

A haplotype-complete chromosome-level assembly of octoploid *Urochloa humidicola* cv. Tully reveals multiple genomic compositions and evolutionary histories in the species

Jose J. De Vega ^{1,*}, Alex Durrant,¹ Naomi Irish,¹ Tom Barker,¹ Rosa N. Jauregui ²

¹Earlham Institute, Norwich Bioscience Institutes (NBI), Norwich Research Park NR4 7UZ, United Kingdom

²Alliance of Biodiversity International and CIAT, Km 17 Recta Cali Palmira, Cali AA 6713, Colombia

*Corresponding author: Earlham Institute, Norwich Bioscience Institutes (NBI), Norwich Research Park NR4 7UZ, United Kingdom. Email: jose.de-vega@earlham.ac.uk

We developed a haplotype-resolved, chromosome-scale genome assembly of the *Urochloa humidicola* (Rendle) Morrone & Zuloaga cultivar Tully, an apomictic C4 forage grass cultivated in the tropics worldwide. We assembled a 4.1 Gb genome into 48 chromosomes ($2n = 8x = 48$), capturing 99.5% of the BUSCO markers and annotating 259,254 protein-coding genes. Subgenome assignment revealed an octoploid AABBBC structure with 3 ancestral lineages (A, B, C), and an aneuploid composition of 14A, 22B, and 12C chromosomes. Comparative analyses with related *Urochloa* species identified *U. dictyoneura* and *U. arrecta* as potential progenitors of the B and C subgenomes, respectively. The likely progenitor of the A subgenome remains an unknown wild species from the Humidicola clade. Analysis of long terminal repeats (LTR) retrotransposons and gene collinearity further indicated a close relationship between A and B ancestries and a distinct evolutionary path for C. Competitive read mapping across additional *U. humidicola* accessions supported multiple evolutionary histories in the species, with ABBBBB (lacking C ancestry), being the most common. We found that previously described subpopulation structures can be explained by the presence or absence of C ancestry, and that sexual *U. humidicola* are likely to be autopolyploid from the B ancestry. The genome is available as assembly GCA_965614515.2. This assembly provides the first complete reference for *U. humidicola* and reveals a multi-ancestral origin and reticulated evolution in *U. humidicola*. It provides a foundation for studying complex polyploid evolution, regulation of apomixis and biological nitrification inhibition, and molecular breeding strategies.

Keywords: *Urochloa humidicola*; genome assembly; polyploid; apomixis

Introduction

Urochloa humidicola is a perennial C4 tropical forage grass of major economic and ecological importance in tropical livestock systems. Native to African savannas and now widely cultivated across Latin America, Africa, Asia, and Oceania, it is particularly valued for its persistence in low-fertility and waterlogged soils, and tolerance to grazing (Ferreira et al. 2021). Despite its agronomic success, *U. humidicola* poses exceptional challenges for genetic improvement. The species exhibits complex polyploidy (Higgins et al. 2022; Tomaszewska et al. 2023), with cytotypes ranging from hexaploid to nonaploid ($2n = 6x-9x = 36-54$).

U. humidicola reproduces predominantly through gametophytic apomixis, producing clonal seeds without fertilization (da Costa Lima Moraes et al. 2023). Apomixis has long been regarded as a transformative trait that could fix heterosis through clonal seed production, enabling elite hybrid genotypes in important seed crops to be perpetually reproduced through biotechnological approaches (Underwood and Mercier 2022). *U. humidicola* and related *Urochloa* species serve as an excellent model for exploring the genomic basis of apomixis in polyploid grasses (Cervantes-Díaz et al. 2024).

Furthermore, *U. humidicola* exhibits biological nitrification inhibition (BNI) activity, a unique physiological trait that suppresses soil nitrification and mitigates nitrogen losses following fertilizing (Nuñez et al. 2018; Egenolf et al. 2020; Vázquez et al. 2020). Understanding the genetic basis of BNI offers a promising route to develop biological inhibitors to reduce nitrogen leakage in arable systems (Issifu et al. 2024).

U. humidicola holds significant potential for climate-smart livestock production. Molecular breeding, such as marker-assisted and genomic selection, can accelerate the development of cultivars with improved nutritional quality and increased productivity (Ferreira et al. 2021). The intensification of livestock systems can reduce the land footprint of pasture expansion in the tropics, and enhanced fiber digestibility in *Urochloa* forages has been linked to lower methane emissions from ruminants, aligning breeding goals for this species with global targets to decrease greenhouse gas emissions from agriculture (Arango et al. 2020).

Here, we report a haplotype-complete, chromosome-level genome assembly of the apomictic *U. humidicola* cv. Tully. This reference genome provides the first comprehensive view of the subgenomic structure and evolutionary relationships within this

complex polyploid species. It represents a valuable resource for the forage genomics community, enabling future studies on genome evolution, the genetic control of apomixis and BNI, and the deployment of genomic tools for breeding improved climate-resilient forages.

Methods and materials

Sample collection, DNA extraction, and sequencing

U. humidicola cv. Tully plants were grown for 8 wk before DNA was extracted from seeds of accession APG 58384, obtained from the Australian Pastures Genebank. Tully is an apomictic genotype found worldwide; it was introduced into Australia in 1952 by J.F. Miles as accession CPI 16707 (AusTRCF 16707) from Rietondale Experimental Stations, Pretoria, South Africa (−25.7283, 28.236, 1,300 m elevation, 700 mm rainfall). It was later introduced to Fiji and Papua New Guinea; from there, clonal material was reintroduced from Papua New Guinea to the Tully region of North Queensland in 1973, and it was officially registered for commercial release as “Tully” in 1981 (Miles 1996). The seed of cv. Tully (APG 58384) used for sequencing comes from this reintroduced material. In America, the cultivated material is known as “pasto humidicola” or “Chetumal” (CIAT 679, PI 499378, BRA 001627, INIAP-NAPO 701), which are apomictic clones of the sequenced accession.

After 8 wk, leaf material was harvested following 2 d in dark conditions. Extraction of high molecular weight (HMW) DNA was performed using the Nucleon PhytoPure kit, with a variation of the recommended protocol optimized for DNA fragment size, yield, and purity. 1 g snap-frozen leaf tissue was ground under liquid nitrogen for 9 min, with the mortar refilled with LN2 between every 90 s of grinding. The resulting powder was thoroughly mixed into Reagent 1 (with 4 µl of 100 mg/ml RNase A) in a 15 ml tube using a bacterial spreader loop. The sample was incubated in a ThermoMixer at 37 °C for 30 min. Reagent 2 was then added, and the tube was gently inverted until the contents appeared homogeneous, at which point the tube was incubated in a ThermoMixer at 65 °C for 10 min.

Following the post-lysis ice incubation, 2 ml of chloroform and 300 µl of resin were added, and the sample was mixed on a 3D platform rocker for 10 min at room temperature. Subsequently, the sample was centrifuged at 1,300 g for 10 min. The upper phase was mixed with 1 volume of 25:24:1 phenol:chloroform:isoamyl alcohol, mixed on a 3D platform rocker for 10 min at 4 °C, and centrifuged (3,000 g) for 10 min. The upper phase was transferred to another 15 ml Falcon tube and precipitation proceeded as recommended by the manufacturer's protocol. The final elution was left open in the chemical safety cabinet for 1.5 h to allow residual ethanol to evaporate, after which the DNA sample was left at room temperature overnight.

11.5 µg of gDNA was manually sheared with the Megaruptor 3 instrument (Diagenode, P/N B06010003) according to the Megaruptor 3 operations manual, with a speed setting of 31 and input of 18 ng/µl in 150 µl. The sheared sample then underwent AMPure® PB bead (PacBio®, P/N 100-265-900) purification and concentration (0.45 $\times$, eluted in 95 µl of EB) before undergoing library preparation using the SMRTbell® Express Template Prep Kit 2.0 (PacBio®, P/N 100-983-900). The HiFi library was prepared according to the HiFi protocol version 03 (PacBio®, P/N 101-853-100), and the final library was size fractionated using the SageELF® system (Sage Science®, P/N ELF0001), 0.75% cassette (Sage Science®, P/N ELD7510), with a 4.55 h program. The

library was quantified by fluorescence (Invitrogen Qubit™ 3.0, P/N Q33216), and the size of the library fractions was estimated from a smear analysis performed on the FEMTO Pulse® System (Agilent, P/N M5330AA).

The loading calculations for sequencing were completed using the PacBio® SMRT®Link Binding Calculator v 11.1.0.166339. The 2 largest ELF fractions proceeded to sequencing as separate libraries, and Sequencing primer 3.2 was annealed to the adapter sequence of each HiFi library. Each library was bound to the sequencing polymerase with the Sequel® II Binding Kit v3.2. Calculations for primer and polymerase binding ratios were kept at default values for the library type. Sequel® II DNA internal control 3.2 was spiked into each fraction at the standard concentration before sequencing. The sequencing chemistry used was Sequel® II Sequencing Plate 2.0 (PacBio®, P/N 101-820-200) and the Instrument Control Software v11.0.1.162970. The library was sequenced on the Sequel IIe across 3 Sequel II SMRT®cells 8 M. The parameters for sequencing per SMRT cell were adaptive loading (set to 0.85 target for 2 h), 30-h movie, 2-h pre-extension time, and 90pM on plate loading concentration.

Genome assembly and quality check

Unitigs and contigs were assembled from the raw HiFi reads using HiFiasm v0.20 (Cheng et al. 2021). The quality of the assembly was assessed using several measures; contiguity was evaluated with Quast v. 5.0.2 (Gurevich et al. 2013), assembly completeness was assessed with BUSCO v5.3.2 analysis (Seppey et al. 2019) using the poales v10 database and OMArk v0.3.0 (Nevers et al. 2025) using the whole OMA database (LUCA.h5); and Merquy v1.3 (Rhie et al. 2020) was used to generate kmer completeness metrics and kmer spectra plots to confirm the assembly captured all the content from the reads.

Scaffolding through synteny within the genome

Firstly, the contigs from each haplotype assembled by HiFiasm, haplotype 1 (hap1), and haplotype 2 (hap2) were compared separately using RagTag v2.1.0 (Alonge et al. 2022) to the “primary contigs” produced by HiFiasm. These “primary contigs” were constructed by HiFiasm with the criteria of maximizing continuity, resulting in longer sequences but mixed haplotypes. For each hap1 or hap2, the PAF file internally generated by RagTag (by running Minimap 2.22 with the “-x asm5” option) was visualized as a dotplot with dotPlotly, using a minimum alignment length of 10 kb (Supplementary file 1). Links in the AGP file with query or target having more than one alignment were manually filtered out.

Secondly, we compared the resulting hap1 and hap2 to each other using RagTag, as before. Then generated a new dotplot from the alignment file and manually filtered any entries from the AGP where the query or target had more than one alignment.

Thirdly, we created a synthetic genome of *Urochloa fusca* with 6 chromosomes (instead of the real 9) by combining chromosomes 2 and 4, chromosomes 5 and 6, and chromosomes 1 and 7 from the available *U. fusca* genome v1.0 downloaded from Phytozome 13. The *humidicola* clade has a base chromosome number of 6, unlike most Paniceae clades, which have 9. The $x = 6$ base number in *humidicola* is well established (Jungmann et al. 2010; Vigna et al. 2016). A previous analysis (Worthington et al. 2019) using a genetic map from a biparental population to compare *U. humidicola* to *Setaria italica* ($x = 9$), a Paniceae reference. This analysis revealed the 3 chromosome fusion events in *U. humidicola* (1 + 7, 2 + 4, 5 + 6). We then compared haplotype 1 and haplotype 2 from the previous step to this synthetic genome and selected the *U. humidicola*

sequence that showed the longest alignment to each chromosome with the synthetic *U. fusca* genome. Finally, we compared the hap1 and hap2 resulting from the second step against the 6 “longest” sequences from *U. humidicola* using RagTag, following the same procedure as before; first analyzing each haplotype independently, then together. In total, 225 contigs were scaffolded into 48 chromosomes. Dotplots comparing the final genome to different relatives were done similarly with Minimap v2.22 and dotPlotly.

Gene annotation

The genome was soft-masked using the EI-Repeat pipeline (EI-CoreBioinformatics; <https://github.com/EI-CoreBioinformatics/eirepeat>). The gene annotation was completed with Braker v3.0.8 (Gabriel et al. 2024), with default options, using as inputs the soft-masked genome, the proteins from the reference genomes of *Setaria italica*, *S. viridis*, *Sorghum bicolor*, *Panicum halli*, and *U. fusca*, all downloaded from Phytozome v13, and all RNA-seq libraries from *U. humidicola* accessions from our previous paper (Higgins et al. 2022). The resulting annotation from Braker v3 (Gabriel et al. 2024) was used as input to Mikado v2.3.3 (Venturini et al. 2018), together with the Stringtie assemblies generated by Braker and an exon/intron junctions file generated with Portcullis v1.2.4 (Mapleson et al. 2018) from the same RNA-seq libraries. The output from Mikado was used as the final gene annotation.

Subgenome assignment and ancestry

We used SubPhaser v20230401 (Jia et al. 2022; Zhang et al. 2024) to assign chromosomes into subgenomes (A, B, C) by counting kmers and test the enrichments of subgenome-specific kmers (differential kmers), identify and compare LTR-RTs by running LTRharvest and LTRfinder followed by TESorter for LTR classification, and compare haplotype blocks by alignment between chromosomes with Minimap2 to identify homoeologous blocks. Subphaser required the reference and a list of chromosomes certain to be homologous (eg, A_hap1, B_hap1, C_hap1). As we were not certain of the ancestry of some chromosomes, only the 18 chrs in hap1 were included initially. Additional homologous triads were entered iteratively until all haplotypes were included.

We downloaded the protein sequences of the “Angiosperms 353” (A353) markers from the “Tree of Life” project (Johnson et al. 2019). We also retrieved reads from the European Bioinformatics Institute (EBI)’s ENA (European Nucleotide Archive) for these markers, which were generated for 18 *Urochloa* species in 2 prior studies (Baker et al. 2022; Masters et al. 2024). We preprocessed them with Trim Galore v0.6.10 (Krueger 2021) with default options and trimmed 5 bps from each read end. Using the A353 protein sequences and the clean reads for each species, we employed Hybpiper (Johnson et al. 2016) to assemble the protein “Angiosperms 353” sequences for the 18 *Urochloa* species. In parallel, we generated the cv. Tully proteome based on the gene annotation and divided it into subsets by subgenome and haplotype (A hap1, A hap2, B hap1, etc.). A hap 3 and B hap 4 were excluded, as they were not present in all 6 chromosomes. We then compared these subsets from *U. humidicola* against the “Angiosperms 353” targets using Diamond v2.0.9 (Buchfink et al. 2021) and selected all queries with over 90% coverage using Seqtk v1.4.122, to extract the A353 homologs in each *U. humidicola* subgenome and haplotype. Finally, these subsets extracted from *U. humidicola*, along with the 18 sets assembled by Hybpiper (all corresponding to “Angiosperms 353” homologs), were clustered and analyzed with Orthofinder v3.0 to construct a phylogenetic

tree, which was visualized using the online tool iTOL v6 (Letunic and Bork 2024).

Composition in other *U. humidicola* accessions

Public DNA-seq data were downloaded from EBI’s ENA, including accessions CIAT 16888 (SRR8321580), CIAT 26155 (SRR16327314), and CIAT 26146 (13 read files from PRJNA509199), which is a sexual accession. These 3 datasets were preprocessed using Trim Galore v0.6.10 (default parameters), then aligned to the assembled chromosomes with Bowtie v2.4.1 using the options “-X 500 -no-unal -no-mixed -no-discordant -no-contain -no-overlap”, ensuring only concordant paired reads were retained. The aligner’s default behavior, where reads are aligned to their “top best” position, was applied. Since our alignment parameters only permit a read pair to align to their “top best” position, this results in a competitive mapping analysis where reads mainly map to their original subgenome for composition estimation. A coverage table was generated from the alignment files using Samtools coverage v1.51.1 (Danecek et al. 2021).

We also aligned the 32 RNA-seq libraries from *U. humidicola* accessions in our previous paper (Higgins et al. 2022) to the assembled chromosomes using Tophat v2.0.13 (Trapnell et al. 2009). The behavior of this aligner is comparable to Bowtie but is designed for RNA reads. A coverage table was generated from the “accepted hits” alignment file using Samtools v1.51.1.

Results and discussion

Genome assembly

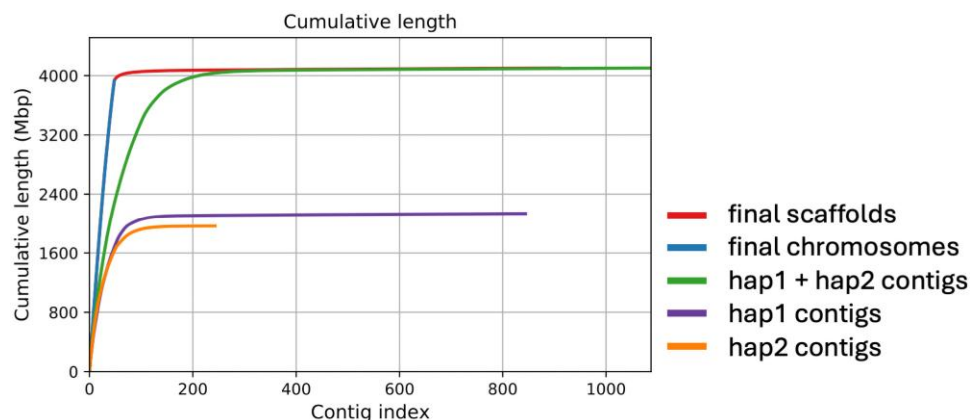
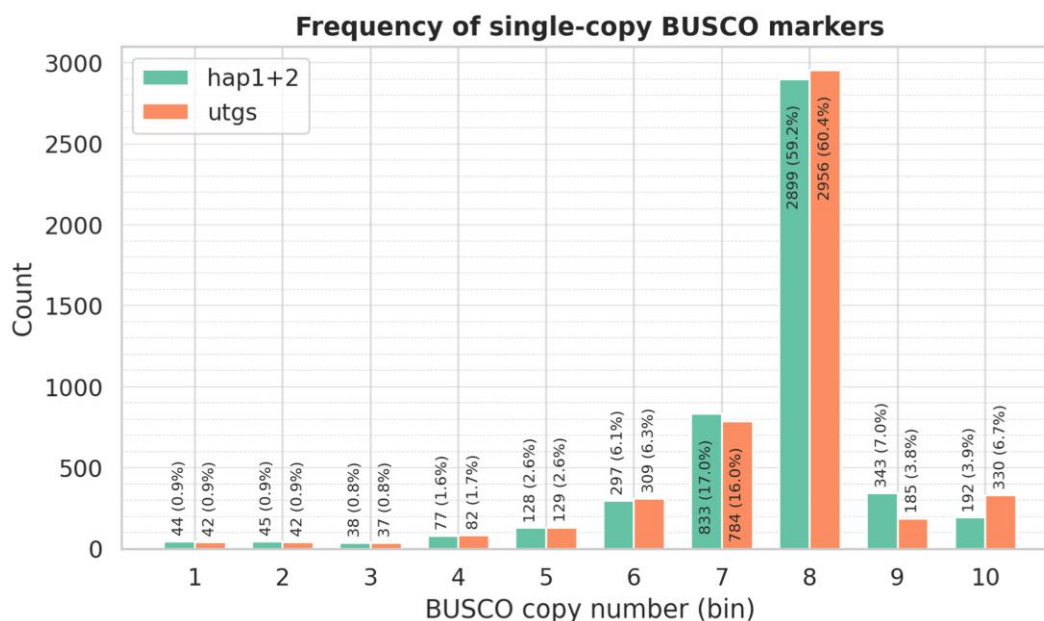
We generated a genome assembly from PacBio HiFi reads alone, using the assembler’s default settings. The resulting primary unitigs (p_utg) had a total length of 4,193 Gbp across 8,785 unitigs. The concatenated contigs from both haplotype files (hap1 + hap2) consisted of 1,087 contigs with a total length of 4,102 Gbp, and a N50 of 32,75 Mbp. 87% of the genome (3.569 Gb) was assembled into 115 contigs longer than 10 Mbp, and the longest 200 contigs accounted for 97% (3.98 Gb) of the genome (Table 1 and Fig. 1).

We aligned the Poaceae BUSCO markers to both unitigs and contigs, achieving 99.5% alignment success, with 99.1% being duplicated in both cases. We counted how many times each single-copy BUSCO marker aligned in the contigs (Fig. 2); 59.2% of the markers aligned 8 times (2899 markers out of 4875), indicating that most haplotypes were present 8 times. Assuming a base chromosome number of 6, this led to an initial estimation of 48 chromosomes. We did not observe significant differences between haplotypes (Supplementary Table 1).

Since the contiguity of the contigs was high (Table 1), and we expected 48 chromosomes, we concluded that most chromosomes would be contained within a few contigs each. Consequently, we decided to scaffold them using synteny, as detailed in the methods; firstly, comparing the hap1 and hap2 contigs to the primary contigs assembled by Hifiasm, then by comparing the “hap1” to the “hap2” homologs to each other, and finally by comparing to the longest assembled haplotype in each chromosome (often a full chromosome at this stage). In total, 225 contigs were scaffolded in 48 chromosomes. Three chromosomes were assembled in 1 contig and 9 more in 2 contigs. Only 6 chromosomes were assembled in 7 or more contigs. While most polyploid assemblies use high-throughput chromosome conformation capture (Zeng et al. 2024), such as Hi-C or Omni-C, the clear separation between the 3 subgenomes resulted in a highly contiguous assembly with HIFI alone that enabled our strategy.

Table 1. Summary statistics of the *Urochloa humidicola* cv. Tully genome at different stages, unitigs, contigs, and chromosomes.

Statistics	Final genome	Chromosomes	Contigs hap1 + hap2	Contigs hap1	Contigs hap2	Primary contigs	Unitigs
Number contigs	910	48	1,087	844	243	754	8,785
Number contigs (≥ 25 kbp)	888	48	1,065	823	242	726	7,233
Number contigs (≥ 50 kbp)	357	48	534	333	201	277	2,423
Largest contig (Mb)	106.79	106.79	98.63	98.63	95.29	93.65	56.63
Total length (Gbp)	4.102	3.93	4.102	2.136	1.966	2.317	4.193
Total length (≥ 25 kbp)	4.101	3.93	4.101	2.135	1.966	2.317	4.16
Total length (≥ 50 kbp)	4.081	3.93	4.081	2.117	1.964	2.299	3.993
N50	86.792	88.436	32.755	32.755	31.873	55.243	85.232
N75	73.268	73.95	18.507	19.622	17.884	28.097	3.793
L50	22	21	41	23	19	17	132
L75	35	33	83	44	39	32	314
GC (%)	45.76	45.73	45.76	45.72	45.81	45.71	45.76
N's	10,263,851	9,114,431	10,246,151	6,011,896	4,234,255	0	0
N's per 100 kbp	250.23	231.91	249.8	281.47	215.4	0	0

**Fig. 1.** Distribution of cumulative sequence lengths before and after anchoring across different subsets: within the contigs of each haplotype (hap1, hap2), the combined contigs (hap1 + hap2), and in the 48 chromosomes after anchoring, as well as unanchored contigs. Sequences are ordered from longest to shortest along the x-axis.**Fig. 2.** Distribution of BUSCO copy numbers across unitigs (utgs) and contigs (hap1 + hap2). The histogram shows the number of alignments for each BUSCO marker, indicating that about 59% of the markers are present 8 times, consistent with an octoploid genome structure.

We annotated 259,254 genes and 277,99 transcripts. 249,665 genes (96.3%) were in chromosomes, and the rest were in contigs. The OMArk's "completeness assessment" for the annotated

proteome was 3.89% complete unique, 93.68% complete duplicated, and 2.42% missing; the "consistency assessment" was 85.48% consistent, 6.52% inconsistent, 8% unknown, and 0%

contaminants. However, this analysis is limited because the closest relative in the database was the not-so-close *Setaria italica* genome. We also annotated repeat elements (Table 2 and Supplementary Table 2).

Differences between ancestries

Chromosomes were assigned to ancestries based on differential k-mer analysis among chromosome sets. The SubPhaser tool clusters the subgenomes using k-means, from which it constructs a hierarchical dendrogram and heatmap (Fig. 3a) based on the presence-absence variation of these differential k-mers, as well as a PCA (Fig. 3b). The composition of cv. Tully was AABBBBCC for chromosomes 1 to 4 and AAABBBCC for chromosomes 5 and 6. This can also be expressed as 14A 22B 12C or 12A + 2 24B-2 12C.

Table 2. Annotation of repeat elements in the genome.

		Count	Length (bp)	%
LTR	Copia	150,248	198,461,341	5.06
	Gypsy	375,261	614,491,907	15.67
	Unknown	239,025	236,103,239	6.02
TIR	CACTA	180,577	109,309,372	2.79
	Mutator	171,399	80,258,412	2.05
	PIF_Harbinger	108,412	32,453,696	0.83
	Tc1_Mariner	159,314	53,031,681	1.35
	hAT	84,620	30,158,398	0.77
nonTIR	Helitron	689,211	365,259,293	9.32
Totals	Repeats	2,158,067	1,719,527,339	43.85
	Genome		3,921,057,692	100.00

A haplotype analysis based on chromosome alignments with Minimap2 allowed to identify conserved homoeologous blocks. It showed the shared synteny between subgenomes (Fig. 4). The A and B copies of each chromosome were highly similar (green links in Fig. 4). The C subgenome exhibited lower synteny to the previous (red links in Fig. 4).

Subphaser was also used to analyze the LTR-RT elements by subgenome by enrichment analysis of LTR-RTs detected by several classification tools. The insertion age was estimated per subgenome (Fig. 3c), which supported a closer relation between subgenomes A and B than with C. We also obtained phylogenetic trees of subgenome-specific LTR/Gypsy and LTR/Copia elements (Fig. 5). In most LTR clades, there was a clear pattern where A and B clustered together, sometimes mixed, and C clustered separately.

The clear clustering of A and B subgenomes in both differential k-mer and LTR analyses indicates that these 2 ancestries diverged more recently from each other than from the C genome donor. The shared peaks of amplification for Copia and Gypsy elements suggest that A and B lineages experienced similar transpositional bursts, possibly during a common polyploidization event. In contrast, the C subgenome shows a different repeat composition and older insertion ages, which align with a long history of independent evolution before its introgression. Collectively, these findings support a stepwise allopolyploid origin for *U. humidicola*, with the A–B component forming first, followed by the later addition of the diverging C genome.

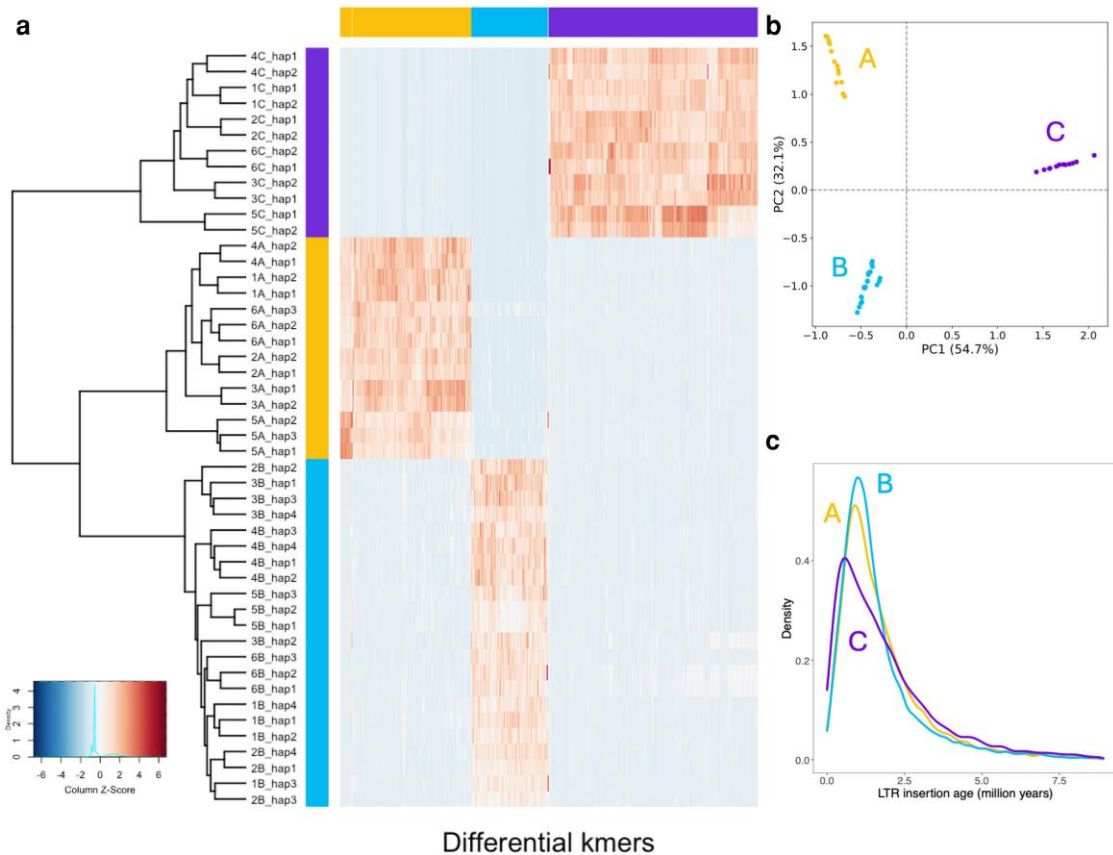


Fig. 3. Subgenome classification of the 48 chromosomes in *U. humidicola*. a) Hierarchical clustering heatmap based on differential Kmers, and b) Principal-component analysis (PCA) based on differential Kmers. Both analyses differentiate chromosomes into 3 subgenomes (A, B, and C). c) Insertion age distribution of LTR retrotransposons reveals overlapping peaks for A and B and older insertions for C, pointing to distinct evolutionary histories.

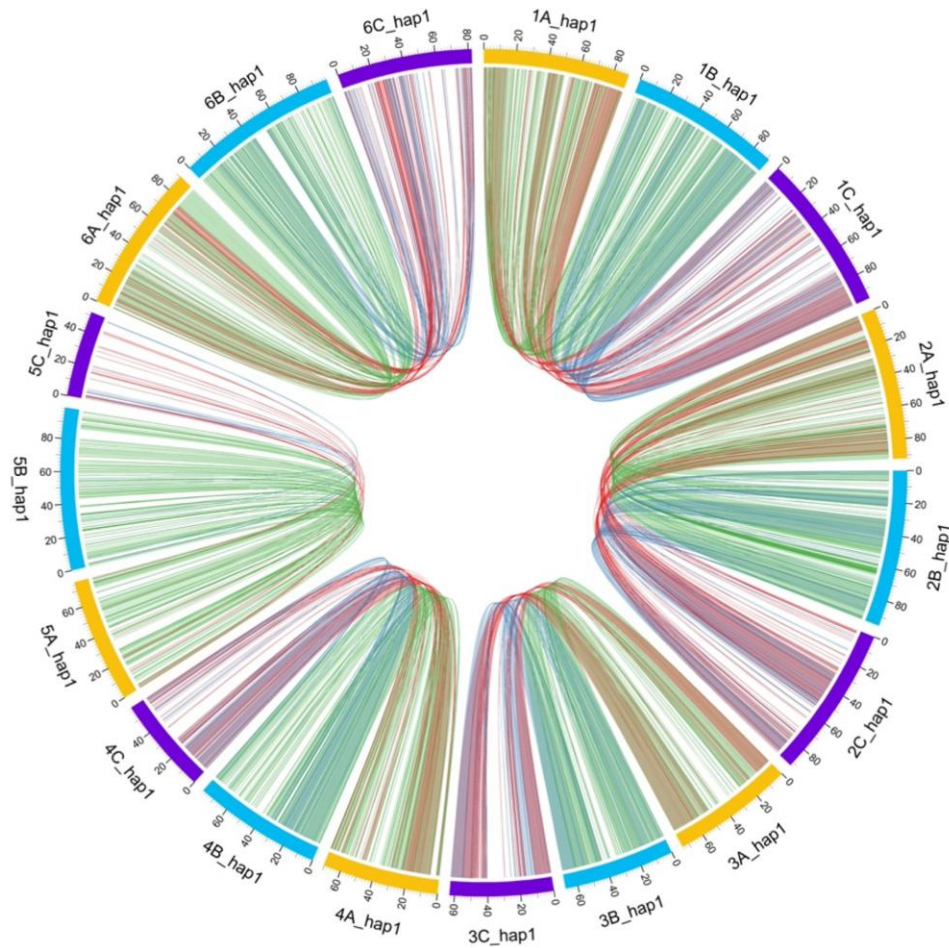


Fig. 4. Comparative synteny among subgenomes. Circos plot illustrating shared syntenic blocks among the A, B, and C subgenomes. Green ribbons represent A-B synteny, yellow ribbons denote B-C, and red ribbons show A-C relationships; reduced collinearity of C-subgenome chromosomes is evident near centromeres.

Ancestral origins

Most *Urochloa* species, particularly wild ones, lack WGS data. This led us to use the “Angiosperms 353” (A353) markers from the “Tree of Life” project, which had been generated for at least 15 *Urochloa* species (Baker et al. 2022; Masters et al. 2024). For this, we identified and extracted the homologs to the A353 marker from the new genome in each *U. humidicola* subgenome and haplotype. The tree we obtained (Fig. 6) consistently placed the B subgenome with *U. dictyoneura* and separated it from other wild relatives in the same clade. This suggests that *U. dictyoneura* is the ancestor of the B subgenome. These 2 species are known to intercross and share the same base chromosome number of 6. The A subgenome was included in the Humidicola clade but was not one of the wild relatives for which we had markers. The C subgenome is phylogenetically distant from the previous and is likely *U. arrecta* from the Mutica clade (Fig. 6).

This phylogenetic framework implies a historical hybridization across clade boundaries, underscoring the reticulate nature of *Urochloa* evolution. These relationships provide a robust basis for revising taxonomic relationships within the species and genus. Our results highlight an evolutionary trajectory of repeated hybridization events, at least as complex as in the Brizantha complex.

Genomic composition in *U. humidicola*

We aligned the publicly available DNA-seq and RNA-seq data from *U. humidicola* to the reference genome. Our alignment parameters allowed each read to align only to its “top best” position (Fig. 7). The approach was experimentally validated when we analyzed the downloaded RNA-seq data from cultivar Tully, where the observed proportions of reads (Fig. 7) were 23% for the C subgenome, 28% for the A subgenome, and 49% for the B subgenome. These closely matched the expected proportions: 25% for the C subgenome (12/48 chromosomes), 29% for the A subgenome (14/48), and 46% for the B subgenome (22/48). CIAT-26155 (SRR16327314) underwent whole-genome resequencing and was identified as hexaploid through k-mer analysis (Tomaszewski et al. 2023). Among the reads, 15,646,090 aligned to a chromosome from the A subgenome (33.32%), 31,076,742 to a chromosome from the B subgenome (66.19%), and 230,974 to a chromosome from the C subgenome (0.005%). The ratio of reads between the A and B subgenomes was clearly 1A:2B, supporting a composition of 12A and 24B (AABBBB) for CIAT-26155. Conversely, accession 16888, expected to be an aneuhexaploid (Worthington et al. 2019), showed about 40% of the reads aligned to A subgenome chromosomes and roughly 60% to B subgenome chromosomes, resulting in a 1A:1.5B ratio. This indicates an overrepresentation of A chromosomes compared to B, but could also be explained

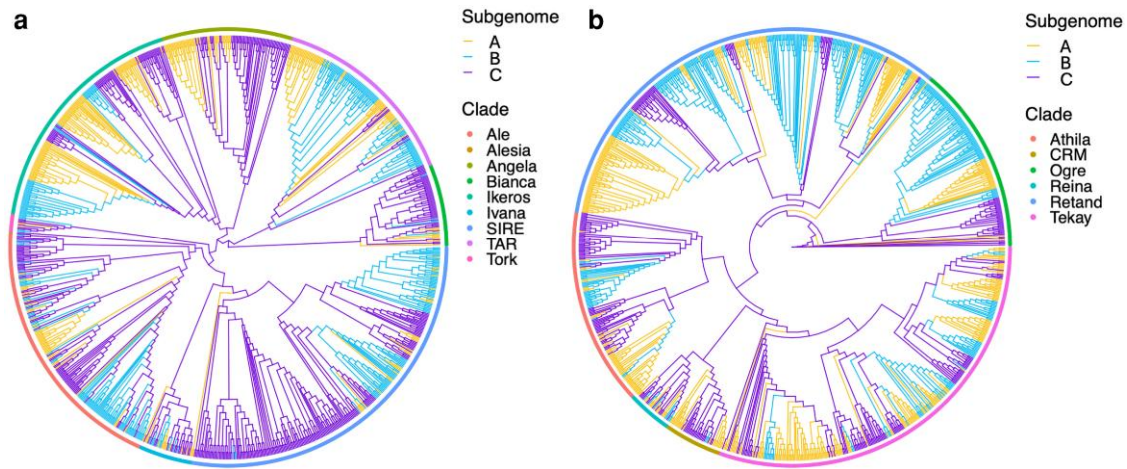


Fig. 5. Distribution and phylogeny of LTR retrotransposons. Bar plots show the relative abundance of Gypsy (a) and Copia (b) elements in each subgenome, and maximum-likelihood trees depict their relationships. Clustering of A and B LTR lineages and separation of C elements support a shared transpositional history between A and B subgenomes.

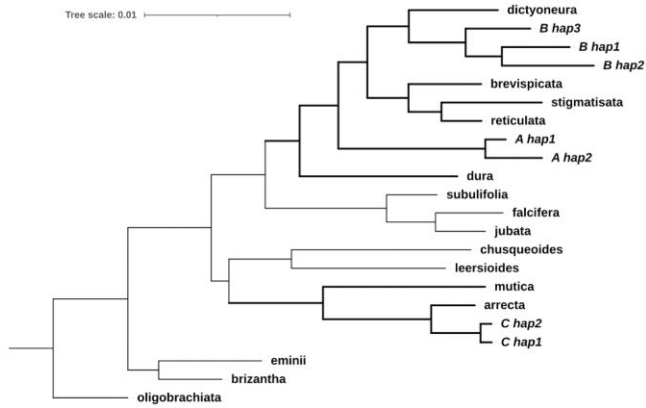


Fig. 6. Phylogenetic placement of *U. humidicola* subgenomes haplotypes (named by ancestry and haplotype number, eg, A_hap1) among related *Urochloa* species (named by species name omitting genus, eg, *dura* for *Urochloa dura*). The maximum-likelihood tree, constructed using 353 protein markers ("Angiosperms 353"), shows the B subgenome grouping with *U. dictyoneura* and the C subgenome with *U. arrecta*, while A subgenome sequences cluster independently within the Humidicola clade.

by unbalanced marker density, as this accession was genotyped using genotyping-by-sequencing (GBS). A 1A:1.5B ratio would be equivalent to the 14A 22B composition observed in the genome of cultivar Tully.

The competitive mapping results for the sexual accession CIAT-26146 (Worthington et al. 2019) were markedly different, with only 10% of reads aligning to the A subgenome and 90% to the B subgenome (Fig. 7), resulting in a 1:9 ratio between A and B. This accession was also genotyped using GBS, as before, there may be unbalanced marker density. Accepting the likely hexaploidy (Worthington et al. 2019), this supports an autopolloid background (BBBBBB), ranging from none to up to 3 A chromosomes (0-3A, 36-33B).

RNA-seq reads (Higgins et al. 2022) from 4 accessions (Fig. 7), including cv. Tully itself (labelled as CIAT-679), aligned to the C subgenome with a frequency of around 23%. None of the remaining accessions aligned to the C subgenome (<0.01% of the reads). These 4 accessions aligned to the A and B subgenomes with

similar proportions, so we should assume a composition similar to cv. Tully. In Higgins et al. 2022, these 4 accessions were grouped into a different subpopulation than the other accession, called "humidicola-2" (Higgins et al. 2022). This evidence supports that these subpopulations were split by the presence or absence of C ancestry.

Among the remaining accessions, most displayed a ratio close to 1A:2B (about 33% reads aligned to A and 66% reads aligned to B), indicating that the most common composition is AABBBB. However, their ratios ranged from 1A:1.25B in CIAT-16887 to 1A:3B in CIAT-16876 (Fig. 7). While coverage analysis from RNAseq is noisier than WGS, as it reflects allelic expression ratios rather than homolog ratios; this variation could equally indicate differing ploidy levels beyond hexaploidy; for example, a composition of AABBBBBB would match the ratio observed in CIAT-16876.

The diversity of subgenomic ratios among accessions indicates ongoing genomic fluidity in *U. humidicola* populations, reflecting recurrent hybridization and uneven chromosome transmission. Such variation may underlie differences in reproductive mode, as apomictic accessions maintain fixed heterozygosity, whereas sexual forms may reshuffle subgenomic composition through residual meiotic pairing.

Evolution of *U. humidicola*

We propose a model (Fig. 8) that summarizes the inferred origin and diversification of *U. humidicola* cytotypes and accessions based on the analysis of this genome reference (Fig. 6) and competitive read-mapping against it (Fig. 7). The ancestors *U. dictyoneura* ($2n = 4x = 24$; B subgenome), wild *Urochloa* from the Humidicola clade (A subgenome), and *U. arrecta* ($2n = 4x = 36$; C subgenome) are shown at the top. *U. dictyoneura* is likely autopolyploid (reference), but it could also be allotetraploid.

Initial hybridization between *U. dictyoneura* (or a very closely related species) and the unknown A genome donor produced the AABBBB-type hexaploids, which are the most common types in the genebank collection. Selected examples include accessions CIAT 26155, CIAT 16888, and CIAT 16887, and variation in the ratios obtained reflecting differential retention of A-origin chromosomes, as observed in cv. Tully. A second lineage derived from *U. dictyoneura* alone generated BBBBBB cytotypes, such as the sexual accession CIAT-26146 (H031). The sexual accession has

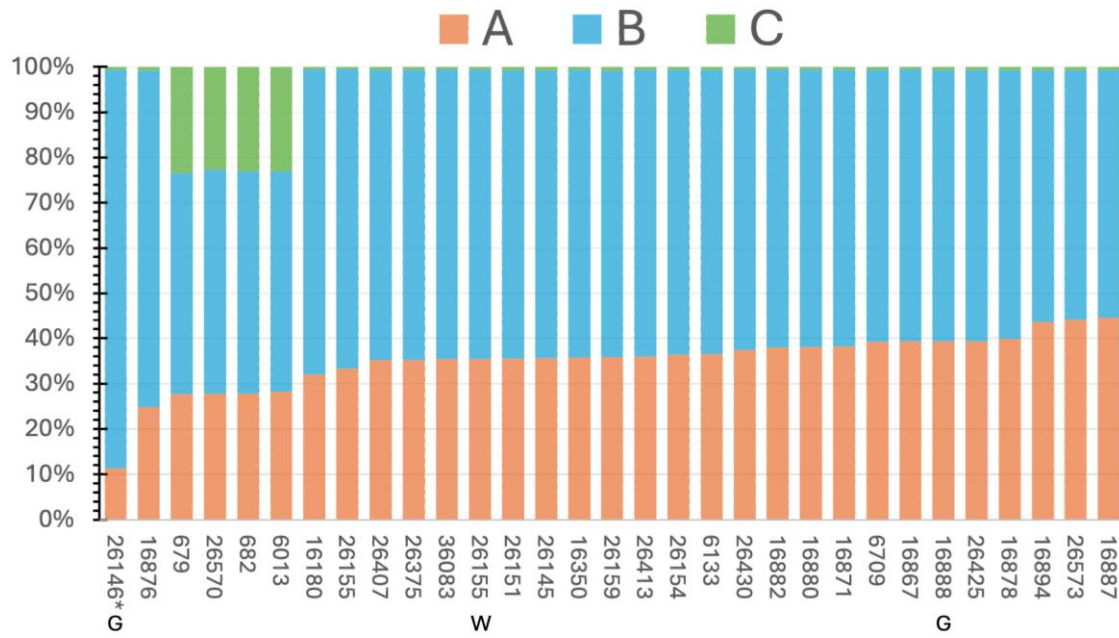


Fig. 7. Competitive read mapping of public *U. humidicola* accessions to subgenomes. Stacked bar plots show the proportion of reads aligning to chromosomes of the A, B, and C subgenomes for each accession. They were genotyped using RNA-seq, except 26155 (labelled W for WGS), and 26146 and 16888 (labelled G for GBS). The sexual accession is marked with an asterisk. These ratios mainly reflect the variability in subgenome copy numbers and ploidy levels across different accessions.

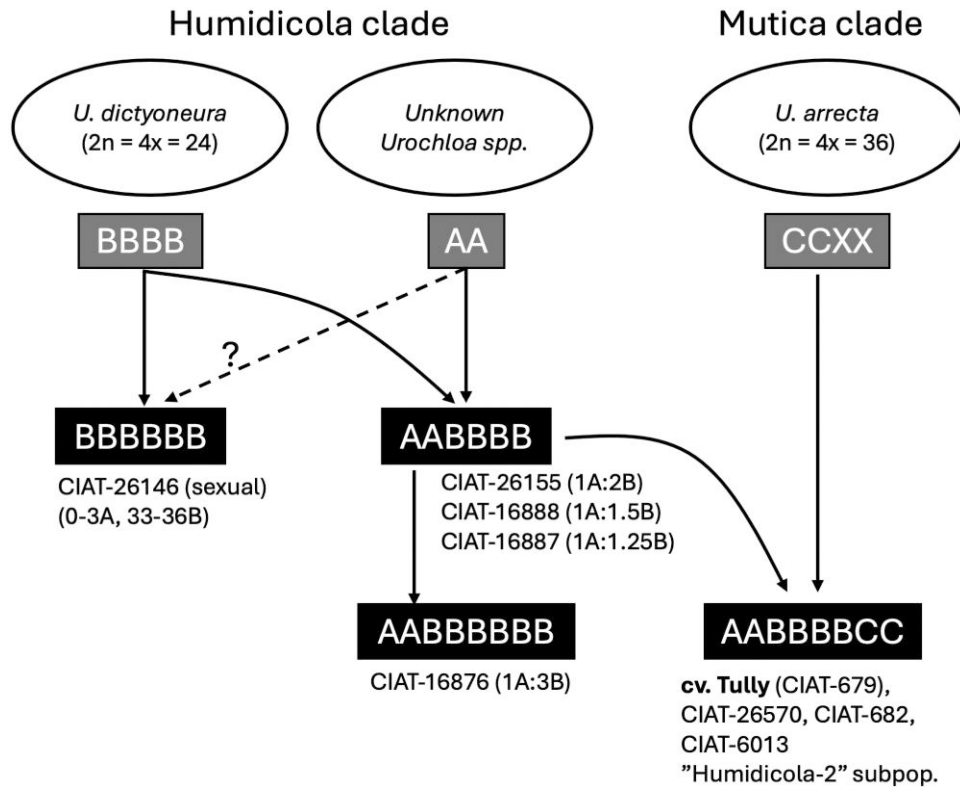


Fig. 8. Evolutionary model based on the phylogenetic placements of the *U. humidicola* subgenomes (Fig. 6) and inferred composition of *U. humidicola* accessions (Fig. 7). The model aims to explain the composition among *Urochloa humidicola* accessions and putative events leading to different *U. humidicola* cytotypes (black boxes), reflecting the species' complex reticulated evolution. The B-genome donor *U. dictyoneura*, an A-genome wild *Urochloa* from the Humidicola clade contributed to the common hexaploid (AABBBB), as well as increasing ploidy levels, such as octoploid (AABBBBBB). While *U. arrecta* (C genome) contributed to AABBBBCC lineages, including the apomictic cultivar Tully sequenced.

been suggested to be autopolyploid from analyses of meiotic pairing in 2 independent biparental populations (Worthington et al. 2019; Ferreira et al. 2021; da Costa Lima Moraes et al. 2023).

Further hybridization among AABBBB cytotypes likely produced octoploid forms (AABBBBBB), exemplified by CIAT-1673 (12A: 36B), and potentially other ploidy levels.

A subsequent introgression event involving the named C genome from *U. arrecta* (or a very closely related species) into the AABBBB background led to the creation of the ABC lineage exemplarized by cultivar *U. humidicola* cv. Tully (AABBBC), which serves as the archetype of the “humidicola-2” subpopulation. This subpopulation also includes CIAT-26570, CIAT 682, and CIAT 6013. These accessions retain genetic signatures of the A, B, and C subgenomes, aligning with the structure observed in this cv. Tully assembly. Jungmann et al. (2010) first noted 2 genetically divergent groups within *U. humidicola*. Higgins et al. (2022) later confirmed and formalized this pattern, naming 2 genomic clusters: “humidicola-1” (AB ancestry) and “humidicola-2” (ABC ancestry).

Although no direct chromosome count exists for cv. Tully, recent cytogenetic work provides strong supporting evidence for an octoploid (8x) cytotype within this genomic background. Tomaszewska et al. (2023) used metaphase imaging and FISH to show that the related accession CIAT 16867 is octoploid ($2n = 8x + 2$), demonstrating that 8x cytotypes occur within the ABC lineage. Consistent with this result, our competitive read-mapping analysis (Fig. 7) shows that Tully (CIAT 679) and CIAT 16867 share nearly identical A–B–C subgenome proportions, strongly supporting the interpretation that cv. Tully is also octoploid. Broader cytogenetic surveys of *U. humidicola* showed that the species forms a polyploid series of 6x, 7x, 8x, and 9x cytotypes (Jungmann et al. 2010; Damasceno et al. 2023). Earlier reports of nonaploid ($2n = 54$) accessions largely correspond to commercial germplasm (Boldrini et al. 2011), but cv. Tully itself was never explicitly counted in these studies. Genome-size profiles and FISH patterns (Damasceno et al. 2023) also support the existence of distinct ploidy classes, with octoploids showing intermediate genome sizes between 6x and 9x cytotypes.

Our results indicate that recurrent hybridization, partial diploidization, and asymmetric subgenome retention have shaped the mosaic ancestry of *U. humidicola*. The inferred evolutionary model (Fig. 8) reflects the broader pattern of reticulate evolution often observed across tropical forage grasses, where polyploidy and apomixis repeatedly facilitate the stabilization of interspecific hybrids (Hojsgaard and Hörandl 2019). Such processes parallel those seen in *Paspalum* (Schedler et al. 2023) and *Poa* (Raggi et al. 2015), suggesting convergent evolutionary solutions underpin the ecological success of polyploid apomictic tropical grasses.

Conclusion

The identification of distinct A, B, and C ancestries in *U. humidicola* now enables subgenome-specific analyses of gene expression, methylation, and apomixis regulation, which were previously obscured by the lack of a complete reference. Researchers working with hexaploid *U. humidicola* panels or families simply need to remove the C ancestry chromosomes to obtain an accurate reference for their system. The haplotype-complete assembly offers a framework to locate and characterize the chromosomal regions associated with apomixis, previously mapped only through low-resolution markers (Worthington et al. 2019). Beyond evolutionary insights, the genome provides practical targets for breeding. The clear delineation of subgenomes allows genome-wide association studies and marker development for traits such as forage digestibility and BNI. These advances align with global efforts to enhance tropical pasture productivity while reducing methane emissions and nitrogen losses, positioning *U. humidicola* as a model for genomic strategies toward low-carbon livestock systems. This genome represents a foundational resource for the *Urochloa* research community. It supports future comparative genomics across tropical forage species, informs

efforts to engineer apomixis in crops, and contributes to global initiatives for climate-resilient agriculture.

Data availability

All data are deposited in the SRA (Sequence Read Archive) under accession PRJEB90424. The genome assembly, together with its gene annotation, was deposited in ENA (European Nucleotide Archive) with accession GCA_965614515.2 (https://www.ebi.ac.uk/ena/browser/view/GCA_965614515.2). The scripts used are available at <https://github.com/DeVegaGroup/Assembly-and-analysis-Urochloa-humidicola-genome>.

The annotation (GTF and GFF formats) and assembly are available in multiple formats under https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/965/614/515/GCA_965614515.1_Urohum_v2/.

Supplemental material available at G3 online.

Acknowledgments

All the authors approved this manuscript. We thank the Australian Pastures Genebank and the CIAT CGIAR Genebank platform in Colombia for providing the accessions. The authors would like to acknowledge the support of the Transformative Genomics group at the Earlham Institute, especially Drs. Sacha Lucchini, Leah Catchpole, Karim Gharbi, and Chris Watkins for their comprehensive support and roles in platform management and supervision. We also acknowledge the invaluable support from the Earlham Institute’s research office, BDE and support teams, the Norwich Bioscience Institute’s research computing and contracts teams, and horticultural services at the John Innes Centre.

Funding

This study was partially funded by the Biotechnology and Biological Sciences Research Council (BBSRC), part of UK Research and Innovation (UKRI), via Earlham Institute’s Strategic Programme Grant (Decoding Biodiversity) BBX011089/1, and its constituent work package BBS/E/ER/230002B (WP2 Genome Enabled Analysis of Diversity to Identify Gene Function, Biosynthetic Pathways, and Variation in Agri/Aquacultural Traits). Funding was also received from the BBSRC’s funded “Core Strategic Programme Grant” (Genomes to Food Security) BB/CSP1720/1 and its constituent work package BBS/E/T/000PR9818 (WP1 Signatures of Domestication and Adaptation), as well as BB/X011089/1. Funding was also received from the Biotechnology and Biological Sciences Research Council (BBSRC) as part of UK Research and Innovation, Core Capability Grant BB/CCG1720/1.

Conflicts of interest

None declared.

Literature cited

Alonge M et al. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23:258. <https://doi.org/10.1186/s13059-022-02823-7>.

- Arango J et al. 2020. Ambition meets reality: achieving GHG emission reduction targets in the livestock sector of Latin America. *Front Sustain Food Syst.* 4:65. <https://doi.org/10.3389/fsufs.2020.00065>.
- Baker WJ et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst Biol.* 71:301–319. <https://doi.org/10.1093/sysbio/syab035>.
- Boldrini KR et al. 2011. Meiotic behavior in nonaploid accessions of *Brachiaria humidicola* (Poaceae) and implications for breeding. *Genet Mol Res.* 10:169–176. <https://doi.org/10.4238/vol10-1gmr990>.
- Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods.* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Cervantes-Díaz CI, Oyama K, Cuevas E. 2024. Ecological and evolutionary implications of hybridization, polyploidy and apomixis in angiosperms. *Bot Sci.* 102:1043–1061. <https://doi.org/10.17129/botsci.3583>.
- Cheng H et al. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- da Costa Lima Moraes A et al. 2023. Advances in genomic characterization of *Urochloa humidicola*: exploring polyploid inheritance and apomixis. *Theor Appl Genet.* 136:238. <https://doi.org/10.1007/s00122-023-04485-w>.
- Damasceno AG et al. 2023. Physical mapping of ribosomal DNA sites and genome size in polyploid series of *Urochloa humidicola* (Rendle) Morrone & Zuloaga (Poaceae). *Bot Lett.* 170:634–643. <https://doi.org/10.1080/23818107.2023.2192274>.
- Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience.* 10:giab008. <https://doi.org/10.1093/gigascience/ giab008>.
- Egenolf K et al. 2020. Brachialactone isomers and derivatives of *Brachiaria humidicola* reveal contrasting nitrification inhibiting activity. *Plant Physiol Biochem.* 154:491–497. <https://doi.org/10.1016/j.plaphy.2020.06.004>.
- EI-CoreBioinformatics. eirepeat: eirepeat - EI Repeat Identification pipeline [accessed 2025 Oct 19]. <https://github.com/EI-CoreBioinformatics/eirepeat>
- Ferreira RCU et al. 2021. An overview of the genetics and genomics of the *Urochloa* species most commonly used in pastures. *Front Plant Sci.* 12:770461. [10.3389/fpls.2021.770461](https://doi.org/10.3389/fpls.2021.770461).
- Gabriel L et al. 2024. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* 34:769–777. <https://doi.org/10.1101/gr.278090.123>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Higgins J et al. 2022. Diverged subpopulations in tropical *Urochloa* (Brachiaria) forage species indicate a role for facultative apomixis and varying ploidy in their population structure and evolution. *Ann Bot.* 130:657–669. <https://doi.org/10.1093/aob/mcac115>.
- Hojsgaard D, Hörandl E. 2019. The rise of apomixis in natural plant populations. *Front Plant Sci.* 10:358. [10.3389/fpls.2019.00358](https://doi.org/10.3389/fpls.2019.00358).
- Issifu S et al. 2024. Biological nitrification inhibitors with antagonistic and synergistic effects on growth of ammonia oxidisers and soil nitrification. *Microb Ecol.* 87:143. <https://doi.org/10.1007/s00248-024-02456-2>.
- Jia K-H et al. 2022. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytol.* 235:801–809. <https://doi.org/10.1111/nph.18173>.
- Johnson MG et al. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl Plant Sci.* 4:1600016. <https://doi.org/10.3732/apps.1600016>.
- Johnson MG et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst Biol.* 68:594–606. <https://doi.org/10.1093/sysbio/syy086>.
- Jungmann L et al. 2010. Genetic diversity and population structure analysis of the tropical pasture grass *Brachiaria humidicola* based on microsatellites, cytogenetics, morphological traits, and geographical origin. *Genome.* 53:698–709. <https://doi.org/10.1139/g10-055>.
- Krueger F, et al. 2021. TrimGalore. <https://doi.org/10.5281/zenodo.5127899>.
- Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52:W78–W82. <https://doi.org/10.1093/nar/gkae268>.
- Mapleson D, Venturini L, Kaithakottil G, Swarbreck D. 2018. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience.* 7:giy131. <https://doi.org/10.1093/gigascience/giy131>.
- Masters LE et al. 2024. Phylogenomic analysis reveals five independently evolved African forage grass clades in the genus *Urochloa*. *Ann Bot.* 133:725–742. <https://doi.org/10.1093/aob/mcae022>.
- Miles JW. 1996. *Brachiaria: biology, agronomy, and improvement*. Cali, Colombia: CIAT Publication.
- Nevers Y et al. 2025. Quality assessment of gene repertoire annotations with OMARK. *Nat Biotechnol.* 43:124–133. <https://doi.org/10.1038/s41587-024-02147-w>.
- Núñez J et al. 2018. Biological nitrification inhibition activity in a soil-grown biparental population of the forage grass, *Brachiaria humidicola*. *Plant Soil.* 426:401–411. <https://doi.org/10.1007/s11104-018-3626-5>.
- Raggi L et al. 2015. Understanding genetic diversity and population structure of a *Poa pratensis* worldwide collection through morphological, nuclear and chloroplast diversity analysis. *PLoS One.* 10:e0124709. <https://doi.org/10.1371/journal.pone.0124709>.
- Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21:245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Schedler M et al. 2023. Alternative evolutionary pathways in *Paspalum* involving allotetraploidy, sexuality, and varied mating systems. *Genes (Basel).* 14:1137. <https://doi.org/10.3390/genes14061137>.
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editors. *Gene prediction: methods and protocols*. Springer. p. 227–245 [accessed 2024 Sept 27]. https://doi.org/10.1007/978-1-4939-9173-0_14.
- Tomaszewska P et al. 2023. Complex polyploid and hybrid species in an apomictic and sexual tropical forage grass group: genomic composition and evolution in *Urochloa* (Brachiaria) species. *Ann Bot.* 131:87–108. <https://doi.org/10.1093/aob/mcab147>.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 25:1105–1111. <https://doi.org/10.1093/bioinformatics/btp120>.
- Underwood CJ, Mercier R. 2022. Engineering apomixis: clonal seeds approaching the fields. *Annu Rev Plant Biol.* 73:201–225. <https://doi.org/10.1146/annurev-arplant-102720-013958>.
- Vázquez E et al. 2020. Gross nitrogen transformations in tropical pasture soils as affected by *Urochloa* genotypes differing in biological nitrification inhibition (BNI) capacity. *Soil Biol Biochem.* 151:108058. <https://doi.org/10.1016/j.soilbio.2020.108058>.

- Venturini L et al. 2018. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*. 7:giy093. <https://doi.org/10.1093/gigascience/giy093>.
- Vigna BBZ et al. 2016. Evidence of allopolyploidy in *Urochloa humidicola* based on cytological analysis and genetic linkage mapping. *PLoS One*. 11:e0153764. <https://doi.org/10.1371/journal.pone.0153764>.
- Worthington M et al. 2019. Translocation of a parthenogenesis gene candidate to an alternate carrier chromosome in apomictic *Brachiaria humidicola*. *BMC Genomics*. 20:41. <https://doi.org/10.1186/s12864-018-5392-4>.
- Zeng X et al. 2024. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat Plants*. 10:1184–1200. <https://doi.org/10.1038/s41477-024-01755-3>.
- Zhang R-G, Shang H-Y, Jia K-H, Ma Y-P. 2024. Subgenome phasing for complex allopolyploidy: case-based benchmarking and recommendations. *Brief Bioinform*. 25:bbad513. <https://doi.org/10.1093/bib/bbad513>.

Editor: T. Pyhäjärvi