gb, equating roughly to 7x for a 500 Mb genome.

The raw sequence data were demultiplexed to the plate level using six base i7 barcodes with Illumina bcl2fastq software (v.2.20; Illumina, Inc.). Sequence data were demultiplexed with fgbio DemuxFastqs v.1.5.0-b01fc04-SNAPSHOT (http://fulcrumgenomics.github.io/fgbio/tools/latest/DemuxFastqs.html) to the sample level in bam file format using the eight base inline barcodes in read 1 according to the guide provided by Twist Bioscience (https://www.twistbioscience.com/resources/guideguideline/twist-96-plex-library-preparation-kit-sample-demultiplexing-guide).

The demultiplexed sequences were then aligned to the Raleigh reference genome using bwa-mem v.0.7.17-r1188 (H. Li, 2013). Picard Tools v.2.26 (http://broadinstitute.github.io/picard/) SamToFastq was used to convert the input bam files to the fastq format read by bwa-mem. The output was streamed into Picard Tools MergeBamAlignment so that the pseudo molecular barcodes could be recovered from the input bam files. Duplicate aligned sequences were then identified using Picard Tool MarkDuplicates. Bcftools mpileup v.1.15.1 (H. Li, 2011) was then used to calculate allelic depths at each variant only for reads that formed a proper pair with a minimum mapping quality of 60 and a minimum base quality of 20. This was parallelized by chromosome. Bcftools filter was then used to filter for SNPs at least five bases away from an indel using these flags –SnpGap 5:indel –include 'TYPE' = "snp" & (INFO/AD[0]+INFO/AD[1]) > = 564 & (INFO/AD[0]+INFO/AD[1]) < = 1128 & (INFO/AD[1]/(INFO/AD[0]+INFO/AD[1])) > 0.02 & (INFO/AD[2]/(INFO/AD[0]+INFO/AD[1])) < 0.01′, with an average depth between 6x and 12x per sample at a given SNP. Filtering procedures considered the depth of the first alternate allele being greater than 2% of the depth of the reference and first alternate allele combined; and, to prevent rare sequencing errors from affecting the process, which would create a third allele at otherwise useable SNPs, preventing rare sequencing errors contributing to <1% of the overall depth from leading to rejection of otherwise good SNPs. A custom awk script was used to call genotypes for each SNP for samples with a minimum depth of 7x. The genotypes were called as heterozygous if the lower depth allele comprised at least 20% of the depth in that individual. The resulting SNPs were then filtered for minimum minor allele frequency (MAF) of 5%, and F < IT > between −0.2 and 0.2. The F < IT > (index of panmixia) was calculated with a combination of bcftools query and a custom awk command. Imputation was then performed with Beagle 4.1 (Browning & Browning, 2016) using the parameters "window = 2000 overlap = 500 err = 0.001 ne = 10000" (Figure S1).

The diversity panel's SNP data were filtered for minor allele frequencies (<5%). Using plink v.1.9, principal component analysis and multidimensional scaling (MDS) plots were generated (Chang et al., 2015; Purcell et al., 2007). Plots were visualized in the R Statistical Software package and colored according to type (cultivar, PI, and breeding line) and population parentage (for those from a known segregating cross) (Henry & Wickham, 2020; Pedersen et al., 2020; R Core Team, 2022; RStudio, 2020; Slowikowski, 2021; Villanueva & Chen, 2019; Wickham, 2019, 2021; Wickham & Wickham, 2017; Wickham et al., 2017, 2018). The panel was analyzed with fastStructure v.1.0 (Raj et al., 2014) and plots were visualized in R with the sample names added to the graph using the most likely cluster number ($k = 3$).

# 3 | RESULTS

## 3.1 | Genome sequencing, assembly, and evaluation

Expected size of the genome was determined using flow cytometry of Raleigh conducted in two separate labs and reported genome sizes of 0.98 and 1.00 pg/2C (Figure S2A). Based on this, the diploid genome size was estimated to be 978 Mb. Results of *k*-mer evaluation using the CCS reads with Genomescope gave a diploid genome size of 760 Mb.

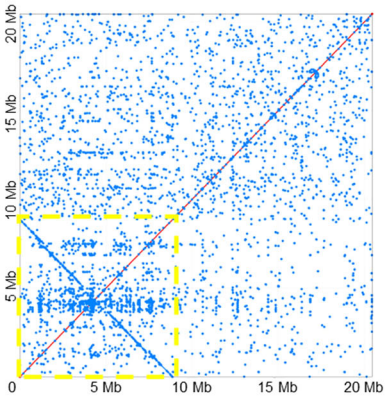
The CTAB method was found to produce high-quality high molecular weight (HMW)-DNA, which proved difficult to obtain with other methods. Whole-genome sequencing was performed on the HMW-DNA using PacBio CCS technology, resulting in 1,942,961 consensus reads totaling 16.7 Gb, with an average length of 8.6 kb and an approximate sequencing depth of 33x across the genome. The genome was assembled using HiCanu (Nurk et al., 2020) and Hifiasm (H. Cheng & al, 2020) using the results of the lab estimated haplotype genome size of 500 Mb. However, different runs with the same set of reads and using results of other genome size estimates as the "expected size" parameter made no difference to the final assembly size. Therefore, the better assembly was determined by comparing assembly statistics (Table S1) and biological expectations (pg/2C genome size, chromosome number, etc.). While HiCanu was able to partition out the haplotypes reasonably well, with the primary haplotype genome size of 429.751 Mb and the secondary haplotype genome size of 422.018 Mb consisting of less contigs overall than the Hifiasm assembly; the assembly was still made up of much smaller contigs than the Hifiasm assembly (Table S1). The largest contig in HiCanu was 9.08 Mb, while the largest in the Hifiasm was 30.37 Mb. The primary Hifiasm assembly was closer to the expected haplotype genome size with a final assembly size of 462.71 Mb. The HiCanu assembly was found to have a higher number of potential contig mis-assemblies than the Hifiasm assembly when aligned to the linkage maps (details below). Therefore, we utilized the Hifiasm assembly as the draft for downstream analysis.

Both the primary and secondary haplotypes were aligned to the two published linkage maps (Yu et al., 2018, 2020) generated from $F_1$ populations involving Raleigh as a parent. When the SNP marker sequences or polymorphs from the SSR maps were mapped back, 85% of the SNPs from the Raleigh × Seville map and 57% of the Raleigh × PI 410353 map were able to be localized to the assembly. Three contigs in the primary haplotype exhibited repeating blocks of SNPs that contained both alleles or SNP polymorphisms within a contig (Figure 2A). These three contigs with identified mis-assemblies around repeating blocks (Figure 2B) were manually broken between the identified block and the

A.

| RLG1 | | | |
|---|---|---|---|
| SNP | Genetic Position | Contig | Contig position |
| SNP41541 | 55.775 | utg000014l | 42727 |
| SNP51740 | 53.151 | utg000014l | 116066 |
| SNP2820 | 59.761 | utg000014l | 523652 |
| SNP29943 | 52.244 | utg000014l | 578548 |
| SNP9608 | 57.019 | utg000014l | 1819703 |
| SNP13530 | 57.553 | utg000014l | 2372621 |
| SNP3740 | 58.856 | utg000014l | 2428030 |
| SNP46801 | 50.702 | utg000014l | 2920383 |
| SNP19956 | 57.762 | utg000014l | 3936602 |
| SNP10127 | 57.978 | utg000014l | 4748031 |
| SNP55196 | 57.289 | utg000014l | 4805701 |
| SNP5406 | 61.69 | utg000014l | 4855219 |
| SNP46801 | 50.702 | utg000014l | 5851436 |
| SNP3740 | 58.856 | utg000014l | 6319974 |
| SNP13530 | 57.553 | utg000014l | 6375433 |
| SNP9608 | 57.019 | utg000014l | 6932129 |
| SNP29943 | 52.244 | utg000014l | 8298334 |
| SNP2820 | 59.761 | utg000014l | 8353272 |
| SNP51740 | 53.151 | utg000014l | 8748421 |
| SNP41541 | 55.775 | utg000014l | 8821370 |
| SNP19306 | 55.512 | utg000014l | 8874132 |
| SNP18321 | 51.609 | utg000014l | 8997532 |

B.




C.

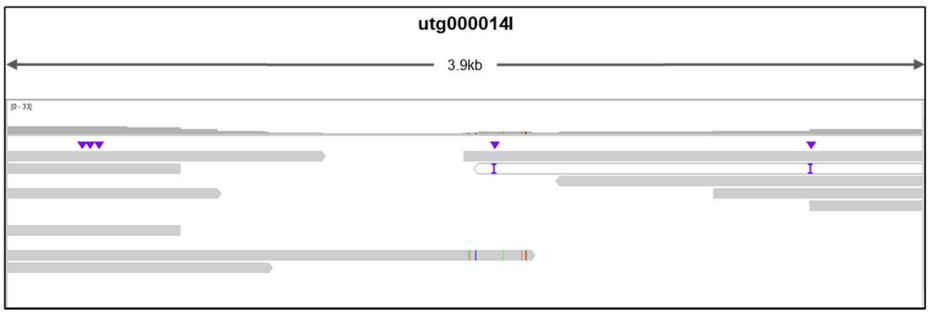



**FIGURE 2** Visualization of presence of both haplotypes of a single nucleotide polymorphism (SNP) block inside of one contig and the alignment of CCS consensus reads to the assembly. (A) Portion of table showing the SNP IDs, position on the linkage map (genetic position), contig ID, and position of alignment for the contig. The duplicated block is highlighted yellow in the SNP ID column. (B) Alignment of contig to itself. The same highlighted double-haplotype block in (A) is outlined in yellow to aid in visualization. (C) Example of a low-coverage section between the two haplotype blocks of SNPs. Colored lines on the lower of two overlapping reads indicate multiple mismatches between the read and the assembly. ID, identification.

unduplicated blocks, where there was low assembly support (usually a single overlap between two reads), or where one of the reads showed multiple mismatched base pairs, suggesting that it belonged to the secondary haplotype (Figure 2C), or where there was a break in alignment between the reads and the contigs. Mismatched reads were used to manually edit the removed contig piece. Cut contig pieces were moved to the secondary haplotype file.

Further scaffolding was completed on the scaffolded, gap-filled, and polished assembly using the Hi-C reads. Scaffolding was run on the combined haplotype genomes, which highlighted chromosome pairs, representing a complete copy of each chromosome in each of the primary and secondary haplotype genomes (Figure 3). The primary assembly improved by 11.8% and the secondary by 3.3%.

The Hi-C scaffolded primary and secondary haplotypes were again evaluated for quality with BUSCO, Merqury, and Bbtools, and alignment to the genetic maps and the *S. italica* reference genome (Doust et al., 2009). The primary assembly resulted in a genome size of 455.27 Mb with over 97% of the genome in nine scaffolds representing the expected nine haplotype chromosomes and with over 97% of the total assembly in scaffolds greater than 50 kb (Table 2). The secondary haplotype also contained 90% of the genome in nine chromosomes with a total size of 401.53 Mb. Merqury results showed that most of the *k*-mer content within the read data was captured in the final assembly (Figures S2B,C), with an overall completeness of 97% (and the primary (v1) and secondary (v1a) haplotypes having *k*-mer completeness scores of 87.8% and 81.2%, respectively). BUSCO

**TABLE 2** Statistics of the final primary and secondary haplotypes genome assemblies.

| | St. Augustinegrass cultivar Raleigh genome assembly | |
|---|---|---|
| | **Primary haplotype** | **Secondary haplotype** |
| **Genome size (Mb)** | 455.268 | 401.527 |
| **Main genome contig total** | 812 | 2716 |
| **Contig N50/L50 (Mb)** | 12/12.064 | 151/819.896 |
| **Contig N90/L90 (Mb)** | 40/3.473 | 554/0.148 |
| **Max contig length (Mb)** | 30.366 | 4.436 |
| **Main genome scaffold total** | 631 | 539 |
| **Scaffold N50/L50 (Mb)** | 4/53.133 | 4/46.632 |
| **Scaffold N90/L90 (Mb)** | 9/38.128 | 8/33.768 |
| **Max scaffold length (Mb)** | 70.895 | 59.333 |
| **% in scaffolds > 50 kb** | 97.57 | 98.44 |
| **Adjusted LAI score** | 17.61 | 17.36 |
| **BUSCO score ($n$ = 4896)** | C:98.7% [S:94.8%, D:3.9%], F:0.8%, M:0.5% | C:97.2% [S:94.2%, D:3.0%], F:0.9%, M:1.9% |
| **Protein-coding genes** | 33,836 | 28,618 |

Abbreviations: BUSCO, benchmarking universal single-copy orthologs; LAI, long terminal repeat assembly index.

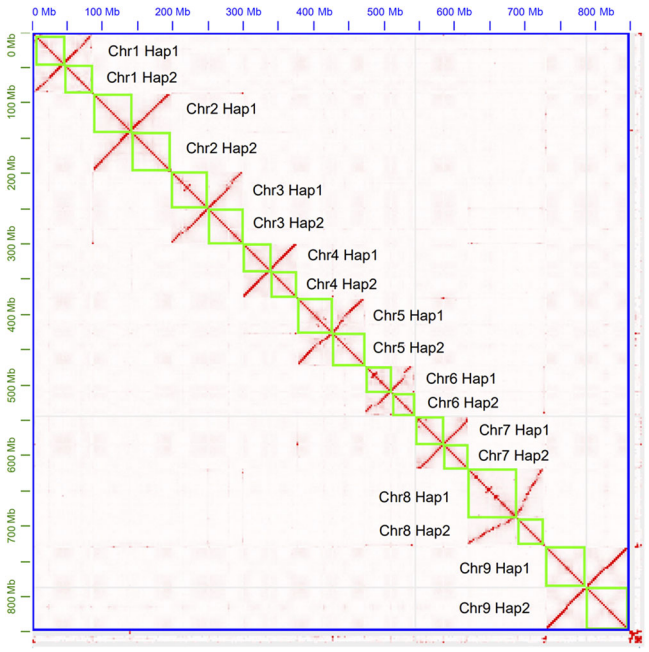



**FIGURE 3** Hi-C Map of the overall diploid assembly containing both primary and secondary haplotypes. Green boxes show boundaries of scaffolds. Scaffold identification is shown beside the boxes. "Hap1" refers to primary haplotype, and "Hap2" refers to secondary haplotype.

assessments of the combined primary and secondary haplotypes indicated 97% complete orthologs, but with nearly all of those being duplicated. Individually, the primary assembly contained 98.3% complete orthologs and the secondary 95.9% complete orthologs, with each also having 4.5% and 2.5% of duplicated orthologs, respectively. Long terminal repeat assembly index (LAI) analysis of the assemblies with long terminal repeat (LTR)-retriever (Ou & Jiang, 2018) indicated the assemblies were of reference quality status (Ou et al., 2018). The primary had an adjusted LAI of 17.61 and the secondary 17.36. Regarding telomeric repeats in the assembly, the primary haplotype had four chromosomes with telomeres at both ends (chr2, chr3, chr5, and chr9) and four chromosomes with a single telomere (chr1, chr4, chr6, and chr7), while the secondary haplotype had one chromosome with both telomeres assembled and five chromosomes with a single telomere. Taken together, all results agreed with the overall high quality of the new reference assembly for St. Augustinegrass and that the primary haplotype is the most complete haploid version of the genome.

Alignment to the genetic maps showed high consistency between the maps and both haplotype assemblies. The Raleigh × Seville genetic map showed a near perfect correlation between the SNP linkage and the alignment positions on the assembly (Figure 4A,B). Meanwhile, the Raleigh × PI 410353 map exhibited two lines of correlation between the map and the assembly alignment (Figure S3). Alignment to the corresponding chromosomes of *S. italica* (Doust et al., 2009), the most closely related available genome, demonstrated a reasonable level of agreement between the genomes with a few inverted regions (Figure 4C). Therefore, all scaffolds were assigned to a chromosome number utilizing this alignment information.

Synteny analysis of individual chromosomes of *S. secundatum* Raleigh haplotypes revealed a high number of structural variants in the form of insertions (182,749), deletions (189,207), and duplications (13,799); however, the number of inversions (46) and translocations (871) was relatively low (Figure 5). SyRI identified 1585 highly diverged, but syntenic regions between haplotypes, which may be due to the high level of heterozygosity in this species. Using the alignment

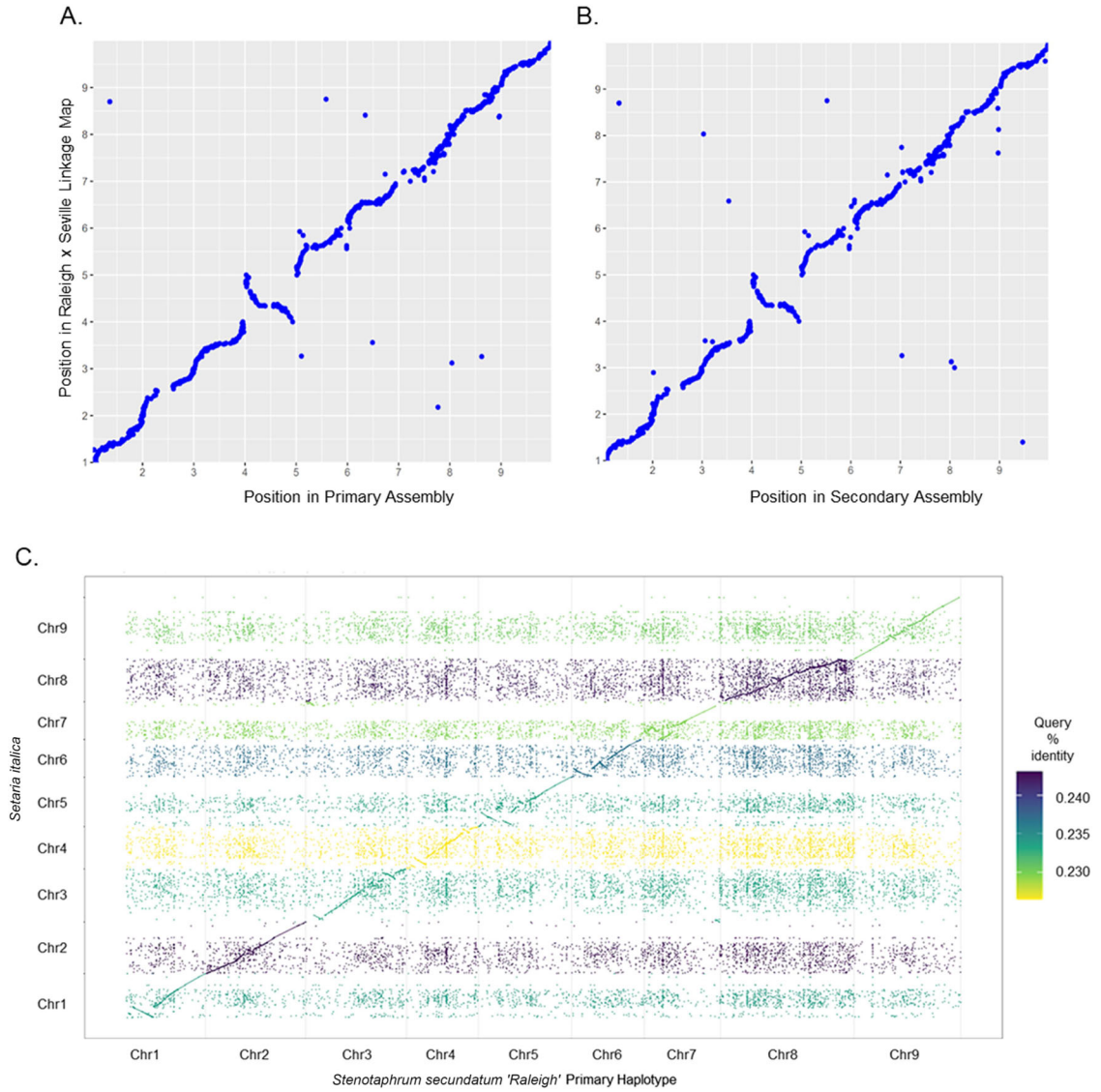



**FIGURE 4** Alignments of the *Stenotaphrum secundatum* assembly to published information. (A) Correlation of markers from the Raleigh × Seville genetic map (*Y*-axis) to the physical position of makers aligned to the final St. Augustinegrass primary haplotype (*X*-axis) (Yu et al., 2018). (B) Correlation of markers from the Raleigh × Seville genetic map (*Y*-axis) to the physical position of markers aligned to the final St. Augustinegrass secondary haplotype (*X*-axis). (C) Alignment dot plot between the St. Augustinegrass primary haplotype assembly and the reference genome for *Setaria italica*.

between haplotypes, an intragenomic SNP rate (total SNPs/total bp in reference haplotype) of 0.005 was calculated, corresponding to one SNP per 194 bp. Finally, 10.67% of the primary and 2.15% of the secondary haplotypes did not align to each other, which can be seen as gaps in the synteny plot.

## 3.2 | Genome annotation

Roughly 49% of the diploid genome was masked, which breaks down to 50.03% of the primary and 48.61% of the secondary (Table S2). In the final assembly, 62,454 protein-coding genes were predicted on the overall diploid assembly (both haplotypes). Of the total protein-coding genes, there were 33,836 gene models in the primary assembly and 28,618 in the secondary assembly (Table 2), which equated to BUSCO (poales_odb10) transcriptome completeness scores of 96.5% and 91.1% for the primary and secondary assemblies, respectively. Proteome quality analysis with OMArk demonstrated that the *S. secundatum* Raleigh genome annotation was comparable to other high-quality, curated grass family species assemblies from the Ensembl database, with very few partially mapped or fragmented orthologs (Figure S4). These results provided further validation for the high quality of the genome annotation presented here. Aside from

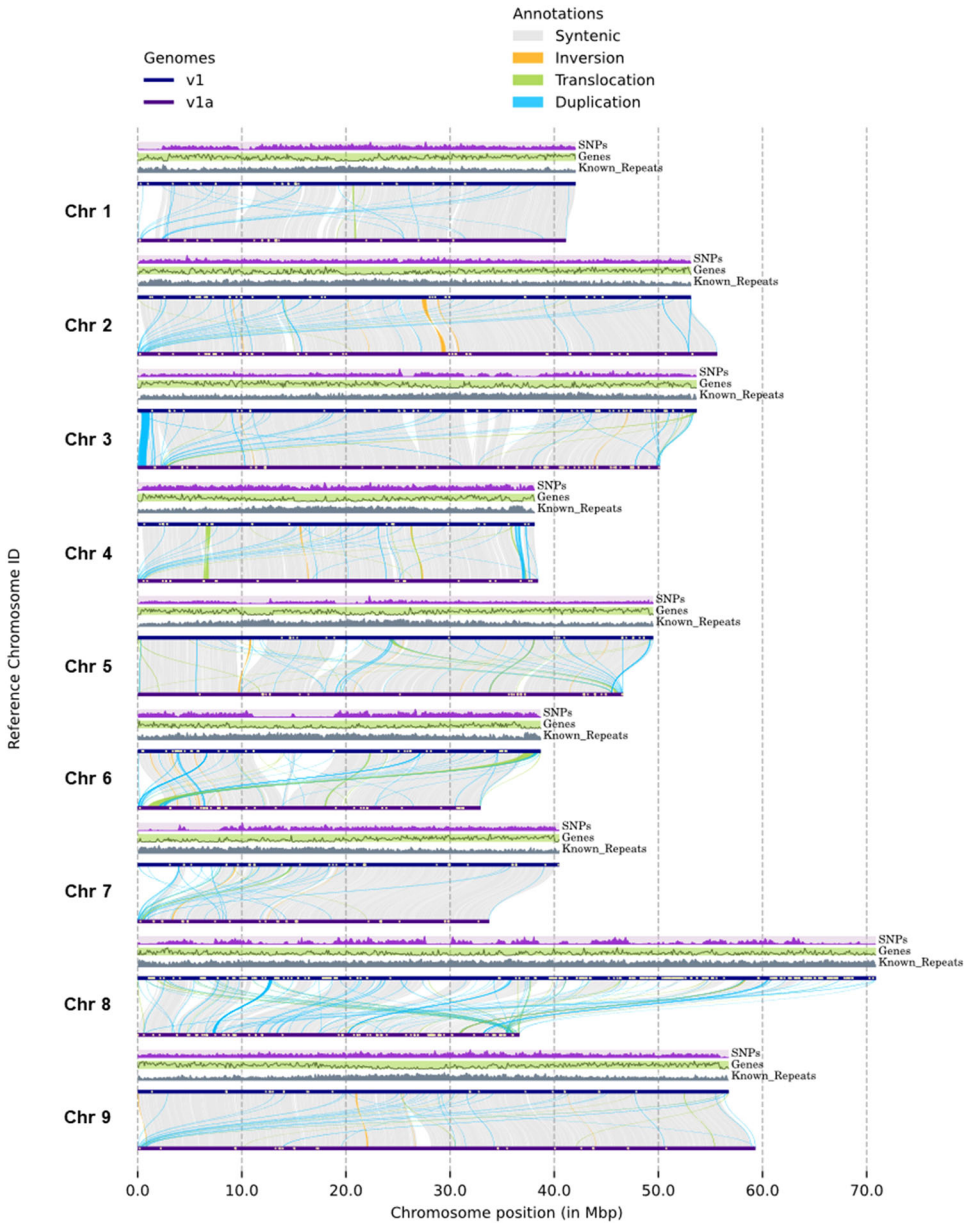



**FIGURE 5** Synteny plots between the *Stenotaphrum secundatum* 'Raleigh' haplotype assemblies v1 (primary) and v1a (secondary). Pale yellow dots along the genome sequence bars indicate predicted nucleotide-binding leucine-rich repeat (NLR) regions from NLR-Annotator. The single nucleotide polymorphisms (SNPs) track shows density of SNPs generated from the low-coverage reads. The known_repeats track shows the density of primary haplotype (v1) repeats that were assigned to a known repeat family (no unknown repeats).

protein-coding genes, there were also 20,685 ncRNA genes annotated in the full genome assembly, which divided up into 11,881 ncRNA genes in the primary and 8,804 ncRNA genes in the secondary assemblies. Finally, a total of 1160 genomic regions were annotated as putative NLR loci by NLR-Annotator, and 899 genes were determined to be NLR genes by NLRtracker across both haplotypes (Table S3). Analysis of the overlap between the NLR gene loci and genomic NLR loci revealed that the majority (66.9%) of genomic NLR loci had an overlap with one predicted NLR gene, indicating the high likelihood that those loci genuinely harbor NLR genes. Additionally, the NLR loci had a strong tendency to cluster together within the genome (Figure 5; Figure S5).

Identifying the overlaps between the intragenomic SNPs and predicted gene loci in the primary haplotype revealed 312,251 (13.3%) of the SNPs were located within the boundaries of 28,030 genes, corresponding to 82.8% of the genes predicted in the primary haplotype.

The EnTap software was used to assign protein domains, putative gene functions, GO terms, and KEGG pathways to each of the gene models using multiple databases. A total

of 62,451 (99.99%) genes were functionally annotated in the entire set of gene models, which demonstrates the high quality of the annotation presented in this study. Most of the genes models (41,435) were annotated with the RefSeq plant database, followed by the UniProt TrEMBL (16,741), NCBI nr (2886), and UniProt Swiss-Prot (713) databases. A total of 53,800 genes were annotated with the EggNOG database.

Orthologous analysis performed with seven grass species and *Arabidopsis* as input revealed a total of 9338 orthologous clusters containing proteins from all species and a steep drop down to 3585 clusters containing only grass species proteins (Figure 6A), which could represent clusters of core angiosperm and core grass family genes. Similarity comparisons between the species indicated that St. Augustinegrass shared the most orthologous clusters with *Urochloa decumbens* (Figure 6B). *S. secundatum* had 1403 single-species orthologous clusters, which were enriched for GO terms involved in stress response and reproduction (Figure 6C), and 840 singletons. Even though *S. secundatum* had the second highest number of proteins out of all the species included in the orthologous analysis, it had the lowest proportion of singletons (1.34%).

## 3.3 | Diversity analysis

The panel of 79 St. Augustinegrass accessions produced an average sequencing depth, calculated from the total number of bases sequenced divided by the estimated haplotype genomes size (500 Mb), of 7.12x with a maximum depth of 14.62x and a minimum of 0.87x. Reads for each sample were assessed for average depth of coverage, proportion of genome, and gene space coverage across multiple genomes. Alignment of the reads to the genome assembly revealed a mostly consistent coverage across all of the chromosomes with some lower coverage spots that could be due to genomic structure or sample sequencing variation (Figure S6). From the initially discovered 760,159 SNPs, curation for an MAF of 0.05 resulted in a total of 605,038 remaining SNPs (Table 3). Heterozygosity was determined by the percentage of SNPs that were heterozygous in a sample, which averaged 29.8% across the panel. There appeared to be minimal impact of sequencing depth on clustering (Figure S7A).

MDS plots showed the samples clustering into three loosely defined groups (Figure 7A). The left cluster contained a mixture of breeding lines, cultivars, and PIs, where nearly all of the breeding lines were progeny of a cross between Raleigh and Seville and fell between the two parents (Figure 7B). Cultivars FloraVerde, TamStar, DALSA 1618, and Seville fell into the middle cluster away from the rest of the cultivars. The rest of the middle cluster is made up of breeding lines, a majority of which resulted from a cross between Raleigh (located in the first cluster) and PI 410353 (located on the far right) (Figure S7B). In the far right, PI 410353, PI 410360, PI 410361, and PI 410364 clustered away from the other PI. PI 410355 did not belong to any of the clusters, though it was closest to the middle cluster. The chooseK script in FASTSTRUCTURE identified a two cluster structure ($k = 2$) as the appropriate number of model components that maximizes the marginal likelihood of the model complexity (Figure 7C).

**TABLE 3** Summary of variant discovery in a panel of 79 St. Augustinegrass diverse accessions.

| Metric | Number |
|---|---|
| Total SNPs | 760,159 |
| Number of SNPs after MAF of 0.05 | 605,038 |
| Average sequencing depth (x) | 7.12 |
| Minimum sample sequencing depth (x) | 0.87 |
| Maximum sample sequencing depth (x) | 14.62 |

*Note*: Genotyping by sequencing was performed using the RipTide library protocol.
Abbreviations: MAF, minor allele frequency; SNP, single nucleotide polymorphism.

# 4 | DISCUSSION

We selected the Raleigh cultivar to create a reference genome for *S. secundatum* due to its importance in breeding for cold tolerance in the species. The estimated genome size of the diploid Raleigh cultivar (2n = 2x = 18) with flow cytometry was 978 Mb as determined by two independent labs in this study. Other estimates produced similar (slightly higher) results, including prior flow cytometry results from Arumuganathan et al. (1999), which gave a genome size 1.08 pg/2C (1056.24 Mb). However, these estimates were ~22% larger than the *k*-mer estimation based on reads. Differences between the *k*-mer run and the flow cytometry experiments may have been caused by collapsing of repeat regions in the *k*-mer analysis leading to a smaller estimation. The results of these estimations and comparisons to our lab-based results and those of previously conducted experiments argue against the reliance on read-based estimations when selecting the best draft assemblies. However, as noted above, using different haploid genome sizes did not appear to affect the overall size of the resulting genome assembly from either assembly program used.

Two haplotype assemblies were constructed to represent a diploid chromosome-level genome. The primary assembly (455.27 Mb) was more complete than the secondary assembly (401.53 Mb in 539 scaffolds). The final primary assembly (including scaffolds outside of assembled chromosomes) is only 6.89% shorter than the expected haploid genome size based on flow cytometry. Both haplotypes present higher

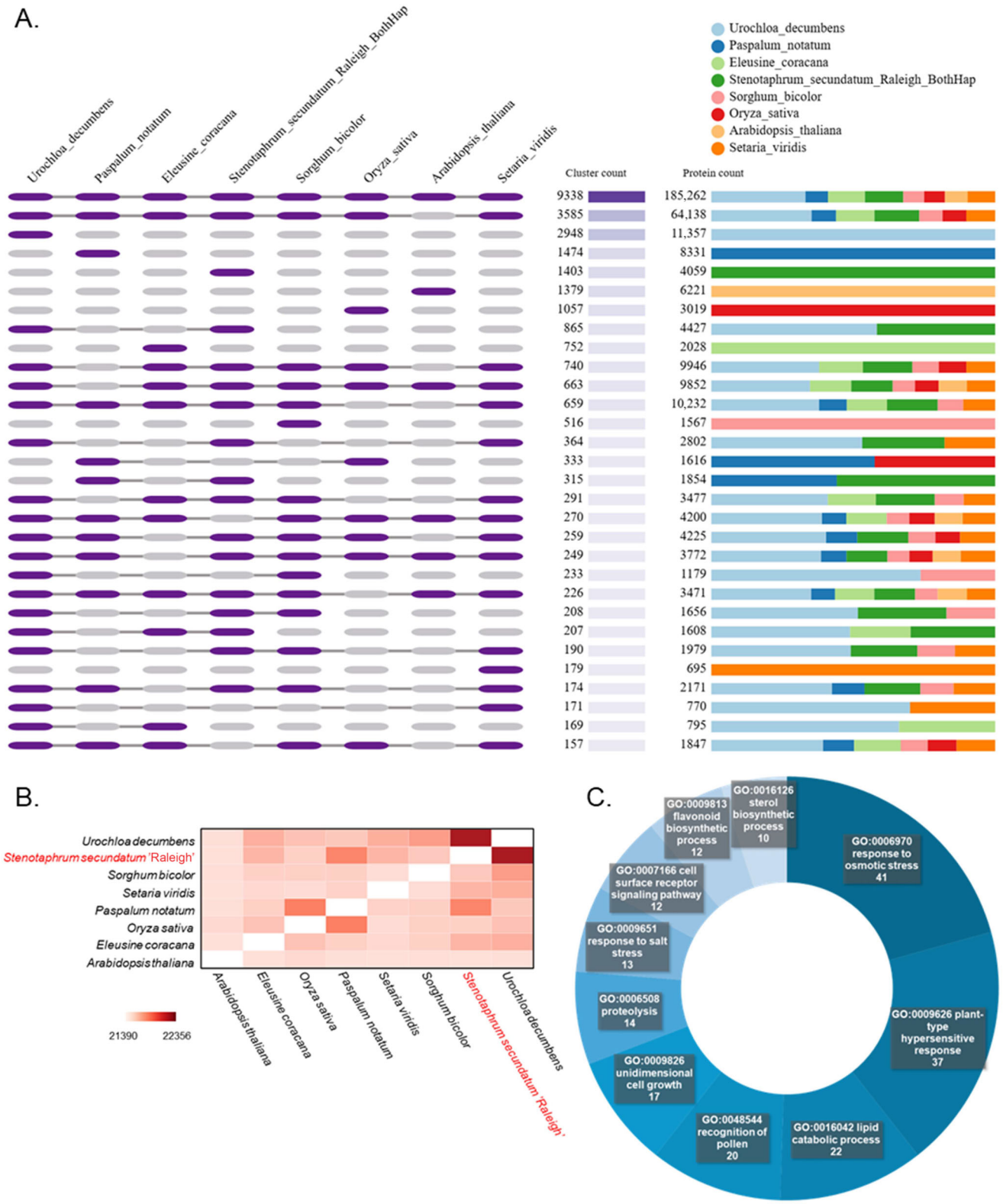



**FIGURE 6** Orthologous analysis results for *Stenotaphrum secundatum* 'Raleigh', six other grass species, and *Arabidopsis*. (A) Occurrence plot showing orthologous clustering results. All genes represented in the plot were assigned to a cluster (singletons are not shown). (B) Similarity matrix based on number of shared orthologous clusters between species. (C) Top 10 enriched biological process gene ontology (GO) terms in *S. secundatum*-only clusters.

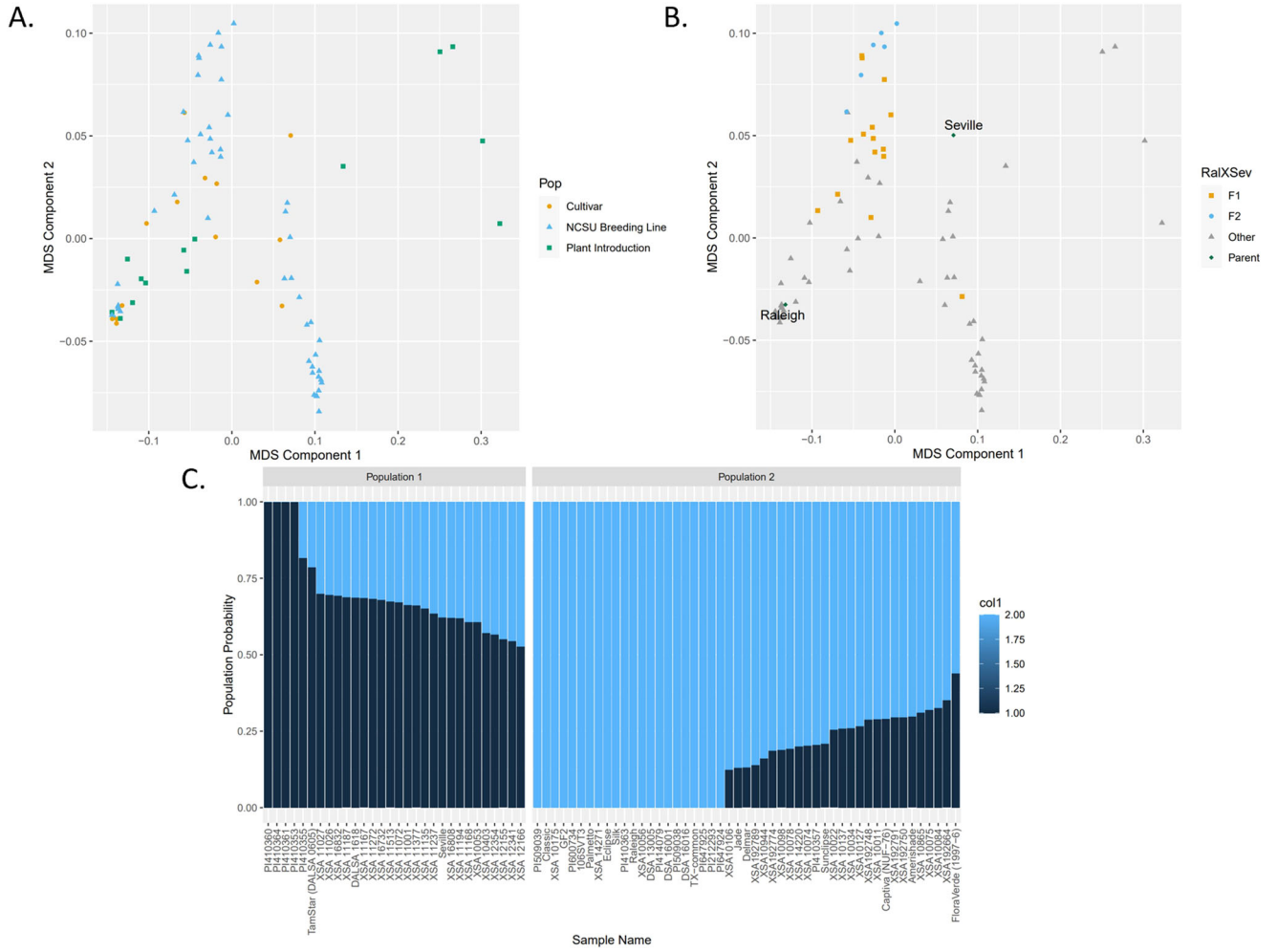



**FIGURE 7** Analysis of the diversity in the panel based on single nucleotide polymorphism (SNP) data. (A) Multidimensional scaling (MDS) plot of the diversity panel. Samples are colored according to their classification as cultivar, breeding line, or plant introduction (PI). (B) MDS plot of the panel with the breeding lines from a controlled cross colored to show the generational relationship with the parents from which they originated. Parents, Raleigh, and Seville are labeled. (C) Population structure plot of the panel derived from FASTSTRUCTURE, shown at $k = 2$ which maximized model complexity. NCSU, North Carolina State University.

quality statistics than two previously published major warm-season turfgrass assemblies, those of African bermudagrass (*Cynodon transvaalensis*) (Cui et al., 2021) and zoysiagrass (*Zoysia japonica*) (Tanaka et al., 2016), with both higher contig lengths in the *de novo* assembly and scaffold lengths in the finished assembly. Although the African bermudagrass and zoysiagrass assemblies met the necessary LAI scores for draft genomes at 5 and 7.20, respectively, the LAI for the Raleigh genome met the requirement for reference genome quality, 17.61 for the primary and 17.36 for the secondary haplotypes, respectively. While we cannot present a fully haplotype-resolved assembly of the genome without access to the parental genomes of the genotype to assist in properly segregating the haplotypes, long-read sequencing has allowed us to assemble phased sequences at the chromosome scale that contain most of the expected genome size. A future project sequencing the progeny of Raleigh could make use of the sequencing data from this project, along with that of the other parent, to achieve a fully phased assembly, either by using parental whole-genome assemblies or by using trio-binning as detailed in Delorean et al. (2023) in the case of only Illumina data being available for the other parent.

We demonstrate alignments to two published linkage maps, where the first linkage map by Yu et al. (2018) was developed from a population of 115 $F_1$ Raleigh × Seville hybrids spanning 1241.7 centimorgans (cM) across the nine groups representing the nine chromosomes with 2871 SNP markers. The second map by Yu et al. (2020) from a Raleigh × PI 410353 population of 153 F1 hybrids with 2257 SNPs spanned 916.63 cM. While the alignment to the Raleigh × Seville map showed a 1:1 correlation between the assembly and the linkage order, the alignment to the Raleigh × PI 410353 was not as good. In order to allow more informative

SNPs to be used in scaffolding, we weighted the map appearing to be higher quality (R × S, where R × S is the Raleigh × Seville linkage map) higher during ALLMAPS scaffolding. We did confirm that allowing ALLMAPS to assign different weights to the maps did not significantly affect the scaffolding. Visualization of the aligned SNPs from the linkage maps showed reasonable coverage across the genome, but contained some gaps including Chr2 for the Raleigh × Seville map and Chr5 for Raleigh × PI 410353 map. The importance of knowing where there are low densities of SNPs from the linkage map in a region causes difficulties in determining the location of traits attributed to the genes in those regions. Density of the annotated genes was fairly consistent across the entire genome, but decreased in the middle of a majority of the chromosomes, which corresponded to higher density in the presence of repeats, indicating the possible location of the centromeres of the chromosomes. A low density of SNPs also coincided with low gene density and high repeat density across the Raleigh chromosomes.

Importantly, *S. secundatum* is believed to have high levels of heterozygosity due to its outcrossing nature, though previously we were unable to quantify that heterozygosity (Mason, 2015; Yu et al., 2022). Here, we were able to quantify the St. Augustinegrass cultivar Raleigh's heterozygosity using the two haploid assemblies. At the assembly level, that is, comparing the primary haplotype to the secondary haplotype, Raleigh showed a heterozygosity level of 0.34% under *k*-mer comparisons by mummer and between 0.94% and 0.96% as estimated with Genomescope. By comparison, mummer predicted 0.12% and Genomescope 0.33%–0.41% heterozygosity for a pepper $F_1$ hybrid (Delorean et al., 2023), whereas other species like cultivated ginger show a much lower level of heterozygosity (0.04%) as estimated from SNP data derived from Illumina reads (S.-P. Cheng et al., 2021). Based on these comparisons, we can reasonably conclude that Raleigh is truly a hybrid with heterozygosity levels higher than many other plant species.

Furthermore, intragenomic comparison of Raleigh haplotypes via SyRI allowed us to determine that this cultivar has an average of one SNP per 194 bp of genome sequence. Many studies have reported SNP density rates across the Poaceae family, but averages around one SNP for every 20–175 bp (Barre et al., 2022); one SNP per 175 bp, 127 bp, and 25–33 bp have been reported for hexaploid wheat (Jordan et al., 2015), switchgrass (Bahri et al., 2018), and Italian ryegrass (Y. Chen et al., 2025), respectively. However, unlike the above comparison, in this case one needs to be careful directly comparing the reported values, as methods utilized to derive these densities vary widely. Overall, however, this indicates St. Augustinegrass Raleigh is nearby the density of other grasses, and further work would need to be done to directly compare the SNP densities directly.

We also explored the structural variations between haplotypes and determined that genome duplications and copy number variations heavily contributed to the intragenomic diversity of St. Augustinegrass Raleigh. Potential future studies could involve investigating and comparing the heterozygosity levels of other well-performing St. Augustinegrass cultivars, as well as the dissection of the duplicated and copy number variation regions identified in this study to determine their biological implications. Additionally, future investigation into potential issues surrounding the sequencing and assembly of chromosome 8 in this species is warranted.

Running annotation on heterozygous individuals like grasses has proven difficult (Brůna et al., 2021). Some of the other methods of filtering, such as those proposed for peanut (J. Li et al., 2022), do not work as well for grasses, leading to an overprediction of genes when compared to other published related genomes (Benson et al., 2023; Doust et al., 2009; Udall et al., 2019). In annotating this genome, we set about choosing a set of annotating and filtering steps that provide reasonable results when compared to annotations of similar species, while also maintaining a high quality of the gene structures. We ran gene prediction with the new version of GeMoMa due to the ability to include RNA-seq data as additional evidence on top of the protein evidence used for homology-based prediction (Keilwagen et al., 2018). The inclusion of *de novo* transcripts assembled with StringTie (Pertea et al., 2015) resulted in more genes predicted at the end of filtering than with GeMoMa alone. Our results were shown to be in line with the annotation results of other heterozygous grass species. To that end, the high proportion of predicted genes in the primary haplotype that contained allelic SNPs (82.8% genes) provided further support to the high heterozygosity reported for St. Augustinegrass. We also utilized orthologous analysis for comparison of protein orthology across multiple grass genomes. Interestingly, St. Augustinegrass shared the highest degree of orthology with *Urochloa decumbens*. This species is taxonomically classified into a different subtribe (Melinidinae), as opposed to the species we expected to share the most orthology with St. Augustinegrass, *Setaria viridis*, which is in the same subtribe, *Stenotaphrum* genus, as St. Augustinegrass (Cenchrinae) (Delfini et al., 2023). Another notable result was that the *S. secundatum*-only clusters were enriched for GO terms involved in stress response and reproduction, possibly indicating an ability for this species or cultivar to more efficiently respond to biotic and abiotic stressors. These results, together with the NLR genes identified in the genome, provide a foundation for future phytopathological studies in St. Augustinegrass.

This is the first study to anchor SNPs to a genome reference in *Stenotaphrum* germplasm. Results of the structural investigation and MDS plots yielded two major clusters with a majority of previously sampled cultivars clustering

similarly to previous work. One notable difference compared with recent previous papers is some cultivars, such as Texas Common, Raleigh, Palmetto, Seville, and Jade, falling within the same cluster and PIs clustering away from cultivars (Milla-Lewis et al., 2013; Mulkey et al., 2014). This difference is likely due to the difference in the panels and marker contribution to the analyses. Previous studies contained representatives from the different ploidy levels present in St. Augustinegrass (Milla-Lewis et al., 2013; Mulkey et al., 2014), whereas this study focused on diploid representatives only. However, while breaking into separate clusters, the mentioned cultivars remain tightly linked to those they were most closely clustered to in previous papers.

Increased understanding of the diversity in a germplasm pool is important to the development of cultivars in an efficient manner. Some of the older cultivars and collections do not have pedigree information, so knowledge of the similarities between germplasm samples assists a breeder in selecting crosses to prevent inbreeding depression. Availability of low-coverage sequencing reads that give enough data to identify relationships and differences introduces a more cost-effective method for researchers to use in future studies. It should be noted that low-quality SNP markers can affect future genetics and breeding work, so it is important that the methods utilized here showed a good correlation with previous studies, and thus a similar SNP discovery pipeline as used here with low-coverage data can generate high-quality variants for future breeding and research. Though a larger panel could dive into a more in-depth analysis of the relationships between the different genotypes across germplasm. Furthermore, it would be useful to determine if the short-read method used in this study can be used in comparing diploids to polyploids.

The availability of a reference genome allows for the identification of the true genomic location of SNPs in linkage maps, and will ultimately assist in the identification of genes underlying QTL related to traits of agronomic interest. Species that are heterogeneous and outcrossing have documented difficulty in development of SNPs with inflated rates of false positives, which affects downstream analyses (Bu et al., 2023; Mason, 2015; You et al., 2012). With the creation of a reference genome, some of these issues can be resolved. This will allow for the expansion of the molecular toolkits that are available to breeders for more efficient selection. Future studies would benefit from sequencing of additional genomes that represent more of the diversity in the species. The genomic toolbox for the species would benefit from the addition of more genotypes to the panel, which would give further insights into the variety of genotypes, genes present, and genomic structural diversity of species.

## 5 | CONCLUSIONS

We present a high-quality, haplotype-resolved genome assembly, the first genome available for an important and highly valuable turfgrass species, *S. secundatum*. As part of this research, we also include a high-quality genome annotation and information on the variation available in breeding material for this species. This study enables the start of a genomics-enabled breeding era for St. Augustinegrass.

## AUTHOR CONTRIBUTIONS

**Ashley N. Schoonmaker**: Data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Ashley G. Yow**: Formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Xingwang Yu**: Investigation; methodology; supervision; writing—review and editing. **Rocio van der Laat**: Investigation; writing—review and editing. **Jeffrey C. Glaubitz**: Data curation; formal analysis; methodology; writing—review and editing. **Kim Thorsted**: Investigation; writing—review and editing. **Matthew D. Robbins**: Methodology; software; visualization; writing—review and editing. B. **Shaun Bushman**: Methodology; software; supervision; writing—review and editing. **Sheron A. Simpson**: Investigation; writing—review and editing. **Brian E. Scheffler**: Investigation; methodology; resources; supervision; writing—review and editing. **Nathan P. Lynch**: Investigation; writing—review and editing. **Thomas G. Ranney**: Investigation; supervision; writing—review and editing. **Susana Milla-Lewis**: Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing—original draft; writing—review and editing. **Amanda M. Hulse-Kemp**: Conceptualization; formal analysis; funding acquisition; methodology; project administration; resources; software; supervision; writing—original draft; writing—review and editing.

## ACKNOWLEDGMENTS

The authors would like to thank Drs. Dongyan Zhao, Meng Lin, and Moira Sheehan from Breeding Insight for their thoughts and support on RipTide sequencing. We would like to thank the Phytozome team of Sumaira Zaman, Tomas Bruna and David Goodstein for assisting with making the resources available on the Phytozome database. The project is supported by the USDA Specialty Crops Research Initiative, Project 2019-51181-30472 (XY, SM-L, and AMH-K). Additional support was provided by USDA Agricultural Research Service CRIS Projects 6066-21310-005-00D and 2080-21000-018-000-D. Support for RipTide sequencing was provided by Breeding Insight (RRID: SCR_026645), a USDA-ARS initiative previously hosted by Cornell

University under Cooperative Agreements (8062-21000-043-004-A, 8062-21000-052-002-A, and 8062-21000-052-003-A) and currently hosted at the University of Florida, Gainesville, under a Cooperative Agreement (8062-21000-052-020-A).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

All data related to reference genome sequencing of *S. secundatum* 'Raleigh' can be found in NCBI under BioProject PRJNA736077 for the primary assembly and PRJNA980716 for the alternate assembly. The final genome assembly containing the primary and secondary haplotypes respectively can be found at NCBI (JAVFYC000000000 and JAVFYD000000000) and at Phytozome: https://phytozome-next.jgi.doe.gov/info/961. Raw whole genome sequencing data for all individuals was uploaded to NCBI under BioProject PRJNA1007876. PacBio CCS reads were deposited in the NCBI short read archive (SRA) under BioProject PRJNA736077. Raw Hi-C reads were deposited in NCBI SRA under accession SRR25592757. Short genomic reads were deposited in the NCBI SRA under accession SRR25691094. Raw RNA-seq reads were deposited in NCBI SRA under accession numbers: SRR25534696, SRR25534684, SRR25534688, SRR25534682, SRR25534687, SRR25534697, SRR25534678, SRR25534692, SRR25534681, SRR25534679, SRR25534680, SRR25534690, SRR25534694, SRR25534685, SRR25534693, SRR25534686, SRR25534691, SRR25534695, SRR25534683, and SRR25534689. Variant data in the form of variant call format file was deposited at the USDA AgDataCommons (10.15482/USDA.ADC/25234093). Additional details including code and parameters used in this project are available at: https://github.com/USDA-ARS-GBRU/Ssec_Raleigh/, specifically access to annotation code is available at: https://github.com/USDA-ARS-GBRU/Grass_annotation_pipeline/.

## ORCID

*Ashley N. Schoonmaker* https://orcid.org/0009-0006-0926-6042
*Ashley G. Yow* https://orcid.org/0000-0002-2498-0133
*Xingwang Yu* https://orcid.org/0000-0003-0382-9474
*Rocio van der Laat* https://orcid.org/0009-0006-2357-9248
*Jeffrey C. Glaubitz* https://orcid.org/0009-0009-4086-1853
*Matthew D. Robbins* https://orcid.org/0000-0002-5467-4452
*B. Shaun Bushman* https://orcid.org/0000-0003-3489-1668
*Sheron A. Simpson* https://orcid.org/0000-0003-0288-3452
*Brian E. Scheffler* https://orcid.org/0000-0003-1968-8952
*Nathan P. Lynch* https://orcid.org/0000-0001-8311-9063
*Thomas G. Ranney* https://orcid.org/0000-0001-8201-9984
*Susana Milla-Lewis* https://orcid.org/0000-0001-8524-5039
*Amanda M. Hulse-Kemp* https://orcid.org/0000-0001-9670-9433

## REFERENCES

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, *23*(1), 258. https://doi.org/10.1186/s13059-022-02823-7

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arumuganathan, K., Tallury, S. P., Fraser, M. L., Bruneau, A. H., & Qu, R. (1999). Nuclear DNA content of thirteen turfgrass species by flow cytometry. *Crop Science*, *39*(5), 1518–1521. https://doi.org/10.2135/cropsci1999.3951518x

Bahri, B. A., Daverdin, G., Xu, X., Cheng, J.-F., Barry, K. W., Brummer, E. C., & Devos, K. M. (2018). Natural variation in genes potentially involved in plant architecture and adaptation in switchgrass (*Panicum virgatum* L.). *BMC Evolutionary Biology*, *18*(1), Article 91. https://doi.org/10.1186/s12862-018-1193-2

Barre, P., Asp, T., Byrne, S., Casler, M., Faville, M., Rognli, O. A., Roldan-Ruiz, I., Skøt, L., & Ghesquière, M. (2022). Genomic prediction of complex traits in forage plants species: Perennial grasses case. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic prediction of complex traits* (Vol., 2467, pp. 521–541). Springer. https://doi.org/10.1007/978-1-0716-2205-6_19

Bateman, D. R. (1980). *Notice to sod producers and growers relative to the naming and release of the new St. Augustine cultivar 'Raleigh'*. North Carolina Agricultural Research Service.

Benson, C. W., Sheltra, M. R., Maughan, P. J., Jellen, E. N., Robbins, M. D., Bushman, B. S., Patterson, E. L., Hall, N. D., & Huff, D. R. (2023). Homoeologous evolution of the allotetraploid genome of *Poa annua* L. *BMC Genomics*, *24*(1), 350. https://doi.org/10.1186/s12864-023-09456-5

Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., Pinto, B. L., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G. A., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunic, I., Llinares-López, F., … Bateman, A. (2025). InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Research*, *53*(D1), D444–D456. https://doi.org/10.1093/nar/gkae1082

Brown, M. R., Manuel Gonzalez De La Rosa, P., & Blaxter, M. (2025). tidk: A toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, *41*(2), btaf049. https://doi.org/10.1093/bioinformatics/btaf049

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, *98*(1), 116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with

GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

Bu, M., Xu, M., Tao, S., Cui, P., & He, B. (2023). Evaluation of different SNP analysis software and optimal mining process in tree species. *Life*, *13*(5), 1069. https://doi.org/10.3390/life13051069

Busey, P. (1986). Morphological identification of St. Augustinegrass cultivars. *Crop Science*, *26*(1), 28–32. https://doi.org/10.2135/cropsci1986.0011183X002600010006x

Busey, P., Broschat, T. K., & Center, B. J. (1982). Classification of St. Augustinegrass. *Crop Science*, *22*, 469–473. https://doi.org/10.2135/cropsci1982.0011183X002200030008x

Bushnell, B. (2014). *BBTools software package*. http://sourceforge.net/projects/bbmap

Caballero, M., & Wegrzyn, J. (2019). gFACs: Gene filtering, analysis, and conversion to unify genome annotations across alignment and gene prediction frameworks. *Genomics, Proteomics & Bioinformatics*, *17*(3), 305–310. https://doi.org/10.1016/j.gpb.2019.04.002

Carbajal Melgar, E. M. (2017). *Breeding for improved St. Augustinegrass through ploidy manipulation and identification of gray leaf spot-resistant germplasm* [Master's thesis]. North Carolina State University.

Casler, M. D., & Duncan, R. R. (2003). *Turfgrass biology, genetics, and breeding*. Wiley.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Y., Kiesbauer, J., Copetti, D., Frei, D., Frey, J. E., Grieder, C., Kölliker, R., & Studer, B. (2025). Chromosome-level haplotype-resolved genome assembly provides insights into the highly heterozygous genome of Italian ryegrass (*Lolium multiflorum* Lam.). *The Plant Genome*, *18*(3), e70079. https://doi.org/10.1002/tpg2.70079

Cheng, H., & al, E. (2020). *Hifiasm: A haplotype-resolved assembler for accurate Hifi reads*. https://www.nature.com/articles/s41592-020-01056-5

Cheng, S.-P., Jia, K.-H., Liu, H., Zhang, R.-G., Li, Z.-C., Zhou, S.-S., Shi, T.-L., Ma, A.-C., Yu, C.-W., Gao, C., Cao, G.-L., Zhao, W., Nie, S., Guo, J.-F., Jiao, S.-Q., Tian, X.-C., Yan, X.-M., Bao, Y.-T., Yun, Q.-Z., … Mao, J.-F. (2021). Haplotype-resolved genome assembly and allele-specific gene expression in cultivated ginger. *Horticulture Research*, *8*, 188. https://doi.org/10.1038/s41438-021-00599-8

Cui, F., Taier, G., Li, M., Dai, X., Hang, N., Zhang, X., Wang, X., & Wang, K. (2021). The genome of the warm-season turfgrass African bermudagrass (*Cynodon transvaalensis*). *Horticulture Research*, *8*(1), 93. https://doi.org/10.1038/s41438-021-00519-w

Delfini, C., Aliscioni, S. S., Acosta, J. M., Pensiero, J. F., & Zuloaga, F. O. (2023). An update of the Cenchrinae (Poaceae, Panicoideae, Paniceae) and a new genus for the subtribe to clarify the dubious position of a species of *Panicum* L. *Plants*, *12*(4), 749. https://doi.org/10.3390/plants12040749

Delorean, E. E., Youngblood, R. C., Simpson, S. A., Schoonmaker, A. N., Scheffler, B. E., Rutter, W. B., & Hulse-Kemp, A. M. (2023). Representing true plant genomes: Haplotype-resolved hybrid pepper genome with trio-binning. *Frontiers Plant Science*, *14*, 1184112. https://doi.org/10.3389/fpls.2023.1184112

Doust, A. N., Kellogg, E. A., Devos, K. M., & Bennetzen, J. L. (2009). Foxtail millet: A sequence-driven grass model system. *Plant Physiology*, *149*(1), 137–141. https://doi.org/10.1104/pp.108.129627

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

Garrison, E., & Marth, G. (2010). *Haplotype-based variant detection from short-read sequencing*. arXiv.

Genovesi, A. D., Jessup, R. W., Engelke, M. C., & Burson, B. L. (2009). Interploid St. Augustinegrass [*Stenotaphrum secundatum* (Walt.) Kuntze] hybrids recovered by embryo rescue. *In Vitro Cellular and Developmental Biology—Plant*, *45*(6), 659–666. https://doi.org/10.1007/s11627-009-9230-x

Goel, M., Sun, H., Jiao, W.-B., & Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, *20*(1), Article 277. https://doi.org/10.1186/s13059-019-1911-0

Graham, S. E., Tome Gouveia, B., Carbajal, E. M., van der Laat, R., & Milla-Lewis, S. R. (2022). Using base index for selection of St. Augustinegrass breeding lines evaluated in multienvironment trials for turfgrass quality traits and stress tolerance in North Carolina. *Crop Science*, *62*(5), 2011–2023. https://doi.org/10.1002/csc2.20755

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

Hart, A. J., Ginzburg, S., Xu, M. S., Fisher, C. R., Rahmatpour, N., Mitton, J. B., Paul, R., & Wegrzyn, J. L. (2020). EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources*, *20*(2), 591–604. https://doi.org/10.1111/1755-0998.13106

Henry, L., & Wickham, H. (2020). Purrr: Functional Programming Tools (R package Version 0.3.4). [Computer software]. CRAN.

Hodges, A. W., Haydu, J. J., van Blokland, P. J., & Bell, A. P. (1994). *Contribution of the turfgrass industry to Florida's economy, 1991–92: A value-added approach* (ER 94-1 92). University of Florida.

Horn, G. C., Dudeck, A. E., & Toler, R. W. (1973). *'Floratam' St. Augustinegrass: A fast growing new variety for ornamental turf resistant to decline and chinch bugs* (Circular No. S-224). Florida Agricultural Experiment Station, University of Florida.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K. L., Consortium, I., Sharpe, A. G., Sidebottom, C. H., Hall, N., Toomajian, C., Close, T., Dubcovsky, J., Akhunova, A., Talbert, L., Bansal, U. K., … Akhunov, E. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology*, *16*(1), Article 48. https://doi.org/10.1186/s13059-015-0606-4

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., & Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, *19*(1), 189–189. https://doi.org/10.1186/s12859-018-2203-5

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kimball, J. A., Isleib, T. G., Reynolds, W. C., Zuleta, M. C., & Milla=Lewis, S. R. (2016). Combining ability for winter survival and turf quality traits in St. Augustinegrass. *HortScience*, *51*, 810–815. https://doi.org/10.21274/HORTSCI.51.7.810

Kimball, J. A., Tuong, T. D., Arellano, C., Livingston, D. P., & Milla-Lewis, S. R. (2018). Linkage analysis and identification of quantitative trait loci associated with freeze tolerance and turf quality traits in St. Augustinegrass. *Molecular Breeding*, *38*, Article 67. https://doi.org/10.1007/s11032-018-0817-y

Kourelis, J., Sakai, T., Adachi, H., & Kamoun, S. (2021). RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. *PLoS Biology*, *19*(10), e3001124. https://doi.org/10.1371/journal.pbio.3001124

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*, Article R12. https://doi.org/10.1186/gb-2004-5-2-r12

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Singh, U., Bhandary, P., Campbell, J., Arendsee, Z., Seetharam, A. S., & Wurtele, E. S. (2022). Foster thy young: Enhanced prediction of orphan genes in assembled genomes. *Nucleic Acids Research*, *50*(7), e37. https://doi.org/10.1093/nar/gkab1238

Long, J. A., & Bashaw, E. C. (1961). Microsporogenesis and chromosome numbers in St. Augustinegrass. *Crop Science*, *1*(1), 41–43. https://doi.org/10.2135/cropsci1961.0011183x000100010013x

Ludwig, A., Pippel, M., Myers, G., & Hiller, M. (2022). DENTIST—Using long reads for closing assembly gaps at high accuracy. *GigaScience*, *11*, giab100.

Mapleson, D., Venturini, L., Kaithakottil, G., & Swarbreck, D. (2018). Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience*, *7*(12), giy131. https://doi.org/10.1093/gigascience/giy131

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mason, A. S. (2015). Challenges of genotyping polyploid species. In J. Batley (Ed.), *Plant genotyping: Methods and protocols* (pp. 161–168). Springer. https://doi.org/10.1007/978-1-4939-1966-6_12

McBee, G. G., & Holt, E. C. (1966). Shade Tolerance studies on bermudagrass and other turfgrasses. *Agronomy Journal*, *58*(5), 523–525. https://doi.org/10.2134/agronj1966.00021962005800050021x

McCoy, N. L., Toler, R. W., & Horne, C. W. (1970). St. *Augustine decline (SAD): A virus disease of St. Augustinegrass* (L-940). Texas Agricultural Extension Service, Texas A&M University.

Milla-Lewis, S. R., Zuleta, M. C., Van Esbroeck, G. A., Quesenberry, K. H., & Kenworthy, K. E. (2013). Cytological and molecular characterization of genetic diversity in *Stenotaphrum*. *Crop Science*, *53*(1), 296–308. https://doi.org/10.2135/cropsci2012.04.0234

Mulkey, S. E., Zuleta, M. C., Keebler, J. E., Schaff, J. E., & Milla-Lewis, S. R. (2014). Development and characterization of simple sequence repeat markers for St. Augustinegrass. *Crop Science*, *54*, 401–412. https://doi.org/10.2135/cropsci2013.04.0246

Nevers, Y., Warwick Vesztrocy, A., Rossier, V., Train, C.-M., Altenhoff, A., Dessimoz, C., & Glover, N. M. (2025). Quality assessment of gene repertoire annotations with OMArk. *Nature Biotechnology*, *43*(1), 124–133. https://doi.org/10.1038/s41587-024-02147-w

Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., & Koren, S. (2020). HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *BioRxiv*. https://doi.org/10.1101/2020.03.14.992248

Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research*, *46*, e126. https://doi.org/10.1093/nar/gky730

Ou, S., & Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, *176*(2), 1410–1422. https://doi.org/10.1104/pp.17.01310

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, *20*(1), Article 275. https://doi.org/10.1186/s13059-019-1905-y

Paysan-Lafosse, T., Andreeva, A., Blum, M., Chuguransky, S. R., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Llinares-López, F., Meng-Papaxanthos, L., Colwell, L. J., Grishin, N. V., Schaeffer, R. D., Clementel, D., Tosatto, S. C. E., Sonnhammer, E., Wood, V., & Bateman, A. (2025). The Pfam protein families database: Embracing AI/ML. *Nucleic Acids Research*, *53*(D1), D523–D534. https://doi.org/10.1093/nar/gkae997

Pedersen, T. L., Pedersen, M. T. L., LazyData, T., Rcpp, I., & Rcpp, L. (2020). ggforce: Accelerating ggplot2 (R package Version 0.3) [Computer software]. CRAN.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Piper, C. V. (1919). Rhizoctonia in lawns and pastures. *Phytopathology*, *9*, 89–92.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American*

*Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org10.1534/genetics.114.164350

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), Article 1432. https://doi.org/10.1038/s41467-020-14998-3

Reinert, J. A., & Kerr, S. H. (1973). Bionomics and control of lawn chinch bugs. *Bulletin of the Entomological Society of America*, *19*(2), 91–92.

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality and phasing assessment for genome assemblies. *BioRxiv*. https://doi.org/10.1101/2020.03.15.992941

RStudio. (2020). RStudio: Integrated development for R. RStudio, PBC. http://www.rstudio.com

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., & Allard, R. W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences*, *81*(24), 8014–8018. https://doi.org/10.1073/pnas.81.24.8014

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slowikowski, K. (2021). *ggrepel: Automatically position non-overlapping text labels with 'ggplot2'* . (R package version 0.9.1). [Computer software]. CRAN.

Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H., Yu, G., Baggs, E., Witek, A. I., Yadav, I., Krasileva, K. V., Jones, J. D. G., Uauy, C., Keller, B., Ridout, C. J., & Wulff, B. B. H. (2020). The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiology*, *183*(2), 468–482. https://doi.org/10.1104/pp.19.01273

Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., & Wang, Y. (2023). OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research*, *51*(W1), W397–W403. https://doi.org/10.1093/nar/gkad313

Tanaka, H., Hirakawa, H., Kosugi, S., Nakayama, S., Ono, A., Watanabe, A., Hashiguchi, M., Gondo, T., Ishigaki, G., Muguerza, M., Shimizu, K., Sawamura, N., Inoue, T., Shigeki, Y., Ohno, N., Tabata, S., Akashi, R., & Sato, S. (2016). Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Research*, *23*(2), 171–180. https://doi.org/10.1093/dnares/dsw006

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., & Lu, J. (2015). ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biology*, *16*(1), Article 3. https://doi.org/10.1186/s13059-014-0573-1

Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M., The Gene Ontology Consortium. (2023). The gene ontology knowledgebase in 2023. *Genetics*, *224*(1), iyad031. https://doi.org/10.1093/genetics/iyad031

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Udall, J. A., Long, E., Hanson, C., Yuan, D., Ramaraj, T., Conover, J. L., Gong, L., Arick, M. A., Grover, C. E., Peterson, D. G., & Wendel, J. F. (2019). De novo genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. *G3: Genes, Genomes, Genetics*, *9*(10), 3079–3085. https://doi.org/10.1534/g3.119.400392

Vendelbo, N. M., Mahmood, K., Steuernagel, B., Wulff, B. B. H., Sarup, P., Hovmøller, M. S., Justesen, A. F., Kristensen, P. S., Orabi, J., & Jahoor, A. (2022). Discovery of resistance genes in rye by targeted long-read sequencing and association genetics. *Cells*, *11*(8), 1273. https://doi.org/10.3390/cells11081273

Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L., & Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, *7*(8), giy093. https://doi.org/10.1093/gigascience/giy093

Villanueva, R. A. M., & Chen, Z. J. (2019). *ggplot2: Elegant graphics for data analysis*. Taylor & Francis.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wickham, H. (2019). Stringr: Simple, Consistent Wrappers For Common String Operations (R package Version 1.4.0). CRAN.

Wickham, H. (2021). Forcats: Tools For Working With Categorical Variables (Factors) (R package Version 0.5. 1) [Computer software]. CRAN.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). Dplyr: A Grammar of Data Manipulation (R package Version 0.7. 6) [Computer software]. CRAN. https://CRAN.R-Project.Org/Package=Dplyr

Wickham, H., Hester, J., Francois, R., Jylänki, J., & Jørgensen, M. (2017). Readr: Read rectangular text data (R package Version 1.1.1) [Computer software]. R Foundation for Statistical Computing.

Wickham, H., & Wickham, M. H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. (R package Version 1.1.1). [Computer software]. CRAN.

Winstead, C. W., & Ward, C. Y. (1974). Persistence of southern turfgrasses in a shade environment. In E. C. Roberts (Ed.), *Proceedings of the Second International Turfgrass Research Conference* (pp. 221–230). ASA, CSSA, SSSA.

Xia, E., Li, F., Tong, W., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., Tai, Y., She, G., Sun, J., Cao, H., Gao, Q., Li, Y., Deng, W., Jiang, X., Wang, W., Chen, Q., Zhang, S., … Wan, X. (2019). The tea plant reference genome and improved gene annotation using long-read and paired-end sequencing data. *Scientific Data*, *6*(1), Article 122. https://doi.org/10.1038/s41597-019-0127-1

You, F. M., Deal, K. R., Wang, J., Britton, M. T., Fass, J. N., Lin, D., Dandekar, A. M., Leslie, C. A., Aradhya, M., & Luo, M.-C. (2012). Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. *BMC Genomics*, *13*(1), Article 354. https://doi.org/10.1186/1471-2164-13-354

Yu, X., Kimball, J. A., & Milla-lewis, S. R. (2018). High density genetic maps of St. Augustinegrass and applications to comparative genomic analysis and QTL mapping for turf quality traits. *BMC Plant Biology*, *18*, Article 346. https://doi.org/10.1186/s12870-018-1554-4

Yu, X., Lara, N. A. H., Carbajal, E. M., & Milla-Lewis, S. R. (2022). QTL mapping of morphological characteristics that correlated to drought tolerance in St. Augustinegrass. *PLoS ONE*, *17*(5), e0268004. https://doi.org/10.1371/journal.pone.0268004

Yu, X., Mulkey, S. E., Zuleta, M. C., Arellano, C., Ma, B., & Milla-Lewis, S. R. (2020). Quantitative trait loci associated with gray leaf spot resistance in St. Augustinegrass. *Plant Disease*, *104*(11), 2799–2806. https://doi.org/10.1094/PDIS-04-20-0905-RE

Zhang, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S., & Ma, Y. (2022). TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. *Horticulture Research*, *9*, uhac017. https://doi.org/10.1093/hr/uhac017

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Schoonmaker, A. N., Yow, A. G., Yu, X., van der Laat, R., Glaubitz, J. C., Thorsted, K., Robbins, M., Bushman, S., Simpson, S. A., Scheffler, B. E., Lynch, N., Ranney, T. G., Milla-Lewis, S., & Hulse-Kemp, A. M. (2026). A whole-genome assembly of St. Augustinegrass and visualizing diversity within the species. *The Plant Genome*, *19*, e70144. https://doi.org/10.1002/tpg2.70144