bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

The repetitive DNA sequence landscape and DNA methylation in chromosomes of an apomictic tropical forage grass, *Cenchrus ciliaris*

Priyanka Rathore[1], Trude Schwarzacher[2,3], J.S. (Pat) Heslop-Harrison[2,3,*], Vishnu Bhat[1] and Paulina Tomaszewska[2,4]

[1)]Department of Botany, Faculty of Science, University of Delhi, Delhi, India

[2)]Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, United Kingdom

[3)] Key Laboratory of Plant Resources Conservation and Sustainable Utilization / Guangdong Provincial, Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, 510650, China

[4)] Department of Genetics and Cell Physiology, Faculty of Biological Sciences, University of Wrocław, 50-328 Wrocław, Poland

[*)]Corresponding author. E-mail: phh4@leicester.ac.uk and phh@molcyt.com

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Abstract

Cenchrus ciliaris is an apomictic, allotetraploid pasture grass widely distributed in tropical and subtropical regions of Africa and Asia. In this work, we aim to investigate the genomic organization and characterize the nature of repetitive DNA sequences in this species. Because of the apomictic propagation, various aneuploid genotypes are found and we analysed here a $2n=4\times+3=39$ accession. The physical mapping of Ty1-*copia* and Ty3-*gypsy* retroelements through fluorescence *in situ* hybridization with global assessment of 5-methylcytosine DNA methylation through immunostaining revealed the genome-wide distribution pattern of retroelements and their association with DNA methylation. About a third of Ty1-*copia* sites overlapped or spanned centromeric DAPI positive heterochromatin, while the centromeric regions and arms of some chromosomes were labeled with Ty3-*gypsy*. Most of the retroelement sites overlapped with 5-methycytosine signals, except some Ty3-*gypsy* on the arms of chromosomes which did not overlap with anti-5-mC signals. Universal retrotransposon probes did not distinguish genomes of *C. ciliaris* showing signals in pericentromeric regions of all 39 chromosomes, unlike highly abundant repetitive DNA motifs found in survey genome sequences of *C. ciliaris* using graph-based clustering. Probes developed from RepeatExplorer clusters gave strong signals mostly in pericentromeric regions of about half of the chromosomes, and we suggested that they differentiate the two ancestral genomes in the allotetraploid *C. ciliaris* likely having different repeat sequence variants amplified before the genome came together in the tetraploid.

Key words: genome organization, retrotransposons, repetitive DNA sequences, DNA methylation pattern, apomixis, polyploidy, allotetraploid, buffelgrass

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Introduction

Cenchrus ciliaris L. (buffelgrass, buffel-grass, or African foxtail grass; syn. *Pennisetum ciliare*) is an apomictic, allotetraploid perennial C_4 grass native to tropical and subtropical arid regions of Africa and Western Asia. Predominantly distributed in warmer tropical regions, *C. ciliaris* grows in a range of harsh conditions (Marshall et al., 2012; Kharrat-Souissi et al., 2013, 2014). *Cenchrus* species are the predominant pasture grasses in India due to their high fodder value (Meena and Nagar, 2019). They are of great ecological importance due to persistence in native arid ecosystems, associated with tolerance to heavy grazing and drought, with deep root systems, and they are also used for erosion control (Miller et al., 2010). *C. ciliaris* is recognized as an invasive species in Australia (where it is a 'declared plant'), South America and the USA (particularly in Arizona and the Sonoran Desert) and became a threat to natural ecosystems (Jackson, 2005; Miller et al., 2010) both by displacement of native plants and by intensifying wildfires. It is important to study the genetic characteristics and genomic organization of *C. ciliaris* to understand the association between biological, agricultural, and ecological characteristics.

Cenchrus ciliaris, with a base chromosome number of ×=9, includes tetra-, penta-, hexa- and septa-ploid cytotypes (Visser et al., 1998; Kharrat-Souissi et al., 2013; Dhaliwal et al., 2018), and aneuploid accessions with 2*n*=32, 40, 43, 44 and 48 have also been found (Burson et al., 2012; Carloni-Jarrys et al., 2018). Most accessions are apomictic, although sexual plants are also found occasionally in nature (Fisher et al., 1954; Yadav et al., 2012; Marshall et al., 2012), with a DNA-based SCAR marker reported by Dwivedi et al. (2007). Apomictic seed production preserves superior genotypes, including heterozygotes and polyploids (Spillane et al., 2004; Bhat et al., 2005; Abdi et al., 2016) and can found new polyploid lineages. The character is well

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

67 represented in polyploid grasses including members of Paniceae grass genera such as *Panicum,*
68 *Pennisetum, Cenchrus* and *Paspalum* (Worthington et al., 2016; Higgins et al., 2021;
69 Tomaszewska et al., 2021b). In *Urochloa* (*Brachiaria*), there is a robust molecular marker
70 discriminating sexual and apomictic plants. The genes involved in *C. ciliaris* have been
71 examined and compared with other species (Roche et al., 1999; Akiyama et al., 2005; Yadav et
72 al., 2012). Like the widely grown *Urochloa* (*Brachiaria*) tropical forage grasses (Tomaszewska
73 et al., 2021a, b), apomicts at all ploidies produce seeds as there is no meiosis, nor necessity for
74 restitution of diploid pairing behaviour. Although there is at best weak evidence that polyploid
75 plants as a group outperform diploids either in wild or agricultural settings (particularly where
76 selection is strong enough to remove mildly deleterious mutations), there are many prospects to
77 exploit polyploidy systematically using additional gene alleles and increased genetic variability
78 which enhance adaptability (Brits et al., 2003; Bhat et al., 2005; Kharrat-Souissi et al., 2014).

79 *Cenchrus ciliaris* has a genome size of approximately 3,000 Mbp in the tetraploid species, 50%
80 larger than the 1,971 Mbp reported for the genome of *C. purpureus* (Elephant or Napier grass;
81 2*n*=4×=28) and the more distantly related but smaller 1,580 Mbp *C. americanus* (2*n*=2×=14;
82 syn. *Pennisetum americanum*) genome (Yan et al., 2021). A major proportion of plant genomes
83 consists of repeated sequences such as ribosomal DNAs, tandem repeats, satellites, and
84 transposable elements which play an important role in shaping of genome organization (Heslop-
85 Harrison and Schwarzacher, 2011; Macas et al., 2011; Biscotti et al., 2015). Yan et al. (2021)
86 reported the composition of the *C. purpureus* genome as comprising 66.32% repetitive
87 sequences, similar to the proportion of repeats in *C. americanus* (77.2%). Ty3-*gypsy* and Ty1-
88 *copia* are two super-families of LTR-retroelements widely distributed at the centromeric region

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

89 of chromosomes in many members of Poaceae and terminal heterochromatic chromosomal
90 regions in *Allium cepa*, respectively (Kumar et al., 1997).

91 Members of the major groups of repetitive elements, tandemly repeated satellites and
92 transposable elements (Class I retrotransposons and Class II DNA transposons) differ in their
93 abundance and genomic distribution in the genome. Tandemly repeated motifs may be located in
94 sub-terminal, intercalary or centromeric regions of chromosomes, or, apart from the 5S and 45S
95 rDNA arrays, not be abundant. In contrast, all plants have relatively abundant retrotransposons,
96 often with a dispersed distribution (Heslop-Harrison et al., 1997), often dispersed throughout the
97 euchromatin but absent from specialized regions such as heterochromatin, centromere, telomeres
98 and nuclear organization regions (NOR). Retrotransposon families may also be restricted to
99 characteristic genomic locations. In *Arabidopsis thaliana* many Ty3-*gypsy* retroelements are
100 present in pericentromeric heterochromatin region (Brandes et al., 1997b). Centromeric
101 retroelements in maize (CRM) is a lineage of Ty3-*gypsy* retroelements (Jin et al., 2004), and the
102 diploid oat *Avena longiglumis* (2*n*=2×=14; 3,850 Gb genome size) has specific LTR
103 retrotransposons located at the centromeres (Liu et al., 2022), in all cases with other retroelement
104 families dispersed along all chromosomes. Some of these repetitive elements may have a role in
105 centromere functionality, although they may also accumulate in regions of the genome where
106 there are few genes.

107 The activity of retroelements is modulated by epigenetic mechanism through DNA methylation
108 (Slotkin and Martienssen, 2007; Le et al., 2015), where methyl group is covalently added to the
109 fifth carbon of cytosine residues to form 5-methylcytosine (5-mC) (Moore et al., 2013).
110 Expression profile of retroelements can be correlated with the differential methylation pattern
111 (Castilho et al., 1999). DNA methylation is mainly distributed in heterochromatic regions in

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

112 genome that are rich in repetitive sequences consisting of retroelements (Vining et al., 2012).
113 Comparison of methylome between the tissues or of different stages can be studied to know
114 genes regulated through methylation and the pathways involved in response to stress and
115 developmental conditions (Li et al., 2020). Whole genome methylation analysis can reveal the
116 distribution pattern of methylation in the genome and the associated genes or region. In *A.*
117 *thaliana*, genome-wide methylation analysis found that a third of genes are methylated within
118 their transcribed regions while in rice about 16% of genes are enriched with methylation
119 (Zilberman et al., 2007; He et al., 2010; Vining et al., 2012). Genome methylation analysis in
120 *Populus trichocarpa* also indicated that Ty3-*gypsy* retroelements which are abundant in
121 centromeric, pericentromeric and heterochromatic regions in plants are enriched with DNA
122 methylation (Douglas and Di Fazio, 2010; Vining et al., 2012). Therefore, the methylation status
123 of retroelements in the genome can be assessed by combining the physical mapping of
124 retroelements through fluorescent *in situ* hybridization (FISH) with assessment of 5-
125 methylcytosine through immunostaining which enables us to understand the genome-wide
126 distribution pattern of retroelements and their association with the methylation.

127 In this study, we aimed to characterize the nature of repetitive DNA in *Cenchrus ciliaris*,
128 investigate the genomic organization, including tandemly repeated satellite DNA and LTR
129 retroelements, and examine genome wide methylation patterns in *C. ciliaris*, aiming to
130 characterize features related to its contribution to the genome, evolution and possible variation in
131 the genomes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Materials and Methods

Fixation of plant material

Seeds of *Cenchrus ciliaris* (CcA7-5) were germinated in Petri dishes at 25° C for three days and root tips were collected. Seedling root tips were pretreated in alpha-bromonaphthalene for two hours to accumulate metaphases and then fixed in ethanol: glacial acetic acid (3:1) solution and stored at 4° C (Schwarzacher and Heslop-Harrison, 2000).

Preparation of chromosome spreads

Root tips were washed twice in enzyme buffer (10 mM citric acid/sodium citrate) for 15 min and then digested with enzyme solution (20 U mL^{-1} cellulase, 10 U mL^{-1} Onozuka RS cellulase, and 20 U mL^{-1} pectinase in enzyme buffer) for 45 min at 37° C. After digestion, root-tips were again washed in enzyme buffer. The meristematic tissues were dissected and squashed in 60 % acetic acid under coverslip by using light thumb pressure. Slides were frozen on dry-ice for 5-10 min and then coverslips were removed with a razor blade. Slides were air dried and used for fluorescent *in situ* hybridization.

Isolation of genomic DNA and designing of probes for repetitive sequences

Genomic DNA was isolated from the apomictic genotype *C. ciliaris* (CcA7-5) using the Qiagen DNeasy Plant mini kit. Primers were designed from conserved regions of the reverse transcriptase domain of LTR retroelements (Ty1-*copia* and Ty3-*gypsy*) and 45S and 5S ribosomal DNA genes using BLAST search to identify the conserved region (Sayers et al., 2019; Table 1). Whole genome sequencing data from a Chinese variety of *C. ciliaris* (deposited in SRA Sequence Read Archive under accession SRR8666664, Illumina HiSeq 2500; Nevill et al., 2020)

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

were downloaded and used to discover the most abundant repetitive DNA sequences. Similarity-based clustering, repeat identification and classification of a subset of raw reads were performed using RepeatExplorer2 (Novak et al., 2013) after quality-trimming, excluding overlapping read pairs, and interlacing. The extracted clusters were analysed by BLAST search to check for repeat identification (Sayers et al., 2019). The primers were designed (Table 1) using Primer3 (Rozen and Skaletsky, 1999). Conserved regions were amplified from genomic leaf DNA in a standard polymerase chain reaction (PCR) using specific primers synthesized commercially (Sigma). PCR products were labeled with biotin-16-dUTP or digoxigenin-11-dUTP using BioPrime CGH array labelling kit (Invitrogen), and then purified using BioPrime Purification Module (Invitrogen). Oligonucleotide probes were synthesized incorporating fluorescent labels (Table 1; Sigma).

Fluorescence *in situ* hybridization

Slides were re-fixed in ethanol:glacial acetic acid (3:1) for 30 min, dehydrated in 100% ethanol twice for 5 min and air dried. Slides were treated with RNase (100µg/ml) in 2X SSC (for 1 hr at 37^{o} C in a humid chamber and then washed twice in 2X SSC for 5 min at room temperature. Then slides were incubated in 0.01M HCL for 2 min, treated with pepsin (5 µg mL^{-1}) in 0.01M HCl for 10 min at 37 ^{o}C in humid chamber, rinsed in distilled water for 1 min and washed in 2X SSC for 5 min. Slides were re-fixed in 4% formaldehyde at room temperature for 10 min, washed in 2X SSC for 2 min and then in 2X SSC for 10 min. Slides were then dehydrated in an ethanol series (70%, 85% and 100% ethanol, 2 min each).

Hybridization mixture consisted of probes (2 ng µL^{-1} each), 50% formamide (v/v), 2X SSC, 10% dextran sulphate (v/v), 200 ng µL^{-1} salmon sperm DNA, 0.125mM EDTA and 0.125% SDS for total volume of 40 µL was denatured at 83^{o} C for 10 min and then cooled down on ice for 10

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

min. The hybridization mixture was then added to slides and placed in a hybridization oven at 73° C for 6 min to denature probe and chromosomes and then cooled to 37° C and left at that temperature overnight for hybridization.

Post hybridization washes were performed to remove unincorporated and weakly hybridized probe. Slides were washed twice in 2X SSC for 2 min at 42° C and then washed in 0.1X SSC at 42° C for 2 min and then for 10 min at the same temperature before a final wash in 2X SSC for 5 min at room temperature. To detect probes, slides were incubated in detection buffer (4X SSC containing 0.2% Tween 20) for 5 min, blocked by incubation with 5% BSA (bovine serum albumin) in detection buffer for 10 min at 37° C in humid chamber. Hybridization sites were detected by incubation with streptavidin conjugated to Alexa 594-conjugated streptavidin antibody or FITC (fluorescein isothiocyanate)-conjugated anti-digoxigenin antibody in detection buffer with 5% BSA for 1 hr at 37° C in humid chamber. Slides were washed in detection buffer thrice at 40° C for 10 min each. Chromosomes were then counterstained with DAPI (4',6-diamidino-2-phenylindole) in antifade solution (AF1, Citifluor).

Methylation analysis using anti-5-methylcytosine through immunostaining

Immunostaining was performed to visualize how whole-genome cytosine methylation is distributed in different chromosomal regions of *C. ciliaris* genome. Slides were pre-treated with RNase and pepsin, dehydrated in ethanol series and air dried according to FISH protocol. Slides were then blocked by 1% BSA (w/v) in 1X PBS buffer/0.5% Tween 20 for 30 min at room temperature and incubated with primary anti-5-methylcytosine antibody (anti-5-mC; 1:500 in 1X PBS buffer/0.5% Tween 20) for 1 hr at 37° C in a humid chamber. Slides were then washed in 1X PBS buffer/0.5% Tween 20 twice for 5 min each at room temperature and incubated with

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Alexa 488-conjugated goat anti-mouse secondary antibody (1:500) in 1X PBS buffer/0.5% Tween 20. Slides were then washed twice in 1X PBS buffer/0.5% Tween 20 for 5 min at room temperature and stained with DAPI in antifade solution (AF1, Citifluor).

Photography and imaging

Slides were examined with a Nikon Eclipse 80i epifluorescence microscope equipped with DS-QiMc monochromatic camera, and NIS-Elements v.2.34 software (Nikon). Images were analysed using Adobe Photoshop software.

Results

Chromosomes of *Cenchrus ciliaris*

Most plants of the apomictic tetraploid *C. ciliaris* were $2n=4\times=36$ but one hyperploid $2n=4\times=39$ plant was found (**Fig. 1A**); all chromosomes were submetacentric and of similar length. Three major signals of 5S rDNA at intercalary regions were present (**Fig. 1B**). The 45S rDNA probe located at four sub-terminal regions (**Fig. 1C, D arrows**), also visible as the satellite at the nucleolar organizing region (NORs by DAPI staining).

The degenerate PCR product used as a 'universal' Ty1-*copia* **(Table 1)** probe labelled broad regions of chromosomes, with sites often overlapping or spanning centromeric DAPI positive heterochromatin (**Fig 2A i**). Along with the centromeric regions, arms of most chromosomes were labeled with Ty3-*gypsy*, but most distal ends lacked signal (**Fig 2B i**). Probe Gy105 (**Table 1**), designed from a sequence including a Ty3-*gypsy* element ED545466 identified in a BAC by Conner et al. (2008) (**Fig. 1D**, green signal) labelled regions near centromeres on most but not all chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Immunostaining with anti-5-methylcytosine antibody showed genome-wide methylation patterns in apomictic *C. ciliaris* along chromosomes, with variability in patterns observed. About a third of chromosome arms were more weakly labelled particularly at the ends of the arms, most centromeric regions showed small gaps, and a few chromosomes were strongly labelled (**Figs. 2A ii** and **2B ii**).

Chromosomal locations of the anti-5-methylcytosine antibody were compared with retroelement distribution by sequential labelling of the same metaphase preparations. Anti-5-mC clusters overlapped the centromeric Ty1-*copia* probe signals of most of the chromosomes (**Fig. 2A iii**). Immunostaining of chromosomal spread previously probed with Ty3-*gypsy* retrotransposons showed methylation signals of anti-5-mC overlapped with Ty3-*gypsy* retroelements on most of pericentromeric and centromeric regions of chromosomes. However, some Ty3-*gypsy* signals on the arms of chromosomes did not overlap with anti-5-mC signals (**Fig. 2B iii; arrows**). Along chromosome arms, anti-5-mC signal was generally more widespread than either retroelement.

Chromosomal organization of highly abundant repetitive DNA motifs

Graph-based clustering (RepeatExplorer) of survey sequence of *Cenchrus ciliaris* (SRR8666664; Nevill et al., 2020) identified highly abundant repetitive DNA motifs in the genome (**Fig. 3A**). Cluster1 (**Supplementary Data Table S1; Fig. 3B**), accounted for 2.5% of the *C. ciliaris* genome, and is a putative low-confidence satellite of 140bp monomer length (**Fig. 3C**), part of which is homologous to a similarly sized satellite reported in *Setaria viridis* (Thielen et al., 2020). Two oligonucleotide probes (44 and 45 bp long) were designed from two different parts of this sequence: CenchStCL1ConsS and CenchEnCL1ConsRC (**Table 1**). Cluster58 (**Supplementary Data Table S1; Fig. 3B, C**) is identified as a satellite with a repeat unit of

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

1326bp that represents a much smaller proportion, 0.061% of the genome, and PCR primers and a 44bp probe were designed for this sequence (**Table 1**). Cluster22 (**Supplementary Data Table S1; Fig. 3B, C**) is identified as a 4807bp Ty1-*copia* retrotransposon, and the oligonucleotide probe CenchCL22Copia was designed to the RNaseH domain (RC) from this sequence (**Table 1**). The probes were localized on metaphase chromosomes of *C. ciliaris* using fluorescence *in situ* hybridization.

CenchEnCL1ConsRC probe gave strong signals mostly in pericentromeric regions of about twenty chromosomes from each metaphase, and the remaining chromosomes showed very weak signals (**Fig. 4A**). Twenty very strong signals of CenchStCL1ConsS probe were also detected in pericentromeric regions of chromosomes, and some minor signals, mostly intercalary, were visible (**Fig. 4B**). CenchStCL1ConsS collocalized with CenchCL22Copia signals (**Fig. 4C**). Both probes are good chromosome markers distinguishing about half the chromosomes. Amplified CenchCL58 probe and commercially synthesized CenchCL58_869 showed dispersed signals along all chromosomes from metaphase (**Fig. 4D**). Stronger signals of CenchCL58_869 probe collocalized with CenchEnCL1ConsRC (**Fig. 4E**), with 19 weaker and 20 stronger centromeric sites.

Discussion

Aneuploidy and polyploidy in *Cenchrus ciliaris* – consequences for forage grasses

C. ciliaris is an allotetraploid with chromosome number 2*n*=4×=36 (Akiyama et al., 2011) and a primary base number of ×=9. In this study, we reported a new chromosomal count of *C. ciliaris* with 2*n*=4×+3=39. Aneuploidy in *C. ciliaris* has been widely reported (Visser et al., 1998), with

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Burson et al. (2012) finding 99 aneuploids from 568 accessions, with different chromosome numbers from 2*n*=18 to 2*n*=56 (Vij and Chaudhary, 1981). Hyperploidy or monosomic addition lines are well-known in tetraploids and hexaploids in many Poaceae members, including *C. ciliaris* (2*n*=6×+1=55). Meiotic irregularities have been widely reported in *C. ciliaris*, presumably because of the close genomic relationships in the polyploid (Visser et al., 1998; Burson et al., 2012). Where seed production is continued through apomixis, the predominant pathway in *C. ciliaris*, there is low selective pressure for euploids. Polyploidization has significant conservational and environmental consequence where one species can become invasive due to adaptation of a species in a new locality and can also change reproductive biology (Thompson and Pellmyr, 1992; John et al., 2004). Thus, characterization of the genome constitution of apomictic *C. ciliaris* is important both genetically and environmentally.

Aneuploid and polyploid cytotypes are widely distributed within tropical forage grasses, including species of the genera *Cenchrus, Urochloa, Panicum, Pennisetum,* and *Paspalum* (Carloni-Jarrys et al., 2018; Worthington et al., 2019; Tomaszewska et al. 2021a, b). Polyploids can arise from spontaneous chromosome doubling following non-disjunction after meiotic or mitotic divisions (Gallo et al., 2007; Burson et al., 2012; Worthington et al., 2019), and often involve fertilization with unreduced (2*n*) gametes. Many species, such as those in the Triticeae tribe, have mechanisms that control chromosome pairing (Luo et al., 1996; Sepsi et al. 2017) and restore diploid meiotic behaviour and fertilization, with elimination of unpaired chromosomes. In contrast, where seed is produced through apomixis, many ploidies and aneuploid lines can be maintained and propagated unchanged (maintaining also any heterozygosity or heterosis in the maternal genotype; Lippman and Zamir, 2007). Apomixis enables propagation of outstanding heterozygous genotypes without vegetative propagation or seeds, which would be especially

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

valuable for domesticated grass breeding (Ozias-Akins and van Dijk, 2007). Along with the $2n=4\times+3=39$ *C. ciliaris* reported here, there are many examples of aneuploids in apomicts at various ploidies such as *Urochloa humidicola*, where an aneuploid accession with $2n=8\times+2$ (or $9\times-4) = 50$ has been found (Tomaszewska et al., 2021a). However, unlike tropical forage grasses, including segmental allopolyploids (Jessup et al., 2003; Worthington et al., 2019; Tomaszewska et al. 2021b), cereal grain-crop species tend to exhibit variation in levels of ploidy and relatively low genome homology, which constitute a barrier for crossbreeding with apomicts and make it impossible to achieve agronomically acceptable genotypes. Identification of loci that firstly control heterotic phenotypes and secondly parthenogenesis or apomixis in tropical forage grasses is important for forage grass improvement, and further work may lead to suitable systems to exploit apomixis in grain crops, regarded as a significant challenge for wheat and rice (Dresselhaus et al., 2001; Conner and Ozias-Akins, 2017; Rathore et al. 2020), thus fixing heterozygosity and removing the requirement for sexual seed production.

Chromosomal localization of highly abundant repetitive DNA motifs

Various approaches have been used to measure the abundance and genome organization of repetitive DNA in the plant genome, and assemblies from short reads will never give the amount present because of collapse of similar repeats during assembly (e.g. Yan et al., 2021). Graph-based clustering of unassembled raw reads, using RepeatExplorer (Novak et al., 2013) is proving to be a reference-free approach to characterize the number and abundance of all repeat families in a genome. Analysis of the frequency of k-mers – oligonucleotide sequences k bases long, where k is typically 16 to 64bp – has also proved a robust method to identify repetitive motifs (Liu et al., 2019; Tomaszewska et al., 2021b).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Here, we used both specific probes developed using graph-based sequence clustering compared to 'universal' conserved regions of Ty1-*copia* and Ty3-*gypsy* retroelements. Similarly to *Urochloa* tropical forage grasses, where genomic DNA and retroelements used as probes showed no substantial differences between genomes (Santos et al., 2015; Tomaszewska et al., 2021b), parallel universal retroelement repeats used in this study were relatively equally abundant over the whole *Cenchrus ciliaris* genome (see **Fig. 3**). The Gy105 probe (see **Fig. 1D**) labelled some chromosomes more weakly. In contrast, Cluster1 (putative satellite) and Cluster22 (LTR-copia) probes showed clear differential labelling of about half the chromosomes (see **Fig. 4**), and we suggest differentiate the two ancestral genomes in the allopolyploid *C. ciliaris* having different retroelement (and/or satellite) sequence variants amplified before the genome came together in the tetraploid. As might be expected, the 'universal' retroelement probes amplified abundant sequences from both genomes and did not show any genome differentiation (see **Fig. 2**).

Methylation pattern and major repeat distribution

The genome distribution and organization of retroelements is important for understanding retroelements dynamics, movement or amplification, chromosomal structure, and the evolutionary processes. In this study, centromeric DAPI-staining bands (corresponding to constitutive heterochromatin; Siljak-Yakovlev et al., 2002) collocated with Ty1-*copia* probes, suggesting accumulation of these retroelements at heterochromatic regions. Both Ty1-*copia* and Ty3-*gypsy* tended to cluster at specific regions on most of chromosomes (along with some intercalary and dispersed sites) contrasting with other species (Brandes et al., 1997a) while Ty1-*copia* was more uniformly distributed on the euchromatin regions in the genome. Whether the non-uniform distributions of retroelements is due to insertion-site preference in the genome

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

(Belyayev et al., 2001), perhaps of particular element subfamilies, or purging from genome regions, will need further study.

A combination of *in situ* hybridization with immunostaining using anti-5-mC allows correlation of the distribution of DNA methylation and sequence location across the genome (Sepsi et al., 2014). 5-mC has been extensively studied for its role in regulation of gene expression, genome imprinting, and suppression of transposable elements (Zhang et al., 2010). In this study, anti-5-methylcytosine immunofluorescence overlapped with most of the Ty1-*copia* elements and Ty3-*gypsy*, showing their location predominantly in methylated genomic regions associated with lower transcriptional activity. With the majority of 5-mC signal associated with retrotransposon-rich regions, there was no further differentiation of groups of chromosomes, and, in particular, no indication of differences between ancestral genomes in the polyploid. In contrast, differential methylation of parental genomes has been observed in intraspecific *A. thaliana* hybrids and implicated in heterosis as a key epigenetic factor (Lauss et al. 2018). However, in rice tetraploid hybrids, Li et al. (2019) reported overall methylome stability with regional variation of cytosine methylation states, in part associated with gene expression changes. Methylation repatterning, often stimulated by hybridization and/or polyploidization, can potentially lead to increased genetic variation through facilitating somatic recombination, which could be adaptive in apomictic lineages (Verhoeven et al., 2010). Changes in methylation pattern were observed both in the newly formed and in the established apomicts of *Taraxacum*, suggesting that considerable methylation variation can increase over short evolutionary time scales.

Author Contributions

Conceptualization, all authors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Funding acquisition, PR, VB, PHH and PT.

Investigation and Methodology, PR, TS and PT

Supervision, VB and PHH

Writing original draft review & editing, all authors.

All authors have read and agreed to the published version of the manuscript.

Funding acknowledgement

P.R. was supported under the Newton Bhabha Ph. D placement programme fellowship coordinated by Department of Biotechnology, India and British Council, United Kingdom (BTIIN/UKJDBT-BC/2017-18). This work was supported under the RCUK-CIAT Newton-Caldas Initiative "Exploiting biodiversity in *Brachiaria* and *Panicum* tropical forage grasses using genetics to improve livelihoods and sustainability", with funding from UK's Official Development Assistance Newton Fund awarded by UK Biotechnology and Biological Sciences Research Council (BB/R022828/1). P.T. has received support from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreements No 844564 and No 101006417 for analysis of polyploid chromosomal evolution.

Data availability statement

Previously published sequence data is available under accession SRR8666664 (Nevill et al. 2020). Results of the RepeatExplorer2 analysis of these sequence reads are archived on Figshare at http://dx.doi.org/10.25392/leicester.data.19798966.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

References

Abdi, S., Dwivedi, A., and Bhat, V. (2016). Harnessing apomixis for heterosis breeding in crop improvement. In: VR Rajpal, SR Rao, SN Raina (ed.) Molecular breeding for sustainable crop improvement. Springer, Cham. Pp. 79-99.

Akiyama, Y., Goel, S., Conner, J. A., Hanna, W. W., Yamada-Akiyama, H., and Ozias-Akins, P. (2011). Evolution of the apomixis transmitting chromosome in Pennisetum. *BMC Evolutionary Biology* 11:289.

Akiyama, Y., Hanna, W. W., and Ozias-Akins, P. (2005). High-resolution physical mapping reveals that the apospory-specific genomic region (ASGR) in *Cenchrus ciliaris* is located on a heterochromatic and hemizygous region of a single chromosome. *Theoret. Appl. Genet.* 111, 1042–1051.

Belyayev, A., Raskina, O., and Nevo, E. (2001). Chromosomal distribution of reverse transcriptase-containing retroelements in two Triticeae species. *Chromosome Research* 9, 129-136.

Bhat, V., Dwivedi, K. K., Khurana, J. P., and Sopory, S. K. (2005). Apomixis: An enigma with potential applications. *Current Science* 10, 1879-93.

Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (2015) Repetitive DNA in eukaryotic genomes. Chromosome Research 23, 415-420.

Brandes, A., Heslop-Harrison, J. S., Kamm, A., Kubis, S., Doudrick, R. L., *et al.* (1997a). Comparative analysis of the chromosomal markers and genomic organization of *Ty*1-*copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* 33, 11–21.

Brandes, A., Thompson, H., Dean, C., Heslop-Harrison, J. S. (1997b). Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. *Chromosome Research* 5(4), 238-46.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

394 Brits, G., Trytsman, M., Smith, A., Ndlovu, S., Mogale, E., and Molifi, R. (2003). Screenings of
395 subtropical grass species for drought tolerance, pp. 1348–1350, in Proceedings of the VII
396 International Rangelands Congress, 26 July–1 August, Durban, South Africa.

397 Burson, B. L., Actkinson, J. M., Hussey, M. A., and Jessup, R. W. (2012). Ploidy determination
398 of buffel grass accessions in the USDA National Plant Germplasm System collection by flow
399 cytometry. *South African Journal of Botany* 79, 91–95.

400 Carloni-Jarrys, E. J., Acosta-Bragato, M. C., and Grunberg-Fraga, K. (2018). Nuclear DNA
401 content and ploidy level in apomictic buffelgrass genotypes. *Revista Fitotecnia Mexicana* 41(1),
402 23-29.

403 Castilho, A., Neves, N., Rufini-Castiglione, M., Viegas, W., and Heslop-Harrison, J. S. (1999).
404 5-Methylcytosine distribution and genome organization in Triticale before and after treatment
405 with 5-azacytidine. *Journal of Cell Science* 112, 4397-4404.

406 Chang, K.-D., Fang, S.-A., Chang, F.-C., and Chung, M.-C. (2010) Chromosomal conservation
407 and sequence diversity of ribosomal RNA genes of two distant *Oryza species*. *Genomics* 96,
408 181–190.

409 Conner, J. A., Goel, S., Gunawan, G., Cordonnier-Pratt, M.-M., Johnson, V. E., Liang, C., Wang,
410 H., Pratt, L., Mullet, J. E., Debarry, J., Yang, L., Bennetzen, J. L., Klein, P. E. and Ozias-Akins,
411 P. (2008). Sequence analysis of bacterial artificial chromosome clones from the apospory-
412 specific genomic region of *Pennisetum* and *Cenchrus*. *Plant Physiology* 147 (3), 1396-1411.

413 Conner, J. A., and Ozias-Akins, P. (2017). Apomixis: engineering the ability to harness hybrid
414 vigor in crop plants. *Plant Germline Development* 17-34.

415 Dhaliwal, A., Dhaliwal, R. S., Kaur, N., and Gupta, R.C. (2018). Cytomorphological study in
416 genus *Cenchrus* L.: an important medicinal plant from north India (Family: Poaceae). *Cytologia*
417 83(1), 45-52.

418 Douglas, C., and DiFazio, S. (2010). Genetics and genomics of *Populus*. In: S. Jansson, R.
419 Bhalerao, and A. Groover (eds.) The *Populus* genome and comparative genomics. Berlin:
420 Springer-Verlag, 67-90.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

421 Dresselhaus, T., Carman, J. G., and Savidan, Y. (2001). Genetic engineering of apomixis in
422 sexual crops: A critical assessment of the apomixis technology. In: Y. Savidan, J. G. Carman,
423 and T. Dresselhaus (eds.) Flowering of Apomixis: From Mechanisms to Genetic Engineering,
424 (Mexico: CIMMYT, IRD, European Commission DG VI), pp. 229–243.

425 Dwivedi, K. K, Bhat, S. R., Bhat, V., Bhat, B. V., and Gupta, M. G. (2007). Identification of a
426 SCAR marker linked to apomixis in buffelgrass (*Cenchrus ciliaris* L.). *Plant Science* 172(4),
427 788-795.

428 Fisher, W. D., Bashaw, E. C., and Holt, E. C. (1954). Evidence for apomixis in *Pennisetum*
429 *ciliare* and *Cenchrus setigerus*. *Agronomy Journal* 46, 401–404.

430 Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R., and Kumar, A. (1992b). Ty1-
431 copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids*
432 *Res.* 20, 3639-3644.

433 Flavell, A. J., Smith, D. B., and Kumar, A. (1992a). Extreme heterogeneity of Ty1-copia group
434 retrotransposons in plants. *Mol Gene Genet* 231, 233-242.

435 Friesen, N., Brandes, A., and Heslop-Harrison, J. S. (2001). Diversity, origin and distribution of
436 retrotransposons (gypsy and copia) in conifers. *Molecular Biology and Evolution* 18, 1176-1188.

437 Gallo, P. H., Micheletti, P. L., Boldrini, K. R., Risso-Pascotto, C., Pagliarini, M. S., and do
438 Valle, C. B. (2007). 2*n* gamete formation in the genus *Brachiaria* (Poaceae: Paniceae).
439 *Euphytica* 154, 255-260.

440 He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen,
441 F., Qi, Y., Chen, R., and Deng, X.-W. (2010). Global epigenetic and transcriptional trends
442 among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22, 17-33.

443 Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G.,
444 Kamm, A., Doudrick, R. L., Schwarzacher, T., Katsiotis, A., and Kubis, S. (1997) The
445 chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and
446 their implications for genome evolution. *Genetica* 100(1), 197-204.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

447 Heslop-Harrison, J. S., and Schwarzacher, T. (2011). Organisation of the plant genome in
448 chromosomes. *The Plant Journal* 66, 18-33.

449 Higgins, J., Tomaszewska, P., Pellny, T. K., Castiblanco, V., Arango, J., Tohme, J.,
450 Schwarzacher, T., Mitchell, R. A., Heslop-Harrison, J. S., de Vega, J. J. (2021). The five
451 *Urochloa* spp. Used in development of tropical forage cultivars originate from defined
452 subpopulations with differentiated gene pools. https://doi.org/10.1101/2021.07.21.453213

453 Hirochika, H., and Hirochika, R. (1993). Tyl-copia group retrotransposons as ubiquitous
454 components of plant genomes. *Jpn J Genet* 68: 35-46.

455 Jackson, J. (2005). Is there a relationship between herbaceous species richness and buffel grass
456 (*Cenchrus ciliaris*)? *Austral Ecology* 30, 505–517. Doi.org/10.1111/j.1442-9993.2005.01465.x

457 Jessup, R. W., Burson, B. L., Burow, G., Wang, Y.-W., Chang, C., Li, Z., Paterson, A. H., and
458 Hussey, M. A. (2003). Segmental allotetraploidy and allelic interactions in buffelgrass
459 (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping.
460 *Genome* 46, 304–313.

461 Jin, W., Melo, J. R., Nagaki, K., Talbert, P. B., Henikoff, S., Dawe, R. K., *et al.* (2004). Maize
462 centromeres: organization and functional adaptation in the genetic background of oat. *Plant Cell*
463 16, 571–81.

464 John, N. T., Nuismer, S. L., and Merg, K. (2004). Plant polyploidy and the evolutionary ecology
465 of plant/animal interactions. *Biological Journal of the Linnean Society* 82, 511-519.

466 Kharrat-Souissi, A., Siljak-Yakovlev, S., Brown, S. C., Baumel, A., Torre, F., and Chaieb, M.,
467 (2014). The polyploid nature of *Cenchrus ciliaris* L. (*Poaceae*) has been overlooked: new
468 insights for the conservation and invasion biology of this species – a review. *The Rangeland*
469 *Journal* 36, 11–23.

470 Kharrat-Souissi, A., Siljak-Yakovlev, S., Brown, S., and Chaieb, M. (2013). Cytogeography of
471 *Cenchrus ciliaris* (Poaceae) in Tunisia. *Folia Geobotanica* 48, 95–113. Doi:10.1007/s12224-
472 012-9137-x

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

473 Kumar, A., Pearce, S. R., McLean, K., Harrison, G., Heslop-Harrison, J. S., *et al.* (1997). The
474 Ty1-*copia* group of retrotransposons in plants: genomic organization, evolution, and use as
475 molecular markers. *Genetica* 100, 205–217.

476 Lauss, K., Wardenaar, R., Oka, R., van Hulten, M. H., Guryev, V., Keurentjes, J. J., Stam, M.,
477 and Johannes, F. (2018). Parental DNA methylation states are associated with heterosis in
478 epigenetic hybrids. *Plant Physiology* 176(2), 1627-45.

479 Le, T. N., Miyazaki, Y., Takuno, S., and Saze, H. (2015). Epigenetic regulation of intragenic
480 transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acid Res.*
481 43(8), 3911-3921.

482 Li, R., Hu, F., Li, B. *et al.* (2020). Whole genome bisulfite sequencing methylome analysis of
483 mulberry (*Morus alba*) reveals epigenome modifications in response to drought stress. *Sci*
484 *Rep* 10: 8013 https://doi.org/10.1038/s41598-020-64975-5.

485 Li, N., Xu, C., Zhang, A., Lv, R., Meng, X., Lin, X., Gong, L., Wendel, J. F., and Liu, B. (2019).
486 DNA methylation repatterning accompanying hybridization, whole genome doubling and
487 homoeolog exchange in nascent segmental rice allotetraploids. *New Phytologist* 223, 979-992.

488 Lippman, Z. B., and Zamir, D. (2007). Heterosis: revisiting the magic. *TRENDS in Genetics*
489 23(2), 60-66.

490 Liu, Q., Li, X., Zhou, X., Li, M., Zhang, F., Schwarzacher, T., and Heslop-Harrison, J. S.
491 (2019). The repetitive DNA landscape in *Avena* (Poaceae): chromosome and genome evolution
492 defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biology* 19 (1):
493 226. Doi.org/10.1186/s12870-019-1769-z

494 Liu, Q., Yuan, H., Li, M., Wang, Z., Cui, D., Ye, Y., Sun, Z., Tan, X., Schwarzacher, T., and
495 Heslop-Harrison, J. S. (2022). Chromosome-scale genome assembly of the diploid oat *Avena*
496 *longiglumis* reveals the landscape of repetitive sequences, genes and chromosome evolution in
497 grasses. https://doi.org/10.1101/2022.02.09.479819

498 Luo, M. C., Dubcovsky, J., and Dvořák, J. (1996) Recognition of homeology by the wheat Ph1
499 locus. *Genetics* 144(3), 1195-203.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Macas, J., Kejnovský, E., Neumann, P., Novák P., Koblížková A., and Vyskot B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioceous plant *Silene latifolia*. *PloS One* 6, e27335.

Marshall, V. M., Lewis, M. M., and Ostendorf, B. (2012). Buffel grass (*Cenchrus ciliaris*) as an invader and threat to biodiversity in arid environments: A review. *Journal of Arid Environments* 78, 1-12.

Meena, S., and Nagar, R. P. (2019). Seasonal effects on fodder yield and quality in *Cenchrus* species under semi-arid climate. *Indian Journal of Agricultural Sciences* 89(8), 1287-1292.

Miller, G., Friedel, M., Adam, P., and Chewings, V. (2010). Ecological impacts of Buffel grass (*Cenchrus ciliaris* L.) invasion in central Australia: does field evidence support a fire-invasion feedback? *The Rangeland Journal* 32, 353–365. Doi:10.1071/RJ09076

Moore, L. D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Nevill, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., Van Leeuwen, S., Boykin, L. M., and Small, I. (2020). Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* 16(1), 1-8.

Novak, P., Neumann, P., Pech, J., Steinhais, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. *Bioinformatics* 29, 792–793.

Ozias-Akins, P., and van Dijk, P. J. (2007). Mendelian genetics of apomixis in plants. *Annu. Rev. Genet.* 41, 509–37.

Rathore, P., Raina, S. N., Kumar, S., and Bhat, V. (2020). Retro-element Gypsy-163 is differentially methylated in reproductive tissues of apomictic and sexual plants of *Cenchrus ciliaris*. *Frontiers in Genetics* 11:795.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

525 Roche, D., Cong, P., Chen, Z., Hanna, W. W., Gustine, D. L., Sherwood, R. T., and Ozias-Akins,
526 P. (1999). An apospory-specific genomic region is conserved between Buffelgrass (*Cenchrus*
527 *ciliaris* L.) and *Pennisetum squamulatum* Fresen. *Plant J* 19(2), 203-208.

528 Rozen S, and Skaletsky H. (1999) Primer3 on the WWW for general users and for biologist
529 programmers. *Bioinformatics Methods and Protocols* 132, 365–386.

530 Santos, F. C., Guyot, R., do Valle, C. B., Chiari, L., Techio, V. H., Heslop-Harrison, P., and
531 Vanzela, A. L. L. (2015). Chromosomal distribution and evolution of abundant retrotransposons
532 in plants: *gypsy* elements in diploid and polyploid *Brachiaria* forage grasses. *Chromosome*
533 *Research* 23, 571–582.

534 Sayers, E. W., Agarwala, R., Bolton, E. E., et al. (2019). Database resources of the National
535 Center for Biotechnology Information. *Nucleic Acids Research* 47: D23–D28.

536 Schwarzacher, T., and Heslop-Harrison, J. S. (2000). Practical *in situ* hybridization. *Oxford:*
537 *Bios.* 203+xii pp.

538 Sepsi, A., Higgins, J. D., Heslop-Harrison, J. S., and Schwarzacher, T. (2017) CENH3
539 morphogenesis reveals dynamic centromere associations during synaptonemal complex
540 formation and the progression through male meiosis in hexaploid23exaploidy wheat. The Plant
541 Journal 89, 235-249.

542 Sepsi A, Shiura, H., Okamoto, A., Sasaki, H., and Abe, K. (2014). Whole-mount MeFISH: a
543 novel technique for simultaneous visualization of specific DNA methylation and protein/RNA
544 expression. *PloS one 9*(4): e95750. https://doi.org/10.1371/journal.pone.0095750

545 Siljak-Yakovlev, S., Cerbah, M., Coulaud, J., Stoian, V., Brown, S. C., Jelenic. S., and Papes. D.
546 (2002). Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika*
547 and *Picea abies*. *Theoretical and Applied Genetics* 104, 505–512.

548 Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation
549 of the genome. Nat. Rev. Genet. 8, 272–285.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Spillane, C., Curtis, M. D., and Grossniklaus, U. (2004). Apomixis technology development—virgin births in farmers' fields? *Nature Biotechnology* 22, 687–691.

Thielen, P. M., Pendleton, A. L., Player, R. A., Bowden, K. V., Lawton, T. J., and Wisecaver, J. H. (2020). Reference genome for the highly transformable *Setaria viridis* ME034V. *G3 Genes Genomes Genetics*, 10(10), 3467-3478.

Thompson, J. N., and Pellmyr, O. (1992). Mutualism with pollinating seed parasites amid co-pollinators: constraints on specialization. *Ecology* 73, 1780-91.

Tomaszewska, P., Pellny, T. K., Hernández, L. M., Mitchell, R. A. C., Castiblanco, V., de Vega, J. J., Schwarzacher, T., and Heslop-Harrison, J. S. (2021a). Flow cytometry-based determination of ploidy from dried leaf specimens in genomically complex collections of the tropical forage grass *Urochloa* s. l. *Genes* 12: 957.

Tomaszewska, P., Vorontsova, M. S., Renvoize, S. A., Ficinski, S. Z., Tohme, J., Schwarzacher, T., Castiblanco, V., de Vega, J. J., Mitchell, R. A. C., Heslop-Harrison, J. S. (2021b). Complex polyploid and hybrid species in an apomictic and sexual tropical forage grass group: genomic composition and evolution in *Urochloa* (*Brachiaria*) species. *Annals of Botany* mcab147, https://doi.org/10.1093/aob/mcab147.

Verhoeven, K. J. F., Dijk, J. D., Biere, A. (2010). Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages. *Molecular Ecology* 19, 315-324.

Vij, S. P., and Chaudhary, J. D. (1981). Cylological investigation in three species of *Cenchrus* L. (Gramineae). *Cytologia* 46, 661-669.

Vining, K. J., Pomraning, K. R, Wilhelm L. J., Priest, H. P., Pellegrini, M., Mockler, T. C., Freitag, M., and Strauss S. H. (2012). Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* 13: 27 http://www.biomedcentral.com/1471-2164/13/27.

Visser, N. C., Spies, J. J., and Venter, H. J. T. (1998). Aneuploidy in *Cenchrus cilliaris* (Poaceae, Panicoideae, Paniceae): Truth or fiction? *S. Afc. J. Bot*. 64(6), 337-345.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

576 Worthington, M., Ebina, M., Yamanaka, N., Heffelfinger, C., Quintero, C., Zapata, Y. P., Perez,
577 J. G., Selvaraj, M., Ishitani, M., Duitama, J., de la Hoz, J. F., Rao, I., Dellaporta, S., Tohme, J.,
578 and Arango, J. (2019) Translocation of a parthenogenesis gene candidate to an alternate carrier
579 chromosome in apomictic *Brachiaria humidicola*. *BMC Genomics* 20: 41.

580 Worthington, M., Heffelfinger, C., Bernal, D., Quintero, C., Zapata, P., Perez, J. G., De Vega, J.,
581 Miles, J., Dellaporta, S., and Tohme, J. (2016) A parthenogenesis gene candidate and evidence
582 for segmental allopolyploidy in apomictic *Brachiaria decumbens*. *Genetics* 203(3), 1117-1132.

583 Yadav, B., Anuj, C., Kumar, S., Gupta, M. G., and Bhat V. (2012). Genetic linkage maps of the
584 chromosomal regions associated with apomictic and sexual modes of reproduction in *Cenchrus*
585 *ciliaris*. *Molecular Breeding* 30(1): 239-50.

586 Yan, Q., Wu, F., Xu, P., Sun, Z., Li, J., Gao, L., Lu, L., Chen, D., Muktar, M., Jones, C., and Yi,
587 X. (2021). The elephant grass (*Cenchrus purpureus*) genome provides insights into
588 anthocyanidin accumulation and fast growth. *Molecular ecology resources* 21(2): 526.

589 Zhang, M., Kimatu, J. N., Xu, K., and Liu, B. (2010). DNA cytosine methylation in plant
590 development. *Genetics and Genomics* 37, 1-12. Doi.org/10.1016/S1673-8527(09)60020-5

591 Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S. (2007). Genome-wide
592 analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between
593 methylation and transcription. *Nat Genet* 39,61-69.

594

595

596

597 Tables

598 **Table 1.** List of PCR primers and commercial oligonucleotide probes used in the study.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Primer name	Primer sequence	Oligonucleotide length [bp]	Product length [bp]	References
CopiaQMDVKT _F	CARATGGARGTNAARAC	17	290	Hirochika and Hirochika, 1993 Flavell et al., 1992a,b
CopiaYVDDML_R	ARCATRTCRTCNACRTA	17		
GypsyRT1RMCVDYR_	MRNATGTGYGTNGAYTAYMG	20	410	Friesen et al., 2001
GypsyRT4YAKLSKC_R	RCAYTTNSWNARYTTNGCR	19		
18S rDNA_F	CGAACTGTGAAACTGCGAATGGC	23	900	Chang et al., 2010
18S rDNA_R	TAGGAGCGACGGGCGGTGTG	20		
Universal conserved 5S	[Cy3] GTTAAGCGTGCTTGGGTGAGA GTAGTACTACGATGGGTGACC	42	-	This paper
CC_C105F76	CGAGTAGCCCAAATCCAGGA	20	590	This paper
CC_C105R665	TTGCTGAACCAACTCCCTGT	20		
CC_C105F76	CGAGTAGCCCAAATCCAGGA	20	919	This paper
CC_C105R994	CAAGGAAGGGGAAACAACGG	20		
CenchCL58_361F	GGCTGGGGGCGATAAGTAG	19	350	This paper
CenchCL58_710R	CAAGAAGGCGCAGATTGTGG	20		
CenchCL58_869	[Cy3]CTGGAGGTTGTAATTCCAAGGC CACGCATCTCAGAAGGGAGTCG	44	-	This paper
CenchStCL1ConsS	[Cy3]CCATTTTCACAACTTCGGTACC CTGATGTAGTGCATTCATGCAC	44	-	This paper
CenchEnCL1ConsRC	[FAM]GCCTACCCCATTAGATCCCAA ACGATGTTTGAGAGTGTTTAGGAG	45	-	This paper
CenchCL22Copia	[FAM]GGGGTTAGAAATAGGACCCA ACAAGAAAAACAATATGATGTAAC CGCGAGG	51	-	This paper

599

600 Legends to figures

601 **Figure 1.** Number of chromosomes and rDNA sites in *Cenchrus ciliaris*.

602 (A) Metaphase spread of *C. ciliaris* after DAPI staining (blue) showing a chromosome count of

603 $2n=4\times+3=39$.

604 (B) Metaphase chromosomes labeled with 5S rDNA fluorescing red showing three major signals

605 (arrows).

606 (C) Fluorescent *in situ* hybridization with 45S rDNA probe fluorescing green (arrows).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

(D) 45S rDNA probe (red) showing four major hybridization sites and Gy105 probe (green) hybridizing to parts of most chromosomes.

Scale bar = 10μm in (A)

Figure 2. Metaphase chromosomes of *C. ciliaris* (2*n*=4×+3=39) after FISH (red fluorescing, i) and immuno-staining with anti-5-methylcytosine (green fluorescing, ii) and as overlap image (iii).

(A) The 'universal' Ty1-*copia* probe shows concentrated signal around the centromeres of all chromosomes while 5-mC signal is seen overlapping them but also further along the chromosome arms. 5-mC strength is variable between chromosomes with most showing a small gap at the centromeres.

(B) The 'universal' Ty3-*gypsy* probe shows a dispersed signal over the chromosome arms, but is absent from the most distal parts. 5-mC signal covers most Ty3-gypsy signal but is weak or absent in some cases (arrows).

Scale bar = 10μm

Figure 3. Consensus sequences used to design primers and probes for fluorescence *in situ* hybridization.

A) RepeatExplorer graph showing number of raw sequencing reads forming abundant clusters; about 50% of the sequences form clusters, with less than 20 motifs representing a third of the genome.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

B) Graphical 2D projection of the structure of three graph-based clusters found in *Cenchrus ciliaris* genome using RepeatExplorer. Each node represents one of the sequence reads. The placement of the nodes reflects sequence similarity and overlaps. Cluster1 and Cluster58 are tandem repeats while cluster22 is a LTR-copia element.

C) Sequence LOGO representations showing lengths of repetitive motifs and conservation of sequences. Cluster1 is 140bp tandem repeat; Cluster58 consensus is 1326bp and Cluster22 4807bp.

Figure 4. Distribution of highly abundant repetitive DNA motifs on chromosomes of *C. ciliaris* (2*n*=4×+3=39).

(A) Tandem repeat CenchEnCL1ConsRC probe labeled commercially with FAM (green) showed strong signals in pericentromeric regions of about half of the chromosomes. The remaining chromosomes showed weak signals.

(B) Twenty very strong and 4 weaker red signals of Tandem Repeat CenchStCL1ConsS probe labeled commercially with Cy3 (red).

(C) Three pictures of the same metaphase showing collocalization of Cluster1 and Cluster22 signals: (i) CenchStCL1ConsS (labeled with Cy3; red) showing signal on about half of the chromosomes, (ii) CenchCL22Copia (labeled with FAM; green) with signal on all chromosomes, (iii) merged.

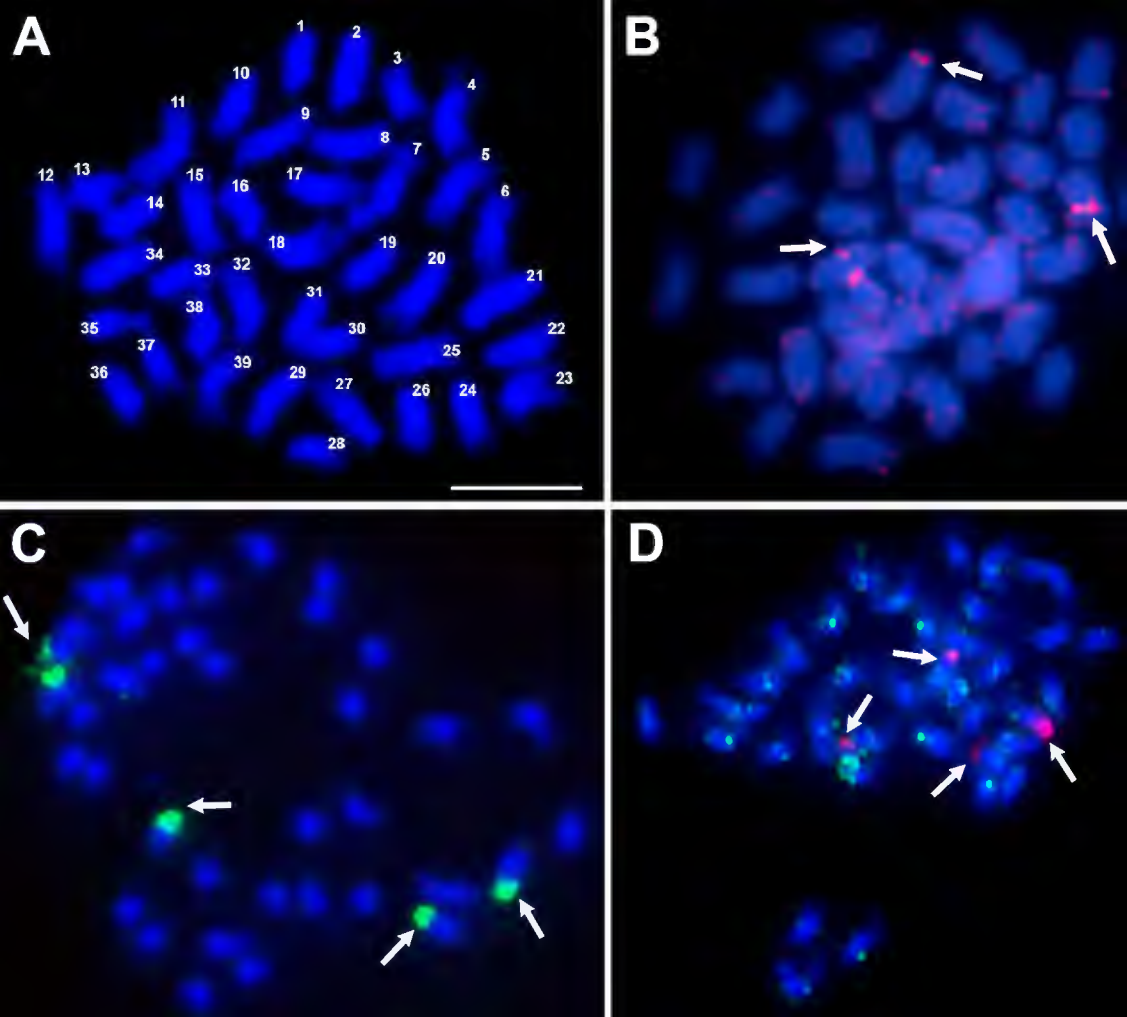
(D) Metaphase chromosomes showing dispersed signals of amplified CenchCL58 probe (detected with FITC; green) and commercially synthesized CenchCL58_869 (Cy3, red).

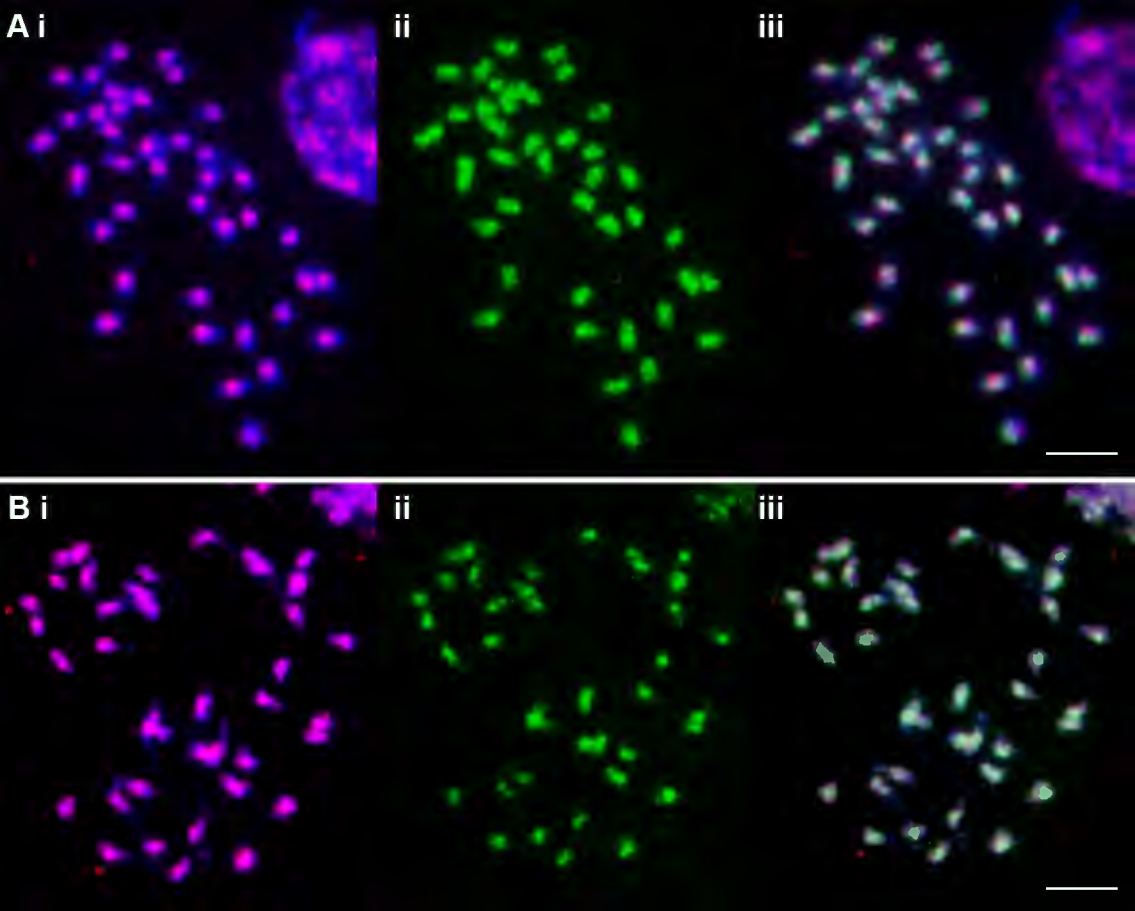
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

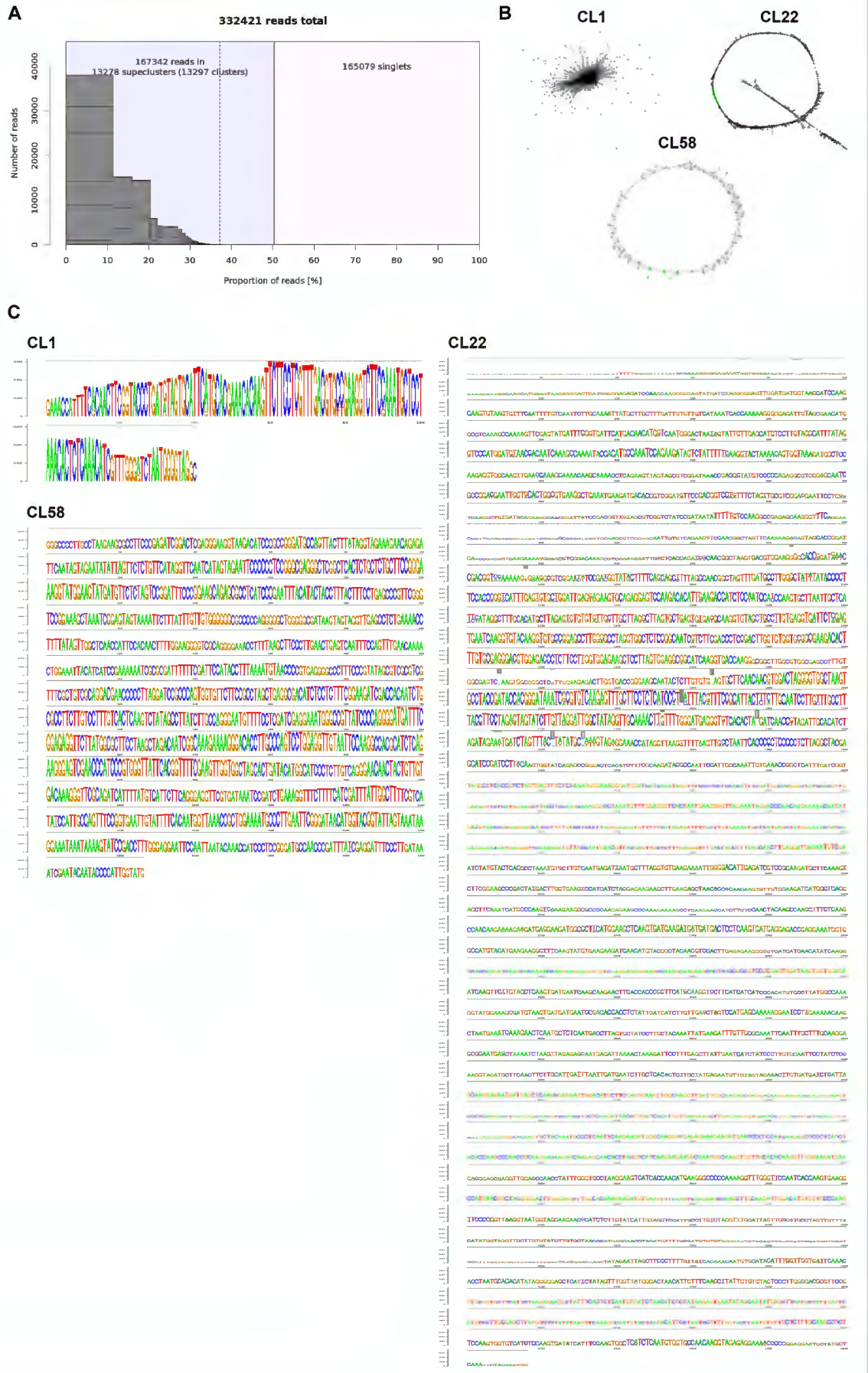
649 (E) Dispersed signals of CenchCL58_869 probe (Cy3, red) along chromosomes. Some stronger

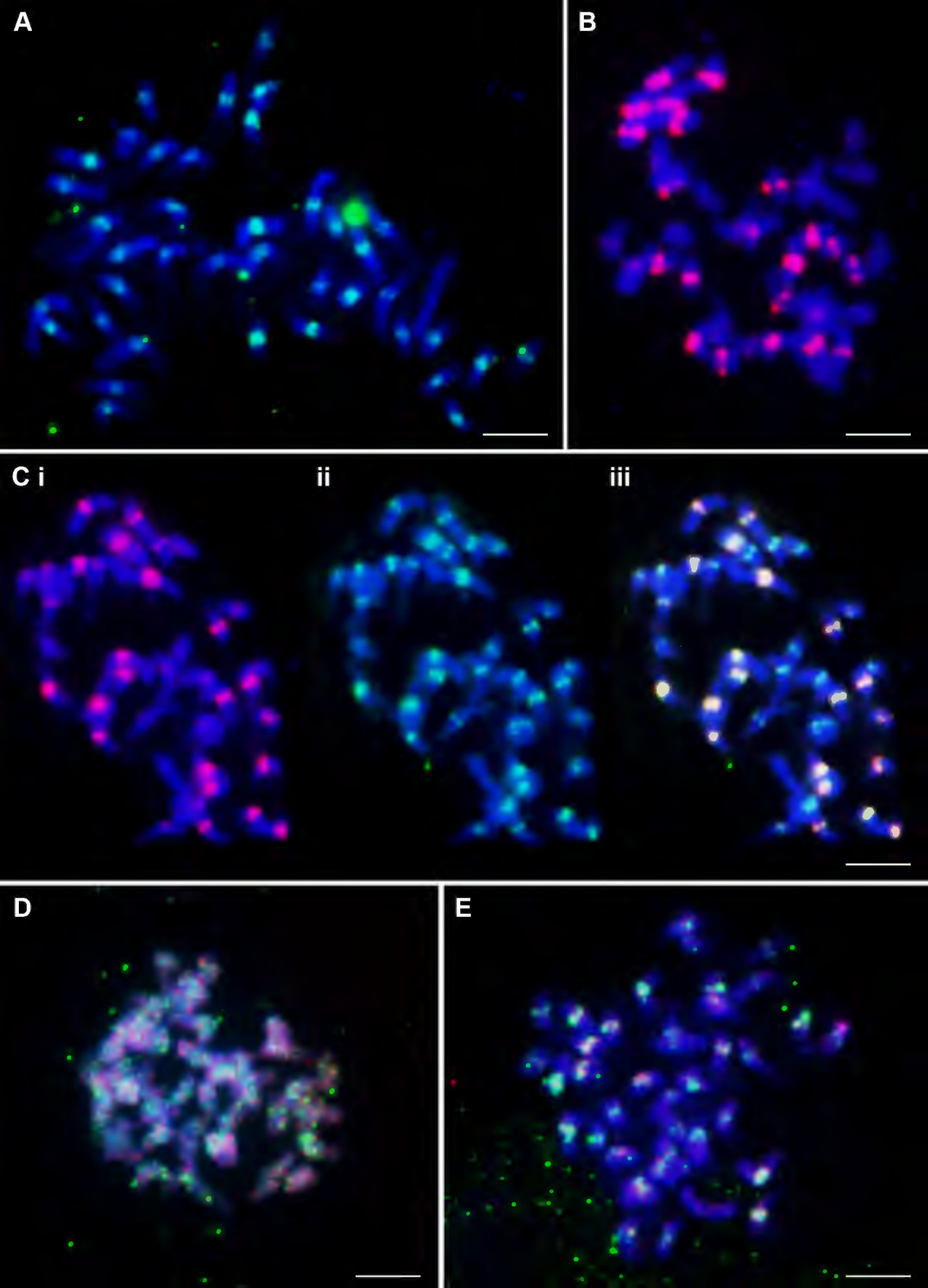
650 signals collocalize with CenchEnCL1ConsRC (FAM; green).

651 Scale bar = 10μm


A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
B
C
D


A i
ii
iii
B i
ii
iii


A
332421 reads total
167342 reads in
13278 supeclusters (13297 clusters)
165079 singlets
Number of reads
Proportion of reads [%]
B
CL1
CL22
CL58
C
CL1
CL58
CL22


A
B
C i
ii
iii
D
E

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Supplementary Data Table S1. List of the most abundant clusters extracted from the Chinese variety of *Cenchrus ciliaris* SRR8666664 using RepeatExplorer.

Name	Length [bp]	GC [%]	Sequence	Description
Cluster1	140	45.7	GAAACCATTTTCACAACTTCGGTACCCTGATGTAGTGCATTCATGCACGAAACAC AAGTTCTTCCTCGTTTGACGTGCCGAAGGTCTTCCAAATGCTCCTAAACACTCTCA AACATCGTTTGGGATCTAATGGGGTAGGC	Labelled oligonucleotide probes ordered are from 44bp (5 to 48; 46.7% GC) and RC of 45bp (96 to 140; 46.7% GC)
Cluster58	1326	47.4	GGGCCCCTTGCCTAAGAAGCGCCTTCCCGAGATCGGACTCGAGGGAAGGTAAGA CATCCCGCCGGGATGCCAGTTACTTTATAGGTAGAAAGAACAGAGATTCAATAG TAGAATATATTAGTTCTCTGTTCATAGGTTCAATCATAGTAGAATTCCCCCCTCCG GGCGAGGGCTCGGCTCACTCTGCTCTGCTTCCGGGAAAGGTATGGAAGTATGATG TTCTCTAGTCCGGATTTCCCGGAAGCAGAGCGCCTCATCCCGAATTTACATACTA CCTTTACTTTCCTGAGCCCGTTCGGGTCCGGAAAGCTAAATCGGAGTAGTAAAAT TCTTTATTTGTTGTGGGGGGCGCCCCCAGGGGGCTGGGGGCGATAAGTAGTAGGT TGAGCCTCTGAAAACCTTTTATAGTTGGCTCAACCATTCCACAACTTTTGGAAGG GGTCGCAGGGGAAACCTTTTAAGCTTCCCTTGAACTGAGTCAATTTCCAGTTTGA ACAAAACTGGAAATTACATCATCCGAAAAAATCGCGCGATTTTTTCGATTCCATA CCTTTAAAATGTAACCCCGTGAGGGGCCCTTTCCCGTATAGCGTCGCGTCGTTTC GCTGTCGCAGGACGAACCCCCTTAGGATCCGCCCAGTGGTGTTCTTCCGCCTAGC TGAGGCGACATCTCCTCTTTCGGAAGATCGACCACAATCTGCGCCTTCTTGTCCTT TGTCACTCAAGTCTATAGCCTTATCTTGCCAGGGAATGTTTTCCTCGATCGAGGA AATGGGCCGTTATCCCCAGGGGATGATTTCGGAGAGGTTCTTATGGCGCTTCCTA AGCTAGACAATCGGCAAAGAAAAGGGACACCTTGCCCAGTCCTGGAGGTTGTAA TTCCAAGGCCACGCATCTCAGAAGGGAGTCGAACCCATCCCGTGGGTTATTCACG GTTTTCGAAGTTGGTGGCTAGCACTGATACATGGCATCCCTCTTGTCAGGGAACA ACTACTGTTGTGACAAAGGGTTCGCAGATCATTTTATGTCATTCTTCAGGGAGGT TCGTGATAAATCCGATCTGAAAGGTTTCTTTTCATCGATTTATTGGCTTTTCGTCA TATCCATTGCCAGTTTCCGGTGAATTGTATTTTCACAATGGTTAAACCGCTGGAA AATGCCCTTGAATTCGGGATAACATGGTACGGTATTAGTAAATAAGGAAATAAA TAAAAGTATCCGACCTTTGGGAGGAATTCCAATTAATACAAACCATCCCTCCGGG ATGCCAACCCGATTTATCGAGGATTTCCCTTGATAAATCGAATACAATACCCCAT TGGTATG	Labelled oligonucleotide probe ordered is from 44bp (867 to 910; 54.5% GC)
Cluster22	4807	45.6	ACTTGCTAGATTCCAATAATATGAGTACCCTTTGGACACTTTTTGGAGATTTTATG CCAAAGGGGGAGAGATAGTGGAGAGATCGTGGAGATTGGAGTCAAAGCAAGGC AAGCATGGAGTAAGGGGGAGTGATCGGGGAGAGATCCAAGCAAGGGGGAGTAT GATCTAGGGGGAGTTGGATGATGGTAAGCATCCAAGCAAGTGTAAGTGTTTCAA TTTTGTCAATTCTTGCAAAATTATGCTTGCTTTGATTGTGTTGTCATAAATCACCA AAAAGGGGGAGATTGTAGCGAACATGGCCTCAAAGCCAAAAGTTCGAGTATGAT TTCGGTGATTCATGACAACATGGTCAATGGGACTAATAGTATTGTTGAGCATGTC CTTGTAGGCATTTATAGGTCCCATGGATGTAACCACAATCAAAGCCAAAATACCA CATGCCAAATCCAGAAGATAGTCTATTTTTCAAGGTACTAAAACAGTGGTAAAG ATGGCTCCAAGAGGTGGCAAGTTGAACCAAAGCAAAACAAGCAAAACCTCAGA AGTTAGTAGCGTCGGATAAACCGACGCTATGTCCCCCAGAGGCGTCGGAGCAAT CGCCGGACGAATTGGTGCACTGGCGTGAAGGCTGAAATGAAGATGACACCGTCG GATGTTCCGACGGTCCTGTTTCTAGTTGCGTCGGACGAATTCCTGATTCAAGGCT GTGATGCAGAAGGCATCGGATTATCCGACGCTTGTAGCGTCGGTCTATCCGATAA TATTTTTGTCCAAGGCCGAGAGCAAGGGTTTCAGGAACCCTTCAGCACCGGAAG ATCCGACGCCCCGTCGGTGCAAGCGTCGGAGCAATTCTGTCAGAAGGTCCAACG GCTAGTTCAAAAAGACAGTAGCACCGGATGAACCGACGGTGAAGAAAAGGCAG CGTCGGACAAAGCGTCGGACAGATTGCGTCAGCAGATCGCAACGGCTAGGTGAC GTGGAAGGGGCACCGGATAAACCGACGGTCCAAAAAGGGAGGCGTCGGAATAT CCGATGGTATACTTTTCAGCAGCTTTTAGCCAACGGCTAGTTTGATGCCTTGGGCT ATATATACCCCTTCCACCCGGTCATTTGAGTGTGCTGGATTGAGGAGAAGTGCAG AGGAGTCCAAGACACATTGAAGACCATCTCCAATCCATCCAAGTGCTTAATTGCT CATTCATAGGCTTTGCACATGCTTAGAGTGTGTGTGTTGCTTGCTTAGGCTTAGTG CTGAGTGTGAGGCAAGGTGTAGCTGCCTTGTGAGGTGATTCTCGAGTGAATCAAG GTGTACAAGGTGTGCCGGAGCCTTGGGGCCTAGGTGGCTCTCCGGCAATCGTCTT CGACCCTCCGACTTGGTGTGGTGCGGCGAAGACACTTTGTGCGAGGGACGTGGA GACCCTCTTCCTTGGTGGAGAAGCTCCTTAGTGGAGGCGGCATCAAGGTGACCAA GGGCGGCTTGGCGTGGCGAGCCTTTGTGGCGAGTCAAGTGGCGGGCTCGTTGGG	Labelled oligonucleotide probe designed to RNaseH domain (RC) is from 51bp (2156 to 2206; 43.1% GC)

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

			ACTAGGGGTGGCTAAGTGCCTACCGATACCACGGGATAAATCGCGTGTCAAGAG TTTGCTTCCTCTCATCCCTATTTACGTTTCCGCATTACTTTCTTGCAATCCTTGTTT GCCTTTACCTTCCTAGAGTAGTATCTTGTTAGGATTGGCTATAGGTTGCAAAACTT GTTTTGGGATGAGGGTGTCACACTAGATCAACCGTAGATTGCACATCTAGATAGA AATGATCTAGTTTACTATATGCAAAGTAGAGGAAACCATAGGTTAAGTTTTTAAG TTGCCTAATTCACCCCCTCCCCCTCTTAGGCTACGAGCATCCGATCCTTACAAGTG GTATCAGAGCCGGGACTCACATCTTTCGCAAGATACGCCAATTCCATTGGCAAAT TGTGAAACCGGCTCATTTGATCGGTTAGGCTTCACCGCCTAGTGAGTTGCTCTAA AGGGGAAAGGGATGGATCATCAAATAGGTTTCCCTCCTCCACATTTCGATGGCAC GGGCTATCCACGTTGGAAAGTTCTCATGGAAGCTTACCTCCAAGCAAAGGGCCTA AATGTTTGGAGGGTCACTAATGAAGGGGTTAGAAATAGGACCCAACAAGAAAAA CAATATGATGTAACCGCGAGGAGCATCCTTTTGTCCTCTCTTAGTGAAGATGTTTT TCATAGAGTTTTTACTTGTGAAAATGCTCATGAGCTATGGAAAACTATCAAAGAG AATAATGAGGGTTCAAAAGATGTTTCCAATGAACGTTACGAGTTGCTCTTTGAGG GTTTTACTAGCTTTAAGCAACTTGAGGATGAAAATGTCGAATCTATGTACTCACG CCTAAATGTGCTTGTCAATGAGATTAATGCTTTAGGTGTGAAGAAAATTGGGGAC ATTGAGATCGTCCGCAAGATGCTTCAAAGTCTTCGGAAGCCCGACTATGACTTGG TGAAGTCCATCATCTACGAGAAGAAGCTTGAAGAGCTAACACCACAACAAGTGT TGTGCAAGATCATGGCTCACGAGCTTCGAATCATGCCCAAGTCAAAGAAGGCGC CCCAAGAGAAGCCCAAAGAAAAGCCTCAAGAACCATCTTCTCCAACTACAAGCC AAGCTTTCTCAAGCCAACAAGAAAAGAAGATGAGGAAGATGGCGCTTCATGGAA GCTCAAGTGATGAAGATGATGATGACTCCTCAAGTGATGAGGAGACCGAGGAAA TGGTGGCCATGTACATGAAGAAGGCTTTCAAGTATGTGAAGAAGATCAACATGT ACGGCTACAACGTCCACTTGAGAGAAGGGCGTCATCATCAACATATCAAGGTCA CCAAGATCAAGCACAAAACCAAGAAGAAGGTGGTCAAGGAGAGGAAGCCAAAG AAAGAAGCACTCACCGCCGTCCGCGAGTGGATAAGTGGTGGAGAATCAAGTTCT TGTAGCTCAAGTGATGAATCAAGCAAGAACTTCACCACCCGCTTCATGCAAGGTC CTTCATCATCATCCCACATGTGCCTTATGGCCAAAGGTATGGAAAGCGATGTAAG TGATGATGAATCCGACACCACCTCTATTGATGATCTTGTTGAGCTAGTCCATGAG CAAAAAGGAATCCTTGAAAAACAAGCTAATGAAATCAAAGAACTCAATGCTCTC AATGACCTTAGTGCTATCCTTGCTACAAATTATGAAGATTTGTTGCGCAAATTCA ATTTGCTTTGCAAGGAGCGCAATGAGCTAAAATCTAAGTTAGAGAGCAATGAGA TTAAAACTAAAGATTCCTTTGAGCTTATTGAATCATCTATCCCTTGTGCAATTCCT ATCTCCAAGGTAGATGCTTCAACTTCTTGCATTGATTTAATTGATGAATCTTGCTC ACACTCTTGCTATGAGAATGTTGTAGTAGAAACTTGTGATGATCTCATTAGCAAG GAGAATGATGAGCTCAAGCGAGAAGTGGACATGCTTCGAGTGAAGTTGGCAAGG TTGATGGGCAAGAGCGAGGACAAGGAGGAACAAGAGGAAGTCCCTCAAGACAT GAAGAGCAACGTCCAACCTTCTCAAGATAACCGTGATCACATGGTGAAGAAGCT TGAGAAGGGTGAAACCGTGACTTGCTACAAGTGTCATGAAGAAGGGCACAAGTT CTACAAATGCCCTCAATTCAAGAAGATGGGCAAGGGCGAGAGAAAGAAGATGA ACCCCTCCATCAAGAGCTCCCTCATCTACACCAAGCCCAACCGCAAGAACAAGA GCAAGAGCAACACCTATGTCATCAAGAAGAAGGCAAATGGCAAGGTGGTTGCAC ACAAGGTTGGGAAGATGAAGACGGAGCGAGGTTGGAGCCAACCTATTTGGGTGC CTAAGGAAGTCATCACCAACATGAAGGGCCCCCAAAAGGTTTGGGTTCCAATCA CCAAGTGAAGGCCATGAACGGCTACGGGGACTTGGGAGGCTTGGCACAAAGAAG ATGTGAAGTTTCAAGTCAAGAAGCCAAGCTTGCAAGATTGGACATGTGTTGCCCA AGTTCCCCGGTTAAGGTAATGGTAGCAAGAACCCATCTCTTGTATCATTGTAGCT CCATTGCCTTGTCTAGGTTTGCATTAGTTGTCATGCCTAGTTGTTTACATATGGTA GGTTGCTTGTGTATCTTGTGCTAACCCATGAGCAACCTACATGTTTTGCCATGTGT GTAGGAAATCTACACTTTGTTCTTTCATGGTGCATGCATTTCTTATGTTATCCTTT GCATCATGAGCAAGCTATAGAATTAGCTTCCCTTTTGTTATCACAAACAATGTGC ATACATTTGGTTGGTGATTCAAACACCTAATGCACACATATAGGGGGAGCTCATT CTATAGTTTGGTTATGGGACTAACATTGTTTCAAGCTTATTGTGTGTAGTCCCTTG GCGACGCGTTCCGGTGCGGTGCTTCATTCTAAGAGAATGCTATTTCACTCTCAAT GTGATCTAACATCTCTATAAGAGTGAATATAGCAATATGAGCTTCATGCTATTGA GTCATGTCCTTGCGAGCTTATGTCTCTATATAAGTTCAAAGGCAATGTACTCATA CATTCCTAATCCTTTTGGTACCTAATGTGTTTCTCTTTGCAAGCCTCTTCCAAGTG GTGTCATCTCCAAGTGATATCATTTCCAAGTGGCTCATCTCAATGTGGTGCCAAC AAGGTAGAGAGGAAAACCCCCGGAGGAATGCTATGCTCAAATTCTACAAGTGG	

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.25.493456; this version posted May 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.