CAIO TÚLIO RODRIGUES CORRÊA

CYTOGENOMIC CHARACTERIZATION OF THE REPETITIVE DNA IN THE ‘BRIZANTHA’ AGAMIC COMPLEX OF *UROCHLOA* P. BEAUV. (POACEAE)

LAVRAS – MG
2023

CAIO TÚLIO RODRIGUES CORRÊA

CYTOGENOMIC CHARACTERIZATION OF THE REPETITIVE DNA IN THE 'BRIZANTHA' AGAMIC COMPLEX OF *UROCHLOA* P. BEAUV. (POACEAE)

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Botânica Aplicada, área de concentração em Botânica Aplicada, para a obtenção do título de Doutor.

Profa. Dra. Vânia Helena Techio

Orientadora

LAVRAS – MG
2023

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Corrêa, Caio Tulio Rodrigues.

Cytogenomic characterization of the repetitive DNA in the 'brizantha' agamic complex of *Urochloa* P. Beauv. (Poaceae) / Caio Tulio Rodrigues Corrêa. - 2023.

95 p. : il.

Orientador(a): Vânia Helena Techio.

Tese (doutorado) - Universidade Federal de Lavras, 2023.

Bibliografia.

1. Citogenética. 2. DNA Repetitivo. 3. Forrageiras. I. Techio, Vânia Helena. II. Título.

CAIO TÚLIO RODRIGUES CORRÊA

CYTOGENOMIC CHARACTERIZATION OF THE REPETITIVE DNA IN THE 'BRIZANTHA' AGAMIC COMPLEX OF *UROCHLOA* P. BEAUV. (POACEAE)

CARACTERIZAÇÃO CITOGENÔMICA DO DNA REPETITIVO NO COMPLEXO AGÂMICO 'BRIZANTHA' DE *UROCHLOA* P. BEAUV. (POACEAE)

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Botânica Aplicada, área de concentração em Botânica Aplicada para a obtenção do título de Doutor.

APROVADA em 03 de março de 2023

Dra. Ludmila Cristina de Oliveira	Institute of Plant Molecular Biology, República Tcheca
Dra. Mariana Alejandra Báez	University of Bonn, Alemanha
Dra. Magdalena Vaio Scvortzoff	Universidad de la Republica, Uruguai
Dra. Giovana Augusta Torres	ICN/UFLA

Profa. Dra. Vânia Helena Techio

Orientadora

LAVRAS – MG
2023

Aos pesquisadores brasileiros, que seguem fazendo ciência de qualidade em meio a tantas dificuldades,

DEDICO

AGRADECIMENTOS

À Universidade Federal de Lavras e ao Instituto de Ciências Naturais, pela educação e estrutura concedida para a realização desse trabalho.

À FAPEMIG, CAPES pelo financiamento do projeto e ao CNPq pela concessão das bolsas de estudo. Às Embrapa Gado de Corte e Gado de Leite, pela parceria e disponibilização das plantas. À Fundação Fiocruz pela estrutura e serviço de sequenciamento.

A todos os professores e professoras do programa de Botânica Aplicada, por contribuírem com a minha formação. Em especial a professora Dra. Marinês Pires, pela coordenação, atenção e disponibilidade.

À professora Dra. Vânia Helena Techio, agradeço a orientação, apoio, confiança e dedicação. Obrigado pelos ensinamentos e pelo esforço de me orientar mesmo com todas as dificuldades impostas pela pandemia.

À professora Dra. Giovana Augusta Torres, agradeço por me iniciar no mundo científico e continuar contribuindo com meu crescimento profissional.

À professora Dra. Magdalena Vaio, agradeço todo o auxílio e disponibilidade. Obrigado por ter me apresentado as ferramentas bioinformáticas sem as quais meu trabalho não seria possível.

A todos os colegas do laboratório de Citogenética Vegetal pela convivência e parceria. Obrigado Yasmim e Isabella pelos vários favores. Obrigado Marco Tulio pelas dicas de protocolo e pelas sondas compartilhadas. Obrigado Bruna pela ajuda nos inúmeros testes de PCR. Obrigado Lucas, Raquel e Ana Gabriela pela amizade dentro e fora do laboratório, com quem sempre posso contar para uma boa prosa e cerveja gelada.

Aos meus amigos e amigas de Lavras, em especial Luiza, Patrícia e Lauren, pelos 11 anos de amizade e convivência, participando de todas as minhas realizações e dificuldades.

À minha família. Em especial à minha mãe, Maria Emília, por todo o amor e incentivo. Obrigado por sempre fazer de tudo para que eu realizasse meus sonhos. Agradeço à minha irmã Karen e ao Átila, pelo amor, companheirismo, e todos os momentos partilhados.

Muito obrigado a todas e todos!

“Prefiro ter perguntas que não podem ser respondidas
do que respostas que não podem ser questionadas”
– Richard Feynman

RESUMO

Urochloa P. Beauv. (syn. *Brachiaria* (Trin.) Griseb.) é um gênero africano de gramíneas perenes extensivamente cultivado no Brasil como forrageira. *Urochloa brizantha*, *U. decumbens* e *U. ruziziensis* estão dentre as espécies de maior relevância econômica. Essas três espécies formam o complexo agâmico 'brizantha', que inclui séries poliploides e hibridações interespecíficas resultando em relações genômicas complexas. Grande parte dos genomas de gramíneas é composto de DNA repetitivo, portanto investigar sua composição e organização nesse gênero permite um maior entendimento da história evolutiva, além de subsidiar programas de melhoramentos na seleção de genótipos para cruzamentos. O presente estudo objetivou caracterizar a fração repetitiva dos genomas de citótipos diploides e tetraploides de *U. brizantha*, *U. decumbens* e *U. ruziziensis*, identificando as principais sequências e entendendo sua organização através da construção de sondas repetitivas para mapeamento físico dos cromossomos por meio da hibridização *in situ* fluorescente (FISH). Os resultados revelaram que o DNA repetitivo compreende 56-65% dos genomas de *Urochloa*, sendo retrotransposons LTR TY3/*Gypsy* os elementos mais frequentes. As proporções das principais linhagens de retrotransposons e transposons de DNA variaram entre as espécies, assim como o número de famílias de DNA satélite (satDNA), que apresentou variações intra e interespecíficas. *Urochloa brizantha* exibiu um perfil mais diverso de satDNA, com um maior número de monômeros únicos e repetições espécie-específicas. Os três satélites mais abundantes (UroSat-1a, UroSat-2a e UroSat-3) foram selecionados para mapeamento nos cromossomos dessas espécies, via FISH. Sinais conspícuos de UroSat-1a foram identificados em todos os centrômeros das espécies avaliadas, enquanto a distribuição de marcas de UroSat-2a e UroSat-3 variaram em número e posição entre elas. Os resultados encontrados validam o uso de DNA satélite como marcadores citogenéticos no complexo agâmico 'brizantha' em *Urochloa* e evidenciam relações genômicas entre diferentes espécies e ploidias.

Palavras chave: Poliploidia. Brachiaria. Gramíneas forrageiras. Hibridização in situ.

ABSTRACT

Urochloa P. Beauv. (formerly classifed as *Brachiaria (Trin.) Griseb.*) is an African genus of perennial grasses that is extensively cultivated in Brazil for cattle nutrition. *Urochloa brizantha*, *U. decumbens* and *U. ruziziensis* are among the most economically relevant species. These three species form the 'brizantha' agamic complex, which includes polyploid series and interspecific hybridizations resulting in complex genomic relationships. Repetitive DNA is a large component of grass genomes, therefore investigating its composition and organization in the genus would provide more insight on its evolutionary history and would allow more informed decisions to be made in breeding programs when selecting genotypes for crosses. The present study aimed to characterize the repetitive fraction of the genomes in diploid and tetraploid cytotypes of *U. brizantha*, *U. decumbens* and *U. ruziziensis*, identifying major sequences and understanding their organization by physically mapping repetitive probes onto chromosomes through fluorescent *in situ* hybridization (FISH). Results revealed that repetitive DNA comprises 56-65% of *Urochloa* genomes and that LTR TY3/*Gypsy* retrotransposons are most frequent elements. The proportions of major retrotransposons and DNA transposons lineages varied among all species, as well as the number of satellite DNA (satDNA) families, which presented intra and interspecific variation. *Urochloa brizantha* presented a more diverse satDNA profile, with a larger number of unique monomers, and species-specific repeats. The three most abundant satellites (UroSat-1a, UroSat-2a, and UroSat-3) were selected to be mapped to chromosomes via FISH. Conspicuous signals of UroSat-1a were present in all centromeres of the evaluated species, while the distribution of UroSat-2a and UroSat-3 marks varied in number and position. Our findings validate the use of satellite DNA as cytogenetic markers in *Urochloa* 'brizantha' agamic complex and highlight genomic relationships among different species and ploidy levels.

Keywords: Polyploidy. Brachiaria. Forage grasses. In situ hybridization.

SUMÁRIO

PRIMEIRA PARTE

1 ANTECEDENTES E JUSTIFICATIVA

O gênero *Urochloa* P. Beauv. (sin. *Brachiaria* (Trin.) Griseb.) é amplamente cultivado como forrageira no Brasil e apresenta grande importância econômica na criação de bovinos para corte e produção de leite. Devido a sua ampla distribuição e boa adaptação ao clima e solo brasileiros, as espécies de *Urochloa* têm sido empregadas em programas de melhoramento que visam a seleção de genótipos que reúnam características desejáveis para a realização de cruzamentos inter e intraespecíficos (JANK et al., 2014).

Complexos agâmicos são sistemas formados por plantas poliploides que se reproduzem primariamente por sementes assexuadas (apomixia), mas que incluem espécies diploides que facilitam a reprodução sexual (COYNE e ORR, 2004). *Urochloa brizantha* (Hoschst. ex A. Rich) R.D. Webster, *U. decumbens* (Stapf) R.D. Webster e *U. ruziziensis* (R. Germ. & C.M. Evrard) Morrone & Zuloaga formam o complexo agâmico '*brizantha*' (LUTTS; NDIKUMANA; LOUANT, 1991; RENVOIZE; CLAYTON; KABUYE, 1996; RENVOIZE; MAASS, 1993), sendo as duas primeiras espécies predominantemente apomíticas e tetraploides (2n=4x=36) (NANI et al., 2016; MORAES et al., 2021), mas com representantes diploides (CORRÊA et al., 2020). *Urochloa ruziziensis*, por sua vez, é exclusivamente diploide (2n=2x=18) e de reprodução sexual (VALLE; SAVIDAN, 1996; NANI et al., 2016).

Estimativas de conteúdo de DNA para acessos diploides de *U. brizantha* (870 Mbp), *U. decumbens* (704 Mbp) e *U. ruzizensis* (732 Mbp) revelaram que seus genomas possuem tamanhos proporcionais aos tetraploides de *U. brizantha* (1720 Mbp) e *U. decumbens* (1855 Mbp). Estudos meióticos evidenciariam a ocorrência de alotetraploidia nas cultivares Marandu (*U. brizantha*) e Basilisk (*U. decumbens*) (MENDES et al., 2006; MENDES-BONATO et al., 2001, 2002a, 2002b, 2006; MENDES-BONATO; PAGLIARINI; VALLE, 2006), o que foi posteriormente confirmado por hibridização *in situ* genômica (GISH), subsidiando a proposta a constituição genômica BBB^1B^1 e $B^1B^1B^2B^2$ para *U. brizantha* e *U. decumbens* e B^2B^2 para *U. ruziziensis*, (PAULA; SOUZA SOBRINHO; TECHIO, 2017).

Correa et al. (2020), também com o uso da GISH, investigaram a ancestralidade dos genomas das duas cultivares tetraploides, comparando seus genomas com as respectivas diploides *U. brizantha* e *U. decumbens*, e com *U. ruziziensis.* Os autores verificaram alta homologia entre os genomas dos genótipos diploides de *U. brizantha* e *U. decumbens* e as sondas genômicas de ambas evidenciaram o genoma B^1 quando hibridizadas com as cultivares tetraploides. Nesse sentido, foram determinadas as constituições genômicas B^1B^1 para *U. brizantha* (2x) e B^1'B^1' para *U. decumbens* (2x), e indicadas como potenciais doadoras ancestrais do genoma B^1 para as alotetraploides.

Posteriormente, Tomaszewska et al. (2023) investigaram as relações de ancestralidade no complexo 'brizantha' usando sondas genoma-específicas e propuseram a constituição dos subgenomas de acessos tetraploides de *U. brizantha* (B^bB^bDD), *U. decumbens* (B^bDRR) e *U. ruziziensis* (RR). Ambas as nomenclaturas evidenciam que *U. ruziziensis* e *U. decumbens* 2x seriam ancestrais na origem alotetraploide de *U. decumbens*, sendo que Paula et al. (2017) e Correa et al. (2020) inferiram uma hibridação direta entre *U. decumbens* 2x e *U. ruziziensis 2x*, enquanto Tomaszewska et al (2023) propuseram que *U. decumbens 4x* teria se originado a partir de dois eventos de hibridação [(*U. brizantha* 2x x *U. decumbens* 2x) x *U. ruziziensis 2x*].

As diferenças genômicas percebidas em nível cromossômico são, geralmente, reflexos das diferenças na composição do DNA repetitivo, visto que a fração repetitiva pode representar mais de 80% do genoma em espécies de Poaceae (LING et al., 2013; SCHNABLE et al., 2009), compondo, pelo menos, metade do genoma em *U. ruziziensis* (WORTHINGTON et al., 2021). Portanto, o estudo de sequências de DNA repetitivo pode fornecer informações sobre a evolução dos genomas de plantas e subsidiar estudos filogenéticos e de relação genômica (JIANG, 2019 (HESLOP-HARRISON et al. 2023).

As sequências de DNA repetitivo podem ser classificadas de acordo com sua estrutura e localização nos cromossomos, incluindo (i) repeats em tandem e (ii) sequências dispersas. Os repeats em tandem representam o maior componente das regiões heterocromáticas e estão associadas a funções específicas nos cromossomos, como a formação de centrômeros e

telômeros (GARRIDO-RAMOS, 2021). As sequências dispersas são encontradas em regiões diversas ao longo do genoma, sendo majoritariamente representadas pelos elementos transponíveis, que são divididos em transposons de DNA e retrotransposons e subdivididos de acordo com sua estrutura, composição e mecanismo de transposição (WELLS; FESCHOTTE, 2020). Os elementos transponíveis são capazes de se replicar nos genomas de maneira independente ao ciclo celular, frequentemente codificando proteínas e sequências regulatórias não codantes que promovem sua propagação (WELLS; FESCHOTTE, 2020).

O estudo do DNA repetitivo tem sido facilitado com o uso do sequenciamento de nova geração (NGS) e da plataforma RepeatExplorer, que permite a identificação de famílias de repeats em espécies sem um genoma de referência (NOVÁK et al., 2013). Uma vez identificadas as classes de repeats, é possível desenhar primers e construir sondas específicas, que podem ser mapeadas fisicamente nos cromossomos por meio da técnica de hibridização *in situ* fluorescente (FISH).

Em *Urochloa*, somente *U. ruziziensis* possui o genoma completamente sequenciado (WORTHINGTON et al., 2021) e a quantidade de sequências de outras espécies de *Urochloa* depositadas em bancos de dados genéticos é relativamente pequena, o que limita o detalhamento dos genomas das diferentes espécies. A maior parte dos estudos com o gênero envolvendo DNA repetitivo se resumem ao mapeamento físico de sequências de DNA ribossomal (AKIYAMA; YAMADA-AKIYAMA; EBINA, 2010; NIELEN et al., 2009) e retrotransposons centroméricos (NANI et al., 2016; SANTOS et al., 2015).

A caracterização citogenômica do DNA repetitivo em *Urochloa* pode ampliar o conhecimento sobre a estrutura e função dessas sequências em genomas vegetais, possibilitar inferências evolutivas, permitir a construção de sondas genoma-específicas, além de subsidiar a escolha de parentais para cruzamentos em programas de melhoramento, com o delineamento de relações genômicas entre espécies.

Diante deste cenário, o presente estudo objetivou caracterizar a fração repetitiva dos genomas de citótipos diploides e tetraploides de *U. brizantha*, *U. decumbens* e *U. ruziziensis*, identificando as principais sequências e entendendo sua composição, com a identificação das

diferentes linhagens de DNA repetitivo e suas proporções genômicas, e sua organização, através da construção de sondas repetitivas para mapeamento físico dos cromossomos por meio da FISH.

2 REFERENCIAL TEÓRICO

2.1 O DNA Repetitivo em genomas eucariotos

O genoma de organismos eucariotos é composto por sequências de DNA de cópia única e sequências de DNA repetitivo. O DNA repetitivo consiste em unidades (monômeros) de sequências de nucleotídeos, com tamanho variável desde dinucleotídeos (como por exemplo a repetição GAGAGA) a unidades maiores que 10.000 pares de bases, que se repetem desde centenas a centenas de milhares de vezes nos genomas (HESLOP-HARRISON; SCHMIDT, 2012).

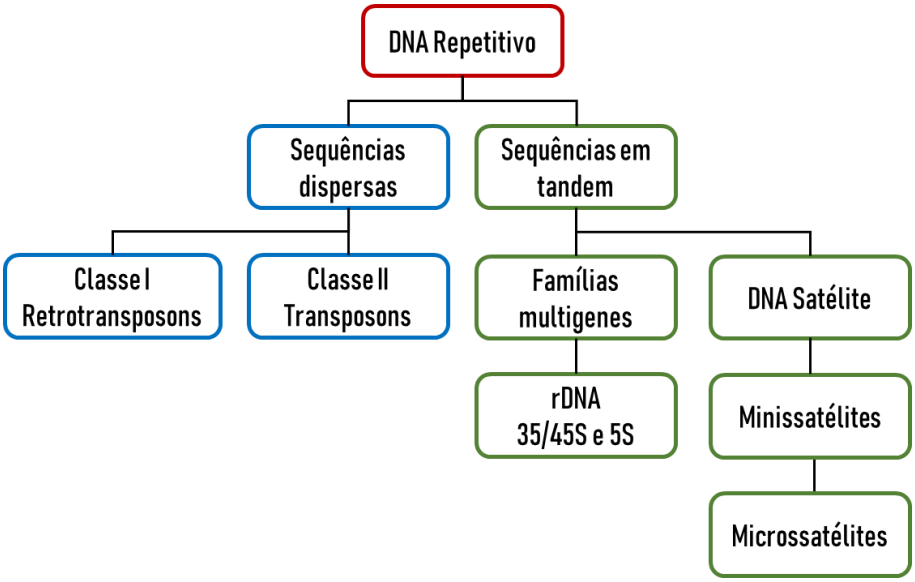
Muito antes do conhecimento sobre sequências de DNA, já era possível distinguir regiões contrastantes nos cromossomos, identificadas como eucromatina e heterocromatina. Posteriormente, foi reconhecido que várias das regiões heterocromáticas dos cromossomos são compostas por DNA repetitivo (BISCOTTI; OLMO; HESLOP-HARRISON, 2015). A existência da fração repetitiva do DNA foi inicialmente classificada por Briiten e Kohne (1968), que dividiram as sequências em moderadamente ou altamente repetitivas, de acordo com seu grau de repetibilidade.

Apesar da porção repetitiva dos genomas incluir genes funcionais, como as sequências de DNA ribossômico (DNAr) 35S e 5S, a maior parte do DNA repetitivo foi por muito tempo considerada ‘DNA lixo’, sem nenhuma função conhecida (FREELING et al., 2015). Atualmente, com a ampla utilização de plataformas de sequenciamento de nova geração (NGS) e mais de 50 anos de estudos genômicos, a funcionalidade do DNA repetitivo nos genomas tem sido cada vez mais evidenciada (BISCOTTI; OLMO; HESLOP-HARRISON, 2015; TANG, 2017).

O conteúdo de DNA repetitivo no genoma de eucariotos é altamente variável. Em animais, sequências repetitivas podem ocupar uma porção relativamente pequena do genoma, como em *Caenorhabditis elegans*, que possui cerca de 16% de DNA não codante (STEIN et al., 2003), ou podem ser predominantes, como em humanos, cuja fração repetitiva ocupa mais de dois terços do genoma (LIEHR, 2021). Plantas tendem a possuir, no geral, uma maior proporção de DNA repetitivo, que pode representar até 92% do genoma, como é o caso da cebola (FU et al., 2019). Nesse contexto, a presença de DNA repetitivo em genomas eucariotos é o principal fator relacionado ao paradoxo do valor C, que se refere à falta de correlação entre tamanho do genoma e complexidade dos organismos (FREELING et al., 2015).

O DNA repetitivo pode ser dividido em várias classes de sequências (Figura 1), que se distinguem em sua organização e localização ao longo dos cromossomos, apesar de existirem formas de arranjo intermediário. Os elementos repetitivos são primariamente divididos em sequências repetitivas em tandem e sequências repetitivas dispersas.

Figura 1: Classificação dos elementos repetitivos do genoma nuclear de plantas



Fonte: adaptado de Heslop-Harrison & Schmidt (2012).

2.1.1 Repetições em tandem

A repetição em tandem se caracteriza por unidades de sequências de DNA organizadas adjacentes umas às outras em arranjos monótonos. Satélites e famílias multigênicas são exemplos dessas sequências (HESLOP-HARRISON; SCHMIDT, 2012).

2.1.1.1 Famílias Multigênicas

Famílias multigênicas são grupos de genes parálogos, que tipicamente exibem sequências e funções relacionadas. Uma família multigênica é produzida quando um gene é copiado uma ou mais vezes por eventos de duplicação gênica, como poliploidização, recombinação ectópica ou retrotransposição, por exemplo. Duplicações podem ocorrer várias vezes e produzir numerosas cópias de um gene particular. O tamanho das famílias pode variar de dois até centenas de unidades (ROSSELLÓ; MARAVILLA; ROSATO, 2022).

Os genes ribossomais são o exemplo mais conhecido de famílias multigênicas. O RNA ribossômico (RNAr) cumpre um papel vital na síntese proteica, constituindo o principal componente estrutural e catalítico dos ribossomos (ROSSELLÓ; MARAVILLA; ROSATO, 2022). Em eucariotos, o DNAr é constituído por repeats em tandem, sendo três dos quatro genes que codificam RNAr localizados na região organizadora do nucléolo (RON) em um ou mais cromossomos. Cada unidade de repetição contém a subunidade maior 28S, a subunidade menor 18S, o gene 5.8S, dois espaçadores externos transcritos (ETS), dois espaçadores internos transcritos (ITS1 e ITS2) e um espaçador não transcrito (NTS). Dessa maneira a unidade 5’-ETS-18S-ITS1-5.8S-ITS2-28S-ETS-3’ é organizada em cópias sequenciais que variam de 39 a 19.300 em animais e de 150 a 26.000 em plantas (PROKOPOWICH; GREGORY; CREASE, 2003). Esses genes são transcritos no nucléolo pela RNA polimerase I em um transcrito primário, que é menor em plantas (35S) do que em animais (45S) (ROSSELLÓ; MARAVILLA; ROSATO, 2022).

O DNAr 5S é o quarto gene que codifica rRNA, que compreende repetições em tandem do gene, separados por um NTS. Na maioria dos eucariotos, os genes de rRNA 5S se encontram em clusters separados dos blocos de rDNA 35S (arranjo tipo S), mas em alguns

casos, como em espécies de Asteraceae, cerca de 30% das gimnospermas (GARCIA; KOVAŘÍK, 2013) e em um acesso de *Urochloa humidicola* (DAMASCENO et al., 2023), o rDNA 5S se localiza ligado às unidades 35S (arranjo tipo L).

Os diferentes componentes do rDNA evoluem, em geral, em taxas distintas. O rDNA 18S é um dos genes de evolução mais lenta conhecido, ao contrário dos espaçadores, que são sequências de evolução rápida, com o NTS evoluindo mais rapidamente do que os ITSs e ETSs. O gene do rRNA 28S também tem evolução relativamente lenta (ROSSELLÓ; MARAVILLA; ROSATO, 2022). Devido ao seu alto grau de conservação em eucariotos, os rDNAs ribossomais são amplamente utilizados em descrições cariotípicas, apresentando variações em número, tamanho, localização cromossômica e atividade transcricional entre espécies (GARCIA et al., 2017).

2.3.1.2 DNA Satélite

O termo DNA satélite (satDNA) foi cunhado por Pech et al. (1979), se referindo a ‘sequências de DNA altamente repetitivas organizadas em tandem’, pois essas sequências, quando submetidas à centrifugação isopícnica, formavam picos ‘satélites’ em relação à banda de DNA principal.

As famílias de satDNA diferem em relação à sua localização, tamanho dos repeats, número de cópias e sequência de nucleotídeos. O DNA satélite é o principal componente da heterocromatina, que é encontrada predominantemente nas regiões pericentroméricas e subteloméricas dos cromossomos. Adicionalmente, o satDNA é um componente frequente de centrômeros, que também podem ser ocupados por elementos transponíveis (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012).

Monômeros de satDNA podem variar de dois a milhares de pares de bases, mas a maioria das espécies apresentam monômeros de 135-195 e 315-375 nucleotídeos, tamanhos similares ao necessário para contornar um ou dois nucleossomos (GARRIDO-RAMOS, 2021). Convencionalmente, repetições de 2 a 10 pares de bases são chamadas de microssatélites, monômeros de 10 a 100 pb são chamados de minissatélites e os satélites

propriamente ditos possuem unidades de repetição que variam de centenas a milhares de pares de bases (GARRIDO-RAMOS, 2021).

Os microssatélites, também conhecidos como repetições de sequência simples (SSRs), são geralmente organizados em arranjos menores que 1 kb e estão presentes na maioria dos genomas eucariotos (GARRIDO-RAMOS, 2021). O número de repetições de um loco de microssatélite tipicamente varia entre cinco e 40, sendo possível repetições mais longas. Os arranjos de microssatélites são encontrados amplamente dispersos nos genomas, em regiões codificantes ou não, apesar de ser possível a formação de grandes blocos em regiões heterocromáticas (GARRIDO-RAMOS, 2021). Dinucleotídeos são o tipo mais comum de microssatélites para várias espécies e repetições contendo múltiplos de três nucleotídeos (repeats trinucleotídeos e hexanucleotídeos) são mais abundantes em regiões codificantes (GARRIDO-RAMOS, 2021). O dinucleotídeos mais comum no genoma humano é $(CA)_n/(GT)_n$, ao passo que os repeats $(GA)_n/(CT)_n$ e $(AT)_n/(TA)_n$ são mais comuns em plantas (LIEHR, 2021).

Os repeats de microssatélites são tipicamente flanqueados por sequências de DNA de cópia única ou de baixo número de cópias. Ao passo que essas sequências flanqueadoras possuem certo grau de conservação, loci de microssatélites tendem a variar no número de repeats em táxons relacionados, o que os torna marcadores polimórficos altamente informativos. Por esse motivo, os microssatélites têm sido cada vez mais utilizados como marcadores moleculares para mapeamento genético e análises genômicas (VIEIRA et al., 2016).

Minissatélites também possuem polimorfismo em relação ao número de repetições, podendo ser utilizados como marcadores. No entanto, o maior número de variantes de microssatélites em relação a minissatélites fez com que o primeiro se tornasse o marcador preferencial em estudos genômicos (ADHIKARI et al., 2017). O DNA minissatélite tende a se concentrar nas regiões subteloméricas e pericentroméricas dos cromossomos de plantas como o trigo (LANG et al., 2019), centeio (XI et al., 2020), abóbora (MATYÁŠEK et al., 2019) e ginseng (WAMINAL et al., 2018).

Os DNAs satélites não se diferenciam apenas no tamanho do monômero, mas também no número de cópias e abundância no genoma. Em uma mesma espécie, as unidades de repetição não são estritamente idênticas entre si, mas exibem polimorfismos nas sequências. Entretanto, esses repeats são mais semelhantes entre si do que quando comparados a repeats de espécies diferentes, seguindo um padrão conhecido como evolução concertada. Nesse modelo de evolução, mecanismos de permuta desigual e conversão gênica atuam homogeneizando mutações ocorridas em um repeat para outras unidades de repetição (BELYAYEV et al., 2019).

Existem diferentes níveis de variações intra e interespecífica nas sequências satélites, dependentes dos fatores evolutivos atuantes. A homogeneização intraespecífica de repeats é acompanhada por uma rápida divergência entre repeats de diferentes espécies e geralmente leva à ocorrência de satDNAs espécie-específicos. Por outro lado, existem famílias de DNA satélite amplamente conservadas em um mesmo gênero, família ou até mesmo em taxa de maior ordem. Nesse contexto, as diferenças entre sequências de satDNA podem ser analisadas e utilizadas para inferir relações filogenéticas em diversos grupos (HESLOP-HARRISON; SCHMIDT, 2012).

2.1.2 Sequências repetitivas dispersas – Elementos transponíveis.

Ao contrário dos arranjos em tandem, as sequências repetitivas dispersas se espalham por todo o genoma, intercaladas com outras sequências e distribuídas na maioria dos cromossomos. Os elementos transponíveis (TEs) são os principais representantes dos elementos dispersos (HESLOP-HARRISON; SCHMIDT, 2012). Essas sequências foram originalmente identificadas por Barbara McClintock no genoma do milho, que os descreveu como agentes hipermutagênicos que agiam em genes individuais, tendo um grande impacto na estrutura do genoma (MCCLINTOCK, 1948).

Posteriormente à sua descoberta, elementos transponíveis foram encontrados em todos os organismos, incluindo todas as espécies vegetais investigadas. Em várias plantas com genomas grandes e complexos, os elementos transponíveis compõem mais de 50% do DNA nuclear. Os elementos transponíveis compartilham duas propriedades básicas: a capacidade de

se moverem de um local para outro do genoma, e a capacidade de amplificar seu número de cópias no genoma, motivo pelo qual são frequentemente chamados de DNA egoísta ou parasita (ÅGREN; CLARK, 2018).

Devido à grande abundância e diversidade dos elementos transponíveis nos genomas, Wicker et al. (2007) propuseram uma classificação unificada para eucariotos, baseada nos mecanismos de transposição, similaridade de sequências e relações estruturais. Essa classificação divide os elementos transponíveis hierarquicamente em classe, subclasse, ordem, superfamília e família ou linhagem.

Segundo a classificação de Wicker et al. (2007), as classes são separadas de acordo com a presença ou ausência de um RNA intermediário no processo da transposição. As subclasses distinguem elementos que fazem cópias deles próprios daqueles que deixam o sítio doador para se reintegrarem em outro local. As ordens marcam as principais diferenças no mecanismo de inserção e, consequentemente, na organização geral e enzimologia dos elementos transponíveis. As superfamílias compartilham a estratégia de replicação, mas distinguem-se por características gerais, como estrutura de proteínas ou domínios não codificadores e também são diferenciadas pela presença e tamanho do TSD (do inglês target site duplication), um pequeno repeat que é gerado nas duas extremidades de um TE durante a inserção. Famílias ou linhagens são definidas pela conservação das sequências de DNA e as subfamílias com base em dados filogenéticos e podem, em casos específicos, distinguir elementos autônomos de não autônomos, segundo a integridade do elemento.

Neumann et al. (2019) atualizaram a classificação de Wicker et al. (2007) com base em análises filogenéticas de domínios proteicos de elementos transponíveis, que incluíram mais de 13.000 elementos pertencentes a mais de 80 espécies de diversos grupos de plantas. A nova classificação (Tabela 1) manteve a categorias propostas por Wicker et al. (2007) e trouxe alterações em algumas linhagens das superfamílias Copia e Gypsy.

A ubiquidade dos elementos transponíveis em todos os organismos vivos sugere uma origem antiga desses DNAs móveis. O conceito de DNA egoísta ou parasita sugere que a habilidade de amplificação dentro dos genomas poderia ser selecionada para qualquer

sequência, desde que isso não diminuísse significativamente o fitness do hospedeiro, o que possibilitaria origens múltiplas e independentes desses elementos (BENNETZEN, 2000).

Tabela 1: Classificação de elementos transponíveis

Classe I (Retrotransposons)			Classe II (Transposons de DNA)	
Ordem	Superfamília	Linhagem	Ordem	Superfamília
LTR	*Copia*	*Ale*	Subclasse I	
		Alesia	TIR	Tc1-Mariner
		Angela		hAT
		Bianca		Mutator
		Bryco		Merlin
		Lyco		Transib
		Gymco-I-IV		P
		Ikeros		PiggyBac
		Ivana		PIF-Harbinger
		Osser		CACTA
		SIRE	Crypton	Crypton
		TAR	Subclasse II	
		Tork	Helitron	Helitron
	Gypsy	*CRM*	Maverick	Maverick
		Chlamyvir		
		Galadriel		
		Tcn1		
		Reina		
		Athila		
		Tat I-III		
		Ogre		
		Retand		
		Phygy		
		Selgy		
	Bel-Pao			
	Retrovirus			
	ERV			
DIRS	DIRS			
	Ngaro			
	Viper			
PLE	Penelope			
LINE	R2			
	RTE			
	Jockey			
	L1			
	I			
SINE	tRNA			
	TSL			
	5S			

Fonte: adaptado de Wicker et al. (2007) e Neumann et al. (2019).

A atividade de elementos transponíveis pode causar alterações funcionais quando inseridos dentro da região codificante de um determinado gene, podendo inativá-lo ou alterar seu produto e, quando inserido dentro ou próximos a regiões promotoras, podem alterar os padrões de expressão gênica (KASHKUSH; FELDMAN; LEVY, 2003). Os TEs podem ainda causar quebras cromossômicas, deslocamento de outras porções de DNA durante a transposição, recombinação ilegítima e, quando replicativos, aumento do tamanho do genoma do hospedeiro (FEDOROFF, 2000).

2.1.3 DNA Centromérico

O centrômero, ou constrição primária, é a região cromossômica que separa o braço curto e o braço longo, e na qual as fibras do fuso se ligam durante a divisão celular. Centrômeros podem ser classificados de maneira generalizada em monocêntricos, nos quais um centrômero é montado em uma única região localizada, ou holocêntricos, no qual o centrômero se estende por todo o comprimento do cromossomo. A maioria dos eucariotos tem cromossomos monocêntricos, geralmente contendo quilobases a megabases de DNA (MCKINLEY; CHEESEMAN, 2016).

Em eucariotos, a região centromérica é definida epigeneticamente, de modo que sequências específicas de DNA não são estritamente necessárias para a formação de um centrômero funcional, sendo a presença da histona centromérica CENP-A/CENH3 a característica unificadora dos centrômeros eucarióticos (OLIVEIRA; TORRES, 2018). No entanto, existem preferências evolutivas e funcionais para estruturas específicas de DNA que contribuem para a função do centrômero, sendo uma delas o DNA repetitivo (SHATSKIKH et al., 2020).

Em geral, centrômeros contêm grandes blocos de DNA satélite que podem se estender a megabases de comprimento (MCKINLEY; CHEESEMAN, 2016). Satélites centroméricos possuem monômeros de 150-180 pb, tamanho coincidente com o necessário para que um segmento de DNA complete uma volta em torno de um nucleossomo (LERMONTOVA et al., 2015).

Ao contrário das repetições teloméricas, que são altamente conservadas em eucariotos, sequências centroméricas se divergem mais rapidamente entre espécies por estarem mais susceptíveis a eventos de recombinações desiguais, transposição e excisão (LERMONTOVA et al., 2015). Embora as sequências centroméricas evoluam rapidamente, elas tendem a ser homogeneizadas dentro do mesmo genoma, o que faz com que os centrômeros de todos os cromossomos do complemento tenham uma composição de DNA centromérico similar (KAWABE; NASUDA, 2005). No entanto, existem casos de alta variação entre sequências centroméricas dentro de uma mesma espécie, como em potato (*Solanum tuberosum*), cujos centrômeros de cada um de seus 12 cromossomos possuem uma composição exclusiva (ZHANG et al., 2014).

Comparações citogenéticas entre espécies relacionadas revelaram que alguns centrômeros adotam novas posições ao longo do tempo evolutivo, fenômeno conhecido como reposicionamento de centrômero (DAWE; HIATT, 2004; ROCCHI et al., 2012). Inicialmente, os novos centrômeros, ou neocentrômeros, são formados sobre sequências de DNA não repetitivo, mas posteriormente adquirem o DNA satélite específico da espécie, se tornando eventualmente indistinguíveis dos centrômeros "antigos" (ROCCHI et al., 2012). A aquisição de repetições em tandem por neocentrômeros é acompanhada de um decaimento das mesmas sequências no local ancestral, o que evidencia a contribuição dessas repetições para a função centromérica (KALITSIS; CHOO, 2012). Especula-se que o DNA satélite funcione como uma plataforma para a ligação do cinetócoro, recrutando histonas centroméricas e marcas epigenéticas típicas de heterocromatina (MCKINLEY; CHEESEMAN, 2016).

O DNA satélite desempenha um papel fundamental nos centrômeros de gramíneas, sendo o tipo de repetição em tandem mais abundante nos genomas desse grupo (BILINSKI et al., 2016). Quando comparados com centenas de espécies de animais e plantas, os satélites centroméricos de gramíneas incluindo cevada, arroz, milho, milheto, *Panicum* e *Paspalum*, possuem alta conservação no comprimento de monômero (sendo 156 bp o tamanho mais frequente) e de sequência de nucleotídeos (MELTERS et al., 2013; QI et al., 2019). Além disso, gramíneas também apresentam em média 2-3 variantes de satélites centroméricos,

sendo possível a coexistência de diferentes sequências dentro de um único bloco de repetições em tandem em um mesmo centrômero (MELTERS et al., 2013).

Elementos transponíveis também podem fazer parte da composição de centrômeros. Em geral, os retrotransposons compõem grande parte dos genomas vegetais e divergem rapidamente em novas linhagens ao longo do tempo (LERMONTOVA et al., 2015). No entanto, essas repetições tendem a ser sub representadas em centrômeros, e algumas linhagens são altamente conservadas, como a linhagem CRM, que já foi observada em várias gramíneas como milho (JIANG et al., 2003), sorgo (MILLER et al., 1998), cevada (STING et al., 1998) e *Urochloa* (NANI et al., 2016).

2.1.4 Uso de sequências repetitivas como marcadores citogenéticos

Marcadores citogenéticos são estruturas cromossômicas visíveis por microscopia, que permitem a identificação de regiões específicas em um cromossomo. Devido ao alto grau de compactação do DNA em cromossomos metafásicos, sequências repetitivas são tradicionalmente usadas como marcadores citogenéticos, uma vez que o grande número de cópias nos genomas permite uma melhor visualização (JIANG, 2019).

Em espécies vegetais, bandas heterocromáticas e a região organizadora no nucléolo foram extensivamente utilizadas como marcadores cromossômicos em estudos empregando técnicas de bandamento (ZOSHCHUK; BADAEVA; ZELENIN, 2003). Após o desenvolvimento da FISH, os genes de DNA ribossomal 35S e 5S se tornaram os marcadores mais populares em cariotipagem, devido a sua onipresença em genomas eucariotos (GARCIA et al., 2017). No entanto, as sequências ribossomais tendem a se distribuir em um número relativamente baixo de loci, sendo necessário outros marcadores para a identificação dos cromossomos.

O avanço das tecnologias de sequenciamento e as análises bioinformáticas facilitaram o desenvolvimento de novos tipos de marcadores. Em espécies que possuem o genoma completamente sequenciado e montado, é possível identificar oligonucleotídeos únicos de cada cromossomo com mais facilidade, o que permite a criação de sondas cromossomo-

específicas, que podem ser mapeadas através das técnicas pintura cromossômica (RIED et al., 1998), oligo-FISH (BRAZ et al., 2018) e chromosome barcoding (IDZIAK et al., 2014). No entanto, o custo do sequenciamento completo ainda limita as aplicações da oligo-FISH para estudos que envolvem muitas espécies ou para espécies menos conhecidas.

Alternativamente, monômeros de elementos repetitivos podem ser facilmente identificados em pequenas frações (<1%) sequenciadas dos genomas vegetais com análises de clustering como realizada pela pipeline RepeatExplorer (NOVÁK et al., 2013), o que permite a construção de sondas para espécies sem um genoma de referência, com um menor custo. Essa abordagem tem sido utilizada para mapear cromossomos em várias gramíneas, como em espécies de *Festuca*, (EBRAHIMZADEGAN et al., 2019; KŘIVÁNKOVÁ et al., 2017), *Brachypodium* (LI et al., 2018), *Avena* (LIU et al., 2019) e *Urochloa* (TOMASZEWSKA et al., 2023).

Em *Festuca*, análises de clustering identificaram 18 sequências de DNA satélite, das quais cinco foram utilizados como marcadores citológicos, o que permitiu a detecção de satélites específicos de cromossomos B em *F. pratensis* (EBRAHIMZADEGAN et al., 2019). Em *Avena*, a mesma estratégia foi usada para identificar e mapear sequências de retrotransposons específicas dos genomas A C e D em oito espécies (LIU et al., 2019).

Em *Urochloa*, além do mapeamento de genes ribossomais em várias espécies (NIELEN et al., 2009; AKIYAMA et al. 2010; NANI et al., 2016; PAULA; SOUZA SOBRINHO; TECHIO, 2017; MORAES et al. 2021), sequências repetitivas identificadas por análises de clustering e k-mer foram utilizadas para mapear cromossomos relacionados a subgenomas de acessos poliploides de *U. brizantha* (CIAT 26032), *U. decumbens* (CIAT 26305), *U. humidicola* (CIAT 26155) e de um acesso diploide de *U. ruziziensis* (CIAT 26162), permitindo a elucidação de relações genômicas entre dois complexos agâmicos no gênero (TOMASZEWSKA et al., 2023).

O sequenciamento completo de *U. ruziziensis* (acesso CIAT 606) revelou que 51% de seu genoma é composto por DNA repetitivo, com predominância de retrotransposons da superfamília *Gypsy* (23.9%) e *Copia* (9.5%) (WORTHINGTON et al., 2021). As proporções

de repetições observadas foram similares ao encontrado para *Setaria italica* (L.) P. Beauv., uma espécie amplamente utilizada como referência em estudos de gramíneas tropicais. Os elementos repetitivos identificados no genoma de *U. ruziziensis* podem ser potencialmente utilizados como marcadores citogenéticos, mas ainda não foram realizados estudos focando no mapeamento físico dessas repetições em espécies de *Urochloa*.

2.2 Aspectos botânicos e taxonômicos de *Urochloa*

O gênero *Urochloa* P. Beauv. (sin. *Brachiaria* (Trin.) Griseb.) pertence à tribo Paniceae, subfamília Panicoideae e família Poaceae. *Urochloa* foi originalmente descrito por Palisot de Beauvois, em 1812, como um gênero paleotropical de gramíneas, que compreendia cerca de 12 espécies (TORRES GONZÁLEZ; MORTON, 2005). No entanto, revisões taxonômicas posteriores incluíram espécies previamente pertencentes ao gênero *Brachiaria* em *Urochloa* e atualmente, os dois gêneros são considerados sinônimos (SALARIATO et al., 2010; VELDKAMP, 2004).

Brachiaria foi descrita inicialmente em 1826, por Trinius, como uma subdivisão do gênero *Panicum* L. Posteriormente, o grupo foi elevado a gênero por Grisebach, em 1853. A orientação das espiguetas e a posição da gluma inferior foram definidas por Nash (1903) como características diagnósticas do gênero (MORRONE; ZULOAGA, 1992). Quando ainda eram tratados como gêneros distintos, os limites taxonômicos de *Brachiaria* e *Urochloa* se fundamentavam na morfologia floral. As espécies de *Brachiaria* foram descritas como anuais ou perenes com inflorescências em forma de espigueta, dispostas em racemos unilaterais ao longo de uma ráquis central filiforme (RENVOIZE; CLAYTON; KABUYE, 1996). As espiguetas foram caracterizadas como sésseis ou pediceladas, individuais ou em pares, com gluma inferior adjacente à ráquis, em posição adaxial (CLAYTON; RENVOIZE, 1986). Espécies de *Urochloa* também foram descritas como anuais ou perenes, com inflorescências racemosas e espiguetas dispostas individualmente ou em pares, mas com gluma oposta à ráquis, em posição abaxial (CLAYTON; RENVOIZE, 1986).

A confiabilidade dos caracteres utilizados para separar *Brachiaria* de *Urochloa* foi questionada por vários autores. Um dos argumentos é que a diferença na orientação das

espiguetas (adaxial, com gluma inferior adjacente à ráquis em *Brachiaria*; e abaxial, com gluma inferior oposta à ráquis em *Urochloa*) só é válida para espiguetas individuais em ramos primários (CLAYTON; RENVOIZE, 1986). Quando em pares, as espiguetas dos dois gêneros possuem orientação adaxial (MORRONE; ZULOAGA, 1992).

Como resposta à problemática taxonômica, Veldkamp (2004) propôs a transferência da maior parte das espécies de *Brachiaria* para *Urochloa*, com exceção de três espécies (*B. eruciformes*, *B. malacodes* e *B. schoenfelderi*) que foram realocadas ao novo gênero *Moorochloa* Veldkamp. Posteriormente, filogenias moleculares utilizando região ITS (TORRES GONZÁLEZ; MORTON, 2005) e genes cloroplastidiais (SALARIATO et al., 2010) corroboraram a revisão de Veldkamp, mostrando que *Brachiaria* e *Urochloa* não formam clados distintos. Atualmente, *Urochloa* compreende 87 espécies, sendo a maior parte delas encontrada no continente africano, que é considerado seu centro de origem (POWO, 2023).

No Brasil, o gênero foi introduzido oficialmente em 1952, com a espécie *Urochloa decumbens* (Stapf) R.D. Webster. Entretanto, seu estabelecimento em larga escala só ocorreu após as importações de *Urochloa ruziziensis* (R. Germ. & C.M. Evrard) Morrone & Zuloaga, *Urochloa brizantha* (Hochst. ex A. Rich.) R.D. Webster a partir de 1965. Na década de 70, houve um enorme aumento das áreas de pastagens cultivadas com as espécies *U. decumbens*, *U. humidicola* (Rendle) Morrone & Zuloaga, *U. brizantha* e *U. ruziziensis*, que se adaptaram bem ao clima e solo brasileiros (ALVIM; BOTREL; XAVIER, 2002).

Devido a ocorrência natural de hibridações interespecíficas entre espécies poliploides apomíticas e diploides sexuais, algumas espécies de *Urochloa* foram organizadas em complexos agâmicos: *U. brizantha*, *U. decumbens* e *U. ruziziensis* foram alocadas no complexo agâmico ‘brizantha’, e *U. humidicola* e *U. dictyoneura* pertencem ao complexo agâmico ‘humidicola’ (LUTTS; NDIKUMANA; LOUANT, 1991; RENVOIZE; MAASS, 1993).

As relações filogenéticas entre *U. brizantha*, *U. decumbens* e *U. ruziziensis* – demonstram controvérsias. De acordo com o banco de dados botânico Plants of the World, de

Kew Gardens (POWO, 2023), *U. decumbens* e *U. ruziziensis* não são aceitas como espécies, sendo ambas sinonímias de *U. eminii*. No entanto, os últimos estudos filogenéticos no gênero as trataram como táxons distintos e não fizeram menção a sinonímia com *U. eminii* (HIGGINS et al., 2022; PESSOA-FILHO; MARTINS; FERREIRA, 2017; SALARIATO et al., 2010; TRIVIÑO et al., 2017a), logo no presente trabalho também trataremos *U. decumbens* e *U. ruziziensis* como espécies distintas.

Renvoize et al. (1996) em um estudo baseado na morfologia das espiguetas, classificaram 83 espécies de *Urochloa* em nove grupos diferentes, sendo *U. brizantha*, *U. decumbens* e *U. ruziziensis* pertencentes ao mesmo grupo. Os autores destacaram ainda uma maior proximidade entre *U. brizantha* e *U. decumbens*, havendo dificuldade de distinção entre as duas e ainda, reidentificaram *U. decumbens* cv. Basilisk como *U. brizantha*.

A proximidade entre *U. brizantha* e *U. decumbens* também foi evidenciada por estudos filogenéticos baseados em marcadores RAPD (AMBIEL et al. 2008, 2010) e genes cloroplastidiais (PESSOA FILHO et al. 2017), que sugeriu que a linhagem de *U. ruziziensis* teria divergido de *U. brizantha* e *U. decumbens* há 5.67 milhões de anos, seguida de uma divergência mais recente entre estas por volta de 1.6 milhões de anos atrás.

Em contrapartida, Triviño et al. (2017), em um estudo de diversidade genética e estrutura populacional baseado em marcadores microssatélites com 601 acessos diploides e poliploides de *Urochloa*, verificaram maior proximidade em *U. decumbens* e *U. ruziziensis*, em relação à *U. brizantha* e determinou que a cultivar Basilisk é de fato mais próxima dos outros acessos tetraploides de *U. decumbens*. Essa relação foi corroborada mais recentemente por Higgins et al. (2022), que realizaram análises de diversidade genética e estrutura populacional em 111 acessos diploides e poliploides de cinco espécies de *Urochloa* (*U. brizantha*, *U. decumbens*, *U. humidicola*, *U. máxima* e *U. ruziziensis*). Os autores concluíram que a poliploidia é um fator fundamental nas relações filogenéticas do complexo agâmico ‘brizantha’, visto que se observou que acessos tetraploides de *U. decumbens* e *U. brizantha* são filogeneticamente mais próximos entre si do que de suas formas diploides, e que acessos diploides de *U. decumbens* são próximos a *U. ruziziensis*. Os autores também apontaram que

um único evento de poliploidização tenha originado as formas tetraploides de *U. brizantha* e *U. decumbens*, uma vez que as análises mostraram um agrupamento de acessos das duas espécies com ancestralidade comum, mas alta diferenciação genética em relação a agrupamentos contendo apenas acessos de *U. brizantha.*

2.3 Caracterização citogenética de *Urochloa*

Estudos citogenéticos em *Urochloa* identificaram números básicos de cromossomos x=5, 8, 10 e 12 (CHRISTOPHER; ABRAHAM, 1976), x=6 (RISSO-PASCOTTO; PAGLIARINI; VALLE, 2006) e x=7 e 9 (DARLINGTON; WYLLIE, 1956), sendo estes últimos encontrados com maior frequência entre as espécies (VALLE; PAGLIARINI, 2009; BASAPPA; MUNIYAMMA; CHINNAPPA, 1987; SPIES et al., 1991; VALLE; SINGH; MILLER, 1987). Valle e Pagliarini (2009) revisaram os números cromossômicos do gênero descritos na literatura e encontraram espécies variando de 2n=14 a 2n=90, sendo algumas delas caracterizadas pela ocorrência de citótipos, tais como *U. brizantha* (2n=2x=18, 2n=4x=36, 2n=5x=45; 2n=6x=54), *U. decumbens* (2n=2x=18; 2n=4x=36), *U. ruziziensis* (2n=2x=18) e *U. humidicola* (2n=6x=36 ou 2n=9x=54).

Contagens cromossômicas para *Urochloa* iniciaram com levantamentos gerais para gramíneas tropicais entre décadas de 1930-60 (AVDULOV, 1931; BROWN, 1951; DE WET, 1960; DE WET; ANDERSON, 1956; MOFFETT; HURCOMBE, 1949). Os primeiros estudos citológicos dedicados exclusivamente a espécies *Urochloa*, ainda identificadas como *Brachiaria*, foram realizados por Sotomayor-Rios et al. (1970), que relataram os números cromossômicos 2n=36 e 2n=18, e viabilidade polínica de 67% e 88% para *Urochloa* sp. (tanner grass) e *U. ruziziensis*, respectivamente. Posteriormente, Ferguson e Crowder (1974) realizaram contagens cromossômicas para 16 acessos de *U. ruziziensis* e confirmaram 2n=18 cromossomos e comportamento reprodutivo da espécie como predominantemente sexual e alógamo.

A partir dos anos 80, a demanda por análises citogenéticas em *Urochloa* cresceu com a intensificação da coleta de germoplasma e estabelecimento de programas de melhoramento de *Urochloa* no Brasil (VALLE; PAGLIARINI, 2009). Nesse contexto, Penteado et al. (2000)

avaliaram por meio de citometria de fluxo a quantidade de DNA e ploidia de 435 acessos pertencentes a 13 espécies de *Urochloa*, evidenciando a existência de citótipos em seis espécies (*U. brizantha*, *U. humidicola*, *U. decumbens*, *U. jubata*, *U. nigropedata* e *U. dictyoneura*).

Bernini e Marin-Morales (2001) realizaram a primeira caracterização cariotípica em *Urochloa*, na qual foram descritos números cromossômicos, números básicos e assimetria cariotípica para 12 acessos de *U. brizantha* (2n = 36), *U. decumbens* (2n = 18, 36), *U. humidicola* (2n = 36, 42, 54), *U. jubata* (2n = 36) e *U. ruziziensis* (2n = 18). A variação de número cromossômico em *U. humidicola* revelou a existência do número básico x = 7 para espécie, diferente do número x = 9 predominante no gênero.

Mais recentemente, Tomaszewska et al. (2023) determinaram os níveis de ploidia para espécies dos complexos agâmicos ‘brizantha’ e ‘humidicola’ por meio de citometria de fluxo. Os autores encontraram acessos diploides (2n=18, 6.5%), tetraploides (2n=36, 64.8%), pentaploides (2n=45, 27.4%) e um acesso hexaploide para *U. brizantha* (2n=54). Para *U. decumbens*, foram encontrados acessos diploides (2n=18, 40.9%), tetraploides (2n=36, 56.8%) e um acesso hexaploide (2n =54). O número básico x=6 foi determinado para o complexo ‘humidicola’ e os níveis hexaploide (2n=36, 28.6%), heptaploide (2n=42, 61.1%), nonaploide (2n=54, 5.5%) e um acesso octaploide ou nonaploide com aneuploidias (2n=8x=48+2 ou 2n=9x=54-4) foram verificados para *U. humidicola* e um único acesso heptaploide (2n=42) de *U. dictyoneura* foi identificado.

Os relatos de variações de ploidia e número cromossômico em *Urochloa*, assim como a busca por genitores compatíveis para cruzamentos estimularam uma série de estudos meióticos no gênero. Morfologia polínica e irregularidades meióticas, como segregação assíncrona, desinapse, aderências cromossômicas, foram descritos predominantemente para acessos poliploides de *U. decumbens* (MENDES-BONATO et al., 2001, 2002b; RICCI et al., 2011; SOUZA et al., 2015), *U. brizantha* (MENDES-BONATO et al., 2002a; PAGLIARINI et al., 2012), *U. humidicola* (BOLDRINI et al., 2009; CALISTO et al., 2008), acessos diploides e artificialmente tetraploidizados de *U. ruziziensis* (MORAIS; SOUZA

SOBRINHO; TECHIO, 2018; PAULA et al., 2016), e híbridos interespecíficos (MENDES et al., 2006; MENDES-BONATO et al., 2006; MENDES-BONATO; PAGLIARINI; VALLE, 2006, 2007; SOUZA-KANESHIMA et al., 2010).

A utilização da técnica de FISH em estudos citogenéticos possibilitou o mapeamento físico de DNAr em espécies de *Urochloa* (Tabela 2). No complexo agâmico 'brizantha', a FISH revelou variações nos números de loci de DNAr 5S entre acessos tetraploides e pentaploides de *U. brizantha*, ao passo que os números de loci de DNAr 35S variam apenas em acessos tetraploides (AKIYAMA et al., 2010; MENDES, 2020; NANI et al., 2016; NIELEN et al., 2009). Também foram observadas variações nos loci de DNAr entre diferentes plantas e entre diferentes amostras de uma mesma planta da cultivar tetraploide Marandu (*U. brizantha*), e a presença de um par heteromórfico para comprimento cromossômico (NANI et al., 2016). Não há relatos de variação numérica em sítios de DNAr em *U. decumbens* e *U. ruziziensis* (AKIYAMA et al., 2010; MENDES, 2020; NANI et al., 2016).

O mapeamento dos loci de DNAr também foram relatados para *Urochloa humidicola* e *U. dictyoneura* (Tabela 2), que compõem o complexo agâmico 'humidicola' (DAMASCENO et al., 2023, MENDES, 2020). *Urochloa humidicola* apresenta alta instabilidade no número de loci de DNAr, com diferenças entre indivíduos no mesmo acesso e dentro da mesma planta, além de variações de ploidia, número cromossômico e ocorrência de aneuploidia (DAMASCENO et al., 2023).

2.4 Relações genômicas no complexo agâmico 'brizantha'

As técnicas FISH e GISH foram usadas concomitantemente para investigar relações genômicas entre as espécies *U. brizantha* (4x), *U. decumbens* (4x), *U. ruziziensis* (2x e 4x) e entre os híbridos H963 (*U. ruziziensis* x *U. decumbens*), H1863 (*U. ruziziensis* x *U. brizantha*) e Mulato II [(*U. ruziziensis* x *U. decumbens*) x *U. brizantha*] (PAULA; SOUZA SOBRINHO; TECHIO, 2017). Os resultados indicaram alta homologia entre os genomas de *U. brizantha*, *U. decumbens* e *U. ruziziensis*, sendo *U. decumbens* a intermediária na relação de parentesco das três espécies. Os autores confirmaram a alopoliploidia de *U. brizantha* e *U.*

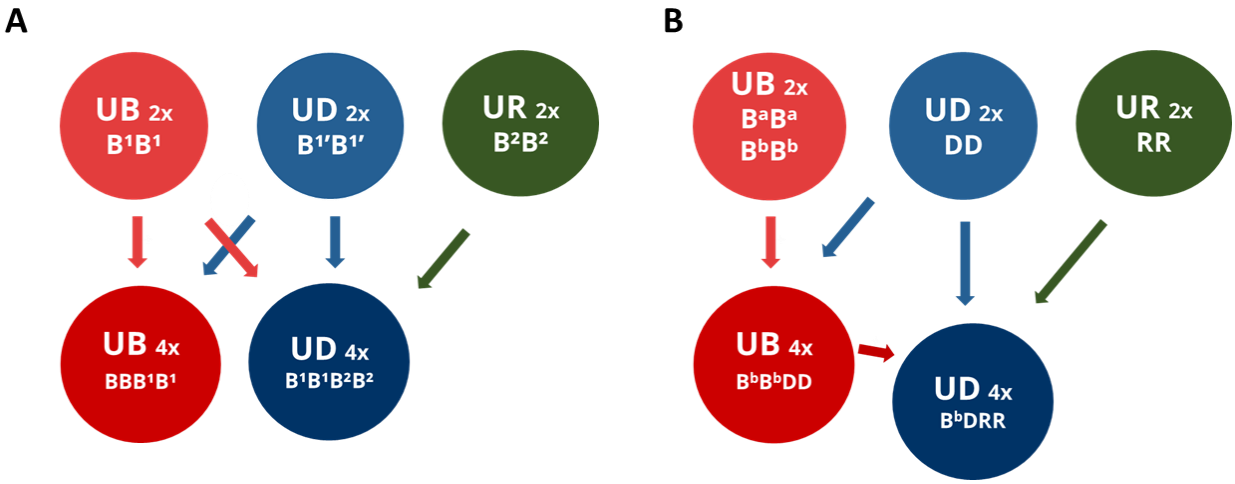
decumbens previamente evidenciada por estudos meióticos (MENDES et al., 2006; MENDES-BONATO et al., 2001, 2002a, 2002b, 2006; MENDES-BONATO; PAGLIARINI; VALLE, 2006) e propuseram a constituição genômica das duas espécies e de *U. ruziziensis*, que foram caracterizadas com os genomas BBB_1B_1, $B_1B_1B_2B_2$ e B_2B_2, respectivamente (Figura 2A). Os autores destacaram que os genomas B, B_1 e B_2 foram considerados homeólogos, com menor afinidade entre B e B_2.

Tabela 2: Número cromossômico e números de loci de DNAr em espécies de *Urochloa*

Espécie	Num. Cromossômico / Ploidia	DNAr 5S	DNAr 35S	Referência
U. arrecta	2n = 4x = 36	4	4	Moraes et al. (2021)
U. brizantha	2n = 2x = 18	2	2	Nielen et al. (2009)
	2n = 2x = 18	2	2	Mendes (2020)
	2n = 4x = 36	3	2	Nielen et al. (2009)
	2n = 4x = 36	6	4	Akiyama et al. (2010)
	2n = 4x = 36	4	4	Akiyama et al. (2010)
	2n = 4x = 36	3,5,6	2,3,4	Nani et al. (2016)
	2n = 5x = 45	5	5	Akiyama et al. (2010)
	2n = 5x = 45	8	5	Moraes et al. (2021)
U. decumbens	2n = 2x = 18	4	2	Mendes (2020)
	2n = 4x = 36	7	4	Nani et al. (2016)
U. dictyoneura	2n = 4x = 24	8	4	Mendes (2020)
U. humidicola	2n = 6x = 54	6	6	Akiyama et al. (2010)
	2n = 6x = 36	9	5	Damasceno et al. (2023)
	2n = 6x = 36+1	13	6	Damasceno et al. (2023)
	2n = 7x = 42	11	7	Damasceno et al. (2023)
	2n = 9x = 54	8	7	Damasceno et al. (2023)
U. ruziziensis	2n = 2x = 18	4	2	Akiyama et al. (2010)
	2n = 2x = 18	4	2	Nani et al. (2016)

Fonte: do Autor (2023).

Figura 2: Propostas de constituição genômica para o complexo agâmico ‘brizantha’, de acordo com Paula et al. (2017) e Correa et al. (2020) (A), e Tomaszewska et al. (2023) (B). Setas indicam possíveis relações de ancestralidade nos processos de alopoliploidização.



UB: *U. brizantha*, UD: *U. decumbens*, UR: *U. ruziziensis*.

Fonte: do autor (2023).

As relações genômicas observadas por Paula et al. (2017) para o complexo agâmico ‘brizantha’ foram reiteradas em um estudo citogenômico que incluiu acessos diploides e tetraploides de *U. brizantha* e *U. decumbens*, utilizando a GISH para investigar os possíveis genomas ancestrais envolvidos no processo de alotetraploidização dessas espécies (CORRÊA et al., 2020). A GISH recíproca entre *U. brizantha* (2x) e *U. decumbens* (2x) mostrou alta homologia genômica e apontou para uma maior variedade de classes de DNA repetitivo no genoma da primeira espécie. Os autores indicaram que ambos os acessos diploides poderiam ter sido doadoras ancestrais do genoma B^1 para as cultivares tetraploides.

Tomaszewska et al. (2023) propuseram uma segunda nomenclatura genômica para o complexo agâmico ‘brizantha’ (Figura 2B), baseada no mapeamento de sequências repetitivas e hibridizações genômicas com sondas de acessos tetraploides diferentes daqueles utilizados nos estudos de Paula et al. (2017) e Corrêa et al. (2020). Foram designadas as letras B, D e R para os genomas de *U. brizantha* (B^bB^bDD), *U. decumbens* (B^bDRR) e *U. ruziziensis* (RR), respectivamente.

Apesar das diferenças na nomenclatura genômica, similaridades foram evidenciadas nas relações genômicas do grupo, incluindo o compartilhamento de um subgenoma entre acessos tetraploides de *U. brizantha* e *U. decumbens* e entre *U. decumbens* 4x e *U. ruziziensis*. No entanto, a proposta de Tomaszewska et al. (2023) sugere uma maior complexidade no genoma da tetraploide *U. decumbens*, na qual foram observados cromossomos com marcações comuns entre *U. brizantha* e *U. decumbens* (B^b), e cromossomos com marcas exclusivas de *U. decumbens* (D) e *U. ruziziensis* (R). Esse padrão de marcas indicou que *U. decumbens* 4x teria sido originado por dois eventos de hibridação [*U. brizantha* 2x × *U. decumbens* 2x) × *U. ruziziensis 2x*], diferente de um único evento (*U. decumbens* 2x × *U. ruziziensis*), como proposto por Paula et al. (2017) e Corrêa et al. (2020).

Considerando que o presente trabalho analisou os mesmos acessos e cultivares de *U. brizantha*, *U. decumbens* e *U. ruziziensis* utilizados nos estudos de Paula et al. (2017) e Corrêa et al. (2020), todas as inferências relacionadas as relações genômicas entre essas espécies foram baseadas na nomenclatura proposta por esses autores.

REFERÊNCIAS

ADHIKARI, S. et al. Application of molecular markers in plant genome analysis: a review. **The Nucleus**, v. 60, n. 3, p. 283–297, 1 dez. 2017.

ÅGREN, J. A.; CLARK, A. G. Selfish genetic elements. **PLoS Genetics**, v. 14, n. 11, 1 nov. 2018.

AKIYAMA, Y.; YAMADA-AKIYAMA, H.; EBINA, M. Morphological diversity of chromosomes bearing ribosomal DNA loci in *Brachiaria* species. **Grassland Science**, v. 56, n. 4, p. 217–223, 2010.

ALVIM, M. J.; BOTREL, M. DE A.; XAVIER, D. F. As principais espécies de *Brachiaria* utilizadas no País. **Embrapa gado de Leite**, p. 3–6, 2002.

AMBIEL, A. C. et al. Agrupamento de acessos e cultivares de três espécies de *Brachiaria* por RAPD. **Acta Scientiarum - Agronomy**, v. 30, n. 4, p. 457–464, 2008.

AVDULOV, N. P. Karyo-systematische Untersuchungen der Familie Gramineen. **Bull Appl Bot Gen Plant Breed Suppl**, v. 34, 1931.

BELYAYEV, A. et al. Natural history of a satellite DNA family: From the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. **International Journal of Molecular Sciences**, v. 20, n. 5, 1 mar. 2019.

BENNETZEN, J. L. Transposable element contributions to plant gene and genome evolution. **Plant Molecular Biology**, v. 42, n. 1, p. 251–269, 2000.

BERNINI, C.; MARIN-MORALES, M. A. Karyotype analysis in *Brachiaria* (Poaceae) species. **Cytobios**, v. 104, p. 157–171, 2001.

BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415–420, 2015.

BOLDRINI, K. et al. Origin of a polyploid accession of *Brachiaria humidicola* (Poaceae: Panicoideae: Paniceae). **Genetics and Molecular Research**, v. 8, n. 3, p. 888–895, 2009.

BRAZ, G. T. et al. Comparative oligo-FISH mapping: An efficient and powerful methodology to reveal karyotypic and chromosomal evolution. **Genetics**, v. 208, n. 2, p. 513–523, 1 fev. 2018.

BRIITEN, R. J.; KOHNE, D. E. Repeated sequences in DNA. **Science**, v. 161, n. 3841, p. 529–540, 1968.

BROWN, W. V. Chromosome numbers of some texas grasses. **Bulletin of Torrey Botany Club**, v. 78, n. 4, p. 292–299, 1951.

CALISTO, V. et al. Desynapsis and precocious cytokinesis in *Brachiaria humidicola* (Poaceae) compromise meiotic division. **J. Genet**, v. 87, n. 4, p. 27–31, 2008.

CLAYTON, W. D.; RENVOIZE, S. A. Genera Graminum. Grasses of the World. **Development**, v. 13, n. 87, p. 389, 1986.

CORRÊA, C. T. R. et al. GISH-based comparative genomic analysis in *Urochloa* P. Beauv. **Molecular Biology Reports**, v. 47, n. 2, p. 887–896, 2020.

COYNE, A. J., ORR, H. A. **Speciation**. Sunderland: Sinauer Associates, 2004.

DAMASCENO, A.G.; FERREIRA, M.T.; SOARES, I.C.; BARRIOS, S.C.; VALLE, C.B.; TECHIO, V.H. Physical mapping of ribosomal DNA sites and genome size in polyploid series of *Urochloa humidicola* (Rendle) Morrone & Zuloaga (Poaceae). **Botany Letters**, p. 1-10, 2023.

DAWE, R. K.; HIATT, E. N. Plant neocentromeres: fast, focused, and driven. **Chromosome Research**, v. 12, n. 6, p. 655–669, 2004.

DE WET, J. M. J. Chromosome Numbers and Some Morphological Attributes of Various South African. **Wet Source: American Journal of Botany**, v. 47, n. 1, p. 44–49, 1960.

DE WET, J. M. J.; ANDERSON, L. J. Chromosome Numbers in Transvaal Grasses. **Cytologia**, v. 21, p. 1–10, 1956.

EBRAHIMZADEGAN, R.; HOUBEN, A.; MIRZAGHADERI, G. Repetitive DNA landscape in essential A and supernumerary B chromosomes of *Festuca pratensis* Huds. **Scientific Reports**, v. 9, n. 1, 1 dez. 2019.

FEDOROFF, N. Transposons and genome evolution in plants. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 13, p. 7002–7007, 2000.

FERGUSON, J. E.; CROWDER, L. V. Cytology and Breeding Behavior of *Brachiaria ruziziensis* Germain Et Evrard 1. **Crop Science**, v. 14, n. 6, p. 893–895, nov. 1974.

FREELING, M. et al. A solution to the c-value paradox and the function of junk DNA: The genome balance hypothesis. **Molecular Plant**, v. 8, n. 6, p. 899–910, 1 jun. 2015.

FU, J. et al. Identification and characterization of abundant repetitive sequences in *Allium cepa*. **Scientific Reports**, v. 9, n. 1, 1 dez. 2019.

GARCIA, S. et al. Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. **Plant Journal**, v. 89, n. 5, p. 1020–1030, 1 mar. 2017.

GARCIA, S.; KOVAŘÍK, A. Dancing together and separate again: Gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. **Heredity**, v. 111, n. 1, p. 23–33, jul. 2013.

GARRIDO-RAMOS, M. A. The Genomics of Plant Satellite DNA. Em: **Satellite DNAs in Physiology and Evolution**. p. 103–143, 2021.

HESLOP-HARRISON, J.S.; SCHMIDT, T. Plant Nuclear Genome Composition. **Encyclopedia of Life Sciences**, p. 1–8, 2012.

HESLOP-HARRISON, J. S., SCHWARZACHER, T.; LIU, Q. Polyploidy: its consequences and enabling role in plant diversification and evolution. **Annals of Botany** 131, 1-10, 2023.

HIGGINS, J. et al. Diverged subpopulations in tropical *Urochloa* (*Brachiaria*) forage species indicate a role for facultative apomixis and varying ploidy in their population structure and evolution. **Annals of Botany**, 17 nov. 2022.

IDZIAK, D.; HAZUKA, I.; POLIWCZAK, B.; WISZYNSKA, A.; WOLNY, E.; HASTEROK, R. Insight into the karyotype evolution of Brachypodium species using comparative chromosome barcoding. **PLoS One**, v. 9, n. 3, p. e93503, 2014.

JANK, L. et al. The value of improved pastures to Brazilian beef production. **Crop and Pasture Science**, v. 65, p. 1132–1137, 2014.

JIANG, J. et al. A molecular view of plant centromeres. **Trends in Plant Science**, v. 8, n. 12, p. 570–575, 2003.

JIANG, J. Fluorescence in situ hybridization in plants: recent developments and future applications. **Chromosome Research**, v. 27, n. 3, p. 153–165, 1 set. 2019.

KALITSIS, P.; CHOO, K. H. A. The evolutionary life cycle of the resilient centromere. **Chromosoma**, v. 121, n. 4, p. 327–340, ago. 2012.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. **Nature Genetics**, v. 33, n. 1, p. 102–106, 2003.

KAWABE, A.; NASUDA, S. Structure and genomic organization of centromeric repeats in Arabidopsis species. **Molecular Genetics and Genomics**, v. 272, n. 6, p. 593–602, fev. 2005.

KŘIVÁNKOVÁ, A. et al. Repetitive DNA: A Versatile Tool for Karyotyping in *Festuca* pratensis Huds. **Cytogenetic and Genome Research**, v. 151, n. 2, p. 96–105, 2017.

LANG, T. et al. Genome-wide distribution of novel Ta-3A1 mini-satellite repeats and its use for chromosome identification in wheat and related species. **Agronomy**, v. 9, n. 2, 29 jan. 2019.

LERMONTOVA, I. et al. Centromeric chromatin and its dynamics in plants. **Plant Journal**, v. 83, n. 1, p. 4–17, 1 jul. 2015.

LI, Y. et al. Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. **Plant Journal**, v. 93, n. 6, p. 1088–1101, 1 mar. 2018.

LIEHR, T. Repetitive elements in humans. **International Journal of Molecular Sciences**MDPI AG, 2 fev. 2021.

LING, H. Q. et al. Draft genome of the wheat A-genome progenitor *Triticum urartu*. **Nature**, v. 496, n. 7443, p. 87–90, 2013.

LIU, Q. et al. The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. **BMC Plant Biology**, v. 19, n. 1, 30 maio 2019.

LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The repetitive DNA content of eukaryotic genomes. Em: **Repetitive DNA**. v. 7p. 1–28, 2012.

LUTTS, S.; NDIKUMANA, J.; LOUANT, B. P. Fertility of *Brachiaria ruziziensis* in interspecific crosses with *Brachiaria decumbens* and *Brachiaria brizantha*: meiotic behavior, pollen viability and seed set. **Euphytica**, v. 57, p. 267–274, 1991.

MATYÁŠEK, R. et al. Intragenomic heterogeneity of intergenic ribosomal DNA spacers in *Cucurbita moschata* is determined by DNA minisatellites with variable potential to form non-canonical DNA conformations. **DNA Research**, v. 26, n. 3, p. 273–286, 1 jun. 2019.

MCCLINTOCK, B. Mutable loci in maize. **Carnegie Inst Wash Year Book 47**, p. 155–169, 1948.

MCKINLEY, K. L.; CHEESEMAN, I. M. The molecular basis for centromere identity and function. **Nature Reviews Molecular Cell Biology**, v. 17, n. 1, p. 16–29, 1 jan. 2016.

MELTERS, D. P.; BRADNAM, K. R.; YOUNG, H. A.; TELIS, N.; MAY, M. R.; RUBY, J. G.; SEBRA, R.; PELUSO, R.; EID, J.; RANK, D.; GARCIA, J.F. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome biology**, v. 14, n. 1, p. 1-20, 2013.

MENDES, D. V. et al. Cytological evidence of natural hybridization in *Brachiaria brizantha* Stapf (Gramineae). **Botanical Journal of the Linnean Society**, v. 150, p. 441–446, 2006.

MENDES, L. M. Análise cariotípica em acessos diploides de *Urochloa brizantha* e *Urochloa decumbens* e poliploides de *Urochloa dictyoneura*. Master thesis. **Federal University of Lavras**, 2020.

MENDES-BONATO, A. B. et al. Meiotic instability in invader plants of signal grass *Brachiaria decumbens* Stapf (Gramineae). **Acta Scientiarum**, v. 23, n. 2, p. 619–625, 2001.

MENDES-BONATO, A. B. et al. Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). **Euphytica**, v. 125, p. 419–425, 2002a.

MENDES-BONATO, A. B. et al. Unusual cytological patterns of microsporogenesis in *Brachiaria decumbens*: Abnormalities in spindle and defective cytokinesis causing precocious cellularization. **Cell Biology International**, v. 26, n. 7, p. 641–646, 2002b.

MENDES-BONATO, A. B. et al. Cytogenetic evidence for genome elimination during microsporogenesis in interspecific hybrid between *Brachiaria ruziziensis* and *B. brizantha* (Poaceae). **Genetics and Mole**, v. 29, n. 4, p. 711–714, 2006.

MENDES-BONATO, A. B.; PAGLIARINI, M. S.; VALLE, C. B. Abnormal spindle orientation during microsporogenesis in an interspecific *Brachiaria* (Gramineae) hybrid. **Genetics and Molecular Biology**, v. 29, n. 1, p. 122–125, 2006.

MENDES-BONATO, A. B.; PAGLIARINI, M. S.; VALLE, C. B. DO. Meiotic Arrest Compromises Pollen Fertility in an Interspecific Hybrid between *Brachiaria ruziziensis* x *Brachiaria decumbens* (Poaceae: Paniceae). **Brazilian Archives of Biology and Technology**, v. 50, p. 831–837, 2007.

MILLER, J. T. et al. Retrotransposon-Related DNA Sequences in the Centromeres of Grass Chromosomes. **Genetics**, v. 150, n. 4, p. 1615–1623, 1998.

MOFFETT, A. A.; HURCOMBE, R. Chromosome numbers of South African grasses. **Heredity**, v. 3, p. 369–373, 1949.

MORAIS, L. C.; SOUZA SOBRINHO, F.; TECHIO, V. H. Comparative microsporogenesis between diploid and tetraploid plants of *Brachiaria ruziziensis* and their progenies. **South African Journal of Botany**, v. 119, p. 258–264, 1 nov. 2018.

MORRONE, O.; ZULOAGA, F. O. Revision de las especies sudamericanas nativas e introducidas de los generos *Brachiaria* y *Urochloa* (Poaceae: Panicoideae: Paniceae). **Darwiniana**, v. 31, n. 1/4, p. 43–109, 1992.

NANI, T. F. et al. Physical map of repetitive DNA sites in *Brachiaria* spp.: Intravarietal and interspecific polymorphisms. **Crop Science**, v. 56, n. 4, p. 1769–1783, 2016.

NASH, G. V. *Brachiaria*. Em: SMALL, J. K. (Ed.). **Flora of southeastern United States**. New York: The New Era Printing Company, 1903.

NEUMANN, P. et al. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA**, v. 10, n. 1, p. 1–17, 2019.

NIELEN, S. et al. Physical mapping of rDNA genes corroborates allopolyploid origin in apomictic Brachiaria brizantha. **Sexual Plant Reproduction**, v. 23, n. 1, p. 45–51, 2009.

NOVÁK, P. et al. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792–793, 2013.

OLIVEIRA, L. C.; TORRES, G. A. Plant centromeres: genetics, epigenetics and evolution. **Molecular Biology Reports,** v.45, p. 1491-1497. 2018PAGLIARINI, M. S. et al.

Microsporogenesis in *Brachiaria brizantha* (Poaceae) as a selection tool for breeding. **Genetics and Molecular Research**, v. 11, n. 2, p. 1309–1318, 2012.

PAULA, C. M. P. et al. Microsporogenesis analysis validates the use of artificially tetraploidized *Brachiaria ruziziensis* in breeding programs. **Genetics and Molecular Research**, v. 15, n. 3, 19 set. 2016.

PAULA, C. M. P. DE; SOUZA SOBRINHO, F.; TECHIO, V. H. Genomic constitution and relationship in *Urochloa* (Poaceae) species and hybrids. **Crop Science**, v. 57, n. 5, p. 2605–2616, 2017.

PECH, M.; IGO-KEMENES, T.; ZACHAU, H. G. Nucleotide sequence of a highly repetitive component of sat DNA. **Nucleic Acids Research**, v. 7, n. 2, p. 417–432, 1979.

PENTEADO, M. I. DE O. et al. Determinação de poliploidia e avaliação da quantidade de DNA total em diferentes espécies de gênero *Brachiaria*. **Boletim de Pesquisa Embrapa**, v. 11, 2000.

PESSOA-FILHO, M.; MARTINS, A. M.; FERREIRA, M. E. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes. **BMC Genomics**, v. 18, n. 1, p. 516, 2017.

POWO. Plants of the World Online**.** Facilitated by **the Royal Botanic Garden, Kew**, 2023.

PRESTING, G. G. et al. A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. **The Plant Journal**, 1998.

PROKOPOWICH, C. D.; GREGORY, T. R.; CREASE, T. J. The correlation between rDNA copy number and genome size in eukaryotes. **Genome**, v. 46, n. 1, p. 48–50, 2003.

QI, P., EUDY, D.; SCHNABLE, J.C.; SCHMUTZ, J.; RAYMER, P.L.; DEVOS, K.M. High Density Genetic Maps of Seashore Paspalum Using Genotyping-By-Sequencing and Their Relationship to The Sorghum Bicolor Genome. **Sci Rep**. v. 21, n. 9, 12183, 2019

RENVOIZE, S. A.; CLAYTON, W. B.; KABUYE, C. H. S. Morphology, taxonomy and natural distribution of *Brachiaria* (Trin.) Griseb. Em: ***Brachiaria*: biology, agronomy and improvement**. p. 1–15, 1996.

RENVOIZE, S.; MAASS, B. *Brachiaria*. A report to CIAT, Colombia, on the species and specimens held in the germplasm collection. **Kew. Royal Botanical Gardens.**, 1993.

RICCI, G. C. L. et al. Chromosome numbers and meiotic analysis in the pre-breeding of *Brachiaria decumbens* (Poaceae). **Indian Academy of Sciences**, v. 90, n. 2, p. 289–294, 2011.

RIED, T.; SCHRÖCK, E.; NING, Y.; WIENBERG, J. Chromosome painting: a useful art. **Human molecular genetics**, v. 7, n. 10, p. 1619-1626, 1998.

ROCCHI, M. et al. Centromere repositioning in mammals. **Heredity**, 2012.

ROSSELLÓ, J. A.; MARAVILLA, A. J.; ROSATO, M. The Nuclear 35S rDNA World in Plant Systematics and Evolution: A Primer of Cautions and Common Misconceptions in Cytogenetic Studies. **Frontiers in Plant Science.** Frontiers Media S.A, 2022.

SALARIATO, D. L. et al. Molecular phylogeny of the subtribe Melinidinae (Poaceae: Panicoideae: Paniceae) and evolutionary trends in the homogenization of inflorescences. **Molecular Phylogenetics and Evolution**, v. 56, n. 1, p. 355–369, 2010.

SANTOS, F. C. et al. Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. **Chromosome Research**, v. 23, n. 3, p. 571–582, 2015.

SCHNABLE, P. S. et al. The B73 maize genome: Complexity, diversity, and dynamics. **Science**, v. 326, n. 5956, p. 1112–1115, 2009.

SHATSKIKH, A. S. et al. Functional Significance of Satellite DNAs: Insights From Drosophila. **Frontiers in Cell and Developmental Biology.** Frontiers Media S.A., 2020.

SOTOMAYOR-RIOS, A.; SCHANK, S. C.; WOODBURY, R. Cytology and Taxonomic Description of Two Brachiarias (Congograss and Tannergrass). **Journal of Agriculture of the University of Puerto Rico**, v. 54, p. 390–400, 1970.

SOUZA, V. F. et al. Meiotic behavior of *Brachiaria* decumbens hybrids. **Genetics and Molecular Research**, v. 14, n. 4, p. 12855–12865, 21 out. 2015.

SOUZA-KANESHIMA, A. M. et al. Meiotic behaviour in the first interspecific hybrids between *Brachiaria brizantha* and *Brachiaria decumbens*. **Plant Breeding**, v. 129, n. 2, p. 186–191, abr. 2010.

STEIN, L. D. et al. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. **PLoS Biology**, v. 1, n. 2, 2003.

TANG, S. J. Potential role of phase separation of repetitive DNA in chromosomal organization. **Genes**, v. 8, n. 10, 18 out. 2017.

TOMASZEWSKA, P. et al. Flow cytometry-based determination of ploidy from leaf specimens in genomically complex collections of the tropical forage grass *Urochloa*. **Genes**, v. 12, n. 7, 2021.

TOMASZEWSKA, P. et al. 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**, 2023.

TORRES GONZÁLEZ, A. M.; MORTON, C. M. Molecular and morphological phylogenetic analysis of *Brachiaria* and *Urochloa* (Poaceae). **Molecular Phylogenetics and Evolution**, v. 37, n. 1, p. 36–44, 2005.

TRIVIÑO, N. J. et al. Genetic Diversity and Population Structure of *Brachiaria* Species and Breeding Populations. **Crop Science**, v. 57, n. 5, p. 2633–2644, 2017a.

TRIVIÑO, N. J. et al. Genetic Diversity and Population Structure of *Brachiaria* Species and Breeding Populations. **Crop Science**, v. 57, n. 5, p. 2633–2644, 2017b.

VALLE, C. B.; PAGLIARINI, M. S. Biology, Cytogenetics, and Breeding of *Brachiaria*. Em: **Genetic resources, chromosome engineering, and crop improvement**. v. 2p. 103–143, 2009.

VELDKAMP, J. F. Miscellaneous notes on mainly Southeast Asian Gramineae. **Reinwardtia**, v. 12, p. 135–140, 2004.

VIEIRA, M. L. C. et al. Microsatellite markers: What they mean and why they are so useful. **Genetics and Molecular Biology**, v. 39, n. 3, p. 312–328, 2016.

WAMINAL, N. E. et al. Characterization of chromosome-specific microsatellite repeats and telomere repeats based on low coverage whole genome sequence reads in *Panax ginseng*. **Plant Breeding and Biotechnology**, v. 6, n. 1, p. 74–81, 2018.

WELLS, J. N.; FESCHOTTE, C. **A Field Guide to Eukaryotic Transposable Elements**. **Annual Review of Genetics.** Annual Reviews Inc., 23 nov. 2020.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews**, v. 8, n. 12, p. 973, 2007.

WORTHINGTON, M. et al. A new genome allows the identification of genes associated with natural variation in aluminium tolerance in *Brachiaria* grasses. **Journal of Experimental Botany**, v. 72, n. 2, p. 302–319, 2021.

XI, W. et al. ND-FISH-positive oligonucleotide probes for detecting specific segments of rye (*Secale cereale* L.) chromosomes and new tandem repeats in rye. **Crop Journal**, v. 8, n. 2, p. 171–181, 2020.

ZHANG, H., et al. Boom-bust turnovers of megabase-sized centromeric DNA in Solanum species: rapid evolution of DNA sequences associated with centromeres. **The Plant Cell**, 26(4), pp.1436-1447, 2014.

ZOSHCHUK, N. V; BADAEVA, E. D.; ZELENIN, A. V. History of Modern Chromosomal Analysis. Differential Staining of Plant Chromosomes. **Russian Journal of Developmental Biology**, v. 34, n. 1, p. 112, 2003.

SEGUNDA PARTE – ARTIGOS

ARTIGO 1: Repetitive DNA landscape in the 'brizantha' agamic complex of *Urochloa*

Norma NBR 6022 (ABNT 2018)

ABSTRACT

Urochloa P. Beauv. [syn. *Brachiaria* (Trin.) Griseb.] is a genus of tropical grasses largely used as forages for their good nutritional value and adaptability. *Urochloa brizantha* (Hoschst. ex A. Rich) R.D. Webster, *U. decumbens* (Stapf) R.D. Webster and *U. ruziziensis* (R. Germ. & C.M. Evrard) Morrone & Zuloaga are among the most economically important species, frequently used for genetic crosses in breeding programs. The three species are very closely related, forming an agamic complex where the predominantly apomictic tetraploids *U. brizantha* and *U. decumbens* can pollinate the sexual diploid *U. ruziziensis*. Cytogenomic approaches have detected allopolyploidization events in *U. brizantha* and *U. decumbens*, but the genomic relationships and ancestral progenitors still need to be elucidated. Therefore, the present study aims to characterize and compare the repetitive DNA landscape of the *Urochloa* agamic complex species, including diploid and tetraploid cytotypes of *U. brizantha* and *U. decumbens*, and *U. ruziziensis*. Our results show that repetitive DNA comprises 56-65% of *Urochloa* genomes and *Gypsy* retrotransposons is the most frequent repeat superfamily. The most representative lineages of transposable elements were similar in the three species, while satellite DNA (satDNA) families varied across and within species. *Urochloa brizantha* presented a more diverse satDNA profile, with a larger number of distinct monomers, and species-specific repeats. The distribution of satellite families provided more insight on the genomic constitution previously determined for the allotetraploids *U. brizantha* (BBB^1B^1) and *U. decumbens* ($B^1B^1B^2B^2$), pointing to *U. brizantha* 2x as a putative ancestral parent in the hybridization event of its tetraploid counterpart.

Keywords: Polyploidy, *Brachiaria*, retrotransposon, satellite DNA, tandem repeat, forage grasses.

1 INTRODUCTION

Urochloa [syn. *Brachiaria* (Trin.) Griseb.] is the most economically relevant genus of tropical forages, being extensively used for livestock feeding in South American countries (JANK et al., 2014). The genus includes 88 species, distributed throughout tropical and subtropical regions worldwide (POWO, 2023). *Urochloa* species display wide variation of chromosome numbers (2n =14 to 2n = 90), ploidy levels (x = 2, 4, 5, 6, 8, 9 and 10), and basic chromosome numbers (n = 6, 7 and 9) (VALLE; PAGLIARINI, 2009).

The species *U. brizantha*, *U. decumbens* and *U. ruziziensis* include wild and cultivated plants that represent valuable germplasm for breeding programs. *Urochoa brizantha* and *U. decumbens* are predominantly apomictic polyploids, but sexual reproduction and hybridization may occur as they form an agamic complex with the sexual diploid *U. ruziziensis*, known as the 'brizantha' complex (LUTTS; NDIKUMANA; LOUANT, 1991; RENVOIZE; MAASS, 1993).

The existence of cytotypes in *U. brizantha* (2n = 18, 36, 45, 54) and *U. decumbens* (2n = 18, 36, 45) germplasm limits the combinations for genetic crosses since ploidy differences often lead to unfertile progeny (NITTHAISONG et al., 2019). Nonetheless, stable hybrids have been produced involving tetraploids forms *U. brizantha* and *U. decumbens* using the artificially tetraploidized *U. ruziziensis* as pollen donor (TIMBÓ et al., 2014).

Meiotic studies indicated an allopolyploid origin for tetraploid cultivars of *U. brizantha* and *U. decumbens* (MENDES et al., 2006; MENDES-BONATO et al., 2001, 2002, 2006), which was later confirmed by a cytogenomic approach using the genomic *in situ* hybridization (GISH) technique (PAULA; SOUZA SOBRINHO; TECHIO, 2017). The genomic probing allowed the determination of the genomic constitutions BBB^1B^1 and $B^1B^1B^2B^2$ for the allotetraploids *U. brizantha* and *U. decumbens*, respectively, and B^2B^2 for the diploid *U. ruziziensis*, evidencing shared (sub)genomes among the species (PAULA; SOUZA SOBRINHO; TECHIO, 2017). A second comparative GISH study investigated the ancestry of the allotetraploid subgenomes by comparing them with diploid cytotypes of *U. brizantha* and *U. decumbens*, which exhibited high homology levels with each other and with

the tetraploid genomes, suggesting that they are related to B^1 genome, being potential ancestors in the allotetraploidy events (CORRÊA et al., 2020)

The genomic relationship of the 'brizantha' complex was further investigated using genome-specific probes and a new nomenclature was proposed for tetraploid accessions of *U. brizantha* (B^bB^bDD), *U. decumbens* (B^bDRR), and *U. ruziziensis* (RR) (Tomaszewska et al., 2023). Both nomenclatures indicate that *U. ruziziensis* and *U. decumbens* 2x would be ancestors in the allotetraploidization of *U. decumbens*, although the proposal of Paula et al. (2017) and Corrêa et al. (2020) infered one hybridization event between *U. decumbens* 2x and *U. ruziziensis*, while Tomaszewska et al. (2023) indicated that *U. decumbens* 4x would have originated from two hybridization events [(*U. brizantha* 2x x *U. decumbens* 2x) x *U. ruziziensis*].

Although the use of genomic probes can be very efficient to highlight homologous regions of different genomes, they cannot determine the specific DNA sequences shared among them. Nonetheless, it is implied that most of the genomic differences perceived at the chromosome level is associated with the repetitive fraction of the DNA, as it can represent up to 92% of plant genomes (FU et al., 2019). Therefore, the characterization of the repetitive portion is useful to highlight overall genomic differences and similarities among closely related species.

Repetitive sequences can also be physically mapped to chromosomes via fluorescent *in situ* hybridization (FISH) to enhance the karyotyping process. In the ‘brizantha’ complex, the ribosomal genes 35S and 5S (MORAES et al., 2021; NANI et al., 2016), low-copy genes (NANI et al., 2018), retrotransposon lineages *CRM* (NANI et al., 2016), *Athila, Del*, and *Tat*, isolated from transcriptome data (SANTOS et al., 2015), as well as 50-mer and tandem repeats (TOMASZEWSKA et al. 2023) have been used as probes on the agamic complex species. However, an in-depth characterization of the repeat profile on multiple species is still lacking for *Urochloa*.

Next-generation sequencing and the development of the RepeatExplorer pipeline (NOVÁK et al., 2013) have facilitated the study of repetitive DNA by allowing the

identification and quantification of repeat families in species without a reference genome. Once characterized, these repeat families can be used for several purposes, such as understanding genomic relationships between related species (GAIERO et al., 2019), tracking genomes in natural and artificial hybrids (HEITKAM et al., 2020) and investigating evolutionary patterns of repetitive DNA classes (RUIZ-RUANO et al., 2016).

The aim of the study was to characterize and compare the repetitive DNA landscape of the '*brizantha'* agamic complex of *Urochloa*, including diploid and tetraploid cytotypes of *U. brizantha* and *U. decumbens*, and *U. ruziziensis*. We identified and compared major repeat classes and evaluated if repeat lineages are linked to specific genomes and subgenomes in these species.

2 MATERIAL AND METHODS

2.1 Plant Material

Plants from two diploid accessions and two tetraploid cultivars of *Urochloa brizantha* and *Urochloa decumbens*, donated by EMBRAPA Gado de Corte, Campo Grande – MS, Brazil and EMBRAPA Gado de Leite, Juiz de Fora – MG, Brazil (Table 1) were cultivated in greenhouse environment for DNA extraction.

Table 1: *Urochloa* species, accessions, chromosome number and genome size.

Species	Accession	Chromosome number	Genome size	Source
U. brizantha	B105	2n = 2x = 18[1]	870 Mbp[1]	EMBRAPA Gado de Corte
U. brizantha	cv. Marandu	2n = 4x = 36[2]	1720 Mbp[3]	EMBRAPA Gado de Leite
U. decumbens	D04	2n = 2x = 18[1]	704 Mbp[1]	EMBRAPA Gado de Corte
U. decumbens	cv. Basilisk	2n = 4x = 36[2]	1855 Mbp[3]	EMBRAPA Gado de Leite
U. ruziziensis	CIAT 26162	2n = 2x = 18[4]	732 Mbp[4]	CIAT

[1]Corrêa et al. (2020), [2]Paula et al. (2017), [3]Penteado et al. (2000), [4]Worthington et al. (2021).

Source: from the Author (2023).

2.2 DNA Extraction and Low-coverage Sequencing

Genomic DNA was extracted from fresh young leaves using the DNEasy® Plant Mini Kit (QIAGEN Inc). The integrity of the extracted DNA was assessed on an agarose gel 1%. The DNA was quantified and evaluated for purity (260/280 absorbance) in a Nanovue spectrophotometer (Biochrom Ltd). We also included sequencing reads from the *U. ruziziensis* accession CIAT 26162, obtained by (WORTHINGTON et al., 2021), available online at the European Nucleotide Archive.

The genomic libraries were constructed using the Nextera™ DNA Flex Library Prep kit, with paired-end fragments of 2x300 bp for all samples. The genomic sequencing was performed by the Oswaldo Cruz Foundation (Fiocruz), René Rachou Institute, Belo Horizonte – MG, Brazil, using the Illumina HiSeq 2500®.

2.3 Read processing and Clustering Analysis

The pre-processing of reads and clustering analysis for identification and characterization of DNA repetitive families were performed for each accession individually using the RepeatExplorer pipeline (NOVÁK et al., 2013), implemented on the galaxy server. The reads from *U. brizantha* and *U. decumbens* were trimmed and quantity-filtered to same length (200 bp). The reads from *U. ruziziensis* had already been trimmed to 100 bp by the original authors. Low quality reads, with 95% of bases below cut-off phred score of 10 were discarded. The sequences were sampled to reach a genomic coverage of 0.1x, so that the number of reads analyzed were proportional to the genome sizes (Table 1). After processing, the number of reads inputted were 435,000, 860,000, 352,000, 927,500 and 732,000 for *U. brizantha* accession B105, *U. brizantha* cv. Marandu, *U. decumbens* accession D04, *U. decumbens* cv. Basilisk, and *U. ruziziensis*, respectively.

The graph-based clustering analysis was performed with the default parameters of 90% similarity over 55% of read length. Overlapping reads were clustered, with each cluster representing a different lineage of repetitive DNA. The genomic proportion of each cluster was estimated based on the total number of reads analyzed, after the removal of plastid and

mitochondrial DNA. The most representative clusters, containing ≥ 0.01% of the total of reads, were automatically annotated (Table S1) based on the REXdb database of plant sequences (NEUMANN et al., 2019). Clusters that were not automatically classified were examined for the graph shape and compared against the NCBI database and GIRI RepBase databases, using the BLAST tool (Table S1).

2.4 Satellite Analysis

The TAREAN (Tandem Repeat Analyzer) tool (NOVÁK et al., 2017) was used for the identification of satellite DNA. The similarity and identity of satellites monomers from the different samples were compared via local alignment (Table S2), using the Repeat Masker platform (SMIT; HUBLEY; GREEN, 2020). The satellites were classified based on Ruiz-Ruano et al. (2016) classification. Satellites with >95% identity were considered the same variant, identity between 80% and 95% were considered different variants within the same family and identity < 80% were considered different variants. Satellite families were numbered according to their relative abundance in the genomes and different letters differentiate variants within a family.

A phylogenetic tree was constructed to infer the evolutionary history of the satellite families. All monomers from the *Urochloa* species and a satellite sequence from *Cenchrus americanus* (UroSat-1A; KAMM et al., 1994) a related grass species used as outgroup, were aligned using MUSCLE algorithm and the tree was created using the Maximum Likelihood method, with 500 bootstrap replications, in the MEGA X software (KUMAR et al., 2018).

The satellite monomers were aligned with repetitive sequences previously identified for *Urochloa* species, including *U. brizantha*, *U. decumbens*. *U. humidicola*, *U. maxima* and *U. ruziziensis* (TOMASZEWSKA et al. 2023), and with repeats from other grass species related to *Urochloa* (ANANIEV; PHILLIPS; RINES, 1998; KAMM et al., 1994).

3 RESULTS

3.1 Repeats Proportion

The estimated repetitive DNA proportion ranged from 56.26% in *U. brizantha* 2x to 65.52% in *U. brizantha* 4x (Table 2). No significant correlation (R = 0.769, *P* = 0.128) was found between the proportion of repetitive DNA and genome size. In the diploid cytotype of *U. brizantha*, 250,665 reads were classified as repetitive DNA, grouped into 35,764 (clusters, from which 316 were considered the most representative (top clusters) for containing at least 0.01% of the total of sequences each. The tetraploid cytotype displayed 573,645 repeats, grouped into 69,042 clusters, with 263 top clusters. In the diploid cytotype of *U. decumbens*, 224,715 reads represented repetitive DNA, grouped into 24,208 clusters and 267 top clusters. The tetraploid cytotype had 610,784 repetitive reads, in 71,403 clusters and 259 top clusters. In *U. ruziziensis*, 442,300 reads were repeats grouped into 36,854 clusters and 249 top clusters.

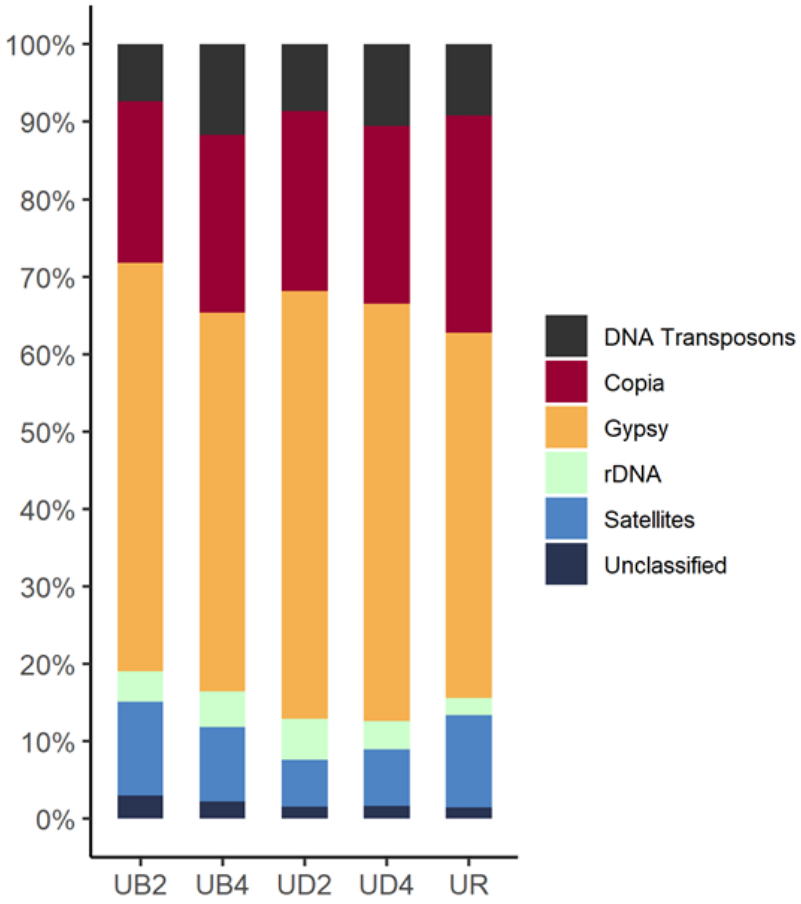
3.2 Transposable Elements

The annotated proportions of repetitive DNA of the analyzed species are displayed in Table 2 and Figure 1 and 2. Long-terminal repeat (LTR) retrotransposons are the most common repetitive element in the *Urochloa* genomes (23.5–34%). The most frequent LTR superfamily was *Ty3/Gypsy* (16.8-23.9%) and the percentages of the different repeat families within *Ty3/Gypsy* varied among the three species (Figure 2A). In *U. brizantha* and *U. decumbens*, *Athila* was the most frequent *Gypsy* lineage, followed by *Retand* and *Tekay*. In *U. ruziziensis*, *Retand* and *Athila* had similarly high proportions.

Urochloa brizantha had a larger fraction of *Ogre* elements, while *U. decumbens* 2x and *U. ruziziensis* 2x had noticeably lower proportions. CRMs (Centromeric Repeats, originally identified in Maize) were also present in all genomes, being more frequent in the diploids.

The second most abundant superfamily of LTRs was *Ty1/Copia* (6.6-11%). Among *Ty1/Copia* elements (Figure 2B), Sireviruses (SIRE) were much more frequent than other lineages, being especially high in *U. ruziziensis*. The percentages of *Bianca* and *Angela* retrotransposons were higher in the tetraploid cytotypes of both species. The tetraploid cytotype of *U. decumbens* had six-fold more *Tork* elements than the other samples.

Figure 1: Repetitive fraction of *Urochloa* genomes. UB2: *U. brizantha*, diploid; UB4: *U. brizantha*, tetraploid; UD2: *U. decumbens*, diploid; UD4: *U. decumbens*, tetraploid. UR: *U. ruziziensis.*



Source: from the Author (2023).

DNA transposons represented 2.4–4.7% of the genomes analyzed (Table 2). *CACTA* was the most common transposon superfamily in all *Urochloa* species, seemingly higher in the tetraploids, followed by *Harbinger* and *Mutator* (Figure 2C). The number of *Helitron* and

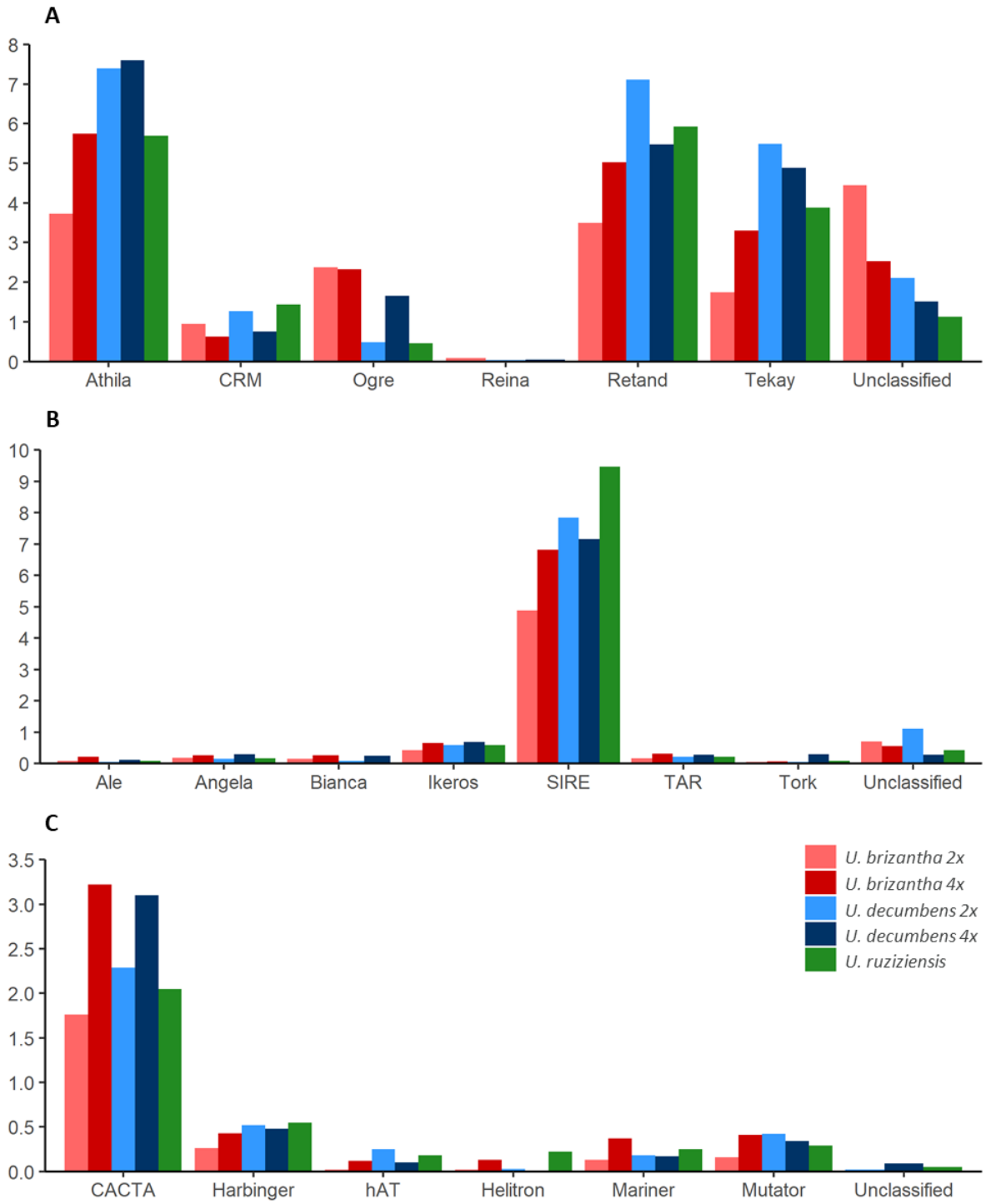
Mariner repeats were especially larger in *U. brizantha* 4x and *U. ruziziensis 2x*, while *hAT* elements were more frequent in *U. decumbens* 4x and *U. ruziziensis 2x*.

Table 2: Genomic proportions (%) of repeat classes and families in *U. brizantha* diploid (UB2) and tetraploid (UB4), *U. decumbens* diploid (UD2) and tetraploid (UD4), and *U. ruziziensis* (UR).

		UB2	UB4	UD2	UD4	UR
Class I Retroelements	Lineage					
LTR	(unclassified)	1.69	2.52	0.41	1.27	3.25
Ty1/Copia	(unclassified)	0.71	0.55	1.11	0.28	0.42
	Ale	0.08	0.21	0.05	0.11	0.08
	Angela	0.19	0.26	0.15	0.29	0.17
	Bianca	0.15	0.27	0.08	0.25	0.00
	Ikeros	0.42	0.66	0.59	0.69	0.59
	SIRE	4.88	6.82	7.84	7.16	9.47
	TAR	0.16	0.32	0.21	0.28	0.21
	Tork	0.06	0.07	0.05	0.30	0.08
Total Ty1/Copia		6.65	9.16	10.08	9.36	11.02
Ty3/Gypsy	(unclassified)	4.45	2.53	2.10	1.51	1.13
	Athila	3.73	5.74	7.39	7.60	5.70
	CRM	0.95	0.63	1.27	0.76	1.44
	Ogre	2.37	2.33	0.48	1.66	0.46
	Reina	0.08	0.00	0.03	0.05	0.00
	Retand	3.50	5.03	7.11	5.48	5.93
	Tekay	1.74	3.30	5.49	4.89	3.88
Total Ty3/Gypsy		16.82	19.56	23.87	21.95	18.54
Total LTR		25.16	31.24	34.36	32.58	32.81
Other	LINE	0.05	0.12	0.09	0.45	0.10
	SINE	0.04	0.05	0.00	0.04	0.07
	MITE	0.02	0.03	0.00	0.02	0.00
	Caulimovirus	0.05	0.13	0.00	0.02	0.00
Total Retrotransposons		25.32	31.57	34.45	33.11	32.98
Class II	(unclassified)	0.00	0.00	0.03	0.09	0.05
DNA transposons	CACTA	1.76	3.22	2.29	3.10	2.05
	hAT	0.02	0.12	0.25	0.10	0.18
	Mutator	0.16	0.41	0.42	0.34	0.29
	Harbinger	0.26	0.43	0.52	0.48	0.55
	Mariner	0.13	0.37	0.18	0.17	0.25
	Helitron	0.02	0.13	0.02	0.01	0.22
Total DNA Transposons		2.35	4.68	3.71	4.29	3.59
Tandem repeats						
	Satellites	3.86	3.83	2.62	2.97	4.65
	rDNA	1.25	1.85	2.30	1.50	0.90
Total tandem repeats		5.11	5.68	4.92	4.47	5.55
Annotated repeats		32.78	41.93	43.08	41.87	42.12
Total repetitive DNA		56.26	65.52	62.48	64.78	57.75

Source: from the Author (2023).

Figure 2: Proportion of *Gypsy* retrotransposons (A), *Copia* retrotransposons (B) and DNA transposons (C) in *Urochloa* genomes



Source: from the Author (2023).

3.3 Satellite DNA

Tandem repeats are mainly represented by ribosomal genes (1.2–2.3%) and satellite DNA (2.6–4.6%; Table 2). Eleven satellite families were identified based on local alignments, three of which (UroSat-1, UroSat-2 and UroSat-6) exhibited different variants. The lengths of satellite monomers ranged from 78 to 744, with 158 and 360-380 being the most common (Table S3). The A-T content was similar within the satellite families, with an average of 56.6% (Table S3).

Urochloa brizantha had a higher number and proportion of satellites than *U. decumbens* (Table 2, Figure 3A). In *U. brizantha*, all 11 satellite families identified were present, with 14 variants present in both cytotypes, and 10 variants shared between them. Four exclusive satellites were found in *U. brizantha* 2x, while one was found in *U. brizantha* 4x. In *U. decumbens*, only three satellite families were identified in the diploid cytotype (UroSat-1, UroSat-3 and UroSat-8), all shared with *U. decumbens* 4x, which had a total of seven satellites. *Urochloa ruziziensis* exhibited the smallest number and largest proportion of satellites among the three species. The satellites found in *U. ruziziensis* (UroSat-1a, UroSat-3, UroSat-8) were the same ones found in *U. decumbens* 2x, which were shared among the three species.

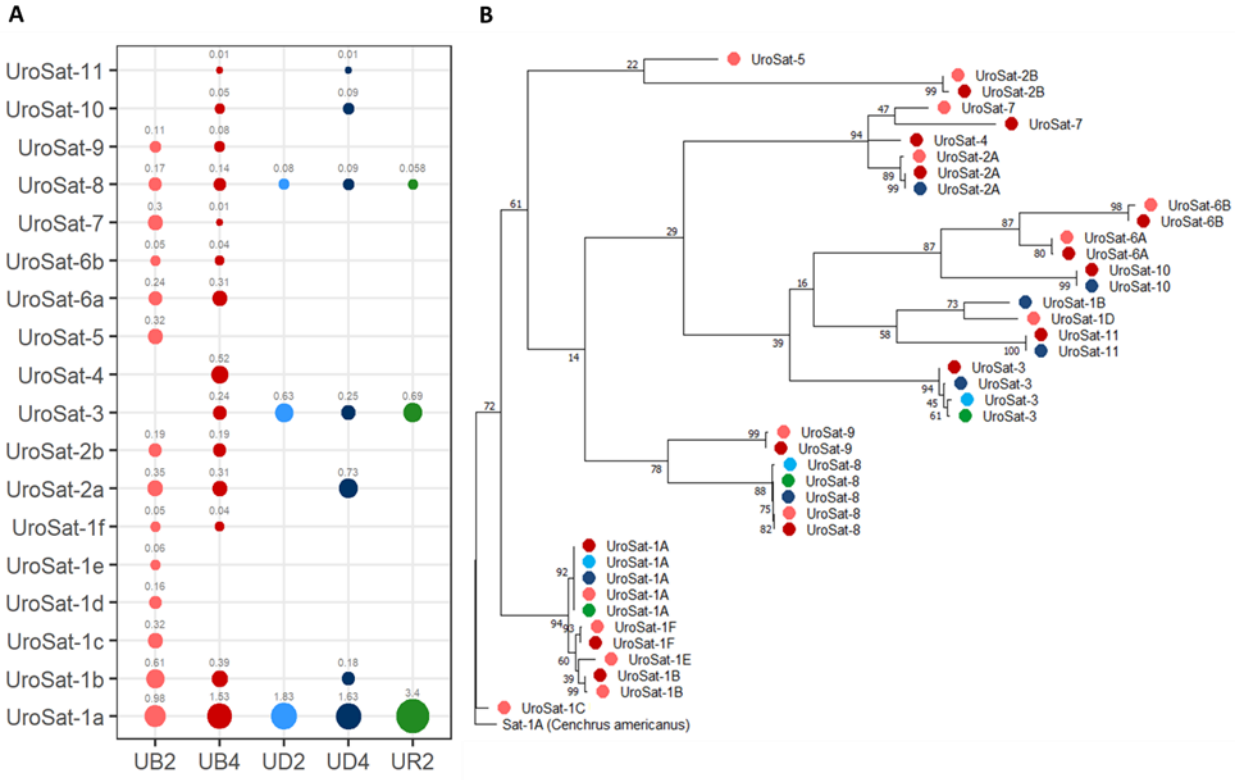
Satellite variants were only observed for UroSat-1, UroSat-2 and UroSat-6. UroSat-1 was the most diverse family of satDNA, including six variants. Four of these variants (UroSat-1c, UroSat-1d, UroSat-1e, and UroSat-1f) were unique to *U. brizantha*, three of which were exclusive to diploid cytotype.

The phylogenetic analysis of the satellite families (Figure 3B) demonstrates that most of the variants have diversified before speciation events. The most conserved variant (UroSat-1a) was positioned at one of the most basal clades. Clades differentiated at species level were only observed for *U. brizantha.* Clades with specific ploidy levels were absent.

The comparison of the identified satellite monomers with tandem repeats previously reported for *Urochloa* revealed that nine variants were highly similar to 21 repeats from *U.*

brizantha, *U. decumbens*, and *U. humidicola* (Table 3). Alignment identities ranged from 81.82 to 100% over query coverages of 28-100%. UroSat-1a was also related to centromeric satellites from maize, *Cenchrus americanus* and *Tripsacum dactyloides* (Table 3).

Figure 3: (A) Genomic proportion of satellite DNA families in *Urochloa* species. The size of the circles corresponds to the relative abundance the satellite families in each genome. (B) Maximum likelihood phylogenetic tree of satellite variants.



Bootstrap values are shown near nodes. UB2 – *U. brizantha* 2x, UB4 – *U. brizantha* 4x, UD2 – *U. decumbens* 2x, UD4 – *U decumbens* 4x, UR2 – *U. ruziziensis*.

Source: from the Author (2023).

4 DISCUSSION

Repetitive DNA is the main component of *Urochloa* genomes, which is consistent with what has been reported for other cultivated grasses, like maize (70-80%), wheat (~80%) and barley (~85%) (GARBUS et al., 2015; MEYERS; TINGEY; MORGANTE, 2001; WICKER et al., 2017). Overall, the total proportions of the major repeat classes

(retrotransposons, transposons and tandem repeats) were similar among all genomes, except for the diploid cytotype of *U. brizantha*, which exhibited lower percentages for all classes. *Urochloa brizantha* 2x also displayed a larger proportion of transposable elements unidentified by repeat databases, thereby having a more diverse repeat profile.

Table 3: Alignments between the identified satellite sequences and reported repeats.

Satellite	Length	Reported repeat	Length	Species	Query coverage[a]	Identity[b]
UroSat-1a	158 bp	TCL2[1]	158 bp	*U. brizantha*	92%	98.63%
		CL1[1]	158 bp	*U. decumbens*	83%	100.00%
		TCL12[1]	158 bp	*U. brizantha*	93%	100.00%
		CL1[1]	158 bp	*U. brizantha*	98%	98.02%
		CL4[1]	158 bp	*U. brizantha*	98%	97.50%
		CL100[1]	567 bp	*U. humidicola*	68%	91.74%
		Sat-1A[2]	70 bp	*C. americanus*	38%	81.82%
		CentT57[3]	156 bp	*T. dactyloides*	82%	91.60%
		Cent-C-158a[4]	158 bp	*Zea mays*	46%	75.68%
UroSat-1b	158 bp	CL67[1]	158 bp	*U. decumbens*	100%	98.37%
		CL44[1]	158 bp	*U. brizantha*	100%	98.46%
		CL9[1]	157 bp	*U. brizantha*	100%	98.36%
		TCL2[1]	158 bp	*U. brizantha*	59%	91.49%
UroSat-1e	314 bp	CL121[1]	157 bp	*U. brizantha*	75%	93.08%
		CL72[1]	154 bp	*U. humidicola*	46%	90.67%
		CL9[1]	157 bp	*U. humidicola*	28%	84.62%
UroSat-1f	158 bp	CL9[1]	158 bp	*U. brizantha*	96%	90.79%
UroSat-2a	361 bp	CL34[1]	361 bp	*U. brizantha*	77%	95.36%
UroSat-2b	362 bp	CL35[1]	362 bp	*U. brizantha*	98%	96.83%
UroSat-7	566 bp	CL100[1]	156 bp	*U. brizantha*	100%	99.54%
UroSat-8	740 bp	CL101[1]	761 bp	*U. decumbens*	99%	94.90%
		CL77[1]	744 bp	*U. brizantha*	100%	100.00%
UroSat-9	498 bp	CL83[1]	418 bp	*U. brizantha*	82%	94.83%

[a]Query coverage indicates the proportion of the sequence included in the alignment; [b]Identity indicates the proportion of identical bases; [1]Tomaszewska et al. (2023); [2]Kamm et al. (1994); [3]Melo and Dawe (2003); [4]Ananiev et al. (1998).

Source: from the Author (2023).

Despite having nearly double the genome size, the repetitive DNA content in the polyploids was not proportionally larger than in the diploids. Considering the structural function of repetitive DNA, polyploidization events may create a selection pressure on genomes to maintain a balanced ratio of repeats and genes in a species. Such pressures may be rooted in preventing excessive activity of retrotransposons (WANG et al., 2021), optimizing chromosome mobility during cell division (SCHUBERT; OUD, 1997), and minimizing the nutrient cost of nucleic acid maintenance (KANG; WANG; HUANG, 2015).

4.1 Interspersed repeats

Retrotransposon activity has been pointed as the main driver of genome size increases in grasses, as the copy-and-paste transposition mechanism of these elements can lead to the rapid accumulation of repeats (VICIENT et al., 2001). This seems to be valid for *Urochloa*, considering that most of the repeats identified in this study are retrotransposons, predominantly *Gypsy* elements. In addition to the large genomic proportion of *gypsy* LTRs, RNAseq data have shown that they also represent near half of the transcriptome of *U. decumbens* 4x (SANTOS et al., 2015) and 24% of the *U. ruziziensis* genome (WORTHINGTON et al., 2021).

Athila elements, the most abundant *Gypsy* lineage identified in our study, have previously exhibited high expression levels in *Urochloa*, and have been associated with centromeric and pericentromeric regions of all chromosomes in *U. brizantha* 4x, *U. decumbens* 2x and *U. ruziziensis*, by FISH-mapping (SANTOS et al., 2015). *Tat* and CRM elements have also been mapped to *Urochloa* chromosomes, the former being scattered throughout interstitial regions and the latter located in the centromere, although a few chromosomes of *U. brizantha* 4x and *U. decumbens* 2x lacked the signal (NANI et al., 2016; SANTOS et al., 2015). Other *Gypsy* elements previously reported in *U. decumbens* transcriptome (SANTOS et al., 2015), such as *Del* and *Galadriel*, were not identified in our study.

The *Copia* superfamily was mainly composed of *Sirevirus* (SIRE), an ancient plant-specific retrotransposon lineage, which was also the most frequent LTR in *Urochloa*. These

elements are highly conserved across the plant kingdom and have been found in many grass species, including rice, sorghum and maize, in which they occupy 21% of the genome (BOUSIOS et al., 2012).

DNA transposons were predominantly represented by *CACTA* elements, which had a four-fold larger proportion compared to other repeats. The *CACTA* superfamily, named after the terminal inverted repeat sequence that flanks these elements, is a diverse group of transposons, including several lineages, such as *CAC1* from *Arabidopsis thaliana*, *Caspar* from Triticeae and *En/Spm* from maize (GIERL, 1996; MLURA et al., 2001; WICKER et al., 2003). All the *CACTA* repeats identified in our study were related to maize's *En/Spm* lineage, which indicates the conservation of these elements in the Panicoideae subfamily.

4.2 Satellite analysis

By using local alignment matrices, we were able to distinguish 11 satellite families for *U. brizantha*, *U. decumbens* and *U. ruziziensis*. The occurrence of UroSat-1, UroSat-3 and UroSat-8 in the three species indicates that these variants are more ancestral than the other satellites, possibly having originated prior to speciation events. Additionally, the high sequence similarity of UroSat-1a across the different *Urochloa* species, and its proximity to UroSat-1a from *Cenchrus americanus* (KAMM et al., 1994), suggests that these satellites have a conserved function in these species' genomes, possibly associated with centromeres. This is supported by the high sequence identity (75%) in alignments between UroSat-1 and a centromeric satellite from maize (ANANIEV; PHILLIPS; RINES, 1998).

Urochloa brizantha and *U. decumbens* shared more satellites with each other than with *U. ruziziensis*. This finding corroborates the most recent phylogenetic analysis for *Urochloa* that traced the divergence between *U. ruziziensis* and the other two species to 5.6 million years ago, followed by a more recent divergence between *U. brizantha* and *U. decumbens* around 1.6 mya (PESSOA-FILHO; MARTINS; FERREIRA, 2017).

The larger number of satellites and the occurrence of species-specific satellites in *U. brizantha* (UroSat-1f, 2b, 6a, 6b, 7 and 9) indicates that its divergence from *U. decumbens*

was accompanied by a great diversification of repeats. This repeat diversity is more evident when comparing the diploid cytotype of *U. brizantha*, which presented four exclusive satellites (UroSat-1c, 1d, 1e and 5), with *U. decumbens* 2x and *U. ruziziensis*, as both presented only three satellites and none of them was exclusive. The higher repeat diversity in *U. brizantha* 2x had already been indicated by the reciprocal GISH performed on the diploid cytotypes (CORRÊA et al., 2020), where its genomic probe fully hybridized with all chromosomes of *U. decumbens* 2x, while the opposite probing produced only ~65% of hybridization.

It has been established that the tetraploid cytotypes of *U. brizantha* (BBB^1B^1) and *U. decumbens* ($B^1B^1B^2B^2$) have originated via allopolyploidization (MENDES et al., 2006; MENDES-BONATO et al., 2006), both sharing the subgenome B^1B^1 (PAULA; SOUZA SOBRINHO; TECHIO, 2017). In this context, the two satellites found exclusively in the tetraploids (UroSat-10 and UroSat-11) may be associated with the subgenome B^1B^1.

The distribution of satellite families provided some insight on the genomic constitution of *U. brizantha.* The diploid cytotype was initially associated with the genome B^1B^1 due to the similar hybridization patterns of its genomic DNA on both *U. brizantha* 4x (BBB^1B^1) and *U. decumbens* 4x ($B^1B^1B^2B^2$) via GISH probing (CORRÊA et al., 2020). However, most of the satellite families found in *U. brizantha* 2x were not present in *U. decumbens* 4x, which indicates that the homologous regions highlighted by the GISH may be mostly due to the high TE composition. Post-polyploidization genomic rearrangements may have also affected the satellitome of *U. decumbens*, leading to the elimination of parental sequences. Nonetheless, the occurrence of six species-specific satellites in *U. brizantha* indicates that the diploid cytotype is indeed close to one of the subgenomes of its tetraploid counterpart.

Considering that the repeat characterization provides a greater resolution than GISH regarding the genomic constitution of related species, our new findings indicate that *U. brizantha* 2x is closer to the BB genome and it may have the ancestral parent involved in the allotetraploidization of *U. brizantha*.

The genome B^2, identified in *U. ruziziensis* (PAULA; SOUZA SOBRINHO; TECHIO, 2017) could not be distinguished by the satellite composition, since the three variants found were shared among all species. However, the distribution of these repeats may vary at the chromosome level and further studies including the physical mapping of the identified satellites are required to understand their organization on *Urochloa* genomes.

5 CONCLUSIONS

- The genomes of the *Urochloa* 'brizantha' complex is primarily composed of repetitive DNA;
- *Gypsy* retrotransposons are the most representative repeat lineage in *Urochloa*;
- Satellite DNA composition varied among the 'brizantha' complex species and therefore are suitable for comparative purposes and for being used as FISH probes in karyotyping studies;
- UroSat-1a is a putative centromeric repeat;
- *Urochloa brizantha* 2x is a putative progenitor of *U. brizantha* 4x;

ACKNOWLEDGEMENTS

The authors thank the Coordination for the Improvement of Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), the Research Foundation of the State of Minas Gerais (FAPEMIG) and the Oswaldo Cruz Foundation (Fiocruz) for funding the project, and the CNPq for providing the PhD scholarships.

REFERENCES

ANANIEV, E. V; PHILLIPS, R. L.; RINES, H. W. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. **Proceedings of the National Academy of Sciences**, v. 95, p. 13073–13078, 1998.

BOUSIOS, A. et al. MASiVEdb: The Sirevirus Plant Retrotransposon Database. **BMC Genomics**, v. 13, n. 1, p. 1–10, 2012.

CORRÊA, C. T. R. et al. GISH-based comparative genomic analysis in *Urochloa* P. Beauv. **Molecular Biology Reports**, v. 47, n. 2, p. 887–896, 2020.

FU, J. et al. Identification and characterization of abundant repetitive sequences in *Allium cepa*. **Scientific Reports**, v. 9, n. 1, 1 dez. 2019.

GAIERO, P. et al. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. **Annals of Botany**, v. 123, n. 3, p. 521–532, 2019.

GARBUS, I. et al. Characterization of repetitive DNA landscape in wheat homeologous group 4 chromosomes. **BMC Genomics**, v. 16, n. 1, 12 maio 2015.

GIERL, A. The En/Spm transposable element of maize. Em: **Transposable Elements**. [s.l: s.n.]. p. 145–159.

HEITKAM, T. et al. Satellite DNA landscapes after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. **Plant Journal**, v. 103, n. 1, p. 32–52, 2020.

JANK, L. et al. The value of improved pastures to Brazilian beef production. **Crop and Pasture Science**, v. 65, p. 1132–1137, 2014.

KAMM, A.; SCHMIDT, T; HESLOP-HARRISON, J.S. Molecular and physical organization of highly repetitive, undermethylated DNA from Pennisetum glaucum. **Molecular and General Genetics MGG**, 244, p. 420-425, 1994.

KANG, M.; WANG, J.; HUANG, H. Nitrogen limitation as a driver of genome size evolution in a group of karst plants. **Scientific Reports**, v. 5, 25 jun. 2015.

KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 1 jun. 2018.

LIU, Q. et al. The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. **BMC Plant Biology**, v. 19, n. 1, p. 1–17, 2019.

LUTTS, S.; NDIKUMANA, J.; LOUANT, B. P. Fertility of *Brachiaria ruziziensis* in interspecific crosses with *Brachiaria decumbens* and *Brachiaria brizantha*: meiotic behavior, pollen viability and seed set. **Euphytica**, v. 57, p. 267–274, 1991.

MENDES, D. V. et al. Cytological evidence of natural hybridization in *Brachiaria brizantha* Stapf (Gramineae). **Botanical Journal of the Linnean Society**, v. 150, p. 441–446, 2006.

MENDES-BONATO, A. B. et al. Meiotic instability in invader plants of signal grass *Brachiaria* decumbens Stapf (Gramineae). **Acta Scientiarum**, v. 23, n. 2, p. 619–625, 2001.

MENDES-BONATO, A. B. et al. Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). **Euphytica**, v. 125, p. 419–425, 2002.

MENDES-BONATO, A. B. et al. Cytogenetic evidence for genome elimination during microsporogenesis in interspecific hybrid between *Brachiaria ruziziensis* and *B. brizantha* (Poaceae). **Genetics and Mole**, v. 29, n. 4, p. 711–714, 2006.

MEYERS, B. C.; TINGEY, S. V.; MORGANTE, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. **Genome Research**, v. 11, n. 10, p. 1660–1676, 2001.

MLURA, A. et al. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. **Nature**, v. 411, n. 6834, p. 212–214, 10 maio 2001.

MORAES, I. DE C. et al. Karyotype analysis and mode of reproduction of two species of *Urochloa* P. Beauv. **Crop Science**, v. 61, n. 5, p. 3415–3424, 1 set. 2021.

NANI, T. F. et al. Physical map of repetitive DNA sites in *Brachiaria* spp.: Intravarietal and interspecific polymorphisms. **Crop Science**, v. 56, n. 4, p. 1769–1783, 2016.

NANI, T. F. et al. Location of low copy genes in chromosomes of *Brachiaria* spp. **Molecular Biology Reports**, v. 45, n. 2, p. 109–118, 1 abr. 2018.

NEUMANN, P. et al. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA**, v. 10, n. 1, p. 1–17, 2019.

NITTHAISONG, P. et al. Pentaploid apomicts by interspecific hybridization between diploid urochloa ruziziensis and tetraploid apomictic *U. decumbens*. **Crop Science**, v. 59, n. 4, p. 1648–1656, 1 jul. 2019.

NOVÁK, P. et al. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792–793, 2013.

NOVÁK, P. et al. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Research**, v. 45, n. 12, 2017.

PAULA, C. M. P. DE; SOUZA SOBRINHO, F.; TECHIO, V. H. Genomic constitution and relationship in Urochloa (Poaceae) species and hybrids. **Crop Science**, v. 57, n. 5, p. 2605–2616, 2017.

PESSOA-FILHO, M.; MARTINS, A. M.; FERREIRA, M. E. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes. **BMC Genomics**, v. 18, n. 1, p. 516, 2017.

POWO. **Plants of the World Online. Facilitated by the Royal Botanic Garden, Kew**, 2023.

RENVOIZE, S.; MAASS, B. *Brachiaria*. A report to CIAT, Colombia, on the species and specimens held in the germplasm collection. **Kew. Royal Botanical Gardens.**, 1993.

RUIZ-RUANO, F. J. et al. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, n. January, p. 1–14, 2016.

SANTOS, F. C. et al. Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. **Chromosome Research**, v. 23, n. 3, p. 571–582, 2015.

SCHUBERT, I.; OUD, J. L. There Is an Upper Limit of Chromosome Size for Normal Development of an Organism. **Cell**, v. 88, p. 515–520, 1997.

SMIT, A. F. A.; HUBLEY, R.; GREEN, P. **RepeatMasker 4.1.1**. Disponível em: <http://www.repeatmasker.org/>. Acesso em: 22 jan. 2021.

TIMBÓ, A. L. DE O. et al. Obtaining tetraploid plants of ruzigrass (*Brachiaria ruziziensis*). **Revista Brasileira de Zootecnia**, v. 43, n. 3, p. 127–131, 2014.

TOMASZEWSKA, P. et al. 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**, 2023.

VALLE, C. B.; PAGLIARINI, M. S. Biology, Cytogenetics, and Breeding of *Brachiaria*. Em: **Genetic resources, chromosome engineering, and crop improvement**. v. 2p. 103–143, 2009.

VICIENT, C. M. et al. Active retrotransposons are a common feature of grass genomes. **Plant Physiology**, v. 125, n. 3, p. 1283–1292, 2001.

WANG, X. et al. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. **Plant Journal**, v. 107, n. 4, p. 1003–1015, 1 ago. 2021.

WICKER, T. et al. CACTA transposons in triticeae. A diverse family of high-copy repetitive elements. **Plant Physiology**, v. 132, n. 1, p. 52–63, 2003.

WICKER, T. et al. The repetitive landscape of the 5100 Mbp barley genome. **Mobile DNA**, v. 8, n. 1, p. 1–16, 2017.

WORTHINGTON, M. et al. A new genome allows the identification of genes associated with natural variation in aluminium tolerance in *Brachiaria* grasses. **Journal of Experimental Botany**, v. 72, n. 2, p. 302–319, 2 fev. 2021.

ARTIGO 2: Physical mapping of satellite DNA sequences in the 'brizantha' agamic complex of *Urochloa* (Poaceae)

Norma NBR 6022 (ABNT 2018)

ABSTRACT

Urochloa P. Beauv. (syn. *Brachiaria*) is a genus of tropical grasses largely cultivated in South America for cattle nutrition. *Urochloa brizantha*, *U. decumbens* and *U. ruziziensis* are among the most economically relevant species, but their complex genomic constitution as consequence of natural hybridizations and polyploidy is a constraint for breeding programs. Repetitive DNA usually represents a large portion of grass genomes, comprising numerous lineages of transposable elements and satellite DNA (satDNA) families. SatDNA repeats can be used as cytogenetic markers to improve karyotype descriptions and to study the behavior of multiple genomes in allopolyploids such as *U. brizantha* and *U. decumbens*. Therefore, this study aimed to physically map satellite DNA sequences on *Urochloa* chromosomes, characterize their organization and distribution among species and cytotypes, and evaluate their potential as chromosome markers. We used fluorescent *in situ* hybridization (FISH) to map three satellite sequences (UroSat-1a, UroSat-2a, and UroSat-3) on chromosomes of diploid and tetraploid accessions of *U. brizantha*, *U. decumbens*, and *U. ruziziensis*. Our results revealed that UroSat-1a is a centromeric repeat, present in all chromosomes of all species. The number of chromosome pairs bearing UroSat-2a and UroSat-3 varied among species, and there was an association between UroSat-2a and a tertiary constriction in the tetraploid *U. decumbens*. Our findings validate the use of satellite sequences as cytogenetic markers for *Urochloa* species, allowing the identification of chromosome pairs, highlighting the composition of specific chromosome regions, and demonstrating the role of satDNA in chromosome organization and evolution in the genus.

Keywords: Polyploidy, *in situ* hybridization, tandem repeat, forage grasses, FISH, repetitive DNA.

1 INTRODUCTION

The African forage genus *Urochloa* P. Beauv. occupies more than 80% of the 112 million hectares of cultivated pastures in Brazil, playing a crucial role in the beef and dairy industry (EMBRAPA, 2022; IBGE, 2017). As in other grasses, polyploidy and natural hybridizations were part of the evolutionary history of *Urochloa* (ESTEP et al., 2014), resulting in a wide range of chromosome numbers (2n = 14, 24, 36, 42, 45, 54, 72, 90) and ploidy levels (2x, 4x, 5x, 6x, 9x) (VALLE; PAGLIARINI, 2009). Interspecific hybridization is a common breeding strategy for *Urochloa*, but the prevalence of asexual reproduction through apomixis, and the large variation of chromosome numbers complicate possible crossing combinations (JANK et al., 2014; PAULA; SOUZA SOBRINHO; TECHIO, 2017).

Three of the most economically important *Urochloa* species, *U. brizantha*, *U. decumbens* and *U. ruziziensis* form an agamic complex known as the 'brizantha' complex, in which the former two are primarily apomictic tetraploids (2n=4x=36), and the latter is a sexual reproducing diploid (2n=2x=18) (VALLE; SAVIDAN, 1996). Molecular phylogenies have shown that these three species are very closely related, with an earlier speciation of *U. ruziziensis* around 5 million years ago (mya), followed by the divergence of *U. brizantha* and *U. decumbens* 1.6 mya (PESSOA-FILHO; MARTINS; FERREIRA, 2017; TRIVIÑO et al., 2017). Breeding programs have been able to perform interspecific crosses involving these three species through the artificial tetraploidization of *U. ruziziensis* (TIMBÓ et al., 2014), but hybrids tend to present meiotic abnormalities and aneuploidy (MORAES et al., 2019; ROCHA et al., 2019).

Extensive evidence, including meiotic studies (MENDES-BONATO et al., 2002a, 2002b, 2006; RICCI et al., 2011) and cytomolecular approaches (CORRÊA et al., 2020; PAULA; SOUZA SOBRINHO; TECHIO, 2017; ROCHA et al., 2019), has demonstrated that the tetraploidy of *U. brizantha* and *U. decumbens* has a hybrid origin. Paula et al. (2017) proposed the genomic constitution BBB^1B^1 and $B^1B^1B^2B^2$ for the allotetraploids *U. brizantha* and *U. decumbens*, respectively, and B^2B^2 for the diploid *U. ruziziensis*, using the same letter to highlight the high degree of homoeology between the genomes, and the superscripts to

differentiate the subgenomes. Tomaszewska et al. (2023) proposed a second genomic nomenclature, using the upper-case letters B, D, R for *U. brizantha* (B^bB^bDD), *U. decumbens* (B^bDRR) and *U. ruziziensis* (RR), based on genome-specific repetitive DNA probes. Despite the different nomenclatures, both studies indicate that the *U. brizantha* and *U. decumbens* have shared subgenomes, while only *U. decumbens* seem to have a common genome with *U. ruziziensis.*

Considering the complexity of genome constitutions within the *Urochloa* agamic complexes, the development of cytogenetic markers is essential to provide more detailed karyotype descriptions. Repetitive sequences are great chromosome markers as the high number of copies allows better visualization under the microscope via fluorescent *in situ* hybridization (FISH) (GUERRA, 2012). More specifically, satellite DNA is a good choice for FISH probes since its clustered organization results in conspicuous signals, as opposed to the dispersed marks characteristic of transposable elements. In *Urochloa*, ribosomal genes (AKIYAMA; YAMADA-AKIYAMA; EBINA, 2010; MENDES, 2020; NANI et al., 2016), the retrotransposon lineages *Athila*, *CRM*, *Del*, and *Tat* (NANI et al., 2016; SANTOS et al., 2015), and 50-mer repeats (TOMASZEWSKA et al., 2023) have been mapped to chromosomes, but studies using satellite probes are incipient.

Correa et al. (2023, unpublished) characterized the repetitive DNA composition of the agamic complex species through genome-skimming, identifying 11 satellite families among diploid and tetraploid forms of *U. brizantha* and *U. decumbens,* as well as *U. ruziziensis*. Information on genomic abundance and sequence identity among species was provided for each satellite, but the organization of the repeats throughout the chromosome complements is yet to be understood. The satellite sequences identified in the study varied among species, being suitable as cytogenetic markers. Thus, this study aimed to physically map satellite DNA sequences on *Urochloa* chromosomes, characterize their organization and distribution among species and cytotypes, and evaluate their potential as chromosome markers.

2 MATERIAL AND METHODS

2.1 Plant material

The studies were conducted using two diploid accessions (2n = 2x = 18) and two tetraploid cultivars (2n = 4x = 36) of *U. brizantha* and *U. decumbens*, and one diploid cultivar of *U. ruziziensis* (2n = 2x = 18) provided by EMBRAPA Gado de Corte, Campo Grande – MS, Brazil, and EMBRAPA Gado de Leite, Juiz de Fora - MG, Brazil.

2.2 Chromosome preparations

To obtain mitotic metaphases, root tips were pre-treated with cycloheximide (12.5 mg/L) for 2h at room temperature and fixed in ethanol/acetic acid (3:1) solution. Cell wall digestion was performed in an enzyme solution (0.7% cellulase Onozuka R10, 0.7% cellulase Sigma-Aldrich, 1% pectolyase Sigma-Aldrich, and 1% cytohelicase Sigma-Aldrich) for 90 min at 37 °C. Slides were prepared according to the cell dissociation technique and air-dried (Dong et al. 2001).

2.3 Physical mapping of satellite DNA

To physically map satellite DNA sequences onto *Urochloa* chromosomes, we selected monomers previously identified via clustering analysis (Correa et al., unpublished) and designed probes for the satellites with higher genomic proportions (Table 1). Primers were designed for the consensus sequences of the most abundant satellites (UroSat-1a, UroSat-2a and UroSat-3; Supplementary Material) using the online tool Primer3 (KÕRESSAAR et al., 2018).

The genomic DNA (gDNA) was isolated from fresh leaves via CTAB protocol (Doyle and Doyle 1990). The probes for the satellites UroSat-1a, UroSat-2a and UroSat-3 were obtained via PCR amplification using digoxigenin-labeled nucleotides and gDNA from *U. brizantha* cv. Marandu, *U. decumbens* cv. Basilisk, and *U. decumbens* D04, respectively. PCR amplification was performed in a total volume of 25 μL containing 12.5 μL of Cellco® Taq Pol Master Mix (0.05 U/μL of Taq DNA polymerase, 0.3 mM of $MgCl_2$, 0.4 mM of each

dNTP), 1 μL of digoxigenin-11-dUTP (0.8 mM), 0.5 μL of each primer (10 μM), 1 μL of DNA template (100 ng/μL), and 9.5 μL of ddH_2O. The PCR parameters were the same for all satDNA probes: an initial denaturation at 95ºC for 2 min, followed by 30 cycles of denaturation at 95ºC for 25 sec, annealing at 54ºC for 1 min, and extension at 68 ºC for 2 min. We also used a 35S rDNA probe (pTa71) as a positive control in the fluorescent *in situ* hybridization (FISH) assay. The 35S probe was labeled with biotin-16-dUTP via nick translation reaction.

For the FISH, chromosome slides were denatured in 70% formamide at 85ºC for 1 min and 25s, and dehydrated in ethanol series (70, 90 and 100%) for 5 min each. A hybridization mixture containing 50% formamide, 10% dextran sulfate, 2x saline sodium citrate (SSC) buffer (pH 7.0), 50-100 ng of digoxigenin-labeled satellite probe, and 50-100 ng of biotin-labeled 35S rDNA probe was denatured at 95ºC and applied to the slides. The hybridization was performed in humid chamber at 37ºC for 24h. The hybridized slides were washed in SSC 2x for 5 min at room temperature followed by a stringency wash in SSC 2x at 42ºC for 10 min. The probes were detected using the antibodies anti-digoxigenin-rhodamine and Alexafluor® 488-streptavidin at 37ºC for 1h. The chromosomes were stained with DAPI (4′,6-diamidino-2-phenylindole)/Vectashield® and images were captured by a QImaging Retiga EXi CCD camera attached to a fluorescence microscope Olympus BX 60.

Karyograms and idiograms were constructed using Adobe Photoshop® CC 2019. For each sample, 10 metaphases pooled from a minimum of three FISH assays were used to describe the organization of the satellite sequences. Homologous chromosomes were paired based on FISH signals and morphometric characteristics, considering previous karyotypes reported by Nani et al. (2016) and Mendes (2020).

3 RESULTS

The probing of satellite DNA successfully revealed the organization of UroSat-1a, UroSat-2a and UroSat-3 on *Urochloa* chromosomes (Fig. 1-3, Table 1). None of the probed satellites overlapped with rDNA 35S sites in any species. UroSat-1a was detected in the centromeric region of all chromosomes of all species and cytotypes evaluated (Fig. 1A, C;

Fig. 2A, C; Fig. 3A). The UroSat-2a probe produced signals located on terminal regions on the short arm of the chromosomes, while the location of UroSat-3 varied among the terminal region of short arms, long arms, or both arms. UroSat-2a was present as strong and well-defined bands in *U. brizantha* 2x and 4x, and *U. decumbens* 4x, and as unspecific weaker signals in *U. decumbens* 2x and *U. ruziziensis* (Fig. S1). UroSat-3 was present in all species and cytotypes, apart from *U. brizantha* 2x.

Table 1: Distribution of FISH signals for UroSat-1a, UroSat-2a, and UroSat-3 in *Urochloa*

Species	UroSat-1a	UroSat-2a	UroSat-3
U. brizantha 2x	18 C	3 TS	no signals
U. brizantha 4x	36 C	4 TS	4 TSL + 2 TS + 1 TL*
U. decumbens 2x	18 C	unspecific signals	5 TSL + 3 TS + 1 TL
U. decumbens 4x	36 C	2 TS + 1 TL	7 TSL + 2 TS + 1 TL
U. ruziziensis	18 C	unspecific signals	6 TSL + 2 TS

Signals were found on centromeric (C) and terminal (T) regions of short arms (S) and long arms (L). *Heteromorphic pair.

In *U. brizantha*, three pairs of UroSat-2a signals were observed in the diploid form (Fig. 1B), while the tetraploid form presented four (Fig. 1D). UroSat-3 was found on six chromosome pairs in *U. brizantha* 4x, and it was hemizygous in the pair bearing the rDNA 35S site (Fig. 1E). This probe did not produce any markings on the diploid cytotype.

In *U. decumbens*, UroSat-3 signals were found in all chromosomes of the diploid cytotype (Fig. 2B), including the pair containing the 35S site. The same probe produced signals on eight chromosomes of *U. decumbens* 4x (Fig. 2E). In the tetraploid *U. decumbens*, one of chromosome pairs containing the 35S site exhibited a heteromorphic tertiary constriction associated with an UroSat-2a signal (Fig. 3D). Signals were also found on two additional chromosome pairs. UroSat-2 signals were variable and inconsistent in terms of number and location in the *U. decumbens* 2x (Fig. S1), therefore it was not possible to assign them to specific chromosome pairs.

Figure 1: Karyotypes and idiograms of of *U. brizantha* diploid (2n=2x=18) and tetraploid (2n=4x=36) probed by UroSat-1a (A, C), UroSat-2a (B, D), UroSat-3 (E) and 35S rDNA. The arrow heads indicate heteromorphic chromosome pairs (white, yellow) and a tertiary constriction (yellow).

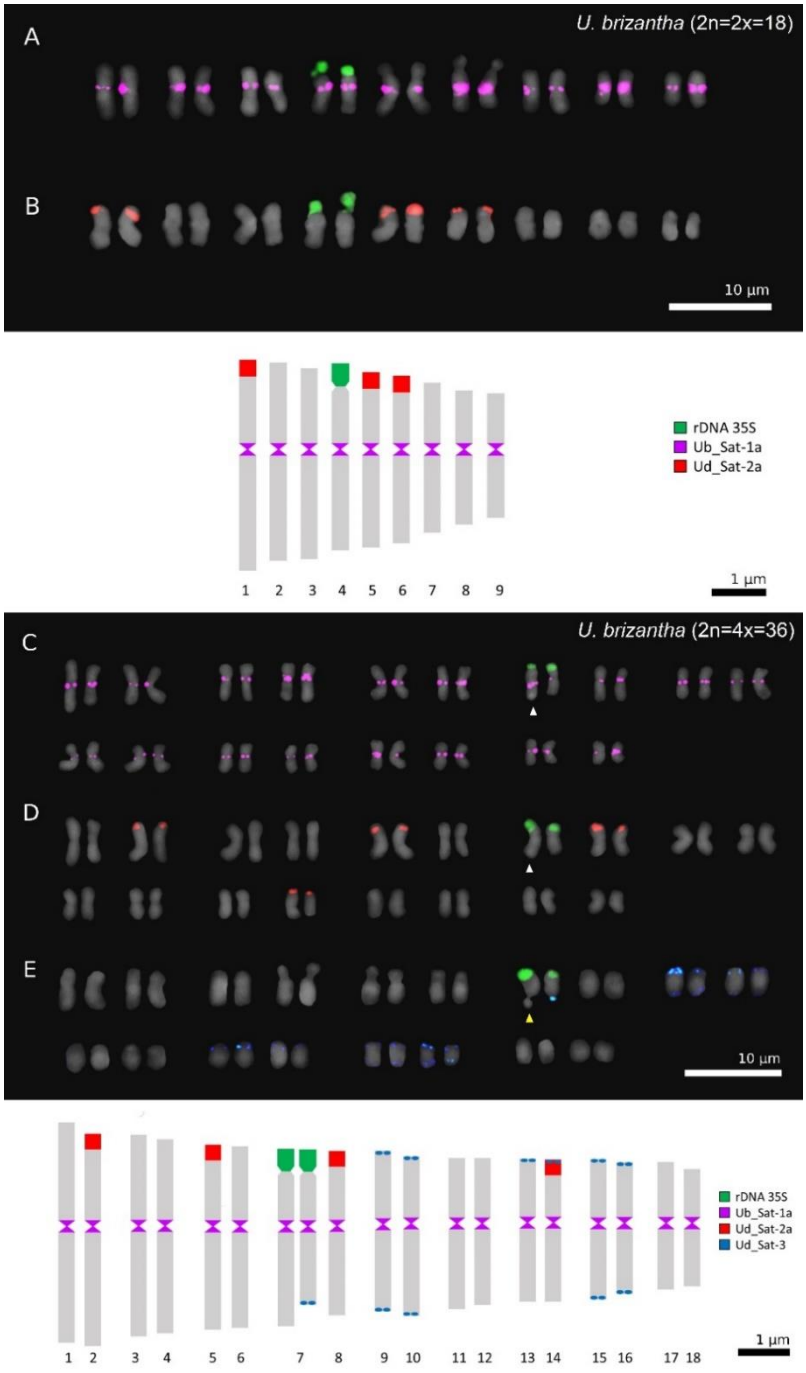
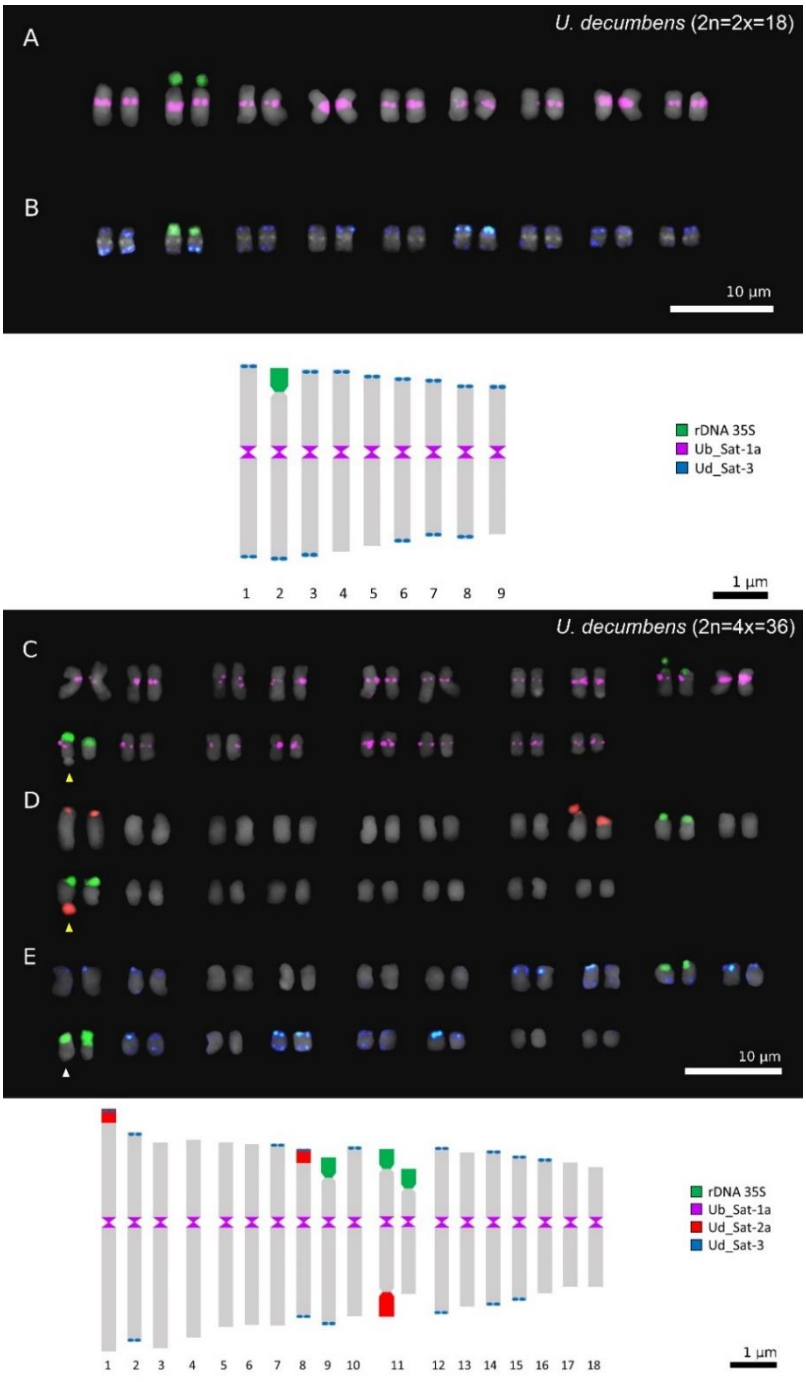
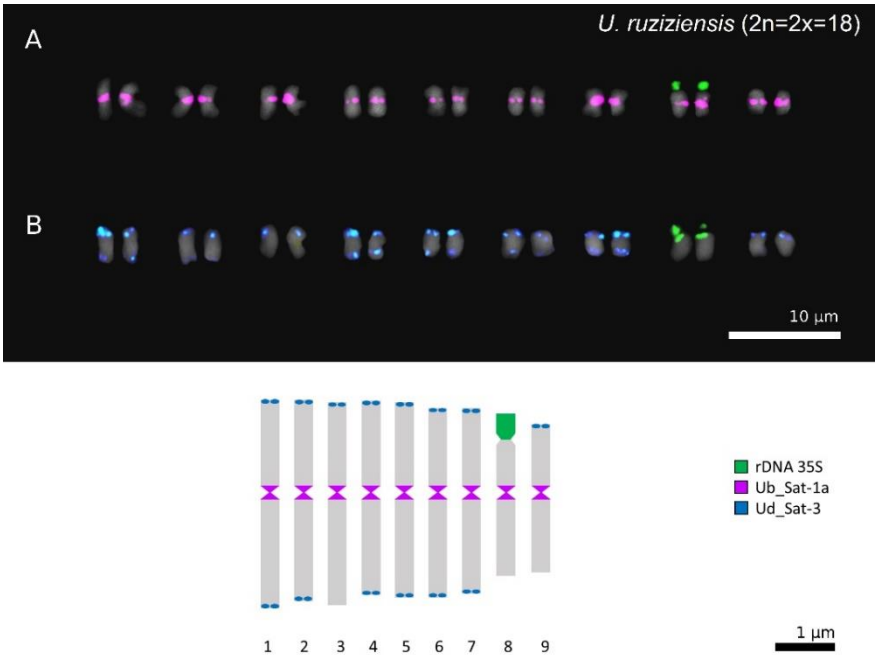


Figure 2: Karyotypes and idiograms of *U. decumbens* diploid (2n=2x=18) and tetraploid (2n=4x=36) probed by UroSat-1a (A, C), UroSat-2a (D), UroSat-3 (B, E), and 35S rDNA. The arrow heads indicate heteromorphic chromosome pairs (white, yellow) and a tertiary constriction (yellow).



In *U. ruziziensis*, UroSat-3 signals were visible on all chromosomes apart from the pair bearing the 35S site (Fig. 3B). Similar to *U. decumbens* 2x, the UroSat-2 probe produced unspecific and variable signals (Fig. S1).

Figure 3: Karyotypes and idiogram of *U. ruziziensis* (2n = 2x = 18) probed by UroSat-1a (A), UroSat-3 (B), and 35S rDNA.



4 DISCUSSION

The physical mapping of satellite sequences on the *Urochloa* species studied was consistent with the satellitome characterized *in silico*, in which UroSat-1a was present in all species and cytotypes, UroSat-3 was absent in *U. brizantha* 2x, but present in the other cytotypes and species, and UroSat-2a was present only in *U. brizantha* (2x and 4x), and *U. decumbens* 4x. However, we observed highly variable and unspecific UroSat-2a signals in *U. decumbens* 2x and *U. ruziziensis*, where this sequence had not been identified *in silico*. These signals may have been produced by probes binding to degraded DNA segments with some degree of homology with the satellite sequence, as often occurs with FISH probing (JIANG, 2019; WANG; XU; GUO, 2001).

The probing of UroSat-2a allowed the differentiation of three chromosome pairs in *U. brizantha* 2x (1, 5 and 6), four pairs in *U. brizantha* 4x (2, 5, 8 and 14), and two pairs in U. decumbens (1 and 8) given that the third pair bearing the signal also contained a 35S rDNA site. The disproportional number of sites between diploid and tetraploid cytotypes of *U. brizantha* and *U. decumbens* may be a consequence of the allopolyploidization events involving these two species. Such variation has also been observed for 35S and 5S rDNA sites (AKIYAMA; YAMADA-AKIYAMA; EBINA, 2010; MENDES, 2020; NANI et al., 2016).

The heteromorphisms for chromosome length observed in the tetraploids *U. brizantha* and *U. decumbens*, and the tertiary constriction in *U. decumbens* had already been reported by Nani et al. (2016), while the tertiary constriction in *U. brizantha* 4x was not detected in the same study, although the authors described the pair as heteromorphic for chromosome length. Our results for *U. decumbens* demonstrate that satellite DNA is a major component of the tertiary constriction, as shown by its co-localization with UroSat-2a. Interestingly, UroSat-2a was not present in *U. brizantha*'s tertiary constriction, which indicates that its formation does not depend on sequence composition alone or it may be associated with a different repeat.

Tertiary constrictions have been reported for other grass species, such as barley (GECHEFF, 1976), rye (JENKINS et al., 2005), gamagrass (KOO; JIANG, 2008), and *Lolium temulentum*, *L. persicum*, and *L. remotum* (SANTOS, 2022), often associated with chromosomal breaks. Tandem repeats are also considered frequent breakage sites, and hotspots for chromosome rearrangements (ARLT et al., 2006), which might explain the presence of satellite sequences like UroSat-2a close to a tertiary constriction.

The presence of UroSat-1a exclusively on the centromeric region of chromosomes in all the genotypes suggests that this sequence is an essential component of *Urochloa* centromeres. Satellite DNA is a common constituent of plant centromeres, and several centromeric satellites have been identified in grass species, such as maize, wheat, and oat (JIN et al., 2004; OLIVEIRA; TORRES, 2018; SU et al., 2019). Additional evidence supporting the centromeric role of UroSat-1a is the high level of conservation of this satellite among *Urochloa* species, having 100% of sequence identity among satellites independently

identified via bioinformatics analysis in *U. brizantha*, *U. decumbens*, and *U. ruziziensis*, as well as the high similarity between UroSat-1a and centromeric satellites from maize (~75%) and *Cenchrus americanus* (~82%) (CORREA et al., 2023, unpublished). This high conservation of centromeric satellites among grasses has also been demonstrated in comparative analysis involving different species, including barley, rice, maize, millet, and *Panicum* (MELTERS et al. 2013).

The centromeric role of UroSat-1a is likely extended to other *Urochloa* species beyond the '*brizantha*' agamic complex, as it is highly similar (92% identity) to a tandem repeat (CL100) identified in an accession of *U. humidicola* in a different study (CORREA et al., 2023, unpublished; TOMASZEWSKA et al., 2023). However, the repeat was not mapped to *U. humidicola* chromosomes, thereby its physical location is yet to be confirmed. Nonetheless, our findings for UroSat-1a enables its use as a FISH probe in several applications, such as karyotype descriptions, characterization of centromeres, tracking chromosome breaks and rearrangements, comparative cytogenetic analyses, and studies on chromosome evolution.

The distribution of UroSat-3 signals in *U. ruziziensis* and the two cytotypes of *U. decumbens* highlights the similar repetitive DNA profile between the two species, which is consistent with genome skimming data (CORREA e al., unpublished), and *in situ* hybridization studies (CORRÊA et al., 2020; PAULA; SOUZA SOBRINHO; TECHIO, 2017; TOMASZEWSKA et al., 2023). The tetraploid *U. brizantha* also presented a similar pattern of UroSat-3 signals. Both allopolyploidization models proposed for the 'brizantha' agamic complex suggested that the tetraploids *U. brizantha* and *U. decumbens* share a subgenome most likely inherited from a diploid *U. decumbens* parent (PAULA; SOUZA SOBRINHO; TECHIO, 2017; TOMASZEWSKA et al., 2023). In this scenario, the chromosomes marked by the UroSat-3 probe are probably derived from this shared genome, considering that this satellite is not present in the diploid *U. brizantha*. Alternatively, UroSat-3 may be a more ancestral sequence, originated before the speciation events, and eventually eliminated in *U. brizantha* 2x.

Overall, the distribution of UroSat-2a and UroSat-3 was more similar between the tetraploids cytotypes of *U. brizantha* and *U. decumbens* than between tetraploids and diploids withing the same species, considering than UroSat-3 was not present in *U. brizantha* 2x and UroSat-2a was absent in *U. decumbens* 2x. These patterns are consistent with the genetic structure of the 'brizantha' agamic complex, which has shown that tetraploid accessions of *U. brizantha* and *U. decumbens* are phylogenetically closer to each other than to their diploid counterparts due the reproductive barriers caused by the polyploidization events (HIGGINS et al., 2022).

5 CONCLUSIONS

- UroSat-1a, UroSat-2a and UroSat-3 were validated as cytogenetic markers for 'brizantha' agamic complex of *Urochloa*;
- Satellite DNA is distributed in (peri)centromeric and terminal chromosomal regions in *Urochloa;*
- UroSat-1a was confirmed as a centromeric satellite in the 'brizantha' agamic complex;
- Satellite DNA is associated with chromosome heteromorphisms and tertiary constrictions in *Urochloa*;
- Polyploidization and hybridization events directly affected the distribution of satellite DNA in *Urochloa*.
- The similar organization of satDNA between the tetraploids *U. brizantha* and *U. decumbens*, and between the diploids *U. decumbens* and *U. ruziziensis* may be a result of the reproductive isolation imposed by ploidy differences within the agamic complex.

ACKNOWLEDGEMENTS

This study was supported by Coordination for the Improvement of Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), the Research Foundation of the State of Minas Gerais (FAPEMIG), and the Oswaldo Cruz Foundation (Fiocruz).

REFERENCES

AKIYAMA, Y.; YAMADA-AKIYAMA, H.; EBINA, M. Morphological diversity of chromosomes bearing ribosomal DNA loci in *Brachiaria* species. **Grassland Science**, v. 56, n. 4, p. 217–223, 2010.

ARLT, M. F. et al. Common fragile sites as targets for chromosome rearrangements. **DNA Repair**, v. 5, n. 9–10, p. 1126–1135, 8 set. 2006.

CORRÊA, C. T. R. et al. GISH-based comparative genomic analysis in *Urochloa* P. Beauv. **Molecular Biology Reports**, v. 47, n. 2, p. 887–896, 2020.

CORRÊA, C. T. R. et al. Repetitive DNA landscape in the 'brizantha' agamic complex of *Urochloa*. **Manuscript in preparation**, 2023.

DONG, F. et al. The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. **Genome**, n. 44, p. 729–734, 2001.

DOYLE, J.; DOYLE, J. L. Isolation of Plant DNA from fresh tissue. **Focus**, v. 12, n. 13, p. 39–40, 1990.

EMBRAPA. Brazil creates its first *Brachiaria ruziziensis* grass cultivar. Disponível em: https://www.embrapa.br/en/busca-de-noticias/-/noticia/68876481/brazil-creates-its-first-brachiaria-ruziziensis-grass-cultivar. Acesso em: 7 abr. 2022.

ESTEP, M. C. et al. Allopolyploidy, diversification, and the Miocene grassland expansion. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 42, p. 15149–15154, 21 out. 2014.

GECHEFF, K. I. A new translocation in barley resulting in a cytologically marked karyotype. **Mutation Research**, v. 36, p. 241–244, 1976.

GUERRA, M. Cytotaxonomy: the end of childhood. **Plant Biosystems**, v. 146, n. 3, p. 703–710, 2012.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Censo Agropecuario 2017**, 2017.

JANK, L. et al. The value of improved pastures to Brazilian beef production. **Crop and Pasture Science**, v. 65, p. 1132–1137, 2014.

JENKINS, G. et al. Strategies for the study of meiosis in rye. **Cytogenetic and Genome Research**, 2005.

JIANG, J. Fluorescence in situ hybridization in plants: recent developments and future applications. **Chromosome Research**, v. 27, n. 3, p. 153–165, 1 set. 2019.

JIN, W. et al. Maize centromeres: Organization and functional adaptation in the genetic background of oat. **Plant Cell**, v. 16, n. 3, p. 571–581, 2004.

KOO, D. H.; JIANG, J. Extraordinary tertiary constrictions of *Tripsacum dactyloides* chromosomes: Implications for karyotype evolution of polyploids driven by segmental chromosome losses. **Genetics**, v. 179, n. 2, p. 1119–1123, jun. 2008.

KÕRESSAAR, T. et al. Primer3-masker: Integrating masking of template sequence with primer design software. **Bioinformatics**, v. 34, n. 11, p. 1937–1938, 2018.

MENDES, L. M. Análise cariotípica em acessos diploides de *Urochloa brizantha* e *Urochloa decumbens* e poliploides de *Urochloa dictyoneura* (Poaceae). Master thesis. **Federal University of Lavras**, 2020.

MENDES-BONATO, A. B. et al. Unusual cytological patterns of microsporogenesis in *Brachiaria decumbens*: Abnormalities in spindle and defective cytokinesis causing precocious cellularization. **Cell Biology International**, v. 26, n. 7, p. 641–646, 2002a.

MENDES-BONATO, A. B. et al. Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). **Euphytica**, v. 125, p. 419–425, 2002b.

MENDES-BONATO, A. B. et al. Cytogenetic evidence for genome elimination during microsporogenesis in interspecific hybrid between *Brachiaria ruziziensis* and *B. brizantha* (Poaceae). **Genetics and Mole**, v. 29, n. 4, p. 711–714, 2006.

MORAES, I. DE C. et al. Characterization of aneuploidy in interspecific hybrid between *Urochloa ruziziensis* (R. Germ. & Evrard) Crins and *Urochloa decumbens* (Stapf) R. D. Webster. **Molecular Biology Reports**, v. 46, n. 2, p. 1931–1940, 2019.

NANI, T. F. et al. Physical map of repetitive DNA sites in *Brachiaria* spp.: Intravarietal and interspecific polymorphisms. **Crop Science**, v. 56, n. 4, p. 1769–1783, 2016.

OLIVEIRA, L. C.; TORRES, G. A. Plant centromeres: genetics, epigenetics and evolution. **Molecular Biology Reports,** v.45, p. 1491-1497. 2018.

PAULA, C. M. P. DE; SOUZA SOBRINHO, F.; TECHIO, V. H. Genomic constitution and relationship in *Urochloa* (Poaceae) species and hybrids. **Crop Science**, v. 57, n. 5, p. 2605–2616, 2017.

PESSOA-FILHO, M.; MARTINS, A. M.; FERREIRA, M. E. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes. **BMC Genomics**, v. 18, n. 1, p. 516, 2017.

RICCI, G. C. L. et al. Chromosome numbers and meiotic analysis in the pre-breeding of *Brachiaria decumbens* (Poaceae). **Indian Academy of Sciences**, v. 90, n. 2, p. 289–294, 2011.

ROCHA, M. J. DA et al. Comparative meiosis and cytogenomic analysis in euploid and aneuploid hybrids of *Urochloa* P. Beauv. **Chromosome Research**, 2019.

SANTOS, F. C. et al. Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. **Chromosome Research**, v. 23, n. 3, p. 571–582, 2015.

SANTOS, Y. D. Cytogenetic and genomic aspects of wild species of the *Lolium-Festuca* complex. PhD thesis. **Federal University of Lavras**, 2022.

SU, H. et al. Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. **Plant Cell**, v. 31, n. 9, p. 2035–2051, 2019.

TIMBÓ, A. L. DE O. et al. Obtaining tetraploid plants of ruzigrass (*Brachiaria ruziziensis*). **Revista Brasileira de Zootecnia**, v. 43, n. 3, p. 127–131, 2014.

TOMASZEWSKA, P. et al. 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**, 2023.

TRIVIÑO, N. J. et al. Genetic Diversity and Population Structure of *Brachiaria* Species and Breeding Populations. **Crop Science**, v. 57, n. 5, p. 2633–2644, 2017.

VALLE, C. B.; PAGLIARINI, M. S. Biology, Cytogenetics, and Breeding of *Brachiaria*. Em: **Genetic resources, chromosome engineering, and crop improvement**. v. 2p. 103–143, 2009.

VALLE, C. B.; SAVIDAN, Y. H. Genetics, cytogenetics and reproductive biology of Brachiaria. Em: **Brachiaria: biology, agronomy and improvement**. p. 163–180, 1996.

WANG, Y.; XU, Z.; GUO, X. A centromeric satellite sequence in the Pacific oyster (Crassostrea gigas Thunberg) identified by fluorescence in situ hybridization. **Marine Biotechnology**, v. 3, n. 5, p. 486–492, 2001.

APÊNDICE

Table S1: Cluster table from RepeatExplorer with information on cluster and supercluster numbers, number of reads, genomic proportions, and sequence annotation

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN_annotation	Final_annotation
1	8	4249	1.01%	Putative satellites (high confidence)	DNAsat/158
2	7	3678	0.87%	Other	Class_I/LTR/Ty3_gypsy
3	1	3366	0.80%	Other	Class_I/LTR/Ty1_copia/SIRE
4	2	3260	0.77%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
5	3	2948	0.70%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
6	1	2779	0.66%	Other	Class_I/LTR/Ty1_copia/SIRE
7	13	2637	0.63%	Putative satellites (high confidence)	DNAsat/158
8	3	2509	0.60%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
9	1	2504	0.59%	Other	Class_I/LTR/Ty1_copia/SIRE
10	1	2422	0.57%	Other	Class_I/LTR/Ty1_copia/SIRE
11	10	2322	0.55%	Other	Class_I/LTR
12	15	2176	0.52%	Other	Class_I/LTR/Ty3_gypsy
13	4	2033	0.48%	Other	Class_I/LTR/Ty3_gypsy
14	4	1967	0.47%	Other	Class_I/LTR/Ty3_gypsy
15	6	1795	0.43%	Other	rDNA/45S_rDNA
18	1	1772	0.42%	Other	Class_I/LTR/Ty1_copia/SIRE
17	5	1776	0.42%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
16	6	1787	0.42%	Other	rDNA/45S_rDNA
19	9	1729	0.41%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
20	3	1675	0.40%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
21	2	1664	0.39%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
22	6	1649	0.39%	Other	rDNA/45S_rDNA
23	1	1600	0.38%	Other	Class_I/LTR/Ty1_copia/SIRE
24	5	1585	0.38%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
25	1	1540	0.37%	Other	Class_I/LTR/Ty1_copia/SIRE
26	16	1527	0.36%	Putative satellites (high confidence)	DNAsat/360
28	2	1466	0.35%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
27	9	1470	0.35%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
29	12	1406	0.33%	Other	Class_I/LTR
30	17	1373	0.33%	Putative satellites (high confidence)	DNAsat/155
31	18	1371	0.33%	Putative satellites (high confidence)	DNAsat/365
33	14	1332	0.32%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
32	19	1343	0.32%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
34	12	1300	0.31%	Other	Class_I/LTR/Ty3_gypsy
36	11	1270	0.30%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
35	20	1285	0.30%	Putative satellites (high confidence)	DNAsat/567
37	2	1224	0.29%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
41	4	1126	0.27%	Other	Class_I/LTR
39	5	1154	0.27%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
40	23	1148	0.27%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
43	1	1110	0.26%	Other	Class_I/LTR/Ty1_copia/SIRE
42	14	1111	0.26%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
44	2	1104	0.26%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila

Table S1: continued

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN_annotation	Final_annotation
45	2	1100	0.26%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
48	26	1023	0.24%	Putative satellites (high confidence)	DNAsat/361
49	4	969	0.23%	Other	Class_I/LTR
50	27	956	0.23%	Other	Class_I/LTR/Ty3_gypsy
51	28	936	0.22%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
52	2	932	0.22%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
53	29	931	0.22%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
57	21	886	0.21%	Other	Class_I/LTR/Ty1_copia/SIRE
56	31	891	0.21%	Other	Class_I/LTR/Ty3_gypsy
58	33	879	0.21%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
55	11	905	0.21%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
59	34	874	0.21%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
60	35	866	0.21%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
61	36	857	0.20%	Other	Class_I/LTR/Ty1_copia
63	38	845	0.20%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
65	5	834	0.20%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
66	40	821	0.19%	Putative LTR elements	Class_I/LTR/Ty1_copia/Angela
68	2	805	0.19%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
67	41	808	0.19%	Putative satellites (high confidence)	DNAsat/361
69	42	770	0.18%	Other	Class_I/LTR/Ty1_copia/SIRE
71	44	759	0.18%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
72	45	747	0.18%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
73	46	746	0.18%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
70	43	764	0.18%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
74	47	700	0.17%	Other	Class_I/LTR/Ty3_gypsy
78	32	666	0.16%	Other	Class_I/LTR/Ty1_copia/SIRE
77	50	668	0.16%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
75	48	684	0.16%	Putative satellites (low confidence)	DNAsat/314
79	51	646	0.15%	Other	Class_I/LTR/Ty1_copia/Bianca
82	54	615	0.15%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
84	56	595	0.14%	Other	Class_I/LTR/Ty1_copia/Ikeros
85	57	595	0.14%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
86	58	574	0.14%	Other	Class_II/Subclass_1/TIR/PIF-Harbinger
88	60	547	0.13%	Other	Class_I/LTR/Ty1_copia/TAR
90	11	534	0.13%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
91	62	533	0.13%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
92	63	528	0.13%	Other	Unclassified
95	65	506	0.12%	Other	Class_I/LTR/Ty1_copia
93	7	523	0.12%	Other	Class_I/LTR/Ty3_gypsy
94	64	517	0.12%	Other	Class_I/LTR/Ty3_gypsy
99	69	491	0.12%	Other	Class_I/LTR/Ty3_gypsy
101	71	486	0.12%	Putative satellites (low confidence)	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
97	67	504	0.12%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
98	68	497	0.12%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
100	70	489	0.12%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
108	77	444	0.11%	Other	Class_I/LTR/Ty1_copia/SIRE
103	73	478	0.11%	Other	Class_I/LTR/Ty3_gypsy
104	3	468	0.11%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
107	76	447	0.11%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
105	74	461	0.11%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
102	72	483	0.11%	Other	Class_II/Subclass_1/TIR/Tc1-Mariner

Table S1: continued

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN_annotation	Final_annotation
106	75	457	0.11%	Putative satellites (high confidence)	DNAsat/744
113	82	401	0.10%	Other	Class_I/LTR/Ty1_copia/Ikeros
111	80	422	0.10%	Other	Class_I/LTR/Ty3_gypsy
110	79	435	0.10%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
116	10	399	0.09%	Other	Class_I/LTR
121	4	361	0.09%	Other	Class_I/LTR/Ty3_gypsy
119	2	371	0.09%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
115	84	400	0.09%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
117	85	381	0.09%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
118	86	371	0.09%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
125	91	348	0.08%	Other	Class_I/LTR/Ty1_copia/Ikeros
128	21	331	0.08%	Other	Class_I/LTR/Ty1_copia/SIRE
123	89	353	0.08%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
122	88	355	0.08%	Other	Unclassified
127	93	334	0.08%	Other	Unclassified
132	7	298	0.07%	Other	Class_I/LTR
133	97	294	0.07%	Other	Class_I/LTR/Ty1_copia
129	94	312	0.07%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
130	95	311	0.07%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
134	98	290	0.07%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila
131	96	300	0.07%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
136	100	280	0.07%	Putative satellites (high confidence)	DNAsat/314
144	108	249	0.06%	Other	Class_I/LTR/Ty3_gypsy
138	102	268	0.06%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
142	106	254	0.06%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
137	101	270	0.06%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
145	109	248	0.06%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
141	105	256	0.06%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
139	103	259	0.06%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
146	110	236	0.06%	Other	Class_II/Subclass_1/TIR/MuDR_Mutator
158	121	209	0.05%	Other	Class_I/LINE
151	115	218	0.05%	rDNA	Class_I/LTR/Cassandra
152	32	216	0.05%	Other	Class_I/LTR/Ty1_copia
154	117	216	0.05%	Other	Class_I/LTR/Ty1_copia
153	116	216	0.05%	Other	Class_I/LTR/Ty1_copia/Ikeros
161	124	201	0.05%	Other	Class_I/LTR/Ty1_copia/SIRE
155	118	212	0.05%	Other	Class_I/LTR/Ty3_gypsy
157	120	210	0.05%	Other	Class_I/LTR/Ty3_gypsy
164	127	194	0.05%	Other	Class_I/LTR/Ty3_gypsy
162	125	197	0.05%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
149	113	219	0.05%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
147	111	229	0.05%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
150	114	219	0.05%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
163	126	195	0.05%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
148	112	226	0.05%	Other	Class_II/Subclass_1/TIR/MuDR_Mutator
160	123	205	0.05%	Putative satellites (high confidence)	DNAsat/158
156	119	210	0.05%	Putative satellites (high confidence)	DNAsat/78
178	141	153	0.04%	Other	Class_I/LTR/Ty1_copia/Tork
165	128	184	0.04%	Other	Class_I/LTR/Ty3_gypsy
168	131	181	0.04%	Other	Class_I/LTR/Ty3_gypsy
173	136	163	0.04%	Other	Class_I/LTR/Ty3_gypsy
176	139	154	0.04%	Other	Class_I/LTR/Ty3_gypsy
170	133	172	0.04%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
180	143	150	0.04%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila

Table S1: continued

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN annotation	Final annotation
182	145	148	0.04%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
172	135	168	0.04%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
181	144	149	0.04%	Other	Class_I/SINE/SINE2
167	130	182	0.04%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
169	132	179	0.04%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
174	137	156	0.04%	Other	Class_II/Subclass_1/TIR/MuDR_Mutator
171	134	171	0.04%	Other	Class_II/Subclass_1/TIR/PIF-Harbinger
177	140	153	0.04%	Other	Class_II/Subclass_1/TIR/PIF-Harbinger
175	138	155	0.04%	Putative satellites (high confidence)	DNAsat/1529
179	142	150	0.04%	Other	Unclassified
183	146	148	0.04%	Other	Unclassified
193	156	132	0.03%	Other	Caulimovirus
186	149	142	0.03%	Other	Class_I/LTR/Ty1_copia
190	153	135	0.03%	Other	Class_I/LTR/Ty1_copia
196	159	129	0.03%	Other	Class_I/LTR/Ty1_copia
200	162	126	0.03%	Other	Class_I/LTR/Ty1_copia
206	168	109	0.03%	Other	Class_I/LTR/Ty1_copia
185	148	146	0.03%	Other	Class_I/LTR/Ty1_copia/Ale
199	1	127	0.03%	Other	Class_I/LTR/Ty1_copia/SIRE
188	151	138	0.03%	Other	Class_I/LTR/Ty1_copia/TAR
189	152	136	0.03%	Other	Class_I/LTR/Ty3_gypsy
192	155	134	0.03%	Other	Class_I/LTR/Ty3_gypsy
194	157	131	0.03%	Other	Class_I/LTR/Ty3_gypsy
201	163	123	0.03%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
204	166	118	0.03%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
205	167	110	0.03%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
198	161	129	0.03%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
207	169	107	0.03%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
197	160	129	0.03%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
184	147	146	0.03%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
187	150	141	0.03%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
191	154	134	0.03%	Other	Unclassified
195	158	130	0.03%	Other	Unclassified
202	164	121	0.03%	Other	Unclassified
203	165	118	0.03%	Other	Unclassified
221	182	83	0.02%	Other	Caulimovirus
208	10	105	0.02%	Other	Class_I/LTR
223	184	82	0.02%	Other	Class_I/LTR
238	199	72	0.02%	Other	Class_I/LTR
242	203	67	0.02%	Putative LTR elements	Class_I/LTR
246	207	64	0.02%	Other	Class_I/LTR
214	175	89	0.02%	Other	Class_I/LTR/Ty1_copia
229	190	79	0.02%	Other	Class_I/LTR/Ty1_copia
231	192	77	0.02%	Other	Class_I/LTR/Ty1_copia
237	198	72	0.02%	Other	Class_I/LTR/Ty1_copia/Ikeros
245	206	65	0.02%	Other	Class_I/LTR/Ty1_copia/Ikeros
209	170	105	0.02%	Other	Class_I/LTR/Ty3_gypsy
234	195	74	0.02%	Other	Class_I/LTR/Ty3_gypsy
235	196	73	0.02%	Other	Class_I/LTR/Ty3_gypsy
239	200	71	0.02%	Other	Class_I/LTR/Ty3_gypsy
215	176	89	0.02%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
225	186	80	0.02%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
226	187	80	0.02%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
247	208	64	0.02%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay

Table S1: continued

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN_annotation	Final_annotation
211	172	100	0.02%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
216	177	88	0.02%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
227	188	80	0.02%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Ogre
220	181	84	0.02%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
222	183	83	0.02%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
241	202	70	0.02%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
244	205	66	0.02%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
224	185	82	0.02%	Other	Class_II/Subclass_1/TIR/hAT
236	197	72	0.02%	Other	Class_II/Subclass_1/TIR/MuDR_Mutator
210	171	103	0.02%	Other	Class_II/Subclass_1/TIR/PIF-Harbinger
212	173	97	0.02%	Other	Class_II/Subclass_1/TIR/PIF-Harbinger
228	189	79	0.02%	Other	Class_II/Subclass_1/TIR/Tc1-Mariner
217	178	87	0.02%	Other	Class_II/Subclass_2/Helitron
230	191	78	0.02%	Other	Stowaway MITE
213	174	89	0.02%	Other	Unclassified
218	179	86	0.02%	Other	Unclassified
219	180	85	0.02%	Other	Unclassified
232	193	76	0.02%	Other	Unclassified
233	194	75	0.02%	Other	Unclassified
240	201	71	0.02%	Other	Unclassified
243	204	66	0.02%	Other	Unclassified
248	209	64	0.02%	Other	Unclassified
305	266	46	0.01%	Other	Class_I/LTR/Ty1_copia/Ale
250	211	63	0.01%	Other	Class_I/LTR/Ty1_copia/Ikeros
270	231	56	0.01%	Other	Class_I/LTR/Ty1_copia/Tork
271	232	56	0.01%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
272	233	55	0.01%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
313	274	43	0.01%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/CRM
307	268	45	0.01%	Other	Class_I/LTR/Ty3_gypsy/chromovirus/Tekay
273	234	55	0.01%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
279	240	52	0.01%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
288	249	49	0.01%	Other	Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Tat/Retand
258	219	60	0.01%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
289	250	49	0.01%	Other	Class_II/Subclass_1/TIR/EnSpm_CACTA
297	258	48	0.01%	rDNA	rDNA/5S_rDNA
249	210	63	0.01%	Other	Unclassified
251	212	62	0.01%	Other	Unclassified
252	213	62	0.01%	Other	Unclassified
253	214	62	0.01%	Other	Unclassified
254	215	61	0.01%	Other	Unclassified
255	216	61	0.01%	Other	Unclassified
256	217	61	0.01%	Other	Unclassified
257	218	60	0.01%	Other	Unclassified
259	220	60	0.01%	Other	Unclassified
260	221	60	0.01%	Other	Unclassified
261	222	60	0.01%	Other	Unclassified
262	223	59	0.01%	Other	Unclassified
264	225	57	0.01%	Other	Unclassified
265	226	57	0.01%	Other	Unclassified
266	227	57	0.01%	Other	Unclassified
267	228	57	0.01%	Other	Unclassified
268	229	56	0.01%	Other	Unclassified
269	230	56	0.01%	Other	Unclassified
274	235	54	0.01%	Other	Unclassified

Table S1: continued

Cluster	Supercluster	N. reads	Genomic [%]	TAREAN_annotation	Final_annotation
275	236	53	0.01%	Other	Unclassified
276	237	53	0.01%	Other	Unclassified
277	238	53	0.01%	Other	Unclassified
278	239	52	0.01%	Other	Unclassified
280	241	51	0.01%	Other	Unclassified
281	242	51	0.01%	Other	Unclassified
282	243	51	0.01%	Other	Unclassified
283	244	51	0.01%	Other	Unclassified
284	245	51	0.01%	Other	Unclassified
285	246	51	0.01%	Other	Unclassified
286	247	50	0.01%	Other	Unclassified
287	248	49	0.01%	Other	Unclassified
290	251	49	0.01%	Other	Unclassified
291	252	49	0.01%	Other	Unclassified
292	253	49	0.01%	Other	Unclassified
293	254	48	0.01%	Other	Unclassified
294	255	48	0.01%	Other	Unclassified
295	256	48	0.01%	Other	Unclassified
296	257	48	0.01%	Other	Unclassified
298	259	47	0.01%	Other	Unclassified
299	260	47	0.01%	Other	Unclassified
300	261	47	0.01%	Other	Unclassified
301	262	46	0.01%	Other	Unclassified
302	263	46	0.01%	Other	Unclassified
303	264	46	0.01%	Other	Unclassified
304	265	46	0.01%	Other	Unclassified
306	267	45	0.01%	Other	Unclassified
308	269	45	0.01%	Other	Unclassified
309	270	44	0.01%	Other	Unclassified
310	271	44	0.01%	Other	Unclassified
311	272	43	0.01%	Other	Unclassified
312	273	43	0.01%	Other	Unclassified
314	275	43	0.01%	Other	Unclassified
315	276	43	0.01%	Other	Unclassified
316	277	43	0.01%	Other	Unclassified

Table S2: Pairwise alignments of satellite monomers, with percentagem of sequence identity and alignment scores

Repeat 1	Repeat 2	Identity (%)	E.value	Max.Score	Satellite
Ur2x_CL2_158	Ud2x_CL5_158	100.00	1.65E-79	282	Sat1A
Ur2x_CL2_158	Ub4x_CL5_158	100.00	1.55E-73	262	Sat1A
Ur2x_CL2_158	Ub4x_CL128_158	91.22	2.99E-57	209	Sat1A
Ur2x_CL2_158	Ub2x_CL13_158	88.46	9.16E-45	168	Sat1A
Ur2x_CL2_158	Ub2x_CL4_158	100.00	5.07E-29	114	Sat1A
Ur2x_CL2_158	Ub2x_CL114_314	85.97	5.41E-35	134	Sat1A
Ur2x_CL2_158	Ub4x_CL23_158	89.66	1.66E-41	156	Sat1A
Ur2x_CL2_158	Ud4x_CL5_158	100.00	2.99E-38	145	Sat1A
Ur2x_CL2_158	Ub2x_CL137_158	87.50	1.89E-15	70.7	Sat1A
Ur2x_CL2_158	Ud4x_CL59_158	88.24	7.05E-21	87.8	Sat1A
Ur2x_CL2_158	Ub2x_CL56_314	71.88	2.63E-07	42.8	Sat1A
Ur2x_CL2_158	Ub2x_CL21_155	83.33	1.12E-05	38.3	Sat1A
Ud2x_CL5_158	Ur2x_CL2_158	100.00	1.65E-79	282	Sat1A
Ud2x_CL5_158	Ub4x_CL5_158	100.00	1.89E-72	259	Sat1A
Ud2x_CL5_158	Ub4x_CL128_158	91.28	8.57E-58	211	Sat1A
Ud2x_CL5_158	Ub2x_CL13_158	88.64	7.52E-46	171	Sat1A
Ud2x_CL5_158	Ub2x_CL114_314	80.56	0.002	31	Sat1A
Ud2x_CL5_158	Ub2x_CL4_158	100.00	4.16E-30	118	Sat1A
Ud2x_CL5_158	Ub4x_CL23_158	87.18	6.18E-09	49.1	Sat1A
Ud2x_CL5_158	Ud4x_CL5_158	100.00	2.99E-38	145	Sat1A
Ud2x_CL5_158	Ub2x_CL137_158	91.09	3.65E-37	142	Sat1A
Ud2x_CL5_158	Ud4x_CL59_158	88.24	7.05E-21	87.8	Sat1A
Ud2x_CL5_158	Ub2x_CL56_314	71.88	2.63E-07	42.8	Sat1A
Ud2x_CL5_158	Ub2x_CL21_155	83.33	1.12E-05	38.3	Sat1A
Ud4x_CL5_158	Ub2x_CL4_158	100.00	1.89E-72	259	Sat1A
Ud4x_CL5_158	Ub2x_CL114_314	88.46	4.44E-55	201	Sat1A
Ud4x_CL5_158	Ud4x_CL59_158	88.51	8.03E-52	191	Sat1A
Ud4x_CL5_158	Ub2x_CL137_158	88.24	6.17E-47	174	Sat1A
Ud4x_CL5_158	Ub2x_CL137_158	95.46	3.90E-05	36.5	Sat1A
Ud4x_CL5_158	Ub4x_CL5_158	100.00	2.62E-45	168	Sat1A
Ud4x_CL5_158	Ub4x_CL23_158	89.08	1.66E-41	157	Sat1A
Ud4x_CL5_158	Ur2x_CL2_158	100.00	2.99E-38	145	Sat1A
Ud4x_CL5_158	Ud2x_CL5_158	100.00	2.99E-38	145	Sat1A
Ud4x_CL5_158	Ub2x_CL13_158	85.85	9.79E-32	124	Sat1A
Ud4x_CL5_158	Ub4x_CL128_158	89.77	4.16E-30	119	Sat1A
Ud4x_CL5_158	Ub2x_CL56_314	74.55	6.18E-09	49.1	Sat1A
Ud4x_CL5_158	Ub2x_CL21_155	76.62	6.18E-09	48.2	Sat1A
Ub4x_CL5_158	Ur2x_CL2_158	100.00	1.55E-73	262	Sat1A
Ub4x_CL5_158	Ud2x_CL5_158	100.00	1.89E-72	259	Sat1A
Ub4x_CL5_158	Ud2x_CL5_158	100.00	0.006	28.3	Sat1A
Ub4x_CL5_158	Ub2x_CL4_158	99.07	8.03E-52	191	Sat1A
Ub4x_CL5_158	Ub4x_CL128_158	90.37	3.41E-50	186	Sat1A
Ub4x_CL5_158	Ub2x_CL114_314	85.24	7.52E-46	170	Sat1A
Ub4x_CL5_158	Ud4x_CL5_158	100.00	2.62E-45	168	Sat1A
Ub4x_CL5_158	Ub2x_CL13_158	87.18	1.04E-37	144	Sat1A
Ub4x_CL5_158	Ub4x_CL23_158	88.35	1.89E-34	132	Sat1A

Table S2: continued

Repeat 1	Repeat 2	Identity (%)	E.value	Max.Score	Satellite
Ub4x_CL5_158	Ub2x_CL137_158	89.66	1.45E-29	117	Sat1A
Ub4x_CL5_158	Ud4x_CL59_158	89.16	2.16E-27	110	Sat1A
Ub4x_CL5_158	Ub2x_CL56_314	73.50	1.77E-09	50.9	Sat1A
Ub4x_CL5_158	Ub2x_CL21_155	78.69	7.53E-08	45.5	Sat1A
Ub2x_CL4_158	Ud4x_CL5_158	100.00	1.89E-72	259	Sat1A
Ub2x_CL4_158	Ub2x_CL114_314	86.28	9.79E-13	61.7	Sat1A
Ub2x_CL4_158	Ub4x_CL5_158	99.07	8.03E-52	191	Sat1A
Ub2x_CL4_158	Ur2x_CL2_158	100.00	5.07E-29	114	Sat1A
Ub2x_CL4_158	Ud4x_CL59_158	87.79	3.20E-44	165	Sat1A
Ub2x_CL4_158	Ud2x_CL5_158	100.00	4.16E-30	118	Sat1A
Ub2x_CL4_158	Ub2x_CL137_158	88.24	2.02E-40	152	Sat1A
Ub2x_CL4_158	Ub4x_CL23_158	90.57	1.55E-16	74.3	Sat1A
Ub2x_CL4_158	Ub4x_CL128_158	90.59	4.16E-30	118	Sat1A
Ub2x_CL4_158	Ub2x_CL13_158	91.05	4.75E-23	95.1	Sat1A
Ub2x_CL4_158	Ub2x_CL56_314	74.55	6.18E-09	49.1	Sat1A
Ub2x_CL4_158	Ub2x_CL21_155	76.62	6.18E-09	48.2	Sat1A
Ud4x_CL13_361	Ub4x_CL33_361	100.00	0	652	Sat2A
Ud4x_CL13_361	Ub2x_CL20_360	95.47	2.50E-132	460	Sat2A
Ud4x_CL13_361	Ub4x_CL16_370	69.55	1.00E-29	119	Sat2A
Ud4x_CL13_361	Ub2x_CL24_567	71.19	0.004	31	Sat2A
Ud4x_CL13_361	Ub4x_CL202_566	71.43	2.21E-06	41	Sat2A
Ur2x_CL13_379	Ud2x_CL13_378	96.30	1.55E-179	616	Sat3
Ur2x_CL13_379	Ud4x_CL40_378	95.96	2.81E-138	479	Sat3
Ur2x_CL13_379	Ub4x_CL46_379	95.16	9.20E-113	394	Sat3
Ud2x_CL13_378	Ur2x_CL13_379	96.30	1.55E-179	616	Sat3
Ud2x_CL13_378	Ud4x_CL40_378	95.25	1.45E-135	470	Sat3
Ud2x_CL13_378	Ub4x_CL46_379	95.14	3.90E-111	389	Sat3
Ub2x_CL13_158	Ub4x_CL23_158	100.00	0.02	26.5	Sat1B
Ub2x_CL13_158	Ub4x_CL128_158	92.86	2.99E-57	208	Sat1B
Ub2x_CL13_158	Ud4x_CL59_158	96.52	8.03E-52	190	Sat1B
Ub2x_CL13_158	Ud2x_CL5_158	88.64	7.52E-46	171	Sat1B
Ub2x_CL13_158	Ub2x_CL114_314	85.07	9.16E-45	168	Sat1B
Ub2x_CL13_158	Ur2x_CL2_158	88.46	9.16E-45	168	Sat1B
Ub2x_CL13_158	Ub2x_CL137_158	88.28	1.12E-43	164	Sat1B
Ub2x_CL13_158	Ub4x_CL5_158	87.81	5.07E-10	52.7	Sat1B
Ub2x_CL13_158	Ud4x_CL5_158	85.85	9.79E-32	124	Sat1B
Ub2x_CL13_158	Ub2x_CL4_158	91.05	4.75E-23	95.1	Sat1B
Ub2x_CL13_158	Ub2x_CL56_314	72.06	0.000136	34.6	Sat1B
Ub2x_CL13_158	Ub2x_CL21_155	75.68	0.02	27.4	Sat1B
Ub4x_CL16_370	Ub4x_CL33_361	69.55	1.03E-29	119	Sat4
Ub4x_CL16_370	Ud4x_CL13_361	69.55	1.03E-29	119	Sat4
Ub4x_CL16_370	Ub2x_CL24_567	78.15	9.62E-24	98.7	Sat4
Ub4x_CL16_370	Ub2x_CL20_360	68.18	1.17E-22	95.1	Sat4
Ub4x_CL16_370	Ub4x_CL202_566	77.97	1.53E-08	49.1	Sat4
Ub4x_CL23_158	Ub2x_CL13_158	97.92	1.19E-68	247	Sat1B
Ub4x_CL23_158	Ud4x_CL59_158	96.55	6.18E-09	49.1	Sat1B

Table S2: continued

Repeat 1	Repeat 2	Identity (%)	E.value	Max.Score	Satellite
Ub4x_CL23_158	Ub2x_CL137_158	90.85	5.41E-54	198	Sat1B
Ub4x_CL23_158	Ub2x_CL114_314	88.31	5.41E-54	197	Sat1B
Ub4x_CL23_158	Ub4x_CL128_158	100.00	9.17E-07	41	Sat1B
Ub4x_CL23_158	Ud2x_CL5_158	89.83	1.36E-42	159	Sat1B
Ub4x_CL23_158	Ud4x_CL5_158	91.67	5.07E-10	52.7	Sat1B
Ub4x_CL23_158	Ur2x_CL2_158	87.81	5.07E-10	52.7	Sat1B
Ub4x_CL23_158	Ub4x_CL5_158	88.35	1.89E-34	132	Sat1B
Ub4x_CL23_158	Ub2x_CL4_158	90.57	1.55E-16	74.3	Sat1B
Ub4x_CL23_158	Ub2x_CL21_155	76.92	0.000136	33.7	Sat1B
Ub4x_CL23_158	Ub2x_CL56_314	72.22	0.02	27.4	Sat1B
Ub2x_CL20_360	Ub4x_CL33_361	95.47	2.50E-132	460	Sat2A
Ub2x_CL20_360	Ud4x_CL13_361	95.47	2.50E-132	460	Sat2A
Ub2x_CL20_360	Ub4x_CL16_370	68.18	1.14E-22	95.1	Sat2A
Ub2x_CL20_360	Ub2x_CL24_567	66.09	1.22E-09	52.7	Sat2A
Ub2x_CL20_360	Ub4x_CL202_566	66.50	1.48E-08	48.2	Sat2A
Ub2x_CL20_360	Ub2x_CL48_361	83.33	0.048	26.5	Sat2A
Ub2x_CL21_155	Ub2x_CL56_314	79.36	2.25E-33	129	Sat1C
Ub2x_CL21_155	Ud4x_CL5_158	76.62	6.05E-09	48.2	Sat1C
Ub2x_CL21_155	Ub2x_CL4_158	76.62	6.05E-09	48.2	Sat1C
Ub2x_CL21_155	Ub4x_CL5_158	78.69	7.37E-08	45.5	Sat1C
Ub2x_CL21_155	Ur2x_CL2_158	83.33	1.09E-05	38.3	Sat1C
Ub2x_CL21_155	Ud2x_CL5_158	83.33	1.09E-05	38.3	Sat1C
Ub4x_CL33_361	Ub4x_CL33_361	90.00	0.014	28.3	Sat2A
Ub4x_CL33_361	Ud4x_CL13_361	100.00	0	652	Sat2A
Ub4x_CL33_361	Ub2x_CL20_360	95.47	2.50E-132	460	Sat2A
Ub4x_CL33_361	Ub4x_CL16_370	69.55	1.00E-29	119	Sat2A
Ub4x_CL33_361	Ub2x_CL24_567	71.65	1.93E-13	64.4	Sat2A
Ub4x_CL33_361	Ub4x_CL202_566	71.43	2.21E-06	41	Sat2A
Ub4x_CL31_361	Ub2x_CL30_361	99.14	4.23E-180	618	Sat5A
Ub4x_CL31_361	Ub4x_CL125_525	85.71	2.21E-06	41.9	Sat5A
Ub2x_CL24_567	Ub4x_CL202_566	99.32	3.97E-151	522	Sat6
Ub2x_CL24_567	Ub4x_CL16_370	78.15	1.49E-23	98.7	Sat6
Ub2x_CL24_567	Ub4x_CL33_361	71.65	3.07E-13	64.4	Sat6
Ub2x_CL24_567	Ud4x_CL13_361	71.65	3.07E-13	64.4	Sat6
Ub2x_CL24_567	Ub2x_CL20_360	66.09	1.94E-09	52.7	Sat6
Ud4x_CL40_378	Ub4x_CL46_379	96.95	8.02E-158	544	Sat3
Ud4x_CL40_378	Ur2x_CL13_379	95.96	2.80E-138	479	Sat3
Ud4x_CL40_378	Ud2x_CL13_378	95.25	1.45E-135	470	Sat3
Ub4x_CL46_379	Ub4x_CL46_379	92.68	2.48E-12	61.7	Sat3
Ub4x_CL46_379	Ud4x_CL40_378	96.95	8.05E-158	544	Sat3
Ub4x_CL46_379	Ur2x_CL13_379	95.16	9.20E-113	394	Sat3
Ub4x_CL46_379	Ud2x_CL13_378	95.14	3.91E-111	389	Sat3
Ub2x_CL30_361	Ub4x_CL31_361	99.14	4.23E-180	618	Sat5A
Ub2x_CL30_361	Ub4x_CL125_525	85.71	2.21E-06	41.9	Sat5A
Ub4x_CL63_362	Ub2x_CL48_361	96.23	1.30E-110	388	Sat-2B
Ub2x_CL48_361	Ub4x_CL63_362	96.23	1.30E-110	388	Sat-2B

Table S2: continued

Repeat 1	Repeat 2	Identity (%)	E.value	Max.Score	Satellite
Ub2x_CL48_361	Ub2x_CL20_360	83.33	0.049	26.5	Sat-2B
Ud4x_CL59_158	Ub4x_CL23_158	99.23	3.19E-63	229	Sat1B
Ud4x_CL59_158	Ub2x_CL114_314	83.87	8.03E-33	127	Sat1B
Ud4x_CL59_158	Ub2x_CL137_158	90.97	4.44E-55	202	Sat1B
Ud4x_CL59_158	Ud4x_CL5_158	88.51	8.03E-52	191	Sat1B
Ud4x_CL59_158	Ub2x_CL13_158	96.52	8.03E-52	190	Sat1B
Ud4x_CL59_158	Ub2x_CL4_158	91.30	0.000136	33.7	Sat1B
Ud4x_CL59_158	Ub4x_CL128_158	93.75	8.58E-39	147	Sat1B
Ud4x_CL59_158	Ud2x_CL5_158	88.24	7.05E-21	87.8	Sat1B
Ud4x_CL59_158	Ur2x_CL2_158	89.54	5.07E-29	115	Sat1B
Ud4x_CL59_158	Ub4x_CL5_158	87.67	5.78E-22	92.4	Sat1B
Ud4x_CL59_158	Ub2x_CL56_314	83.33	0.02	26.5	Sat1B
Ub2x_CL85_744	Ud4x_CL85_744	99.73	0	1312	Sat-7
Ub2x_CL85_744	Ud2x_CL80_740	96.24	0	1039	Sat-7
Ub2x_CL85_744	Ud2x_CL80_740	98.13	1.72E-49	185	Sat-7
Ub2x_CL85_744	Ur2x_CL73_738	98.29	0	905	Sat-7
Ub2x_CL85_744	Ur2x_CL73_738	99.08	9.52E-110	385	Sat-7
Ub2x_CL85_744	Ub4x_CL79_744	100.00	0	874	Sat-7
Ub2x_CL85_744	Ub4x_CL79_744	100.00	2.91E-135	470	Sat-7
Ub2x_CL56_314	Ub2x_CL21_155	79.36	4.78E-33	129	Sat-1F
Ub2x_CL56_314	Ub4x_CL5_158	73.50	3.68E-09	50.9	Sat-1F
Ub2x_CL56_314	Ud4x_CL5_158	74.55	1.28E-08	49.1	Sat-1F
Ub2x_CL56_314	Ub2x_CL4_158	74.55	1.28E-08	49.1	Sat-1F
Ub2x_CL56_314	Ur2x_CL2_158	73.53	1.56E-07	45.5	Sat-1F
Ub2x_CL56_314	Ud2x_CL5_158	73.53	1.56E-07	45.5	Sat-1F
Ub2x_CL56_314	Ub2x_CL114_314	68.16	2.32E-05	37.4	Sat-1F
Ub2x_CL56_314	Ub2x_CL13_158	72.06	0.000283	34.6	Sat-1F
Ub2x_CL56_314	Ub4x_CL23_158	72.22	0.042	27.4	Sat-1F
Ub2x_CL56_314	Ud4x_CL59_158	83.33	0.042	26.5	Sat-1F
Ub4x_CL79_744	Ur2x_CL73_738	100.00	6.44E-17	77	Sat-7
Ub4x_CL79_744	Ub2x_CL85_744	100.00	2.91E-135	470	Sat-7
Ub4x_CL79_744	Ud4x_CL85_744	100.00	8.90E-142	491	Sat-7
Ub4x_CL79_744	Ud2x_CL80_740	98.37	0	636	Sat-7
Ub2x_CL80_498	Ub4x_CL99_498	98.98	7.24E-46	173	Sat-8
Ud4x_CL85_744	Ub2x_CL85_744	99.73	0	1312	Sat-7
Ud4x_CL85_744	Ud2x_CL80_740	100.00	3.80E-45	170	Sat-7
Ud4x_CL85_744	Ur2x_CL73_738	98.64	0	893	Sat-7
Ud4x_CL85_744	Ub4x_CL79_744	100.00	8.90E-142	491	Sat-7
Ud4x_CL84_364	Ud4x_CL84_364	100.00	0	657	Sat-9
Ud4x_CL84_364	Ub4x_CL115_364	100.00	5.56E-128	444	Sat-9
Ud2x_CL80_740	Ud4x_CL85_744	100.00	3.78E-45	170	Sat-7
Ud2x_CL80_740	Ub2x_CL85_744	98.13	1.72E-49	185	Sat-7
Ud2x_CL80_740	Ur2x_CL73_738	99.39	7.75E-168	578	Sat-7
Ud2x_CL80_740	Ub4x_CL79_744	94.72	1.50E-170	588	Sat-7
Ub4x_CL99_498	Ub2x_CL80_498	84.38	0.000131	36.5	Sat-8
Ub2x_CL114_314	Ud4x_CL59_158	89.74	1.78E-57	210	Sat1D

Table S2: continued

Repeat 1	Repeat 2	Identity (%)	E.value	Max.Score	Satellite
Ub2x_CL114_314	Ud4x_CL5_158	85.29	6.65E-06	40.1	Sat1D
Ub2x_CL114_314	Ub2x_CL137_158	89.80	1.12E-53	198	Sat1D
Ub2x_CL114_314	Ub4x_CL23_158	84.42	2.32E-43	163	Sat1D
Ub2x_CL114_314	Ub2x_CL4_158	87.18	4.78E-52	192	Sat1D
Ub2x_CL114_314	Ur2x_CL2_158	81.58	0.000283	34.6	Sat1D
Ub2x_CL114_314	Ud2x_CL5_158	83.54	6.64E-44	165	Sat1D
Ub2x_CL114_314	Ub2x_CL56_314	71.23	0.000283	34.6	Sat1D
Ur2x_CL73_738	Ub4x_CL79_744	98.43	0	1214	Sat-7
Ur2x_CL73_738	Ub2x_CL85_744	98.29	0	905	Sat-7
Ur2x_CL73_738	Ud4x_CL85_744	99.13	2.89E-116	407	Sat-7
Ur2x_CL73_738	Ud2x_CL80_740	96.87	0	688	Sat-7
Ub2x_CL137_158	Ud4x_CL59_158	90.97	4.44E-55	202	Sat1E
Ub2x_CL137_158	Ub4x_CL23_158	90.85	5.41E-54	198	Sat1E
Ub2x_CL137_158	Ub2x_CL114_314	89.80	5.41E-54	198	Sat1E
Ub2x_CL137_158	Ub4x_CL128_158	97.92	3.00E-19	83.3	Sat1E
Ub2x_CL137_158	Ud4x_CL5_158	88.24	6.17E-47	174	Sat1E
Ub2x_CL137_158	Ub2x_CL13_158	88.28	1.12E-43	164	Sat1E
Ub2x_CL137_158	Ud2x_CL5_158	91.09	3.65E-37	142	Sat1E
Ub2x_CL137_158	Ur2x_CL2_158	91.00	1.27E-36	141	Sat1E
Ub2x_CL137_158	Ub4x_CL5_158	89.66	1.45E-29	117	Sat1E
Ub4x_CL115_364	Ud4x_CL84_364	100.00	5.56E-128	444	Sat-9
Ub2x_CL133_78	Ub4x_CL125_525	96.15	3.61E-33	128	Sat5B
Ub4x_CL128_158	Ud2x_CL5_158	91.28	8.57E-58	211	Sat1E
Ub4x_CL128_158	Ur2x_CL2_158	91.22	2.99E-57	209	Sat1E
Ub4x_CL128_158	Ub2x_CL13_158	92.86	2.99E-57	208	Sat1E
Ub4x_CL128_158	Ub2x_CL137_158	97.92	3.00E-19	83.3	Sat1E
Ub4x_CL128_158	Ub4x_CL23_158	93.65	2.30E-52	192	Sat1E
Ub4x_CL128_158	Ub4x_CL5_158	91.30	0.000136	33.7	Sat1E
Ub4x_CL128_158	Ub2x_CL114_314	83.97	3.20E-44	166	Sat1E
Ub4x_CL128_158	Ud4x_CL59_158	93.75	8.58E-39	147	Sat1E
Ub4x_CL128_158	Ud4x_CL5_158	91.43	1.12E-24	100	Sat1E
Ub4x_CL128_158	Ub2x_CL4_158	90.14	1.36E-23	97.8	Sat1E
Ub4x_CL125_525	Ub2x_CL133_78	96.15	2.84E-32	128	Sat5B
Ub4x_CL125_525	Ub4x_CL31_361	85.71	3.24E-06	41.9	Sat5B
Ub4x_CL125_525	Ub2x_CL30_361	85.71	3.24E-06	41.9	Sat5B
Ub4x_CL202_566	Ub2x_CL24_567	99.32	3.97E-151	522	Sat6
Ub4x_CL202_566	Ub4x_CL16_370	78.33	6.76E-09	50.9	Sat6
Ub4x_CL202_566	Ub2x_CL20_360	72.17	2.36E-08	48.2	Sat6
Ub4x_CL202_566	Ub4x_CL33_361	71.43	3.50E-06	41	Sat6
Ub4x_CL202_566	Ud4x_CL13_361	71.43	3.50E-06	41	Sat6

Table S3: Summary of the TAREAN output with information on the identified satellites and corresponding species, accession, cluster number, read count, length, genomic proportion, and AT proportion.

Satellite	Species	Accession/Cultivar	Cluster	Count	Length	Genomic (%)	AT (%)
UroSat-1a	*U. brizantha 2x*	B105	CL4	4249	158	0.98	50
	U. brizantha 4x	Marandu	CL5	13196	158	1.53	50.63
	U. decumbens 2x	D04	CL5	6429	158	1.83	50.63
	U. decumbens 4x	Basilisk	CL5	15117	158	1.63	50.63
	U. ruziziensis 2x	CIAT26162	CL2	18648	158	3.4	50.63
UroSat-1b	*U. brizantha 2x*	B105	CL13	2637	158	0.61	53.8
	U. brizantha 4x	Marandu	CL23	3315	158	0.39	52.53
	U. decumbens 4x	Basilisk	CL59	1696	158	0.18	53.16
UroSat-1c	*U. brizantha 2x*	B105	CL21	1373	155	0.32	53.55
UroSat-1d	*U. brizantha 2x*	B105	CL56	684	314	0.16	50
UroSat-1e	*U. brizantha 2x*	B105	CL114	280	314	0.06	54.14
UroSat-1f	*U. brizantha 2x*	B105	CL137	205	158	0.05	54.43
	U. brizantha 4x	Marandu	CL128	319	158	0.04	53.8
UroSat-2a	*U. brizantha 2x*	B105	CL20	1527	360	0.35	56.11
	U. brizantha 4x	Marandu	CL33	2632	361	0.31	55.68
	U. decumbens 4x	Basilisk	CL13	6787	361	0.73	55.68
UroSat-2b	*U. brizantha 2x*	B105	CL48	808	361	0.19	51.8
	U. brizantha 4x	Marandu	CL63	1658	362	0.19	51.1
UroSat-3	*U. brizantha 4x*	Marandu	CL46	2075	379	0.24	60.69
	U. decumbens 2x	D04	CL13	2201	378	0.63	61.38
	U. decumbens 4x	Basilisk	CL40	2284	378	0.25	60.85
	U. ruziziensis 2x	CIAT26162	CL13	3742	379	0.69	62.53
UroSat-4	*U. brizantha 4x*	Marandu	CL16	4457	370	0.52	50.27
UroSat-5	*U. brizantha 2x*	B105	CL22	1371	365	0.32	58.08
UroSat-6a	*U. brizantha 2x*	B105	CL30	1023	361	0.24	55.96
	U. brizantha 4x	Marandu	CL31	2657	361	0.31	55.68
UroSat-6b	*U. brizantha 2x*	B105	CL133	210	78	0.05	51.28
	U. brizantha 4x	Marandu	CL125	337	525	0.04	51.43
UroSat-7	*U. brizantha 2x*	B105	CL24	1285	567	0.3	54.67
	U. brizantha 4x	Marandu	CL202	102	566	0.01	55.3
UroSat-8	*U. brizantha 2x*	B105	CL85	744	744	0.17	56.72
	U. brizantha 4x	Marandu	CL79	1243	744	0.14	56.72
	U. decumbens 4x	Basilisk	CL85	832	744	0.09	56.72
	U. decumbens 2x	D04	CL80	271	740	0.08	56.76
	U. ruziziensis 2x	CIAT26162	CL73	316	738	0.06	56.64
UroSat-9	*U. brizantha 2x*	B105	CL80	486	498	0.11	53.01
	U. brizantha 4x	Marandu	CL99	672	498	0.08	53.82
UroSat-10	*U. brizantha 4x*	Marandu	CL115	430	364	0.05	59.34
	U. decumbens 4x	Basilisk	CL84	863	364	0.09	59.62
UroSat-11	*U. brizantha 4x*	Marandu	CL213	97	472	0.01	72.88
	U. decumbens 4x	Basilisk	CL215	101	474	0.01	73.21

The satellite monomers used for fluorescent *in situ* hybridization probes are shown below. The bold underline letters are the region targeted by the primers.

A) UroSat-1a (158 bp)

ACACTGGAACACCCGAGCGAAACTCGTACCGGTCATTTTCGAAGTAACGAAATGGTAC
GAAACGCATCCAAACATGAGTCTTGGGTCTAATAAGGTGCATTGGGCG**CGTTCGTTGCG**
AAAAAGTTTCCGGAAGTTCGGTCGCCCAGAATGGTGCATTC

B) UroSat-2a (360 bp)

TAAAAAATCATGATAGCTCTATTCATGTAAATTTGCCTCGCCCTGTGTAGCAGACAACATT
CTGCATCCAATGGAAAAAATGCG**ATTGATTCGGGCTCCAAACG**AATAAGTTATGGAAA
TTCAGGGGGAAGTTAATCAATTTTGTGCATCGAAACCCACTCGGACGCCAAGACTATGGA
CTAAACGTTTGGTCCACAACCCGAAATTGACATATGAGTTAACATACCATGTCTTGGTGTT
CCTCGACT**ACTGCTTGAGATGTGGACCA**TCCATGGTAATGCGGTGGAGCCCTTCTCGTG
AGCTGCGACGCTGCATGCCCAAGGGGTACCTCTATGGACATGTGGGAGTTATTATTTTTTT
C

C) UroSat-3 (360 bp)

TTTCCTCACCAAAAAGCCATCTCACGTCAACCTCTTGGCCAGTAGGTCATCTCACACCCTC
CA**CGTCATGATGTCGTACTGCC**ATGTAGTTTAGGCGTCTATTTAGAGGCCGATAGGTCC
CTTCACGACCATTCTAGGGCCGATATATATCTATTTTCACAATTCCTAAATTTTTCCACCT
ATTTTTGCGATTTTCCAATGAAATAATATTATTTTAAAAATATTGATCTTTGATGTGACTTT
TTTTC**TCCAACACGAGCACTTTGT**TATATAACCGAATATATGCATTTACCATGGCCAAAA
ACGTTGAGATATCATCTCAAAAAAGAAAATAAAAAACAAAAGGACCTGTCGGCCATTGA
AGGGCCGACAGGTCCT

Figure S1: Probing of UroSat-1a (purple), UroSat-2a (red), UroSat-3 (blue) and 35S rDNA (green) on *U. brizantha* diploid (A, F) and tetraploid (B, G), *U. decumbens* diploid (C, I, J, N) and tetraploid (D, G, O), and *U. ruziziensis* (E, K, L, P). Bar: 10 µm.