

# **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<sup>1</sup>, Alex Durrant<sup>1</sup>, Naomi Irish<sup>1</sup>, Tom Barker<sup>1</sup>, Rosa N. Jauregui<sup>2</sup>

<sup>1</sup> Earlham Institute, Norwich Bioscience Institutes (NBI), Norwich Research Park, NR4 7UZ, UK.

<sup>2</sup> Alliance of Biodiversity International and CIAT, Km 17 Recta Cali Palmira, AA 6713, Cali, Colombia.

## **Abstract**

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 AABBBBCC structure with three 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 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 AABBBB (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 fertilisation (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 fertilising (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).

56

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

64

65 Here, we report a haplotype-complete, chromosome-level genome assembly of the apomictic *U.*  
66 *humidicola* cv. Tully. This reference genome provides the first comprehensive view of the subgenomic  
67 structure and evolutionary relationships within this complex polyploid species. It represents a valuable  
68 resource for the forage genomics community, enabling future studies on genome evolution, the genetic  
69 control of apomixis and BNI, and the deployment of genomic tools for breeding improved climate-  
70 resilient forages.

71

72

## 73 **Methods & Materials**

74

### 75 **Sample collection, DNA extraction and sequencing**

76

77 *U. humidicola* cv. Tully plants were grown for 8 weeks before DNA was extracted from seeds of  
78 accession APG 58384, obtained from the Australian Pastures Genebank. Tully is an apomictic genotype  
79 found worldwide; it was introduced into Australia in 1952 by J.F. Miles as accession CPI 16707  
80 (AusTRCF 16707) from Rietondale Experimental Stations, Pretoria, South Africa (-25.7283, 28.236,  
81 1,300 m elevation, 700 mm rainfall). It was later introduced to Fiji and Papua New Guinea; from there,  
82 clonal material was reintroduced from Papua New Guinea to the Tully region of North Queensland in  
83 1973, and it was officially registered for commercial release as “Tully” in 1981 (Miles 1996). The seed

84 of cv. Tully (APG 58384) used for sequencing comes from this reintroduced material. In America, the  
 85 cultivated material is known as “pasto humidicola” or “Chetumal” (CIAT 679, PI 499378, BRA  
 86 001627, INIAP-NAPO 701), which are apomictic clones of the sequenced accession.  
 87  
 88 After 8 weeks, leaf material was harvested following two days in dark conditions. Extraction of HMW  
 89 DNA was performed using the Nucleon PhytoPure kit, with a variation of the recommended protocol  
 90 optimised for DNA fragment size, yield and purity. 1g snap-frozen leaf tissue was ground under liquid  
 91 nitrogen for 9 minutes, with the mortar refilled with LN2 between every 90 second of grinding. The  
 92 resulting powder was thoroughly mixed into Reagent 1 (with 4uL of 100mg/ml RNase A) in a 15ml  
 93 tube using a bacterial spreader loop. The sample was incubated on a ThermoMixer at 37C for 30  
 94 minutes. Reagent 2 was then added and the tube was gently inverted until the contents appeared  
 95 homogeneous, at which point the tube was incubated on a ThermoMixer at 65C for 10 minutes.  
 96 Following the post-lysis ice incubation, 2ml chloroform and 300uL resin was added, and the sample was  
 97 mixed on a 3D platform rocker for 10 minutes at room temperature. From here, the sample was  
 98 centrifuged at 1300g for 10 minutes. The upper phase was mixed with 1 volume of 25:24:1  
 99 Phenol:Chloroform:Isoamyl Alcohol, mixed on a 3D platform rocker for 10 minutes at 4C, and  
 100 centrifuged (3000g) for 10 minutes. The upper phase was transferred to another 15ml Falcon tube and  
 101 precipitation proceeded as recommended by the manufacturer protocol. The final elution was left open  
 102 in the chemical safety cabinet for 1.5 hours to allow residual ethanol to evaporate, after which the DNA  
 103 sample was left at room temperature overnight.  
 104  
 105 11.5µg of gDNA was manually sheared with the Megaruptor 3 instrument (Diagenode, P/N  
 106 B06010003) according to the Megaruptor 3 operations manual, with a speed setting of 31 and input of  
 107 18ng/µl in 150µl. The sheared sample then underwent AMPure® PB bead (PacBio®, P/N 100-265-  
 108 900) purification and concentration (0.45X, eluted in 95 µL of EB) before undergoing library  
 109 preparation using the SMRTbell® Express Template Prep Kit 2.0 (PacBio®, P/N 100-983-900). The  
 110 HiFi library was prepared according to the HiFi protocol version 03 (PacBio®, P/N 101-853-100), and  
 111 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 4.55hr 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 Two 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 8M. The parameters for sequencing per SMRT cell were adaptive loading (set to 0.85 target for 2 hours), 30-hour movie, 2-hour 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 Merqury 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 synten 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 maximising 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 visualised as a dotplot with dotPlotly, using a minimum alignment length of 10 kb. 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 six chromosomes (instead of the real nine) by combining chromosomes 2 and 4, chromosomes 5 and 6, and chromosomes 1 and 7 from the available *U. fusca* genome downloaded from Phytozone. 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 six “longest” sequences from *U. humidicola* using RagTag, following the same procedure as before; first analysing each haplotype independently, then together. In total, 225 contigs were scaffolded into 48 chromosomes.

## Gene annotation

The genome was soft-masked using the EI-Repeat pipeline (EI-CoreBioinformatics). 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 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), identify and compare LTR-RTs, and compare haplotype blocks by alignment between chromosomes. Subphaser required the reference and a list of chromosomes certain to be homoelogenous (e.g. 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 homoelogenous 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 EBI’s ENA for these markers, which were generated for 18 *Urochloa* species in two prior studies (Baker et al. 2022; Masters et al. 2024). We pre-processed them with Trim Galore v0.6.10 (Krueger 2015) 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 six 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.*

196 *humidicola*, along with the 18 sets assembled by Hybpiper (all corresponding to “Angiosperms 353”  
197 homologs), were clustered and analysed with Orthofinder v3.0 to construct a phylogenetic tree, which  
198 was visualised using the online tool itol.

199

200

## 201 **Composition in other *U. humidicola* accessions**

202

203 Public DNA-seq data were downloaded from EBI’s ENA, including accessions CIAT 16888  
204 (SRR8321580), CIAT 26155 (SRR16327314), and CIAT 26146 (13 read files from PRJNA509199),  
205 which is a sexual accession. These three datasets were pre-processed using Trim Galore v.0.6.10  
206 (default parameters), then aligned to the assembled chromosomes with Bowtie v2.4.1 using the options  
207 “-X 500 --no-unal --no-mixed --no-discordant --no-contain --no-overlap”, ensuring only concordant  
208 paired reads were retained. The aligner’s default behaviour, where reads are aligned to their “top best”  
209 position, was applied. Since our alignment parameters only permit a read pair to align to their “top best”  
210 position, this results in a competitive mapping analysis where reads mainly map to their original  
211 subgenome for composition estimation. A coverage table was generated from the alignment files using  
212 Samtools coverage v1.51.1 (Danecek et al. 2021).

213

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

218

219

## 220 **RESULTS AND DISCUSSION**

221

### 222 **Genome assembly**

223

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

230

231 We aligned the Poaceae BUSCO markers to both unitigs and contigs, achieving 99.5% alignment  
232 success, with 99.1% being duplicated in both cases. We counted how many times each single-copy  
233 BUSCO marker aligned in the contigs (figure 2); 59.2 % of the markers aligned eight times (2899  
234 markers out of 4875), indicating that most haplotypes were present eight times. Assuming a base  
235 chromosome number of six, this led to an initial estimation of 48 chromosomes.

236

237 Since 97% of the genome was assembled into 200 contigs (figure 1), and we expected 48 chromosomes,  
238 we concluded that most chromosomes were assembled in just a few contigs (around four on average).  
239 Consequently, we decided to scaffold them using synteny, as detailed in the methods; firstly, comparing  
240 the hap1 and hap2 contigs to the primary contigs assembled by Hifiasm, then by comparing the "hap1"  
241 to the "hap2" homologs to each other, and finally by comparing to the longest assembled haplotype in  
242 each chromosome (often a full chromosome at this stage). In total, 225 contigs were scaffolded in 48  
243 chromosomes. Three chromosomes were assembled end-to-end in one contig, and nine more in two  
244 contigs. Only six chromosomes were assembled in seven or more contigs.

245

246 We annotated 259,254 genes and 277,99 transcripts. 249,665 genes (96.3%) were in chromosomes, and  
247 the rest were in contigs.

248

249

250 **Differences between ancestries**

251

252 Chromosomes were assigned to ancestries based on differential k-mer analysis among chromosome sets.  
 253 The SubPhaser tool clusters the subgenomes using k-means, from which it constructs a hierarchical  
 254 dendrogram and heatmap (figure 3A) based on the presence-absence variation of these differential k-  
 255 mers, as well as a PCA (figure 3B). The composition of cv. Tully was AABBBBCC for chromosomes 1  
 256 to 4, and AAABBBCC for chromosomes 5 and 6. This can also be expressed as 14A 22B 12C, or  
 257 12A+2 24B-2 12C.

258

259 A haplotype analysis based on chromosome alignment showed the shared synteny between subgenomes  
 260 (figure 4). The A and B copies of each chromosome were highly similar (green links in figure 4). The C  
 261 subgenome exhibited lower synteny (red links in figure 4). In all chromosomes, there was no synteny  
 262 with the centromeric regions on the C subgenome chromosomes; this is particularly clear in  
 263 chromosomes 3, 4 and 5. Synteny between C and the other two subgenomes was particularly low in  
 264 chromosome 5.

265

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

272

273 The clear clustering of A and B subgenomes in both differential k-mer and LTR analyses indicates that  
 274 these two ancestries diverged more recently from each other than from the C genome donor. The shared  
 275 peaks of amplification for Copia and Gypsy elements suggest that A and B lineages experienced similar  
 276 transpositional bursts, possibly during a common polyploidization event. In contrast, the C subgenome  
 277 shows a different repeat composition and older insertion ages, which align with a long history of  
 278 independent evolution before its introgression. Collectively, these findings support a stepwise

279 allopolyploid origin for *U. humidicola*, with the A–B component forming first, followed by the later  
280 addition of the diverging C genome.

281

282

## 283 **Ancestral origins**

284

285 Most *Urochloa* species, particularly wild ones, lack WGS data. This led us to use the “Angiosperms  
286 353” (A353) markers from the “Tree of Life” project, which had been generated for at least 15  
287 *Urochloa* species (Baker et al. 2022; Masters et al. 2024). For this, we identified and extracted the  
288 homologs to the A353 marker from the new genome in each *U. humidicola* subgenome and haplotype.  
289 The tree we obtained (figure 6) consistently placed the B subgenome with *U. dictyoneura* and separated  
290 it from other CWRs in the same clade. This suggests that *U. dictyoneura* is the ancestor of the B  
291 subgenome. These two species are known to intercross and share the same base chromosome number of  
292 six. The A subgenome was included in the Humidicola clade, but was not one of the wild relatives for  
293 which we had markers. The C subgenome is phylogenetically distant from the previous, and is likely *U.*  
294 *arrecta* from the Mutica clade (figure 6).

295

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

300

301

## 302 **Genomic composition in *U. humidicola***

303

304 We aligned the publicly available DNA-seq and RNA-seq data from *U. humidicola* to the reference  
305 genome. Our alignment parameters allowed each read to align only to its “top best” position (figure 7).  
306 CIAT-26155 (SRR16327314) underwent whole-genome resequencing and was identified as hexaploid

307 through k-mer analysis (Tomaszewska et al. 2023). Among the reads, 15,646,090 aligned to a  
308 chromosome from the A subgenome (33,32%), 31,076,742 to a chromosome from the B subgenome  
309 (66,19%), and 230,974 to a chromosome from the C subgenome (0,005%). The ratio of reads between  
310 the A and B subgenomes was clearly 1A:2B, supporting a composition of 12A and 24B (AABBBB) for  
311 CIAT-26155. Conversely, accession 16888, expected to be an aneuhexaploid (Worthington et al. 2019),  
312 showed about 40% of the reads aligned to A subgenome chromosomes and roughly 60% to B  
313 subgenome chromosomes, resulting in a 1A:1.5B ratio. This indicates an overrepresentation of A  
314 chromosomes compared to B, but could also be explained by unbalanced marker density, as this  
315 accession was genotyped using Genotyping-By-Sequencing (GBS). An 1A:1.5B ratio would be  
316 equivalent to the 14A 22B composition observed in the genome of cultivar Tully.

317

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

324

325 RNA-seq reads (Higgins et al. 2022) from four accessions (figure 7), including cv. Tully itself (labelled  
326 as CIAT-679), aligned to the C subgenome with a frequency of around 23%. None of the remaining  
327 accessions aligned to the C subgenome (<0.01% of the reads). These four accessions aligned to the A  
328 and B subgenomes with similar proportions, so we should assume a composition similar to cv. Tully. In  
329 Higgins et al. 2022, these four accessions were grouped into a different subpopulation than the other  
330 accession, called “humidicola-2” (Higgins et al. 2022). This evidence supports that these subpopulations  
331 were split by the presence or absence of C ancestry.

332

333 Among the remaining accessions, most displayed a ratio close to 1A:2B (about 33% reads aligned to A  
334 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 (figure 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 re-shuffle subgenomic composition through residual meiotic pairing.

### Evolution of *U. humidicola*

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

Initial hybridisation 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 can reflect additional copies 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. Further hybridisation among AABBBB cytotypes likely produced octoploid forms (AABBBBBB), exemplified by CIAT-1673 (12A 36B).

A subsequent introgression event involving the C genome from *U. arrecta* (or a very closely related species) led to the creation of the cultivar *U. humidicola* cv. Tully (AABBBBCC), 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.

The model demonstrates how recurrent hybridisation, partial diploidisation, and subgenome retention have shaped the mosaic ancestry of *U. humidicola* and explain the genomic diversity observed among apomictic and sexual accessions. The inferred hybridisation network (Figure 8) reflects the broader pattern of reticulate evolution observed across tropical forage grasses, where polyploidy and apomixis repeatedly facilitate the stabilisation of interspecific hybrids. Such processes parallel those seen in *Paspalum* and *Poa*, suggesting convergent evolutionary solutions to reproductive assurance under tropical conditions.

## **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 characterise 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 biological nitrification inhibition (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 Statement**

All data are deposited in the SRA under accession PRJEB90424. The genome assembly, together with its gene annotation, was deposited in ENA with accession GCA\_965614515.2 ([https://www.ebi.ac.uk/ena/browser/view/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>.

### **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.

### **Conflict of Interest**

None

### **Funder Information**

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

419 Grant (Decoding Biodiversity) BBX011089/1, and its constituent work package BBS/E/ER/230002B  
 420 (WP2 Genome Enabled Analysis of Diversity to Identify Gene Function, Biosynthetic Pathways, and  
 421 Variation in Agri/Aquacultural Traits). Funding was also received from the BBSRC's funded "Core  
 422 Strategic Programme Grant" (Genomes to Food Security) BB/CSP1720/1 and its constituent work  
 423 package BBS/E/T/000PR9818 (WP1 Signatures of Domestication and Adaptation), as well as  
 424 BB/X011089/1. Funding was also received from the Biotechnology and Biological Sciences Research  
 425 Council (BBSRC) as part of UK Research and Innovation, Core Capability Grant BB/CCG1720/1.

426

427

## 428 **Literature Cited**

429

430 Alonge M et al. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for  
 431 high-throughput genome editing. *Genome Biol.* 23(1):258

432 Arango J et al. 2020. Ambition meets reality: achieving GHG emission reduction targets in the livestock  
 433 sector of Latin America. *Front Sustain Food Syst.* 4:65

434 Baker WJ et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of  
 435 life. *Syst Biol.* 71(2):301–319

436 Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using  
 437 DIAMOND. *Nat Methods.* 18(4):366–368

438 Cervantes-Díaz CI, Oyama K, Cuevas E. 2024. Ecological and evolutionary implications of  
 439 hybridization, polyploidy and apomixis in angiosperms. *Bot Sci.* 102(4):1043–1061.

440 <https://doi.org/10.17129/botsci.3583>

441 Cheng H et al. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.  
 442 *Nat Methods.* 18(2):170–175

443 da Costa Lima Moraes A et al. 2023. Advances in genomic characterization of *Urochloa humidicola*:  
 444 exploring polyploid inheritance and apomixis. *Theor Appl Genet.* 136(11):238.  
 445 <https://doi.org/10.1007/s00122-023-04485-w>

446 Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience.* 10(2):giab008

447 Egenolf K et al. 2020. Brachialactone isomers and derivatives of *Brachiaria humidicola* reveal  
 448 contrasting nitrification inhibiting activity. *Plant Physiol Biochem.* 154:491–497.  
 449 <https://doi.org/10.1016/j.plaphy.2020.06.004>

450 EI-CoreBioinformatics: eirepeat - EI Repeat Identification pipeline. [accessed 2025 Oct 19].  
 451 <https://github.com/EI-CoreBioinformatics/eirepeat>

452 Ferreira RCU et al. 2021. An Overview of the Genetics and Genomics of the *Urochloa* Species Most  
 453 Commonly Used in Pastures. *Front Plant Sci.* 12 [accessed 2022 June 30].  
 454 <https://www.frontiersin.org/article/10.3389/fpls.2021.770461>

455 Gabriel L et al. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein  
 456 evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* 34(5):769–777

457 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome  
 458 assemblies. *Bioinformatics.* 29(8):1072–1075

459 Higgins J et al. 2022. Diverged subpopulations in tropical *Urochloa* (*Brachiaria*) forage species indicate  
 460 a role for facultative apomixis and varying ploidy in their population structure and evolution. *Ann Bot.*  
 461 130(5):657–669. <https://doi.org/10.1093/aob/mcac115>

462 Issifu S et al. 2024. Biological Nitrification Inhibitors with Antagonistic and Synergistic Effects on  
 463 Growth of Ammonia Oxidisers and Soil Nitrification. *Microb Ecol.* 87(1):143.  
 464 <https://doi.org/10.1007/s00248-024-02456-2>

465 Jia K-H et al. 2022. SubPhaser: a robust allopolyploid subgenome phasing method based on  
 466 subgenome-specific k-mers. *New Phytol.* 235(2):801–809

467 Johnson MG et al. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from  
468 high-throughput sequencing reads using target enrichment. *Appl Plant Sci.* 4(7):1600016

469 Johnson MG et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any  
470 flowering plant designed using k-medoids clustering. *Syst Biol.* 68(4):594–606

471 Krueger F. 2015. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality  
472 and adapter trimming to FastQ files, 516: 517.

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

476 Masters LE et al. 2024. Phylogenomic analysis reveals five independently evolved African forage grass  
477 clades in the genus *Urochloa*. *Ann Bot.* 133(5–6):725–742

478 Miles JW, editor. 1996. *Brachiaria: biology, agronomy, and improvement*. CIAT. (CIAT Publication).

479 Nuñez J et al. 2018. Biological nitrification inhibition activity in a soil-grown biparental population of  
480 the forage grass, *Brachiaria humidicola*. *Plant Soil.* 426(1):401–411

481 Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and  
482 phasing assessment for genome assemblies. *Genome Biol.* 21(1):245

483 Ryan C et al. 2025. A haplotype-resolved chromosome-level genome assembly of *Urochloa decumbens*  
484 cv. Basilisk resolves its allopolyploid ancestry and composition. *G3 GenesGenomesGenetics.*  
485 15(4):jkaf005. <https://doi.org/10.1093/g3journal/jkaf005>

486 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation  
487 Completeness. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols*. Springer; p 227–245  
488 [accessed 2024 Sept 27]. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14). [https://doi.org/10.1007/978-1-](https://doi.org/10.1007/978-1-4939-9173-0_14)  
489 [4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14)

490 Tomaszewska P et al. 2023. Complex polyploid and hybrid species in an apomictic and sexual tropical  
 491 forage grass group: genomic composition and evolution in *Urochloa* (Brachiaria) species. *Ann Bot.*  
 492 131(1):87–108

493 Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq.  
 494 *Bioinformatics.* 25(9):1105–1111

495 Underwood CJ, Mercier R. 2022. Engineering Apomixis: Clonal Seeds Approaching the Fields. *Annu*  
 496 *Rev Plant Biol.* 73(1):201–225. <https://doi.org/10.1146/annurev-arplant-102720-013958>

497 Vázquez E et al. 2020. Gross nitrogen transformations in tropical pasture soils as affected by *Urochloa*  
 498 genotypes differing in biological nitrification inhibition (BNI) capacity. *Soil Biol Biochem.* 151:108058

499 Venturini L et al. 2018. Leveraging multiple transcriptome assembly methods for improved gene  
 500 structure annotation. *GigaScience.* 7(8):giy093

501 Worthington M et al. 2019. Translocation of a parthenogenesis gene candidate to an alternate carrier  
 502 chromosome in apomictic *Brachiaria humidicola*. *BMC Genomics.* 20(1):41

503 Zhang R-G, Shang H-Y, Jia K-H, Ma Y-P. 2024. Subgenome phasing for complex allopolyploidy: case-  
 504 based benchmarking and recommendations. *Brief Bioinform.* 25(1):bbad513.  
 505 <https://doi.org/10.1093/bib/bbad513>

506

507

## Figure legends

Figure 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.

Figure 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 eight times, consistent with an octoploid genome structure.

Figure 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 three 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.

Figure 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.

Figure 5: Distribution and phylogeny of LTR-retrotransposons. Bar plots show the relative abundance of Gypsy and Copia 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.

535 Figure 6: Phylogenetic placement of *U. humidicola* subgenomes among related *Urochloa* species. The  
 536 maximum-likelihood tree, constructed using 353 protein markers (“Angiosperms 353”), shows the B  
 537 subgenome grouping with *U. dictyoneura* and the C subgenome with *U. arrecta*, while A subgenome  
 538 sequences cluster independently within the Humidicola clade.

539  
 540 Figure 7: Competitive read mapping of public *U. humidicola* accessions to subgenomes. Stacked bar  
 541 plots show the proportion of reads aligning to chromosomes of the A, B, and C subgenomes for each  
 542 accession. They were genotyped using RNA-seq except 26155 (labelled W for WGS), and 26146 and  
 543 16888 (labelled G for GBS). These ratios mainly reflect the variability in subgenome copy numbers and  
 544 ploidy levels across different accessions.

545  
 546 Figure 8. Evolutionary model of the composition among *Urochloa humidicola* accessions and putative  
 547 events leading to different *U. humidicola* cytotypes (black boxes), reflecting the species' complex  
 548 reticulated evolution. The B-genome donor *U. dictyoneura*, an A-genome wild *Urochloa* from the  
 549 Humidicola clade contributed to the common hexaploid (AABBBB), as well as increasing ploidies, such  
 550 as octoploid (AABBBBBB). While *U. arrecta* (C genome) contributed to AABBBBCC lineages,  
 551 including the apomictic cultivar Tully sequenced.

552

553

554

555 Table 1. Summary statistics of the *Urochloa humidicola* cv. Tully genome at different stages, unitigs,  
556 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       |

557

558

559

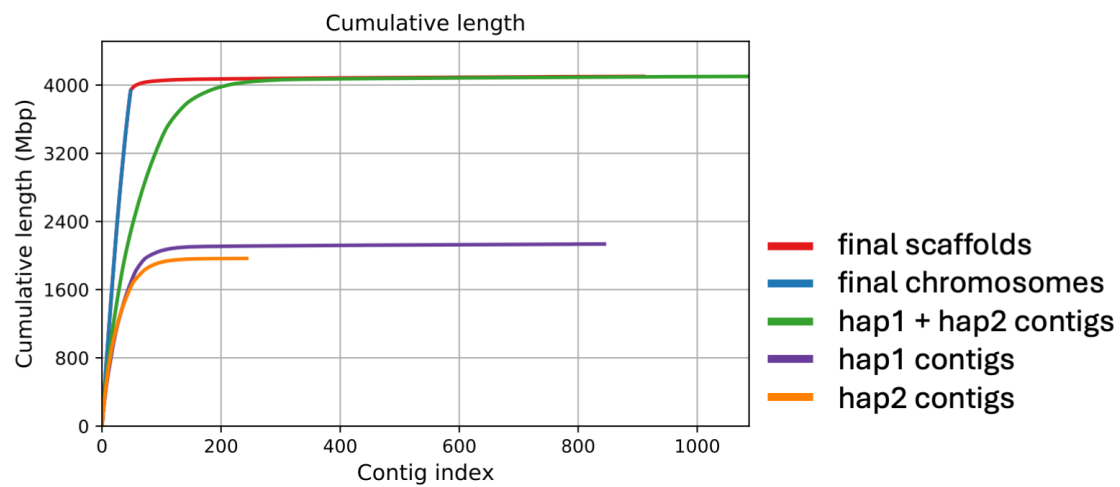


Fig 1

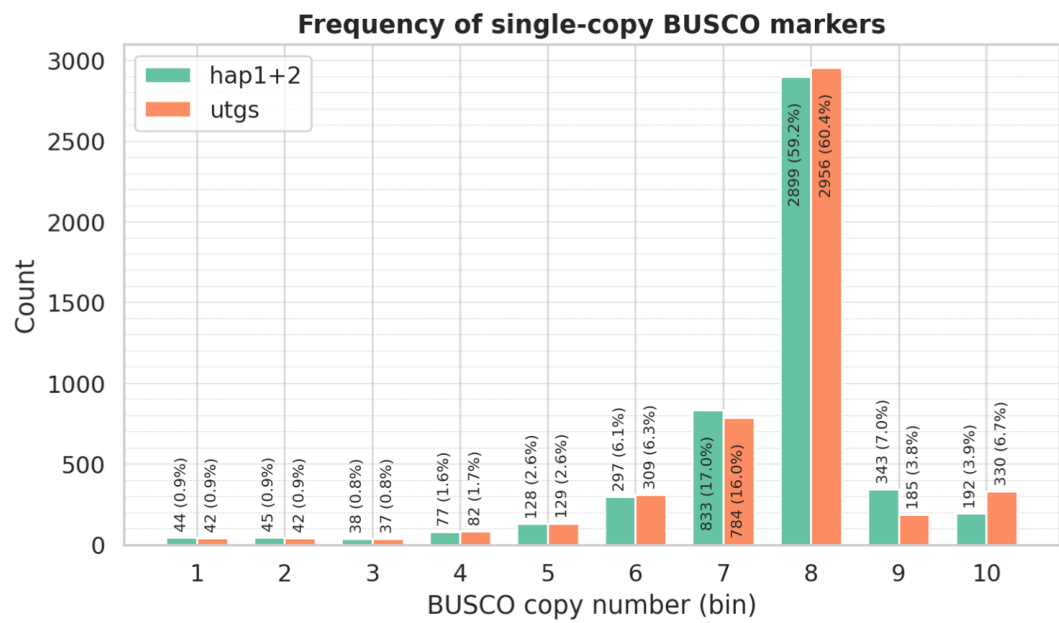


Fig 2

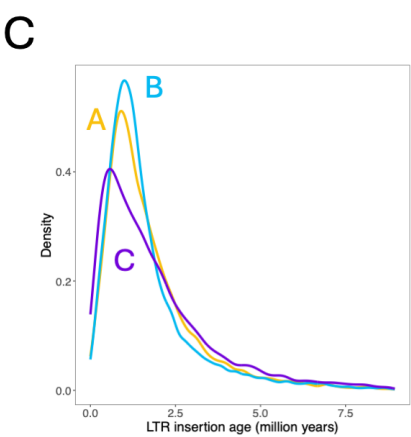
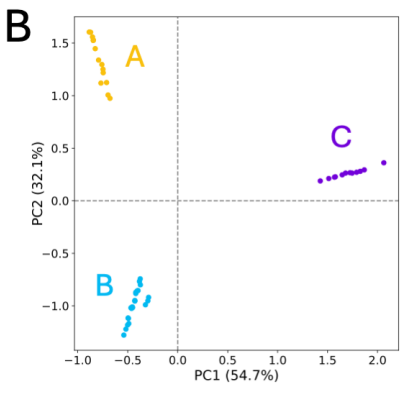
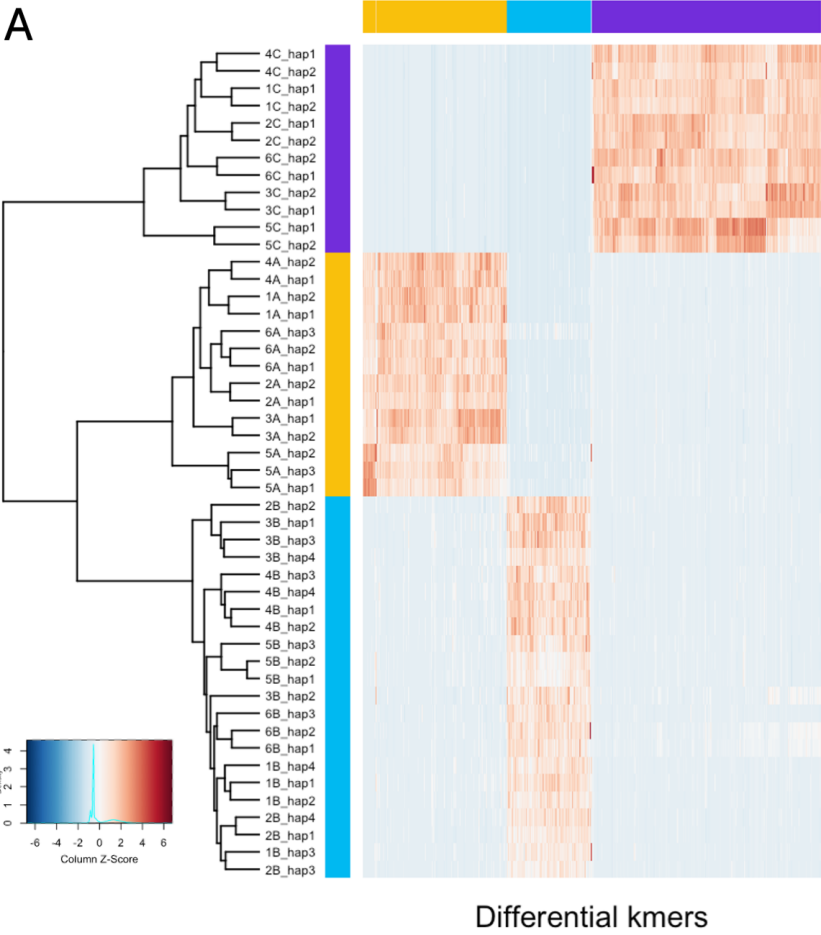


Fig 3

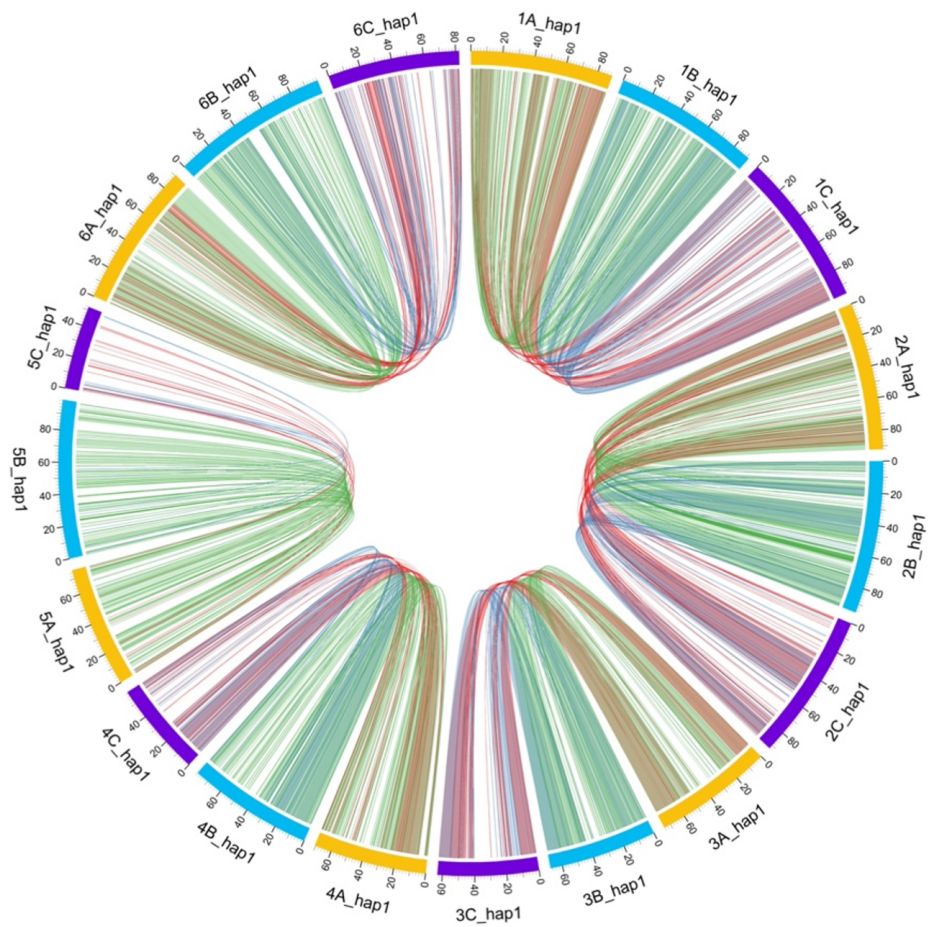


Fig 4

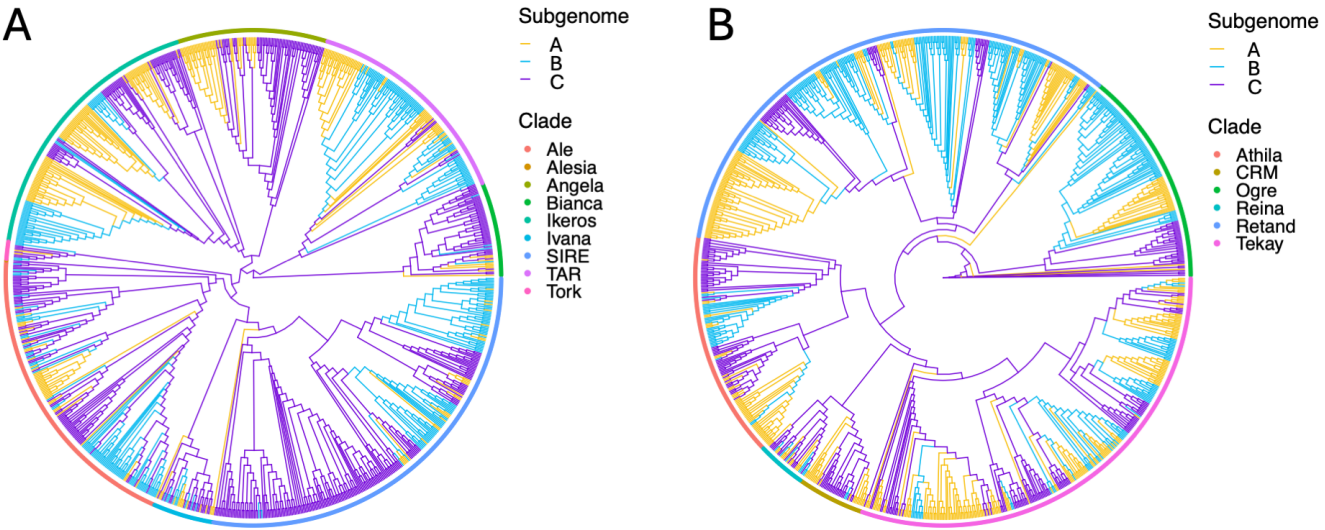


Fig 5

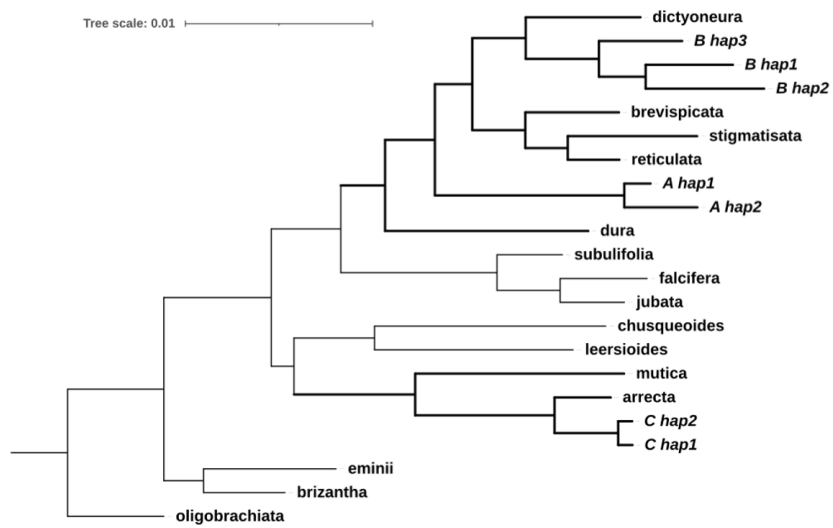


Fig 6

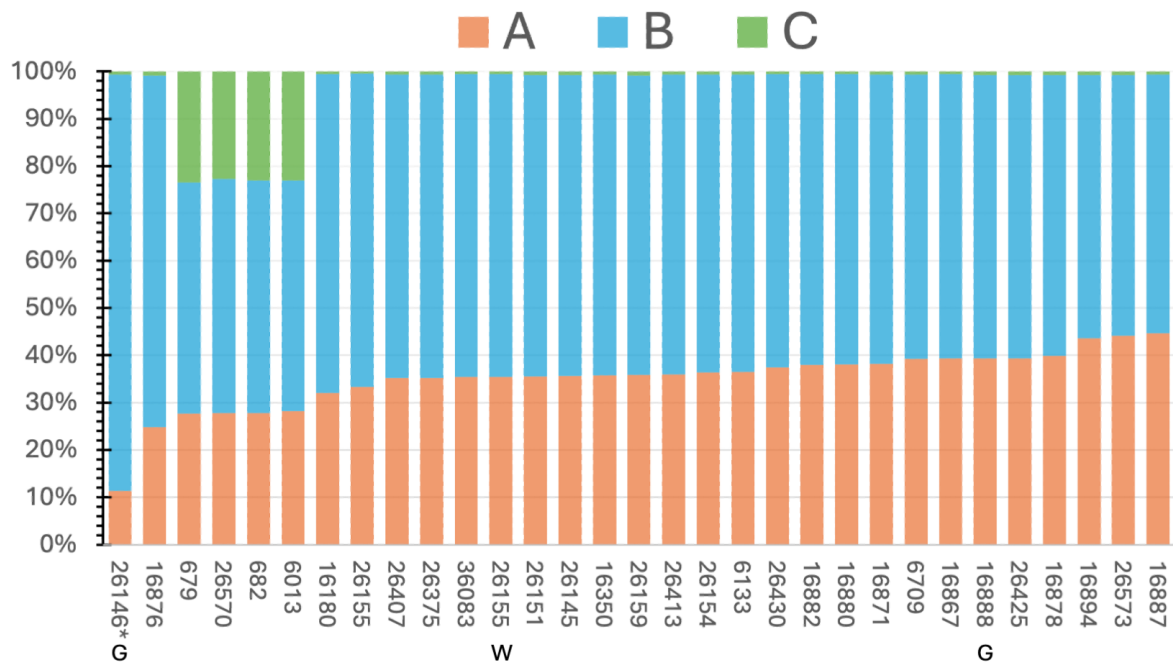


Fig 7

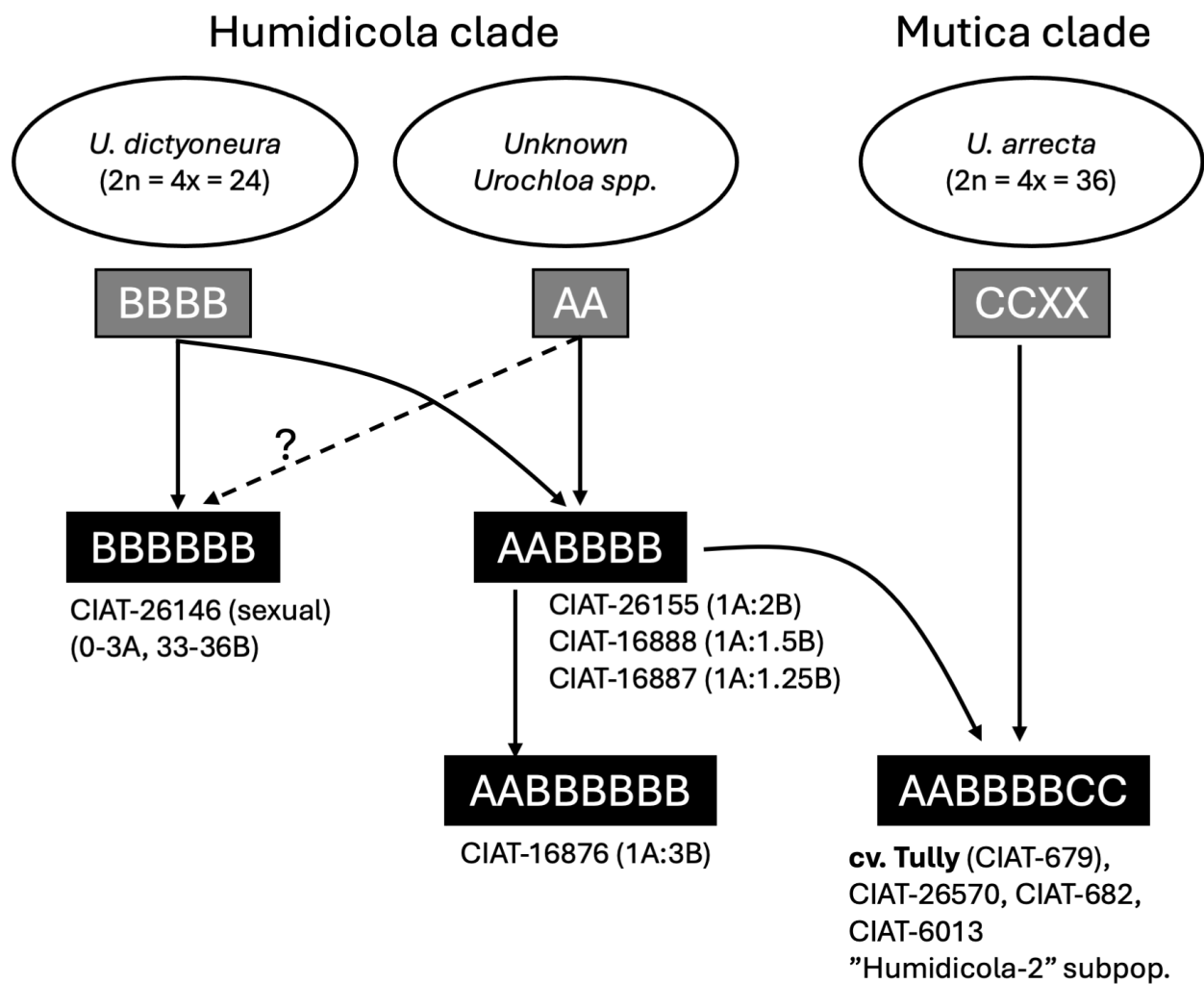


Fig 8