



ALEX JÚNIOR APARECIDO SILVESTRINI

**ANÁLISE COMPARATIVA DO DNA REPETITIVO NUCLEAR
E DO GENOMA CLOROPLASTIDIAL DE ESPÉCIES DE
Cenchrus L. (Poaceae)**

**LAVRAS – MG
2020**

ALEX JÚNIOR APARECIDO SILVESTRINI

**ANÁLISE COMPARATIVA DO DNA REPETITIVO NUCLEAR E DO GENOMA
CLOROPLASTIDIAL DE ESPÉCIES DE *Cenchrus L.* (Poaceae)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Orientadora: Prof^a. Dr^a. Giovana Augusta Torres

Coorientadora: Prof^a. Dr^a. Magdalena Vaio Scvortzoff

**LAVRAS – MG
2020**

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

Silvestrini, Alex Junior Aparecido.

Análise comparativa do DNA repetitivo nuclear e do genoma cloroplastidial de espécies de *Cenchrus* L. (Poaceae) / Alex Junior Aparecido Silvestrini. - 2020.

100 p. : il.

Orientador(a): Giovana Augusta Torres.

Coorientador(a): Magdalena Vaio Scvortzoff.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2020.

Bibliografia.

1. Genética Vegetal. 2. Genômica Comparativa. 3. Evolução Genômica. I. Torres, Giovana Augusta. II. Scvortzoff, Magdalena Vaio. III. Título.

ALEX JUNIOR APARECIDO SILVESTRINI

**ANÁLISE COMPARATIVA DO DNA REPETITIVO NUCLEAR E DO GENOMA
CLOROPLASTIDIAL DE ESPÉCIES DE *Cenchrus L.* (Poaceae)**

**COMPARATIVE ANALYSIS OF NUCLEAR REPETITIVE DNA AND
CHLOROPLASTIDIAL GENOME OF *Cenchrus L.* (Poaceae) SPECIES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADA em 14 de outubro de 2020

Dr^a. Magdalena Vaio Scvortzoff – Universidad de la República - Uruguai

Dr. Luiz Gustavo Rodrigues Souza - UFPE

Prof^a. Dr^a. Giovana Augusta Torres

Orientadora

LAVRAS-MG

2020

À luta pelo acesso à educação pública, ciência e à vida de todxs frente ao desgoverno e a morte diária dos nossos

Dedico.

AGRADECIMENTOS

À Universidade Federal de Lavras e a toda sociedade por ter subsidiado minha formação.

Aos professores do Programa de Pós-graduação em Genética e Melhoramento de Plantas, em especial à minha orientadora, Professora Giovana, por estar me auxiliado desde os anos de iniciação científica, e através dos ensinamentos despertado ainda mais meu interesse pela área.

À Professora Magdalena, que mesmo distante me auxiliou na condução desse trabalho. Sem seu auxílio esse trabalho jamais teria sido concluído. Obrigado.

Aos técnicos do laboratório Felipe e Iara por terem possibilitado a realização dos experimentos necessários durante todos esses anos.

Aos colegas de laboratório pelo apoio e excelente convivência durante todos esses anos. A depender do rumo que irei seguir, vou sentir falta de cada um.

Aos amigos, antigos e novos que vieram com os anos de mestrado, Thaianni, Ana Luisa, Ana Gabriela, Othon, Mariana, Yasmin, Isabela, Marco Túlio, Maroca e todos outros. Mais uma vez, o apoio de vocês fez tudo muito mais fácil. Nossos roles e conversas vão ficar pra sempre marcados.

Aos membros da banca, Gustavo, Magdalena, Vânia e Ludmila por aceitarem participar da conclusão dessa importante etapa.

À minha irmã Bruna e a todos os meus familiares pelo apoio e incentivo.

Aos meus pais por serem a base disso tudo e por terem possibilitado que eu, um aluno oriundo do sistema de ensino básico público, chegasse até aqui.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Seremos sempre **RESISTÊNCIA**

RESUMO

Cenchrus L. é um importante gênero da família Poaceae que abriga diversas espécies de elevada importância agrônômica, principalmente forrageiras de destaque na pecuária nacional, como *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*. Diversos programas de melhoramento genético exploram a capacidade de produção e qualidade dessas forrageiras. Além disso, são interessantes para estudos genômico-evolutivos, uma vez que possuem genomas relativamente grandes e combinações poliploides complexas. Estudos genômicos utilizando sequenciamento de próxima geração (NGS) com baixa cobertura vem se mostrando uma interessante ferramenta para caracterização do germoplasma e estudos de evolução genômica. Esses estudos permitem desde a caracterização da fração repetitiva do genoma de uma espécie, até a montagem de seu genoma cloroplastidial. Sendo assim, o objetivo do presente trabalho é a análise genômica comparativa da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus*, duas espécies com genoma nuclear sequenciado, assim como do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*. Para tanto, o genoma de *C. purpureus* e *C. americanus* foi sequenciado em plataforma Illumina HiSeq™ 4000 e analisado na plataforma RepeatExplorer. Para a análise e montagem do genoma cloroplastidial, além dos genomas sequenciados, foi obtido o genoma cloroplastidial de *C. ciliaris* depositado sob o acesso MH286942 no banco de dados NCBI. A montagem e análise comparativa do cpDNA foram realizadas através de ferramentas implementadas no software Geneious Prime 2020.1.1 e em plataformas específicas como a CHLOROBOX – GeSeq. A fração repetitiva de *C. purpureus* e *C. americanus* corresponde, respectivamente, a 52,23 e 76,82% do genoma, e os elementos repetitivos mais abundantes em ambas são os LTRs. Foram ainda identificadas sequências de DNA satélite em ambos os genomas que correspondem a 2,55 e 4,17% do genoma de cada uma das espécies. A relação ancestral, assim como os ciclos de poliploidização-diploidização na evolução de ambas as espécies, tiveram um papel fundamental na composição da fração repetitiva das mesmas. Os processos evolutivos inerentes às sequências de DNA satélite são a base da diferenciação e amplificação dessa porção genômica de *Cenchrus purpureus* e *Cenchrus americanus*. O genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus cilairis* possui comprimento e composição muito similares, corroborando com a estreita relação filogenética entre essas espécies, assim como com a já conhecida conservabilidade desse genoma entre espécies de gramíneas, principalmente espécies de *Cenchrus*.

Palavras-chave: *Cenchrus*; DNA Repetitivo; Genoma Cloroplastidial; Evolução Genômica.

ABSTRACT

Cenchrus L. is an important genus of the Poaceae family that comprises several species of high agronomic importance, mainly forage prominent in national livestock, such as *Cenchrus purpureus*, *Cenchrus americanus* and, *Cenchrus ciliaris*. Several breeding programs explore the production capacity and quality of these forages. Besides, they are interesting for genomic-evolutionary studies, since they have relatively large genomes and complex polyploid combinations. Genomic studies using next-generation sequencing (NGS) with low coverage are an interesting tool for the characterization of germplasm and studies of genomic evolution. These studies allow characterization of the repetitive fraction of the genome and assembly of its chloroplastidial genome. Therefore, the objective of the present work is the comparative genomic analysis of the repetitive fraction of the genome of *Cenchrus purpureus* and *Cenchrus americanus*, as well as of the chloroplastidial genome of *Cenchrus purpureus*, *Cenchrus americanus* and, *Cenchrus ciliaris*. For that, the genomes of *C. purpureus* and *C. americanus* were sequenced on the Illumina HiSeq™ 4000 platform and analyzed using the platform RepeatExplorer. For the analysis and assembly of the chloroplastidial genome, in addition to sequenced genomes, the chloroplastidial genome of *C. ciliaris* deposited under access MH286942 in the NCBI database was also used. The assembly and comparative analysis of the cpDNA was performed using tools implemented in the Geneious Prime 2020.1.1 software and in specific platforms such as CHLOROBOX - GeSeq. The repetitive fraction of *C. purpureus* and *C. americanus* corresponds to 52.23 and 76.82% of the genome, respectively, and the most abundant repetitive elements in both are the LTRs. Satellite DNA sequences were also identified in both genomes that correspond to 2.55 and 4.17% of the genome of each species, respectively. The ancestral relationship, as well as the cycles of polyploidization-diploidization involved in evolution of both species, played a fundamental role in the composition of their repetitive fraction. The evolutionary processes inherent to satellite DNA sequences are the basis for the differentiation and amplification of this genomic fraction of *Cenchrus purpureus* and *Cenchrus americanus*. The chloroplastidial genomes of *Cenchrus purpureus*, *Cenchrus americanus* and, *Cenchrus ciliaris* are very similar either in length and or composition, corroborating to the close phylogenetic relationship among these species, as well as to the well-known conservability of its genome among grass species, mainly in *Cenchrus* species.

Keywords: *Cenchrus*; Repetitive DNA; Chloroplast Genome; Genome Evolution.

SUMÁRIO

PRIMEIRA PARTE.....	11
1. INTRODUÇÃO GERAL.....	11
2. REFERENCIAL TEÓRICO.....	13
2.1 Considerações botânicas, agronômicas e citogenômicas em <i>Cenchrus L.</i>	13
2.2 Sequências de DNA repetitivo	16
2.2.1 Sequências de DNA repetitivo dispersas - Elementos Transponíveis (ETs).....	17
2.2.1.1 Elementos Transponíveis de Classe I – Transposição por RNA	18
2.2.1.2 Elementos Transponíveis de Classe II – Transposição por DNA	19
2.2.2 Sequências de DNA repetitivo em tandem – DNA Satélite	20
2.3 Análise da fração repetitiva em genomas vegetais.....	22
2.4 Genoma cloroplastidial.....	25
2.4.1 Análise do genoma cloroplastidial em espécies de <i>Cenchrus L.</i>	27
REFERÊNCIAS BIBLIOGRÁFICAS	29
SEGUNDA PARTE.....	37
ARTIGO 1 - Genômica comparativa da fração repetitiva de <i>Cenchrus purpureus</i> e <i>Cenchrus americanus</i> (Poaceae).....	37
ARTIGO 2 - Análise comparativa do genoma cloroplastidial de três espécies relacionadas de <i>Cenchrus L.</i> : <i>Cenchrus purpureus</i> , <i>Cenchrus americanus</i> e <i>Cenchrus</i> <i>ciliaris</i> (Poaceae).....	74

PRIMEIRA PARTE

1. INTRODUÇÃO GERAL

Cenchrus L. (syn. *Pennisetum* Rich.) é um importante gênero da família Poaceae que possui 107 espécies cientificamente aceitas (KEW GARDENS SCIENCE, 2020), distribuídas ao longo de toda faixa tropical e subtropical do planeta (CLAYTON, 2006). A mais recente circunscrição taxonômica do gênero resultou da unificação dos gêneros *Cenchrus* L., *Pennisetum* Rich. e *Odontelytrum* Hack., resultando na inclusão de todos os três gêneros em *Cenchrus* (DONADÍO et al., 2009; CHEMISQUY et al., 2010).

Cenchrus purpureus (Schumach.) Morrone (syn. *Pennisetum purpureum* Schumach.) ou Capim-elefante e *Cenchrus americanus* (L.) Morrone (syn. *Pennisetum glaucum* [(L.) R. Brown]) ou Milheto são espécies cultivadas do gênero que denotam considerável importância agronômica. A primeira por sua elevada produção de massa seca que pode ser destinada à alimentação de gado leiteiro, e a segunda pela considerável produção de grãos utilizados na alimentação humana e presença de características favoráveis ao uso forrageiro (HANNA, 1981; SOBRINHO et al., 2005).

Com o crescimento e expansão da pecuária leiteira (SOUZA et al., 2016), novas variedades são sempre requisitadas por produtores visando o atendimento à demanda de alimentação dos animais (PEREIRA et al., 2003). O melhoramento genético é a principal ferramenta voltada à produção de cultivares melhoradas e adaptadas às demandas cada vez mais exigentes, tanto relacionadas à alimentação humana, quanto animal (FALEIRO; FARIAS NETO; RIBEIRO JUNIOR, 2008).

Diversas estratégias de melhoramento já foram e ainda possuem potencial de emprego em espécies do gênero *Cenchrus*. Um interessante exemplo é a produção do híbrido interespecífico entre *Cenchrus purpureus* ($2n=4x=28$, A'A'BB) e *Cenchrus americanus* ($2n=2x=14$, AA). O híbrido hexaploide obtido após duplicação cromossômica artificial do híbrido triploide obtido do cruzamento, demonstrou grande potencial para utilização na alimentação do gado leiteiro (JAUHAR; HANNA, 1998). Além desse clássico exemplo, diversas outras espécies, tanto selvagens (OSGOOD et al., 1997) quanto domesticadas, como por exemplo *Cenchrus ciliaris* L., outra espécie de grande interesse na utilização como forrageira, possuem potencial para transferência de alelos favoráveis para espécies cultivadas dentro do gênero (SOBRINHO et al., 2005).

Estudos genômicos, principalmente os que envolvem dados oriundos de plataformas NGS - *Next-generation sequencing*, vêm se mostrando de grande valia para tais programas,

apresentando importantes resultados que auxiliam os melhoristas em suas tomadas de decisão (BRACHI et al., 2011; YAMASHITA et al., 2020), assim como também são fontes extremamente úteis para o subsídio de estudos em evolução genômica. Uma abordagem que vem sendo amplamente utilizada é o sequenciamento genômico de baixa cobertura, que permite a caracterização da fração repetitiva, assim como a montagem *de novo* ou por referência de genomas cloroplastidiais (RASMUSSEN; NOOR, 2009; LI et al., 2020).

A célula vegetal possui três genomas distintos: o nuclear, o plastidial e o mitocondrial. O genoma nuclear possui sequências de DNA que podem ser divididas em dois grupos: as sequências de cópia única, que são, em sua maioria, sequências codificadoras ou genes, e as sequências de DNA repetitivo, quando monômeros de uma mesma sequência se repetem diversas vezes (HESLOP-HARRISON; SCHMIDT, 2007). Em plantas, a fração repetitiva pode corresponder a até 90% de seu genoma total (WINTERFELD; RÖSER, 2007) e possui diversas funções, tanto estruturais quanto regulatórias.

Dentre essas funções, está a interferência de determinadas classes de elementos repetitivos em processos de regulação gênica, em rearranjos cromossômicos, e até mesmo em respostas genômicas a estímulos ambientais. Todas essas diferentes facetas, unidas ao dinamismo e conseqüente rápida evolução desses componentes (WICKER et al., 2007), conferem aos mesmos a habilidade de transferir maior plasticidade aos genomas (MEHROTRA; GOYAL, 2014). Como conseqüência de todo esse dinamismo, os elementos repetitivos podem interferir de maneira direta nos resultados obtidos por programas de melhoramento genético vegetal (WEBER et al., 2012), assim como no curso evolutivo das mais diversas espécies (RASKINA et al., 2008; HEITKAM et al., 2020).

Com o avanço das técnicas bioinformáticas tornou-se possível a caracterização da fração repetitiva dos mais diversos genomas (NOVÁK et al., 2013; 2020). Esse tipo de abordagem permite a descrição da composição e classificação de sequências repetitivas, assim como a análise comparativa dessa fração genômica entre diferentes espécies (NOVÁK et al., 2020). Os resultados dessas análises, são, de forma específica, extremamente úteis na caracterização de germoplasma e em estudos genômico-evolutivos (MACAS et al., 2007; NOVÁK et al., 2020; c).

Dentre as outras organelas que possuem um genoma próprio, se encontram os plastídios. Em plantas, o mais comum de ser estudado é o cloroplasto, que é a principal organela responsável pela fotossíntese. Estruturalmente, o genoma cloroplastidial possui três distintas regiões que reúnem desde genes relacionados ao processo fotossintético, até regiões que possuem sequências de DNA repetidas. Entre tais sequências, algumas são ainda

conhecidamente conservadas devido às baixas taxas de inversão e substituição de nucleotídeos (BOCK, 2007).

Diante de tais características e com o avanço das técnicas *NGS*, a montagem do cpDNA em espécies vegetais permite, desde a caracterização de determinado germoplasma, até análises genômicas comparativas entre espécies. Essas análises fornecem dados acerca da evolução do genoma cloroplastidial, assim como marcas específicas para espécie e até mesmo para cultivares (JOH et al., 2017; GUO et al., 2020). Essas marcas podem, por sua vez, ser utilizadas como marcadores moleculares auxiliando na manutenção e identificação de germoplasma, assim como subsidiando análises acerca das relações filogenéticas entre distintos taxos (SASKI et al., 2007; JOH et al., 2017; GUO et al., 2020; PARK et al., 2020).

Diante disso, os objetivos do presente trabalho são o estudo comparativo da fração repetitiva, assim como a montagem e análise comparativa do genoma cloroplastidial de espécies de *Cenchrus* a partir do sequenciamento genômico de baixa cobertura.

2. REFERENCIAL TEÓRICO

2.1 Considerações botânicas, agronômicas e citogenômicas em *Cenchrus* L.

Cenchrus L. é um amplo gênero pertencente à família Poaceae, subfamília Panicoideae, tribo Paniceae que possui espécies distribuídas ao longo de toda a faixa tropical e subtropical do planeta (GUTIÉRREZ, 2015). A mais recente circunscrição taxonômica do gênero resultou da unificação dos gêneros *Cenchrus*, *Pennisetum* Rich. e *Odontelytrum* Hack., a partir de análise morfológica e molecular com dados plastidiais e nucleares. Esse estudo resultou na inclusão de todos os três gêneros em *Cenchrus* (DONADÍO et al., 2009; CHEMISQUY et al., 2010).

Cenchrus possui plantas com comportamento reprodutivo sexual ou apomítico, ciclo de vida anual, bianual ou perene, hábito de crescimento cespitoso ou ereto, e diversas outras características morfológicas extremamente variáveis entre as espécies que o compõe (S. FILGUEIRAS, 1984; GUTIÉRREZ, 2015). Um exemplo dessa variabilidade são as diferenças morfológicas existentes entre as espécies cultivadas *Cenchrus americanus* (syn. *Pennisetum glaucum*) e *Cenchrus purpureus* (syn. *Pennisetum purpureum*).

Cenchrus americanus, também conhecido como milheto, é uma espécie de origem africana (TOSTAIN, 1998) pertencente ao pool gênico primário juntamente de outras duas espécies diploides: *Cenchrus violaceus* (Lam.) Morrone e *Pennisetum mollissimum* Hochst.

(HARLAN AND DE WET, 1971). Possui ciclo de vida anual, alta taxa de fecundação cruzada (JAUHAR; HANNA, 1998), crescimento ereto, caule único capaz de emitir diversos perfilhos, além de possuir lâminas foliares largas e inflorescência paniculada (ALCÂNTARA; BUFARAH, 1988). A espécie se destaca como forrageira por possuir excelente qualidade de matéria seca, resistência a patógenos e boa adaptação a condições de seca; sendo utilizada na alimentação humana devido à sua elevada produção de grãos (PAIVA, 2006)

Cenchrus purpureus, popularmente conhecido como capim-elefante, é originária do continente Africano (BRUNKEN, 1977), pertencente ao pool gênico secundário de *Cenchrus americanus* (HARLAN; THE WET, 1971) e possui ciclo de vida vegetativo perene associado ao modo de reprodução predominantemente alógamo. Apresenta porte elevado, hábito de crescimento cespitoso, colmo cilíndrico, ereto e capaz de formar touceiras, folhas largas e compridas, e inflorescência do tipo panícula primária terminal (ALCÂNTARA; BUFARAH, 1988; ALVIM, 2001). É uma excelente espécie forrageira por possuir elevado potencial produtivo, se adaptar a divergentes condições ambientais e possuir boa palatabilidade, vigor e persistência no campo (SOBRINHO et al., 2005).

Outras espécies do gênero também possuem importância econômica, como é o caso de *Cenchrus ciliaris* L., uma espécie perene, ereta e decumbente, com porte robusto (S. FILGUEIRAS, 1984) e comumente utilizada como forrageira por apresentar boa digestibilidade e alto valor nutritivo (S. FILGUEIRAS, 1984; CAVALCANTE DE OLIVEIRA, 1993).

Espécies selvagens também se destacam dentro do gênero, uma vez que podem servir como fonte de alelos para programas de melhoramento genético de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*. Essas espécies comumente possuem características agronomicamente desejáveis, como tolerância a estresses ambientais, hábito perene de crescimento, maior período vegetativo e até mesmo genes para apomixia (OSGOOD et al., 1997). *Cenchrus violaceus* (Lam.) Morrone (MARTEL et al., 2004), *Cenchrus nervosus* (Nees) Kuntze (ROBERT, 2011) e *Cenchrus orientalis* (Rich.) Morrone (HANNA; DUJARDIN, 1982) são exemplos que se enquadram nessa descrição.

Um interessante exemplo da aplicação do melhoramento genético dentro do gênero foi a produção de um híbrido interespecífico entre *Cenchrus purpureus* ($2n=4x=28$) e *Cenchrus americanus* ($2n=2x=14$). A progênie resultante ($2n=3x=21$) apresenta características agrônomicas intermediárias aos dois parentais e possui boa aceitação pelo rebanho leiteiro, no qual comumente é utilizado (JAUHAR; HANNA, 1998; HANNA et al., 1999). Porém, devido a sua condição triploide, com consequências em sua fertilidade, a utilização e propagação por sementes desse híbrido é inviável. Um método encontrado para restaurar, mesmo que

parcialmente essa fertilidade, foi a duplicação dos cromossomos da progênie utilizando agentes anti-mitóticos como a colchicina (SCHANK et al., 1993; HANNA et al., 1999; BARBOSA et al., 2007). Esse processo resultou em um híbrido sintético hexaploide e fértil, com grande potencial para utilização e propagação seminífera (SCHANK et al., 1993).

Diversos estudos citogenéticos envolvendo tanto os genitores, quanto a progênie resultante dessa hibridação foram realizados. *Cenchrus americanus* é uma espécie diploide com $2n = 2x = 14$ cromossomos, constituição genômica como AA e quantidade de DNA nuclear intermediária de 2,32 Gb (2,37 pg) (JAUHAR; HANNA, 1998; TECHIO et al., 2002; ANDRADE-VIEIRA, 2010). *Cenchrus purpureus* é uma espécie alotetraploide que possui $2n = 4x = 28$ cromossomos, número básico $X=7$ e constituição genômica A'A'BB. A quantidade de DNA nuclear da espécie, também intermediária, é de 2,22 Gb (2,27 pg) (JAUHAR; HANNA, 1998; TECHIO et al., 2002; BARBOSA; DAVIDE; PEREIRA, 2003; ANDRADE-VIEIRA, 2010; DOS REIS et al., 2014).

Graus de homologia cromossômica/genômica entre *Cenchrus purpureus* e *Cenchrus americanus* foram definidos por meio de hibridização *in situ* genômica por Dos Reis e colaboradores (2014). A hibridização de *C. purpureus* (A'A'BB) com sonda genômica (DNAG) de *C. americanus* (AA) revelou que metade dos cromossomos são intensamente marcados na maior parte de sua extensão, indicando se tratar do genoma A' homeólogo ao genoma A de milho. Nos demais cromossomos, provavelmente pertencentes ao genoma B, apenas a região centromérica apresentou sinal de hibridização, indicando similaridade entre as prováveis sequências repetitivas centroméricas dos genomas A e B.

Quando os cromossomos de *C. americanus* foram hibridizados com DNAG de *C. purpureus* todo o conjunto cromossômico da espécie apresentou diferentes intensidades de sinais positivos para homologia, corroborando com a hipótese de homeologia entre os genomas A e A'. O padrão de sinais se mostrou distribuído em toda a extensão cromossômica da espécie, porém com maior intensidade nas regiões centroméricas, indicando que A e A' também possuem sequências repetitivas centroméricas similares.

Segundo Dos Reis e colaboradores (2014) os genomas A e A' derivam de um ancestral comum que possuía genoma A pertencente ao pool gênico primário, junto de *Cenchrus molissimus* e *Cenchrus violaceus*. O genoma A' pode ser considerado um subgenoma de A que divergiu devido a alterações genômicas e estruturais. Os autores mostraram que o genoma B possui certo grau de homeologia com os genomas A e A', mas sua origem permanece desconhecida, necessitando de estudos mais aprofundados e específicos para maiores esclarecimentos.

Existem ainda relatos de *Cenchrus americanus* se tratar de um provável paleopoliploide (MARTEL et al., 2004), originado de um progenitor de constituição AA poliploide, que, após processo de diploidização, se comporta como um diploide (TECHIO; DAVIDE; PEREIRA, 2006). Assim como relatos de que *Cenchrus purpureus*, um aloploiploide recente, surgiu após a diferenciação de *Cenchrus americanus* a partir da hibridação do ancestral de genoma A com um progenitor de genoma B (MARTEL et al., 2004; YAN et al., 2020).

Yan e colaboradores (2020) propõe elevada similaridade entre cromossomos e conseqüentemente o genoma de *Cenchrus purpureus* para com *Cenchrus americanus*, corroborando com Dos Reis e colaboradores (2014), *Setaria italica* e *Brachypodium distachyon* indicando a possível existência de rearranjos cromossômicos, como processos de fusão, que podem ter levado a divergência dessas espécies para com seus ancestrais comuns. Segundo os autores, espécies de *Pennisetum* divergiram há cerca de 15 milhões de anos atrás após a diferenciação dessas para com espécies de *Setaria*, como *S. italica*. Yan e colaboradores ainda propõe que formas ancestrais de *Cenchrus americanus* e *Cenchrus purpureus*, espécies anteriormente alocadas em *Pennisetum*, divergiram há cerca de 15 milhões e 13,66 milhões de anos atrás, respectivamente, devido processos de seleção e domesticação impostos as mesmas. A forma tetraploide e atual de *Cenchrus purpureus* divergiu há cerca de 6,61 milhões de anos atrás, milhares de anos após a diferenciação de *Cenchrus americanus*.

2.2 Sequências de DNA repetitivo

O genoma nuclear de grande parte dos eucariotos é comumente composto por sequências de DNA de cópia única e sequências repetitivas. As últimas, conhecidas também como elementos repetitivos, são tidas como fragmentos de DNA que variam de 2 pb até mais de 10000 pb e que possuem múltiplas cópias no genoma, podendo corresponder a até 90% de seu tamanho total. Essas sequências são conhecidamente responsáveis pela grande variação no tamanho do genoma dos organismos, principalmente os vegetais (HESLOP-HARRISON; SCHMIDT, 2007; MEHROTRA; GOYAL, 2014).

Esse importante componente genômico, por muito tempo ignorado devido à falta de atribuições biológicas bem definidas (BISCOTTI; OLMO; HESLOP-HARRISON, 2015), atualmente vem sendo reconhecido como uma das mais intrigantes frações do genoma devido às diversas funções que pode desempenhar. A exemplo das atribuições que essas sequências desempenham está a regulação gênica através da transcrição de microRNAs, seu envolvimento

no pareamento dos cromossomos e nos rearranjos cromossômicos, e ainda sua relação com respostas genômicas a estímulos ambientais (MEHROTRA; GOYAL, 2014).

Devido a seu dinamismo, sequências de DNA repetitivo são constantemente modificadas em composição e abundância (WICKER et al., 2007). Tais modificações resultam em sua ampla diversidade e consequente variação. Devido a tal variação, tornou-se necessária uma classificação consistente para tais sequências. De forma geral, elas podem ser divididas em dois grandes grupos relacionados à sua organização no genoma: as sequências repetitivas dispersas e as sequências repetitivas em tandem (HESLOP-HARRISON; SCHMIDT, 2007).

2.2.1 Sequências de DNA repetitivo dispersas - Elementos Transponíveis (ETs)

Sendo os maiores componentes individuais do genoma eucarioto, os elementos transponíveis podem corresponder a cerca de 80% do DNA total de uma espécie, possuindo elevada abundância principalmente em espécies vegetais (WICKER et al., 2007). São definidos como fragmentos de DNA capazes de se moverem e se amplificarem no genoma através de diversos mecanismos de transposição. Essa capacidade de transposição, seu frequente acúmulo em algumas regiões genômicas, assim como a sua abundância, são fatores que possibilitam o dinamismo em relação a sua localização no genoma (FESCHOTTE; PRITHAM, 2007).

Todos esses fatores, aliados à enorme diversidade de famílias que compõem esse tipo de sequência, levaram Wicker e colaboradores (2007) a propor um sistema de classificação e nomenclatura unificado para elementos transponíveis, que além de classifica-los, facilita ainda na comparação entre diferentes elementos repetitivos dentro e entre espécies. Os autores classificaram os ETs em classe, subclasse, ordem, superfamília, família e subfamília.

A primeira divisão, denominada como classe, os divide de acordo com a presença ou ausência de um intermediário de RNA no processo de transposição. A subclasse os classifica em elementos que copiam a eles próprios e deixam uma cópia no local de origem, e os que deixam o local de origem e se inserem em outro, sem necessariamente deixar cópias. Em ordem, os ETs são classificados com relação a seu mecanismo de inserção, e na classificação de superfamília são diferenciados por características gerais, tais como sequências que DNA que codificam proteínas estruturais ou presença e tamanho de *TSDs* (*Target Site Duplication*), que são sequências curtas duplicadas em seu sítio de inserção. Nas famílias, os ETs são organizados de acordo com a conservabilidade de suas sequências, e as subfamílias com relação a dados filogenéticos (WICKER et al., 2007).

Na divisão em classes, os elementos transponíveis podem ser classificados como os de Classe I e os de Classe II.

2.2.1.1 Elementos Transponíveis de Classe I – Transposição por RNA

Nos elementos transponíveis incluídos na classe I, um fragmento de DNA é sintetizado via transcriptase reversa a partir de um RNA já transcrito, e esse novo fragmento de DNA é então capaz de se inserir em determinada região do genoma. Devido a esse processo de transposição, os ETs pertencentes a essa classe são também denominados de retrotransposons ou retroelementos, e são ainda os responsáveis pela maior porção dos elementos transponíveis presente nos genomas, uma vez que a cada ciclo de replicação novas cópias de determinados fragmentos de DNA são criadas pelo mecanismo de “cópia e cola” (HESLOP-HARRISON; SCHMIDT, 2007; JURKA et al., 2007; WICKER et al., 2007).

Como nos elementos transponíveis da classe I nenhum fragmento é cortado e transferido para outros sítios, mas apenas “copiados”, não houve necessidade de classificação dos mesmos em subclasses, mas apenas em ordens. Sendo assim, esses elementos foram divididos em cinco ordens de acordo com suas características mecânicas, organizacionais e filogenéticas: os retrotransposons LTR (*Long Terminal Repeat*), DIRS (*Dictyostelium Intermediate Repeat Sequence*), Penelope-Like Elements (PLEs), LINEs (*Long Interspersed Nuclear Elements*) e por fim os SINEs (*Short Interspersed Nuclear Elements*) (WICKER et al., 2007).

Os retrotransposons LTR possuem três ORFs (*Open Reading Frames*) responsáveis pela codificação das proteínas gag, env e pol que estão envolvidas no processo de transposição e inserção do fragmento através das integrases. Essas ORFs são flanqueadas por longas repetições terminais que dão o nome a essa ordem (HESLOP-HARRISON; SCHMIDT, 2007). A ordem LTR é dividida ainda em duas superfamílias principais: *Copia* e *Gypsy*, que se diferem pela ordem em que as ORFs aparecem (HESLOP-HARRISON; SCHMIDT, 2007; WICKER et al., 2007). A divisão em famílias é extremamente extensa e dispersa em diversas espécies, principalmente as vegetais (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998).

Os membros da ordem DIRS não possuem um dos genes para *integrase*, que é comum e presente nas ORFs das outras ordens, além de possuir um gene que codifica uma proteína chamada *tirosina recombinase*, envolvida em seu mecanismo de integração ao genoma, e ausente em outras ordens (JURKA et al., 2007). Essa ordem pode ser ainda dividida em três superfamílias: *DIRS*, *Ngaro* e *VIPER* (WICKER et al., 2007).

A ordem *Penelope* se diferencia das demais, de maneira geral, por codificar uma única ORF e pela semelhança de sua transcriptase reversa com a telomerase, o que não ocorre em outras ordens. Essa ordem é composta ainda por uma única superfamília denominada como *Penelope* (JURKA et al., 2007; WICKER et al., 2007).

Os LINEs e os SINEs são semelhantes entre si (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998) e diferentes de outras ordens por não possuírem as longas sequências repetitivas terminais, sendo por isso chamados por alguns autores de elementos “*Non-LTR*” (HESLOP-HARRISON; SCHMIDT, 2007). Elementos inclusos na ordem LINE são autônomos no processo de transposição e podem ser divididos em cinco superfamílias: *R2*, *L1*, *RTE*, *I* e *Jockey*. Diferentemente dos LINEs, os SINEs são mais curtos e não são autônomos, dependendo, frequentemente, da maquinaria utilizada pelos LINEs para a realização de sua transposição. Os SINEs são divididos ainda em três superfamílias: *tRNA*, *7SL* e *5S* (JURKA et al., 2007). A maioria das ordens não são restritas a uma única espécie, sendo, na verdade, presente em grande parte dos eucariotos e procariotos estudados até o momento (WICKER et al., 2007).

2.2.1.2 Elementos Transponíveis de Classe II – Transposição por DNA

Os ETs presentes na classe II são conhecidos por realizarem sua transposição sem a necessidade de mediação de uma molécula de RNA, como no caso dos retrotransposons (BISCOTTI; OLMO; HESLOP-HARRISON, 2015). Devido a essa característica, estes são chamados apenas de transposons de DNA, uma vez que sua transposição envolve a excisão de um fragmento de DNA de um local de origem e sua consequente inserção em outro, sem deixar cópias. A transposase é a enzima responsável por catalisar tanto as reações de excisão quanto as de inserção (JURKA et al., 2007), e os elementos dessa classe estão presentes, assim como os da classe I, em diversos eucariotos e procariotos (WICKER et al., 2007).

De acordo com a classificação proposta por Wicker e colaboradores (2007), os transposons pertencentes a essa classe podem ainda ser divididos em duas subclasses: Subclasse I e Subclasse II. Os inclusos na subclasse I são caracterizados pelo flanqueamento de longas repetições terminais chamadas de TIRs (*Terminal Inverted Repeats*) (JURKA et al., 2007), o qual dá o nome à ordem em que esses elementos estão inclusos. As TIRs são reconhecidas pela transposase para realização dos cortes na dupla fita de DNA com a posterior transposição do ET. Os elementos inclusos nessa ordem podem ainda ser divididos em nove superfamílias que se diferenciam pelo tamanho das repetições terminais invertidas e das duplicações do sítio alvo:

Tcl-Mariner, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger* e *CACTA*. A maioria dessas superfamílias apresentam apenas uma ORF, que codifica para a transposase (WICKER et al., 2007).

A Subclasse II comporta elementos em que o processo de transposição não envolve a clivagem da dupla fita de DNA, mas sim, de apenas uma delas. Esses ETs podem ainda ser classificados em duas ordens: Helitron e Maverick, no qual ambas abrigam apenas uma superfamília cada, denominadas, respectivamente, de *Helitron* e *Maverick* (WICKER et al., 2007). As duas ordens se diferenciam pelo mecanismo de transposição e por suas sequências flanqueadoras. Os elementos inclusos em Helitron se transpõem através do mecanismo de círculo rolante, que gera uma molécula de DNA de fita dupla replicada, mais especificamente o transposon, e outra molécula de fita simples, e não possuem as TIRs. A ordem dos elementos transponíveis Maverick, também conhecida como Políntrons, possui elementos flanqueados por sequências terminais invertidas e que se transpõem pela excisão de uma única fita de DNA que passa por replicação extracromossômica e posteriormente é reinserida no genoma (JURKA et al., 2007; WICKER et al., 2007).

2.2.2 Sequências de DNA repetitivo em tandem – DNA Satélite

O DNA repetitivo distribuído em tandem é constituído por monômeros arranjados sequencialmente e são frequentemente conhecidos como DNA satélite (HESLOP-HARRISON; SCHMIDT, 2007; GARRIDO-RAMOS, 2017). Esse tipo de elemento repetitivo pode ser dividido de acordo com seu comprimento, suas propriedades e distribuições genômicas em: DNA microssatélite, minissatélite e satélite (GARRIDO-RAMOS, 2017).

Na classe dos DNAs microssatélites, também conhecidos como *Simple Sequence Repeats* (SSRs), as sequências repetidas variam de 1 a 5 pb e correspondem, junto com os minissatélites, que possuem de 10 a 50 pb de comprimento, a uma grande porção da região repetitiva do genoma de diversos eucariotos. Já as sequências de DNA satélite são constituídas por monômeros que possuem de 150 a 400 pb repetidos de centenas a milhares de cópias ao longo do genoma de diversas espécies (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998; GARRIDO-RAMOS, 2017).

O número de repetições dos monômeros que compõem as sequências repetitivas dos DNAs microssatélites são extremamente variáveis dentro e entre espécies. Em contraste com essa variabilidade, os sítios que flanqueiam essas sequências, por sua vez, costumam ser altamente conservados. Sendo assim, essas duas características permitem que o mesmo seja

amplamente utilizado como marcador molecular, tanto na diferenciação de variedades quanto no estabelecimento de relações filogenéticas entre espécies (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998, HESLOP-HARRISON; SCHMIDT, 2007). Esse tipo de elemento repetitivo é ainda o componente majoritário das regiões teloméricas de cromossomos eucariotos, sendo assim imprescindível para a proteção e estabilização dessas estruturas (ZAKIAN, 1995; GARRIDO-RAMOS, 2017).

Os satélites de DNA são elementos que possuem comprimento variável e que estão constantemente associados a regiões de heterocromatina, como os centrômeros e subtelômeros (MEHROTRA; GOYAL, 2014). Devido a essa localização, tais elementos são frequentemente ligados a funções de proteção e estabilização da estrutura cromossômica, assim como ocorre com alguns microssatélites (GARRIDO-RAMOS, 2017).

Muitas famílias de DNA satélite já foram descritas, e em grande parte das vezes sua distribuição genômica ou suas sequências de DNA são específicas para um determinado grupo de espécies, indivíduos ou até mesmo cromossomo-específicas (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998, GARRIDO-RAMOS, 2017). É comum que em apenas um genoma hajam diversas famílias que partilham entre si apenas o arranjo em tandem e a localização heterocromática de suas sequências (PLOHL, 2010, GARRIDO-RAMOS, 2017).

Por mais que algumas dessas famílias possam ser específicas para determinados grupos de espécies, é conhecido que alguns tipos de repetições apresentam sequências conservadas e de função bem conhecida entre diferentes reinos, como é o caso dos DNAs ribossômicos e até mesmo de algumas sequências teloméricas (HESLOP-HARRISON, 2000). Essas características, associadas à rápida evolução desse tipo de sequência, faz com que a mesma possa ser amplamente utilizada em estudos taxonômicos e filogenéticos, principalmente em espécies vegetais, no qual esses elementos são abundantes (MEHROTRA; GOYAL, 2014).

As classes de DNA satélite por muito tempo foram consideradas como entraves no processo de sequenciamento genômico por serem altamente repetidas, sendo muitas das vezes sub representadas em diversas análises bioinformáticas (MEHROTRA; GOYAL, 2014). Porém, com o recente desenvolvimento de softwares e algoritmos específicos para estudos relacionados a esse tipo de sequência (NOVÁK et al., 2013), uma nova visão acerca desses elementos repetitivos surgiu.

Assim, a união de dados obtidos através de sequenciamento com a localização física de alguns elementos repetitivos em cromossomos vegetais se mostra uma excelente ferramenta para se explorar a composição, organização e evolução genômica em diversas espécies, assim como para fornecer excelentes marcas cromossômicas para uma cariotipagem mais precisa.

Alguns trabalhos já relatam a utilização de tais ferramentas e demonstram a enorme aplicabilidade que as mesmas possuem (MACAS; NEUMANN; NAVRÁTILOVÁ, 2007; MACAS et al., 2011; HEITKAM et al., 2020).

2.3 Análise da fração repetitiva em genomas vegetais

Estudos genômicos possuem diversas aplicações, e em sua grande maioria requerem o sequenciamento do DNA nuclear total de determinado organismo. As plataformas de sequenciamento NGS (*Next-generation sequencing*) permitem esse tipo de abordagem, uma vez que são capazes de gerar elevada quantidade de dados em um curto período de tempo. Tais plataformas vêm sendo amplamente utilizadas em estudos envolvendo espécies vegetais, com diversos trabalhos de montagem e análise de genomas sendo cada mais frequentes devido ao baixo custo e facilidade de acesso a essas tecnologias (LI et al., 2019; WANG et al., 2019; LI et al., 2020).

Porém, mesmo sendo capazes de gerar elevada quantidade de dados através das plataformas NGS, os pesquisadores responsáveis pela montagem desses genomas encontram determinadas limitações ao se depararem com genomas muito grandes, que é o caso da maioria das espécies vegetais (CARVALHO; SILVA, 2010). Essa grandeza se deve principalmente à quantidade exacerbada de sequências repetitivas que compõem esses genomas (HESLOP-HARRISON; SCHMIDT, 2007; MEHROTRA; GOYAL, 2014). Mesmo tida como um empecilho para a montagem *in silico* de genomas (TREANGEN; SALZBERG, 2012), a fração repetitiva constantemente chama a atenção de pesquisadores, principalmente por permitir sua análise através do sequenciamento de baixa cobertura (RASMUSSEN; NOOR, 2009) aliada à utilização de recentes softwares e pipelines especificamente desenvolvidas (SAHA et al. 2008; NOVÁK et al. 2013).

Uma dessas pipelines é o *RepeatExplorer*, que possui ferramentas que viabilizam a análise da fração repetitiva do genoma de diversos organismos. O software possui ferramentas para pré-processamento de dados e para realização de análise de *clustering* através da comparação *all-to-all* dos *reads* gerados pelo sequenciamento. Esses *reads* são agrupados com base em sua similaridade de sequências em clusters, gerando gráficos que permitem identificar a qual elemento repetitivo determinado conjunto de sequências pertence. A análise e identificação dos gráficos é feita pela própria ferramenta de *clustering*. É possível também realizar anotação e análise manual dos resultados buscando similaridades através de ferramentas BLAST (NOVÁK et al., 2013).

Diversos trabalhos utilizando diferentes abordagens já foram realizados para caracterização da fração repetitiva em espécies vegetais. A exemplo de *Pisum sativum* L., no qual os autores identificaram maior proporção de elementos transponíveis do que de DNA satélite e ainda propuseram alta similaridade desses elementos com os presentes no genoma de *Medicago truncatula* (MACAS; NEUMANN; NAVRÁTILOVÁ, 2007). Assim como Macas e colaboradores (2015), utilizando do pipeline *RepeatExplorer*, caracterizaram a fração repetitiva e realizaram inferências evolutivas através de dados genômicos de 23 espécies pertencentes a tribo Fabae.

Como uma das principais finalidades da identificação de elementos repetitivos, a localização citogenética dos mesmos possui destaque em diversos trabalhos. A exemplo de Heitkam e colaboradores (2020) que conseguiram definir sequências de DNA satélite como marcas específicas de genomas e sub genomas nos cromossomos de espécies de *Chenopodium* L., auxiliando na cariotipagem dessas espécies, além de obter inferências acerca da evolução genômica no gênero.

Liu e colaboradores (2019), com o objetivo de elucidar questões acerca da evolução genômica e cromossômica em espécies do gênero *Avena* L., analisaram a fração repetitiva de oito espécies através da plataforma *RepeatExplorer*. Em seus resultados, os autores relataram o papel da dinâmica das sequências repetitivas em modificações cromossômicas estruturais que contribuíram para a diferenciação de espécies ancestrais de *Avena* para a espécie *Avena sativa* cultivada nos dias atuais.

Em *Cenchrus*, existem relatos de estudos envolvendo sequências de DNA repetitivo. Ingham e colaboradores (1993) identificaram dois monômeros de DNA repetitivo em *tandem* através de hibridização com posterior eletroforese em gel de agarose dos primers para sequências da família repetitiva *KpnI* em diferentes espécies do gênero. *Cenchrus purpureus* e *Cenchrus squamulatus* possuem um monômero com 140 pb, assim como outro de 160 pb. Em *Cenchrus americanus* foi encontrada apenas a sequência de 140 pb, enquanto que *Cenchrus hohenackeri* (Hochst. ex Steud.) Morrone, *Cenchrus setaceus* e *Cenchrus longisetus* M.C.Johnst, apresentaram apenas o monômero com 160 pb. Os autores propuseram que os monômeros correspondem a 17,4% do genoma nuclear de *Cenchrus purpureus* e 20,3% em *Cenchrus americanus*, indicando que nessa última, tal elemento repetitivo é o mais abundante. Ingham e colaboradores ainda propõem, através de análises comparativas, que o monômero de 140 pb evoluiu a partir do 160 pb por perda de pares de bases.

Kamm e colaboradores (1994) realizaram uma caracterização mais aprofundada do mesmo monômero de 140 pb isolado por Ingham e colaboradores (1993) denominando o

mesmo como *HAE III*. Um dos principais resultados desse trabalho foi o relato da localização centromérica do mesmo em todos os cromossomos de *Cenchrus americanus*. Os autores ainda identificaram a presença desse mesmo satélite em *Cenchrus violaceus* (Lam.) Morrone e *Cenchrus mollissimus* (Lam.) Morrone através de técnicas moleculares. Elamein e colaboradores (2007) relataram a localização centromérica de *HAE III* para seis espécies de *Cenchrus*, dentre as quais a confirmação em *C. violaceus* e *C. mollissimus* havia sido realizada por Kamm e colaboradores (1994).

Mais recentemente, Ishii e colaboradores (2012) corroboraram com o proposto por Elamein e colaboradores (2007) e propuseram a deleção de *HAE III* para *Cenchrus polystachios* (L.) Morrone e *Cenchrus pedicellatus* (Trin.) Morrone durante a especiação das mesmas. Os autores ainda propuseram uma nova nomenclatura para *HAE III* como *Gla's cen* e também identificaram quatro novas sequências satélite presentes em região subtelomérica dos cromossomos de *Cenchrus orientalis*, *Cenchrus longisetus* e *Cenchrus setaceus*. *Cenchrus americanus* apresentou apenas a sequência *Gla's cen* em seu genoma.

Wang, Chengran e colaboradores (2018b), utilizando de dados NGS, realizaram sequenciamento genômico de levantamento (*Survey Sequencing*) em *Cenchrus purpureus* visando identificação de marcadores SSR (*Simple Sequence Repeats*). Dentre os resultados, os autores descreveram a composição da fração repetitiva da espécie, que compreende 71,36% de todo genoma. A classe dos retroelementos é a mais abundante, ocupando cerca de 9,36% do genoma nuclear. Dentro dos retroelementos, os encontrados em maior quantidade são os LTR, mais especificamente os pertencentes à família *Gypsy/DIRS1*. Seguindo dos retrotransposons, estão os transposons, presentes em cerca de 3,66% do genoma. Em relação a classe dos DNAs satélite, os autores relataram apenas o sucesso em encontrar marcadores SSR dentre 98.480 sequências prováveis.

Yan e colaboradores (2020) realizaram o sequenciamento e a montagem em cromossomos do genoma de *Cenchrus purpureus* através de Nanopore *long-reads* e tecnologia *Hi-C*. Com esses resultados os autores conseguiram definir *in silico* cromossomos específicos pertencentes aos genomas A' e B da espécie, além de realizarem a ancoragem de genes nesses genomas, gerando marcas genômico-específicas. Os autores descreveram ainda, de forma geral, a composição da fração repetitiva do genoma da espécie como abrangendo cerca de 60,74% de todo o genoma, com predominância de elementos *Ty3-gypsy* seguidos de *Copia*.

As proporções genômicas da fração repetitiva relatadas por Yan e colaboradores (2020) diferem das relatadas por Wang, Chengran e colaboradores (2018b) pois as técnicas utilizadas para o sequenciamento diferem entre si. Wang, Chengran e colaboradores (2018b) realizaram

apenas um estudo de levantamento genômico em busca de marcadores moleculares, enquanto Yan e colaboradores (2020) realizaram a montagem completa do genoma através de recentes técnicas de sequenciamento que possuem ferramentas específicas para auxiliar na resolução da fração repetitiva de genomas, como a utilização de *long reads* e a técnica *Hi-C* (SCHREIBER; STEIN; MASCHER, 2018). Portanto, os dados de Yan e colaboradores (2020) são baseados em análises que se aproximam mais da real proporção da fração repetitiva em genomas estudados.

Varshney e colaboradores (2017) realizaram o sequenciamento genômico de diversos genótipos de *Cenchrus americanus* através da plataforma Illumina *HiSeq*TM 2000 visando a caracterização do germoplasma, assim como a obtenção de dados acerca de estrutura populacional, diversidade genética e domesticação da espécie. Entre os resultados, os autores relatam uma sucinta caracterização da fração repetitiva do genoma no qual corresponde a 77,2% de todo o DNA. Foi relatado ainda que a classe de elementos repetitivos presente de forma majoritária no genoma, estando de acordo com o relatado para grande parte das espécies vegetais, são os elementos transponíveis da classe LTR, sendo a superfamília *Gypsy* presente em maior quantidade quando comparada a *Copia*.

2.4 Genoma cloroplastidial

A célula vegetal possui, predominantemente, três organelas capazes de armazenar informação genética: o núcleo, a mitocôndria e os plastídios. Cada qual dessas estruturas são capazes, para além de armazenar, também de utilizar da informação genética de forma funcional, o que faz com que a célula vegetal possua no mínimo três diferentes genomas funcionais em sua constituição (WANG et al., 2018b).

O genoma nuclear é o mais bem documentado, sendo sua origem intimamente relacionada ao surgimento da primeira célula (GRAY, 1999). Esse genoma possui em sua constituição sequências codificantes e não codificantes e é o principal genoma diretamente envolvido no crescimento e desenvolvimento dos organismos eucariotos (HESLOP-HARRISON; SCHMIDT, 2007). Já os genomas mitocondriais e plastidiais possuem distintas origens e funcionalidades.

A teoria evolutiva amplamente aceita relacionada a origem mitocondrial e plastidial é a endossimbiótica. Essa teoria surgiu nos anos 1980 e postula que ambas organelas surgiram, em eventos evolutivos distintos, da união simbiótica entre uma célula primitiva e uma eubactéria

fotossintetizante (GRAY, 1999). Nessa teoria a eubactéria originou o que conhecemos hoje como cloroplasto, possuindo inclusive um genoma exclusivo e próprio.

O genoma plastidial está presente em uma gama de distintos plastídios, destacando-se dentre eles, os cloroplastídios ou cloroplastos. Assim como nas mitocôndrias (HANSON; FOLKERTS, 1992) esse genoma possui majoritariamente genes relacionados à manutenção da respiração celular, sendo que o genoma cloroplastidial possui genes envolvidos especificamente no processo fotossintético. Ambas as organelas possuem ainda padrões de herança não mendelianos e em grande parte das vezes relacionados à herança uni-parental (BOCK, 2007).

Nas plantas terrestres, de forma geral, o genoma cloroplastidial possui formato circular contendo em média 120 a 160 Kpb. Esse genoma pode ser ainda dividido em quatro partes: uma grande região denominada como *Large Single Copy Region* (LSC), uma segunda região, menor que a anterior, denominada como *Small Single Copy Region* (SSC) e as *Inverted Repeats Region* (IRa e IRb). Os genes presentes em LSC e SSC possuem apenas uma cópia, sendo que essas duas regiões são separadas pelas IRa e IRb, que por sua vez, são regiões repetidas e invertidas entre si, ou seja, genes contidos nessas regiões possuem duas cópias em um mesmo genoma cloroplastidial (BOCK, 2007).

Além da possibilidade de extração e sequenciamento desse genoma de forma específica, uma outra maneira de estudá-lo é através da montagem genômica utilizando dados de sequenciamento de próxima geração (NGS), tanto através de coberturas elevadas quanto em baixas coberturas. Projetos de sequenciamento, montagem e análise de genomas cloroplastidiais, utilizando sequenciamento genômico de baixa cobertura (LI et al., 2020), assim como sequenciamento específico de alta cobertura, já foram extensamente realizados nos mais diversos gêneros do reino vegetal, tais como *Zea* (MAIER et al., 1995), *Triticum* (OGIHARA et al., 2000) e *Sorghum* (DANG; PRING, 1986), dentre vários outros.

Uma das aplicações diretas desses estudos são trabalhos que examinam as relações filogenéticas entre espécies vegetais. Devido às baixas taxas de inversão e substituição de nucleotídeos, algumas regiões desse genoma são consideradas conservadas entre taxas e podem ser utilizadas para aferições acerca de parentesco e distâncias no ramo da filogenia vegetal (JANSEN; RUHLMAN, 2012).

Além de estudos filogenéticos, o sequenciamento do cpDNA aliado a estudos comparativos entre espécies permite aferições evolutivas acerca do próprio genoma cloroplastidial. Diversas modificações a nível estrutural e molecular como inversões e substituições de nucleotídeos, perda de sequências e adição de novos genes já foram relatadas em análises comparativas utilizando de cpDNA. Todos esses resultados são de extrema

importância para entendimento da evolução desse importante componente vegetal, assim como para melhores esclarecimentos acerca das relações e distâncias filogenéticas entre taxas através das análises filogenômicas (SASKI et al., 2007; LIN et al., 2010).

2.4.1 Análise do Genoma Cloroplastidial em Espécies de *Cenchrus L.*

Algumas espécies de *Cenchrus* possuem seu genoma cloroplastidial sequenciado e anotado, como é o caso de *Cenchrus ciliaris* (BHATT; THAKER, 2018), *Cenchrus americanus* (XU et al., 2019) e *Cenchrus purpureus* (WU; ZHOU, 2019), no qual as duas últimas estavam anteriormente alocadas no extinto gênero *Pennisetum* de acordo com o proposto por Chemisquy e colaboradores (2010).

A análise do genoma plastidial de *Cenchrus ciliaris*, realizada por Bhatt e colaboradores (2018), revelou que o cpDNA da espécie possui 138737 pb de comprimento com conteúdo G/C de 38,6%. A região LSC possui 81053 pb enquanto SSC possui 16108 pb e cada região IR possui 20788 pb. O número de genes codificantes relatado foi de 130. Em análises filogenéticas, a montagem do genoma plastidial de *Cenchrus ciliaris* demonstrou especificidade suficiente para distinguir essa espécie de outras espécies pertencentes a 4 diferentes subfamílias de Poaceae.

Wu e colaboradores (2019), por sua vez, sequenciaram e anotaram o cpDNA de *Cenchrus purpureus*. Como resultados, os autores demonstraram que o genoma cloroplastidial da espécie possui 138199 bp de comprimento divididos em 12386 pb na região SSC; 81161 pb na região LCS e 22326 pb nas regiões IRa e IRb. A porcentagem de G/C é de 38,6% e um total de 136 genes foram anotados. Nas análises filogenéticas realizadas entre *Cenchrus purpureus* e outras sete espécies do gênero, foi demonstrada a elevada proximidade filogenética da espécie com *Cenchrus americanus*. Esses resultados são congruentes com Martel e colaboradores (2004) e Dos Reis e colaboradores (2014) que propõem uma origem comum para essas duas espécies.

O sequenciamento do cpDNA de *Cenchrus americanus* também forneceu diversos dados acerca de sua composição e organização. Xu e colaboradores (2019) relatam, com uma descrição básica, que o genoma completo possui 138.119 pb de comprimento divididos em 12.409 pb na região SSC; 81.034 pb na região LCS e 22.308 pb nas IRa e IRb. O conteúdo de G/C nesse genoma é de 38,6% com o total de 110 genes codificantes. Análises filogenéticas também foram realizadas no qual os autores confirmaram a inserção da espécie na família Poaceae, subfamília Panicoideae. Raveendar e colaboradores (2019) que também descrevem o

genoma cloroplastidial de *Cenchrus americanus*, obtiveram dados similares aos de Xu e colaboradores (2019), corroborando com a descrição estrutural desse genoma.

Mesmo possuindo o genoma cloroplastidial sequenciado e anotado, *Cenchrus ciliaris*, *Cenchrus purpureus* e *Cenchrus americanus* possuem apenas estudos básicos e individuais em relação a estrutura e composição de seu cpDNA. Estudos comparativos e mais detalhados ainda não foram realizados, deixando em aberto questões acerca da evolução comparativa do genoma cloroplastidial dessas espécies que, atualmente, foram classificadas como estritamente relacionadas (CHEMISQUY et al., 2010).

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVIM, M. J. et al. Relatório técnico da Embrapa Gado de Leite 1999-2000. **Embrapa Gado de Leite-Docmentos (INFOTECA-E)**, 2001.
- ALCÂNTARA, P. B.; BUFARAH, G. Plantas forrageiras - gramíneas e leguminosas. São Paulo: **Nobel**, p.162, 1988.
- ANDRADE-VIEIRA, L. F. ANÁLISE CITOMOLECULAR EM HÍBRIDOS DE CAPIM-ELEFANTE E MILHETO. **Tese apresentada ao Programa Pós Graduação em Genética e Melhoramento de Plantas.** (*Pennisetum* sp. Schum., Poaceae). 2010. Universidade Federal de Lavras, 2010.
- BARBOSA, S., VIDE, L. C. D., PEREIRA, A. V., & ABREU, J. C. D. Duplicação cromossômica de híbridos triplóides de *Pennisetum purpureum* e *Pennisetum glaucum*. *Bragantia* v. 66, n. 3, 2007.
- BARBOSA, Sandro; DAVIDE, Lisete Chamma; PEREIRA, Antônio Vander. Cytogenetics of *Pennisetum purpureum* Schumack x *Pennisetum glaucum* L hybrids and their parents. **Cienc. agrotecnologia** v. 27, n. 1, p. 26–35, 2003.
- BHATT, P. P.; THAKER, V. S. The complete chloroplast genome of *Cenchrus ciliaris* (Poaceae). **Mitochondrial DNA Part B: Resources**, v. 3, n. 2, p. 674–675, 2018. doi: 10.1080/23802359.2018.1481795.
- BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Pat. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415–420, 2015. doi: 10.1007/s10577-015-9499-z.
- BOCK, R. Structure, function, and inheritance of plastid genomes. In: **Current**. [s.l.: s.n.]20 p. 29–63. 2007.
- BRACHI, BENJAMIN; MORRIS, GEOFFREY P.; BOREVITZ, JUSTIN O. Genome-wide association studies in plants: the missing heritability is in the field. **Genome biology**, v. 12, n. 10, p. 1-8, 2011.
- BRUNKEN, J. N. Morphometric variation and the classification of *Pennisetum* section Brevivalvula (Gramineae) in tropical Africa. **Botanical Journal of the Linnean Society**, v. 79, n. 1, p. 51-64, 1979.
- CARVALHO, M. C. da C. G. de; SILVA, D. C. G. da. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, n. 3, p. 735–744, 2010.
- CAVALCANTE DE OLIVEIRA, M. Capim buffel: Produção e Manejo nas regiões secas do Nordeste. **Petrolina: EMBRAPA/CPATSA (Circular Técnica)**, p. 18, 1993.
- CHEMISQUY, M. A.; GIUSSANI, L. M.; SCATAGLINI, M. A.; KELLOGG, E. A.; MORRONE, O. Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and

Odontelytrum (Poaceae): A combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. **Annals of Botany**, v. 106, n. 1, p. 107–130, 2010. doi: 10.1093/aob/mcq090.

DANG, L. H.; PRING, D. R. A physical map of the sorghum chloroplast genome. **Plant Molecular Biology**, v. 6, n. 2, p. 119–123, mar. 1986. doi: 10.1007/BF00027305.

DONADÍO, S.; GIUSSANI, L. M.; KELLOGG, E. A.; ZUOLAGA, F. O.; MORRONE, O. A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the trnL-F, rpl16 chloroplast markers. **Taxon**, v. 58, n. 2, p. 392–404, 2009.

DOS REIS, G. B.; MESQUITA, A. T.; TORRES, G. A.; ANDRADE-VIEIRA, L. F.; PEREIRA, A. Vander; DAVIDE, L. C. Genomic homeology between *Pennisetum purpureum* and *Pennisetum glaucum* (Poaceae). **Comparative Cytogenetics**, v. 8, n. 3, p. 199–209, 2014. doi: 10.3897/CompCytogen.v8i3.7732.

FALEIRO, F. G.; FARIAS NETO, A. L. De; RIBEIRO JUNIOR, W. Q. **Pré-melhoramento, melhoramento e pós-melhoramento: estratégias e desafios**. 1. ed. Planaltina: EMBRAPA Cerrados, 2008.

FESCHOTTE, C.; PRITHAM, E. J. DNA transposons and the evolution of eukaryotic genomes. **Annual review of genetics**, v. 41, p. 331–68, 2007. doi: 10.1146/annurev.genet.40.110405.090448.

GARRIDO-RAMOS, M. Satellite DNA: An Evolving Topic. **Genes**, v. 8, n. 9, p. 230, 18 set. 2017. doi: 10.3390/genes8090230.

GRAY, M. W. Evolution of organellar genomes. **Current Opinion in Genetics and Development**, v. 9, n. 6, p. 678–687, 1999. doi: 10.1016/S0959-437X(99)00030-1.

GUO, L., GUO, S., XU, J., HE, L., CARLSON, J. E., & HOU, X. Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. **Industrial Crops and Products**, v. 153, p. 112567, 2020.

GUTIÉRREZ, H. F. Las especies americanas del género *Cenchrus* S.L. (Poaceae, Panicoideae, Paniceae). **Darwiniana**, v. 3, n. 1, p. 114–200, 2015. doi: 10.14522/darwiniana.2015.31.641.

HANNA, W. W. et al. Melhoramento do *Pennisetum purpureum*. **Biologia e manejo do Pennisetum purpureum**. Juiz de Fora - Embrapa Gado de Leite p. 1–6, 1999.

HANNA, W. W. Method of reproduction in napier grass and in the 3X and 6X allopolyploid hybrids with pearl millet. **Crop Science**, Madison, v. 21, p. 123–126, 1981.

HANNA, W. W.; DUJARDIN, M. Apomictic Interspecific Hybrids Between Pearl Millet and *Pennisetum orientale* L. C. Rich. **Crop Science**, v. 22, n. 4, p. 857–859, 1982.

HANSON, M. R.; FOLKERTS, O. Structure and Function of the Higher Plant Mitochondrial Genome. **International Review of Cytology**, v. 141, n. C, p. 129–172, 1992. doi: 10.1016/S0074-7696(08)62065-3.

HARLAN, Jack R.; DE WET, Jan MJ. Toward a rational classification of cultivated plants. **Taxon**, p. 509-517, 1971.

HEITKAM, T.; WEBER, B.; WALTER, I.; LIEDTKE, S.; OST, C.; SCHMIDT, T. Satellite DNA landscapes after allotetraploidisation of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. **The Plant Journal**, n. March, p. tpj.14705, 24 jan. 2020. doi: 10.1111/tpj.14705.

HESLOP-HARRISON, J. (Pat); SCHMIDT, T. Plant Nuclear Genome Composition. **Encyclopedia of Life and Science**, p. 1–8, 2007. doi: 10.1002/9780470015902.a0002014.pub2.

HESLOP-HARRISON, J. S. Comparative Genome Organization in Plants: From Sequence and Markers to Chromatin and Chromosomes. **the Plant Cell Online**, v. 12, n. 5, p. 617–636, 2000. doi: 10.1105/tpc.12.5.617.

INGHAM, L. D.; HANNA, W. W.; BAIER, J. W.; HANNAH, L. C. Origin of the main class of repetitive DNA within selected Pennisetum species. p. 350–356, 1993.

ISHII, T.; MATSUMOTO, N.; TANAKA, H.; ELTAYEB, A. E.; TSUJIMOTO, H. Evolution of subtelomeric and centromeric repetitive sequences in genus *Pennisetum* (Poaceae). **Chromosome Science**, v. 15, n. 3+4, p. 53–59, 2012. doi: 10.11352/scr.15.53.

JAUHAR, P. P.; HANNA, W. W. Cytogenetics and genetics of pearl millet. **Advances in Agronomy**, San Diego, v. 64, p. 1-26, Aug. 1998.

JANSEN, R. K.; RUHLMAN, T. A. Plastid Genomes of Seed Plants. In: **Advances in Photosynthesis and Respiration**. [s.l: s.n.]35 p. 103–126. 2012.

JOH, H. J., KIM, N. H., JAYAKODI, M., JANG, W., PARK, J. Y., KIM, Y. C., ... & YANG, T. J. Authentication of golden-berry *P. ginseng* cultivar ‘Gumpoong’ from a landrace ‘Hwangsook’ based on pooling method using chloroplast-derived markers. **Plant Breeding and Biotechnology**, v. 5, n. 1, p. 16-24, 2017.

JURKA, J.; KAPITONOV, V. V.; KOHANY, O.; JURKA, M. V. Repetitive Sequences in Complex Genomes: Structure and Evolution. **Annual Review of Genomics and Human Genetics**, v. 8, n. 1, p. 241–259, set. 2007. doi: 10.1146/annurev.genom.8.080706.092416.

KALENDAR, R.; GROB, T.; REGINA, M.; SUONIEMI, A.; SCHULMAN, A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. **Theoretical and Applied Genetics**, v. 98, n. 5, p. 704–711, abr. 1999. doi: 10.1007/s001220051124.

KAMM, A.; SCHMIDT, T.; HESLOP-HARRISON, J. S. Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. **MGG Molecular & General Genetics**, v. 244, n. 4, p. 420–425, 1994. doi: 10.1007/BF00286694.

KUBIS, S.; SCHMIDT, T.; HESLOP-HARRISON, J. S. Repetitive DNA elements as a major component of plant genomes. **Annals of Botany**, v. 82, n. SUPPL. A, p. 45–55, 1998. doi: 10.1006/anbo.1998.0779.

LI, S. F., WANG, J., DONG, R., ZHU, H. W., LAN, L. N., ZHANG, Y. L., ... & GAO, W. J. Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant *Asparagus setaceus*. **Horticulture research**, v. 7, n. 1, p. 1-11, 2020.

LI, Q., LI, H., HUANG, W., XU, Y., ZHOU, Q., WANG, S., ... & ZHANG, Z. A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). **GigaScience**, v. 8, n. 6, p. giz072, 2019.

LI, B.; LIN, F.; HUANG, P.; GUO, W.; ZHENG, Y. Development of nuclear SSR and chloroplast genome markers in diverse *Liriodendron chinense* germplasm based on low-coverage whole genome sequencing. **Biological Research**, v. 53, n. 1, p. 1–12, 2020. doi: 10.1186/s40659-020-00289-0.

LIN, C. P.; HUANG, J. P.; WU, C. S.; HSU, C. Y.; CHAW, S. M. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. **Genome Biology and Evolution**, v. 2, n. 1, p. 504–517, 2010. doi: 10.1093/gbe/evq036.

LIU, Q.; LI, X.; ZHOU, X.; LI, M.; ZHANG, F.; SCHWARZACHER, T.; HESLOP-HARRISON, J. S. 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. 226, 30 dez. 2019. doi: 10.1186/s12870-019-1769-z.

MACAS, J.; KEJNOVSKÝ, E.; NEUMANN, P.; NOVÁK, P.; KOBLÍŽKOVÁ, A.; VYSKOT, B. Next Generation Sequencing-Based Analysis of Repetitive DNA in the Model Dioecious Plant *Silene latifolia*. **PLoS ONE**, v. 6, n. 11, p. e27335, 9 nov. 2011. doi: 10.1371/journal.pone.0027335.

MACAS, J.; NEUMANN, P.; NAVRÁTILOVÁ, A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: Comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. **BMC Genomics**, v. 8, 2007. doi: 10.1186/1471-2164-8-427.

MACAS, J.; NOVAK, P.; PELLICER, J.; CIZKOVA, J.; KOBLIZKOVA, A.; NEUMANN, P.; FUKOVA, I.; DOLEZEL, J.; KELLY, L. J.; LEITCH, I. J. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabaeae. **PLoS ONE**, v. 10, n. 11, p. 1–23, 2015. doi: 10.1371/journal.pone.0143424.

MAIER, R. M.; NECKERMANN, K.; IGLOI, G. L.; KOÖSSEL, H. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal of Molecular Biology**, v. 251, n. 5, p. 614–628, 1 ago. 1995. doi: 10.1006/jmbi.1995.0460.

MARTEL, E.; PONCET, V.; LAMY, F.; SILJAK-YAKOVLEV, S.; LEJEUNE, B.; SARR, A. Chromosome evolution of *Pennisetum species* (Poaceae): Implications of ITS phylogeny. **Plant Systematics and Evolution**, v. 249, n. 3–4, p. 139–149, 2004. doi: 10.1007/s00606-004-0191-6.

MEHROTRA, S.; GOYAL, V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. **Genomics, Proteomics and Bioinformatics**, v. 12, n.

4, p. 164–171, 2014. doi: 10.1016/j.gpb.2014.07.003.

NOVAES, E.; DROST, D. R.; FARMERIE, W. G.; PAPPAS, G. J.; GRATTAPAGLIA, D.; SEDEROFF, R. R.; KIRST, M.; KIRST, M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. **BMC genomics**, v. 9, p. 312, 30 jun. 2008. doi: 10.1186/1471-2164-9-312.

NOVÁK, P.; NEUMANN, P.; PECH, J.; STEINHAIŠL, J.; MACAS, J. 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. doi: 10.1093/bioinformatics/btt054.

NOVÁK, PETR; NEUMANN, PAVEL; MACAS, JIŘÍ. Global analysis of repetitive DNA from unassembled sequence reads using *RepeatExplorer2*. **Nature Protocols**, p. 1-32, 2020.

OGIHARA, Y.; ISONO, K.; KOJIMA, T.; ENDO, A.; HANAOKA, M.; SHIINA, T.; TERACHI, T.; UTSUGI, S.; MURATA, M.; MORI, N.; TAKUMI, S.; IKEO, K.; GOJOBORI, T.; MURAI, R.; MURAI, K.; MATSUOKA, Y.; OHNISHI, Y.; TAJIRI, H.; TSUNEWAKI, K. Chinese Spring Wheat (*Triticum aestivum* L.) Chloroplast Genome: Complete Sequence and Contig Clones. **Plant Molecular Biology Reporter**, v. 18, n. 3, p. 243–253, 3 set. 2000. doi: 10.1007/BF02823995.

OSGOOD, R. V., HANNA, W. W., AND TEW, T. L. Hybrid seed production of pearl millet X Napier grass triploid hybrids. **CropSci**. 37,998-999. 1997.

PAIVA, E. A. A. Meiose em híbridos hexaploides de *Pennisetum purpureum* e *Pennisetum glaucum*. 53 f. **Dissertação (Mestrado em genética e melhoramento de plantas)**, Universidade Federal de Lavras, Lavras, 2006.

PARK, I., SONG, J. H., YANG, S., & MOON, B. C. Comparative Analysis of Actaea Chloroplast Genomes and Molecular Marker Development for the Identification of Authentic Cimicifugae Rhizoma. **Plants**, v. 9, n. 2, p. 157, 2020.

PLOHL, M. Those mysterious sequences of satellite DNAs. **Periodicum Biologorum**, v. 112, n. 4, p. 403–410, 2010. doi: 10.14716/ijtech.v6i4.1206.

PEREIRA, Antônio Vander et al. Tendências do melhoramento genético e produção de sementes de forrageiras no Brasil. In: **SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS**, 7., 2003, Lavras. Anais... p. 36–63, 2003.

RASKINA, O.; BARBER, J. C.; NEVO, E.; BELYAYEV, A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. **Cytogenetic and Genome Research**, v. 120, n. 3–4, p. 351–357, maio 2008. doi: 10.1159/000121084.

RASMUSSEN, D. A.; NOOR, M. A. What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). **BMC Genomics**, v. 10, n. 1, p. 382, 2009. doi: 10.1186/1471-2164-10-382.

RAVEENDAR, S.; LEE, G. A.; LEE, K. J.; SHIN, M. J.; LEE, J. R.; LEE, S. Y.; CHO, G. T.; MA, K. H.; CHUNG, J. W. The complete chloroplast genome of pearl millet (*Pennisetum*

glaucum (L.) R. Br.) and comparative analysis within the family poaceae. **Cereal Research Communications**, v. 47, n. 1, p. 1–10, mar. 2019. doi: 10.1556/0806.46.2018.064.

S. FILGUEIRAS, T. O gênero *Cenchrus* L. No Brasil (Gramineae:Panicoideae). **Acta Amazonica**, v. 14, p. 95–127, 1984.

ROBERT, T., KHALFALLAH, N., MARTEL, E., LAMY, F., PONCET, V., ALLINNE, C., ... & SILJAK-YAKOVLEV, S. *Pennisetum*. In: **Wild crop relatives: genomic and breeding resources**. Springer, Berlin, Heidelberg, 2011. p. 217-255.

ROYAL BOTANIC GARDENS KEW. (2020) **Seed Information Database (SID)**. Version 7.1. Available from: <http://data.kew.org/sid/> (October 2020)

SAHA, S.; BRIDGES, S.; MAGBANUA, Z. V.; PETERSON, D. G. Computational Approaches and Tools Used in Identification of Dispersed Repetitive DNA Sequences. **Tropical Plant Biology**, v. 1, n. 1, p. 85–96, 2008. doi: 10.1007/s12042-007-9007-5.

SASKI, C.; LEE, S. B.; FJELLHEIM, S.; GUDA, C.; JANSEN, R. K.; LUO, H.; TOMKINS, J.; ROGNLI, O. A.; DANIELL, H.; CLARKE, J. L. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. **Theoretical and Applied Genetics**, v. 115, n. 4, p. 571–590, 2007. doi: 10.1007/s00122-007-0567-4.

SCHREIBER, M.; STEIN, N.; MASCHER, M. **Genomic approaches for studying crop evolution** *Genome Biology* BioMed Central Ltd., , 21 set. 2018. . doi: 10.1186/s13059-018-1528-8.

SCHANK, S. C., DIZ, D. A., BATES, D. B., & THOMRSOLA, K. E. Genetic improvement of napiergrass and hybrids with pearl millet. **Biomass and Bioenergy** v. 5, n. 1, p. 35–40, 1993.

SOBRINHO, F. D. S.; PEREIRA, A. Vander; JOSÉ, F.; BOTREL, M. A. Avaliação agronômica de híbridos interespecíficos entre capim-elefante e milheto. **Pesquisa Agropecuaria Brasileira**, v. 40, n. 9, p. 873–880, 2005.

SOUZA, K. A. De; ABRANTES, R. S. X.; LAÍS, C.; DE, D.; SOARES, M. A.; SANTOS, C.; VINICIUS, M.; PAULO, L.; FEITOSA, M. Panorama de sistema de produção de leite no Brasil. **Informativo Técnico do Semiárido**, v. 10, n. 2, p. 57–61, 2016.

TECHIO, V. H., DAVIDE, L. C., PEREIRA, A. V., & BEARZOTI, E. Cytotaxonomy of some species and of interspecific hybrids of *Pennisetum* (Poaceae, Poales). **Genetics and molecular biology**, v. 25, n. 2, p. 203-209, 2002.

TREANGEN, T. J.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. **Nature Reviews Genetics**, v. 13, n. 1, p. 36–46, 29 jan. 2012. doi: 10.1038/nrg3117.

TOSTAIN S (1998) Le mil, une longue histoire, hypothèses sur sa domestication et ses migrations. In: **Chastenet N (ed) Plantes et paysages d’Afrique**. Karthala, Paris, pp 461–490

VARSHNEY, R. K.; SHI, C.; THUDI, M.; MARIAC, C.; WALLACE, J.; QI, P.; ZHANG, H.; ZHAO, Y.; WANG, X.; RATHORE, A.; SRIVASTAVA, R. K.; CHITKINENI, A.; FAN, G.; BAJAJ, P.; PUNNURI, S.; GUPTA, S. K.; ... VIGOUROUX, Y.; XU, X. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. **Nature Biotechnology**, v. 35, n. 10, p. 969–976, 18 out. 2017. doi: 10.1038/nbt.3943.

WANG, L., MA, G., WANG, H., CHENG, C., MU, S., QUAN, W., ... & WANG, X. A draft genome assembly of halophyte *Suaeda aralocaspica*, a plant that performs C4 photosynthesis within individual cells. **GigaScience**, v. 8, n. 9, p. giz116, 2019.

WANG, C.; YAN, H.; LI, J.; ZHOU, S. Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum ‘Zise’) and identification of its SSR markers. **Molecular Breeding**, v. 39, n. 194, 2018a. doi: 10.1007/s11032-018-0849-3.

WANG, X.; CHENG, F.; ROHLSSEN, D.; BI, C.; WANG, C.; XU, Y.; WEI, S.; YE, Q.; YIN, T.; YE, N. Organellar genome assembly methods and comparative analysis of horticultural plants. **Horticulture Research**, v. 5, n. 1, p. 1–13, 2018b. doi: 10.1038/s41438-017-0002-1.

WICKER, T.; NARECHANIA, A.; SABOT, F.; STEIN, J.; VU, G. T.; GRANER, A.; WARE, D.; STEIN, N. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. **BMC Genomics**, v. 9, n. 1, p. 518, 31 out. 2008. doi: 10.1186/1471-2164-9-518.

WEBER, N.; HALPIN, C.; HANNAH, L. C.; JEZ, J. M.; KOUGH, J.; PARROTT, W. Editor’s Choice: Crop Genome Plasticity and Its Relevance to Food and Feed Safety of Genetically Engineered Breeding Stacks. **Plant Physiology**, v. 160, n. 4, p. 1842–1853, 1 dez. 2012. doi: 10.1104/pp.112.204271.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973–982, 2007. doi: 10.1038/nrg2165.

WINTERFELD, G.; RÖSER, M. Chromosomal localization and evolution of satellite DNAs and heterochromatin in grasses (Poaceae), especially tribe Aveneae. **Plant Systematics and Evolution** v. 264, n. 1–2, p. 75–100, 2007.

WU, Y.; ZHOU, H. The complete chloroplast genome sequence of *Cenchrus purpureus*. **Mitochondrial DNA Part B: Resources**, v. 4, n. 1, p. 51–52, 2019. doi: 10.1080/23802359.2018.1536451.

XU, J.; SONG, Y.; JING, X.; LI, M. Characterization of the complete chloroplast genome sequence of *Pennisetum glaucum* and its phylogenetic implications. **Mitochondrial DNA Part B: Resources**, v. 4, n. 2, p. 3764–3765, 2019. doi: 10.1080/23802359.2019.1668312.

YAMASHITA, H., UCHIDA, T., TANAKA, Y., KATAI, H., NAGANO, A. J., MORITA, A., & IKKA, T. Genomic predictions and genome-wide association studies based on RAD-seq of quality-related metabolites for the genomics-assisted breeding of tea plants. **Scientific**

reports, v. 10, n. 1, p. 1-10, 2020.

YAN, Q., WU, F., XU, P., SUN, Z., LI, J., GAO, L., ... & YI, X. The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. **Molecular Ecology Resources**, 2020.

ZAKIAN, V. A. Telomeres: Beginning to Understand the End. **Science**, v. 270, n. 5242, p. 1601–1607, 8 dez. 1995. doi: 10.1126/science.270.5242.1601.

SEGUNDA PARTE – ARTIGOS**ARTIGO 1****Genômica comparativa da fração repetitiva de *Cenchrus purpureus* e *Cenchrus americanus* (Poaceae)**

Silvestrini, A. J. A; Vaio, MAGDALENA; Torres, G. A.

RESUMO

Cenchrus L. é um importante gênero da família Poaceae que abriga diversas espécies de elevada importância agrônômica, como *Cenchrus purpureus* e *Cenchrus americanus*, duas espécies que se destacam na produção de forragem e de grãos, respectivamente. Diversos programas de melhoramento genético exploram a capacidade de produção e qualidade dessas forrageiras, assim como, devido à complexa composição genômica entre ambas, as mesmas são de grande interesse para estudos genômico evolutivos. A utilização de dados oriundos de sequenciamento de próxima geração (NGS) aliado a recentes “pipelines” desenvolvidas para caracterização da fração genômica repetitiva, são uma interessante abordagem para caracterização de germoplasma, assim como para estudos genômico-comparativos. Essa abordagem permite desde a identificação de marcadores genômicos ou cromossômicos-específicos, até maiores esclarecimentos sobre as relações genômicas que moldam espécies relacionadas, resultados esses de grande valia para espécies de *Cenchrus*. Portanto, o objetivo do presente trabalho é a caracterização *in silico* e análise comparativa da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus* através do sequenciamento genômico de baixa cobertura. Para tanto, o DNA genômico de *C. purpureus* e *C. americanus* foi extraído e sequenciado em plataforma Illumina HiSeq™ 4000. A análise individual e comparativa da fração repetitiva do genoma de ambas as espécies se deu na plataforma RepeatExplorer com configurações padrão, sendo as demais análises bioinformáticas realizadas através do software Geneious Prime 2020.1.1. A fração repetitiva do genoma de *C. purpureus* e *C. americanus* corresponde a 52,23 e 76,82%, respectivamente, e os elementos repetitivos mais abundantes em ambas as espécies são os LTRs. Foram ainda identificadas sequências de DNA satélite em ambos os genomas que correspondem a 2,55 e 4,17% do genoma de cada uma das espécies, respectivamente. A relação ancestral, assim como os ciclos de poliploidização-diploidização que envolvem ambas as espécies, tiveram um papel fundamental na composição da fração repetitiva das mesmas. Esses ciclos levaram *Cenchrus americanus*, um possível paleopoliploide, a possuir maior abundância de elementos transponíveis quando comparado à *Cenchrus purpureus*, um recente aloploiploide. Os processos evolutivos inerentes às sequências de DNA satélite, como a evolução concertada e teoria da biblioteca genômica de DNA satélites compartilhada entre espécies relacionadas, são a base da diferenciação e amplificação dessa região genômica entre *Cenchrus purpureus* e *Cenchrus americanus*.

ABSTRACT

Cenchrus L. is an important genus of the Poaceae family that comprises several species of high agronomic importance, such as *Cenchrus purpureus* and *Cenchrus americanus*, two species that stand out in the production of forage and grains, respectively. Several breeding programs explore the production capacity and quality of these forages, as well, due to the complex genomic composition between them, they are of great interest for evolutionary genomic studies. The use of data from next-generation sequencing (NGS) combined with recent pipelines developed for the characterization of the repetitive genomic fraction, is an interesting approach for the characterization of germplasm, as well as for comparative genomics studies. This approach allows from the identification of genomic or chromosomal-specific markers, to further clarification on the genomic relationships that shape related species. Therefore, the objective of the present work is the in silico characterization and comparative analysis of the repetitive fraction of the genome of *Cenchrus purpureus* and *Cenchrus americanus* through low coverage genomic sequencing. For that, the genomic DNA of *C. purpureus* and *C. americanus* was extracted and sequenced on the Illumina HiSeq™ 4000 platform. The individual and comparative analysis of the repetitive fraction of the genome of both species took place on the *RepeatExplorer* platform within standard setup, with the other bioinformatics analysis carried out using the Geneious Prime 2020.1.1 software. The repetitive fraction of the genome of *C. purpureus* and *C. americanus* corresponds to 52.23 and 76.82%, respectively, and the most abundant repetitive elements in both species are the LTRs. Satellite DNA sequences were also identified in both genomes that correspond to 2.55 and 4.17% of the genome of each species, respectively. The ancestral relationship, as well as the polyploidization-diploidization cycles that involve both species, played a fundamental role in the composition of their repetitive fraction. These cycles led *Cenchrus americanus*, a possible paleopolyploid, to have a greater abundance of transposable elements when compared to *Cenchrus purpureus*, a recent allopolyploid. The evolutionary processes inherent to satellite DNA sequences, such as concerted evolution and the theory of the genomic library of satellite DNA shared between related species, are the basis for the differentiation and amplification of this genomic region between *Cenchrus purpureus* and *Cenchrus americanus*.

1. INTRODUÇÃO

Cenchrus L. é um amplo gênero pertencente à família Poaceae, subfamília Panicoideae, tribo Paniceae que possui cerca de 107 espécies distribuídas ao longo de toda a faixa tropical e subtropical do planeta (KEW GARDENS SCIENCE, 2020). A mais recente circunscrição taxonômica do gênero resultou da inclusão dos gêneros *Cenchrus*, *Pennisetum* Rich. e *Odontelytrum* Hack em *Cenchrus* (DONADÍO et al., 2009; CHEMISQUY et al., 2010).

O gênero possui diversas espécies de importância agrônômica, sendo as principais *Cenchrus purpureus* (Schumach.) (syn. *Pennisetum purpureum*) ou Capim-elefante ($2n = 4x = 28$, A'A'BB) e *Cenchrus americanus* [(L.) R. Brown, 1810] (syn. *Pennisetum glaucum*) ou Milheto ($2n = 2x = 14$, AA). *Cenchrus purpureus* é conhecido por sua elevada produção de massa seca adequada à alimentação de gado leiteiro, e *Cenchrus americanus* pela considerável produção de grãos utilizados na alimentação humana, assim como pela presença de características favoráveis ao uso forrageiro (HANNA, 1981; SOBRINHO et al., 2005; DOS REIS et al., 2014).

Cenchrus purpureus e *Cenchrus americanus* possuem genomas de tamanhos entre pequeno e intermediário, com cerca de 2,27 pg (2,22 Gb) e 2,37 pg (2,32 Gb), respectivamente (SOLTIS et al., 2003; ANDRADE-VIEIRA, 2010; MICHAEL, 2014), sendo 60 a 80% desses genomas composto por sequências de DNA repetitivo (VARSHNEY et al., 2017; WANG et al., 2018). A fração repetitiva dos genomas pode ser dividida em dois grandes grupos: um grupo de sequências repetidas em tandem e que compõe regiões heterocromáticas em cromossomos eucariotos, denominadas como DNA Satélite e outro de sequências dispersas no genoma denominadas como Elementos Transponíveis. (HESLOP-HARRISON; SCHMIDT, 2007). Em plantas, a fração repetitiva pode chegar a 90% do genoma total e é a classe de sequências que melhor explica a grande variação nos tamanhos genômicos vegetais. Muitos de seus componentes desempenham funções que vão desde processos de regulação gênica até o envolvimento em rearranjos cromossômicos (GARRIDO-RAMOS, 2017).

O dinamismo e consequente rápida evolução dessas sequências (WICKER et al., 2007) faz com que elas confirmem maior plasticidade aos genomas (MEHROTRA; GOYAL, 2014), o que permite às espécies vegetais se adaptarem às mais adversas condições (LEITCH; LEITCH, 2008). Como consequência desse dinamismo, os elementos repetitivos interferem no curso evolutivo das mais diversas espécies (RASKINA et al., 2008) e podem impactar os resultados obtidos por programas de melhoramento genético vegetal.

A utilização de dados oriundos do sequenciamento de próxima geração (NGS) aliado a rotinas específicas implementadas em plataformas para estudo da fração repetitiva do genoma

eucarioto, vêm se mostrando de grande valia para caracterização de germoplasma, assim como para subsidiar estudos acerca de evolução genômica entre espécies. Essas ferramentas permitem a análise, classificação e quantificação das classes e famílias de sequências repetitivas presentes nos mais diversos organismos (NOVÁK et al., 2013).

Esses estudos podem subsidiar o estabelecimento de marcadores moleculares, especialmente a partir de sequências satélites, permitindo a caracterização da diversidade e estrutura genética de germoplasma, bem como servir de base para a identificação de marcas associadas a genes de interesse agrônomo (KALENDAR et al., 1999). A classificação de elementos repetitivos permite ainda o estabelecimento de marcadores genômicos/cromossômicos-específicos, que subsidiam uma cariotipagem mais precisa, e podem auxiliar em estudos de especiação e evolução genômica (RASKINA et al., 2008; HEITKAM et al., 2020).

Cenchrus purpureus e *Cenchrus americanus* são espécies bastante estudadas do ponto de vista genético e genômico-evolutivo. Diversos autores já realizaram análises acerca da composição genômica de ambas. Atualmente, a composição amplamente aceita é de que *Cenchrus americanus* possui genoma AA e *Cenchrus purpureus* genomas A'A'BB, no qual o sub genoma A' é homeólogo de A. (MARTEL et al., 2004; DOS REIS et al., 2014). O genoma B por sua vez não possui origens definidas, estando presente apenas em *Cenchrus purpureus*. A origem de *Cenchrus purpureus* remete a hibridação entre o ancestral comum para com *Cenchrus americanus* e o desconhecido portador do genoma B. Existem relatos de *Cenchrus americanus* se tratar de um possível paleopoliploide, que, após eventos de diploidização, se comporta como diploide, enquanto que *Cenchrus purpureus* se trata de um poliploide recente quando comparado a *Cenchrus americanus* (MARTEL et al., 2004; TECHIO; DAVIDE; PEREIRA, 2006; WANG et al., 2018; YAN et al., 2020).

Além de estudos acerca das relações citogenômicas entre essas espécies, que predominantemente estão relacionados a regiões heterocromáticas e conseqüentemente portadoras de sequências repetitivas, existem descrições gerais sobre fração repetitiva do genoma de ambas a partir de dados NGS, com ênfase nos elementos transponíveis (VARSHNEY et al., 2017; PAUDEL et al., 2018; YAN et al., 2020a). Grande parte desses trabalhos não apresenta detalhes das famílias e super famílias que compõe os elementos repetitivos, assim como não possuem resultados acerca das sequências satélites. Não existem também, até o presente momento, análises genômico-comparativas através de dados NGS, entre espécies de *Cenchrus*.

Todos os estudos genômicos comparativos realizados em *Cenchrus purpureus* e *Cenchrus americanus*, tanto através de técnicas citogenômicas (HELENA; DAVIDE; PEREIRA, 2005; DOS REIS et al., 2014), quanto através de dados NGS (VARSHNEY et al., 2017; PAUDEL et al., 2018; YAN et al., 2020) demonstram elevada similaridade entre o genoma dessas espécies, refletindo a estrita relação filogenética entre as mesmas (DONADÍO et al., 2009; CHEMISQUY et al., 2010). A descrição detalhada e comparativa da fração repetitiva do genoma dessas espécies se faz necessária para desvendar o quão similar é essa fração genômica em *Cenchrus purpureus* e *Cenchrus americanus*, assim como para desenvolver possíveis marcadores genômicos e cromossômicos específicos oriundos dos genomas A, A' e B que possam auxiliar em maiores esclarecimentos acerca da diferenciação e origem dos mesmos.

2. MATERIAL E MÉTODOS

Material genético e sequenciamento genômico

Foram utilizadas para sequenciamento genômico a variedade Pioneiro de *Cenchrus purpureus* e a variedade BN2 de *Cenchrus americanus* (Tabela 1).

Tabela 1 Dados de passaporte *Cenchrus* spp.

Material Vegetal	Origem
<i>Cenchrus purpureus</i>	Banco Ativo de Germoplasma – EMBRAPA Gado de Leite – Juiz de Fora – MG
<i>Cenchrus americanus</i>	Banco Ativo de Germoplasma – EMBRAPA Gado de Leite – Juiz de Fora – MG

Folhas jovens de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2 foram utilizadas para extração de DNA através do kit de extração *DNeasy® Plant Mini Kit* (Qiagen® Inc., Venlo, Netherlands). A avaliação quanto à integridade e qualidade do DNA extraído foi realizada, respectivamente, por meio de Eletroforese em gel de Agarose 1% e Espectrofotômetro *NanoDrop™* (Thermo Scientific™). A montagem da biblioteca foi realizada utilizando o kit *Nextera™ DNA Flex Library Prep* com fragmentos *Pair-ended* de

2x300pb para ambas as espécies. O sequenciamento genômico foi realizado em plataforma Illumina *HiSeq*TM 4000, no Instituto René Rachou – Fiocruz Minas, Belo Horizonte - MG.

Pré-processamento de dados

As etapas de pré-processamento dos *reads* ou sequências, e as análises de *clustering*, foram realizadas na plataforma *RepeatExplorer* (NOVÁK et al. 2013).

A qualidade dos dados brutos de cada espécie foi analisada por meio da ferramenta *FastQC Quality Reports* (BROWN; PIRRUNG; MCCUE, 2017) com parâmetros de configurações padrão.

Após a análise de qualidade, procedeu-se com o corte das sequências através da ferramenta *Trim Sequence*. O corte foi concentrado na manutenção de regiões entre 0 e 190 nt com maior qualidade ou confiabilidade em relação aos nucleotídeos presentes. Posteriormente, foi realizada a filtragem de sequências através da ferramenta *Filter by quality*, com parâmetros de ao menos 90% de uma sequência inteira com *Score Phred 20*. Finalmente foi feita a filtragem para homogeneização do tamanho de cada sequência, sendo mantidas apenas sequências que possuíam ao menos 190 pb de comprimento. Após o descarte das sequências de baixa qualidade, o número de sequências (*reads*) submetidas para análise foi de 1.060.996 para *Cenchrus purpureus* e de 1.273.846 para *Cenchrus americanus*, o que corresponde a aproximadamente 0,1x o genoma de cada espécie.

Análise individual da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus*: *clustering*, anotação, mapeamento e alinhamento de sequências

A análise de *clustering* foi feita por meio da ferramenta *mgblast tool*, na qual todas as sequências foram comparadas umas contra as outras e agrupadas com base em sua similaridade, estabelecida como padrão pela plataforma *RepeatExplorer*, de pelo menos 90% em ao menos 55% de toda a sua extensão.

A partir da análise de agrupamento foram fornecidas a anotação automática dos elementos repetitivos e os gráficos com características topográficas distintas para cada classe (NOVÁK et al., 2013; BELYAYEV et al., 2018). A anotação automática é realizada com base no banco de dados REXdb (NEUMAN et al., 2019).

Elementos transponíveis não classificados foram manualmente verificados contra o banco de dados *Giri Repbase* (BAO; KOJIMA; KOHANY, 2015) e classificados de acordo

com sua similaridade com elementos repetitivos previamente depositados. Todas as sequências classificadas como DNA satélite foram identificadas pelo algoritmo TAREAN (NOVÁK et al., 2017) sendo a estrutura de repetição analisada através de gráfico *Dotplot* na plataforma *Dotmatcher* – EMBOSS (LI et al., 2015) para confirmação de arranjo em tandem. Essas sequências foram submetidas ainda a buscas por similaridade com sequências previamente publicadas na literatura através de ferramentas BLAST em <<https://blast.ncbi.nlm.nih.gov/Blast.cgi>>.

A relação e identidade entre os satélites oriundos das duas espécies foi analisada na plataforma *Repeat Masker* (SMIT et al., 2015) e por meio de alinhamento local múltiplo de sequências através do algoritmo presente em MAFFT (KATO; STANDLEY, 2013). A partir das matrizes de identidade oriundas dos alinhamentos, os DNA satélites foram classificados em famílias e subfamílias, de forma que famílias de DNA satélite que possuem identidade maior que 95% foram classificadas como uma mesma variante, identidade entre 80% e 95% classificadas como diferentes variantes pertencentes à uma mesma família de DNA satélite, e identidade abaixo de 80% foram classificadas como pertencentes a uma mesma super família de DNA satélite (RUIZ-RUANO et al., 2016).

Os DNA satélites foram então organizados e enumerados de acordo com sua abundância para com sua respectiva espécie. Para diferenciação entre famílias e super famílias foram atribuídos sufixos de acordo com as letras iniciais do alfabeto romano para família e as letras finais para cada super família. A ferramenta *Map to Reference* foi utilizada para mapeamento e quantificação de sequências satélite contra o genoma pré-processado de *Cenchrus purpureus* e *Cenchrus americanus*. O pré-processamento foi realizado através do algoritmo presente em BBduk, disponível em <<https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh>>. As ferramentas MAFFT, BBduk e *Map to Reference* foram utilizadas através do software Geneious Prime 2020.1.1 (KEARSE et al., 2012).

Análise filogenética dos domínios RT em famílias de *Ty3-gypsy*

Para a análise filogenética dos domínios conservados da Retrotransposase (RT) de famílias do retroelemento *Ty3-gypsy*, foi utilizada a ferramenta DANTE (NEUMANN et al., 2019) implementada em *RepeatExplorer* para extração das sequências de aminoácidos componentes desses domínios. Como grupo externo foram selecionadas, aleatoriamente, 60 sequências de domínios proteicos RT em *Setaria italica*, uma espécie filogeneticamente relacionada a *Cenchrus* (BENNETZEN et al., 2012), a partir do banco de dados de domínios

proteicos REXdb (NEUMANN et al., 2019). Após a extração, essas sequências foram alinhadas através do algoritmo MUSCLE, e as relações filogenéticas foram calculadas utilizando o método *Neighbor-Joining* com a inclusão de valores *Bootstrap*. Todas as análises foram realizadas através do software MEGA X (KUMAR et al., 2018).

Análise comparativa da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus*

Para eliminar efeito da diferença de ploidia das espécies, *Cenchrus purpureus* tetraploide e *Cenchrus americanus* diploide, foram realizados cálculos para amostragem dos *reads* em que a mesma fosse equivalente ao nível de ploidia Cx das duas espécies (GAIERO et al., 2019).

A análise de *clustering*, a anotação e os alinhamentos foram realizados seguindo as mesmas rotinas descritas para as análises individuais através da utilização da opção de configurações avançadas em *RepeatExplorer 2* com os *reads* de ambas as espécies imputados de forma conjunta. Foram utilizados parâmetros de configuração padrão (NOVÁK et al. 2013).

3. RESULTADOS

Caracterização da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus*

A partir do sequenciamento genômico em plataforma Illumina *HiSeq*TM 4000 foram gerados 1.223.394 *reads* referentes a aproximadamente 0,15x a cobertura do genoma de *Cenchrus purpureus* e 1.479.774 *reads* equivalentes a aproximadamente 0,16x o genoma de *Cenchrus americanus*. Após as etapas de pré-processamento e controle de qualidade, foram utilizadas sequências correspondendo a cerca de 0,1x do genoma de cada uma das espécies, tanto nas análises individuais, quanto na comparativa.

Para *Cenchrus purpureus*, baseado na análise de *clustering*, dos 1.060.996 analisados pelo algoritmo, 759.317 foram classificados como sequências repetitivas sendo agrupados em 75.634 *clusters*, enquanto 301.679 foram classificados como *singlets*, ou seja, *reads* que não foram identificados como elementos repetitivos (Figura 1A). De todos os clusters formados 333 foram tidos como os mais representativos por possuírem mais de 0,01% de abundância, sendo assim identificados pelo algoritmo com base no formato dos gráficos. Portanto, de acordo com

os clusters identificados, e suas respectivas proporções, após desconsiderar *reads* oriundos de sequências cloroplastidiais e mitocondriais, a fração repetitiva do genoma *Cenchrus purpureus* é cerca de 52,23% (Tabela 1, Figura 2).

Em *Cenchrus americanus* foram analisados 1.313.846 *reads* dos quais 1.144.939 foram alocados como sequências repetitivas e 168.907 como *singlets* (Figura 1A). As sequências repetitivas foram agrupadas em 61.975 *clusters*, sendo que destes, 247 foram identificados como elementos repetitivos por corresponderem a mais de 0,01% de abundância. Sendo assim, a fração repetitiva do genoma de *Cenchrus americanus*, desconsiderando sequências oriundas dos genomas cloroplastidiais e mitocondriais, corresponde a cerca de 76,82% (Tabela 1, Figura 2).

Dentre os *clusters* identificados entre as duas espécies, 2,63% são exclusivos de *Cenchrus purpureus*, enquanto 0,65% são exclusivos de *Cenchrus americanus* (Figura 1B). Apesar de todas as classes de elementos transponíveis estarem presentes em ambas as espécies (Tabela 1), em relação às famílias de DNA satélite, nem todas são compartilhadas entre *Cenchrus purpureus* e *Cenchrus americanus* (Figura 6), sendo justamente essa a fração correspondente aos clusters exclusivos presentes em *Cenchrus purpureus*, enquanto os *clusters* exclusivos presentes em *Cenchrus americanus* estão relacionados a *reads* não inclusos nos *clusters* identificados que, apesar de pertencerem a elementos repetitivos já listados, foram agrupados separadamente e caracterizados como ausentes em *Cenchrus purpureus* através da anotação automática presente em *RepeatExplorer 2*.

Dentre as classes de elementos repetitivos identificadas, os retroelementos são a mais abundantemente representada em ambas as espécies, sendo os LTR os que possuem as maiores proporções. Dentro das super famílias de retroelementos, *Ty3-Gypsy* é a mais representativa, ocupando 48,60% do genoma de *Cenchrus americanus* e 28,25% do genoma de *Cenchrus purpureus* (Tabela 1, Figura 2). A família mais abundante dentro de *Ty3-Gypsy*, em ambos genomas, é *Athila*. A segunda super família mais abundante é *Ty1-Copia*, que corresponde a 17,34% e 12,07% dos genomas de *Cenchrus americanus* e *Cenchrus purpureus*, respectivamente. Nesse caso, a família mais representativa é SIRE. Em relação aos transposons, a família mais representativa é *EnSpm-CACTA* em ambas as espécies (Tabela 1, Figura 2).

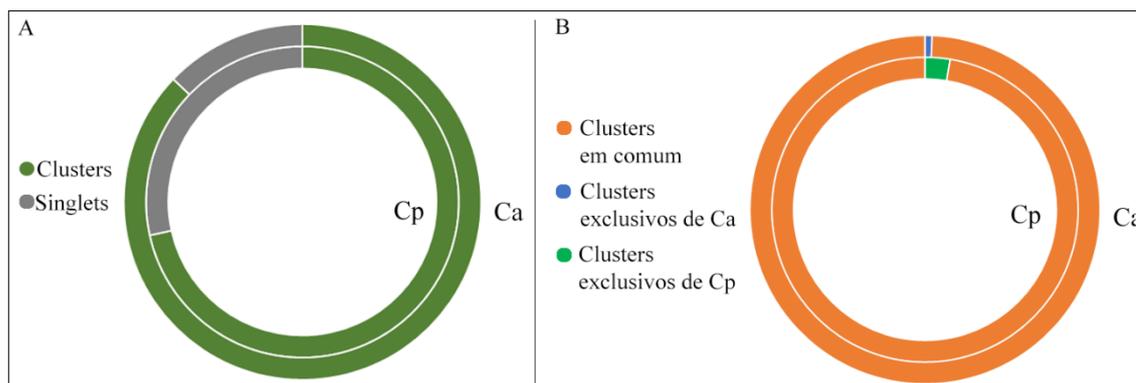


Figura 1 Agrupamento dos *reads* analisados pela plataforma *RepeatExplorer* nos genomas de *Cenchrus purpureus* e *Cenchrus americanus*. (A) Proporção de *clusters* identificados e *singlets* no sequenciamento de baixa cobertura através da plataforma *RepeatExplorer*. (B) Proporção de *clusters* em comum e exclusivos entre *Cenchrus purpureus* e *Cenchrus americanus*. Cp: *Cenchrus purpureus*, Ca: *Cenchrus americanus*.

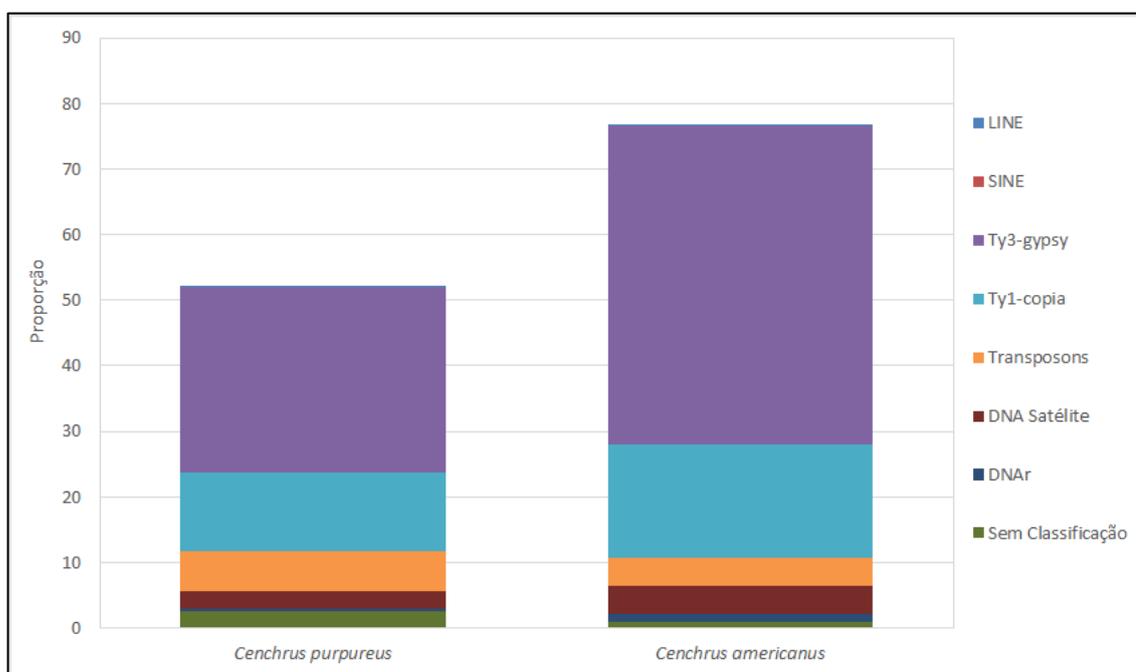


Figura 2 Proporção de elementos repetitivos presentes no genoma de *Cenchrus purpureus* e *Cenchrus americanus*.

A classificação e a abundância das famílias de elementos transponíveis revelaram que *Cenchrus americanus* possui maior quantidade desses elementos, que correspondem a 70,38% de seu genoma, quando comparada a *Cenchrus purpureus* que possui 46,68% (Tabela 1, Figura 2). *Cenchrus americanus* possui maior abundância de elementos Classe I, ou retroelementos, correspondendo a 66,13% de seu genoma, enquanto *Cenchrus purpureus* possui 40,58%. É possível notar que essa situação se repete mesmo dentro das super famílias e famílias desses retroelementos, principalmente em *Ty3-Gypsy*, no qual a grande maioria é mais abundante em

Cenchrus americanus. Algumas exceções estão na super família LINE, na família *Ale*, e em algumas famílias de *Ty1-Copia* (Figura 3). Em relação aos elementos de Classe II ou transposons, a situação se inverte, no qual *Cenchrus purpureus* possui 6,10% e *Cenchrus americanus* 4,25% de todo seu genoma, com a maioria das famílias em maior quantidade na primeira espécie (Tabela 1, Figuras 2 e 3).

Em *Cenchrus americanus*, a proporção do genoma ocupada por sequências que codificam para RNA's ribossomais como os arranjos gênicos de rDNA 5S e rDNA 45S (18S, 5.8S e 26S) é de 1,36%, enquanto que para *Cenchrus americanus* é de 0,51%. As sequências de DNA satélite correspondem a 5,53% e 3,07% do genoma de *Cenchrus americanus* e *Cenchrus purpureus*, respectivamente (Tabela 1, Figura 2).

Tabela 1 Composição da fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus*.

Ordem	Super-família	Família	Abundância		
			<i>C. purpureus</i>	<i>C. americanus</i>	
Elementos Transponíveis					
Classe I - Retrotransposons					
LTR	<i>Ty3-Gypsy</i>	LTR Sem Classificação	0,01%	0,02%	
		<i>Tat/Retand</i>	6,13%	13,96%	
		Chromovirus (<i>Tekay</i> + CRM)	6,90%	12,21%	
		<i>Athila</i>	11,16%	14,38%	
		<i>Ogre</i>	1,76%	6,54%	
		<i>Gypsy</i> Sem Classificação	2,29%	1,50%	
	Total <i>Gypsy</i>	28,25%	48,60%		
	<i>Ty1-Copia</i>	<i>Ale</i>	0,13%	0,02%	
		<i>TAR</i>	0,35%	0,24%	
		<i>SIRE</i>	9,86%	15,36%	
		<i>Ivana</i>	0,01%	0,01%	
		<i>Bianca</i>	0,10%	0,20%	
		<i>Angela</i>	0,11%	0,31%	
		<i>Ikeros</i>	0,16%	0,15%	
		<i>Tork</i>	0,05%	0,05%	
		<i>Copia</i> Sem Classificação	1,25%	0,95%	
	Total <i>Copia</i>	12,07%	17,34%		
	LINE			0,23%	0,10%
	SINE			0,01%	0,05%
Total Retrotransposons			40,58%	66,13%	
Classe II - Transposons					
Subclasse I - TIR	<i>PIF Harbinger</i>		0,37%	0,19%	
	<i>Tc1-Mariner</i>		0,08%	0,06%	
	<i>EnSpm - CACTA</i>		4,59%	3,31%	
	<i>MuDR - Mutator</i>		0,77%	0,47%	
	<i>hAT</i>		0,18%	0,16%	
Subclasse II	<i>Helitron</i>		0,09%	0,05%	
Total Transposons			6,10%	4,25%	
Total Elementos Transponíveis			46,68%	70,38%	
DNA Satélite			2,55%	4,17%	
rDNA 5S			0,015%	0,026%	
rDNA 45S			0,49%	1,33%	
rDNA Total			0,51%	1,36%	
Total DNA Satélite			3,07%	5,53%	
Sem Classificação			2,47%	0,89%	
Total DNA Repetitivo			52,23%	76,82%	

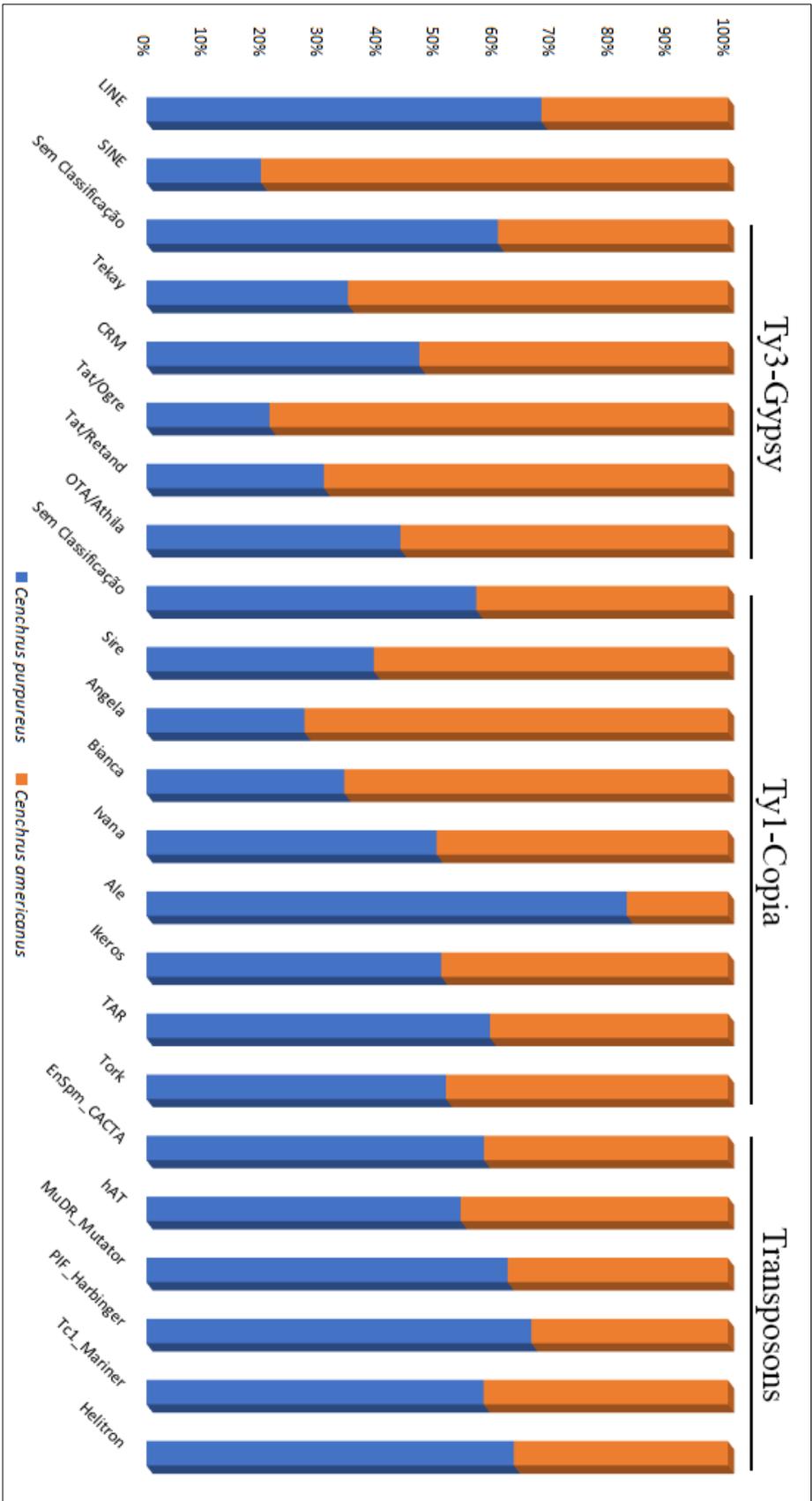


Figura 3 Análise comparativa da abundância das super famílias e famílias de retroelementos e transposons no genoma de *Cnephrys purpureus* e *Cnephrys americanus*.

Filogenia dos domínios da Retrotransposase em famílias de *Ty3-gypsy*

As análises filogenéticas do domínio RT das famílias de *Ty3-gypsy* (Figura 4) presentes no proteoma de *Cenchrus purpureus*, *Cenchrus americanus* e *Setaria italica*, demonstram que linhagens comuns de *Retand* se diferenciaram ao longo do curso evolutivo em clados com linhagens específicas de cada uma das três espécies. Poucos clados apresentam linhagens de *Tekay* diferenciadas a nível de espécie, sendo a grande maioria compartilhada e dispersas em momentos de diferenciação semelhantes entre *Cenchrus purpureus*, *Cenchrus americanus* e *Setaria italica*. Linhagens de *Ogre* e *Athila* são específicas de *Cenchrus*, uma vez que nenhuma delas possui agrupamentos oriundos de *Setaria italica*. É válido mencionar que para *Setaria italica*, utilizada como grupo externo, foi realizada uma amostragem aleatória de *reads* e, portanto, algumas famílias de *Ty3-gypsy* podem estar sub representadas. Poucas linhagens CRM são espécie-específicas, sendo a maioria compartilhadas entre *Cenchrus americanus* e *Cenchrus purpureus*, e *Setaria italica*.

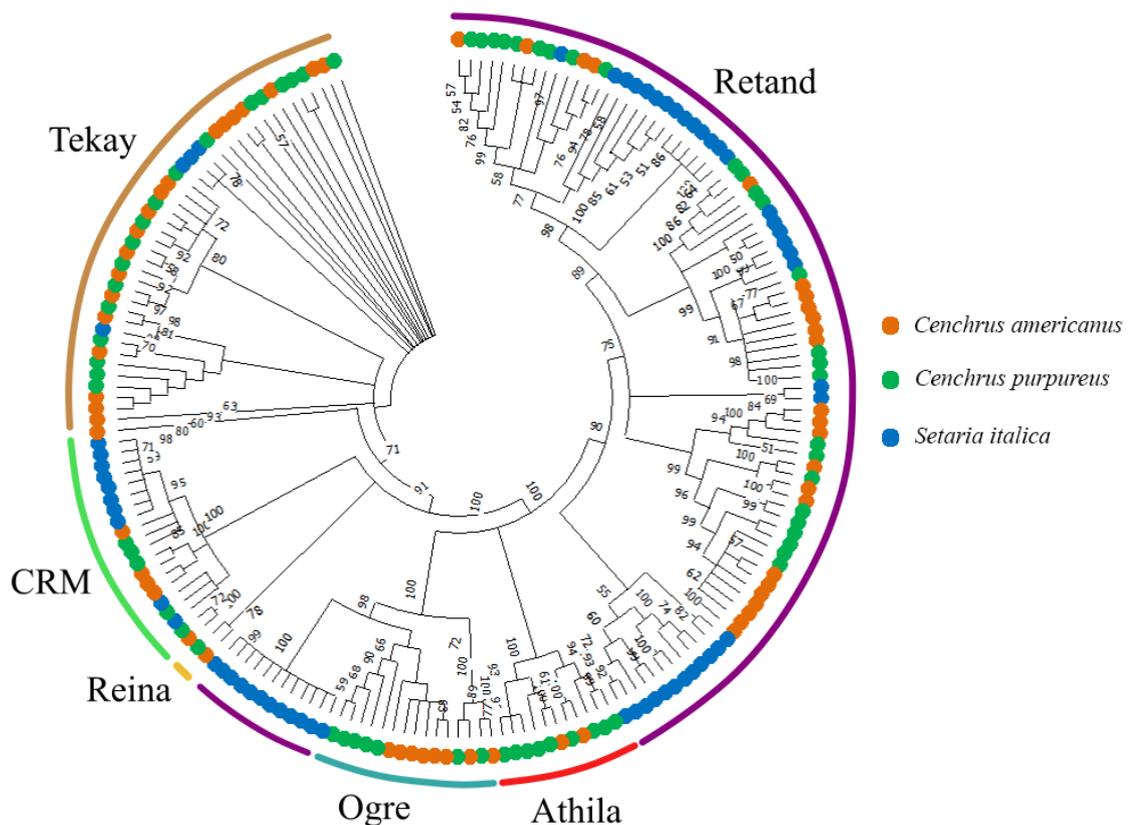


Figura 4 Relações filogenéticas entre os domínios RT de famílias do Retroelemento *Ty3-gypsy*. Valores de *bootstrap* se encontram posicionados nos nós de cada ramo da árvore.

Caracterização e análise comparativa das sequências de DNA satélite identificadas no genoma de *Cenchrus purpureus* e *Cenchrus americanus*

A análise individual da fração repetitiva do genoma de ambas as espécies revelou, inicialmente, 16 *clusters* de DNA satélite, dez oriundos de *Cenchrus purpureus* e seis de *Cenchrus americanus*. Para o estabelecimento de possíveis relações entre essas sequências, cada um dos distintos monômeros foram utilizados em alinhamentos múltiplos locais.

Os DNA satélites identificados foram agrupados em dez famílias e duas super famílias (Tabela 2). Na família 1 (F1; Sufixo A) foram agrupados um total de sete DNA satélites, sendo que dentre esses, três foram identificados como uma mesma variante de um único DNA satélite pertencentes a CpaSat1A. Em CpaSat1A dois satélites identificados no genoma de *Cenchrus americanus* (CpaSat1A_VarCa e Ca_CL15) e um no genoma de *Cenchrus purpureus* (CpaSat1A_VarCp) foram agrupados como mesma variante de um único DNA satélite (CpaSat1A) (Figura 5). Todos os satélites da família F1 possuem comprimento médio entre 137 e 156 pb e quantidades G/C similares. Na família 2 (F2; Sufixo B) foram agrupados dois DNA satélites com tamanhos de 334 pb e 335 pb. O conteúdo G/C desses satélites é de 50% e 46,4% respectivamente. A super família 1 (SF1; Sufixo Y) engloba seis famílias de DNA satélite com distintos comprimentos e porcentagens de G/C, assim como a super família 2 (SF2; Sufixo Z), que por sua vez compreende duas famílias de DNA satélite.

A partir das buscas por similaridade via BLAST, foram identificados apenas dois DNA satélites com homologia com uma sequência satélite presente em *Cenchrus americanus* previamente publicada: *Gla's cen* (KAMM; SCHMIDT; HESLOP-HARRISON, 1994; ISHII et al., 2012). Os satélites que obtiveram *hits* com *Gla's cen* foram CpaSat1A_VarCp e CpurSat3Y, ambos presentes em *Cenchrus purpureus*. Posteriormente, através dos alinhamentos locais, CpurSat3Y foi descartada como um possível satélite pertencente à mesma família de *Gla's cen* por possuir similaridade inferior a 80%. *Gla's cen* está depositado no banco de dados NCBI sob o acesso Z23007.1.

Por *Gla's cen* se tratar de um DNA satélite originalmente isolado do genoma de *Cenchrus americanus* e os resultados BLAST não indicarem homologia dessa sequência com nenhum dos satélites identificados na espécie, foi realizado o mapeamento manual de *Gla's cen* contra os *reads* referentes a *Cenchrus americanus*. Foram identificadas duas sequências de *Cenchrus americanus* relacionadas à *Gla's cen* nos *reads*: uma com identidade superior a 95%, considerada uma mesma variante de *Glas'cen* (CameSat1A), e outra com identidade inferior a

95% e superior a 80%, considerada da mesma família de *Gla's cen*, presente em *Cenchrus americanus* (CpaSat1A_VarCa) e também em *Cenchrus purpureus* (CpaSat1A_VarCp) (Figura 5). CpaSat1A compreende, portanto, três variantes pertencentes a mesma família de *Gla's cen*: CpaSat1A_VarCa, CpaSat1A_VarCp e Ca_CL4. Por Ca_CL4 e CpaSat1A_VarCa serem ambas uma mesma variante dentro de *Cenchrus americanus*, apenas a segunda nomenclatura foi mantida nas análises.

Para efeitos de comparação, em relação à presença ou ausência de cada DNA satélite identificado no genoma de ambas as espécies, foi realizado o mapeamento de todos os monômeros extraídos contra os *reads* de *Cenchrus purpureus* e *Cenchrus americanus* por se tratar de uma abordagem de quantificação mais precisa. As proporções relatadas pela análise através do *RepeatExplorer* e pelo mapeamento direto contra os *reads* oriundos do sequenciamento foram variáveis. Isso se deve ao fato de que análises realizadas pela plataforma *RepeatExplorer* podem subestimar a abundância e a diversidade das sequências de DNA satélite presentes, o que não inviabiliza a sua utilização, uma vez que uma visão geral e estimada da fração repetitiva é proposta (NOVÁK; NEUMANN; MACAS, 2010; GONZÁLEZ et al., 2020).

Tabela 2 Sequências de DNA satélite identificadas no genoma de *Cenchrus purpureus* e *Cenchrus americanus*

F/SF	DNASat	V	Comprimento (pb)	GC (%)	Abundância	
					Cp	Ca
F1	CpaSat1A	3	138	44.2	35629	28201
F1	CpurSat2A	1	148	42.6	33216	25900
SF1	CpurSat3Y	1	156	43.6	31812	25967
F1	CpurSat4A	1	156	41.7	30664	25987
SF1	CpurSat5Y	1	336	42.6	7534	0
SF2	CpurSat6Z	1	154	51.3	7128	11419
SF1	CpurSat7Y	1	372	43.0	6395	0
F2	CpurSat8B	1	334	50.0	1887	11879
SF2	CpurSat9Z	1	528	39.8	1069	0
SF1	CpurSat10Y	1	364	43.7	272	0
F1	CameSat1A	1	137	43.8	31857	26206
F2	CameSat2B	1	335	46.3	1865	11887
SF1	CameSat3Y	1	331	38.7	1435	336
SF1	CameSat4Y	1	1314	48.1	18	220

Legenda: F/SF: Família/Super família. V: Número de variantes. GC (%): Porcentagem de bases GC. Abundância: Número de *reads* mapeados contra cada monômero. Cp: *Cenchrus purpureus*. Ca: *Cenchrus americanus*. CpurSat: Satélites originalmente extraídos do genoma de *Cenchrus purpureus*. CameSat: Satélites originalmente extraídos de *Cenchrus americanus*. CpaSat1A: Satélite que possui variantes em ambos os genomas.

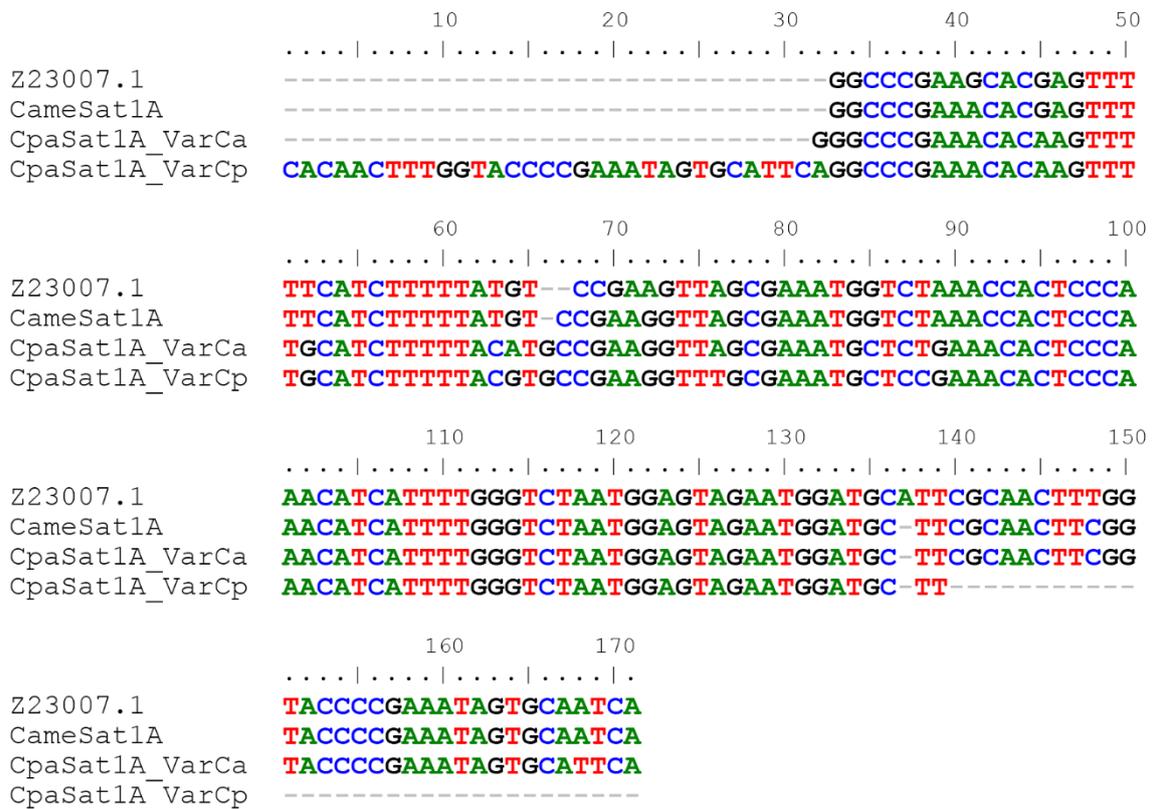


Figura 5 Alinhamento local entre *Gla's cen* e DNA satélites relacionados isolados do genoma de *Cenchrus americanus* e *Cenchrus purpureus*. Z23007.1: *Gla's cen*. CpaSat1A_VarCa: Variante oriunda do genoma de *Cenchrus americanus* inclusa em CpaSat1A. CpaSat1A_VarCp: Satélite oriundo do genoma de *Cenchrus purpureus* incluso em CpaSat1A.

De acordo com os dados obtidos através do mapeamento, *Cenchrus purpureus* possui em seu genoma todos os satélites identificados, inclusive os originalmente extraídos de *Cenchrus americanus*. Em F1 é possível notar um padrão de abundância nos DNA satélites, em que todos obtiveram mais de 33.000 e menos de 35.000 *reads* mapeados, indicando uma grande abundância dessa família no genoma da espécie. Já a família F2 aparece com valores reduzidos de abundância, mas que são relativamente próximos. SF1 possui grande heterogeneidade no padrão de abundância para cada família de DNA satélite, variando de 18 *reads* mapeados para CameSat4Y até 31.812 para CpurSat3Y. SF2 apresenta abundâncias relativamente distintas entre as duas famílias de DNA satélite presentes. *Cenchrus purpureus* possui ainda DNA satélites exclusivos quando comparados aos de *Cenchrus americanus*: CpurSat5Y, CpurSat7Y, CpurSat9Z e CpurSat10Y. Três dos quatro satélites exclusivos de *Cenchrus purpureus* estão compreendidos em SF1, o quarto está em SF2 (Tabela 2, Figura 6).

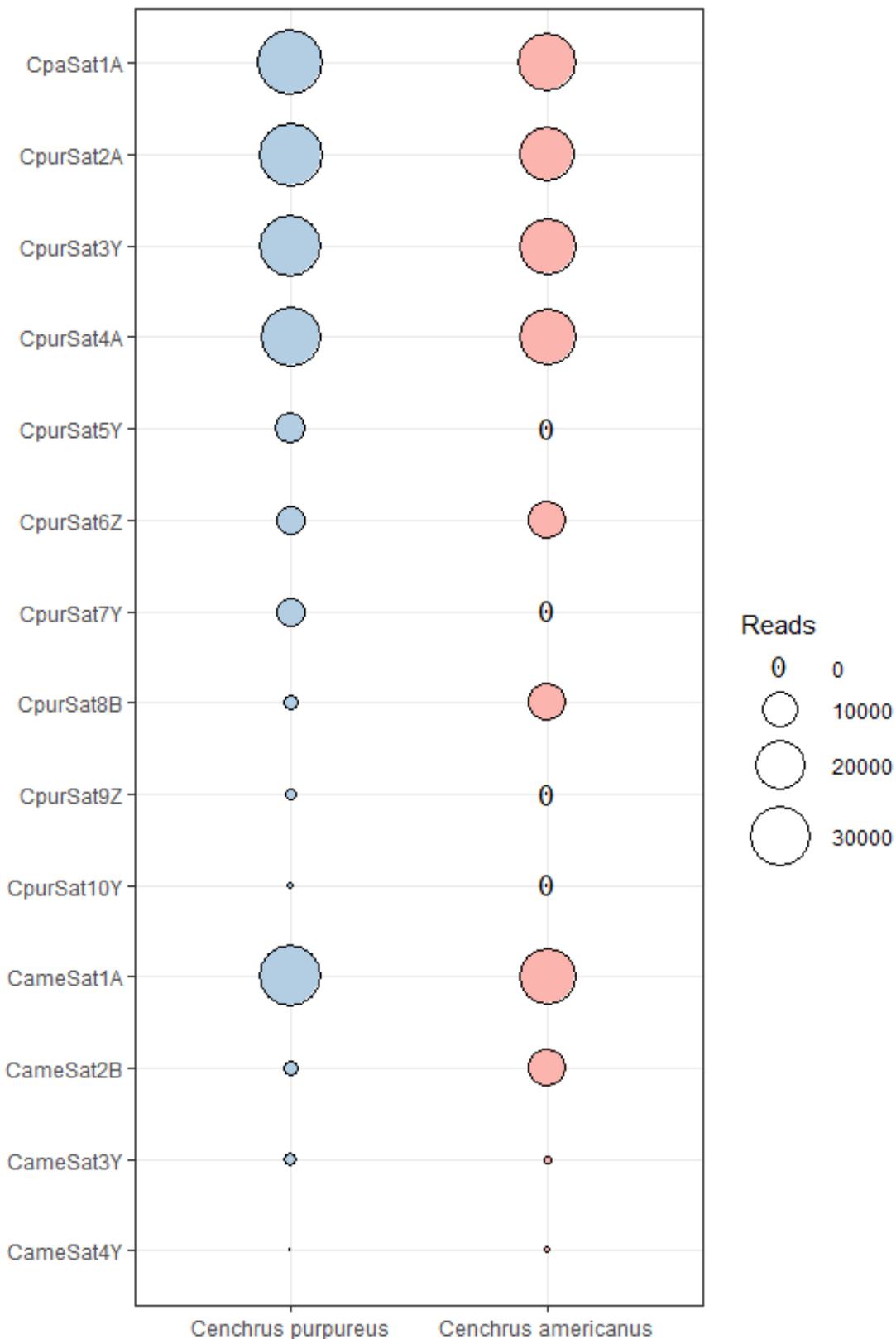


Figura 6 Gráfico *Bubble Chart* evidenciando a abundância de *reads* mapeados em relação aos monômeros de cada DNA satélite identificado para com o genoma de *Cenchrus purpureus* e *Cenchrus americanus*. CpurSat: Satélites originalmente extraídos do genoma de *Cenchrus purpureus*. CameSat: Satélites originalmente extraídos de *Cenchrus americanus*. CpaSat1A: Satélite que possui variantes em ambos os genomas.

4. DISCUSSÃO

Composição e diversidade de elementos transponíveis no genoma de *Cenchrus purpureus* e *Cenchrus americanus*

Avanços nas técnicas de Sequenciamento de Nova Geração (NGS) vêm possibilitando a realização de diversos trabalhos relacionados à caracterização da fração repetitiva, principalmente em genomas de espécies vegetais consideradas não-modelo. Diversas espécies que já possuem sua fração repetitiva caracterizada são de elevada importância agrônoma, como arroz (BILINSKI et al., 2017b; SHENTON et al., 2020), milho (TENAILLON et al., 2011), cevada (WICKER et al., 2017), assim como diversas espécies que são fonte de grãos para alimentação humana e animal (KAPUSTOVÁ et al., 2019). No presente trabalho realizamos o sequenciamento de baixa cobertura e a caracterização completa da fração repetitiva de duas gramíneas que possuem elevada importância para a pecuária global, *Cenchrus purpureus* e para a alimentação humana, *Cenchrus americanus*.

O genoma de espécies vegetais tende a possuir grandes proporções de sequências repetitivas (KIDWELL, 2002; SHENTON et al., 2020), caracterizadas por vezes acima de 80% para genomas considerados relativamente grandes (NOVÁK et al., 2014). As frações de DNA repetitivo de *Cenchrus purpureus* (52,23% - 2,22 Gb) e *Cenchrus americanus* (76,82% - 2,28 Gb) estão dentro da faixa esperada para espécies com tamanho de genoma entre pequeno e intermediário (1,40 Gb – 13,71 Gb) (SOLTIS et al., 2003; NOVÁK et al., 2014). Proporções de sequências repetitivas menores que 80% já foram relatadas para diversas espécies que possuem genomas considerados pequenos ou intermediários (SOLTIS et al., 2003), como algumas coníferas (STEVENS et al., 2016), espécies da ordem Asparagales (HERTWECK, 2013) e também a espécie considerada um fóssil vivo, *Ginkgo biloba* (GUAN et al., 2016).

Dentro das classes que compõe a fração repetitiva, a maior parte do genoma de *Cenchrus purpureus* e *Cenchrus americanus* é composta pelos elementos transponíveis, sendo os retroelementos mais abundantes quando comparados aos transposons de DNA. Os elementos LTR são os mais abundantes em ambas as espécies, como relatado para grande parte dos genomas vegetais, nas quais os retroelementos são os principais responsáveis pelo incremento em tamanho e os LTRs a principal ordem envolvida (WICKER et al., 2007; BENNETZEN; WANG, 2014). Alguns autores já propuseram esse mesmo padrão de abundância para *Cenchrus americanus* (VARSHNEY et al., 2017), *Cenchrus purpureus* (PAUDEL et al., 2018), assim

como para *Cenchrus ciliaris* e *Cenchrus squamulatus* (Fresen.) Morrone (CONNER et al., 2008), outras duas espécies do gênero.

Em relação às super famílias de retrotransposons, *Ty3-gypsy* apresenta valores maiores que duas vezes a abundância de *Ty1-copia* em ambas as espécies, sendo caracterizada como as super famílias mais abundantes nesses genomas. Essa relação é similar às descritas em diversas espécies vegetais (PAMPONÉT et al., 2019; SHENTON et al., 2020), sendo que a maioria das angiospermas apresentam maiores abundâncias de *Ty3-gypsy* em relação em *Ty1-copia*, com raras exceções, como é o caso de algumas espécies do gênero *Quercus* L. (MASCAGNI et al., 2020).

Para a classe dos transposons, a família que apresenta maior abundância em ambas as espécies é *CACTA*, o que corrobora com o proposto por Conner e colaboradores (2008) para outras espécies do gênero. Os autores também propõem baixas abundâncias de transposons da subclasse II em *Cenchrus ciliaris* e *Cenchrus squamulatus*, assim como relatado em nossos resultados (0,09% em *Cenchrus purpureus* e 0,05% em *Cenchrus americanus*).

Através da análise dos *clusters* exclusivos e compartilhados entre *Cenchrus purpureus* e *Cenchrus americanus* é notável a elevada similaridade entre seus genomas, principalmente em relação à fração repetitiva dos mesmos. Essa similaridade está relacionada aos *clusters* identificados como elementos transponíveis nas análises e que são, em sua totalidade, compartilhados entre ambas as espécies. A relação ancestral comum de *Cenchrus americanus* e *Cenchrus purpureus* (MARTEL et al., 2004) pode explicar estes dados, no qual, os genomas e sub genomas que compõe ambas as espécies são intimamente relacionados, tanto entre os homeólogos A (*Cenchrus americanus*) e A' (*Cenchrus purpureus*), quanto entre os mesmos para com B. (HELENA; DAVIDE; PEREIRA, 2005; DOS REIS et al., 2014). Paudel e colaboradores (2018) relataram elevada colinearidade entre o genoma de ambas as espécies, assim como Dos Reis e colaboradores (2014) relatam elevada similaridade, corroborando com tais resultados.

Ao contrário do esperado, *Cenchrus americanus*, uma espécie diploide, possui a fração repetitiva relativamente mais abundante que a de *Cenchrus purpureus*, um alotetraploide. Em relação aos elementos transponíveis, a primeira espécie possui cerca de 1,6 vezes mais abundância ao ser comparada à segunda espécie. As principais hipóteses que podem explicar tais números também residem na ancestralidade dessas espécies, devido ao ancestral comum de ambas ser um provável poliploide (MARTEL et al., 2004; YAN et al., 2020). Para *Cenchrus americanus* existem relatos de que se trata de um paleopoliploide, que, após ciclos de poliploidização-diploidização se comporta como um diploide (MARTEL et al., 2004; TECHIO;

DAVIDE; PEREIRA, 2006). Para *Cenchrus purpureus*, diversos autores propõem se tratar de um alopoliploide recente que se originou após a diferenciação de *Cenchrus americanus* (MARTEL et al., 2004; DOS REIS et al., 2014; YAN et al., 2020), através da hibridação entre o ancestral comum das duas espécies, de genoma A, e um ancestral desconhecido de genoma B.

Processos que resultam em alterações cromossômicas numéricas e/ou estruturais, como a duplicação genômica ou a diploidização acarretam estresses genômicos. Frequentemente, retrotransposons e transposons são ativados e amplificados devido a tais estresses. No caso dos retroelementos, essa ativação causa o aumento exponencial da fração repetitiva e, conseqüentemente, do tamanho do genoma (BELYAYEV, 2014). Essa é uma hipótese que pode explicar *Cenchrus americanus* possuir um genoma relativamente maior e com maiores abundâncias de elementos transponíveis quando comparado a *Cenchrus purpureus*, uma vez que a primeira passou por ciclos de poliploidização-diploidização, enquanto a segunda se mantém como um alopoliploide recente, iniciando o processo de diploidização (TECHIO; DAVIDE; PEREIRA, 2006). A amplificação diferencial de linhagens de retrotransposons pós especiação relatada nas análises filogenéticas envolvendo famílias de *Ty3-gypsy* no presente trabalho, corrobora com essa hipótese, indicando um possível aumento na diversidade e conseqüentemente no número de cópias desses elementos no genoma de *Cenchrus americanus*.

Aumentos da fração repetitiva devido a processos de poliploidização-diploidização também são descritos para espécies de *Nicotiana* (RENNY-BYFIELD et al., 2013) e *Gossypium* (HAWKINS et al., 2006). Existem ainda relatos de alterações cromossômicas e genômico estruturais, como rearranjos e deleções, descritas para *Cenchrus americanus* (MARTEL et al., 2004; DOS REIS et al., 2014; VARSHNEY et al., 2017; PAUDEL et al., 2018), corroborando com os resultados obtidos pela análise comparativa.

Cenchrus purpureus, um recente alopoliploide quando comparado a *Cenchrus americanus*, passou por um recente ciclo de poliploidização (MARTEL et al., 2004; YAN et al., 2020). Poliploides recentes tendem a perder sequências repetitivas que podem levar à diferenciação entre os cromossomos e ao comportamento diploide já descrito para *Cenchrus purpureus* através da formação de associações cromossômicas bivalentes na meiose (JAUHAR, 1981; TECHIO; DAVIDE; PEREIRA, 2006). Esse pode ser um dos caminhos trilhado por *Cenchrus purpureus* em relação a seu ancestral comum com *Cenchrus americanus*. *Cenchrus purpureus* vem possivelmente sofrendo, ao longo de seu período evolutivo, tais processos de perda de sequências, fazendo com que *Cenchrus americanus* sempre possua maiores frações genômicas mesmo que ambos sejam oriundos da diferenciação de um mesmo ancestral poliploide (MARTEL et al., 2004; DOS REIS et al., 2014; VARSHNEY et al., 2017; PAUDEL

et al., 2018). Perdas de sequências em poliploides recentes são descritas para aloploiploides de *Nicotiana* spp. (RENNY-BYFIELD et al., 2012, 2013), e tetraploides de *Arabidopsis lyrata* (KUITTINEN et al., 2004), um evento mais frequente ainda quando se trata de poliploides formados através de processos de hibridação, como é o caso de *Cenchrus purpureus* (WENDEL, 2000).

Por se tratar da super família de retroelementos mais abundante em ambos os genomas, famílias de *Ty3-gypsy* foram utilizadas para análises de diversidade de retrotransposons entre espécies. Linhagens de *Retand*, *CRM* e *Tekay* estão presentes no genoma de *Cenchrus purpureus*, *Cenchrus americanus* e *Setaria italica*, indicando sua presença no ancestral comum entre as três espécies, assim como indicando um possível processo de diversificação e amplificação dessas linhagens pós especiação. Diversos autores propõem uma relação próxima entre espécies de *Setaria* e de *Cenchrus/Pennisetum* (MARTEL et al., 2004; CHEMISQUY et al., 2010), assim como também existem fortes evidências de que *Setaria italica* compartilha um mesmo ancestral com *Cenchrus americanus* (BENNETZEN et al., 2012) e, conseqüentemente, com *Cenchrus purpureus* (MARTEL et al., 2004). Resultados semelhantes, com agrupamento de retrotransposons por linhagens e não espécies, já foram descritos para espécies de *Populus* L. (USAI et al., 2017).

Linhagens de *Athila*, *Ogre* e *Reina* estão presentes apenas em *Cenchrus purpureus* e *Cenchrus americanus*, o que pode indicar uma especificidade das mesmas em *Cenchrus*, assim como um processo de amplificação pós especiação, uma vez que, apesar de partilharem um mesmo ancestral, seus processos de diversificação ocorreram de forma distinta aos de *Setaria* (BENNETZEN et al., 2012). Outra hipótese está relacionada à amostragem aleatória dos domínios proteicos RT no banco de dados REXdb de *Setaria italica*, o que pode ter sub representado as linhagens pertencentes a tais super famílias na espécie.

Analisando mais especificamente, linhagens de *Retand* se diferenciaram entre as três espécies, evidenciando a presença de possíveis marcas espécie-específicas. Um agrupamento *Retand* próximo a tais linhagens específicas apresenta uma mistura entre linhagens de *Cenchrus americanus* e *Cenchrus purpureus*, no qual alguns nós apresentam período de diferenciação anterior ao agrupamento das linhagens específicas de cada uma das espécies. Esses resultados indicam provável localização dessas linhagens que se diferenciaram anteriormente às específicas, nos genomas A de *Cenchrus americanus* e A' de *Cenchrus purpureus*, de acordo com a filogenia e ancestralidade de ambas as espécies. Autores propõem a ancestralidade dos genomas A e A' pautados apenas em um doador de composição A, como um evento anterior à hibridação que originou *Cenchrus purpureus* (MARTEL et al., 2004; YAN et al., 2020). Sendo

assim, com períodos de diferenciação mais antigos, essas linhagens presentes em *Cenchrus purpureus* e *Cenchrus americanus* em um mesmo clado, e, portanto, similares, se diferenciaram anteriormente ao processo de hibridação que originou *Cenchrus purpureus*, estando presentes em ambas as espécies.

Ainda relacionando as linhagens de retrotransposons à ancestralidade das espécies, em *Retand*, as linhagens específicas de *Cenchrus purpureus* se diferenciaram anteriormente às presentes em *Cenchrus americanus*, indicando uma possível localização das mesmas nos cromossomos pertencentes ao genoma B, que se diferenciou de forma independente em relação ao genoma A anteriormente à origem de *Cenchrus purpureus*. (MARTEL et al., 2004; HELENA; DAVIDE; PEREIRA, 2005; DOS REIS et al., 2014). Linhagens de *Retand* em *Setaria italica* divergiram concomitantemente às de *Cenchrus purpureus* e *Cenchrus americanus*. Tais resultados vão de encontro ao proposto por Bennetzen e colaboradores (2012), no qual *Cenchrus americanus* e *Setaria italica* se divergiram em períodos próximos a partir de um mesmo ancestral. Linhagens de *Tekay* não possuem agrupamentos específicos entre espécies, estando assim possivelmente expostas a eventos de diversificação e amplificação distintos para cada uma das três espécies, assim como proposto para outras espécie como arroz, trigo e *Arabidopsis* (WICKER; KELLER, 2007).

Descrição e análise das sequências de DNA Satélite identificadas nos genomas de *Cenchrus purpureus* e *Cenchrus americanus*

Cenchrus americanus e *Cenchrus purpureus*, como recorrentemente relatado para espécies vegetais, possuem baixas abundâncias de sequências satélite (4,17 e 2,55%, respectivamente) quando comparadas aos elementos transponíveis (70,38 e 46,68%). Baixas proporções de sequências de DNA satélite já foram descritas para diversas gramíneas como arroz, milho e sorgo (BILINSKI et al., 2017a). Comparativamente, *Cenchrus americanus* possui maior proporção dessas sequências em relação a *Cenchrus purpureus*. Uma hipótese que possivelmente explica esses resultados diz respeito à perda de sequências repetitivas, inclusive sequências satélite, em poliploides recentes, como é o caso de *Cenchrus purpureus* (SHARMA; RAINA, 2005).

Kamm e colaboradores (1994) analisando a fração repetitiva de *Cenchrus americanus* identificaram um monômero de DNA satélite de aproximadamente 137 pb localizado em região centromérica de todos cromossomos da espécie. Posteriormente, Ishii e colaboradores (2012) analisando esse mesmo satélite, propuseram a denominação do mesmo como *Gla's cen*. Nas

análises de similaridade por BLAST, CameSat1A obteve identidade superior a 95% com *Gla's cen*, indicando ser uma mesma variante deste satélite. Além deste monômero, todos os três agrupados em CpaSat1A, incluindo a variante isolada de *Cenchrus purpureus*, apresentaram *hits* para com *Gla's cen* através das análises BLAST. Os demais satélites presentes na família 1 não obtiveram *hits* em relação ao satélite isolado por Kamm e colaboradores (1994), porém, a partir das identidades obtidas através dos alinhamentos, foram agrupados como pertencentes a mesma família da variante de *Gla's cen* identificada no presente trabalho, CameSat1A. Esses resultados indicam maior similaridade entre os satélites CpaSat1A e CameSat1A do que para com os demais presentes na família 1, tanto em *Cenchrus americanus* quanto em *Cenchrus purpureus*, uma vez que CameSat1A se trata de uma mesma variante de *Gla's cen*.

Uma hipótese que pode explicar as relações entre os satélites presentes na Família 1, especificamente em *Cenchrus americanus*, diz respeito aos elevados níveis de similaridade entre sequências satélite pertencentes a uma mesma família. Essa hipótese se baseia na origem comum e na evolução concertada dessas sequências, onde pode ocorrer baixas ou elevadas taxas de homogeneização e amplificação de mutações entre sequências de monômeros adjacentes em uma mesma família de DNA satélite, levando a eliminação ou fixação dessas variantes (PLOHL, 2010).

Essa teoria é clara de se observar na espécie, uma vez que CpaSat1A e CameSat1A possuem maiores valores de identidade quando comparadas ao restante dos satélites inclusos na família 1, assim como obtiveram *hits* para com *Gla's cen* (uma variante de CpaSat1A). Esses monômeros possivelmente tiveram menores taxas de mutação e maiores taxas de homogeneização, fazendo com que permanecessem com elevado grau de similaridade entre eles. Também é possível observar maiores similaridades em relação ao comprimento e proporção G/C entre CpaSat1A e CameSat1A em relação ao restante de F1. O agrupamento dos outros membros de F1 em uma mesma família, assim como ocorre em F2, também podem ser explicados a luz dessa teoria, porém, nesse caso, algumas mutações foram fixadas ou menos homogeneizadas entre os monômeros. Diversos autores já propuseram esse mecanismo agindo no genoma de diversas espécies vegetais, como Garrido-Ramos e colaboradores (2017) e Samoluk e colaboradores (2009).

A relação entre a variante de *Cenchrus purpureus* presente em CpaSat1A (CpaSat1A_VarCp) para com *Gla's cen*, assim como para com o restante dos monômeros de F1 originalmente isolados de *Cenchrus americanus*, e vice-versa, pode ser explicada pela hipótese de que existe uma “biblioteca” de monômeros de DNA satélites comuns entre espécies filogeneticamente relacionadas, como *Cenchrus purpureus* e *Cenchrus americanus*. Essa teoria

foi inicialmente proposta por Fry e Salser (1977) e posteriormente comprovada por diversos autores (KAPUSTOVÁ et al., 2019; LIU et al., 2019; HEITKAM et al., 2020). Plohl (2010) explica que, a partir dessa “biblioteca”, vários monômeros podem se diferenciar entre espécies, mantendo certa homologia, como acontece com CpaSat1A_VarCp e *Gla's cen*/CameSat1A, ou podendo chegar em certo nível de distinção, no qual, a relação entre as sequências compartilhadas é drasticamente reduzida devido aos efeitos da evolução concertada.

A localização centromérica de *Gla's cen* relatada por Kamm e colaboradores (1994) e Ishii e colaboradores (2012) possivelmente pode ser assumida para as sequências CpaSat1A, e, principalmente, CameSat1A, presentes tanto em *Cenchrus americanus*, quanto em *Cenchrus purpureus*. A primeira hipótese que sustenta essa proposta está relacionada à similaridade da região centromérica e pericentromérica de ambas as espécies relatada por Dos Reis e colaboradores (2014) ao analisar por GISH, os cromossomos de *Cenchrus americanus*, *Cenchrus purpureus* e do híbrido interespecífico, corroborando com a presença de *repeats* similares nessas regiões entre as duas espécies. Outra hipótese que corrobora com essa proposta foi relatada por Melters e colaboradores (2013), no qual os autores sugerem que sequências satélites centroméricas são comumente as mais abundantes na fração repetitiva do genoma de grande parte dos organismos. Essa afirmação é condizente com as abundâncias de CpaSat1A e CameSat1A, no qual o primeiro é o mais abundante entre todos os outros satélites originalmente isolados de *Cenchrus purpureus*, e o segundo o mais abundante entre todos os presentes em *Cenchrus americanus*. A elevada similaridade entre os monômeros de CameSat1A para com *Gla's cen* e CpaSat1A para com CameSat1A também corrobora com o proposto.

A Família 2 possui dois monômeros agrupados, no qual, um deles foi isolado de *Cenchrus purpureus* e o outro de *Cenchrus americanus*. Ambos possuem características similares em relação ao comprimento e proporção de bases G/C, podendo essas relações também serem explicadas pela teoria da “biblioteca” comum de DNA satélites entre as duas espécies. As super famílias 1 e 2, por sua vez, englobam famílias de DNA satélite com características heterogêneas entre si, indicando maior diferenciação dessas famílias devido, provavelmente, a processos de diferenciação entre as mesmas que são explicados a luz da evolução concertada.

A presença de famílias de DNA satélite compartilhadas e exclusivas entre o genoma de *Cenchrus purpureus* e *Cenchrus americanus* está de acordo com os relatos de similaridade na composição genômica dessas duas espécies propostos por Helena e colaboradores (2005), assim como por Dos Reis e colaboradores (2014). Existe a possibilidade ainda de que os satélites exclusivos de *Cenchrus purpureus* estejam presentes apenas no genoma B da espécie, enquanto

os comuns estejam localizados em A de *Cenchrus americanus* e A' de *Cenchrus purpureus*, que são genomas homeólogos. A técnica de Hibridização *in situ* Genômica (GISH) utilizada por Dos Reis e colaboradores (2014), demonstra homologia relacionada principalmente às sequências repetitivas presentes nos cromossomos analisados, aproximando ainda mais nossos resultados com os propostos pelos autores. A evolução independente de sequências satélites presentes em genomas não relacionados pode explicar esses resultados ao considerarmos os genomas A e B, que divergiram de forma independente (PLOHL, 2010). Especificidade de sequências satélite para genomas e sub genomas já foram relatadas por Heitkam e colaboradores (2020) em espécies de *Quenopodium*, assim como por Liu e colaboradores (2019) para espécies de *Avena*.

As abundâncias relacionadas a cada um dos monômeros e das famílias de DNA satélite se enquadram nas hipóteses propostas anteriormente baseadas na “biblioteca” comum de DNA satélites presente em ambas as espécies, assim como com a evolução concertada dessas sequências.

As sequências que codificam para RNA ribossomal, 18S, 5.8S e 26S, conhecidas como rDNA 45S, assim como as sequências de rDNA 5S, são arranjos gênicos organizados em *tandem* e por isso identificados como pertencentes ao grupo de satélites (EICKBUSH; EICKBUSH, 2007). O somatório das proporções de *reads* referentes a rDNA 45S e rDNA 5S são maiores em *Cenchrus americanus* quando comparados a *Cenchrus purpureus* (1,36% e 0,51% respectivamente).

Andrade-Vieira (2010) utilizou da técnica de Hibridização *in situ* fluorescente (FISH) para localizar nos cromossomos de *Cenchrus purpureus* e *Cenchrus americanus* os arranjos gênicos de rDNA 45S e rDNA 5S. A autora descreve para *Cenchrus purpureus* dois pares de sinais de rDNA 45S e apenas um par de rDNA 5S presente nos cromossomos da espécie. Para *Cenchrus americanus* a autora descreveu dois pares de sinais de rDNA 45S e dois pares de sinais de rDNA 5S. A maior quantidade de locos referentes a rDNA 5S em *Cenchrus americanus* pode explicar a maior proporção dessas sequências na espécie, corroborando com nossos resultados.

As proporções relatadas no presente trabalho para as sequências de rDNA 45S são maiores em *Cenchrus americanus* quando comparadas a *Cenchrus purpureus*, apesar de ambas as espécies possuírem o mesmo número de sinais em seus cromossomos como proposto por Andrade-Vieira (2010). Essa não concordância pode ser explicada pela intensidade dos sinais relatados pela autora, no qual, para *Cenchrus americanus*, o sinal de um dos pares

cromossômicos ocupa uma região do braço cromossômico maior em relação a *Cenchrus purpureus*, assim como apresenta maior intensidade, e portanto, maior número de cópias.

5. CONCLUSÕES

A fração repetitiva do genoma de *Cenchrus purpureus* e *Cenchrus americanus* corresponde, respectivamente, a 52,23 e 76,82% do genoma.

Os elementos repetitivos mais abundantes em ambas as espécies foram classificados como LTR, sendo o número de cópias dos elementos pertencentes à super família *Ty3-gypsy* superior aos do elemento *Ty1-gypsy*.

Foram identificadas 10 famílias de DNA satélite presentes nos genomas de *Cenchrus purpureus* e *Cenchrus americanus*, sendo 4 dessas exclusivas de *Cenchrus purpureus*.

Sugere-se que a ancestralidade comum, assim como a similaridade genômica entre *Cenchrus purpureus* e *Cenchrus americanus* e os processos de poliploidização e diploidização inerente a essas espécies, afetaram diretamente a homogeneização e amplificação das famílias de elementos transponíveis presentes no genoma de ambas.

Os processos evolutivos que moldam as relações entre monômeros de DNA satélite estão intrinsicamente envolvidos na diferenciação e no compartilhamento das famílias de DNA satélite identificadas em *Cenchrus purpureus* e *Cenchrus americanus*.

REFERÊNCIAS

- ANDRADE-VIEIRA, L. F. ANÁLISE CITOMOLECULAR EM HÍBRIDOS DE CAPIM-ELEFANTE E MILHETO (*Pennisetum* sp. Schum., Poaceae). **Tese em Genética e Melhoramento de Plantas**. Universidade Federal de Lavras, 2010.
- BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 1, p. 11, 2 jun. 2015. doi: 10.1186/s13100-015-0041-9.
- BELYAYEV, A. Bursts of transposable elements as an evolutionary driving force. **Journal of Evolutionary Biology**, v. 27, n. 12, p. 2573–2584, dez. 2014. doi: 10.1111/jeb.12513.
- BELYAYEV, A.; PAŠTOVÁ, L.; FEHRER, J.; JOSEFIOVÁ, J.; CHRTEK, J.; MRÁZ, P. Mapping of *Hieracium* (Asteraceae) chromosomes with genus-specific satDNA elements derived from next-generation sequencing data. **Plant Systematics and Evolution**, v. 304, n. 3, p. 387–396, 29 mar. 2018. doi: 10.1007/s00606-017-1483-y.
- BENNETZEN, J. L.; SCHMUTZ, J.; WANG, H.; PERCIFIELD, R.; HAWKINS, J.; PONTAROLI, A. C.; ESTEP, M.; FENG, L.; VAUGHN, J. N.; GRIMWOOD, J.; JENKINS, J.; BARRY, K.; LINDQUIST, E.; HELLSTEN, U.; DESHPANDE, S.; WANG, X.; WU, X.; MITROS, T.; TRIPLETT, J.; YANG, X.; YE, C.-Y.; MAURO-HERRERA, M.; WANG, L.; LI, P.; SHARMA, M.; SHARMA, R.; RONALD, P. C.; PANAUD, O.; KELLOGG, E. A.; BRUTNELL, T. P.; DOUST, A. N.; TUSKAN, G. A.; ROKHSAR, D.; DEVOS, K. M. Reference genome sequence of the model plant *Setaria*. **Nature Biotechnology**, v. 30, n. 6, p. 555–561, 13 jun. 2012. doi: 10.1038/nbt.2196.
- BENNETZEN, J. L.; WANG, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. **Annual Review of Plant Biology**, v. 65, p. 505–530, 2014. doi: 10.1146/annurev-arplant-050213-035811.
- BILINSKI, P.; HAN, Y.; HUFFORD, M. B.; LORANT, A.; ZHANG, P.; ESTEP, M. C.; JIANG, J.; ROSS-IBARRA, J. Genomic abundance is not predictive of tandem repeat localization in grass genomes. **PLoS ONE**, v. 12, n. 6, p. 1–12, 2017a. doi: 10.1371/journal.pone.0177896.
- BROWN, J.; PIRRUNG, M.; MCCUE, L. A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. **Bioinformatics**, v. 33, n. 19, p. 3137–3139, 1 out. 2017. doi: 10.1093/bioinformatics/btx373.
- CHEMISQUY, M. A.; GIUSSANI, L. M.; SCATAGLINI, M. A.; KELLOGG, E. A.; MORRONE, O. Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): A combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. **Annals of Botany**, v. 106, n. 1, p. 107–130, 2010. doi: 10.1093/aob/mcq090.
- CONNER, J. A.; GOEL, S.; GUNAWAN, G.; CORDONNIER-PRATT, M.-M.; JOHNSON, V. E.; LIANG, C.; WANG, H.; PRATT, L. H.; MULLET, J. E.; DEBARRY, J.; YANG, L.; BENNETZEN, J. L.; KLEIN, P. E.; OZIAS-AKINS, P. Sequence Analysis of Bacterial Artificial Chromosome Clones from the Apospory-Specific Genomic Region of *Pennisetum*

and *Cenchrus*. **Plant Physiology**, v. 147, n. 3, p. 1396–1411, jul. 2008. doi: 10.1104/pp.108.119081.

DONADÍO, S.; GIUSSANI, L. M.; KELLOGG, E. A.; ZUOLAGA, F. O.; MORRONE, O. A. A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the trnL-F, rpl16 chloroplast markers. **Taxon**, v. 58, n. 2, p. 392–404, 2009.

DOS REIS, G. B.; MESQUITA, A. T.; TORRES, G. A.; ANDRADE-VIEIRA, L. F.; PEREIRA, A. Vander; DAVIDE, L. C. Genomic homeology between *Pennisetum purpureum* and *Pennisetum glaucum* (Poaceae). **Comparative Cytogenetics**, v. 8, n. 3, p. 199–209, 2014. doi: 10.3897/CompCytogen.v8i3.7732.

EICKBUSH, T. H.; EICKBUSH, D. G. Finely orchestrated movements: Evolution of the ribosomal RNA genes. **Genetics**, v. 175, n. 2, p. 477–485, 2007. doi: 10.1534/genetics.107.071399.

FRY, K.; SALSER, W. Nucleotide sequences of HS- α satellite DNA from kangaroo rat *dipodomys ordii* and characterization of similar sequences in other rodents. **Cell**, v. 12, n. 4, p. 1069–1084, 1 dez. 1977. doi: 10.1016/0092-8674(77)90170-2.

GAIERO, P.; VAIO, M.; PETERS, S. A.; SCHRANZ, M. E.; DE JONG, H.; SPERANZA, P. R. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. **Annals of Botany**, v. 123, n. 3, p. 521–532, 15 fev. 2019. doi: 10.1093/aob/mcy186.

GARRIDO-RAMOS, M. Satellite DNA: An Evolving Topic. **Genes**, v. 8, n. 9, p. 230, 18 set. 2017. doi: 10.3390/genes8090230.

GONZÁLEZ, M. L.; CHIAPELLA, J.; TOPALIAN, J.; URDAMPILLETA, J. D. Genomic differentiation of *Deschampsia antarctica* and *D. cespitosa* (Poaceae) based on satellite DNA. **Botanical Journal of the Linnean Society**, p. 1–16, 4 jul. 2020. doi: 10.1093/botlinnean/boaa045.

GUAN, R.; ZHAO, Y.; ZHANG, H.; FAN, G.; LIU, X.; ZHOU, W.; SHI, C.; WANG, J.; LIU, W.; LIANG, X.; FU, Y.; MA, K.; ZHAO, L.; ZHANG, F.; LU, Z.; LEE, S. M.-Y.; XU, X.; WANG, J.; YANG, H.; FU, C.; GE, S.; CHEN, W. Draft genome of the living fossil *Ginkgo biloba*. **GigaScience**, v. 5, n. 1, p. 49, 21 dez. 2016. doi: 10.1186/s13742-016-0154-1.

HAEUSLER, R. A.; ENGELKE, D. R. Spatial organization of transcription by RNA polymerase III. **Nucleic Acids Research**, v. 34, n. 17, p. 4826–4836, 2006. doi: 10.1093/nar/gkl656.

HANNA, W. W. Method of reproduction in napier grass and in the 3X and 6X allopolyploid hybrids with pearl millet. **Crop Science**, Madison, v. 21, p. 123–126, 1981.

HAWKINS, J. S.; KIM, H. R.; NASON, J. D.; WING, R. A.; WENDEL, J. F. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. **Genome Research**, v. 16, n. 10, p. 1252–1261, 2006. doi: 10.1101/gr.5282906.

- HEITKAM, T.; WEBER, B.; WALTER, I.; LIEDTKE, S.; OST, C.; SCHMIDT, T. Satellite DNA landscapes after allotetraploidisation of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. **The Plant Journal**, n. March, p. tpj.14705, 24 jan. 2020. doi: 10.1111/tpj.14705.
- HELENA, T. V.; DAVIDE, L. C.; PEREIRA, A. Vander. Genomic analysis in *Pennisetum purpureum* x *P. glaucum* hybrids. **Caryologia**, v. 58, n. 1, p. 28–33, 18 jan. 2005. doi: 10.1080/00087114.2005.10589428.
- HERTWECK, K. L. Assembly and comparative analysis of transposable elements from low coverage genomic sequence data in Asparagales. **Genome**, v. 56, n. 9, p. 487–494, set. 2013. doi: 10.1139/gen-2013-0042.
- HESLOP-HARRISON, J. (Pat); SCHMIDT, T. Plant Nuclear Genome Composition. **Encyclopedia of Life and Science**, p. 1–8, 2007. doi: 10.1002/9780470015902.a0002014.pub2.
- ISHII, T.; MATSUMOTO, N.; TANAKA, H.; ELTAYEB, A. E.; TSUJIMOTO, H. Evolution of subtelomeric and centromeric repetitive sequences in genus *Pennisetum* (Poaceae). **Chromosome Science**, v. 15, n. 3+4, p. 53–59, 2012. doi: 10.11352/scr.15.53.
- JAUHAR, P. P. Cytogenetics and breeding of pearl millet and related species. New York: A. R. Liss, 289 p., 1981.
- KALENDAR, R.; GROB, T.; REGINA, M.; SUONIEMI, A.; SCHULMAN, A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. **Theoretical and Applied Genetics**, v. 98, n. 5, p. 704–711, abr. 1999. doi: 10.1007/s001220051124.
- KAMM, A.; SCHMIDT, T.; HESLOP-HARRISON, J. S. Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. **MGG Molecular & General Genetics**, v. 244, n. 4, p. 420–425, 1994. doi: 10.1007/BF00286694.
- KAPUSTOVÁ, V.; TULPOVÁ, Z.; TOEGELOVÁ, H.; NOVÁK, P.; MACAS, J.; KARAFIÁTOVÁ, M.; HRIBOVÁ, E.; DOLEŽEL, J.; ŠIMKOVÁ, H. The Dark Matter of Large Cereal Genomes: Long Tandem Repeats. **International Journal of Molecular Sciences**, v. 20, n. 10, p. 2483, 20 maio 2019. doi: 10.3390/ijms20102483.
- KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013. doi: 10.1093/molbev/mst010.
- KEARSE, M.; MOIR, R.; WILSON, A.; STONES-HAVAS, S.; CHEUNG, M.; STURROCK, S.; BUXTON, S.; COOPER, A.; MARKOWITZ, S.; DURAN, C.; THIERER, T.; ASHTON, B.; MEINTJES, P.; DRUMMOND, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012. doi: 10.1093/bioinformatics/bts199.
- KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, v. 115, n. 1, p. 49–63, 2002. doi: 10.1023/A:1016072014259.

KUITTINEN, H.; DE HAAN, A. A.; VOGL, C.; OIKARINEN, S.; LEPPÄLÄ, J.; KOCH, M.; MITCHELL-OLDS, T.; LANGLEY, C. H.; SAVOLAINEN, O. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. **Genetics**, v. 168, n. 3, p. 1575–1584, nov. 2004. doi: 10.1534/genetics.103.022343.

KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C.; TAMURA, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 1 jun. 2018. doi: 10.1093/molbev/msy096.

LEITCH, A. R.; LEITCH, I. J. Genomic Plasticity and the Diversity of Polyploid Plants. **Science**, v. 320, n. 5875, p. 481–483, 25 abr. 2008. doi: 10.1126/science.1153585.

LI, W.; COWLEY, A.; ULUDAG, M.; GUR, T.; MCWILLIAM, H.; SQUIZZATO, S.; PARK, Y. M.; BUSO, N.; LOPEZ, R. The EMBL-EBI bioinformatics web and programmatic tools framework. **Nucleic acids research**, v. 43, n. W1, p. W580-4, 1 jul. 2015. doi: 10.1093/nar/gkv279.

LIU, Q.; LI, X.; ZHOU, X.; LI, M.; ZHANG, F.; SCHWARZACHER, T.; HESLOP-HARRISON, J. S. 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. 226, 30 dez. 2019. doi: 10.1186/s12870-019-1769-z.

MARTEL, E.; PONCET, V.; LAMY, F.; SILJAK-YAKOVLEV, S.; LEJEUNE, B.; SARR, A. Chromosome evolution of *Pennisetum* species (Poaceae): Implications of ITS phylogeny. **Plant Systematics and Evolution**, v. 249, n. 3–4, p. 139–149, 2004. doi: 10.1007/s00606-004-0191-6.

MASCAGNI, F.; VANGELISTI, A.; GIORDANI, T.; CAVALLINI, A.; NATALI, L. A computational comparative study of the repetitive DNA in the genus *Quercus* L. **Tree Genetics & Genomes**, v. 16, n. 1, p. 11, 23 fev. 2020. doi: 10.1007/s11295-019-1401-2.

MEHROTRA, S.; GOYAL, V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. **Genomics, Proteomics and Bioinformatics**, v. 12, n. 4, p. 164–171, 2014. doi: 10.1016/j.gpb.2014.07.003.

MELTERS, D. P.; BRADNAM, K. R.; YOUNG, H. A.; TELIS, N.; MAY, M. R.; RUBY, J. G.; SEBRA, R.; PELUSO, P.; EID, J.; RANK, D.; GARCIA, J. F.; DERISI, J. L.; SMITH, T.; TOBIAS, C.; ROSS-IBARRA, J.; KORF, I.; CHAN, S. W. L. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome Biology**, v. 14, n. 1, p. 1–20, 30 jan. 2013. doi: 10.1186/gb-2013-14-1-r10.

MICHAEL, T. P. Plant genome size variation: bloating and purging DNA. **Briefings in Functional Genomics**, v. 13, n. 4, p. 308–317, 1 jul. 2014. doi: 10.1093/bfgp/elu005.

NAVAJAS-PÉREZ, R.; QUESADA DEL BOSQUE, M. E.; GARRIDO-RAMOS, M. A. Effect of location, organization, and repeat-copy number in satellite-DNA evolution. **Molecular Genetics and Genomics**, v. 282, n. 4, p. 395–406, out. 2009. doi: 10.1007/s00438-009-0472-4.

- NEUMANN, P.; NOVÁK, P.; HOŠTÁKOVÁ, N.; MACAS, J. 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, 3 dez. 2019. doi: 10.1186/s13100-018-0144-1.
- NOVÁK, P.; HŘIBOVÁ, E.; NEUMANN, P.; KOBLÍŽKOVÁ, A.; DOLEŽEL, J.; MACAS, J. Genome-Wide Analysis of Repeat Diversity across the Family Musaceae. **PLoS ONE**, v. 9, n. 6, p. e98918, 16 jun. 2014. doi: 10.1371/journal.pone.0098918.
- NOVÁK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BMC Bioinformatics**, v. 11, n. 1, p. 378, 2010. doi: 10.1186/1471-2105-11-378.
- NOVÁK, P.; NEUMANN, P.; PECH, J.; STEINHAIŠL, J.; MACAS, J. 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. doi: 10.1093/bioinformatics/btt054.
- NOVÁK, P., ÁVILA ROBLEDILLO, L., KOBLÍŽKOVÁ, A., VRBOVÁ, I., NEUMANN, P., & MACAS, J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic acids research**, v. 45, n. 12, p. e111–e111, 2017.
- PAMPONÉT, V. C. C.; SOUZA, M. M.; SILVA, G. S.; MICHELI, F.; DE MELO, C. A. F.; DE OLIVEIRA, S. G.; COSTA, E. A.; CORRÊA, R. X. Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: citogenomic characterization of transposable elements and satellite DNA. **BMC Genomics**, v. 20, n. 1, p. 262, 2 dez. 2019. doi: 10.1186/s12864-019-5576-6.
- PAUDEL, D.; KANNAN, B.; YANG, X.; HARRIS-SHULTZ, K.; THUDI, M.; VARSHNEY, R. K.; ALTPETER, F.; WANG, J. Surveying the genome and constructing a high-density genetic map of napiergrass (*Cenchrus purpureus* Schumach). **Scientific Reports**, v. 8, n. 1, p. 14419, 26 dez. 2018. doi: 10.1038/s41598-018-32674-x.
- PLOHL, M. Those mysterious sequences of satellite DNAs. **Periodicum Biologorum**, v. 112, n. 4, p. 403–410, 2010. doi: 10.14716/ijtech.v6i4.1206.
- PURUGGANAN, M. D.; WESSLER, S. R. Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. **Molecular Ecology**, v. 4, n. 2, p. 265–270, abr. 1995. doi: 10.1111/j.1365-294X.1995.tb00218.x.
- RASKINA, O.; BARBER, J. C.; NEVO, E.; BELYAYEV, A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. **Cytogenetic and Genome Research**, v. 120, n. 3–4, p. 351–357, maio 2008. doi: 10.1159/000121084.
- RENNY-BYFIELD, S.; KOVAŘÍK, A.; CHESTER, M.; NICHOLS, R. A.; MACAS, J.; NOVÁK, P.; LEITCH, A. R. Independent, Rapid and Targeted Loss of Highly Repetitive DNA in Natural and Synthetic Allopolyploids of *Nicotiana tabacum*. **PLoS ONE**, v. 7, n. 5, p. e36963, 14 maio 2012. doi: 10.1371/journal.pone.0036963.

- RENNY-BYFIELD, S.; KOVARIK, A.; KELLY, L. J.; MACAS, J.; NOVAK, P.; CHASE, M. W.; NICHOLS, R. A.; PANCHOLI, M. R.; GRANDBASTIEN, M.-A.; LEITCH, A. R. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. **The Plant Journal**, v. 74, n. 5, p. 829–839, jun. 2013. doi: 10.1111/tpj.12168.
- ROYAL BOTANIC GARDENS KEW. (2020) Seed Information Database (SID). Version 7.1. Available from: <http://data.kew.org/sid/> (October 2020)
- RUIZ-RUANO, F. J.; LÓPEZ-LEÓN, M. D.; CABRERO, J.; CAMACHO, J. P. M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, n. 1, p. 28333, 7 set. 2016. doi: 10.1038/srep28333.
- SAMOLUK, S. S.; ROBLEDO, G.; BERTIOLI, D.; SEJO, J. G. Evolutionary dynamics of an at-rich satellite DNA and its contribution to karyotype differentiation in wild diploid *Arachis* species. **Molecular Genetics and Genomics**, v. 292, n. 2, p. 283–296, 1 abr. 2017. doi: 10.1007/s00438-016-1271-3.
- SHARMA, S.; RAINA, S. N. Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. **Cytogenetic and Genome Research**, v. 109, n. 1–3, p. 15–26, 2005. doi: 10.1159/000082377.
- SHENTON, M.; KOBAYASHI, M.; TERASHIMA, S.; OHYANAGI, H.; COPETTI, D.; HERNÁNDEZ-HERNÁNDEZ, T.; ZHANG, J.; OHMIDO, N.; FUJITA, M.; TOYODA, A.; IKAWA, H.; FUJIYAMA, A.; FURUUMI, H.; MIYABAYASHI, T.; KUBO, T.; KUDRNA, D.; WING, R.; YANO, K.; NONOMURA, K.; SATO, Y.; KURATA, N. Evolution and diversity of the wild rice *Oryza officinalis* complex, across continents genome types, and ploidy levels. **Genome Biology and Evolution**, v. 12, n. 4, p. 413–428, 3 mar. 2020. doi: 10.1093/gbe/evaa037.
- SMIT, AFA, HUBLEY, R & GREEN, P. *RepeatMasker Open-4.0*. 2013-2015 <<http://www.repeatmasker.org>>.
- SOBRINHO, F. D. S.; PEREIRA, A. Vander; JOSÉ, F.; BOTREL, M. A. Avaliação agronômica de híbridos interespecíficos entre capim-elefante e milho. **Pesquisa Agropecuária Brasileira**, v. 40, n. 9, p. 873–880, 2005.
- SOLTIS, D. E.; SOLTIS, P. S.; BENNETT, M. D.; LEITCH, I. J. Evolution of genome size in the angiosperms. **American Journal of Botany**, v. 90, n. 11, p. 1596–1603, nov. 2003. doi: 10.3732/ajb.90.11.1596.
- STEVENS, K. A.; WEGRZYN, J. L.; ZIMIN, A.; PUIU, D.; CREPEAU, M.; CARDENO, C.; PAUL, R.; GONZALEZ-IBEAS, D.; KORIABINE, M.; HOLTZ-MORRIS, A. E.; MARTÍNEZ-GARCÍA, P. J.; SEZEN, U. U.; MARÇAIS, G.; JERMSTAD, K.; MCGUIRE, P. E.; LOOPSTRA, C. A.; DAVIS, J. M.; ECKERT, A.; DE JONG, P.; YORKE, J. A.; SALZBERG, S. L.; NEALE, D. B.; LANGLEY, C. H. Sequence of the *Sugar Pine* Megagenome. 2016. doi: 10.1534. **Genetics**.116.193227.
- TECHIO, V. H.; DAVIDE, L. C.; PEREIRA, A. Vander. Meiosis in elephant grass (*Pennisetum purpureum*), pearl millet (*Pennisetum glaucum*) (Poaceae, Poales) and their

interspecific hybrids. **Genetics and Molecular Biology**, v. 29, n. 2, p. 353–362, 2006. doi: 10.1590/S1415-47572006000200025.

TENAILLON, M. I.; HUFFORD, M. B.; GAUT, B. S.; ROSS-IBARRA, J. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. **Genome Biology and Evolution**, v. 3, n. 1, p. 219–229, 2011. doi: 10.1093/gbe/evr008.

USAI, G.; MASCAGNI, F.; NATALI, L.; GIORDANI, T.; CAVALLINI, A. Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. **Tree Genetics & Genomes**, v. 13, n. 5, p. 96, 19 out. 2017. doi: 10.1007/s11295-017-1181-5.

VARSHNEY, R. K.; SHI, C.; THUDI, M.; MARIAC, C.; WALLACE, J.; QI, P.; ZHANG, H.; ZHAO, Y.; WANG, X.; RATHORE, A.; SRIVASTAVA, R. K.; CHITIKINENI, A.; FAN, G.; BAJAJ, P.; ... BENNETZEN, J. L.; PATERSON, A. H.; OZIAS-AKINS, P.; GRANDO, S.; WANG, J.; MOHAPATRA, T.; WECKWERTH, W.; REIF, J. C.; LIU, X.; VIGOUROUX, Y.; XU, X. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. **Nature Biotechnology**, v. 35, n. 10, p. 969–976, 18 out. 2017. doi: 10.1038/nbt.3943.

WANG, C.; YAN, H.; LI, J.; ZHOU, S. Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum ‘Zise’) and identification of its SSR markers. **Molecular Breeding**, v. 39, n. 194, 2018. doi: 10.1007/s11032-018-0849-3.

WEBER, N.; HALPIN, C.; HANNAH, L. C.; JEZ, J. M.; KOUGH, J.; PARROTT, W. Editor’s Choice: Crop Genome Plasticity and Its Relevance to Food and Feed Safety of Genetically Engineered Breeding Stacks. **Plant Physiology**, v. 160, n. 4, p. 1842–1853, 1 dez. 2012. doi: 10.1104/pp.112.204271.

WENDEL, J. F. Genome evolution in polyploids. **Plant Molecular Biology**, v. 42, n. 1, p. 225–249, 2000. doi: 10.1023/A:1006392424384.

WICKER, T.; KELLER, B. Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. **Genome Research**, v. 17, n. 7, p. 1072–1081, 13 jun. 2007. doi: 10.1101/gr.6214107.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973–982, 2007. doi: 10.1038/nrg2165.

WICKER, T.; SCHULMAN, A. H.; TANSKANEN, J.; SPANNAGL, M.; TWARDZIOK, S.; MASCHER, M.; SPRINGER, N. M.; LI, Q.; WAUGH, R.; LI, C.; ZHANG, G.; STEIN, N.; MAYER, K. F. X.; GUNDLACH, H. The repetitive landscape of the 5100 Mbp barley genome. **Mobile DNA**, v. 8, n. 1, p. 22, 20 dez. 2017. doi: 10.1186/s13100-017-0102-3.

YAN, Q., WU, F., XU, P., SUN, Z., LI, J., GAO, L., ... & YI, X. The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast

growth. **Molecular Ecology Resources**, 2020.

ARTIGO 2

Análise comparativa do genoma cloroplastidial de três espécies relacionadas de *Cenchrus* L: *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris* (Poaceae)

Silvestrini, A. J. A; Vaio, MAGDALENA; Torres, G. A.

RESUMO

Cenchrus L. é um importante gênero da família Poaceae que abriga diversas espécies de elevada importância agrônoma, principalmente forrageiras de destaque na pecuária nacional, como *Cenchrus purpureus* cv. Pioneiro, *Cenchrus americanus* cv. BN2 e *Cenchrus ciliaris*. Diversos programas de melhoramento genético exploram a capacidade de produção e qualidade dessas forrageiras, assim como, devido à complexa relação taxonômica entre as mesmas, elas são de grande interesse para estudos genômico evolutivos. A montagem e análise comparativa de genomas cloroplastidiais é um desses estudos, abrindo caminhos para a filogenômica, assim como gerando marcadores moleculares cultivares-específicos, muito úteis para o manejo de germoplasma. Portanto, o objetivo do presente trabalho é a montagem e análise comparativa do genoma cloroplastidial de *Cenchrus purpureus* cv. Pioneiro, *Cenchrus americanus* cv. BN2 e *Cenchrus ciliaris*. Para tanto, procedeu-se com o sequenciamento genômico de baixa cobertura de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2. Para as análises envolvendo *Cenchrus ciliaris*, o genoma cloroplastidial foi obtido através do banco de dados NCBI. Os genomas foram montados e anotados através do software Geneious Prime 2020.1.1 por referência para com os já depositados em bancos de dados, sendo realizada uma nova anotação de conferência e montagem gráfica através da plataforma CHLOROBOX – GeSeq. As análises comparativas foram realizadas através de alinhamento genômico pelo algoritmo MAUVE e os SNPs e INDELS identificados através de ferramentas implementadas no software Geneious Prime 2020.1.1. O genoma cloroplastidial das três espécies apresenta formato quadripartido com cerca de 105 – 110 genes únicos em cada espécie. As análises comparativas demonstraram estrutura genômica conservada entre as três espécies, com apenas um Bloco Local Colinear (LCB), indicando a ausência de rearranjos estruturais. As demais análises comparativas corroboram com esses resultados. Os dados propostos estão de acordo com as relações filogenéticas entre essas espécies, corroborando com a proximidade já relatada entre espécies anteriormente alocadas em *Pennisetum*, para com *Cenchrus*. Foram identificados ainda 19 SNPs na cultivar Pioneiro de *Cenchrus purpureus* e 6 SNPs na cultivar BN2 de *Cenchrus americanus* que podem auxiliar na identificação e manejo desses genótipos em bancos de germoplasma.

Palavras-chave: *Cenchrus*, cpDNA, Filogenômica, Genômica Comparativa

ABSTRACT

Cenchrus L. is an important genus of the Poaceae family that includes several species of agronomic importance, especially forages of prominence in national livestock, such as *Cenchrus purpureus* cv. Pioneiro, *Cenchrus americanus* cv. BN2, and *Cenchrus ciliaris*. Several Plant Breeding programs explore the production capacity and quality of these forages, as well, due to the complex taxonomic relationship between them, they are of great interest for evolutionary genomic studies. The assembly and comparative analysis of chloroplast genomes is one of these studies, and can be an opening path for phylogenomics, as well as for identification of cultivar-specific molecular markers, very useful for germplasm management. Therefore, the objective of the present study is to perform the assembly and comparative analysis of the chloroplast genome of *Cenchrus purpureus* cv. Pioneiro, *Cenchrus americanus* cv. BN2, and *Cenchrus ciliaris*. For this, we proceeded with the genomic sequencing with low coverage of the genome of *Cenchrus purpureus* cv. Pioneiro and *Cenchrus americanus* cv. BN2. For analysis involving *Cenchrus ciliaris*, its chloroplast genome was found in the NCBI database. The genomes were assembled and annotated using the Geneious Prime 2020.1.1 software by reference to those already published in databases, and an annotation check and graphical assembly was performed through the CHLOROBOX - GeSeq platform. The comparative analysis were carried out through genomic alignment by the MAUVE algorithm and the SNPs and INDELs identified through tools implemented in the Geneious Prime 2020.1.1 software. The chloroplast genome of the three species has a quadripartite structure with about 105 - 110 unique genes in each species. The comparative analyzes showed the conserved genomic structure among the three species, with only one Local Collinear Block (LCB), indicating the absence of rearrangements. The other comparative analyzes corroborate these results. The proposed data available here are in accordance with the phylogenetic relationships between these species, corroborating the proximity already reported between species previously allocated in *Pennisetum*, to *Cenchrus*. 19 SNPs were also identified in the cultivar Pioneiro of *Cenchrus purpureus* and 6 SNPs in the cultivar BN2 of *Cenchrus americanus* that can assist in the identification and management of these genotypes in germplasm banks.

Keywords: *Cenchrus*; cpDNA; Filogenomics; Comparative Genomics.

1. INTRODUÇÃO

Cenchrus L. é um amplo gênero pertencente à família Poaceae, subfamília Panicoideae, tribo Paniceae que possui diversas espécies distribuídas ao longo de toda a faixa tropical e subtropical do planeta (KEW GARDENS SCIENCE, 2020). A mais recente circunscrição taxonômica do gênero resultou da unificação dos gêneros *Cenchrus*, *Pennisetum* Rich. e *Odontelytrum* Hack., com a inclusão dos três em *Cenchrus* (DONADÍO et al., 2009; CHEMISQUY et al., 2010).

O gênero possui diversas espécies de importância agrônômica, a exemplo de *Cenchrus purpureus* (Schumach.) (syn. *Pennisetum purpureum*) ou Capim-elefante, espécie forrageira, *Cenchrus americanus* [(L.) R. Brown, 1810] (syn. *Pennisetum glaucum*) ou Milheto, produtora de grãos utilizados na alimentação humana e uso como forrageira e *Cenchrus ciliaris* L. (syn. *Pennisetum ciliare*) ou Grama *Buffel*, outra espécie largamente utilizada na alimentação animal (HANNA, 1981; SOBRINHO et al., 2005).

O melhoramento genético é a principal ferramenta voltada à produção de cultivares melhoradas e adaptadas às demandas cada vez mais exigentes do crescente e expansivo mercado da pecuária (SOUZA et al., 2016). *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris* são espécies há muito exploradas geneticamente, contando com coleção de genótipos depositados em bancos de germoplasma, como o Banco de Germoplasma da EMBRAPA Gado de Leite. Para o progresso de tais programas, é imprescindível a caracterização do germoplasma disponível.

As relações taxonômicas entre *Pennisetum* Rich. e *Cenchrus* nem sempre foram bem esclarecidas, havendo espécies alocadas e removidas entre os dois gêneros ao longo de diversos estudos filogenéticos (ZULOAGA; MORRONE; GIUSSANI, 2000; DONADÍO et al., 2009; CHEMISQUY et al., 2010). Uma espécie que ilustra de maneira clara essas confusões taxonômicas é *Cenchrus ciliaris*, que sempre esteve à margem dos dois gêneros. Diversos autores já propuseram sua classificação dentro de *Pennisetum* (WIERSEMA, 2019) enquanto outros propuseram em *Cenchrus* (DONADÍO et al., 2009). Apesar de possuírem relações taxonômicas recentemente revisadas e propostas, espécies de *Cenchrus*, assim como as anteriormente alocadas em *Pennisetum*, carecem de estudos filogenômicos.

O cloroplasto é a organela vegetal responsável pela manutenção da fotossíntese e que possui um genoma próprio. O genoma cloroplastidial possui três distintas regiões que reúnem desde genes relacionados ao processo fotossintético, até regiões que portam sequências de DNA repetidas, a *Large Single Copy Region* (LSC) (Do inglês Grande região de cópia única), *Small*

Single Copy Region (SSC) (Do inglês pequena região de cópia única) e as *Inverted Repeats Region* (IRa e IRb) (Do inglês regiões repetidas invertidas). Entre essas regiões e sequências, algumas são conservadas devido às baixas taxas de inversão e substituição de nucleotídeos, sendo úteis em estudos filogenômicos, assim como na obtenção de marcadores moleculares extremamente úteis no manejo de germoplasma (EISEN, 2003; BOCK, 2007). A montagem do genoma cloroplastidial em espécies vegetais é uma interessante estratégia para a realização de tais estudos.

O sequenciamento e a montagem individual do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris* já foram realizados (BHATT; THAKER, 2018; WU; ZHOU, 2019; XU et al., 2019). No entanto, as análises não são detalhadas e não existe relato de análises genômico-comparativas envolvendo o genoma cloroplastidial dessas três espécies. Portanto, uma análise comparativa detalhada do genoma cloroplastidial dessas três espécies é de grande valia para a filogenômica de *Cenchrus*, assim como para a geração de marcadores específicos para as cultivares utilizadas no presente estudo e que são largamente utilizadas em programas de melhoramento genético.

2. MATERIAL E MÉTODOS

Material Genético

Foram utilizadas a variedade Pioneiro da espécie *Cenchrus purpureus* e a variedade BN2 da espécie *Cenchrus americanus*. O genoma cloroplastidial da espécie *Cenchrus ciliaris* (BHATT; THAKER, 2018) foi obtido através do banco de dados NCBI sob o depósito MH286942 (Tabela 1).

Tabela 1 Dados de passaporte *Cenchrus* spp.

Material Vegetal	Origem/Código de acesso NCBI
<i>Cenchrus purpureus</i>	Banco Ativo de Germoplasma – EMBRAPA Gado de Leite – Juiz de Fora – MG
<i>Cenchrus americanus</i>	Banco Ativo de Germoplasma – EMBRAPA Gado de Leite – Juiz de Fora – MG
<i>Cenchrus ciliaris</i>	MH286942

Extração de DNA e Sequenciamento Genômico

Folhas jovens de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2 foram selecionadas e utilizadas para extração de DNA genômico através do kit de extração *DNeasy® Plant Mini Kit* (Qiagen® Inc., Venlo, Netherlands). A avaliação quanto à integridade e qualidade do DNA extraído foi realizada por meio de eletroforese em gel de agarose 1% e em espectrofotômetro *NanoDrop™* (Thermo Scientific™). A montagem da biblioteca foi realizada utilizando o kit *Nextera™ DNA Flex Library Prep* com fragmentos *Pair-ended* de 2x300pb para ambas as espécies. O sequenciamento genômico foi realizado em plataforma Illumina *HiSeq™* 4000.

Montagem por referência e anotação do genoma cloroplastidial

As sequências obtidas, 246 Mb com 1 223 394 *reads* para *Cenchrus purpureus* e 286 Mb com 1 448 458 *reads* para *Cenchrus americanus*, foram pré-processadas através da ferramenta *BBduk*, com configurações padrão de filtragem e corte. As etapas de mapeamento e montagem dos genomas cloroplastidiais foram realizadas utilizando um genoma de referência para cada espécie. Para *Cenchrus purpureus* cv. Pioneiro foi utilizado como referência o genoma cloroplastidial de *Cenchrus purpureus* (WU; ZHOU, 2019) previamente depositado no banco de dados NCBI sob o número de acesso MF594682. Para a espécie *Cenchrus americanus* cv. BN2 foi utilizado o genoma de *Cenchrus americanus* acesso IT144339 (RAVEENDAR et al., 2019) depositado no banco de dados NCBI sob o número KX756179.

O mapeamento, a montagem e a anotação inicial foram realizadas através da ferramenta *Geneious Assembler* presente no software *Geneious Prime 2020.1.1*. (KEARSE et al., 2012). Posteriormente procedeu-se com a checagem da anotação desses genomas utilizando da ferramenta presente em *CHLOROBOX – GeSeq* (TILLICH et al., 2017). Após a montagem do genoma das três espécies, sucedeu-se com o desenho da estrutura circular desses genomas através da ferramenta online *OrganelleGenomeDRAW - OGDRAW* (LOHSE; DRECHSEL; BOCK, 2007).

Análises genômico-comparativas

A primeira análise realizada foi a busca por Polimorfismos de Nucleotídeo Único (SNPs) nas cultivares Pioneiro de *Cenchrus purpureus* e BN2 de *Cenchrus americanus* tendo como referência os genomas cloroplastidiais depositados no NCBI para cada espécie individualmente. Para tanto, foram identificadas regiões de baixa cobertura nos dados oriundos do sequenciamento das duas cultivares afim de que essas regiões pudessem ser mascaradas e não afetassem a busca pelos SNPs. As análises foram realizadas utilizando de algoritmo próprio implementado no software Geneious Prime 2020.1.1, com configurações padrão.

Foi realizado ainda o alinhamento dos três genomas com o objetivo de identificar deleções, substituições e translocações em blocos conservados. A partir do alinhamento foi analisada também a composição das regiões limítrofes entre as três grandes regiões dos genomas. Para o alinhamento genômico foi utilizado o algoritmo MAUVE (DARLING et al., 2004) implementado no software Geneious Prime 2020.1.1.

3. RESULTADOS E DISCUSSÃO

Composição e estrutura do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*.

A montagem do genoma cloroplastidial de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2 foi realizada tendo como referência genomas já publicados de ambas as espécies (RAVEENDAR et al., 2019; WU; ZHOU, 2019). A partir do mapeamento dos genomas já publicados contra os dados oriundos do sequenciamento genômico de baixa cobertura gerados no presente trabalho, foram mapeados 16.670 *reads* em *Cenchrus purpureus* e 97.903 *reads* em *Cenchrus americanus*. Ao relacionar a quantidade de *reads* mapeados e o tamanho já conhecido do genoma cloroplastidial de cada espécie, foram obtidas coberturas genômicas de 36x para o cpDNA de *Cenchrus purpureus* e 212x para o cpDNA de *Cenchrus americanus* (Tabela 1).

O genoma cloroplastidial de *Cenchrus purpureus* cv. Pioneiro, *Cenchrus americanus* cv. BN2 e *Cenchrus ciliaris* apresentam, respectivamente, tamanhos de 138.223 pb, 138.171 pb e 138.737 pb. Tais valores estão de acordo com o já proposto para as três espécies, assim como para outras espécies de *Cenchrus* (BHATT; THAKER, 2018; HYUN et al., 2019;

RAVEENDAR et al., 2019; WU; ZHOU, 2019; WANG, 2020). O genoma de todas as espécies analisadas apresenta estrutura quadripartida como relatado para a grande maioria das angiospermas (SASKI et al., 2007a) (Figuras 1, 2, 3). Essa estrutura está dividida entre as regiões LSC (Grande Região de Cópia Única), com comprimentos de 81.158 pb (*Cenchrus purpureus*), 81.086 pb (*Cenchrus americanus*) e 81.053 pb (*Cenchrus ciliaris*); outra região, SSC (Pequena Região de Cópia Única) com 12.384 pb (*C. purpureus*), 12.409 pb (*C. americanus*) e 13.045 pb (*C. ciliaris*) e duas regiões IRs, ou Regiões Repetidas Invertidas que separam as regiões LSC e SSC, de cerca de 22.340 pb (*C. purpureus*), 22338 pb (*C. americanus*) e 22307 pb (*C. ciliaris*) (Tabela 1). Esses comprimentos estão de acordo com o proposto para espécies de Gramíneas (SASKI et al., 2007a), assim como para espécies de *Cenchrus* (BHATT; THAKER, 2018; HYUN et al., 2019; RAVEENDAR et al., 2019; WU; ZHOU, 2019; WANG, 2020).

A porcentagem de bases nitrogenadas G/C no genoma cloroplastidial das três espécies foi de 38,6% para *Cenchrus purpureus*, 38,5% para *Cenchrus americanus* e 38,6% para *Cenchrus ciliaris* e estão de acordo valores descritos para espécies do gênero (HYUN et al., 2019; RAVEENDAR et al., 2019; WU; ZHOU, 2019; WANG, 2020).

A proporção correspondente a cada uma das regiões em relação a todo o genoma cloroplastidial foi de aproximadamente 58% para LSC, 32% para as regiões IR e 9% para a região SSC nas três espécies (Tabela 1). Essas proporções são inéditas em trabalhos de caracterização do genoma cloroplastidial para *Cenchrus purpureus* e *Cenchrus americanus* e está de acordo com o descrito para espécies inclusas em Poaceae como *Cynodon dactylon* (L.) Pers., *Pharus latifolius* L., entre outras (HUANG et al., 2017).

Tabela 2 Propriedades gerais do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*.

	<i>Cenchrus purpureus</i>	<i>Cenchrus americanus</i>	<i>Cenchrus ciliaris</i>
Acesso GenBank	-	-	MH286942
<i>Reads</i> Mapeados	16.670	97.903	-
Cobertura (x)	36,19	212,57	-
Plastoma (pb)	138.223	138.171	138.737
GC (%)	38,6	38,5	38,6
Região LSC (pb)	81.158	81.086	81.053
Proporção (%)	58,72	58,69	58,42
Região SSC (pb)	12.384	12.409	13.045
Proporção (%)	8,96	8,98	9,40
Regiões IR (pb)	44.681	44.676	44.639
Proporção (%)	32.32530042	32.33384719	32.17526687

Legenda: GC: Porcentagem de bases nitrogenadas G/C. LSC: *Large Single Copy Region*. SSC: *Small Single Copy Region*. IR: *Inverted Regions*.

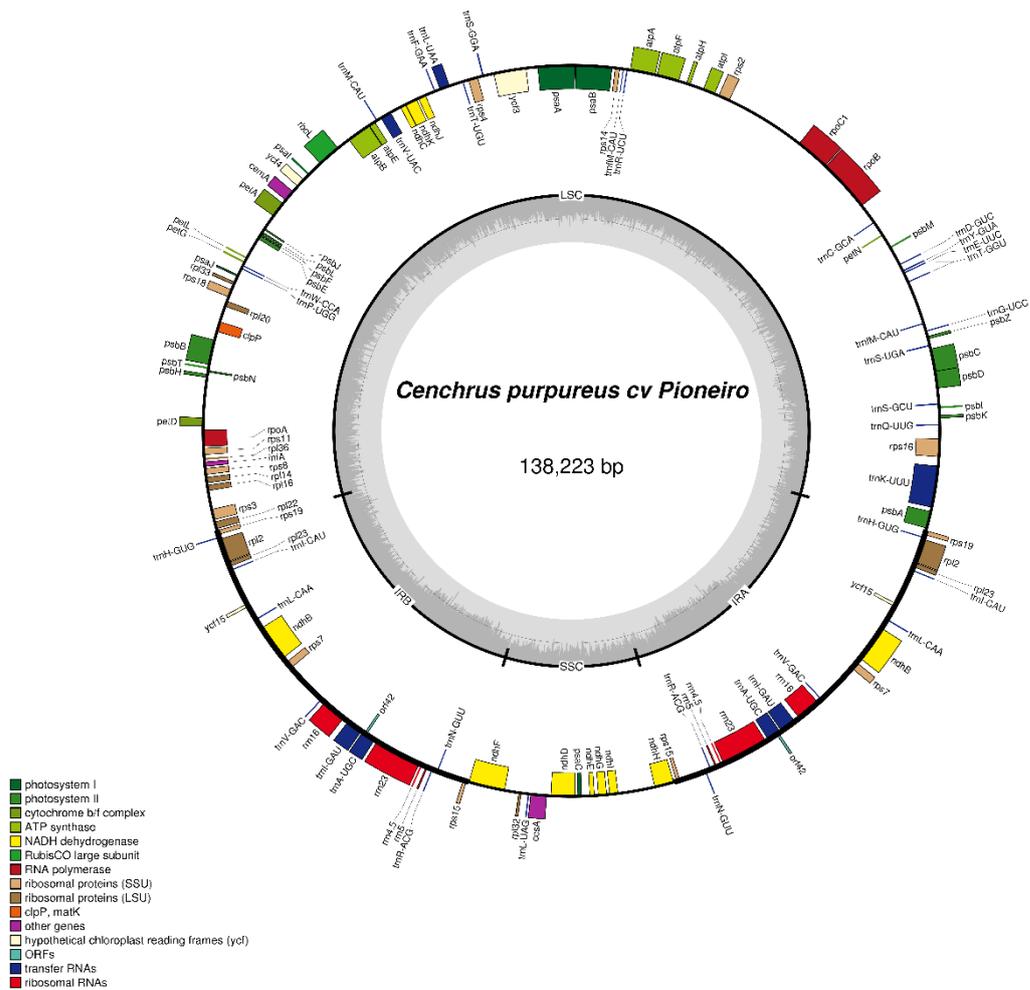


Figura 1 Representação gráfica circular do genoma cloroplastial de *Cenchrus purpureus* cv. Pioneiro. Genes internos ao círculo são transcritos em sentido horário e genes externos ao círculo em sentido anti-horário. Genes de diferentes cores pertencem a diferentes grupos funcionais. A região cinza escuro dentro do círculo corresponde à proporção de GC, enquanto a região cinza clara corresponde a proporção de AT.

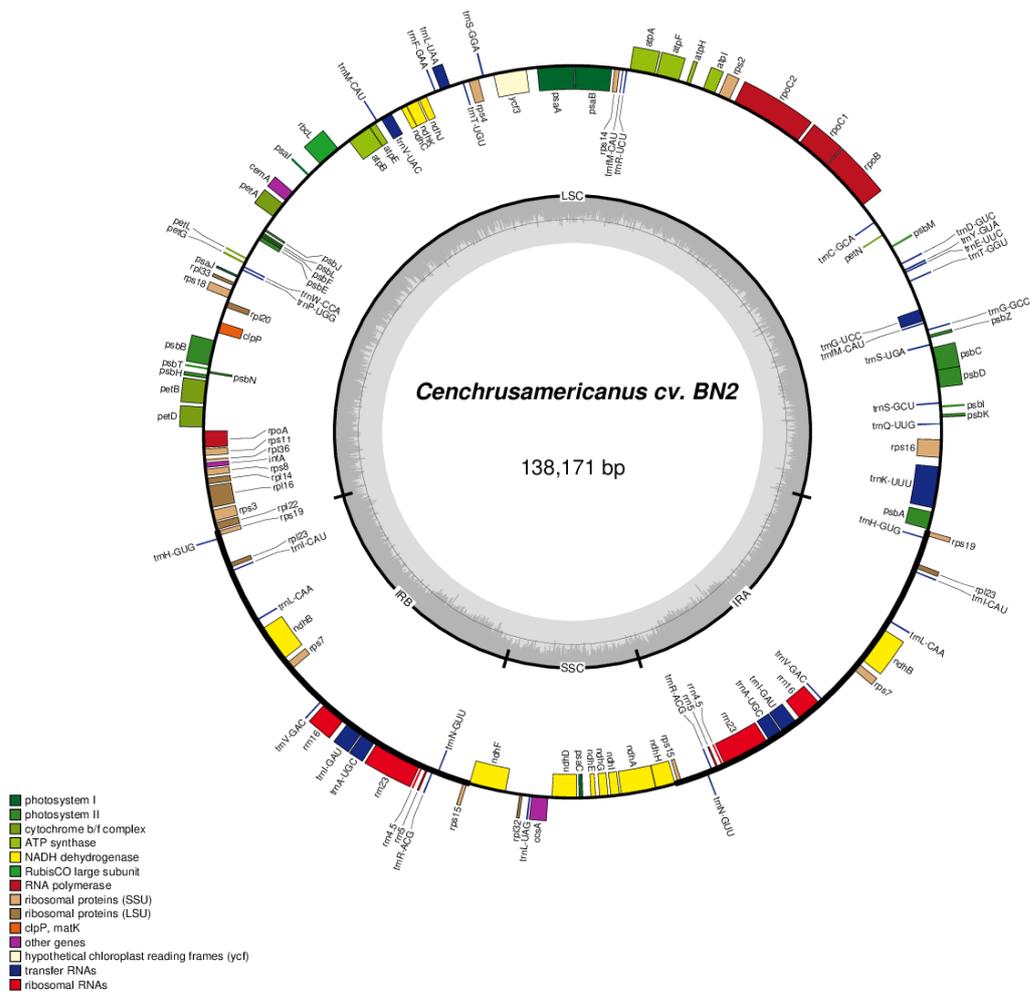


Figura 2 Representação gráfica circular do genoma cloroplastial de *Cenchrus americanus* cv. BN2. Genes internos ao círculo são transcritos em sentido horário e genes externos ao círculo em sentido anti-horário. Genes de diferentes cores pertencem a diferentes grupos funcionais. A região cinza escuro dentro do círculo corresponde à proporção de GC, enquanto a região cinza clara corresponde a proporção de AT.

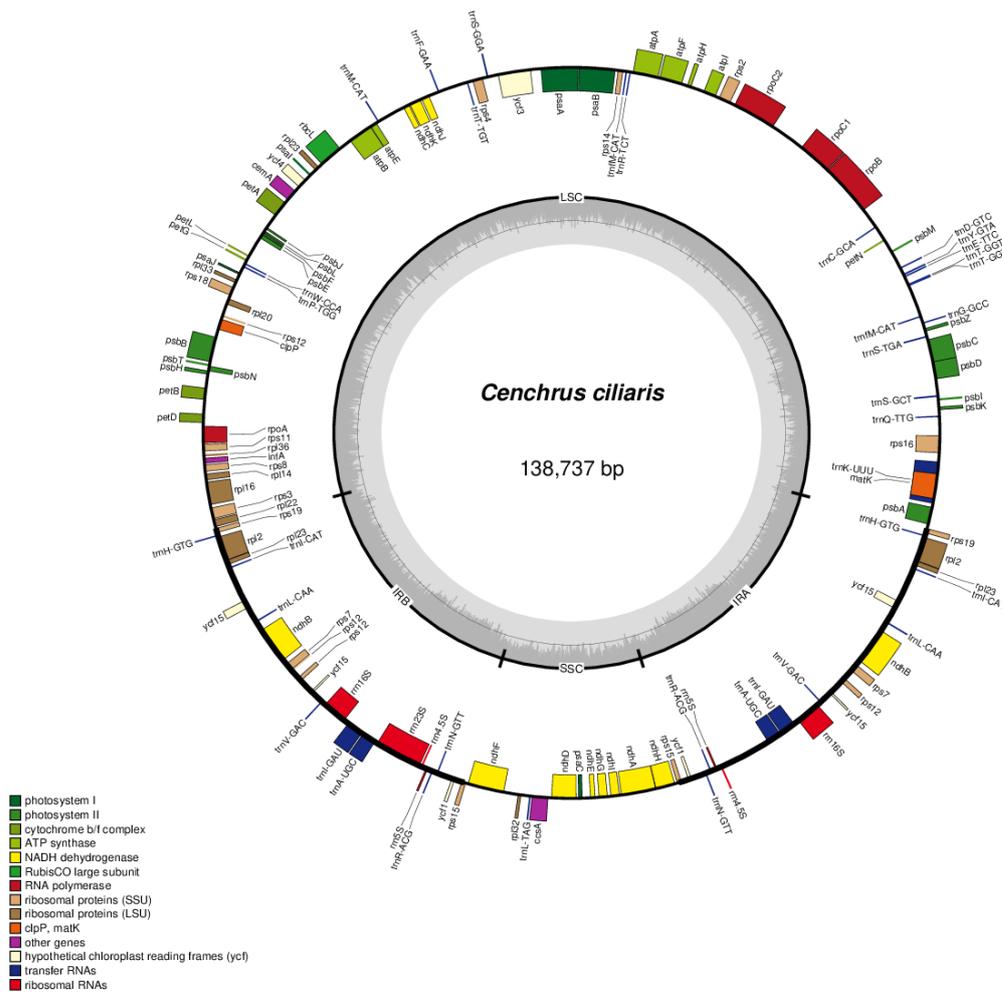


Figura 3 Representação gráfica circular do genoma cloroplastidial de *Cenchrus ciliaris*. Genes internos ao círculo são transcritos em sentido horário e genes externos ao círculo em sentido anti-horário. Genes de diferentes cores pertencem a diferentes grupos funcionais. A região cinza escuro dentro do círculo corresponde à proporção de GC, enquanto a região cinza clara corresponde a proporção de AT.

Analisando a composição desses genomas, foram identificados 105 genes, para *Cenchrus purpureus*, sendo 18 desses duplicados nas IRs (Total de 123 genes), 110 genes para *Cenchrus americanus*, sendo 19 duplicados nas IRs (Total de 129) e 108 genes para *Cenchrus ciliaris* com 23 duplicados nas IRs (Total de 131) (Tabelas , 2, 3 e 4). Essa composição está de

acordo com o relatado para espécies de Gramíneas (SASKI et al., 2007b), mas difere do já proposto para *Cenchrus purpureus* (136 genes) (WU; ZHOU, 2019) e *Cenchrus americanus* (110 genes) quando consideramos o total de genes (RAVEENDAR et al., 2019). Essas diferenças podem ter ocorrido devido à utilização de distintas abordagens entre os trabalhos. Raveendar e colaboradores (2019) utilizaram do genoma de *Avena sativa* L. como referência para montagem do cpDNA de *Cenchrus americanus*, assim como utilizaram apenas da ferramenta implementada no pacote CHLOROBX – GeSeq (TILLICH et al., 2017). O mesmo se aplica ao proposto para *Cenchrus purpureus* por Wu e colaboradores (2019).

Os genomas cloroplastidiais de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris* possuem elevada similaridade em relação à composição e estrutura como está evidenciado na Tabela 1, nas representações gráficas desses genomas (Figuras 1, 2 e 3) e no conteúdo gênico dos mesmos (Tabelas 2, 3 e 4). Esses resultados estão de acordo com os relatos sobre o quão genomas cloroplastidiais podem ser conservados estruturalmente em Gramíneas (SASKI et al., 2007b), e em espécies de *Cenchrus* (HYUN et al., 2019).

Essa similaridade pode ser explicada pela estreita relação filogenética entre essas espécies. Donadío e colaboradores (2009) propõe que espécies de *Cenchrus*, como *Cenchrus ciliaris*, e espécies anteriormente alocadas em *Pennisetum*, como *Cenchrus purpureus* e *Cenchrus americanus* formam um clado monofilético, no qual compartilham um mesmo ancestral comum, e, portanto, são filogeneticamente relacionadas. Chemisquy e colaboradores (2010) também demonstraram a proximidade entre essas taxas, propondo a unificação de *Pennisetum* e *Cenchrus*.

Apesar de a grande maioria dos genes ser compartilhada entre as espécies (Tabelas 2, 3 e 4), corroborando com as hipóteses de elevada conservabilidade entre esses genomas (SASKI et al., 2007b; HYUN et al., 2019), existe ainda certa diferença no número de genes entre *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*. Essa diferença pode ser explicada pela perda de alguns desses genes durante a diferenciação e evolução dessas espécies, como já relatado extensivamente na literatura para os mais diversos gêneros vegetais (MAIER et al., 1995; LI et al., 2013; WANG; CHEN; ZHANG, 2016), inclusive para *Cenchrus americanus*, no qual Raveendar e colaboradores (2019) relatam a perda dos genes *accD*, *ycf1* e *ycf2*, corroborando com o observado na Tabela 3.

Tabela 2 Conteúdo gênico presente no genoma cloroplastidial de *Cenchrus purpureus* cv. Pioneiro.

Categorias	Grupo Gênico	Genes						
Genes da Fotosíntese	Subunidade ribossômica menor	<i>rps2</i>	<i>rps3</i>	<i>rps4</i>	<i>rps7*</i>	<i>rps8</i>	<i>rps11</i>	<i>rps15*</i>
		<i>rps14</i>	<i>rps16</i>	<i>rps18</i>	<i>rps19*</i>			
	Subunidade ribossômica maior	<i>rpl2*</i>	<i>rpl14</i>	<i>rpl16</i>	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23*</i>	<i>rpl32</i>
		<i>rpl33</i>	<i>rpl36</i>					
	RNA Polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i>				
	Genes para RNA Ribossomal	<i>rrn4.5*</i>	<i>rrn5*</i>	<i>rrn16*</i>	<i>rrn23*</i>			
	Genes para RNA de Transferência	<i>trnA-</i>	<i>trnC-</i>	<i>trnD-</i>	<i>trnE-</i>	<i>trnF-</i>	<i>trnG-</i>	<i>trnH-</i>
		<i>UGC*</i>	<i>GCA</i>	<i>GUC</i>	<i>UUC</i>	<i>GAA</i>	<i>UCC</i>	<i>GUG*</i>
		<i>trnI-</i>	<i>trnI-</i>	<i>trnL-</i>	<i>trnL-</i>	<i>trnL-</i>	<i>trnM-</i>	<i>trnM-</i>
		<i>CAU*</i>	<i>GAU*</i>	<i>UAA</i>	<i>CAA*</i>	<i>UAG</i>	<i>CAU</i>	<i>CAU*</i>
		<i>trnN-</i>	<i>trnP-</i>	<i>trnQ-</i>	<i>trnR-</i>	<i>trnR-</i>	<i>trnS-</i>	<i>trnS-</i>
	<i>GUU</i>	<i>UGG</i>	<i>UUG</i>	<i>ACG*</i>	<i>UCU</i>	<i>GCU</i>	<i>UGA</i>	
	<i>trnS-</i>	<i>trnT-</i>	<i>trnT-</i>	<i>trnV-</i>	<i>trnV-</i>	<i>trnW-</i>	<i>trnY-</i>	
	<i>GGA</i>	<i>UGU</i>	<i>GGU</i>	<i>GAC*</i>	<i>UAC</i>	<i>CCA</i>	<i>GUA</i>	
	Fator de Iniciação da Tradução	<i>infA</i>						
	Subunidades do Fotossistema I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>		
	Subunidades do Fotossistema II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>
		<i>psbI</i>	<i>psbJ</i>	<i>psbK</i>	<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>
		<i>psbZ</i>						
Subunidades do Citocromo	<i>petA</i>	<i>petD</i>	<i>petG</i>	<i>petL</i>	<i>petN</i>			
Subunidades ATP-Synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i>	<i>atpH</i>	<i>atpI</i>		
Subunidade maior da RUBisco	<i>rbcL</i>							
Subunidades do NADH	<i>ndhB*</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	
	<i>ndhI</i>	<i>ndhJ</i>	<i>ndhK</i>					
Outros Genes	Proteína de Membrana do Envelope	<i>cemA</i>						
	Gene de síntese do Citocromo Tipo-C	<i>ccsA</i>						
	Protease	<i>clpP</i>						
	Sem função definida	<i>orf42*</i>	<i>ycf3</i>	<i>ycf4</i>	<i>ycf15</i>			

Tabela 3 Conteúdo gênico presente no genoma cloroplastidial de *Cenchrus americanus* cv. BN2.

Categorias	Grupo Gênico	Genes						
Genes da Fotosíntese	Subunidade ribossômica menor	<i>rps2</i>	<i>rps3</i>	<i>rps4</i>	<i>rps7*</i>	<i>rps8</i>	<i>rps11</i>	<i>rps14</i>
		<i>rps15*</i>	<i>rps16</i>	<i>rps18</i>	<i>rps19*</i>			
	Subunidade ribossômica maior	<i>rpl14</i>	<i>rpl16</i>	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23*</i>	<i>rpl32</i>	<i>rpl33</i>
		<i>rpl36</i>						
	RNA Polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i>	<i>rpoC2</i>			
	Genes para RNA Ribossomal	<i>rrn4.5*</i>	<i>rrn5*</i>	<i>rrn16*</i>	<i>rrn23*</i>			
	Genes para RNA de Transferência	<i>trnA-UGC*</i>	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>trnE-UUC</i>	<i>trnF-GAA</i>	<i>trnG-GCC</i>	<i>trnG-UCC</i>
		<i>trnH-GUG*</i>	<i>trnI-CAU*</i>	<i>trnI-GAU</i>	<i>trnK-UUU</i>	<i>trnL-UAA</i>	<i>trnL-CAA*</i>	<i>trnL-UAG</i>
		<i>trnM-CAU*</i>	<i>trnM-CAU*</i>	<i>trnN-GUU*</i>	<i>trnP-UGG</i>	<i>trnQ-UUG</i>	<i>trnR-ACG*</i>	<i>trnR-UCU</i>
		<i>trnS-UGA</i>	<i>trnS-GCU</i>	<i>trnS-GGA</i>	<i>trnT-GGU</i>	<i>trnT-UGU</i>	<i>trnV-GAC</i>	<i>trnV-UAC</i>
	<i>trnW-CCA</i>	<i>trnY-GUA</i>						
Fator de Iniciação da Tradução	<i>infA</i>							
Subunidades do Fotossistema I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>			
Subunidades do Fotossistema II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	
	<i>psbI</i>	<i>psbJ</i>	<i>psbK</i>	<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	
	<i>psbZ</i>							
Subunidades do Citocromo	<i>petA</i>	<i>petD</i>	<i>petG</i>	<i>petL</i>	<i>petN</i>	<i>petB</i>		
Subunidades ATP-Synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i>	<i>atpH</i>	<i>atpI</i>		
Subunidade menor da RUBisco	<i>rbcL</i>							
Subunidades do NADH	<i>ndhA</i>	<i>ndhB*</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	
	<i>ndhH</i>	<i>ndhI</i>	<i>ndhJ</i>	<i>ndhK</i>				
Outros Genes	Proteína de Membrana do Envelope	<i>cemA</i>						
	Gene de síntese do Citocromo Tipo-C	<i>ccsA</i>						
	Protease	<i>clpP</i>						
	Sem função definida	<i>ycf3</i>	<i>ycf1</i>	<i>ycf1</i>	<i>ycf1</i>	<i>ycf1</i>	<i>80*</i>	
		<i>79a*</i>	<i>79b*</i>	<i>80*</i>				

Legenda: Genes duplicados*.

Tabela 4 Conteúdo gênico presente no genoma cloroplastidial de *Cenchrus ciliaris*.

Categorias	Grupo Gênico	Genes						
Genes da Fotosíntese	Subunidade ribossômica menor	<i>rps2</i>	<i>rps3</i>	<i>rps4</i>	<i>rps7*</i>	<i>rps8</i>	<i>rps11</i>	<i>rps12*</i>
		<i>rps15*</i>	<i>rps14</i>	<i>rps16</i>	<i>rps18</i>	<i>rps19*</i>		
	Subunidade ribossômica maior	<i>rpl2*</i>	<i>rpl14</i>	<i>rpl16</i>	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23*</i>	<i>rpl32</i>
		<i>rpl33</i>	<i>rpl36</i>					
	RNA Polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i>	<i>rpoC2</i>			
	Genes para RNA Ribossomal	<i>rrn4.5*</i>	<i>rrn5*</i>	<i>rrn16*</i>	<i>rrn23</i>			
	Genes para RNA de Transferência	<i>trnA-UGC*</i>	<i>trnC-GCA</i>	<i>trnD-GTC</i>	<i>trnE-TTC</i>	<i>trnF-GAA</i>	<i>trnG-GCC</i>	<i>trnH-GTG*</i>
		<i>trnI-CAT*</i>	<i>trnI-GAU*</i>	<i>trnK-UUU</i>	<i>trnL-CAA*</i>	<i>trnL-TAG</i>	<i>trnfM-CAT*</i>	<i>trnM-CAT</i>
		<i>trnN-GTT*</i>	<i>trnP-TGG</i>	<i>trnQ-TTG</i>	<i>trnR-ACG*</i>	<i>trnR-TCT</i>	<i>trnS-TGA</i>	<i>trnS-GGA</i>
		<i>trnT-GGT*</i>	<i>trnT-TGT</i>	<i>trnV-GAC*</i>	<i>trnW-CCA</i>	<i>trnY-GTA</i>		
Fator de Iniciação da Tradução	<i>infA</i>							
Subunidades do Fotossistema I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>			
Subunidades do Fotossistema II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	
	<i>psbI</i>	<i>psbJ</i>	<i>psbK</i>	<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	
	<i>psbZ</i>							
Subunidades do Citocromo	<i>petA</i>	<i>petB</i>	<i>petD</i>	<i>petG</i>	<i>petL</i>	<i>petN</i>		
Subunidades ATP-Synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i>	<i>atpH</i>	<i>atpI</i>		
Subunidade maior da RUBisco	<i>rbcL</i>							
Subunidades do NADH	<i>nadhA</i>	<i>ndhB*</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	
	<i>ndhH</i>	<i>ndhI</i>	<i>ndhJ</i>	<i>ndhK</i>				
Outros Genes	Proteína de Membrana do Envelope	<i>cemA</i>						
	Gene de síntese do Citocromo Tipo-C	<i>ccsA</i>						
	Protease	<i>clpP</i>						
	Maturase	<i>matK</i>						
	Sem função definida	<i>ycf1*</i>	<i>ycf3</i>	<i>ycf4*</i>	<i>ycf15*</i>			

Legenda: Genes duplicados*.

Análise comparativa do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*

A partir do alinhamento genômico foi possível comprovar visualmente a elevada similaridade entre o genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris* relatada pelas análises individuais de estrutura e composição dos mesmos (Figura 4). A linha rosa ao centro, que conecta os três genomas alinhados na Figura 4, indica um único LCB (Bloco Colinear Local), ou seja, entre os três genomas existe um único bloco colinear, sem inversões, translocações, ou nenhum outro tipo de rearranjo estrutural. Esse resultado também indica que a ordem estrutural no qual os genes estão arranados nesses genomas é conservada entre as três espécies. Esses dados corroboram mais uma vez com relatos da elevada similaridade entre os genomas cloroplastidiais de Gramíneas (SASKI et al., 2007a) e de espécies de *Cenchrus* (RAVEENDAR et al., 2019).

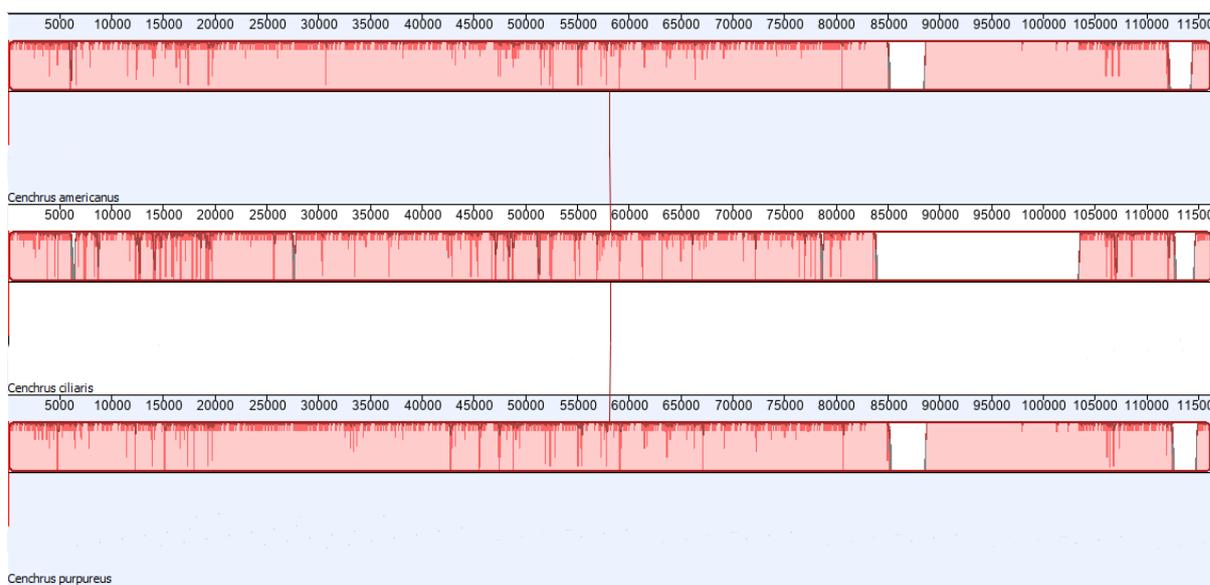


Figura 4 Alinhamento genômico entre os genomas cloroplastidiais de *Cenchrus americanus*, *Cenchrus ciliaris* e *Cenchrus purpureus*. Gráficos em rosa indicam similaridade entre essas regiões genômicas. Regiões descoloridas (em branco), são regiões não alinhadas entre os genomas.

Os gráficos que preenchem a região rosa dos boxes indicam o nível de similaridade entre as regiões genômicas homólogas do genoma das três espécies, e por sua vez, as regiões descoloridas ou brancas, indicam as dissimilaridades entre esses genomas. Tais regiões descoloridas não foram alinhadas, e conseqüentemente não são homólogas entre o genoma cloroplastidial das três espécies, podendo ser consideradas como regiões genômicas exclusivas

de cada espécie, ou ainda regiões que abrigam sequências repetitivas nesses genomas (DARLING et al., 2004). É possível notar que entre *Cenchrus americanus* e *Cenchrus purpureus* existe maior semelhança estrutural entre suas regiões homólogas e não homólogas, o que está de acordo com a maior proximidade filogenética entre as mesmas quando comparadas a *Cenchrus ciliaris*, que possui a estrutura de seu genoma menos conservada entre as três espécies, com a região não homóloga possuindo maior extensão (CHEMISQUY et al., 2010). Resultados similares já foram relatados para diversos grupos de plantas, como espécies da família Magnoliaceae (LI et al., 2019), espécies de algas vermelhas (COSTA et al., 2016) e espécies de *Pinus* L. (PARKS; CRONN; LISTON, 2009).

Foram analisadas comparativamente as junções entre as quatro regiões LSC/IRb/SSC/IRa do genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*, que estão dispostas na Figura 5. De maneira geral, essas regiões apresentam os genes e o arranjo dos mesmos conservados entre as três espécies, como já relatado para grande parte das Angiospermas (SASKI et al., 2007b; LI et al., 2019). Assim como proposto através do alinhamento genômico presente na Figura 4, as maiores similaridades são encontradas entre as junções dos genomas de *Cenchrus purpureus* e *Cenchrus americanus*, que por sinal são idênticas, quando comparados a *Cenchrus ciliaris*, corroborando mais uma vez com a proposição de maior proximidade filogenética entre as duas primeiras espécies, do que entre as mesmas para com *Cenchrus ciliaris*.

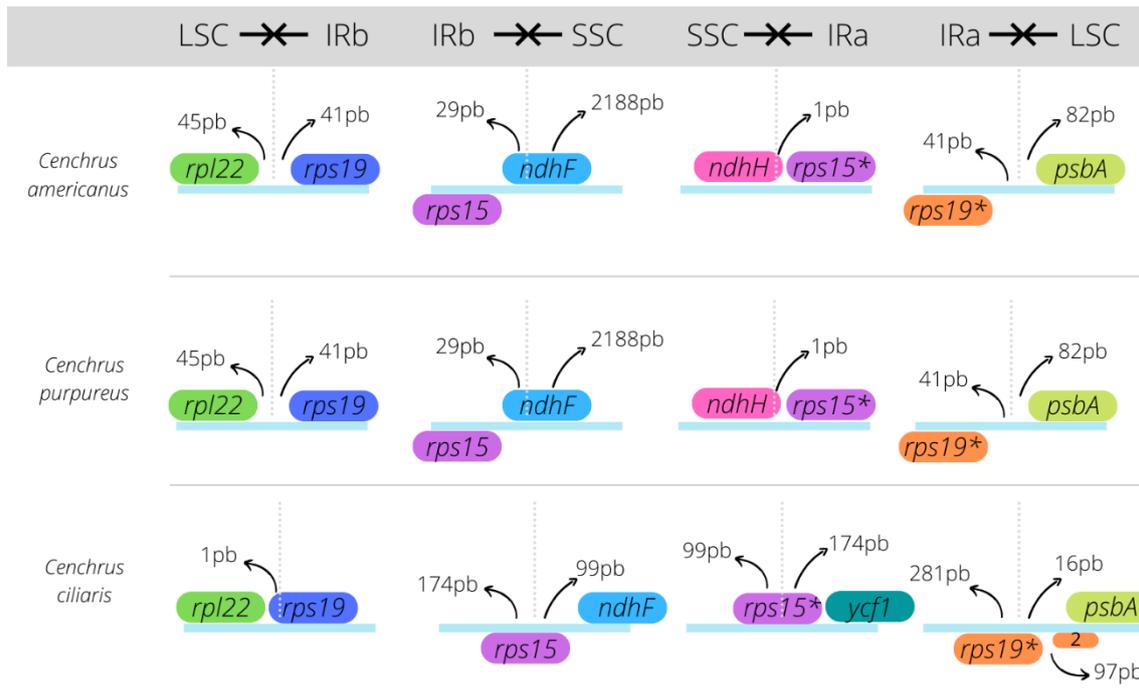


Figura 5 Composição das regiões de junção entre a estrutura quadripartida do genoma cloroplastidial de *Cenchrus americanus*, *Cenchrus purpureus* e *Cenchrus ciliaris*. *Large Single Copy*. SSC: *Small Single Copy*. IR: *Inverted Repeat*. ²Continuação da estrutura de *rps19*. *Gene duplicado. Genes acima da linha azul são transcritos em sentido anti-horário e abaixo da linha azul em sentido horário.

As junções entre as regiões LSC e IRb de *Cenchrus purpureus* e *Cenchrus americanus* possuem os genes *rpl22* em LSC e *rps19* em IRb, separados por uma região intergênica de 86 pb. Essa mesma região em *Cenchrus ciliaris*, apresenta os mesmos genes, porém sem o espaço intergênico, no qual, *rpl22* se encontra dentro de LSC assim como 1 pb de *rps19*, sendo o restante desse último gene presente em IRb. Genes entre regiões de junção, como *rps19* em *Cenchrus ciliaris*, comumente são caracterizados como pseudogenes, assim como relatado por Machado e colaboradores (2017). Diversas espécies de Poaceae possuem esses mesmos genes entre LSC e IRb (HUANG et al., 2017; LIU et al., 2020).

A junção entre IRb e SSC em *Cenchrus purpureus* e *Cenchrus americanus* está localizada 29 pb internamente ao gene *ndhF*, fazendo com que esse gene se comporte como um pseudogene. Situação semelhante já foi descrita para espécies de *Oryza*, no qual a junção IRb/SSC se encontra 308 pb internamente ao gene *ndhF*, também caracterizado como um pseudogene (HUANG et al., 2017). Em *Cenchrus ciliaris* a junção IRb/SSC se encontra 99 pb adentro de uma cópia do gene *rps15*.

A região de junção entre SSC e IRa em *Cenchrus ciliaris* se encontra em uma segunda cópia do gene *rps15*, formando um pseudogene duplicado na espécie. A junção SSC/IRa em *Cenchrus purpureus* e *Cenchrus americanus* compreende o gene *ndhH* presente em sua maior

extensão em SSC e apenas 1 pb em IRa, assim como descrito para outras espécies de Poaceae como *Triticum aestivum* L. e *Avena* L. spp (LIU et al., 2020), no qual o comprimento de inserção de *nadhH* em IRa é variável, sendo caracterizado como um pseudogene em ambas as espécies.

A junção entre IRa e LSC em *Cenchrus purpureus* e *Cenchrus americanus* apresenta uma das cópias do gene *rps19* totalmente inserida em IRa. Esse resultado corrobora com a hipótese de expansão dessa região visando a total abrangência do gene *rps19*, situação comumente descrita para espécies de Gramíneas (SASKI et al., 2007b). Para *Cenchrus ciliaris* por sua vez, a expansão de IRa para abranger o gene *rps19* possivelmente não ocorreu. Nesse caso, o gene se encontra entre as junções IRa/LSC, separado em duas partes presentes em LSC, o que o caracteriza como um pseudogene. Machado e colaboradores (2017) descrevem essa mesma posição, assim como classificam *rps19* como um pseudogene em espécies da ordem Myrtales, indicando essa localização como um evento comum em Angiospermas.

Processos de expansão e contração das regiões IRs são comumente relatados em plantas terrestres (PALMER, 1985), assim como em espécies de Gramíneas (SASKI et al., 2007b) e *Cenchrus* (RAVEENDAR et al., 2019). Esses processos são conhecidamente responsáveis pelas variações em comprimento das regiões IR do genoma cloroplastidial de espécies filogeneticamente relacionadas (PALMER, 1985) e podem explicar as variações entre as regiões de junção dos cpDNA de *Cenchrus ciliaris* em relação a *Cenchrus purpureus* e *Cenchrus americanus*. O menor comprimento das regiões IR em *Cenchrus ciliaris* é uma evidência que pode indicar processos de contração das mesmas na espécie. Um outro indício pode ser visto nas junções presentes em IRa e IRb de *Cenchrus ciliaris*, no qual, comparado a *Cenchrus purpureus* e *Cenchrus americanus*, essas regiões de fronteira se encontram sempre localizadas em genes anteriores aos presentes nessas espécies, refletindo o menor comprimentos das IRs de *Cenchrus ciliaris*. Relatos de contração de IRs através de análises comparativas já foram propostos para espécies de Geraniaceae (GUISINGER et al., 2011) e Apiaceae (PLUNKETT; DOWNIE, 2000), ambas Angiospermas.

A similaridade relatada através das análises comparativas entre o genoma cloroplastidial de *Cenchrus purpureus*, *Cenchrus americanus* e *Cenchrus ciliaris*, corroboram com as relações filogenéticas propostas para essas espécies (DONADÍO et al., 2009; CHEMISQUY et al., 2010; RAVEENDAR et al., 2019). Esses resultados também corroboram os trabalhos que propõe o incremento das análises filogenéticas através de dados genômicos (DELSUC; BRINKMANN; PHILIPPE, 2005; PARKS; CRONN; LISTON, 2009). Futuras análises, através do sequenciamento maciço de genomas cloroplastidiais em outras espécies do gênero, podem

auxiliar ainda mais na elucidação das relações existentes em *Cenchrus* spp através da filogenômica.

Identificação de SNPs e INDELs nos genótipos Pioneiro e BN2 de *Cenchrus purpureus* e *Cenchrus americanus*

Cenchrus purpureus e *Cenchrus americanus* possuem diversos estudos envolvendo a identificação de marcadores moleculares cultivares específicos, como microssatélites (SOUSA AZEVEDO et al., 2012), marcadores RAPD (PEREIRA et al., 2008), izoenzimas, entre outros (CAVALCANTE; LIRA, 2010). Apesar de muitos desses estudos estarem focados em genótipos oriundos do Banco Ativo de Germoplasma da EMBRAPA Gado de Leite – Juiz de Fora – MG, poucos estão relacionados às cultivares Pioneiro e BN2. Além disso, todos os trabalhos até o momento publicados, para ambas as espécies, utilizam de marcadores nucleares, não havendo ainda publicações que relatem marcas identificadas no genoma cloroplastidial dessas espécies.

De acordo com Sigmon e colaboradores (2017) marcadores moleculares oriundos do genoma cloroplastidial possuem diversas vantagens sobre os identificados no genoma nuclear. Uma dessas vantagens é a maior eficiência e robustez dessas marcas, principalmente quando se considera espécies poliplóides, no qual, os marcadores nucleares podem ser afetados pelo número de cópias, assim como pela frequência alélica relativa à sequência de DNA do marcador. Portanto, a identificação de possíveis marcadores moleculares a partir do genoma cloroplastidial de *Cenchrus americanus* cv BN2 e *Cenchrus purpureus* cv. Pioneiro, pode auxiliar em futuros trabalhos de identificação de cultivares, assim como de manejo de germoplasma em programas de melhoramento genético, como o que conduzido pela EMBRAPA Gado de Leite – Juiz de Fora – MG.

A partir dos genomas cloroplastidiais montados de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2, comparados, respectivamente, aos acessos MF594682 de *Cenchrus purpureus* e KX756179 de *Cenchrus americanus* depositados na base de dados NCBI, foram obtidas variações relacionadas a Polimorfismos de Nucleotídeo Único (SNPs) e Inserções/Deleções de Nucleotídeos (INDELs). A partir dessas análises foram obtidos 19 SNPs candidatos e 17 INDELs candidatos para Pioneiro, enquanto que para BN2 foram identificados 6 SNPs candidatos, porém nenhum INDEL foi obtido através das comparações (Tabela 5).

Tabela 5 Polimorfismos de Nucleotídeo Único (SNPs) e Inserções/Deleções (INDELs) identificados no genoma de *Cenchrus purpureus* cv. Pioneiro e *Cenchrus americanus* cv. BN2

Variação	Região do Genoma	<i>Cenchrus purpureus</i>	<i>Cenchrus americanus</i>
		MF594682/Pioneiro	KX756179/BN2
Polimorfismo de Nucleotídeo Único - SNP	Região gênica	4	1
	Região não-gênica	15	5
Total de SNPs		19	6
Inserção/Deleção de Bases - INDEL	Região Gênica	0	0
	Região não-gênica	17	0
Total de INDELs		17	0

Em *Cenchrus purpureus*, foram encontrados 4 SNPs em região gênica, sendo considerados aqui apenas os genes codificadores de proteínas, enquanto que em regiões não gênicas, foram identificados 15 SNPs. O total de SNPs corresponde a 0,13 SNPs/Kb. Em *Cenchrus americanus*, com menor número total de SNPs (0,04 SNPs/Kb), apenas 1 foi identificado em região gênica, enquanto que 5 foram identificados em região não gênica. Esse mesmo padrão foi observado na análise de Inserção e Deleção de nucleotídeos, no qual, a região não gênica apresenta maior frequência dos mesmos em *Cenchrus purpureus*. Maiores valores de SNPs e INDELs em regiões não codificantes quando comparadas às codificantes são uma constante em espécies vegetais, principalmente em gramíneas como *Festuca* L. (QIU et al., 2019) e *Oryza* L. (TANG et al., 2004).

A montagem detalhada e comparativa do genoma cloroplastidial de *Cenchrus americanus* cv. BN2, *Cenchrus purpureus* cv. Pioneiro e *Cenchrus ciliaris*, inédita para tais espécies, demonstra a eficiência da utilização de dados genômicos em estudos filogenéticos. Fornece ainda, informações de polimorfismos estáveis, como os 19 SNPs na cultivar Pioneiro de *Cenchrus purpureus* e 6 SNPs na cultivar BN2 de *Cenchrus americanus*, que podem ser usados em trabalhos evolutivos e na manutenção de germoplasma, após validação.

4. CONCLUSÕES

As três espécies possuem um genoma cloroplastidial similar em estrutura e composição, o que corrobora com a proximidade filogenética das mesmas, assim como com a proposição de similaridade entre espécies anteriormente encaixadas em *Pennisetum* (*Cenchrus americanus* e *Cenchrus purpureus*) e espécies de *Cenchrus* (*Cenchrus ciliaris*), no qual resultou na unificação de ambos os gêneros.

A homologia entre os genomas de *Cenchrus purpureus* e *Cenchrus americanus* é maior que entre elas e *Cenchrus ciliaris*, corroborando com a maior proximidade filogenética entre as duas primeiras espécies do que das mesmas para com a última.

Foram identificados 19 SNPs na cultivar Pioneiro de *Cenchrus purpureus* e 6 SNPs na cultivar BN2 de *Cenchrus americanus*, que após validação, podem ser extremamente úteis no manejo de germoplasma assim como em análises populacionais.

REFERÊNCIAS

- BHATT, P. P.; THAKER, V. S. The complete chloroplast genome of *Cenchrus ciliaris* (Poaceae). **Mitochondrial DNA Part B**, v. 3, n. 2, p. 674–675, 3 jul. 2018. doi: 10.1080/23802359.2018.1481795.
- BOCK, R. Structure, function, and inheritance of plastid genomes. In: **Current**. [s.l.: s.n.] 20 p. 29–63. 2007.
- CAVALCANTE, M.; LIRA, M. de A. Variabilidade genética em *Pennisetum purpureum* Schumacher. **Revista Caatinga**, v. 23, n. 2, p. 153–163, 2010.
- CHEMISQUY, M. A.; GIUSSANI, L. M.; SCATAGLINI, M. A.; KELLOGG, E. A.; MORRONE, O. Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): A combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. **Annals of Botany**, v. 106, n. 1, p. 107–130, 2010. doi: 10.1093/aob/mcq090.
- DARLING, A. C. E.; MAU, B.; BLATTNER, F. R.; PERNA, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. **Genome Research**, v. 14, n. 7, p. 1394–1403, jul. 2004. doi: 10.1101/gr.2289704.
- DELSUC, F.; BRINKMANN, H.; PHILIPPE, H. Phylogenomics and the reconstruction of the tree of life. **Nature Reviews Genetics**, v. 6, n. 5, p. 361–375, maio 2005. doi: 10.1038/nrg1603.
- DONADÍO, S.; GIUSSANI, L. M.; KELLOGG, E. A.; ZUOLAGA, F. O.; MORRONE, O. A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the trnL-F, rpl16 chloroplast markers. **Taxon**, v. 58, n. 2, p. 392–404, 2009.
- EISEN, J. A. Phylogenomics: Intersection of Evolution and Genomics. **Science**, v. 300, n. 5626, p. 1706–1707, 13 jun. 2003. doi: 10.1126/science.1086292.
- F COSTA, J.; LIN, S. M.; MACAYA, E. C.; FERNÁNDEZ-GARCÍA, C.; VERBRUGGEN, H. Chloroplast genomes as a tool to resolve red algal phylogenies: A case study in the Nemaliales. **BMC Evolutionary Biology**, v. 16, n. 1, p. 1–13, 10 out. 2016. doi: 10.1186/s12862-016-0772-3.
- GUISINGER, M. M.; KUEHL, J. V.; BOORE, J. L.; JANSEN, R. K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. **Molecular Biology and Evolution**, v. 28, n. 1, p. 583–600, 1 jan. 2011. doi: 10.1093/molbev/msq229.
- HANNA, W. W. Method of reproduction in napier grass and in the 3X and 6X allopolyploid hybrids with pearl millet. **Crop Science**, Madison, v. 21, p. 123–126, 1981.
- HUANG, Y.-Y.; CHO, S.-T.; HARYONO, M.; KUO, C.-H. Correction: Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. **PLOS ONE**, v. 12, n. 9, p. e0184409, 8 set. 2017. doi: 10.1371/journal.pone.0184409.

HYUN, J.; DO, H. D. K.; JUNG, J.; KIM, J.-H. Development of molecular markers for invasive alien plants in Korea: a case study of a toxic weed, *Cenchrus longispinus* L., based on next generation sequencing data. **PeerJ**, v. 7, n. 11, p. e7965, 11 nov. 2019. doi: 10.7717/peerj.7965.

KEARSE, M.; MOIR, R.; WILSON, A.; STONES-HAVAS, S.; CHEUNG, M.; STURROCK, S.; BUXTON, S.; COOPER, A.; MARKOWITZ, S.; DURAN, C.; THIERER, T.; ASHTON, B.; MEINTJES, P.; DRUMMOND, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012. doi: 10.1093/bioinformatics/bts199.

LI, X.; ZHANG, T.-C.; QIAO, Q.; REN, Z.; ZHAO, J.; YONEZAWA, T.; HASEGAWA, M.; CRABBE, M. J. C.; LI, J.; ZHONG, Y. Complete Chloroplast Genome Sequence of Holoparasite *Cistanche deserticola* (Orobanchaceae) Reveals Gene Loss and Horizontal Gene Transfer from Its Host *Haloxylon ammodendron* (Chenopodiaceae). **PLoS ONE**, v. 8, n. 3, p. e58747, 15 mar. 2013. doi: 10.1371/journal.pone.0058747.

LI, Y.; SYLVESTER, S. P.; LI, M.; ZHANG, C.; LI, X.; DUAN, Y.; WANG, X. The Complete Plastid Genome of *Magnolia zenii* and Genetic Comparison to Magnoliaceae species. **Molecules**, v. 24, n. 2, p. 261, 11 jan. 2019. doi: 10.3390/molecules24020261.

LIU, Q.; LI, X.; LI, M.; XU, W.; SCHWARZACHER, T.; HESLOP-HARRISON, J. S. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. **BMC Plant Biology**, v. 20, n. 1, p. 406, 2 dez. 2020. doi: 10.1186/s12870-020-02621-y.

LOHSE, M.; DRECHSEL, O.; BOCK, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. **Current Genetics**, v. 52, n. 5–6, p. 267–274, 7 nov. 2007. doi: 10.1007/s00294-007-0161-y.

MACHADO, L. de O.; VIEIRA, L. do N.; STEFENON, V. M.; OLIVEIRA PEDROSA, F. de; SOUZA, E. M. de; GUERRA, M. P.; NODARI, R. O. Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. **Genetica**, v. 145, n. 2, p. 163–174, 9 abr. 2017. doi: 10.1007/s10709-017-9954-1.

MAIER, R. M.; NECKERMANN, K.; IGLOI, G. L.; KOÖSSEL, H. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal of Molecular Biology**, v. 251, n. 5, p. 614–628, 1 ago. 1995. doi: 10.1006/jmbi.1995.0460.

PALMER, J. D. Comparative organization of chloroplast genomes. **Annual review of genetics**. Annu Rev Genet, , 1985. . doi: 10.1146/annurev.ge.19.120185.001545.

PARKS, M.; CRONN, R.; LISTON, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. **BMC Biology**, v. 7, n. 1, p. 84, 2009. doi: 10.1186/1741-7007-7-84.

PEREIRA, A. Vander; MACHADO, M. A.; AZEVEDO, A. L. S.; NASCIMENTO, C. S. do; CAMPOS, A. L.; LÉDO, F. J. da S. Diversidade genética entre acessos de capim-elefante obtida com marcadores moleculares. **Revista Brasileira de Zootecnia**, v. 37, n. 7, p. 1216–1221, jul. 2008. doi: 10.1590/S1516-35982008000700011.

PLUNKETT, G. M.; DOWNIE, S. R. Expansion and Contraction of the Chloroplast Inverted Repeat in Apiaceae Subfamily Apioideae. **Systematic Botany**, v. 25, n. 4, p. 648, 1 out. 2000. doi: 10.2307/2666726.

QIU, Y.; HIRSCH, C. D.; YANG, Y.; WATKINS, E. Towards Improved Molecular Identification Tools in Fine Fescue (*Festuca* L., Poaceae) Turfgrasses: Nuclear Genome Size, Ploidy, and Chloroplast Genome Sequencing. **Frontiers in Genetics**, v. 10, n. December, p. 1–12, 6 dez. 2019. doi: 10.3389/fgene.2019.01223.

RAVEENDAR, S.; LEE, G. A.; LEE, K. J.; SHIN, M. J.; LEE, J. R.; LEE, S. Y.; CHO, G. T.; MA, K. H.; CHUNG, J. W. The complete chloroplast genome of pearl millet (*Pennisetum glaucum* (L.) R. Br.) and comparative analysis within the family poaceae. **Cereal Research Communications**, v. 47, n. 1, p. 1–10, mar. 2019. doi: 10.1556/0806.46.2018.064.

ROYAL BOTANIC GARDENS KEW. (2020) Seed Information Database (SID). Version 7.1. Available from: <http://data.kew.org/sid/> (October 2020)

SASKI, C.; LEE, S.-B.; FJELLHEIM, S.; GUDA, C.; JANSEN, R. K.; LUO, H.; TOMKINS, J.; ROGNLI, O. A.; DANIELL, H.; CLARKE, J. L. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. **Theoretical and Applied Genetics**, v. 115, n. 4, p. 571–590, 2 ago. 2007a. doi: 10.1007/s00122-007-0567-4.

SIGMON, B. A.; ADAMS, R. P.; MOWER, J. P. Complete chloroplast genome sequencing of vetiver grass (*Chrysopogon zizanioides*) identifies markers that distinguish the non-fertile ‘Sunshine’ cultivar from other accessions. **Industrial Crops and Products**, v. 108, n. January, p. 629–635, dez. 2017. doi: 10.1016/j.indcrop.2017.07.029.

SOBRINHO, F. D. S.; PEREIRA, A. Vander; JOSÉ, F.; BOTREL, M. A. Avaliação agronômica de híbridos interespecíficos entre capim-elefante e milheto. **Pesquisa Agropecuária Brasileira**, v. 40, n. 9, p. 873–880, 2005.

SOUSA AZEVEDO, A. L.; COSTA, P. P.; MACHADO, J. C.; MACHADO, M. A.; PEREIRA, A. Vander; JOSÉ DA SILVA LÉDO, F. Cross Species Amplification of *Pennisetum glaucum* Microsatellite Markers in *Pennisetum purpureum* and Genetic Diversity of Napier Grass Accessions. **Crop Science**, v. 52, n. 4, p. 1776–1785, jul. 2012. doi: 10.2135/cropsci2011.09.0480.

SOUZA, K. A. De; ABRANTES, R. S. X.; LAÍS, C.; DE, D.; SOARES, M. A.; SANTOS, C.; VINICIUS, M.; PAULO, L.; FEITOSA, M. Panorama de sistema de produção de leite no Brasil. **Informativo Técnico do Semiárido**, v. 10, n. 2, p. 57–61, 2016.

TANG, J.; XIA, H.; CAO, M.; ZHANG, X.; ZENG, W.; HU, S.; TONG, W.; WANG, J.; WANG, J.; YU, J.; YANG, H.; ZHU, L. A comparison of rice chloroplast genomes. **Plant Physiology**, v. 135, n. 1, p. 412–420, 2004. doi: 10.1104/pp.103.031245.

TILLICH, M.; LEHWARK, P.; PELLIZZER, T.; ULBRICHT-JONES, E. S.; FISCHER, A.; BOCK, R.; GREINER, S. GeSeq – versatile and accurate annotation of organelle genomes. **Nucleic Acids Research**, v. 45, n. W1, p. W6–W11, 3 jul. 2017. doi: 10.1093/nar/gkx391.

WANG, Q. The plastid genome of an ornamental grass *Pennisetum villosum* (Poaceae: Paniceae). **Mitochondrial DNA Part B**, v. 5, n. 2, p. 1586–1587, 2 abr. 2020. doi: 10.1080/23802359.2020.1742211.

WANG, W.-C.; CHEN, S.-Y.; ZHANG, X.-Z. Chloroplast Genome Evolution in Actinidiaceae: clpP Loss, Heterogenous Divergence and Phylogenomic Practice. **PLOS ONE**, v. 11, n. 9, p. e0162324, 2 set. 2016. doi: 10.1371/journal.pone.0162324.

WIERSEMA, J.H., TURLAND, N.J., BARRIE, F.R., GREUTER, W., HAWKSWORTH, D.L., HERENDEEN, P.S., KNAPP, S., KUSBER, W.-H., LI, D.-Z., MARHOLD, K., MAY, T.W., MCNEILL, J., MONRO, A.M., PRADO, J., PRICE, M.J. & SMITH, G.F. (eds.) 2018+ [continuously updated]: *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*: Appendices I–VII. Online at < <https://naturalhistory2.si.edu/botany/codes-proposals/> > [accessed 4 April 2020].

WU, Y.; ZHOU, H. The complete chloroplast genome sequence of *Cenchrus purpureus*. **Mitochondrial DNA Part B**, v. 4, n. 1, p. 51–52, 2 jan. 2019. doi: 10.1080/23802359.2018.1536451.

XU, J.; SONG, Y.; JING, X.; LI, M. Characterization of the complete chloroplast genome sequence of *Pennisetum glaucum* and its phylogenetic implications. **Mitochondrial DNA Part B: Resources**, v. 4, n. 2, p. 3764–3765, 2019. doi: 10.1080/23802359.2019.1668312.

ZULOAGA, F. O.; MORRONE, O.; GIUSSANI, L. M. A cladistic analysis of the Paniceae: a preliminary approach. **Grasses: systematics and evolution**, n. June, p. 123–135, 2000.