

**MONTAGEM, CARACTERIZAÇÃO E COMPARAÇÃO DO GENOMA
ORGANELAR DE ESPÉCIES SELVAGENS E CULTIVARES HÍBRIDOS DE
CANA-DE-AÇÚCAR**

DEISE FERREIRA FERNANDES PAES

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO - UENF**

**CAMPOS DOS GOYTACAZES, RJ
FEVEREIRO – 2018**

MONTAGEM, CARACTERIZAÇÃO E COMPARAÇÃO DO GENOMA
ORGANELAR DE ESPÉCIES SELVAGENS E CULTIVARES HÍBRIDOS DE
CANA-DE-AÇÚCAR

DEISE FERREIRA FERNANDES PAES

“Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia.”

Orientadora: Prof^ª. Clícia Grativol Gaspar de Matos

CAMPOS DOS GOYTACAZES, RJ
FEVEREIRO – 2019

MONTAGEM, CARACTERIZAÇÃO E COMPARAÇÃO DO GENOMA
ORGANELAR DE ESPÉCIES SELVAGENS E CULTIVARES HÍBRIDOS DE
CANA-DE-AÇÚCAR

DEISE FERREIRA FERNANDES PAES

“Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia.”

Aprovada em 27 de fevereiro de 2019.

Comissão Examinadora:

Régis Lopes Corrêa (D.Sc., Ciências) - UFRJ

Jorge Hernandez Fernandez (D.Sc., Biologia Celular e Estrutural) - UENF

Fábio Lopes Olivares (D.Sc., Agronomia) - UENF

Prof^a. Clícia Grativol Gaspar de Matos (D.Sc., Química Biológica) – UENF
(Orientadora)

AGRADECIMENTOS

Agradeço primeiramente a Deus pelo dom da vida e por estar sempre ao meu lado diante das provações vivenciadas;

À Doutora, orientadora e acima de tudo professora Clícia Grativol Gaspar de Matos, sem seu apoio e ensinamentos eu não conseguiria alcançar meu tão sonhado objetivo, tenho a plena certeza que Deus colocou essa pessoa tão brilhante, humilde e com potencial gigante em meu caminho, caminho esse que foi galgado com muito sacrifício, mas sempre com grandes expectativas e fé;

Ao Professor Dr. Thiago Motta Venâncio por aceitar colaborar com este projeto cedendo equipamentos de informática e mão de obra humana, na figura do Dr. Filipe Pereira Matteoli que foi essencial nos aprendizados para utilização de softwares de bioinformática, além de ser o revisor desta dissertação, sem vocês meu caminho seria mais tortuoso;

À minha família, em especial meu esposo por tanto apoio, incentivo e disponibilidade de até mudar de cidade para que fosse mais fácil minha realidade, palavras são poucas para expressar tamanha gratidão e amor por você, às minhas três filhas que mesmo sem entender facilitaram minha adaptação à nova vida, muitas vezes me fiz ausente, mas é por vocês e para vocês tudo isso, aos meus pais, Cláudia Mara Gomes Ferreira Fernandes e José Jorge Amim Fernandes que sempre foram à base da nossa família, que sempre me esperam de braços abertos, à minha cunhada Aline Rocha Miranda que nos momentos de desespero estava pronta a me ouvir e me incentivar mais uma vez;

Aos amigos do Laboratório de Química e Função de Proteínas e Peptídeos (LQFPP) com vocês eu aprendi a chorar, mas principalmente a sorrir, e agora me aproprio da celebre frase “o importante é que emoções eu vivi”;

À Universidade Estadual do Norte Fluminense Darcy Ribeiro e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo custeio e financiamento deste trabalho, aos professores da pós-graduação em Biociências e Biotecnologia que dedicaram horas do seu dia preparando material e ensinando as mais diversas disciplinas, a todos os professores da banca examinadora que se dispusera a avaliar este trabalho e colaborar com suas valiosas sugestões;

E por fim a todos que tenham contribuído de forma direta ou indiretamente para que este trabalho fosse realizado da melhor maneira possível. Obrigada!!

“O conhecimento e a persistência são o combustível do Mundo. Ouse, persista... às vezes é necessário um pouco mais de teimosia!”

SUMÁRIO

1. INTRODUÇÃO	1
1.1. Um breve histórico da cana-de-açúcar no Brasil	1
1.2. Filogenia e variabilidade genética da cana-de-açúcar	3
1.3. Genômica	5
1.3.1 Genômica em plantas	6
1.3.2. Genoma e estrutura organelar	7
1.3.2.1 Estrutura genômica mitocondrial	11
1.3.2.2 Estrutura genômica cloroplastidial.....	13
2. OBJETIVOS	16
2.1. Objetivo geral.....	16
2.2. Objetivos específicos	16
3. MATERIAIS E MÉTODOS	17
3.1. Obtenção de bibliotecas de DNA genômico e métodos de busca de DNA organelar	17
3.2. Montagem dos genomas de cloroplasto e mitocôndria de cana-de-açúcar.....	17
3.3. Anotação dos genomas organelares de cana-de-açúcar	18
3.4. Comparação do genoma cloroplastidial e mitocondrial de espécies selvagens e cultivares híbridos de cana-de açúcar	20
3.4.1 Análise de divergência de sequência	20
3.4.2 Caracterização de sequências repetidas	20
3.4.3 Detecção de Indels e sua localização no genoma	20
3.4.4 Análise das regiões IRs de cloroplastos.....	21
4. RESULTADOS	22
4.1. Montagem de genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar	22
4.1.1. Genoma cloroplastidial	22
4.1.2. Genoma mitocondrial.....	23
4.2. Anotação de genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar	24
4.2.1. Genoma cloroplastidial	24
4.2.2. Genoma mitocondrial.....	31

4.3. Análise de divergência de sequência no genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar	42
4.3.1 Cloroplastos	42
4.3.2 Mitocôndrias	42
4.4. Análise de sequências repetitivas no genoma organelar	48
4.4.1. Sequências repetitivas em cloroplastos	48
4.4.2. Sequências repetitivas em mitocôndrias	50
4.5. Detecção de Indels no genoma organelar de cana-de-açúcar	57
4.5.1 <i>Indels</i> e sua localização no genoma de cloroplasto	57
4.5.2 <i>Indels</i> e sua localização no genoma de mitocôndrias	58
4.6. Comparação das regiões IRs de cloroplastos de cana-de-açúcar	59
5. DISCUSSÃO	60
5.1. Montagem e anotação de genomas	60
5.2. Análise de divergência de sequência nucleotídica nas organelas	62
5.3. Análise de regiões repetitivas nos genomas organelares	63
5.4. Análise de inserção/deleção nos genomas organelares	65
5.5. Análise das fronteiras IRs de cloroplastos	66
6. CONCLUSÕES	67
7. REFERÊNCIAS BIBLIOGRÁFICAS	69

LISTA DE FIGURAS

- Figura 1-** Hibridização natural da cana-de-açúcar. A imagem mostra hibridização natural entre o selvagem *S. spontaneum* e *S. officinarum* (proveniente da especiação do selvagem *S. robustum*) originando os cultivares *S. barberi* e *S. sinense*. Adaptado de GRIVET et al., 2006. 4
- Figura 2-** Montagem de sequência genômica utilizando Whole genome shotgun. Adaptada de Green(2001). 6
- Figura 3-** Diagrama esquemático da evolução dos eucariotas. As mitocôndrias surgiram a partir de α -proteobactérias e os cloroplastos surgiram de cianobactérias, ambos foram incorporados através da endossimbiose e contribuíram para complementar os genes encontrados atualmente em núcleos eucarióticos. Adaptada de Timmis et al. (2004). 8
- Figura 4-** Esquema de biogênese e fisiologia de organelas, mostrando suas diferentes características e funções entre cloroplastos e mitocôndrias. DNA cloroplastidial (DNAcP); DNA mitocondrial (DNAMt); Ciclo TCA (ciclo do ácido tricarboxílico). Adaptada de Van Dingenen et al. (2016) 10
- Figura 5-** Estrutura de um genoma mitocondrial de cana-de-açúcar, evidenciando seus dois cromossomos circulares e seus respectivos genes (Shearman et al., 2016). 13
- Figura 6-** Estrutura de um genoma cloroplastidial de cana-de-açúcar. Evidenciando suas quatro partes: Região grande de cópia única(LSC); Região menor de cópia única (SSC); Região repetida e invertida (IRa) e (IRb) (HOANG et al., 2015). 15
- Figura 7-** Esquema de estratégias metodológicas de montagem e anotação do genoma organelar da cana-de-açúcar. A célula vegetal está representada em verde contendo o DNA total, sendo:(N) para DNA nuclear, (M) para DNA mitocondrial e (C) para DNA cloroplastidial. 19
- Figura 8-** Mapa circular do genoma de cloroplasto de cana-de-açúcar. As setas indicam o sentido em que a transcrição ocorre (horário e anti-horário). O círculo interno mostra o conteúdo de GC no genoma e as linhas pretas no círculo externo demarcam as regiões IRs. As letras de A à G representam o genoma de cada espécie *Saccharum*. (A) *S.officinarum*, (B) *S.spontaneum*, (C) RB72454, (D) RB867515, (E) Tambo FEPAGRO, (F) Roxa, (G) B4362..... 31
- Figura 9-** Cromossomos do genoma mitocondrial de cana-de-açúcar . Os genes do lado externo dos cromossomos 1 e 2 estão na cadeia positiva, os genes mostrados dentro do círculo estão localizados na cadeia negativa. O círculo interno de cor cinza em cada

cromossomo representa o conteúdo de GC e os genes representados são classificados na legenda em cores. As letras de A à G representam o genoma mitocondrial de cada espécie <i>Saccharum</i> . (A) <i>S.officinarum</i> , (B) <i>S.spontaneum</i> , (C) RB72454, (D) RB867515, (E) Tambo FEPAGRO, (F) Roxa, (G) B4362.	40
Figura 10- Divergência de sequência baseado em sete genomas de cloroplasto de cana-de-açúcar com referência a <i>Saccharum officinarum</i> IJ76-514. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas (CNS) do genoma são mostradas em rosa.....	45
Figura 11- Divergência de sequência baseado em sete genomas de cana-de-açúcar com referência ao cultivar Khon Kaen 3 cromossomo 1. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas(CNS) do genoma são mostradas em rosa.	46
Figura 12- Divergência de sequência baseado em sete genomas de cana-de-açúcar com referência ao cultivar Khon Kaen 3 cromossomo 2. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas(CNS) do genoma são mostradas em rosa.	47
Figura 13- Total das repetições existentes no genoma cloroplastidial de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	48
Figura 14- Tipos de repetições (Tandem, Palindromicas e Dispersas) no genoma de cloroplasto de espécies selvagens e cultivares híbridos de cana-de-açúcar.	49
Figura 15- Frequência de motivos de SSRs identificados. Ao lado direito do gráfico a legenda em cores representa as sequências nucleotídicas(di-nucleotídeo, tri-nucleotídeo, tetra-nucleotídeo e penta-nucleotídeos) encontradas em cloroplastos de cana-de-açúcar. .	50
Figura 16- Total das repetições existentes no genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar. (A) cromossomo 1, (B) cromossomo 2.....	51
Figura 17- Distribuição dos tipos de repetições (Tandem, Palindromicas e Dispersas) no cromossomo 1 (A) e no cromossomo 2 (B) do genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	53
Figura 18- Frequência de motivos de SSRs identificados. Ao lado direito do gráfico a legenda em cores representa as sequências nucleotídicas(di-nucleotídeo, tri-nucleotídeo, tetra-nucleotídeo, penta-nucleotídeos e hexanucleotídeo) encontradas no cromossomo 1(A) e no cromossomo 2 (B) do genoma mitocondrial de cana-de-açúcar.....	56

Figura 19- Maior porcentagem de indels e sua localização no genoma cloroplastial de cana-de-açúcar. Os números no eixo horizontal separados por hífen abaixo do gráfico indica o início e o fim da localização dos indels dentro do genoma de cada espécie e cultivar especificada. No eixo vertical a esquerda está a maior porcentagem encontrada de indels..... 57

Figura 20- Maior porcentagem de indels e sua localização no cromossomo 1 do genoma mitocondrial de cana-de-açúcar..... 58

Figura 21- Comparação das fronteiras IRs em cloroplastos de cana-de-açúcar. Junção LSC- IRa (JLA); junção IRa-SSC (JSA); junção SSC-IRb (JSB) e junção IRb-LSC (JLB).
..... 59

LISTA DE TABELAS

Tabela 1- Parâmetros de montagem do genoma de cloroplasto de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	22
Tabela 2- Parâmetros de montagem do genoma de mitocôndrias de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	23
Tabela 3- Comparação da anotação do genoma cloroplastidial de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	32
Tabela 4- Comparação da anotação do genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar.....	41

LISTA DE ABREVIACÕES

- AGORA**- Annotator for Genome of Organelle from Referenced sequence Analysis
- BLAST**- Basic Local Alignment Search Tool
- CDS**- Região de codificação (do inglês *coding sequence*)
- CNS**- Sequencia não codificante conservada (do inglês *conserved non-coding sequence*)
- Contigs**- Sequência contínua de fragmentos de DNA
- DGB**- Gafro de Bruijn (do inglês *De Bruijn graph*)
- INDELS**- Deleção ou inserção de pares de base em um genoma
- IR**- Região invertida do genoma de cloroplasto
- GenBank**- banco de dados de anotações de sequências de nucleotídeos
- GS**- genome selection
- GWAS**- genome-wide association studies
- K-mers** - todas as possíveis subsequências (de comprimento k) de uma leitura obtida através do sequenciamento de DNA.
- LSC**- Grande região de cópia única do genoma de cloroplasto
- NCBI**- National Center for Biotechnology Information
- OGDRAW**- Draw Organelle Genome Maps
- ORF**- open reading frame
- READS** – leituras curtas proveniente do sequenciamento Illumina
- RNAr**- ácido ribonucleico ribossomal
- RNA^t**- ácido ribonucleico transportador
- SSC**- pequena região de cópia única do genoma de cloroplasto
- SSR**- sequências simples repetidas

RESUMO

O sequenciamento do genoma de organelas é uma ferramenta essencial na busca de informações moleculares e evolutivas em plantas. Além do genoma nuclear, as plantas possuem genoma mitocondrial e de cloroplasto, que podem fornecer uma ampla visão sobre o relacionamento filogenético entre espécies. O gênero *Saccharum* (cana-de-açúcar) pertencente à família Poaceae, vem sendo muito utilizado na produção de açúcar/álcool. Nos últimos anos, os cultivares RB e SP, lançados através de programas de melhoramento genético, são utilizados para fins agrônômicos e pouco se sabe sobre o impacto da hibridização de espécies selvagens sobre os genomas de organelas destes cultivares. Assim, o objetivo deste trabalho foi comparar os genomas mitocondrial e cloroplastidial de espécies selvagens e cultivares híbridos de cana-de-açúcar, através da montagem, anotação e análises de regiões repetitivas dos genomas organelares. A metodologia envolveu a utilização de ferramentas computacionais para a filtragem, seleção de *reads*, montagem, anotação e análise de divergência de sequências dos genomas de cloroplasto e mitocôndria de espécies selvagens e cultivares híbridos. Os resultados mostraram similaridades de sequências entre as espécies selvagens. Uma dessas similaridades é a menor conservação nos genes cloroplastidiais *ycf2* e *ccsa* comparado aos cultivares híbridos, e também menor conservação no gene mitocondrial *rps2*. Somente *S. officinarum* apresentou o gene ACRS. Os cultivares híbridos apresentaram uma maior identidade de sequência no genoma mitocondrial com o híbrido já publicado Khon Kaen 3. Tanto o genoma mitocondrial quanto o de cloroplasto apresentaram maior divergência de sequência nas regiões não codificantes. Dentre repetições analisadas (tandem, palindrômicas e microssatélites), as de microssatélites foram mais abundantes dentro dos genomas organelares. Os indels detectados no genoma de cloroplasto e no cromossomo 1 das mitocôndrias foram em sua maioria na região intergênica, mas no cromossomo 2 todos os indels detectados estavam dentro do gene *rps4*. O gene *psbA* se mostrou bem conservado nas análises das regiões IRs dos cloroplastos, podendo ter sofrido seleção positiva e os genes *rpl22*, *rps19*, *rps15*, *ndhF* e *ndhH* sofreram compressão ou expansão nos genomas de cloroplastos. Essas análises comparativas contribuem para agregar informações sobre o DNA organelar no gênero *Saccharum*, suas divergências e similaridades para que essas informações possam ser utilizadas em marcadores moleculares, ajudando a compreender os polimorfismos existentes e correlaciona-los com a hibridização no gênero.

Palavras-chave: *Saccharum*, genoma, organelas, montagem, anotação

ABSTRACT

Sequencing of the organelle genome is an essential tool in the search for molecular and evolutionary information in plants. In addition to the nuclear genome, plants have mitochondrial and chloroplast genomes, which can provide a broad view of the phylogenetic relationship between species. The genus *Saccharum* (sugarcane) belonging to the family Poaceae, has been widely used in the production of sugar. In recent years, cultivars RB and SP, introduced through genetic improvement programs, are used for agronomic purposes and little is known about the impact of hybridization of wild species on the organelle genomes of these cultivars. Thus, the objective of this work was to compare the mitochondrial and chloroplast genomes of wild species and hybrid sugarcane cultivars, by assembling, annotating and analyzing repetitive regions of the organelle genomes. The methodology involved the use of computational tools for the filtering, selection of reads, assembly, annotation and analysis of sequence divergence of the chloroplast and mitochondrial genomes of wild species and hybrid cultivars. The results showed similarities of sequences between wild species. One of these similarities is the lower conservation in the chloroplastid genes *ycf2* and *ccsa* compared to the hybrid cultivars, and also less conservation in the mitochondrial *rps2* gene. Only *S. officinarum* presented the ACRS gene. Hybrid cultivars showed a higher sequence identity in the mitochondrial genome with the Khon Kaen 3 hybrid already published. Both the mitochondrial and chloroplast genomes showed greater sequence divergence in the non-coding regions. Amongst the analyzed replicates (tandem, palindromic and microsatellite), those of microsatellites were more abundant within the organellar genomes. The indels detected in the chloroplast genome and chromosome 1 of the mitochondria were mostly in the intergenic region, but on chromosome 2 all the indels detected were within the *rps4* gene. The *psbA* gene was well conserved in the analyzes of the chloroplast IR regions and may have been positively selected and the *rpl22*, *rps19*, *rps15*, *ndhF* and *ndhH* genes underwent compression or expansion in the chloroplast genomes. These comparative analyzes contribute to aggregate information about the organellar DNA in the *Saccharum* genus, its divergences and similarities so that this information can be used in molecular markers, helping to understand the existing polymorphisms and correlating them with the hybridization in the genus.

Key words: *Saccharum*, genome, organelles, assembly, annotation

1. INTRODUÇÃO

1.1. Um breve histórico da cana-de-açúcar no Brasil

O primeiro registro da cana-de-açúcar no Brasil trata-se de quando a cultura foi trazida da Ilha da Madeira por Martin Afonso de Souza em 1522. Sendo introduzida posteriormente por Duarte Coelho Pereira em Pernambuco no ano de 1533 (BASTOS, 1987). Durante o período colonial, a cana-de-açúcar era exclusivamente destinada à produção de açúcar. Somente em 1970, a indústria explorou a potencialidade e o uso da cana-de-açúcar como fonte de energia renovável (SZMRECSÁNYI; MOREIRA, 1991), sendo atualmente modelo no uso do etanol como biocombustível atraindo o interesse internacional (TÁVORA, 2011). O bioetanol ganhou destaque a partir da crise do petróleo, no qual houve a escassez do produto e resultou na conscientização da população sobre os malefícios do uso da gasolina proveniente do petróleo (LEITE; LEAL, 2007). Em consequência do aumento do preço do petróleo, o Brasil criou o programa Proálcool em 1975 pelo decreto nº 76.593 com o intuito de diminuir o uso da gasolina. Os carros foram adaptados e passaram a utilizar o etanol hidratado. Em torno de 1986, os carros movidos a álcool chegaram a 90% das vendas (ANDRIETTA, M. G. S.; ANDRIETTA, S. R.; STECKELBERG, C.; STUPIELLO, 2007). No entanto, o governo diminuiu o compromisso com o programa Proálcool, o que prejudicou a cadeia produtiva do bioetanol, diminuindo o abastecimento e o interesse por carros a álcool no ano de 1999, dessa forma o Brasil passou a exportar mais açúcar e a importar o etanol dos Estados Unidos para misturar na gasolina (Furtado e Scandiffio 2007). A partir de 2003 com a introdução de carros bicompostíveis no Brasil a produção sucroalcooleira aumentou chegando a 30,4 milhões de toneladas de açúcar e 17,8 bilhões de litros de álcool na safra de 2006/2007, aumentando também suas exportações(MINISTÉRIO DA AGRICULTURA, 2007).

Inicialmente a economia de Campos dos Goytacazes girava em torno do gado. Depois surgiram os engenhos, marcados por mão-de-obra escrava. Já no final do século XIX, os engenhos declinaram e emergiram as modernas usinas de açúcar e álcool, que no início enfrentaram dificuldades em termos de mão-de-obra por causa da abolição da escravatura. Os ex-escravos voltaram ao trabalho de forma assalariada (SILVA, 2004). A partir do ano 1930, Campos dos Goytacazes já não apresentava o mesmo vigor vivido de

1835 em diante. Em torno de 1973, a crise atingiu o setor canavieiro fluminense, diferente do Estado de São Paulo que crescia no setor sucroalcooleiro (SILVA, 2004; SOUSA, 1935). A região Norte Fluminense do Estado do Rio de Janeiro destacou-se na indústria sucroalcooleira no passado. A cidade de Campos dos Goytacazes fora a maior produtora de cana-de-açúcar de todo o Estado. Apesar disso, no ano de 2008 apenas quatro usinas encontravam-se em operação (FRANCISCO, 2009).

Nos últimos 40 anos a produtividade canavieira do Brasil alcançou e superou índices mundiais, mesmo havendo grande heterogeneidade entre regiões produtoras (ROSA SAMBUICHI, 2014). A produtividade é maior no Estado de São Paulo, seguido de Minas Gerais e, posteriormente, Goiás. Segundo um levantamento sistemático da produção agrícola de cana-de-açúcar no Brasil em 2016, o Estado de São Paulo obteve 442 milhões de toneladas, o Estado de Minas Gerais produziu 69 milhões de toneladas e Goiás 71 milhões de toneladas (IBGE, 2017). Os cultivares RB (República Brasileira) e SP (São Paulo), lançados através de programas de melhoramento genético, são utilizados para fins agronômicos. A variedade SP70-1143, que foi muito cultivada nas décadas de 80 e 90, sofre caracteristicamente com a presença do fungo causador da ferrugem alaranjada (*Puccinia kuehnii*). Esta variedade foi gradativamente substituída pela RB72454 e, subsequentemente, pela sua progênie RB867515, que atingiu altos índices de cultivo no ano de 2010 (DAL-BIANCO et al., 2012; MATSUOKA et al., 2009). A cultivar RB72454 é proveniente da variedade CP53-76 como mãe e possui pai não identificado. A RB72454 possui adaptabilidade e estabilidade, grande quantidade de sacarose, mas é suscetível principalmente a estrias vermelhas. Sua progênie RB867515 possui crescimento rápido e alta produtividade e é muito resistente, principalmente a seca (HOFFMAN et al., 2008).

A cana-de-açúcar é uma importante matéria prima, utilizada na fabricação de uma grande gama de produtos como o açúcar, etanol, cachaça, melado, rapadura e também na alimentação animal (DIAS, 2011). Além da utilização do polietileno para a fabricação de plásticos biodegradáveis, o que reduz os impactos ambientais (MORSCHBACKER, 2009). O etanol proveniente da cana-de-açúcar produz aproximadamente 9 unidades de energia para cada unidade utilizada em seu processamento (INTERNATIONAL ENERGY AGENCY, 2004), enquanto que o milho e a beterraba na Europa possuem 1,4 e 2,0 unidades, respectivamente (MACEDO; SEABRA; SILVA, 2008; WORLD WATCH INSTITUTE, 2006). Dessa forma a utilização da cana-de-açúcar na produção de energia é mais eficiente do que o milho e a beterraba utilizados para este fim.

1.2. Filogenia e variabilidade genética da cana-de-açúcar

A família Poaceae é bem variada havendo mais de 10.000 espécies (WATSON & DALLWITZ, 1992), e se destaca predominantemente entre a flora de todo mundo, possuindo grande potencial ecológico e econômico (BREMER, 2002). Entre todas as plantas terrestres, as gramíneas somam mais de 20% (SHANTZ, 1954). Além da cana-de-açúcar (*Saccharum spp.*), cereais como sorgo (*Sorghum bicolor*), milho (*Zea mays*), trigo (*Triticum spp.*), cevada (*Hordeum vulgare*), aveia (*Avena sativa*) e arroz (*Oryza sativa*) fazem parte desta família (DANIELS; ROACH, 1987).

A origem da cana-de-açúcar não é totalmente esclarecida, embora a hipótese mais aceita de acordo com a distribuição geográfica, é que a cultura seja proveniente da Nova Guiné, no Pacífico Sul (FORTUCCI; DIVISION, 1999). Estudos neste gênero mostram que a endogamia tende a homozigose, podendo surgir características deletérias, ao passo que o cruzamento entre indivíduos distintos pode evitar tais características indesejáveis (BISON et al., 2004; RODRIGUES et al., 2001; SILVA et al., 2005). Dessa forma, o conhecimento da diversidade genética inter-cultivares é essencial para o melhoramento genético (PALOMINO et al., 2005).

Revisões taxonômicas ocorridas no século passado agruparam espécies que realizam intercruzamento sem gêneros, o que deu origem ao “Complexo Saccharum”. Dentre todas as espécies, apenas *S. spontaneum* e *S. robustum* são consideradas selvagens (DANIELS; ROACH, 1987; MUKHERJEE, 1957). Hibridizações naturais entre *S. spontaneum* e *S. officinarum* deram origem aos cultivares do Norte da Índia (*S. barberi*) e Chinês (*S. sinense*) (GRIVET; GLASZMANN; D’HONT, 2006), como mostra a figura 1.

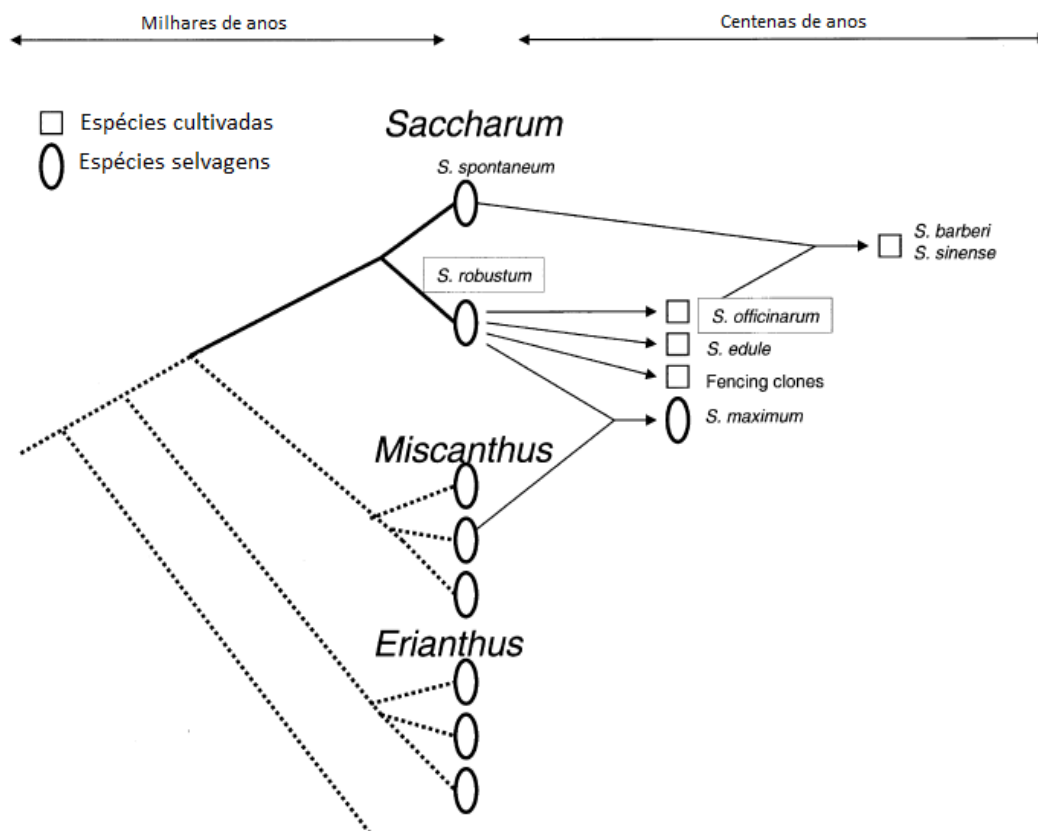


Figura 1- Hibridização natural da cana-de-açúcar. A imagem mostra hibridização natural entre o selvagem *S. spontaneum* e *S. officinarum* (proveniente da especiação do selvagem *S. robustum*) originando os cultivares *S. barberi* e *S. sinense*. Adaptado de GRIVET et al., 2006.

O cruzamento entre as espécies *S. spontaneum* ($2n=36-128$) e *S. officinarum* ($2n=80$) realizado no início do século 20 por melhoristas, resultou em um híbrido (*Saccharum spp.*) que foi retrocruzado com *S. officinarum*, viabilizando o desenvolvimento do comércio açucareiro (EDME et al., 2001; MOORE, 2005; SALLY et al., 2007). Este cruzamento conferiu ao novo cultivar a capacidade de acumular sacarose herdada principalmente de *S. officinarum* e a resistência a estresses bióticos e abióticos herdada de *S. spontaneum* (CHEAVEGATTI-GIANOTTO et al., 2011; GRIVET; ARRUDA, 2002). Segundo Panje e Babu (1960), hipóteses sugerem que uma hibridização natural entre *Sclerostachya* e *Erianthus* seção *Rapidium* ocorreu no norte da Índia dando origem a *S. spontaneum* com cromossomo $2n=40$. Já *S. officinarum* também conhecida como cana nobre, é a única do gênero a ter o número de cromossomos constante ($2n=80$). Possui colmos largos e doces, chegando a 22% de sacarose. Pode apresentar várias cores e folhas largas, chegando a 7cm de limbos (FERRARI, 2010; PRICE, 1955).

1.3. Genômica

Ao longo da história, importantes descobertas contribuíram para o avanço da genética e genômica, tais como a descoberta da hereditariedade (MENDEL, 1865); a descoberta de que o DNA era o material genético hereditário, sua estrutura e seu código genético (AVERY, 1944; NIRENBERG, 1963; WATSON; CRICK, 1953); criação de tecnologias com DNA recombinante (COHEN et al., 1973; JACKSON; SYMONS; BERG, 1972); a determinação da sequência completa do genoma de um bacteriófago de cadeia simples de DNA denominado ϕ X174 (SANGER et al., 1977); e tecnologias para o sequenciamento de DNA de forma automática (LANDER et al., 2001; MAXAM; GILBERT, 1977; SANGER; COULSON, 1975; SMITH et al., 1986). Essas descobertas viabilizaram o sequenciamento de vários genomas de procariotos e eucariotos, sendo o primeiro genoma completamente sequenciado o da bactéria *Haemophilus influenzae Rd* (FLEISCHMANN et al., 1995). Outros genomas de espécies importantes também foram parcialmente sequenciados como *Caenorhabditis elegans* (SULSTON et al., 1992), *Bacillus subtilis* (GLASER et al., 1993) e *Saccharomyces cerevisiae* (LEVY, 1994). Após estes, os genomas de *Escherichia coli K-12* (BLATTNER et al., 1997), *Drosophila melanogaster* (KULKARNI et al., 2002), *Arabidopsis thaliana* (THE ARABIDOPSIS GENOME INITIATIVE, 2000) e o genoma humano (LANDER et al., 2001) foram publicados.

Para desvendar a sequência de nucleotídeos organizados em genomas complexos como o humano, uma grande evolução nos métodos de sequenciamento ocorreu, acarretando a geração da metodologia *Whole Genome Shotgun* – WGS (POP et al., 2004). Nessa técnica, primeiramente ocorre a fragmentação do genoma para obtenção de sequências menores para então, sequenciá-las. Após isso, os fragmentos de DNA sequenciados (*reads*) são montados utilizando softwares com algoritmos que realizam as melhores combinações de sobreposições, a fim de obter *contigs*. A principal estratégia utilizada na montagem de genomas é o Grafo-de-Bruijn (DBG), no qual os *reads* são organizados em grafos - conjunto de vértices e arestas, onde arestas representam as sobreposições e vértices representam os *reads* (LI et al., 2011). Após a montagem dos *contigs*, estes são organizados em *scaffolds* e, por fim, a sequência genômica é reconstruída (GREEN, 2001)(Figura 2).

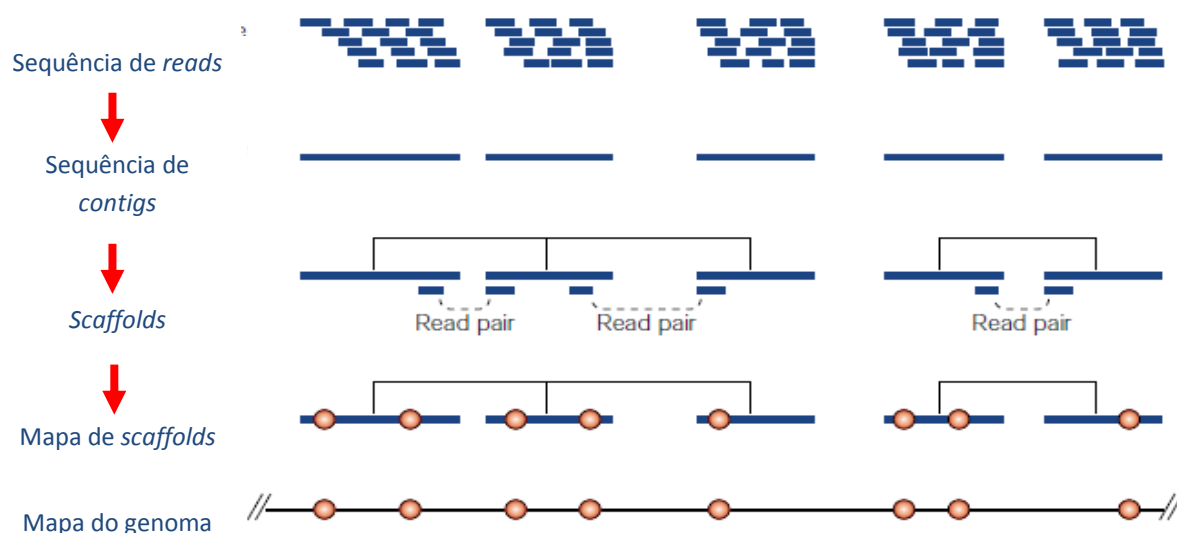


Figura 2- Montagem de sequência genômica utilizando *Whole genome shotgun*. Adaptada de Green(2001).

Em montagens de genomas complexos há a probabilidade de ocorrer erros, principalmente devido às regiões repetitivas e rearranjo de sequências (PHILLIPPY; SCHATZ; POP, 2008). Programas de montagem devem ser cada vez mais aprimorados, para lidarem com a montagem de genomas grandes e repetitivos (LI et al., 2011). Após a montagem de um genoma, é realizada a anotação de genes e sua respectiva função biológica, através da similaridade com sequências genômicas relacionadas (CONESA et al., 2005). Genomas completos ou *drafts* (genomas incompletos) são geralmente depositados no banco de dados GenBank (BENSON et al., 2012), disponibilizado pelo National Center for Biotechnology Information (NCBI). Em 2019, neste banco de dados há cerca de 8114 genomas organelares de animais e 2773 genomas organelares de plantas depositados (NCBI, 2019).

1.3.1 Genômica em plantas

Antes da era genômica, as técnicas utilizadas para a criação de cultivares e o aumento do rendimento da maioria das culturas estavam baseadas nos métodos de seleção fenotípica através da combinação de parentais com características genéticas favoráveis (HALLAUER, 2011; ALLARD, 1960; GRIFFING, 1956). A genômica tem causado um grande impacto no melhoramento de plantas, visto que por meio de tecnologias como *Genome-wide Association Studies* – GWAS ou *Genome Selection* - GS, é possível realizar

uma avaliação mais aprimorada da relação do genótipo com o fenótipo (FEDOROFF, 2010; TESTER; LANGRIDGE, 2010). A primeira planta a ter seu genoma completo sequenciado foi *Arabidopsis thaliana* no ano de 2000 e até hoje é um modelo no estudo de genomas de plantas cultivadas e não cultivadas (THE ARABIDOPSIS GENOME INITIATIVE, 2000).

Cerca de 25% das plantas, ao longo de sua história evolutiva, realizaram cruzamentos naturais com outras espécies, levando ao surgimento de híbridos. A hibridização é a maior responsável pela origem de novas espécies (MALLET, 2005; SOLTIS; SOLTIS, 2009). Muitos híbridos são cultivados atualmente, dentre eles a cana-de-açúcar. Devido à sua origem interespecífica, a cana-de-açúcar possui um dos genomas mais complexos entre as plantas cultivadas (CHEN, 2010). Vários trabalhos já haviam montado regiões do genoma da cana-de-açúcar (GRATIVOL et al., 2014; RIAÑO-PACHÓN; MATTIELLO, 2017), e também vários transcriptomas (CARDOSO-SILVA et al., 2014; HOANG et al., 2017; SANTA BRIGIDA et al., 2016; VETTORE et al., 2003; XU et al., 2018). Em 2018, dois trabalhos anunciaram o sequenciamento do genoma da cultivar R570 e de *Saccharum spontaneum* (GARSMEUR et al., 2018; ZHANG et al., 2018). No primeiro trabalho Garsmeur e colaboradores utilizaram cromossomos artificiais de bactérias (BAC) e sorgo como referência para construir um genoma monoplóide de referência em mosaico. Os autores listaram em torno de 25.316 genes que codificam proteínas. O segundo trabalho traz o sequenciamento de *S. spontaneum* (AP85-441) utilizando a abordagem “*de novo*”, ou seja, não referenciada, dessa forma construíram 32 pseudo-cromossomos, contendo cerca de 35.525 genes (ZHANG et al., 2018). Além dos genomas nucleares, o sequenciamento e a comparação dos genomas organelares também podem contribuir com informações sobre a história evolutiva da cana-de-açúcar, fornecendo um recurso valioso para o melhoramento genético (GUALBERTO et al., 2014). No entanto, pouco ainda se conhece sobre os genomas organelares de cana-de-açúcar.

1.3.2. Genoma e estrutura organelar

Além do genoma nuclear, as plantas possuem genoma mitocondrial e de cloroplasto, que ajudam a fornecer informações sobre espécies específicas (DANIELS; ROACH, 1987). O cloroplasto teria surgido a partir da endossimbiose de cianobactérias e a

mitocôndria de proteobactérias por células eucarióticas (Figura 3). Embora tenham mantido seus genomas, estes passaram por alterações estruturais (GRAY; BURGER; LANG, 1999; STOEBE; KOWALLIK, 1999), levando a um tamanho menor, quando comparado ao genoma nuclear, e a funções essenciais para a célula. Mesmo tendo sofrido reestruturação de seus genomas ao longo da evolução, as organelas ainda possuem mais de 1000 proteínas que garantem sua ancestralidade procariótica (MARTIN et al., 2002). Grande parte dos genes que codificam essas proteínas foi transferida para núcleo, sendo sintetizadas no citosol e redirecionadas as organelas. Estudos apontam que essa transferência é um processo de evolução contínua e que muitos desses genes se tornaram cópias nucleares funcionais que controlam a biogênese de mitocôndrias e cloroplastos, e que outros genes teriam evoluído para controlar ainda outras funções celulares (MARTIN et al., 2002; TIMMIS et al., 2004).

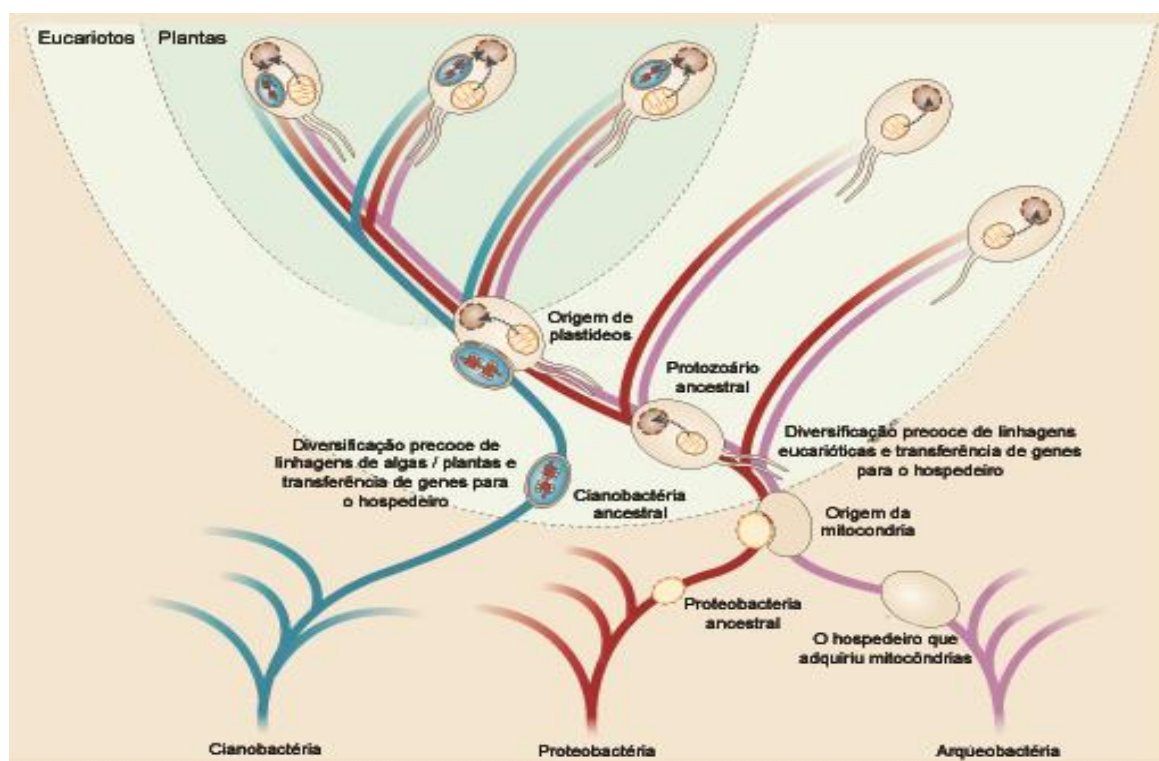


Figura 3- Diagrama esquemático da evolução dos eucariotas. As mitocôndrias surgiram a partir de α -proteobactérias e os cloroplastos surgiram de cianobactérias, ambos foram incorporados através da endossimbiose e contribuíram para complementar os genes encontrados atualmente em núcleos eucarióticos. Adaptada de Timmis et al. (2004).

Os cloroplastos são responsáveis por produzir vários compostos orgânicos através do processo de fotossíntese e sintetizam em torno de 80 proteínas (ALLEN et al., 2011; ROBERTIS, 2001). Já as mitocôndrias são responsáveis pela cadeia respiratória da célula e

possuem cerca de 63 genes codificadores de proteínas em plantas (LANG et al., 1997). Mitocôndrias e cloroplastos parecem ter um papel fundamental no desenvolvimento de folha e raízes, visto que genes codificadores de proteínas encontrados nessas organelas se apresentam diferencialmente expressos durante esse estágio da planta (VAN DINGENEN et al., 2016). Além disso, as mitocôndrias podem estar envolvidas com a proliferação celular e fornecimento de energia para produção de fontes de carbono. Mutantes para genes relacionados à biogênese dessa organela pode causar um fenótipo alterado em plantas ou até letalidade (VAN DINGENEN et al., 2016).

Muitas mutações no genoma nuclear que estão associadas com proteínas direcionadas a mitocôndrias e cloroplastos parecem afetar o desenvolvimento embrionário em plantas (BRYANT et al., 2011). Isso pode ser comprovado por Guillaumot e colaboradores (2017) e por Andrés-Colás e colaboradores (2017), ao descrever classes de proteínas de repetição pentatricopeptídica (PPR) que editam RNA em organelas, através da inserção de mutantes para essas proteínas. Esses autores observaram defeitos de edição nas organelas, o que causou problemas no desenvolvimento e letalidade embrionária, o que reforça a importância organelar no desenvolvimento de plantas (ANDRÉS-COLÁS et al., 2017; GUILLAUMOT et al., 2017). Pesquisadores afirmam que a letalidade embrionária relacionada a mutações no genoma mitocondrial, está relacionada à falta de fornecimento de energia por essa organela (XIU et al., 2016; ZMUDJAK et al., 2013). Cloroplastos podem ter um papel fundamental na embriogênese, antes mesmo de realizar a fotossíntese, visto que mutações em proteínas envolvidas no desenvolvimento e em mecanismos bioquímicos do cloroplasto afetam seriamente embriões em desenvolvimento, causando até sua morte (APUYA et al., 2002; CUSHING et al., 2005). Essas edições de RNA que ocorrem tanto no genoma de cloroplasto, quanto no de mitocôndria são processos pós-transcricionais que ocorrem em gimnospermas e angiospermas (CHEN et al., 2011; RAMAN; PARK, 2015).

Diferentemente das mitocôndrias, os cloroplastos necessitam de luz para completar sua formação e exercer sua função, o corpo prolamelar recebe luz, o que desencadeia a formação das tilacoides a partir da membrana interna que cresce e emite prolongamentos, o líquido dentro dos cloroplastos, o estroma, há também a formação de membranas empilhadas chamadas de grana e a síntese do DNA cloroplastidial (SAKAMOTO; MIYAGISHIMA; JARVIS, 2008). Após sua organização, o cloroplasto está pronto para realizar a biosíntese de aminoácidos e atuar como fonte de carbono. As mitocôndrias

possuem um fluido chamado matriz mitocondrial e possuem dobras na membrana interna denominada de cristas. Nas cristas ocorre o transporte de elétrons, necessário para gerar energia (ATP). As mitocôndrias tem DNA próprio como os cloroplastos, não sofrem grandes mudanças na sua formação e são formadas a partir de fissão de mitocôndrias que já existem (MILLAR et al., 2008). Após a ativação transcricional passam a exercer a função de fotorespiração, a matriz realiza o ciclo do ácido tricarboxílico e essas organelas também fazem sinalização para o núcleo regulando expressão gênica (Figura 4).

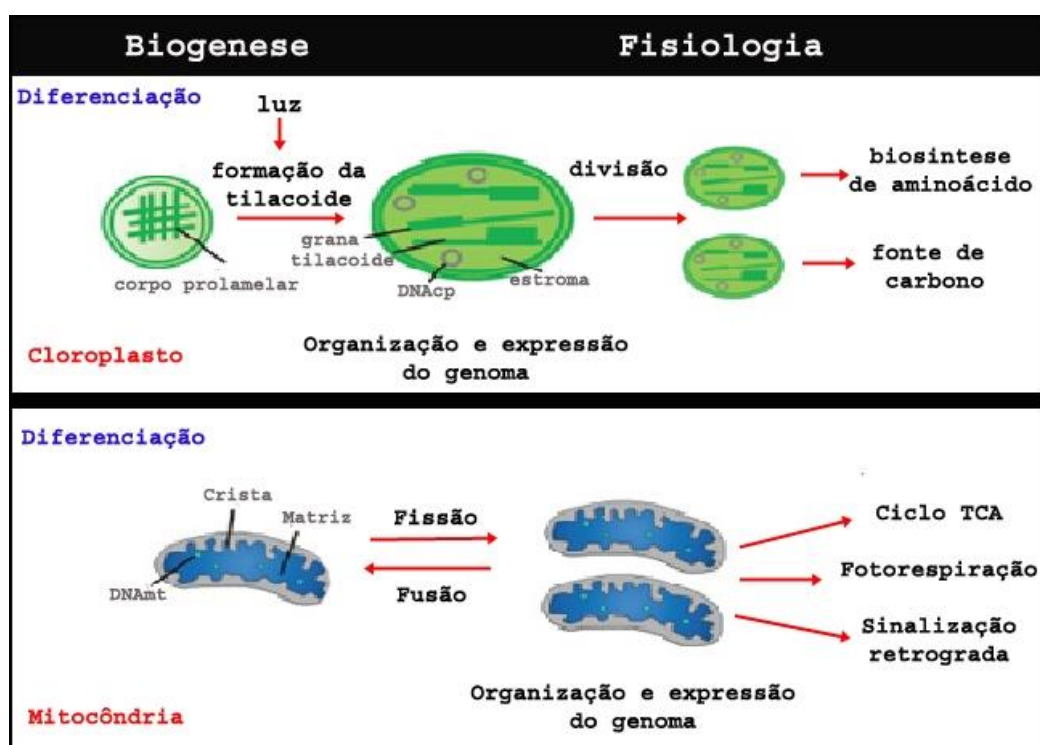


Figura 4- Esquema de biogênese e fisiologia de organelas, mostrando suas diferentes características e funções entre cloroplastos e mitocôndrias. DNA cloroplastidial (DNAcp); DNA mitocondrial (DNAm); Ciclo TCA (ciclo do ácido tricarboxílico). Adaptada de Van Dingenen et al. (2016)

Na atualidade com o avanço da tecnologia para o sequenciamento de nova geração, houve um grande aumento no número de sequenciamentos completos de organelas. Dentre as organelas, o genoma cloroplastidial é o mais sequenciado devido ao seu tamanho menor e por possuir regiões conservadas, diferente do genoma mitocondrial que apresenta grandes variações de tamanho entre as espécies (GRAY; BURGER; LANG, 1999; LANG; GRAY; BURGER, 1999). A estrutura genômica dessas duas organelas apresenta muitas repetições, que podem impactar a montagem (ALVERSON et al., 2010, 2011).

1.3.2.1 Estrutura genômica mitocondrial

A estrutura de um genoma mitocondrial foi definida por Shearman e colaboradores (2016), e é referida como um círculo mestre seguido de mini círculos formados por DNA provenientes de recombinação. Segundo Mower e colaboradores (2012), houve uma convenção para representar o genoma mitocondrial como um único círculo de DNA, no entanto, isso nem sempre é observado. O tamanho deste genoma é variável indo de 200 Kb em *Brassica hirta* (PALMER; HERBON, 1987), 445 Kb em *Saccharum officinarum* (SHEARMAN et al., 2016) à 11.3 Mb em *Silene conica* (SLOAN et al., 2012).

As repetições que fazem parte do genoma mitocondrial estão diretamente relacionadas à recombinação, essas são as responsáveis pelos tamanhos variados de genoma mitocondrial e são essenciais no processo de evolução (CHANG et al., 2013). Em um estudo realizado em milho comprovou-se que regiões metiladas no genoma mitocondrial possuem relação direta com as repetições contidas neste genoma (CLIFTON et al., 2004). Este genoma possui também muitos mecanismos de edição de RNA (HANDA, 2003; MULLIGAN; CHANG; CHOU, 2007), e com a evolução sofreu vários rearranjos e perda gênica (CHANG et al., 2013; LONSDALE; HODGE; FAURON, 1984), embora haja conservação de genes funcionais (DOMBROVSKA; QIU, 2004).

Segundo Handa (2003), espécies relacionadas dividem sítios de edição de RNA mitocondrial, chegando a 90% de compartilhamento. Geralmente as edições ocorrem em maior quantidade nas citosinas para uracilas, o que podem formar códons com funções indeterminadas (HANDA, 2003). Em relação à perda de gene, o genes *rps12*, *sdh2* e *sdh4* foram perdidos em monocotiledôneas (PALMER; HERBON, 1988; WOLFE; LI; SHARP, 1987), e os genes *rps9*, *rps11* e *rps16* também se perderam ao longo da evolução através da diferenciação entre algas e plantas superiores (NOTSU et al., 2002). Apesar dessas perdas, a taxa de mutação do genoma mitocondrial é considerada baixa, e por isso este genoma tem sido utilizado no estudo evolutivo de plantas (MA; GUO; LI, 2012; QIU et al., 2010; SLOAN et al., 2010).

Quanto à herança do genoma mitocondrial em plantas, em sua maioria é uniparental de origem materna (HATTORI et al., 2002; MOGENSEN, 1996). Quando a herança é biparental está diretamente relacionada à hibridização entre espécies distintas (BARR; NEIMAN; TAYLOR, 2005). Estes mesmos autores destacam em sua revisão que assim como as plantas, os animais também possuem estratégias diversas para evitar a transmissão

do genoma mitocondrial de origem paterna ao seu descendente. Aparentemente as mitocôndrias paternas são selecionadas para morrer, após a entrada do espermatozoide no óvulo, sua exposição no exterior do óvulo faz com que as mitocôndrias paternas sofram despolarização, o que as marca para morrer por autofagia (ASHRAFI; SCHWARZ, 2013; WANG; KLIONSKY, 2011). Vários estudos mostram que a permanência da mitocôndria paterna pode ser prejudicial para o desenvolvimento animal (DELUCA AND O'FARREL, 2012; LI; LUO; WANG, 2001; ZHOU et al., 2016). Para humanos acreditava-se que a herança mitocondrial era exclusivamente materna, até que recentemente um estudo em três famílias diferentes comprovou a herança biparental das mitocôndrias, embora esse evento seja raro (LUO et al., 2018).

Em se tratando do genoma mitocondrial de cana-de-açúcar, há apenas um registro de sequenciamento completo, referente a um híbrido moderno, o cultivar Khon Kaen 3. Este apresenta dois cromossomos depositados no *GenBank*, com as chaves de acesso LC107874.1 e LC107875.1. O genoma mitocondrial de cana-de-açúcar possui dois cromossomos circulares 1 e 2, sendo o primeiro de 300.784 pares de bases e o segundo de 144.698 pares de bases, conforme figura 5. Este genoma mitocondrial da cana-de-açúcar possui um total de 111 repetições e fragmentos gênicos provenientes de cloroplasto, dentre eles, genes da via oxidativa e do ribossomo, genes de RNAt, RNAr, genes para síntese de ATP, maturases e NADH desidrogenase, além de 85 genes, sendo 77 genes codificados por um único éxon e 8 genes provenientes de vários éxons (SHEARMAN et al., 2016). É comum a perda de genes do genoma mitocondrial para o genoma nuclear (ADAMS; PALMER, 2003). Embora Shearman et al (2016) demonstraram que o gene *trnL-CAA* tenha sido conservado no genoma mitocondrial da cana-de-açúcar quando comparado com o genoma mitocondrial de sorgo, este gene teria sido perdido para o núcleo, e ao longo da evolução, foi recuperado para o genoma mitocondrial novamente (SHEARMAN et al., 2016).

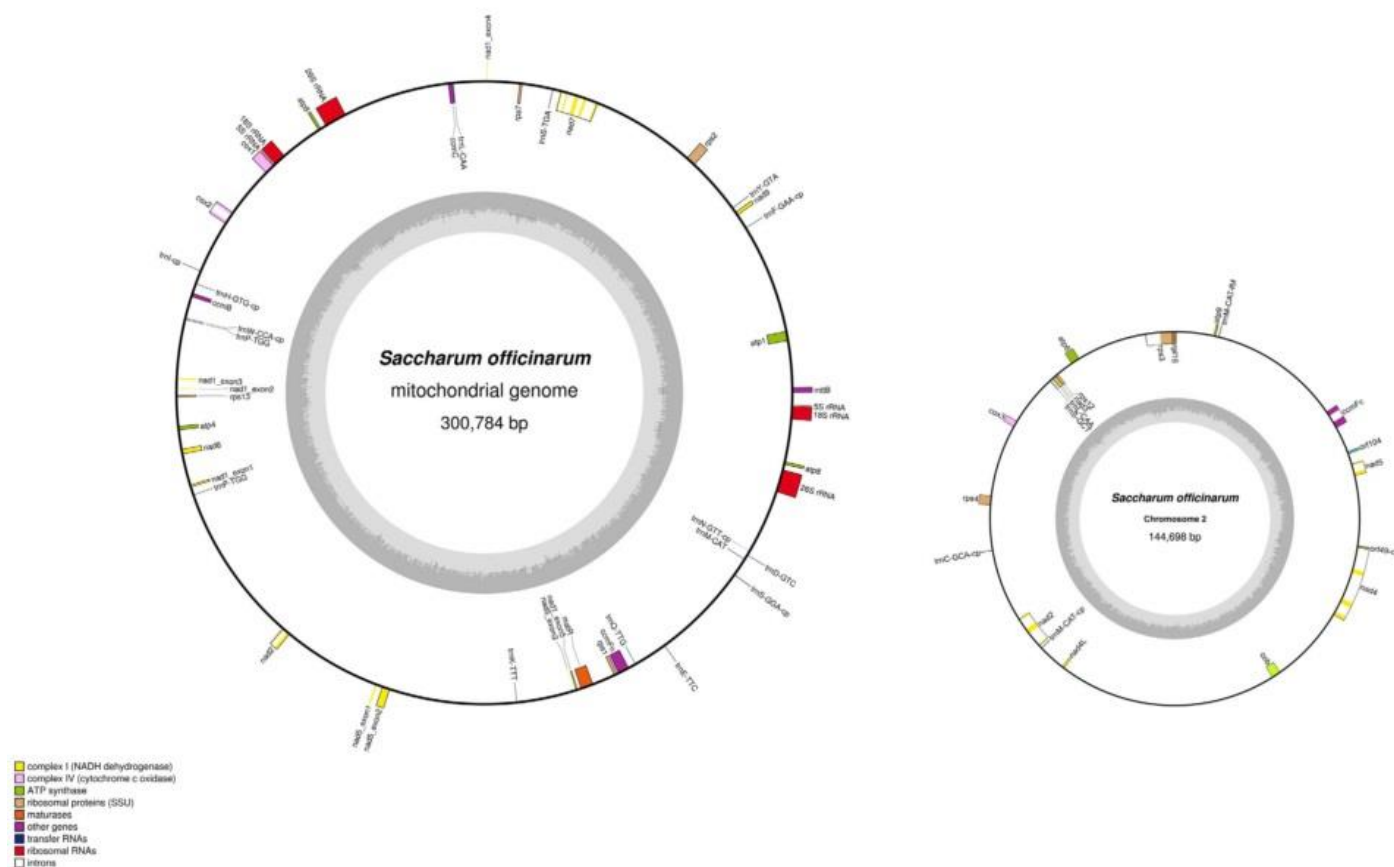


Figura 5- Estrutura de um genoma mitocondrial de cana-de-açúcar, evidenciando seus dois cromossomos circulares e seus respectivos genes (Shearman et al., 2016).

1.3.2.2 Estrutura genômica cloroplastidial

Para se comparar a variabilidade genética dentro e entre populações, as sequências do genoma cloroplastidial tem se tornado uma ferramenta primordial (DUMOLIN-LAPEGUE; PEMONGE; PETIT, 1997; TABERLET et al., 1991). Esses genomas não são recombinantes, geralmente a herança é monoparental materna, além de serem haploides (OLMSTEAD; PALMER, 1994). Apesar de existir casos de herança biparental em cloroplastos, eles são mais raros do que esse tipo de herança em mitocôndrias (HAVEY, 1997; MOGENSEN, 1996; VAILLANCOURT; PETTY; MCKINNON, 2004). No cloroplasto as regiões de microsatélites (cpSSR) tem se destacado em estudos de genética de populações. Oligonucleotídeos universais para a amplificação destas regiões têm sido descritos para monocotiledôneas (PROVAN; MURPHY; MAGGS, 2004) e para dicotiledôneas (WEISING; GARDNER, 1999).

As regiões do genoma de cloroplasto LSC, SSC e IRs são mais variáveis em tamanho dentro do grupo das gimnospermas, mas em relação ao tamanho total do genoma de cloroplastos, as monocotiledôneas apresentaram uma maior variação (XIAO-MING et al., 2017). Essas variações no genoma podem ser devido à perda de genes (WAKASUGI et al., 1994; WOLFE, K. H., MORDEN, C. W. & PALMER, 1992); variação da região repetida invertida – IR (CHUMLEY et al., 2006; GOCKEL; HACHTEL; MELKONIAN, 2000; LIN et al., 2003; PALMER; HERBON, 1987; PARKS; CRONN; LISTON, 2009; WAKASUGI et al., 1994; WOLFE, K. H., MORDEN, C. W. & PALMER, 1992); e variação da região que se localiza entre os genes (GREINER et al., 2008; SHAHID MASOOD et al., 2004; WU et al., 2011).

Desde 2004 pesquisadores investigam o genoma de cloroplasto em cana-de-açúcar (CALSA JÚNIOR et al., 2004). A montagem deste genoma revelou uma configuração com quatro partes, sendo: uma região grande denominada LSC, a região SSC de tamanho menor e duas regiões repetidas invertidas (IRs), que geralmente são denominadas IRa e IRb (Figura 6). Cinco híbridos *Saccharum* tiveram o genoma de cloroplasto sequenciado completamente, foram eles: Q155 (HOANG et al., 2015), NCo310 (ASANO et al., 2004), IJ76-514 (EVANS; JOSHI, 2016), SP80-3280 (CALSA JÚNIOR et al., 2004), RB72454 (EVANS; JOSHI, 2016) e RB867515 (VIDIGAL et al., 2016). Uma espécie selvagem também teve seu genoma sequenciado: *Saccharum spontaneum* clone SES234B (EVANS; JOSHI, 2016).

Os genomas de cloroplasto de cana-de-açúcar publicados possuem em média 140,000 pares de bases e genes relacionados à fotossistemas, síntese de ATP, RNA polimerase, subunidade grande da rubisco, proteínas ribossomais, RNAt, RNAr, ORFs e outros genes. Estes estudos mostram que o genoma de cloroplasto é bem conservado entre a linhagem *Saccharum*, embora alguns cultivares deste gênero apresentem polimorfismo (ASANO et al., 2004; CALSA JÚNIOR et al., 2004; EVANS; JOSHI, 2016; HOANG et al., 2015; VIDIGAL et al., 2016). A análise comparativa dos genomas de organelas pode ser uma ferramenta de grande importância para decifrar e compreender linhagens *Saccharum* híbridas, uma vez que na maioria das plantas o cloroplasto possui origem materna (HARRIS; INGRAM, 1991).

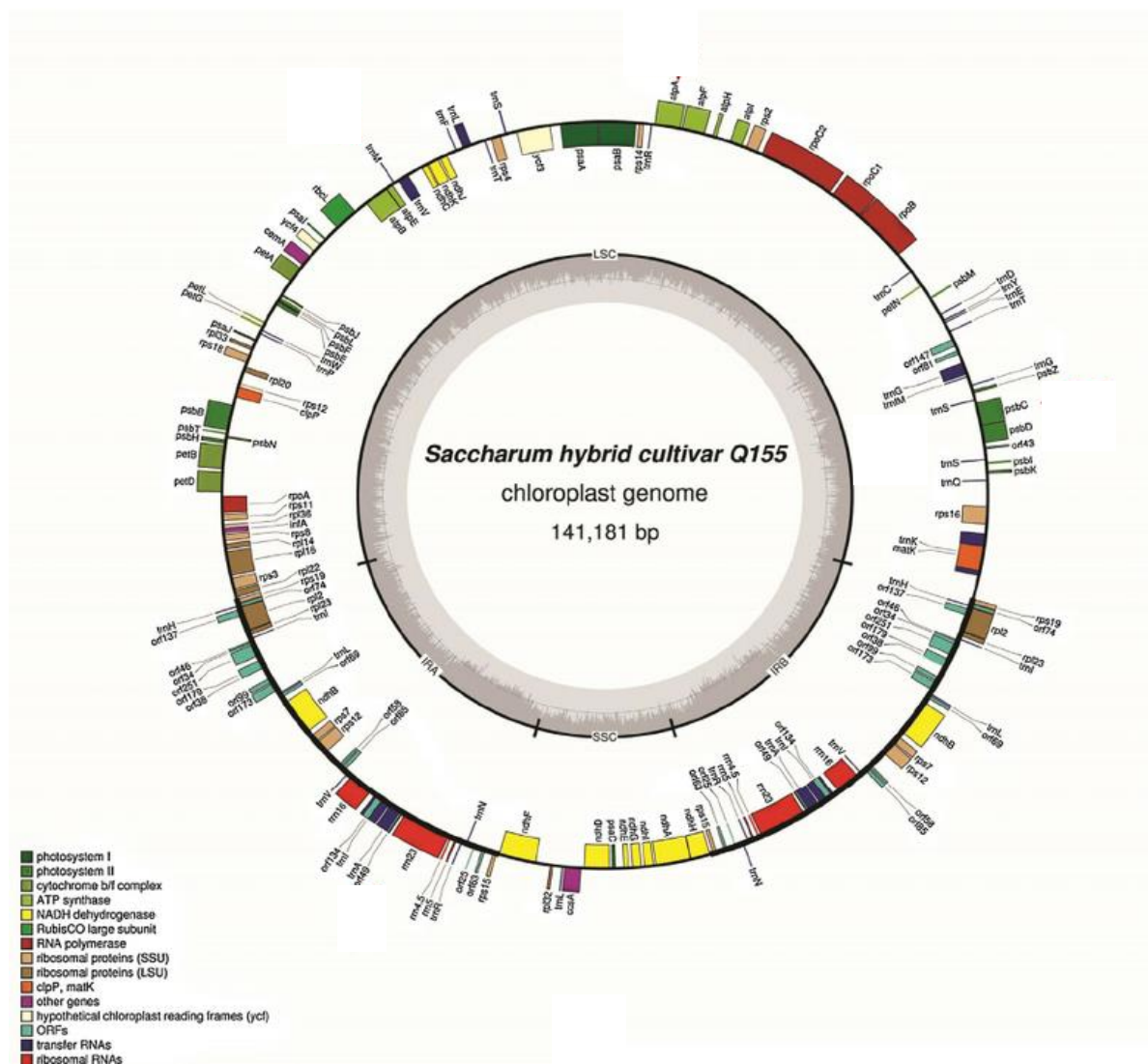


Figura 6- Estrutura de um genoma cloroplasmial de cana-de-açúcar. Evidenciando suas quatro partes: Região grande de cópia única(LSC); Região menor de cópia única (SSC); Região repetida e invertida (IRa) e (IRb) (HOANG et al., 2015).

2. OBJETIVOS

2.1. Objetivo geral

Analisar comparativamente os genomas cloroplastidial e mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar, com enfoque nas regiões conservadas e divergentes entre os genomas de selvagens e híbridos.

2.2. Objetivos específicos

- Montagem e anotação dos genomas de cloroplasto e mitocôndria de espécies selvagens *S. officinarum* (82-72) e *S. spontaneum* (SES205A) e cultivares híbridos RB72454, RB867515, B436269, Tambo FEPAGRO e Roxa;
- Análise da divergência de sequências entre cloroplastos e mitocôndrias de espécies selvagens e cultivares híbridos;
- Caracterização de sequências repetidas (tandem, palindrômicas e sequências simples) de cloroplastos e mitocôndrias de espécies selvagens e cultivares híbridos;
- Identificação de *indels* (inserção/deleção) e sua localização no genoma cloroplastidial e mitocondrial de espécies selvagens e cultivares híbridos;
- Realização de análise comparativa entre as regiões IRs dos cloroplastos de espécies selvagens e cultivares híbridos.

3. MATERIAIS E MÉTODOS

3.1. Obtenção de bibliotecas de DNA genômico e métodos de busca de DNA organelar

As bibliotecas de DNA total das espécies selvagens *Saccharum officinarum* (acesso SRA/NCBI: SRX313496), *Saccharum spontaneum* (acesso SRA/NCBI: SRX313495) e dos cultivares híbridos RB72454, RB867515, Tambo FEPAGRO, Roxa e B4362 foram previamente sequenciadas e filtradas por (GASPAR, 2014). O sequenciamento foi realizado com folhas jovens; 2 de cada genótipo utilizando o sequenciador Illumina GAII protocolo de 100 ciclos paired-end, depois de sequenciadas, os *reads* foram filtrados utilizando-se FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), para selecionar *reads* de qualidade, sendo retidas 90% das bases com qualidade maior ou igual a Q20. Para buscar os *reads* correspondentes ao genoma de cloroplasto e mitocôndria, utilizamos o software BLASTN(ALTSCHUL et al., 1990), realizando alinhamentos de sequências de genomas de cloroplastos e mitocôndrias próximos filogeneticamente com a cana-de-açúcar (*Saccharum*, *Sorghum bicolor*, *Zea mays*, *Miscanthus sinensis*, *Oryza sativa* e *Setaria italica*) disponíveis no Genbank contra os *reads* oriundos de cada genótipo de cana-de-açúcar (selvagens e híbridos). Para isso, usamos $e\text{-value} = 10^{-4}$ e o filtro DUST desligado. Os *reads* alinhados aos genomas de cloroplastos e mitocôndrias foram filtrados para posterior montagem.

3.2. Montagem dos genomas de cloroplasto e mitocôndria de cana-de-açúcar

Os *reads* de cloroplastos separados anteriormente foram testados com o software VelvetOptimizer (ZERBINO; BIRNEY, 2008) com intervalo de k-mer de 29 a 87. A montagem *de novo* foi realizada no software SPADES (BANKEVICH et al., 2012) com k-mer 53, 69 e 77, cov-cutoff 70 e parâmetro *careful* e no SSPACE (BOETZER et al., 2011), executado com parâmetros padrão nos contigs montados no SPADES. Após a montagem, os cloroplastos foram verificados através de alinhamento utilizando o cultivar *Saccharum* híbrido cultivar Q155 já publicado (HOANG et al., 2015). Os cloroplastos de *S. spontaneum* (depositado no genbank sob acesso: LN896360), RB72454 (EVANS; JOSHI, 2016) e RB867515 (VIDIGAL et al., 2016) depositados e publicados também foram utilizados para as demais análises.

Os *reads* mitocondriais foram montados de forma referenciada, através do software MITObim (HAHN; BACHMANN; CHEVREUX, 2013). Os parâmetros foram executados de acordo com a descrição do desenvolvedor no site <https://github.com/chrishah/MITObim>, utilizando como referência o cultivar publicado de cana-de-açúcar, o híbrido Khon Kaen 3 (SHEARMAN et al., 2016). Após a montagem, foi feita a verificação do genoma através de alinhamento com o genoma de referência Khon Kaen 3.

3.3. Anotação dos genomas organelares de cana-de-açúcar

Os genomas foram anotados com Geseq—Annotation of Organellar Genomes (TILLICH et al., 2017) disponível em <https://chlorobox.mpimp-golm.mpg.de/Geseq-app.html>. A conferência da anotação foi feita com o software AGORA - Annotator for Genes of Organelle from the Reference sequence Analysis (JUNG et al., 2018) e, também manualmente com dados da literatura (**Figura 7**). O arquivo *genbank* gerado pelo GeSeq/AGORA foi visualizado no OGDRAW—Organellar Genome DRAW (LOHSE; DRECHSEL; BOCK, 2007). Para detectar RNAs transportadores utilizamos o software TRNAscan-SE (LOWE, 1997) com configuração padrão.

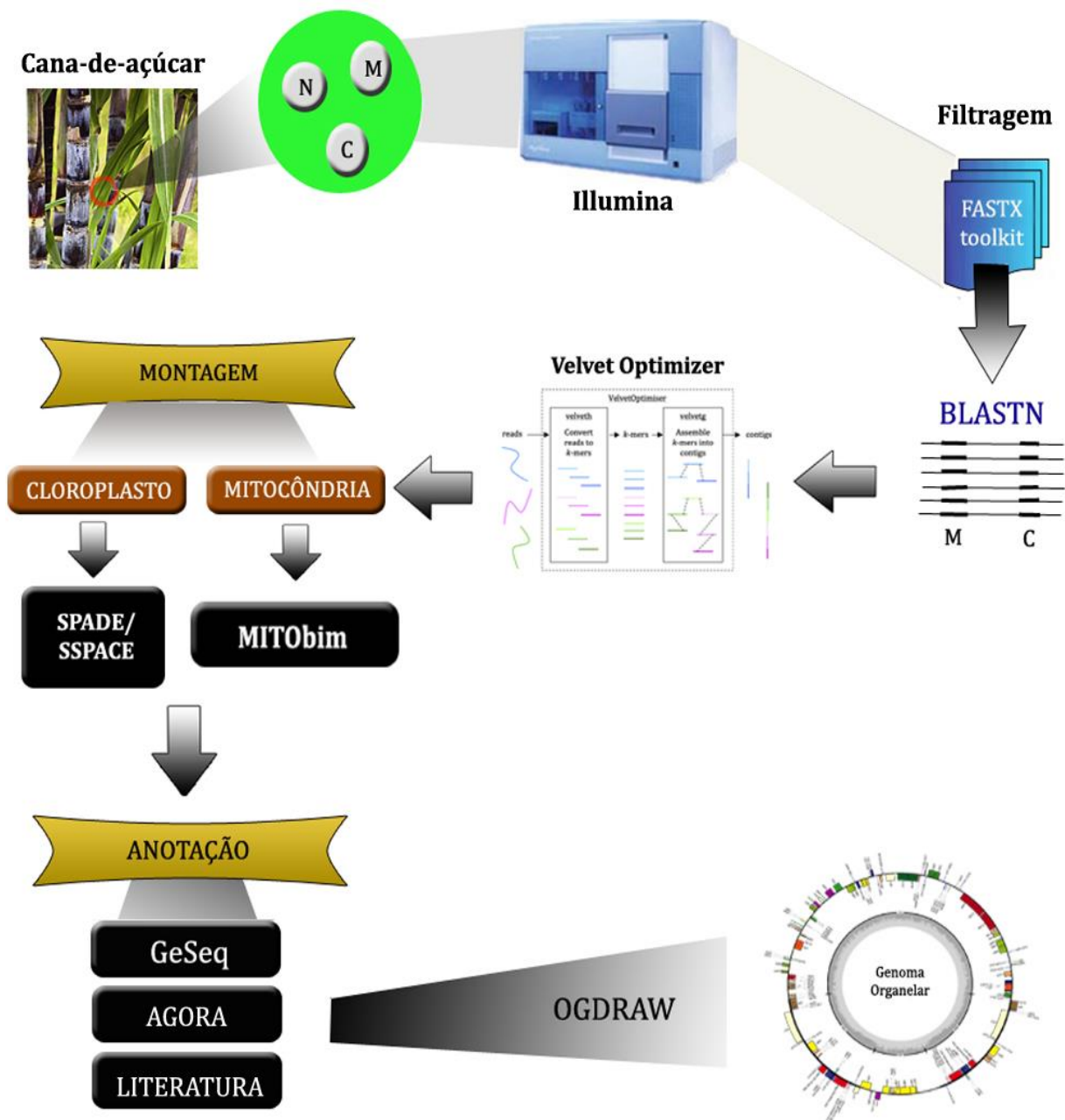


Figura 7- Esquema de estratégias metodológicas de montagem e anotação do genoma organelar da cana-de-açúcar. A célula vegetal está representada em verde contendo o DNA total, sendo: (N) para DNA nuclear, (M) para DNA mitocondrial e (C) para DNA cloroplastidial.

3.4. Comparação do genoma cloroplastidial e mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar

3.4.1 Análise de divergência de sequência

Para analisar a divergência entre as sequências de cultivares híbridos e espécies selvagens de cana-de-açúcar, foram selecionados sete genomas de cana-de-açúcar, sendo cinco cultivares híbridos e duas espécies selvagens e comparados com o software mVISTA (FRAZER et al., 2004), utilizando como referência *Saccharum officinarum* cultivar IJ76-514 (EVANS; JOSHI, 2016) e parâmetro LAGAN(alinhamento múltiplo), para detecção de éxons utilizamos a anotação de cada cultivar.

3.4.2 Caracterização de sequências repetidas

Para calcular os tipos de repetições ao longo do genoma de cloroplasto das espécies selvagens e cultivares híbridos de cana-de-açúcar, foram utilizados três métodos computacionais, sendo: o software Tandem Repeats Finder (BENSON, 1999) para calcular as sequências repetidas em tandem (>10 pares de bases); o software REPuter que foi utilizado para visualizar repetições palindrômicas e dispersas pelo genoma com um tamanho mínimo de repetição de 30 pares de bases, distâncias de edição menores que três pares de bases e duas cópias de repetição com pelo menos 90% de similaridade (KURTZ et al., 2001); e o perl script MISA(BEIER et al., 2017) para calcularmos as repetições de sequências simples (SSRs) seguindo os seguintes parâmetros (1-6 nucleotídeos e limiares de dez, cinco, quatro, três, três e três unidades de repetição para mono-, di-, tri-, tetra-, penta- e hexanucleotídeos SSRs), segundo metodologia descrita por (WU et al., 2018).

3.4.3 Detecção de Indels e sua localização no genoma

Os indels foram detectados utilizando-se o software Tandem Repeats Finder (BENSON, 1999), utilizando os parâmetros 2, 7, 7 para correspondência, incompatibilidade e indels, respectivamente. Para reportar repetição, pontuação mínima de 50 em um tamanho máximo de 500, conforme configuração padrão.

3.4.4 Análise das regiões IRs de cloroplastos

A análise dos genes das regiões de fronteira IRs em cloroplastos foi realizada através de um alinhamento múltiplo e observação dos tamanhos dos genes em pares de bases, onde também se observou a distância relativa dos genes nas fronteiras. Para isso, foram utilizados os arquivos genbank gerados a partir da anotação dos cultivares híbridos e das espécies selvagens de cana-de-açúcar.

4. RESULTADOS

4.1. Montagem de genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar

4.1.1. Genoma cloroplastidial

A montagem *de novo* realizada para o genoma cloroplastidial revelou a porcentagem de GC do genoma de cada cultivar, o número de contigs necessários para cobertura de todo genoma, o maior contig em pares de bases dentre aqueles que cobriram o genoma de cada cultivar e o tamanho total em pares de bases de cada genoma de cloroplasto montado (**Tabela1**).

Tabela 1- Parâmetros de montagem do genoma de cloroplasto de espécies selvagens e cultivares híbridos de cana-de-açúcar

Montagem de Cloroplastos							
Espécies selvagens e cultivares	<i>S. officinarum</i>	<i>S. spontaneum</i>	RB72454	RB867515	Tambo	Roxa	B4362
GC %	37,41	38,43	37,32	38,44	37,50	37,42	37,51
Nº de contigs	5	2	4	1	12	6	9
Maior contig (pb)	106.869	97.413	47.009	141.181	40.076	54.007	35.593
Tamanho total (pb)	142.234	141.162	141.151	141.181	140.470	141.721	141.875

Neste trabalho também montamos *S.spontaneum* (sob acesso: LN896360), RB72454 (EVANS; JOSHI, 2016b) e RB867515(VIDIGAL et al., 2016) atualmente essas sequências já encontram-se publicadas por outros autores. Contudo, utilizamos essas montagens para continuar nossas análises, dando os devidos créditos aos referidos autores.

A espécie selvagem *S. spontaneum* e o cultivar RB867515 apresentaram uma quantidade de GC similar, sendo a maior porcentagem dentre os demais cloroplastos. Um grande número de *contigs* é um indicador de fragmentação no genoma, dessa forma o cultivar Tambo se apresentou mais fragmentado, possuindo 12 *contigs*. Já o cultivar RB867515 apresentou em sua montagem um único *contig*, mostrando uma cobertura integral do genoma. Em relação ao tamanho total do genoma de cloroplasto em pares de base, observam-se tamanhos relativamente próximos indo de 140.470 pb no cultivar tambo à 142.234 na espécie selvagem *Saccharum officinarum*.

4.1.2. Genoma mitocondrial

Na montagem referenciada do genoma mitocondrial os mesmos parâmetros de montagem do genoma de cloroplasto do item 4.1 também foram obtidos (**Tabela 2**).

Tabela 2- Parâmetros de montagem do genoma de mitocôndrias de espécies selvagens e cultivares híbridos de cana-de-açúcar

Montagem de Mitocôndrias							
Espécies selvagens e cultivares híbridos	<i>S. officinarum</i>	<i>S. spontaneum</i>	RB72454	RB867515	Tambo	Roxa	B4362
GC%	43,36	43,64	43,76	43,68	43,73	43,78	43,74
Nº de contigs	2	2	2	2	2	2	2
Maior contig (pb)	305.257	300.855	302.393	302.639	303.667	301.956	302.038
Tamanho total (pb)	453.895	447.331	449.603	448.926	449.366	447.925	447.298

A porcentagem de GC no genoma mitocondrial entre as espécies selvagens e os cultivares híbridos apresentou menos variação do que no genoma cloroplastidial, todos

ficaram na faixa dos 43%, sugerindo uma maior conservação deste parâmetro nas mitocôndrias. Não houve variação no número de contigs, por ser uma montagem referenciada dos dois cromossomos 1 e 2 de cana-de-açúcar, como já demonstrado por Shearman et al (2016). *S. officinarum* apresentou o maior contig, ou seja, o cromossomo 1. Apresentou também maior tamanho total em pares de base com 453.895. O menor tamanho total em pares de bases foi da cultivar B4362 com 447.298.

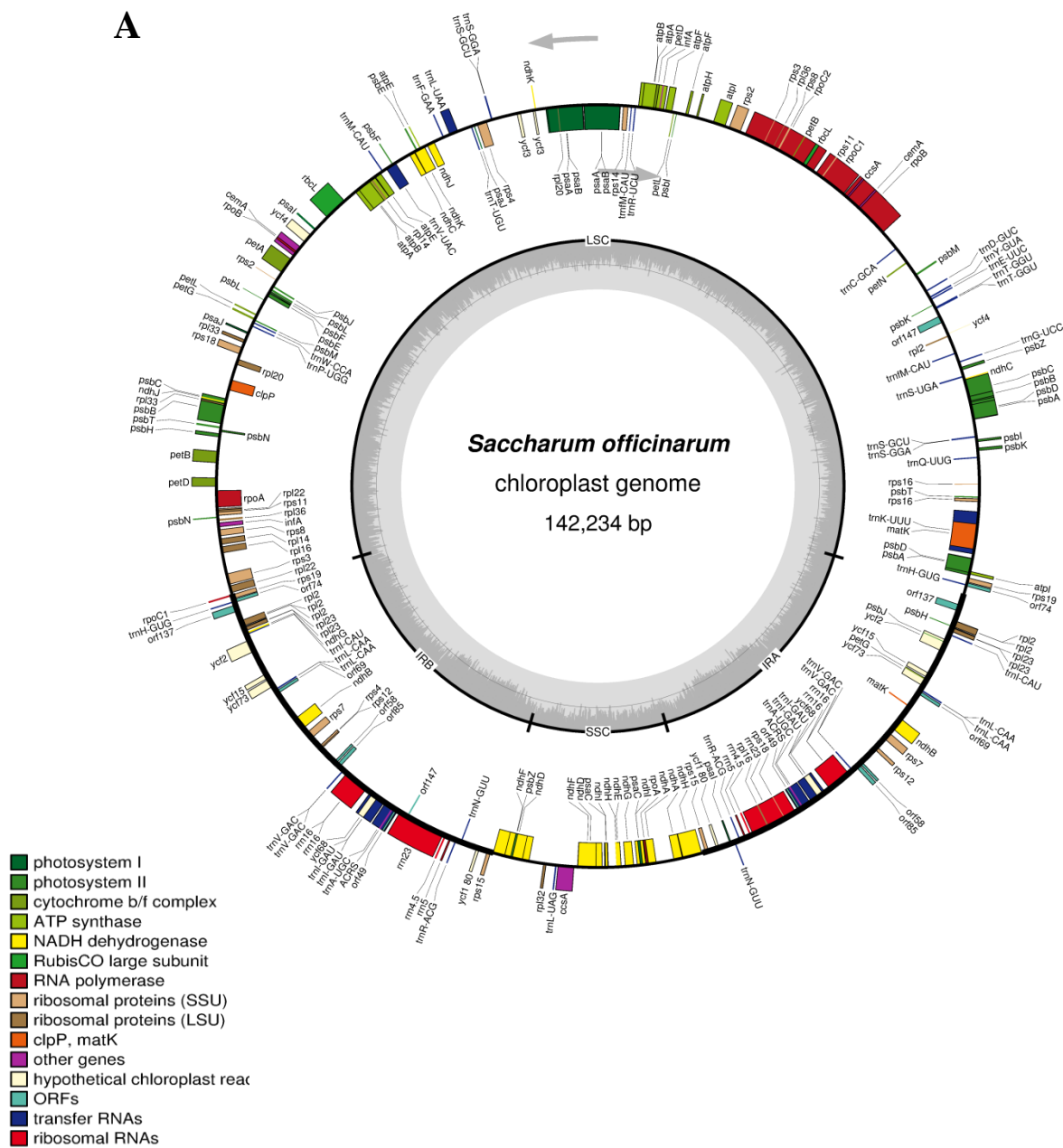
4.2. Anotação de genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar

4.2.1. Genoma cloroplastidial

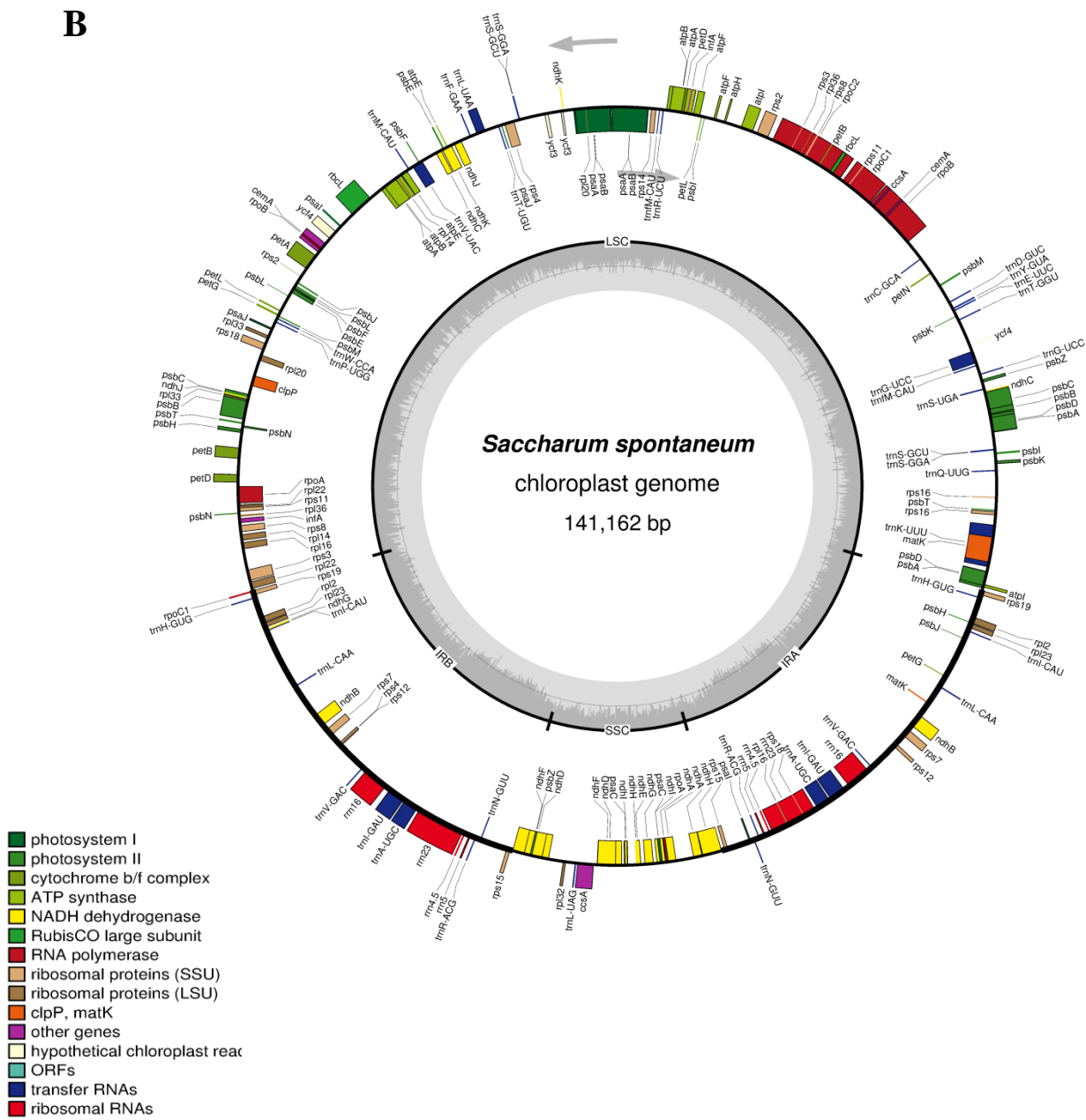
A anotação de todos os genomas cloroplastidiais apresentaram estrutura contendo quatro partes, sendo uma região cópia única grande chamada de LSC, uma pequena região também de cópia única (SSC) e duas regiões IRs (IRa e IRb). Os números de genes que serão mencionados para cada espécie selvagem e para cada cultivar híbrido só foram registrados uma única vez, sem contar os genes que se repetem ao longo do genoma (**Figura 8**).

A espécie selvagem *Saccharum officinarum* (**Figura 8A**) possui 156 genes, sendo 109 sequências codificantes de proteínas (CDS), 8 RNAs ribossomais (RNAr) e 39 RNAs transportadores (RNAt). A espécie selvagem *Saccharum spontaneum* (**Figura 8B**) possui 129 genes, sendo 82 CDS, 8 RNAr e 39 RNAt. No genoma da cultivar RB72454 (**Figura 8C**) identificamos 129 genes com 82 CDS, 8 RNAr e 39 RNAt. A cultivar RB867515 (**Figura 8D**) possui um genoma de cloroplasto com 135 genes, dentre eles 88 de CDs, 8 de RNAr e 39 RNAt. Para a cultivar Tambo FEPAGRO (**Figura 8E**) detectamos 129 genes: 82 CDS, 8 RNAr e 39 RNAt. Foi encontrado no genoma de cloroplasto da cultivar Roxa (**Figura 8F**) um total de 129 genes e um tamanho de 141.721 pares de base. Dos genes encontrados 82 são CDS, 8 são RNAr e 39 são RNAt. Na cultivar B4362 (**Figura 8G**) encontramos 129 genes, onde 82 são CDS, 8 RNAr e 39 de RNAt. Esses dados são comparados na **tabela 3**.

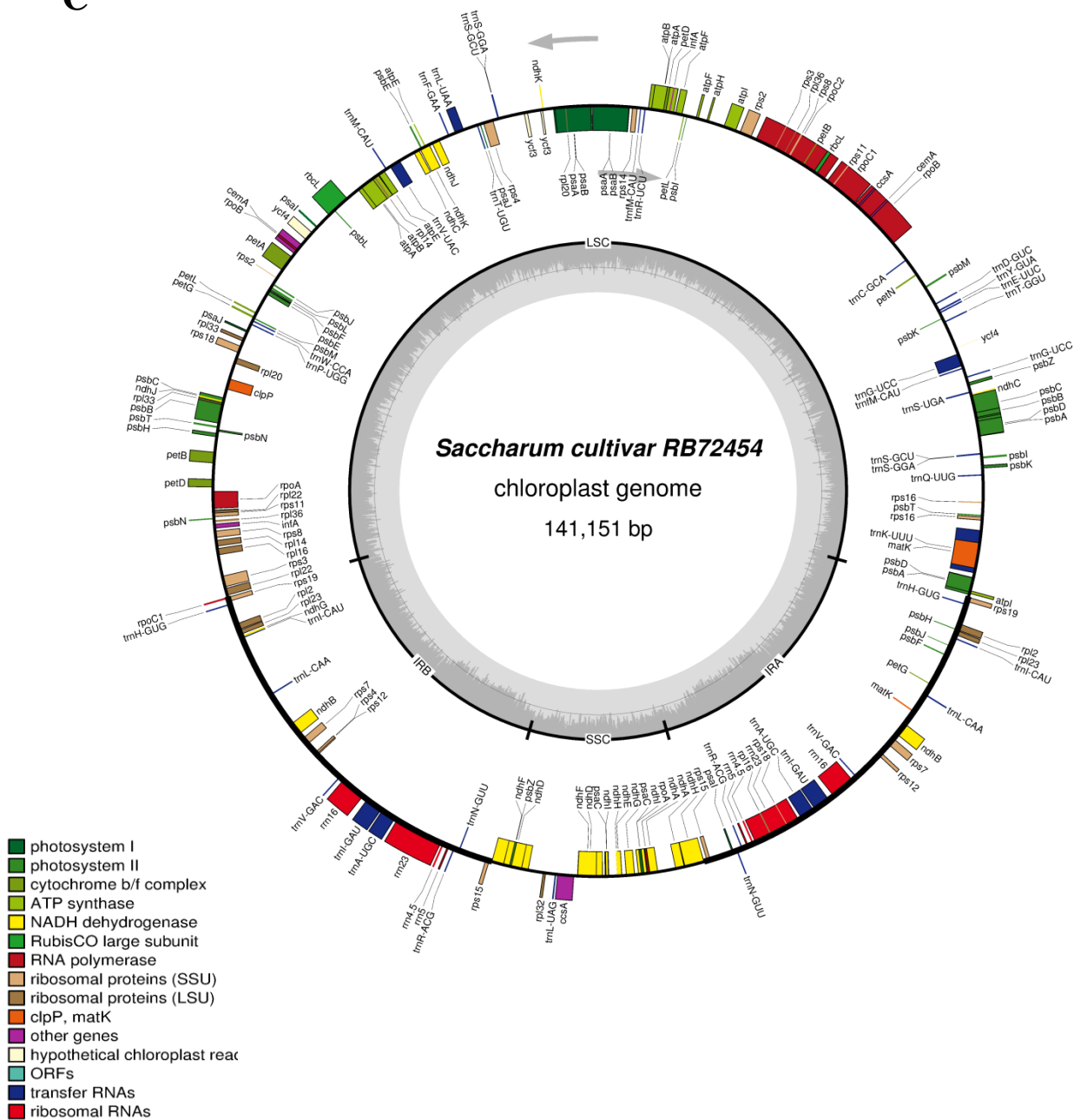
A



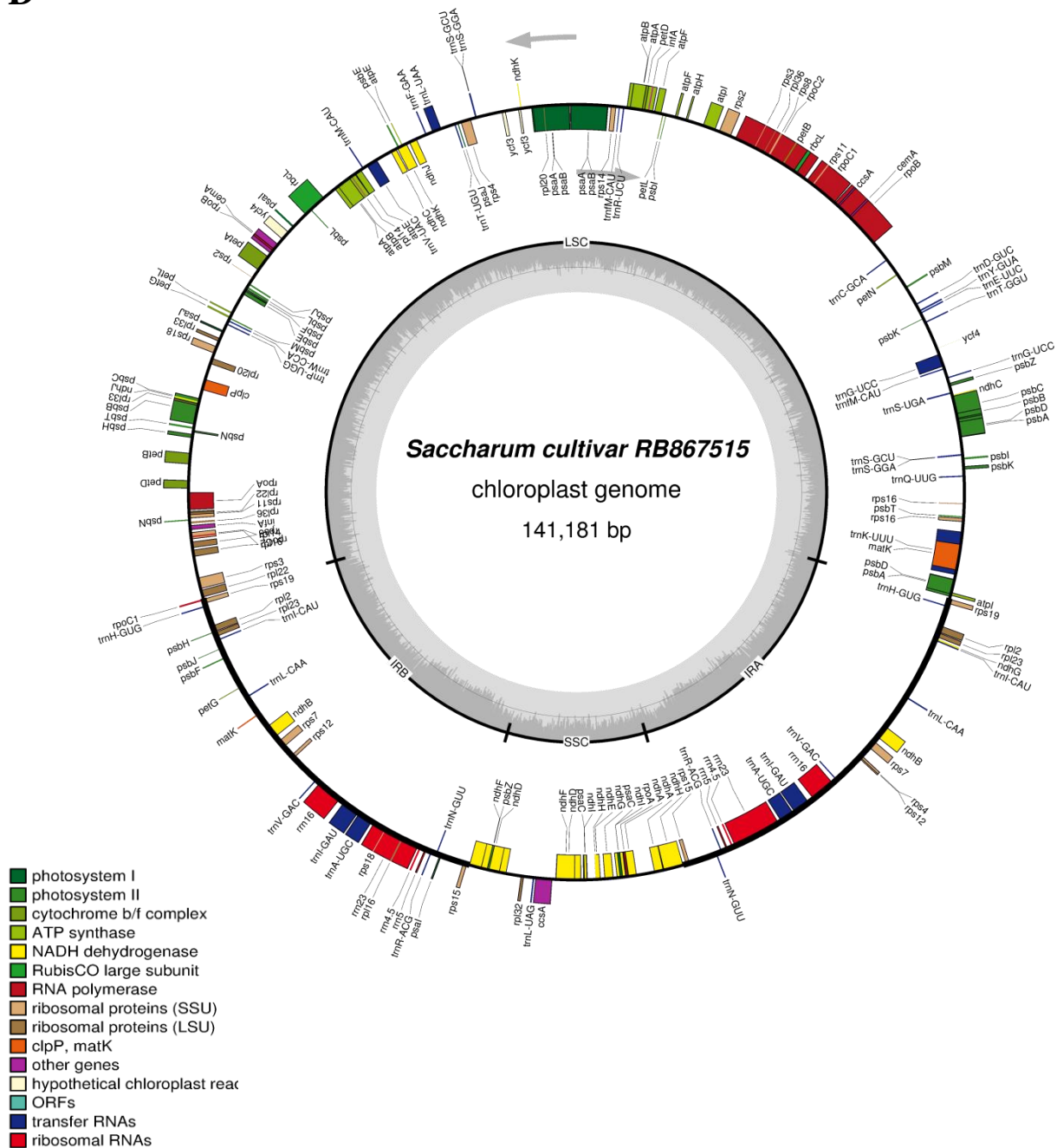
B



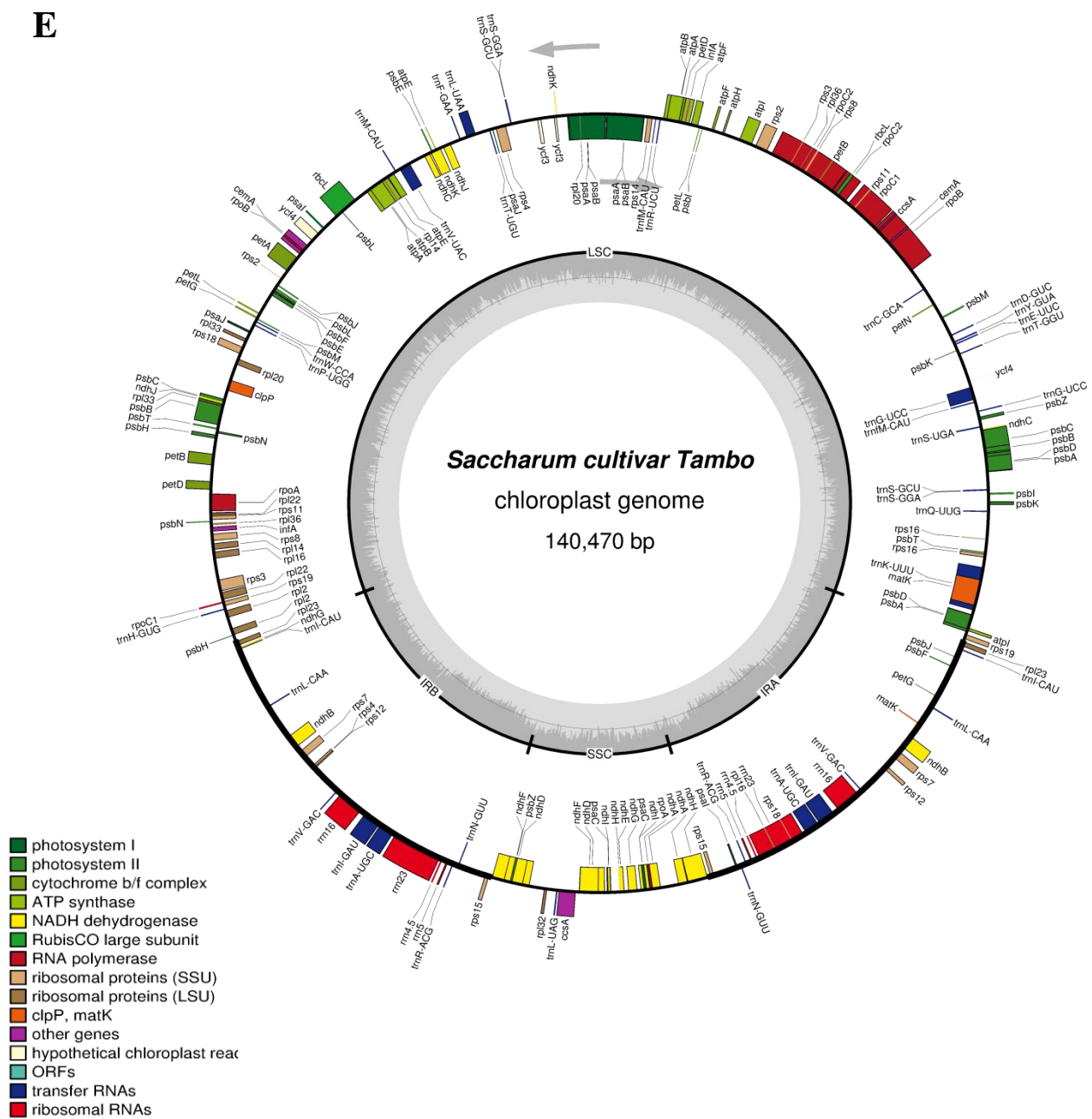
C



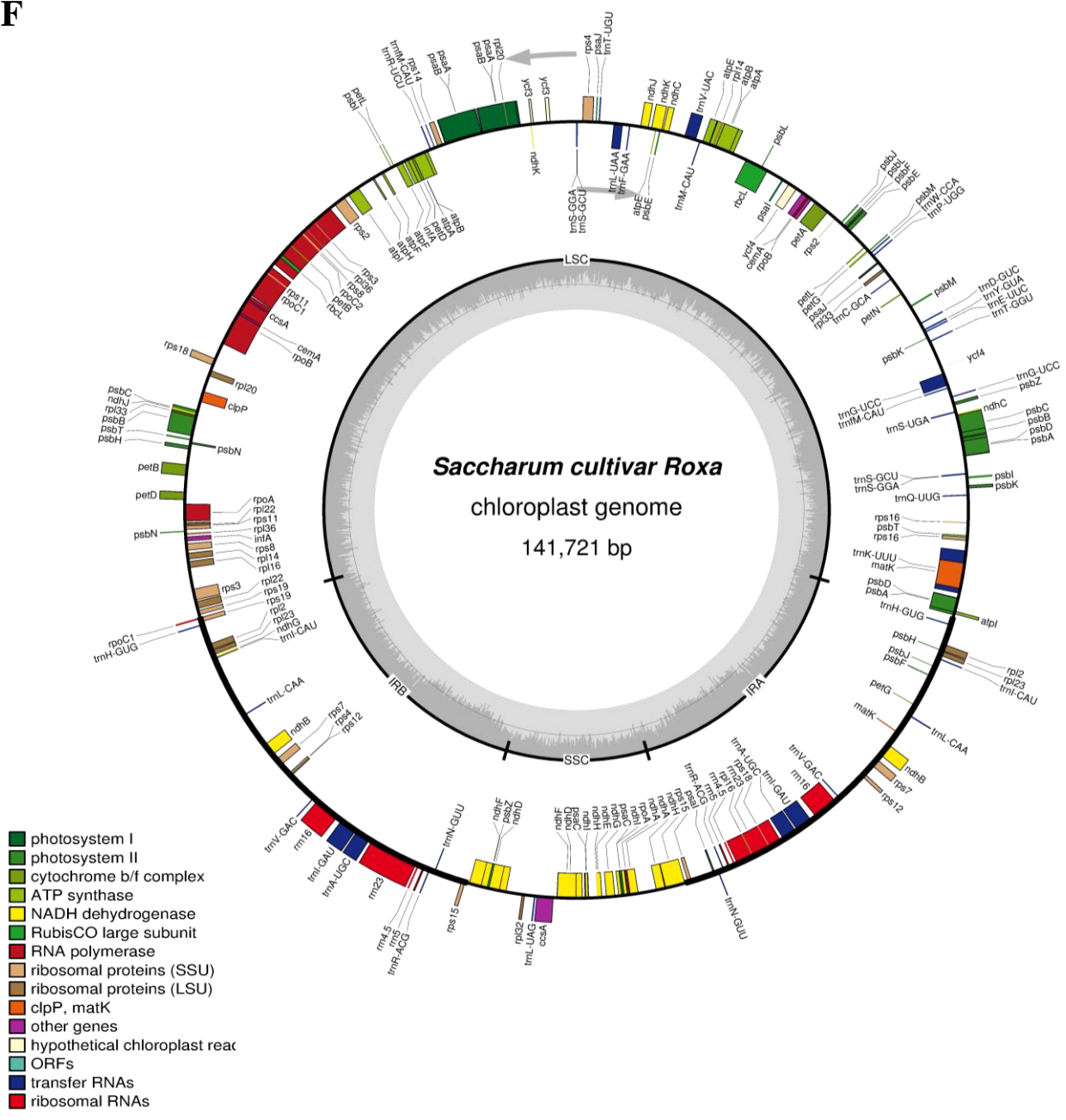
D



E



F



G

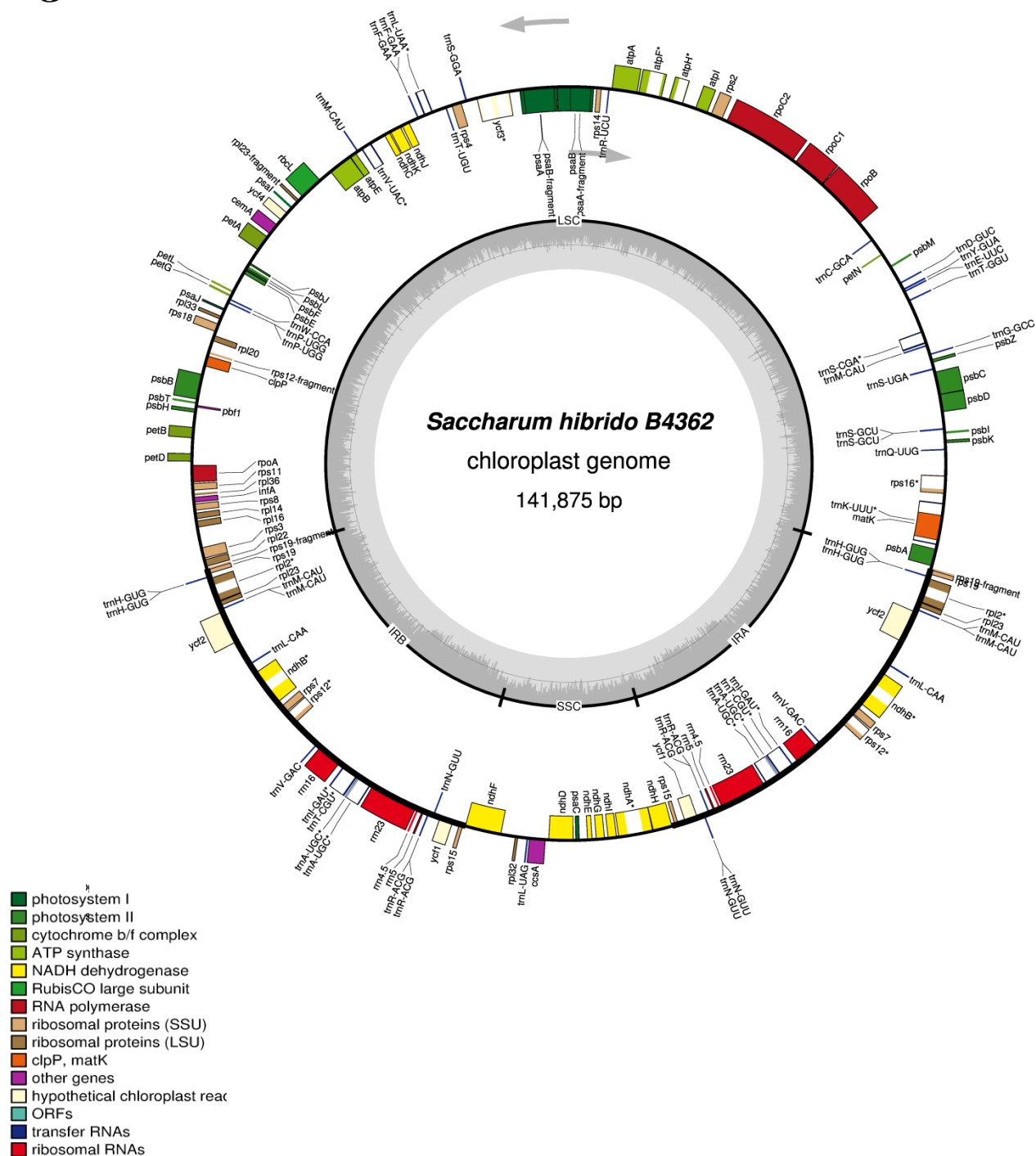


Figura 8- Mapa circular do genoma de cloroplasto de cana-de-açúcar. As setas indicam o sentido em que a transcrição ocorre (horário e anti-horário). O círculo interno mostra o conteúdo de GC no genoma e as linhas pretas no círculo externo demarcam as regiões IRs. As letras de A à G representam o genoma de cada espécie *Saccharum*. (A) *S.officinarium*, (B) *S.spontaneum*, (C) RB72454, (D) RB867515, (E) Tambo FEPAGRO, (F) Roxa, (G) B4362.

Tabela 3- Comparação da anotação do genoma cloroplastidial de espécies selvagens e cultivares híbridos de cana-de-açúcar

Comparação das anotações							
Espécies selvagens e cultivares	<i>S. officinarum</i>	<i>S. spontaneum</i>	RB72454	RB867515	Tambo	Roxa	B4362
CDS	109	82	82	88	82	82	82
RNA _t	39	39	39	39	39	39	39
RNA _r	8	8	8	8	8	8	8
Total de genes	156	129	129	135	129	129	129

As mitocôndrias apresentaram dois cromossomos, sendo o cromossomo 1 de maior tamanho e cromossomo 2 de menor tamanho. Estes cromossomos apresentam genes relacionados à via oxidativa (complexo I, III e IV), ATP sintase, proteínas ribossomais, maturases, RNAs transportadores e RNAs ribossomais, também possuem ORFs e introns. O quantitativo dos genes que serão descritos para cada cultivar abaixo, só foram contados uma única vez, desprezando a repetição do gene nos dois cromossomos (**Figura 9**).

O cromossomo 1 da espécie selvagem *Saccharum officinarum* apresentou 20 CDS, 25 RNA_t e 3 RNA_r; o cromossomo 2 possui 15 CDS e 6 RNA_t, o que gera um total de 69 genes no genoma desta mitocôndria e um tamanho total de 453.895 pares de base (**Figura 9A**).

O genoma mitocondrial da espécie selvagem *Saccharum spontaneum* apresentou 20 CDS, 21 RNA_t e 3 RNA_r. Já para o segundo cromossomo foi encontrado 15 CDS e 5 RNA_t, totalizando um tamanho total de 447.311 pares de base e 64 genes (**Figura 9B**).

Detectamos no cultivar híbrido RB72454 para o cromossomo 1 20 CDS, 17 RNA_t e 3 RNA_r e para o cromossomo 2 15 CDS e 5 RNA_t. Seu tamanho total chegou a 449.603 pares de bases contendo 60 genes (**Figura 9C**).

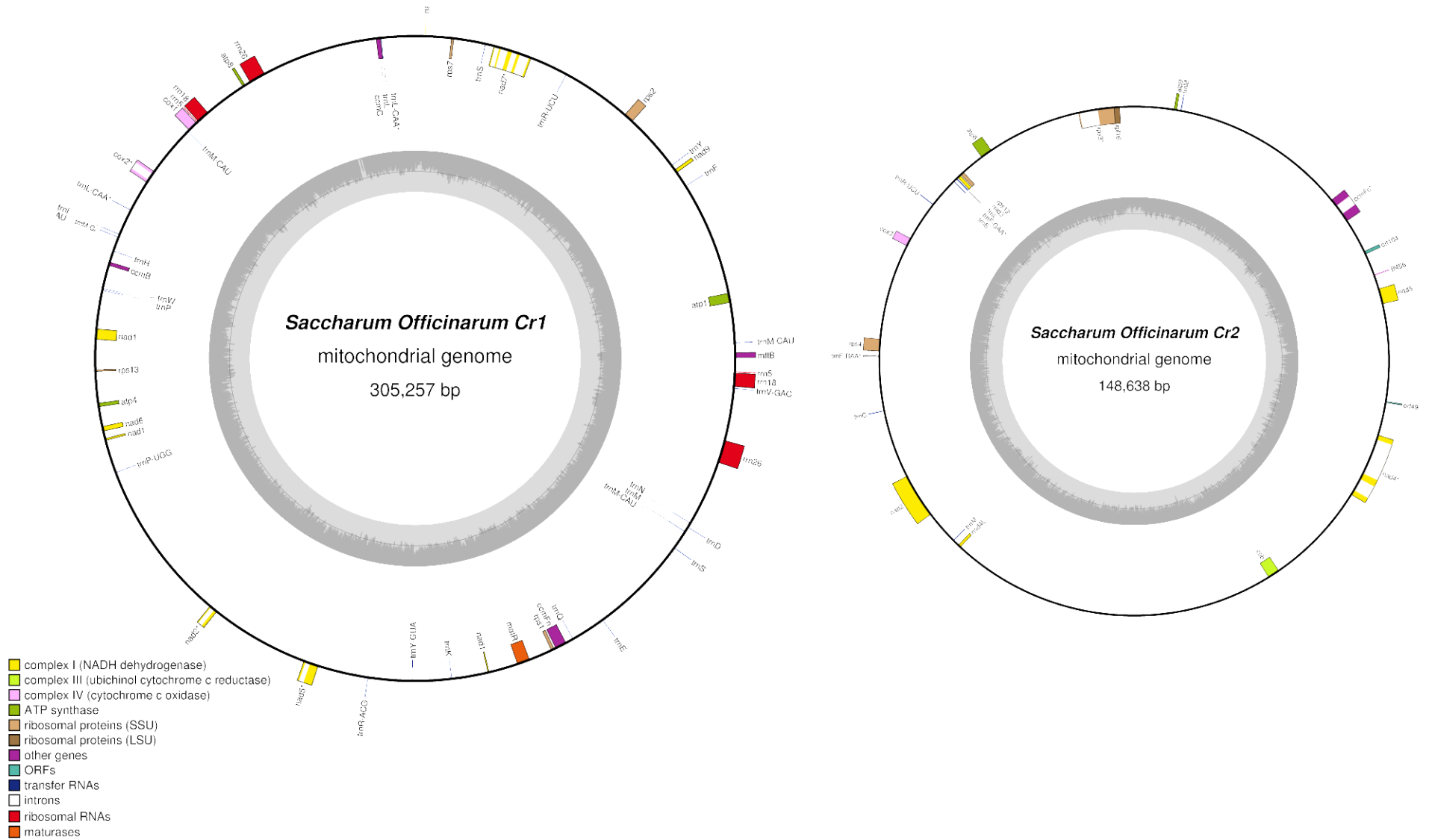
O cromossomo 1 da cultivar RB867515 possui 20 CDS, 20 RNAt e 3 RNAr. O cromossomo 2 possui 15 CDS e 4RNAt. Este genoma apresentou um total de 62 genes e um tamanho total de 448.926 pares de base (**Figura 9D**).

Na cultivar Tambo FEPAGRO para o cromossomo 1 detectamos 20 CDS, 20 RNAt e 3 RNAr. No cromossomo 2 foram 15 CDS e 5 RNAt. Anotamos então para este genoma 63 genes e um tamanho total de 449.366 pares de base (**Figura 9E**).

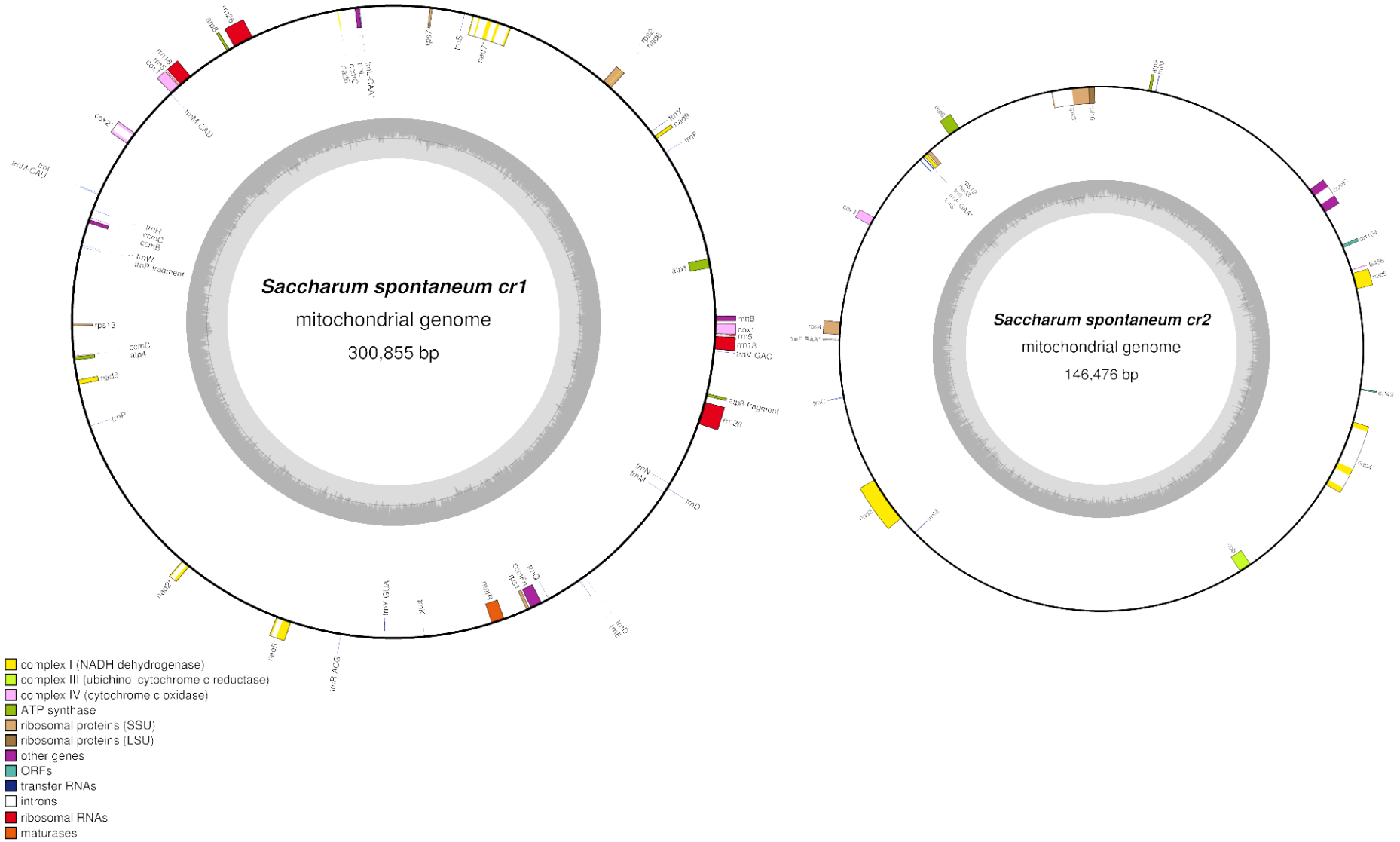
Para a cultivar Roxa anotamos um total de 62 genes, sendo para o cromossomo 1: 20 CDS, 19 RNAt e 3 RNAr, para o cromossomo 2: 15 CDS e 5 RNAt. Seu tamanho total ficou em 447.925 pares de base (**Figura 9F**).

A cultivar B436269 foi anotada com 63 genes, sendo para o primeiro cromossomo 20 CDS, 20 RNAt e 3RNAr. Para o segundo cromossomo foram 15 CDS e 5 RNAt. Em relação ao tamanho, o total foi de 447.298 pares de bases (**Figura 9G**). Esses genes encontrados no cromossomo 1 e no cromossomo 2 no genoma mitocondrial das espécies selvagens e cultivares híbridos de cana-de-açúcar foram comparados conforme **tabela 4**.

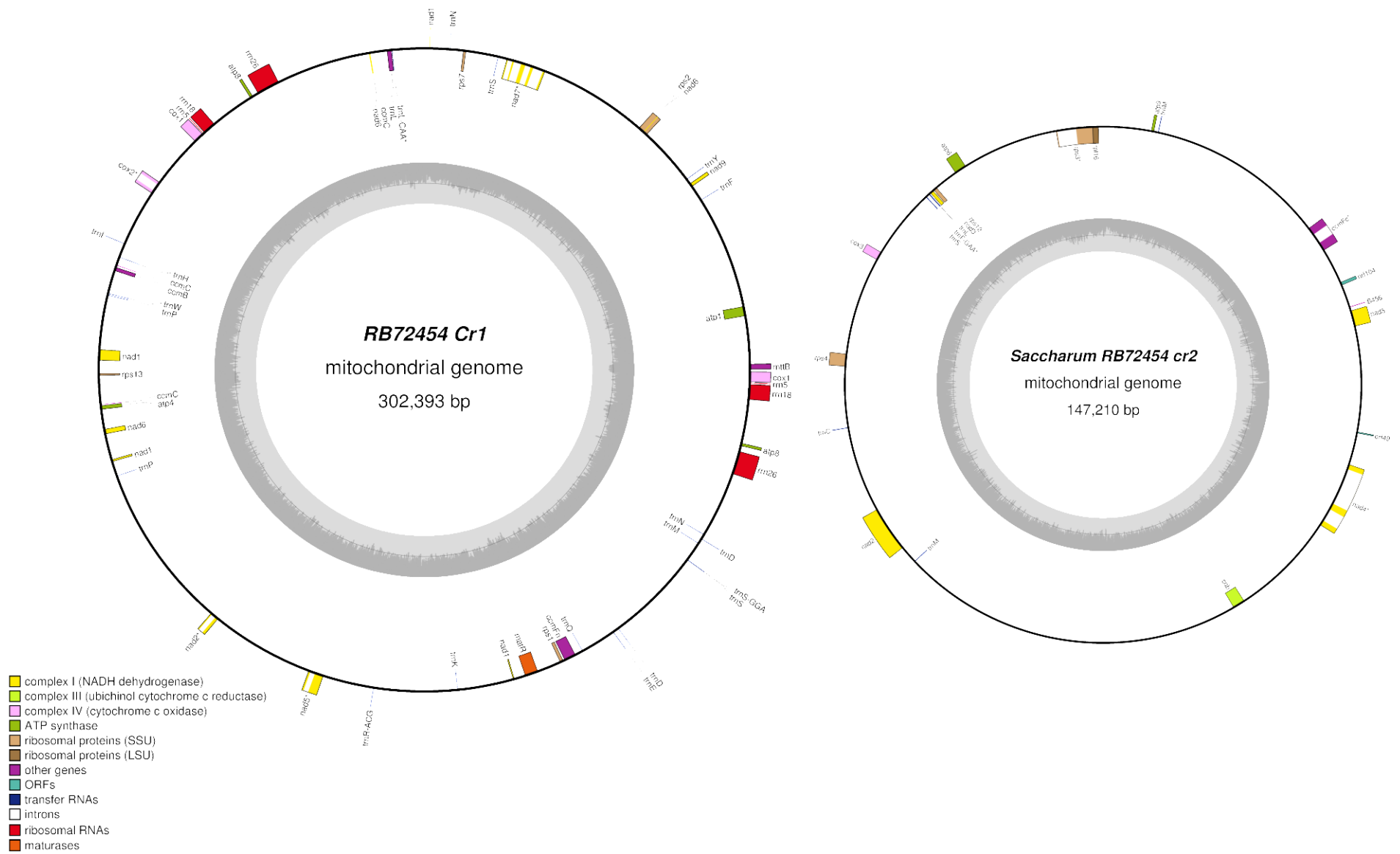
A



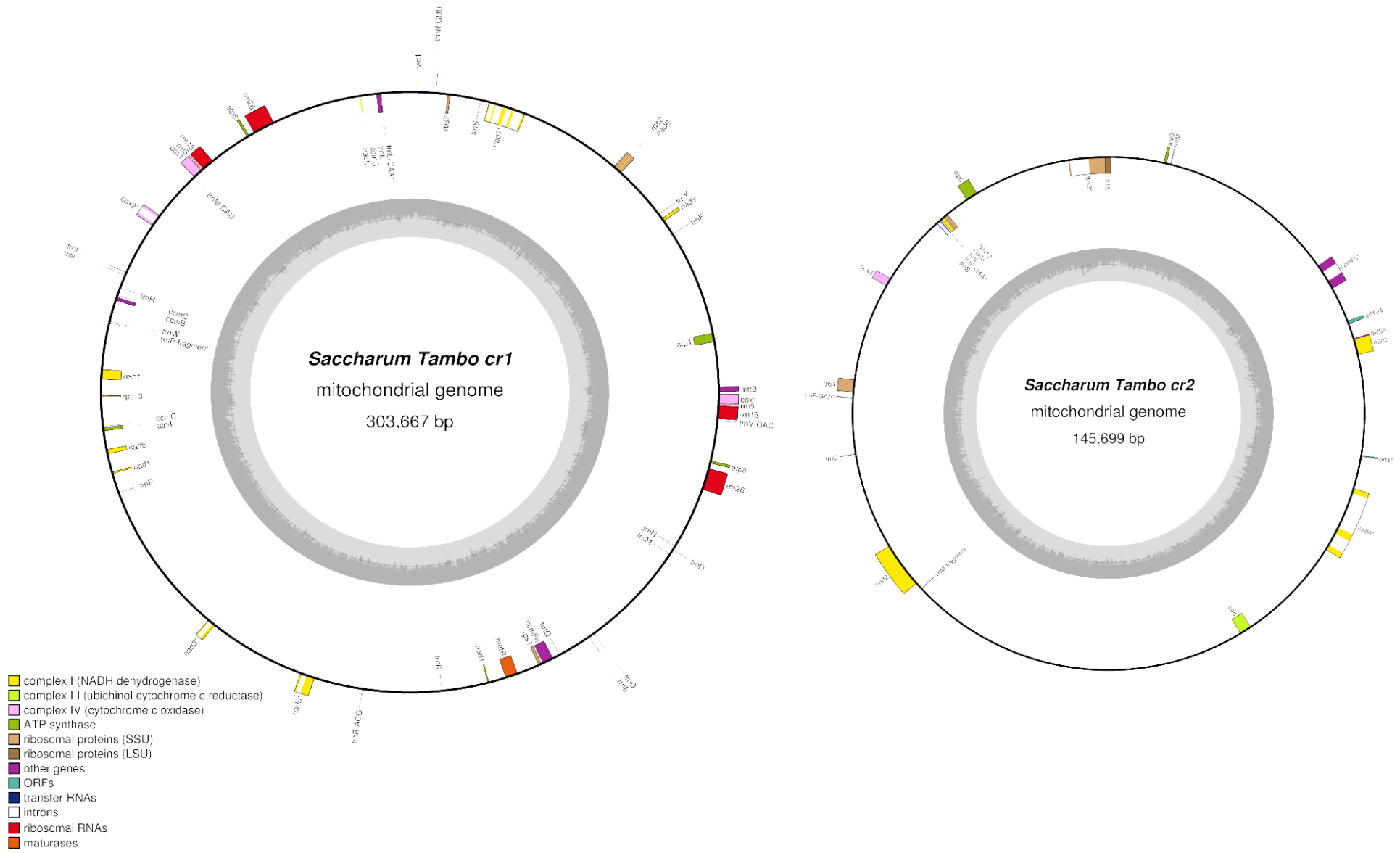
B



C



E



F



G

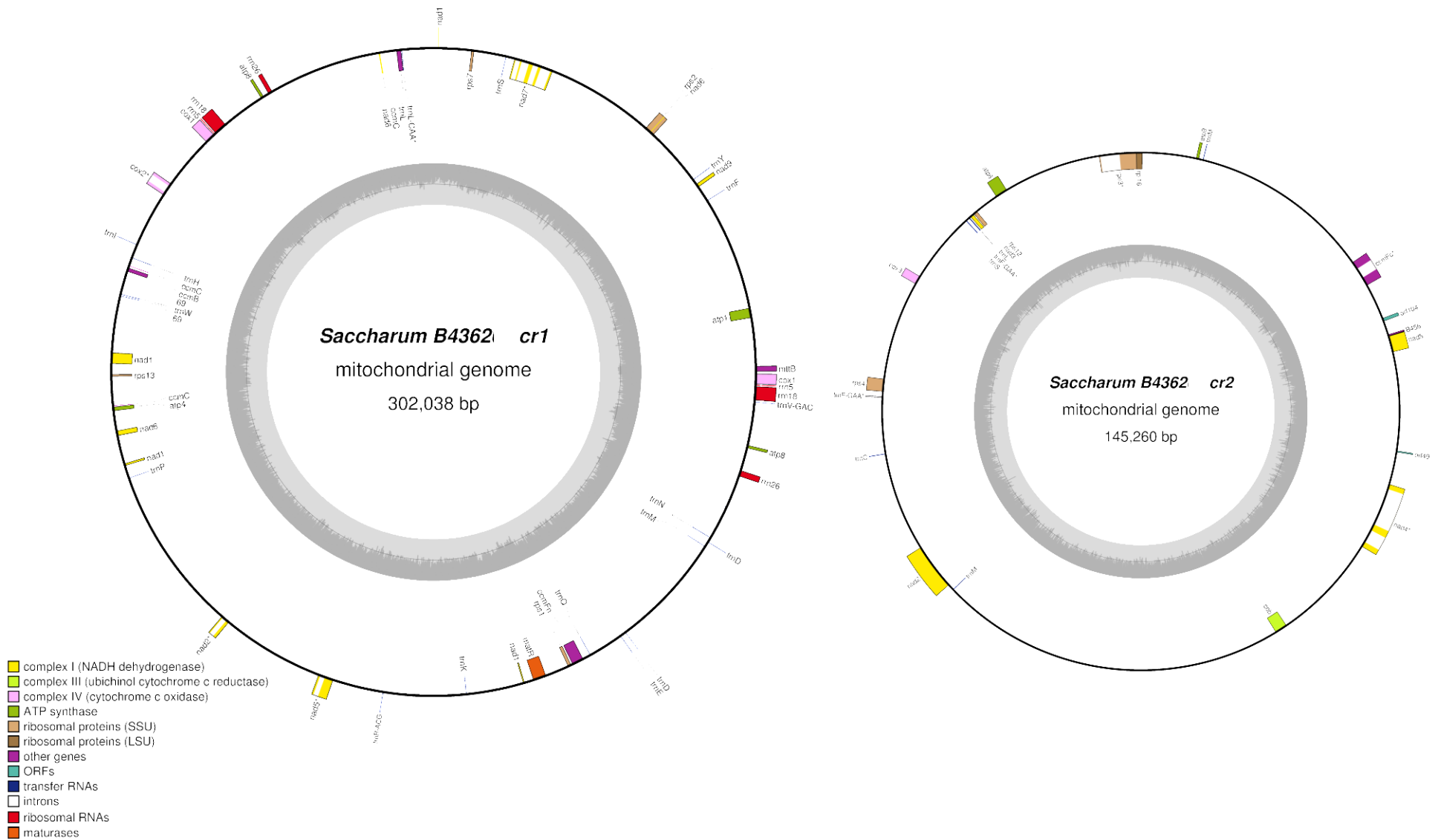


Figura 9- Cromossomos do genoma mitocondrial de cana-de-açúcar . Os genes do lado externo dos cromossomos 1 e 2 estão na cadeia positiva, os genes mostrados dentro do círculo estão localizados na cadeia negativa. O círculo interno de cor cinza em cada cromossomo representa o conteúdo de GC e os genes representados são classificados na legenda em cores. As letras de A à G representam o genoma mitocondrial de cada espécie *Saccharum*. (A) *S.officinorum*, (B) *S.spontaneum*, (C) RB72454, (D) RB867515, (E) Tambo FEPAGRO, (F) Roxa, (G) B4362.

Tabela 4- Comparação da anotação do genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar

Comparação das anotações Cr1							
Espécies selvagens e cultivares	<i>S. officinarum</i>	<i>S. spontaneum</i>	RB72454	RB867515	Tambo	Roxa	B4362
CDS	20	20	20	20	20	20	20
RNA_t	25	21	17	20	20	19	20
RNA_r	3	3	3	3	3	3	3
Cr2							
CDS	15	15	15	15	15	15	15
RNA_t	6	5	5	4	5	5	5
Total de genes	69	64	60	62	63	62	63

4.3. Análise de divergência de sequência no genoma organelar de espécies selvagens e cultivares híbridos de cana-de-açúcar

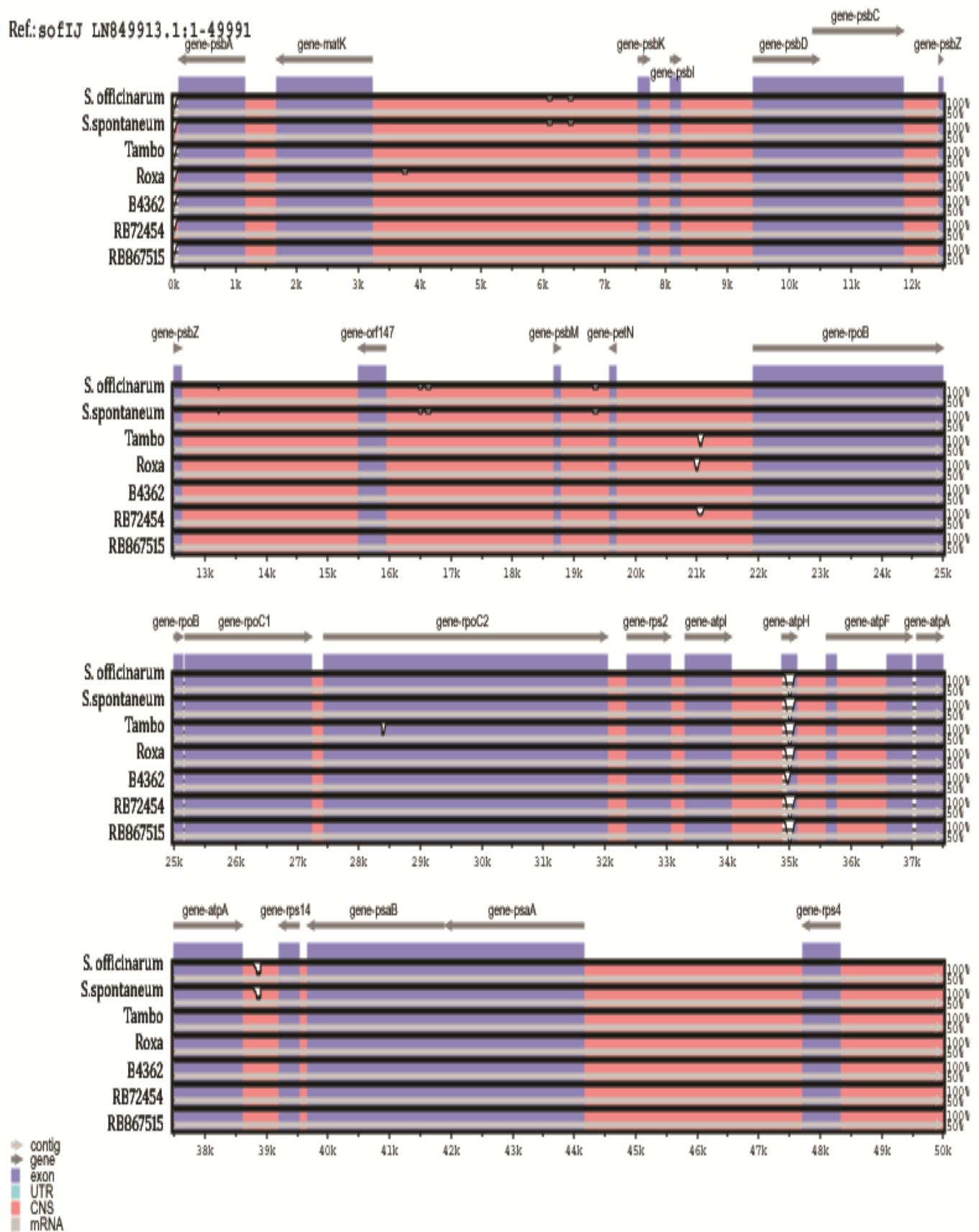
4.3.1 Cloroplastos

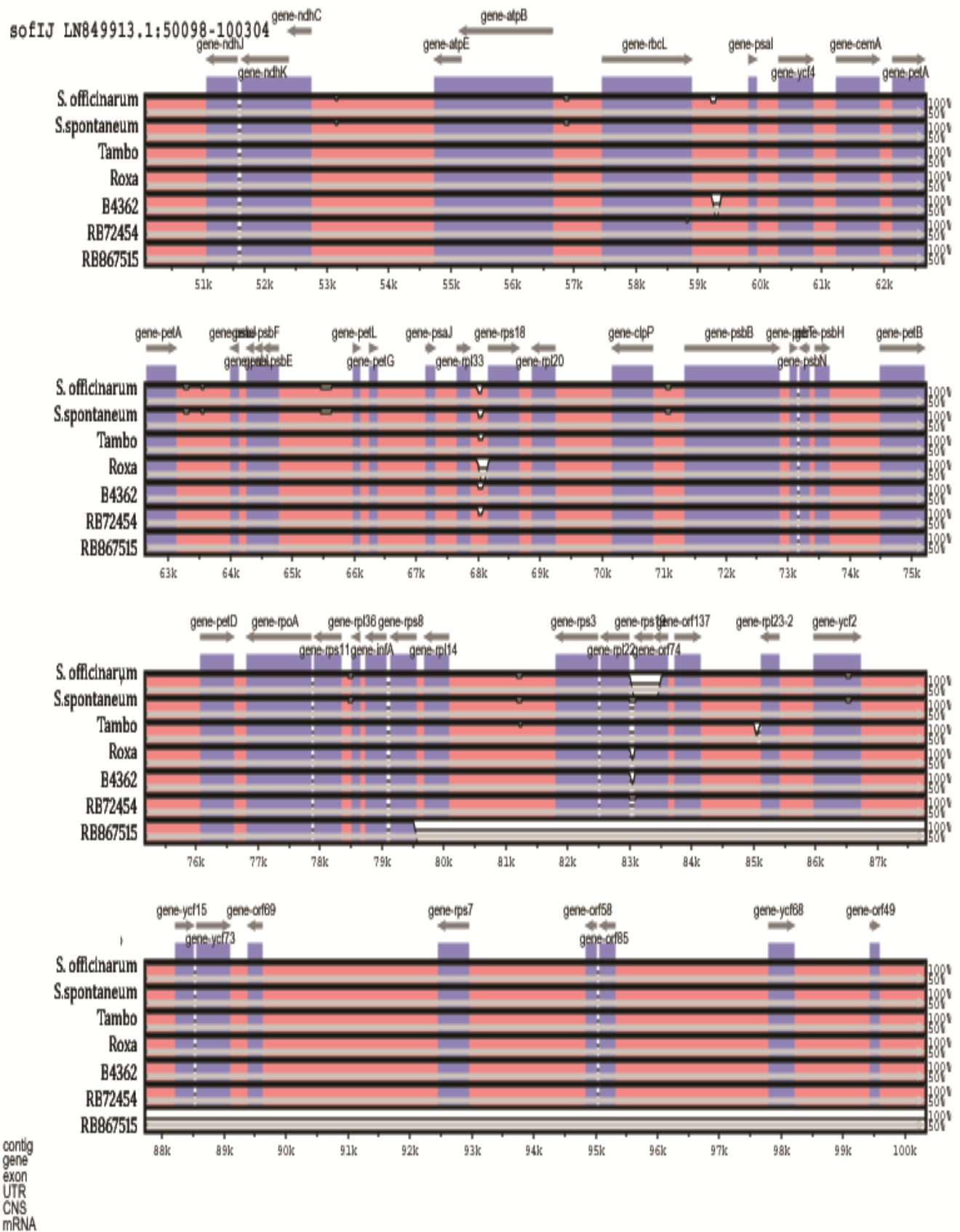
Para analisar a divergência entre os genomas de cloroplastos utilizamos a montagem realizada neste estudo para *S.officinarum*, tambo, roxa e B4362; e para *S.spontaneum*, RB72454 e RB867515 utilizamos a montagem previamente realizada. Os picos que representam divergência entre as sequências de genomas de cloroplastos foram mais acentuados nas regiões não codificantes que aparecem entre os genes. Apesar desses genomas se apresentarem bem conservados, as cultivares Roxa, B4362 e RB867515 possuem grandes regiões com identidades menores que 70%, incluindo as regiões com éxons, comparando com a referência Sof IJ76-514. Os selvagens *S. officinarum* e *S. spontaneum* apresentaram discretas similaridades, um pequeno pico de divergência entre os genes *matK* e *psbL*, na região do gene *ycf2* e do gene *ccsA*, o que sugere que estes genes não são tão conservados nesses selvagens. Apresentaram também divergência na região intergênica entre os genes *atpA* e *rps14*. Tambo e RB72454 também apresentaram um pico de divergência na região entre os genes *petN* e *rpoB*. A cultivar Tambo parece ter uma menor conservação no gene *rpoC2* do que os outros cultivares (**Figura 10**).

4.3.2 Mitocôndrias

A análise de divergência do cromossomo 1 dos genomas mitocondriais mostrou maior divergência na região não-codificante. *S. officinarum* e *S.spontaneum* parecem seguir os mesmos padrões de divergência e conservação. Ambos possuem o gene *rps2* menos conservado comparado com os demais cultivares. O cromossomo 1 das mitocôndrias dos cultivares híbridos são bem similares entre si (**Figura 11**).

O cromossomo 2 também possui alta conservação, com menor identidade de sequências nas regiões intergênicas, embora ocorram pequenos picos de divergência em alguns éxons. Os cultivares híbridos apresentaram uma maior identidade de forma geral com a cultivar Khon Kaen 3 utilizada como referência e *S. officinarum* e *S. spontaneum* apresentaram maior similaridade entre si (**Figura 12**).





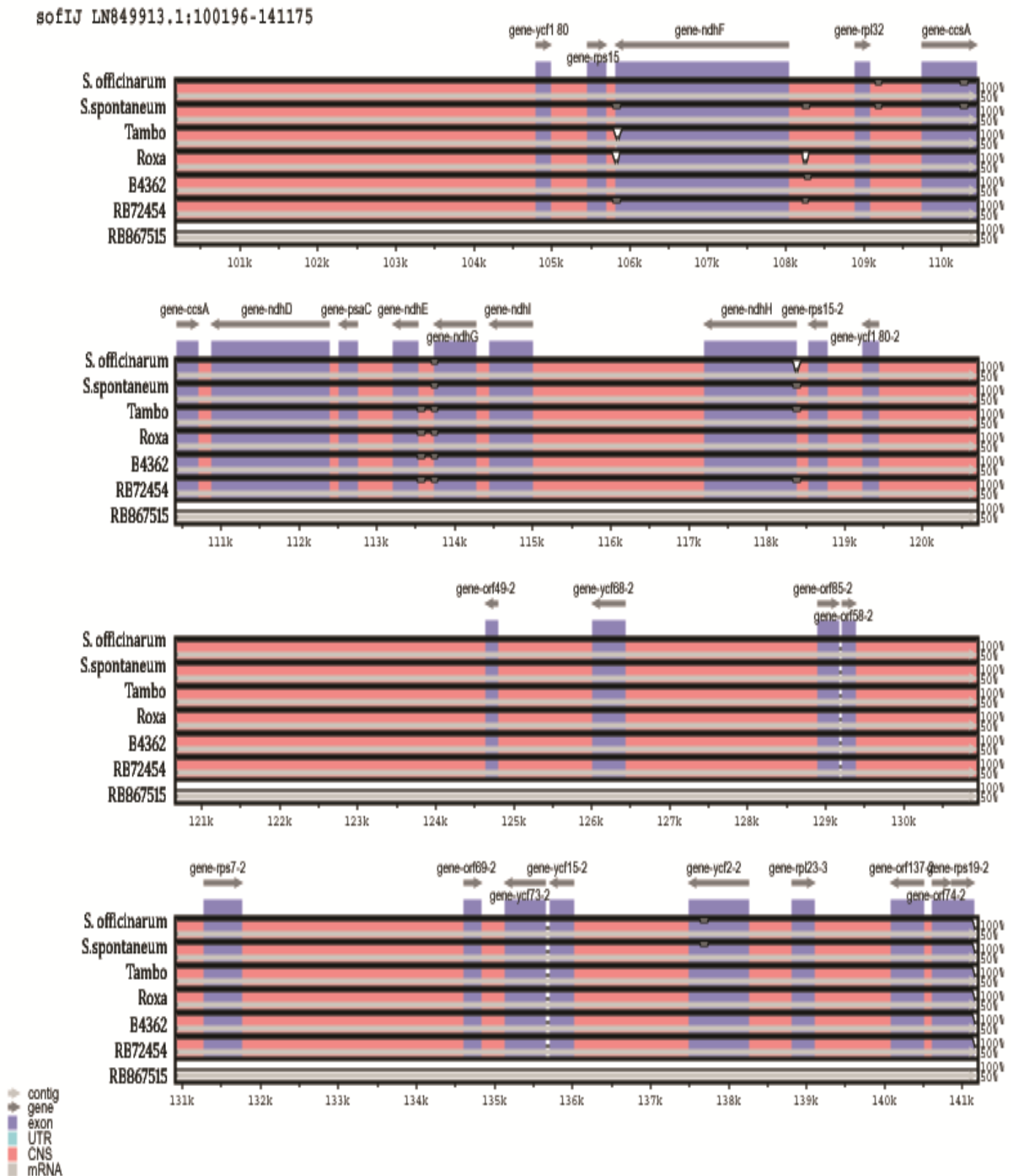


Figura 10- Divergência de sequência baseado em sete genomas de cloroplasto de cana-de-açúcar com referência a *Saccharum officinarum* IJ76-514. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas (CNS) do genoma são mostradas em rosa.

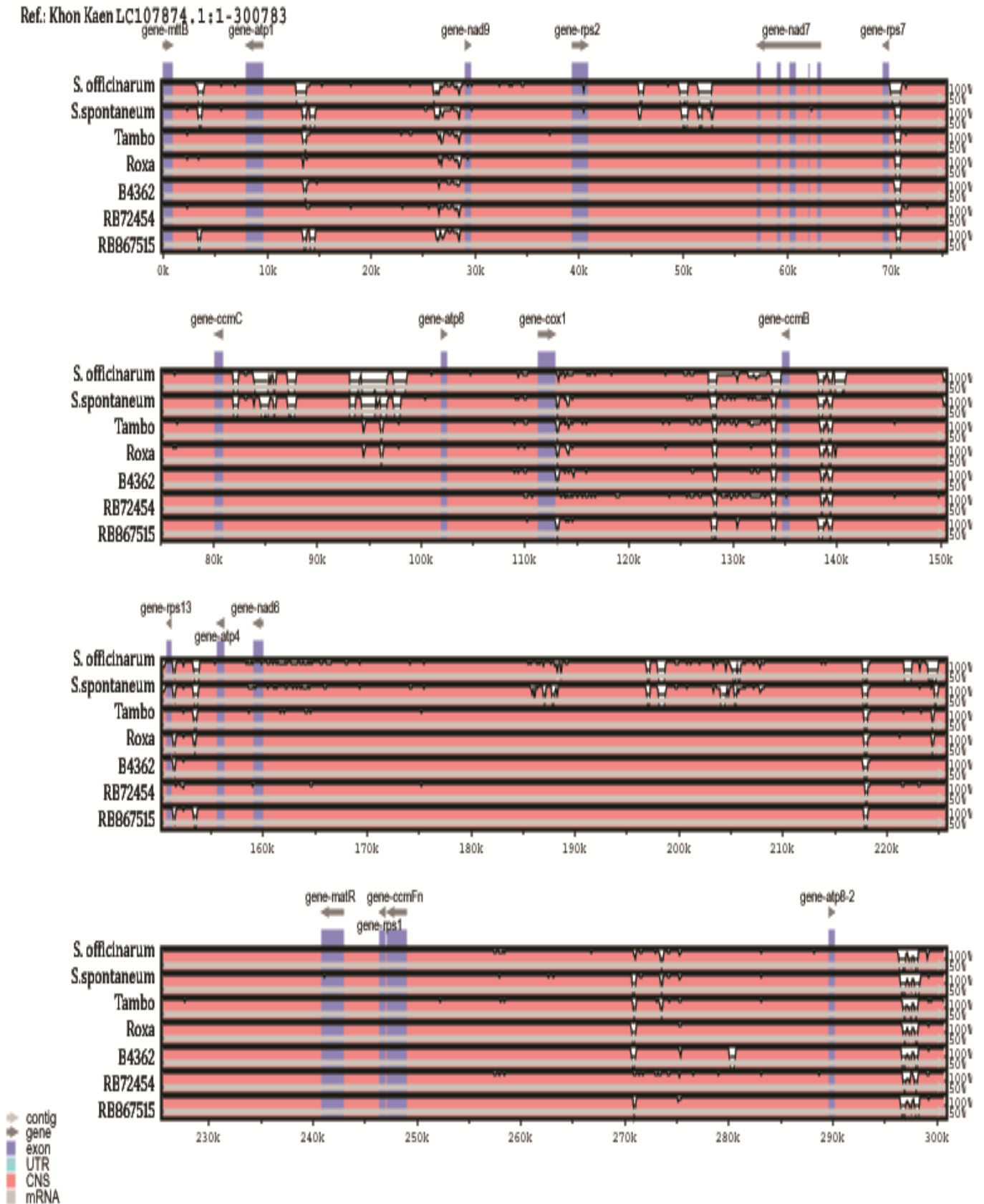


Figura 11- Divergência de sequência baseado em sete genomas de cana-de-açúcar com referência ao cultivar Khon Kaen 3 cromossomo 1. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas(CNS) do genoma são mostradas em rosa.

Ref.: Khon Kaen LC107875.1:1-144697

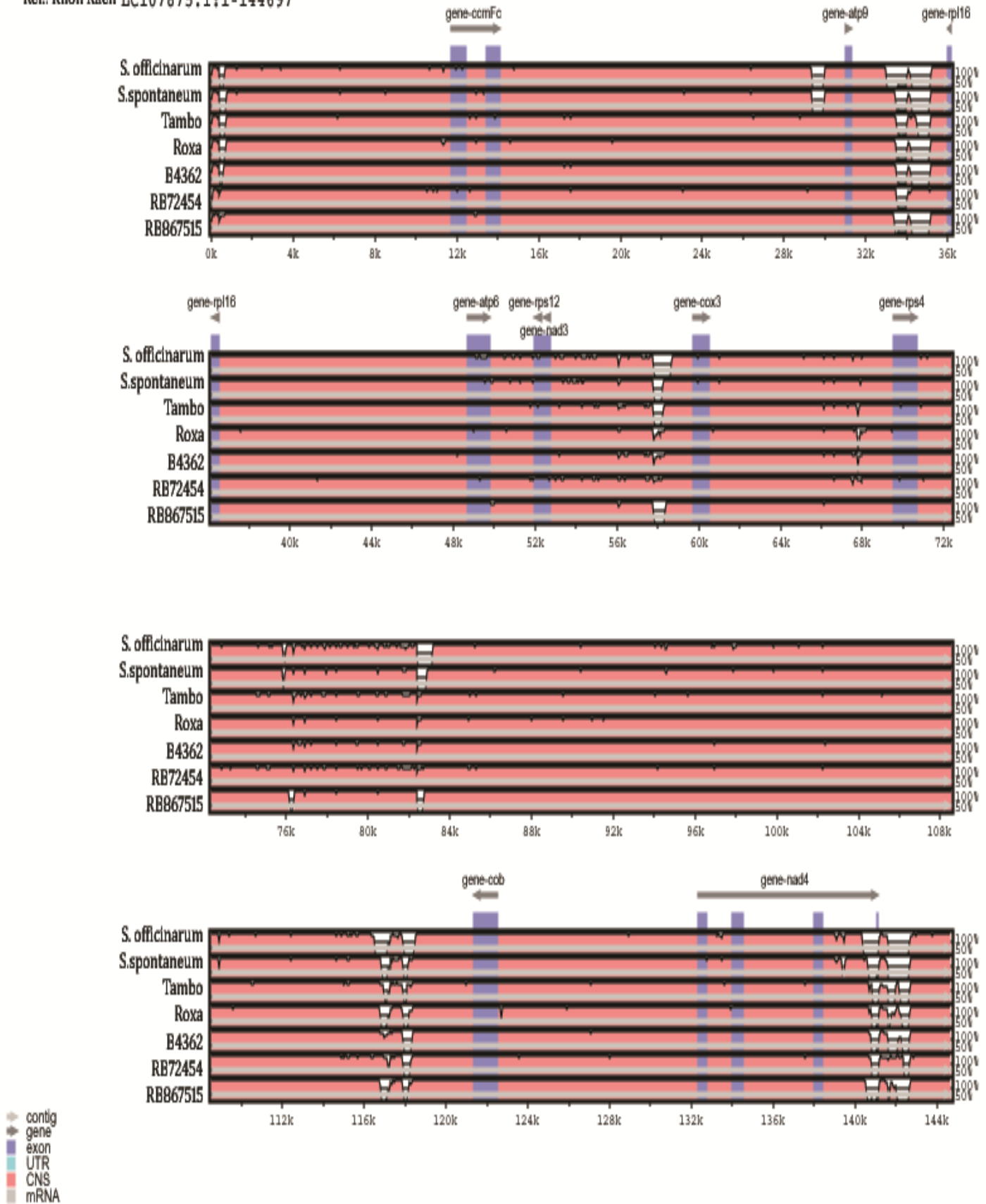


Figura 12- Divergência de sequência baseado em sete genomas de cana-de-açúcar com referência ao cultivar Khon Khaen 3 cromossomo 2. A escala vertical a direita do gráfico é a porcentagem de identidade do alinhamento que vai de 50% a 100%. As sequências não-codificantes conservadas(CNS) do genoma são mostradas em rosa.

4.4. Análise de sequências repetitivas no genoma organelar

4.4.1. Sequências repetitivas em cloroplastos

O genoma de cloroplasto possui muitas sequências repetitivas, dentre elas são as repetições em tandem, palindrômicas e SSRs (sequências simples repetidas) ou microssatélites. Analisamos um total de 724 repetições nos sete genomas de cloroplasto de cana-de-açúcar, distribuídas conforme mostrado na **Figura 13**.

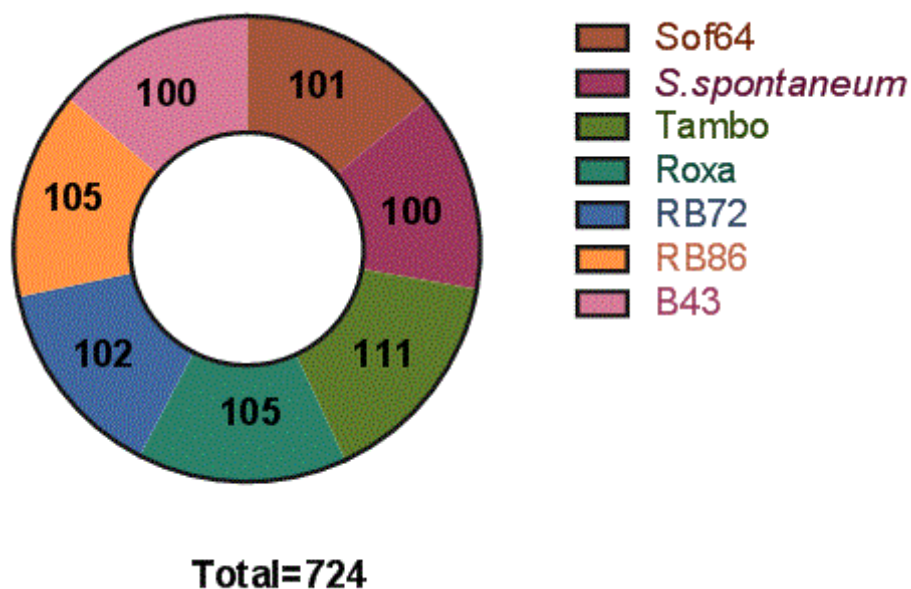


Figura 13- Total das repetições existentes no genoma cloroplastidial de espécies selvagens e cultivares híbridos de cana-de-açúcar.

As repetições dispersas aparecem em maior quantidade no genoma cloroplastidial, somam 46% com 332 repetições. As repetições em tandem ocupam o segundo lugar com 315 repetições no genoma, o que é equivalente a 43% e, por último, as repetições palindrômicas com 77 repetições (11%). A cultivar Tambo obteve um maior número de repetições em tandem chegando à marca de 52 repetições, o que contrasta com *S. spontaneum* que obteve a menor quantidade de repetições (42 repetições em tandem). As repetições palindrômicas parecem ser mais conservadas entre os genomas de cloroplastos de cana-de-açúcar, havendo pouca variação entre as espécies selvagens e os cultivares híbridos, essas repetições ficaram entre 10 a 12 repetições no genoma. As repetições

dispersas também não apresentaram grandes variações, sendo a cultivar Tambo a que apresentou o maior número dentre os demais, 49 repetições (**Figura 14**).

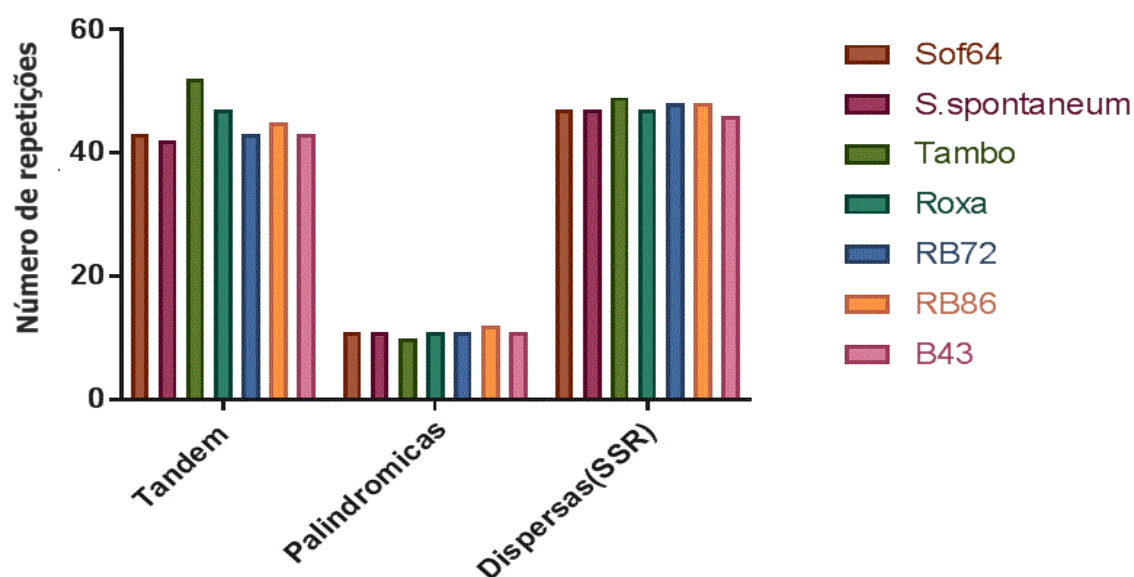


Figura 14- Tipos de repetições (Tandem, Palindromicas e Dispersas) no genoma de cloroplasto de espécies selvagens e cultivares híbridos de cana-de-açúcar.

A análise da distribuição dos SSRs nos cloroplastos mostrou de 2 a 5 unidades de repetição nos sete genomas comparados (**Figura 15**). Em três repetições os motivos mais frequentes são os tetra-nucleotídeos, mas há também a presença de dois penta-nucleotídeos (TATAA e AAAGT). Curiosamente um desses penta-nucleotídeos só aparece no genoma dos dois selvagens (SOF e SSP). Na frequência de quatro repetições todos os cloroplastos apresentaram um tri-nucleotídeos (CTT) e um tetra-nucleotídeo (GTAG). Na quinta frequência de repetição, só ocorreram repetições di-nucleotídicas, sendo que a cultivar Tambo apresentou uma repetição a mais de (AT) do que os outros cloroplastos de cana.

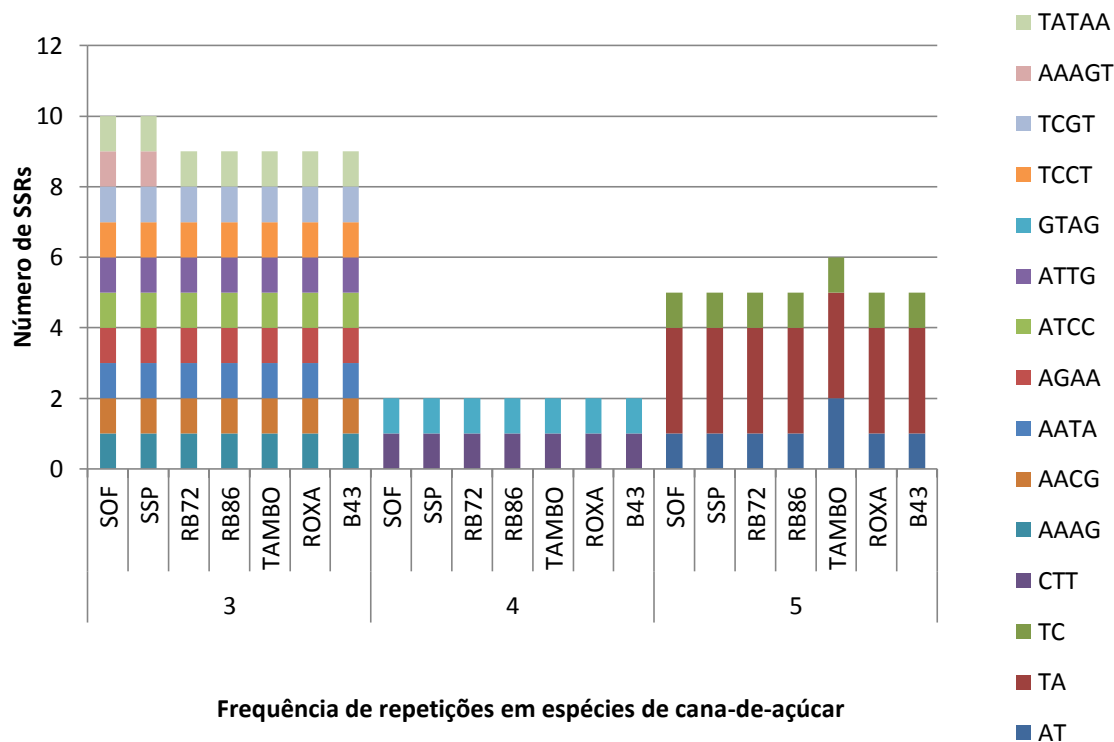
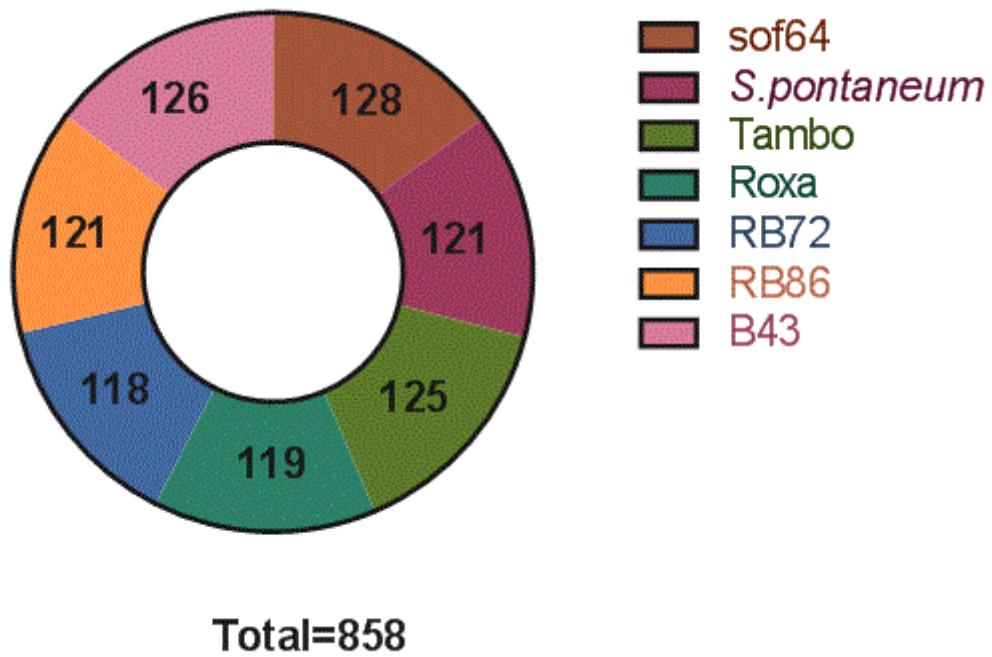


Figura 15- Frequência de motivos de SSRs identificados. Ao lado direito do gráfico a legenda em cores representa as sequências nucleotídicas (di-nucleotídeo, tri-nucleotídeo, tetra-nucleotídeo e penta-nucleotídeos) encontradas em cloroplastos de cana-de-açúcar.

4.4.2. Sequências repetitivas em mitocôndrias

Analisamos as repetições no genoma mitocondrial de forma separada para o cromossomo 1 e para o cromossomo 2. As análises mostraram para o cromossomo 1 um total de 858 repetições (Tandem, palindrômicas e SSRs), distribuídas conforme **Figura 16A**. Para o cromossomo 2 o total de repetições foi de 496, distribuídas pelas espécies selvagens e cultivares híbridos **Figura 16B**.

A



B

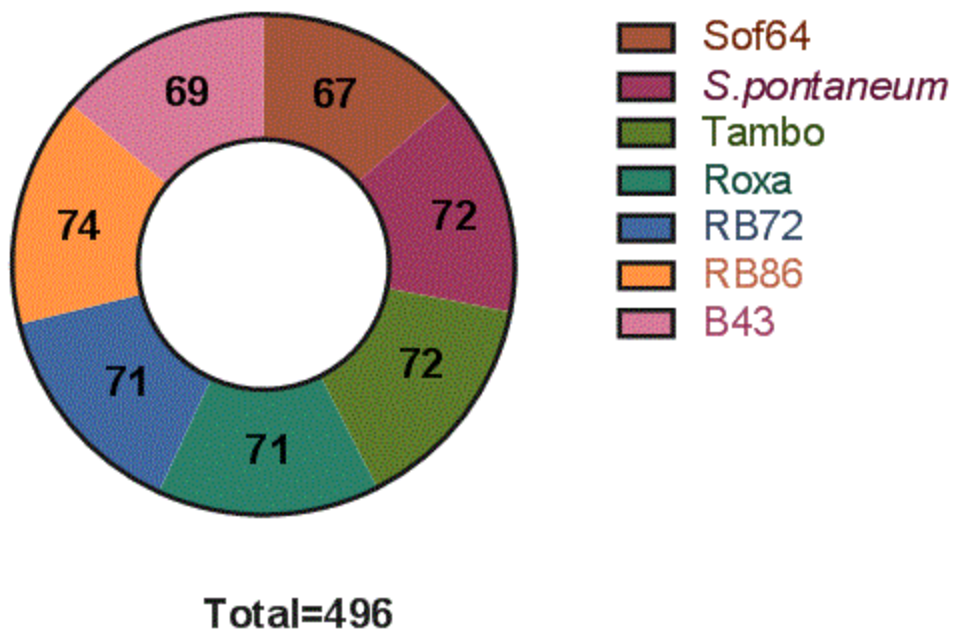
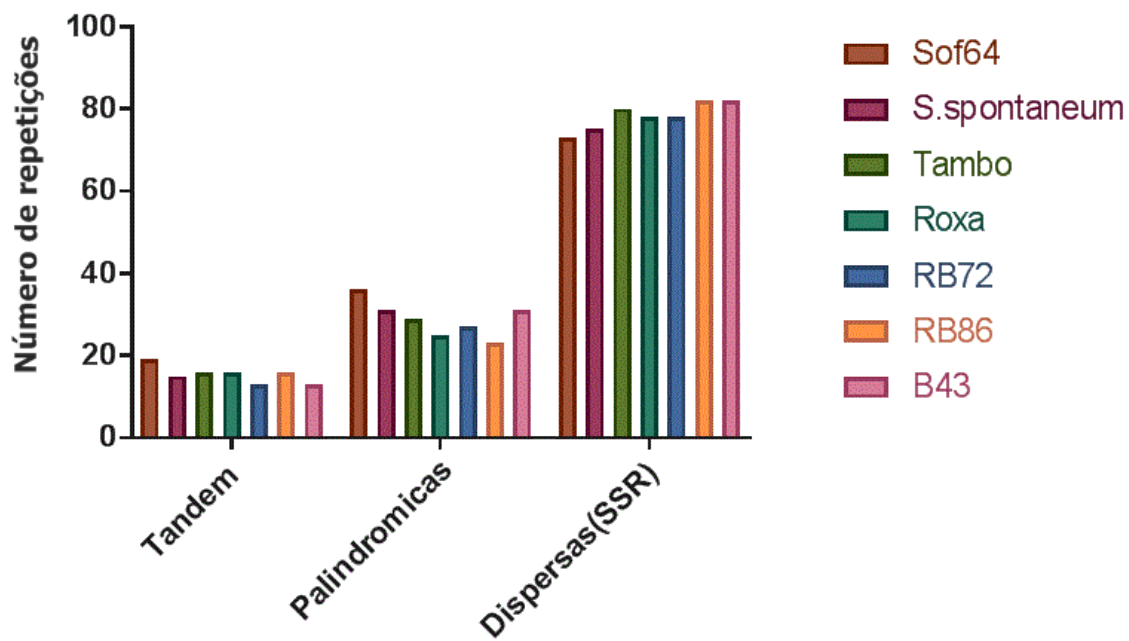


Figura 16- Total das repetições existentes no genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar. (A) cromossomo 1, (B) cromossomo 2.

Assim como nos cloroplastos, o cromossomo 1 mitocondrial também apresentou as repetições dispersas em maior quantidade, sendo 548 repetições (64%). Para as repetições palindrômicas obtivemos 202 repetições, o que corresponde a 23%. Já as repetições em tandem somam 108 repetições, correspondendo a 13% das repetições. *S. officinarum* apresentou a maior quantidade de repetições em tandem dentre os demais com 19 repetições. Tambo e RB72454 tiveram a menor quantidade destas repetições, 13 repetições. Dentre as repetições palindrômicas, *S. officinarum* também apresentou a maior quantidade com 36 repetições. A menor quantidade destas repetições foi de 23 em RB867515. Curiosamente, *S. officinarum* e *S. spontaneum* aparecem com as menores quantidades de repetições dispersas. RB867515 e B4362 apresentaram 82 repetições cada, sendo esta a maior quantidade de repetições nesta classificação (**Figura 17A**).

No cromossomo 2 as repetições em tandem ficaram em torno de 16%, havendo uma pequena faixa de variação desse tipo de repetição, de 11 repetições em Tambo e B4362 a 13 repetições em RB72454 (**Figura 17B**). As repetições palindrômicas ocupam 17% das repetições com um total de 86 repetições, dentre as quais, *S. officinarum* e RB72454 apresentaram a menor quantidade dessas repetições. A Roxa e RB867515 apresentaram o maior número de repetições palindrômicas com 15 repetições cada. Para as repetições dispersas que estão em maior quantidade dentre as repetições, com 329 repetições (67%), o cromossomo 2 das mitocôndrias variaram pouco, de 45 a 49 repetições, sendo a menor quantidade dessas repetições encontrada na B4362 e, Tambo e RB72454 com a maior quantidade de repetições dispersas, 49 repetições cada (**Figura 17B**).

A



B

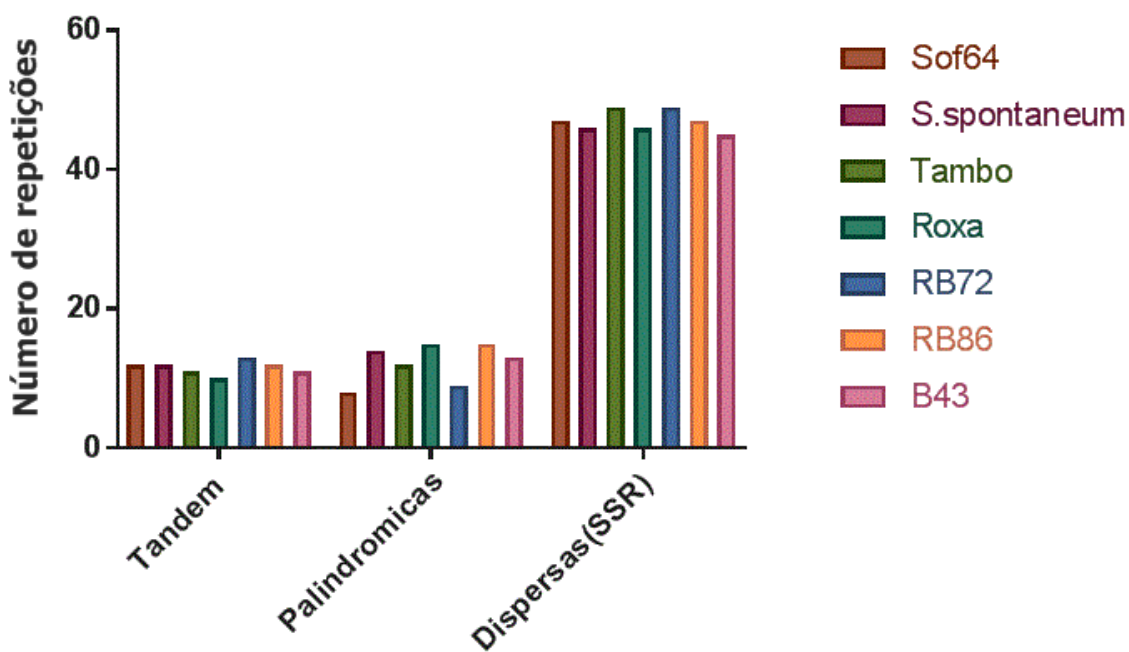
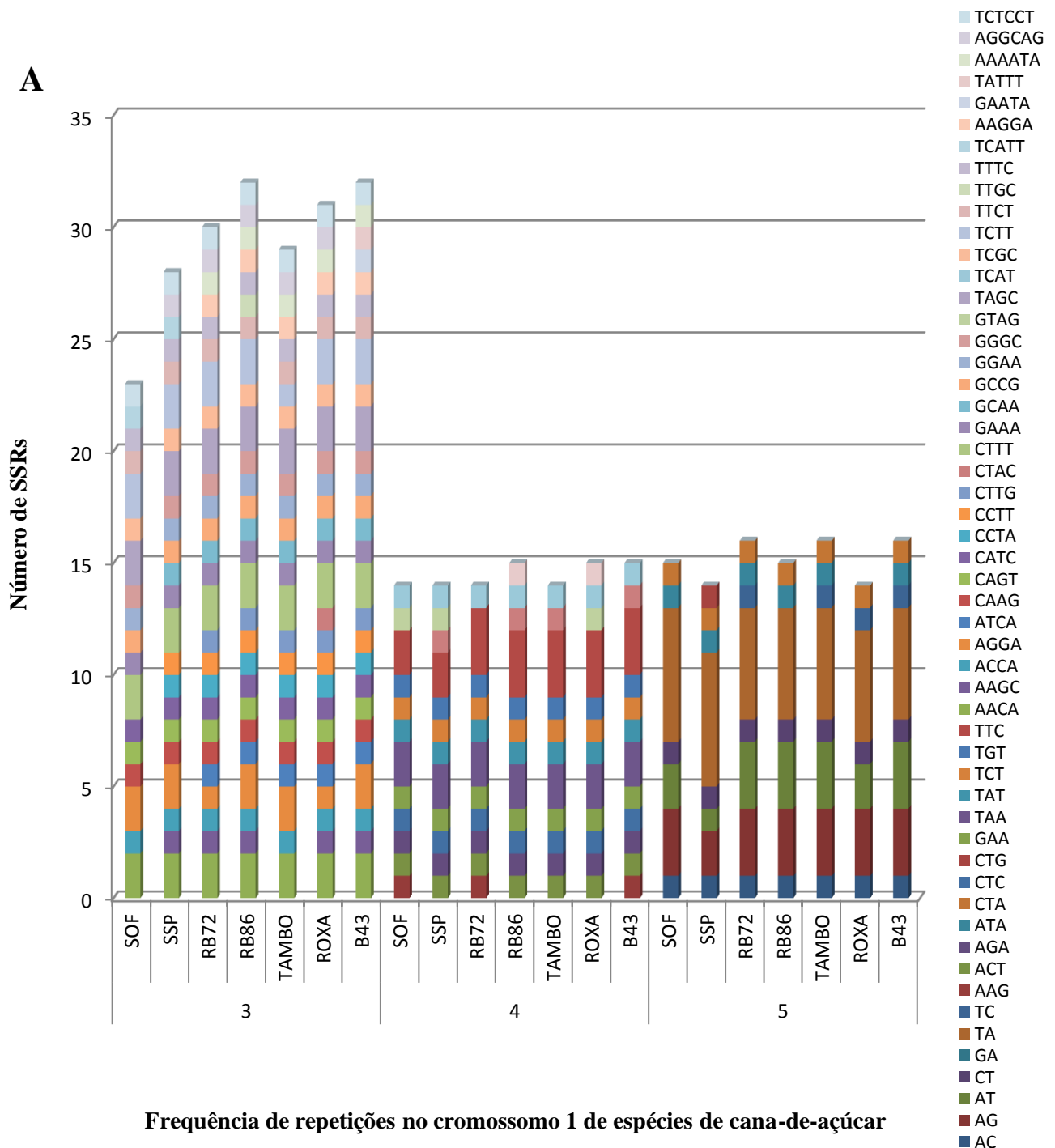


Figura 17- Distribuição dos tipos de repetições (Tandem, Palindromicas e Dispersas) no cromossomo 1 (A) e no cromossomo 2 (B) do genoma mitocondrial de espécies selvagens e cultivares híbridos de cana-de-açúcar.

Analisando a frequência de identidade das repetições dispersas no cromossomo 1 dos genomas mitocondriais, foram identificadas de 2 a 6 unidades de repetição (**Figura 18A**). Na frequência de 3 repetições os motivos mais comuns são os tetra-nucleotídeos, embora haja a presença de quatro penta-nucleotídeos (TATTT, GAATA, AAGGA e TCATT) e três hexa-nucleotídeos (TCTCCT, AGGCAG e AAAATA). Dentre os vários tetra-nucleotídeos mapeados, há um (ATCA) que não aparece nos selvagens *S. officinarum* e *S. spontaneum*. Eles também são os únicos que não possuem um penta-nucleotídeo (AAGGA) identificado nos demais. Por outro lado, esses selvagens possuem uma sequência de penta-nucleotídeo (TCATT) e uma de hexa-nucleotídeos (AAAATA) que os demais não possuem. Ainda na frequência 3 identificamos que a cultivar Roxa é a única a possuir o tetra-nucleotídeo CTAC; a RB867515 é o único a possuir a sequência TTGC; *S. officinarum* e B4362 partilham da ausência de um hexa-nucleotídeo (AGGCAG); e *S. officinarum* e Tambo não possuem um tetra-nucleotídeo (AAGC) que os demais possuem. Na frequência 4, os motivos mais comuns são os tri-nucleotídeos, havendo 3 tetra-nucleotídeos e 1 penta-nucleotídeo. Apenas a RB867515 e a Roxa possuem o penta-nucleotídeo (TATTT); *S. officinarum*, RB727515 e Roxa não possuem a sequência CTAC; *S. officinarum*, *S. spontaneum* e Roxa são as únicas a possuir a sequência GTAG; *S. officinarum*, RB727515 e B4362 são os únicos com o tri-nucleotídeo AAG; e os selvagens *S. officinarum* e *S. spontaneum* possuem um tri-nucleotídeo (TTC) a menos que os cultivares híbridos. Por fim na frequência 5 de repetições do cromossomo 1 os motivos mais frequentes são os di-nucleotídeos, havendo a presença de três tri-nucleotídeos. *S. spontaneum* é o único a possuir o tri-nucleotídeo CTG; a cultivar Roxa é a única que não possui a sequência ATA; *S. officinarum*, *S. spontaneum* e RB867515 compartilham a ausência do di-nucleotídeo TC; *S. spontaneum* possui um AG a menos que os demais; *S. officinarum* e Roxa possuem 2 repetições de AT, *S. spontaneum* possui apenas um AT e os demais possuem 3 repetições de AT; os selvagens possuem um TA a mais que os cultivares híbridos.

Na análise do cromossomo 2 identificamos unidades de repetição de di-nucleotídeo à penta-nucleotídeo (**Figura 18B**). Na frequência 3, foram detectados tetra-nucleotídeos como os mais comuns, embora haja a presença de três penta-nucleotídeos. Os selvagens compartilham algumas similaridades, como possuir um tetra-nucleotídeo (TTCA) e um penta-nucleotídeo (TCTTT) que os demais híbridos não possuem. Eles também partilham a ausência de duas sequências (TGGA e AATAC). A RB72454 e Tambo possuem um tetra-

nucleotídeo (TGAA). Na frequência 4 do segundo cromossomo os motivos tri-nucleotídeos são mais comuns, mas há um único penta-nucleotídeo (AATAC) correspondente a *S. spontaneum*. Dentre os tri-nucleotídeos, somente *S.officinarum* e Tambo possuem a sequência ACT e não possuem a sequência TAC dentre os demais cultivares. Na frequência 5 os motivos mais comuns são os di-nucleotídeos, havendo a presença de um único penta-nucleotídeo (AATAC) correspondente a *S.officinarum*. Dentre as análises de di-nucleotídeos RB72454 e B4362 possuem um di-nucleotídeo (AG) a menos que os demais. Nas demais sequências observa-se os mesmos padrões.



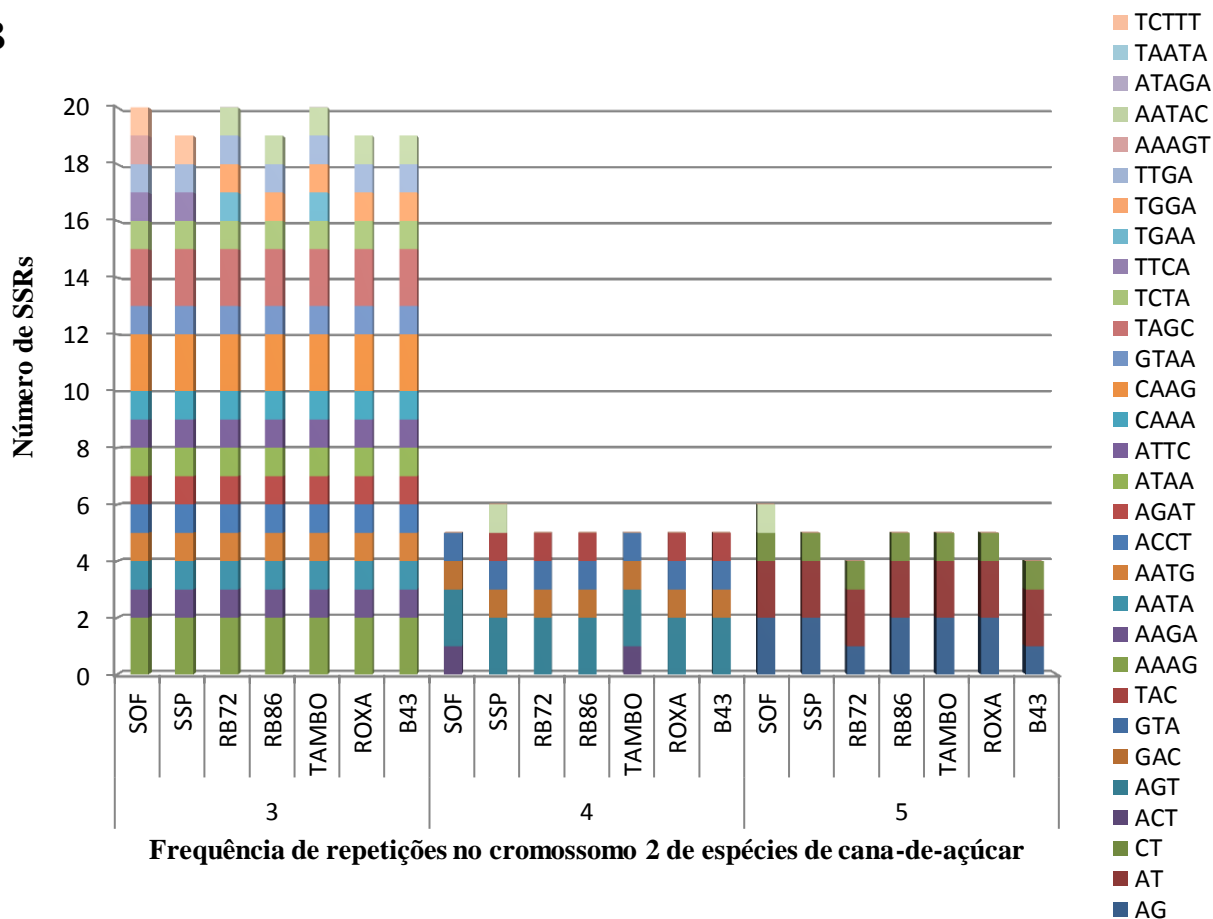
B

Figura 18- Frequência de motivos de SSRs identificados. Ao lado direito do gráfico a legenda em cores representa as sequências nucleotídicas (di-nucleotídeo, tri-nucleotídeo, tetra-nucleotídeo, penta-nucleotídeos e hexanucleotídeo) encontradas no cromossomo 1(A) e no cromossomo 2 (B) do genoma mitocondrial de cana-de-açúcar.

4.5. Detecção de Indels no genoma organelar de cana-de-açúcar

4.5.1 Indels e sua localização no genoma de cloroplasto

As mutações de inserção/deleção (*indels*) no genoma de cloroplasto, principalmente nas regiões intergênicas tem sido utilizadas para estudos filogenéticos em angiospermas (GRAHAM et al., 2000). Na análise de *indels* nos sete genomas de cloroplastos de cana-de-açúcar, a cultivar Tambo apresentou a maior porcentagem de *indels* dentre os demais analisados (17%) (**Figura 19**). Esses *indels* estão localizados na região LSC no espaço intergênico entre os genes *rpl33* e *rps18*. Na cultivar RB867515 somam 13% e estão localizados na região LSC na região intergênica no início do gene *rps18*. A maior porcentagem de *indels* da cultivar Roxa (12%) está localizada na região do gene *rpoC2*, também localizado na LSC. *S. officinarum*, *S. spontaneum*, RB72454 e B4362 obtiveram um total de 11% cada e todos estão localizados na LSC. Em *S. spontaneum* estão na região dos genes *rpoC2* e *ycf1*, e em *S.officinarum*, RB72454 e B4362 estão localizados apenas na região do gene *rpoC2*.

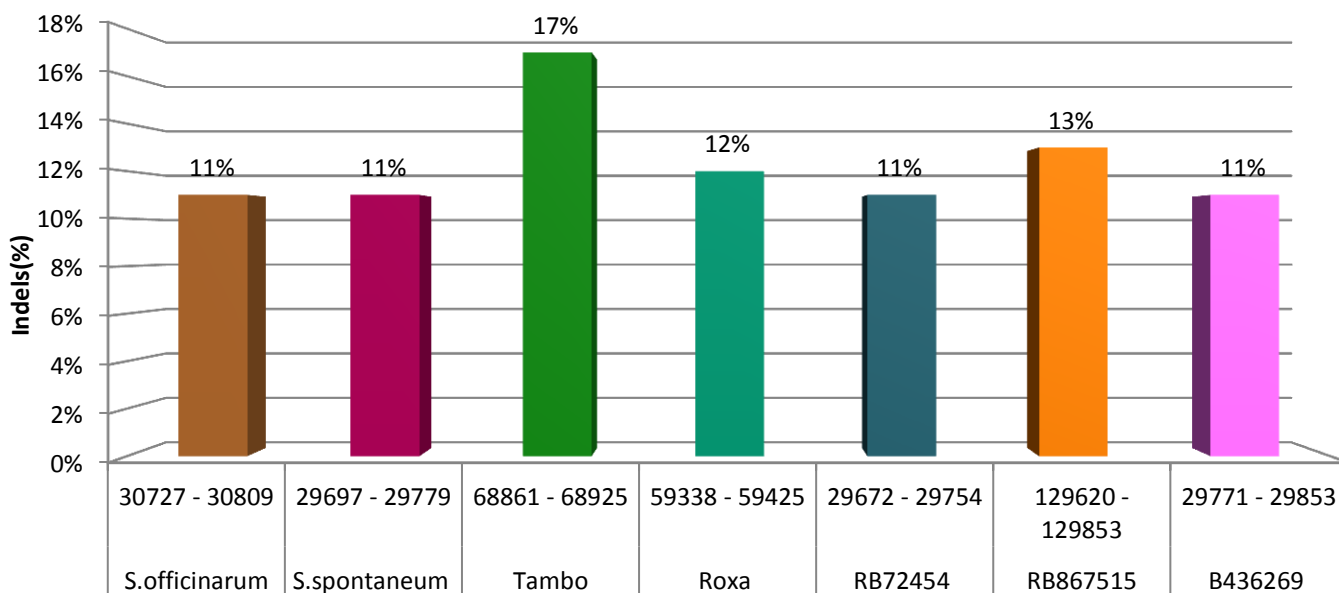


Figura 19- Maior porcentagem de *indels* e sua localização no genoma cloroplasdial de cana-de-açúcar. Os números no eixo horizontal separados por hífen abaixo do gráfico indica o início e o fim da localização dos indels dentro do genoma de cada espécie e cultivar especificada. No eixo vertical a esquerda está a maior porcentagem encontrada de indels.

4.5.2 *Indels* e sua localização no genoma de mitocôndrias

No cromossomo 1 foram encontradas menores porcentagens de *indels* comparado com o cromossomo 2 (**Figura 20**). No primeiro cromossomo Roxa, RB72454, RB867515 e B4362 apresentaram 9% de *indels* no genoma. A localização desses *indels* na RB72454 está entre os genes *trnS-GGA* e *trnM* e na Roxa, RB867515 e B4362 estão localizados entre os genes *trnD* e *trnM*. *S. officinarum* e *S. spontaneum* apresentaram um total de 7% de *indels*, estando localizados entre os genes *ccmC* e *rrn26* para *S.officinarum* e, entre os genes *nad6* e *rrn26* para *S. spontaneum*. Com 6% de *indels* localizados entre os genes *trnM-CAU* e *trnH*, está a cultivar Tambo.

A maior porcentagem de *indels* encontrada para o cromossomo 2 foi de 14% para todas as mitocôndrias analisadas. Todas as localizações desses *indels* nos diferentes cultivares e nas espécies selvagens se localizam dentro do gene *rps4*, mostrando uma alta similaridade entre os *indels* do cromossomo 2 do genoma mitocondrial de espécies do gênero *Saccharum* analisadas neste estudo.

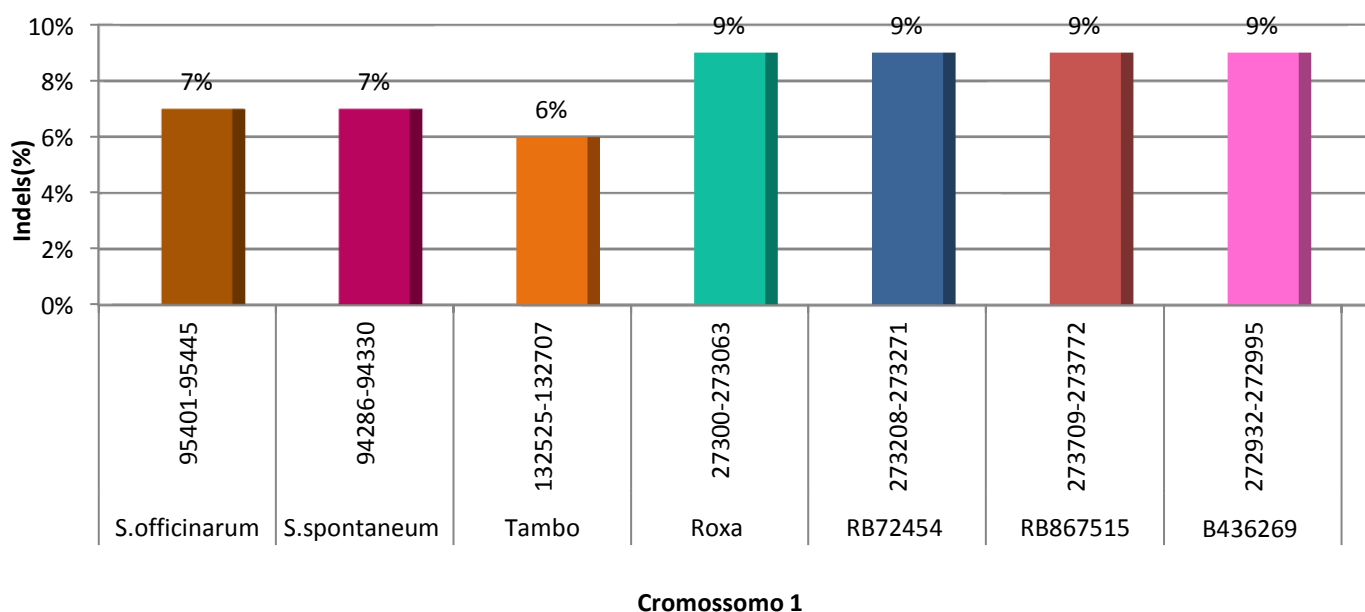


Figura 20- Maior porcentagem de *indels* e sua localização no cromossomo 1 do genoma mitocondrial de cana-de-açúcar.

4.6. Comparação das regiões IRs de cloroplastos de cana-de-açúcar

A comparação das fronteiras IRs de cloroplastos revelou que o gene *psbA*, que faz parte do fotossistema, é bem conservado entre os genótipos de cana-de-açúcar analisados, com 1061 pares de bases (pb) (**Figura 21**). O gene *ndhH*, que está localizado próximo à junção SSC-IRb (JSB), apenas a cultivar RB72454 apresentou uma pequena compressão deste gene de 1 pb. O gene *rpl22* em *S. spontaneum*, B4362 e Tambo possui 20 pb a menos do que nos demais cloroplastos. *S. spontaneum*, Tambo, Roxa e B4362 possuem o gene *rps15* com mesmo tamanho (275 pb). Já *S. officinarum*, RB72454 e RB867515 possuem este mesmo gene com um tamanho menor (236 pb). O gene *rps19* possui variações de tamanho indo de 215 pb em RB72454 a 283 pb em B4362. O gene *ndhF* que está na fronteira IRa – SSC (JSA) apresentou compressão em *S. officinarum*, RB72454 e RB867515 com 2216 pb e expansão nos demais genótipos.

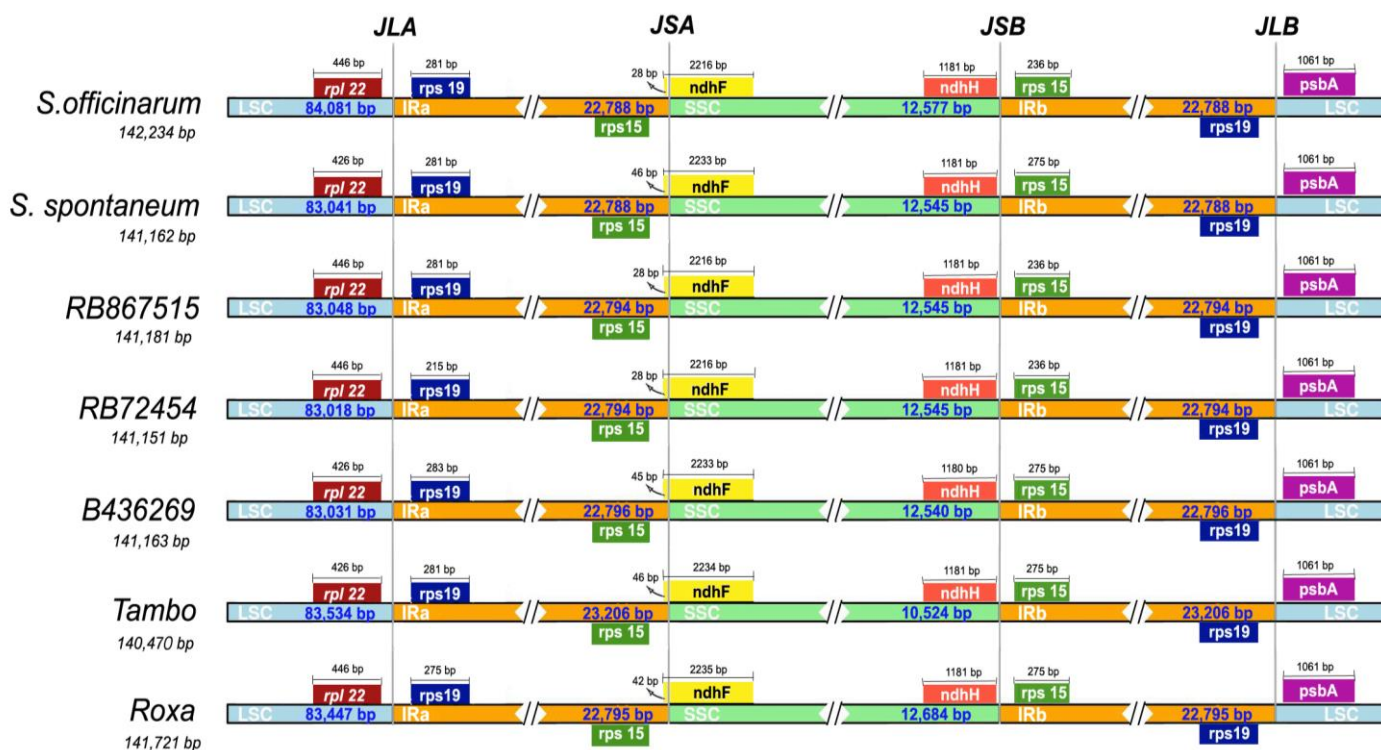


Figura 21- Comparação das fronteiras IRs em cloroplastos de cana-de-açúcar. Junção LSC- IRa (JLA); junção IRa-SSC (JSA); junção SSC-IRb (JSB) e junção IRb-LSC (JLB).

5. DISCUSSÃO

5.1. Montagem e anotação de genomas

Há 37 anos o mundo conheceu a estrutura e organização dos primeiros genomas organelares montados e anotados, os genomas mitocondriais de ratos e humanos (ANDERSON et al., 1981; BIBB et al., 1981). Em 1986, a sequência do genoma cloroplastidial também foi determinada para *Marchantia polymorpha* (OHYAMA et al., 1986) e tabaco (SHINOZAKI et al., 1986). O desenvolvimento cada vez maior da bioinformática contribui para que a montagem e anotação de genomas de organelas sejam feitas de forma menos dispendiosa e em menor tempo (MARIAC et al., 2014; TANG et al., 2014). Assim, o número de sequenciamentos completos (sequenciamento, montagem e anotação) de organelas vem aumentando. A partir dos genomas organelares disponíveis em bancos de dados é possível fazer a comparação entre essas organelas de forma estrutural, genética e evolutiva, correlacionando suas similaridades e ou divergências (BENDICH, 2007; NOSEK; TOMÁŠKA, 2003). Neste trabalho além da montagem e anotação, realizamos também a comparação de espécies selvagens e cultivares híbridos de cana-de-açúcar utilizando genomas de organelas. Esses genomas inspiraram as teorias modernas da evolução molecular (GRAY et al., 2010; LYNCH; KOSKELLA; SCHAACK, 2006), o que os torna cada vez mais importantes e desafiadores para a comunidade científica.

Genomas cloroplastidiais de plantas variam de 120.000 a 160.000 pb e possuem uma estrutura dividida em quatro partes - LSC, SSC, IRa e IRb (PALMER, 1985). A montagem *de novo* realizada para os genomas de cloroplastos de cana-de-açúcar revelou o mesmo padrão estrutural descrito anteriormente e os tamanhos variaram de 140.470 pares de bases na cultivar Tambo a 142.234 pares de bases no selvagem *S.officinarum*, o selvagem *S.spontaneum* apresentou menor tamanho que os cultivares híbridos RB867525, Roxa e B4362 (**Tabela 1**). Já o tamanho total do genoma mitocondrial segue o padrão de tamanho maior dos selvagens em relação aos cultivares híbridos (**Tabela 2**). É importante ressaltar que o DNA pode sofrer modificações epigenéticas, como é o caso da metilação que ocorre na base citosina, onde é adicionada um grupamento metila e geralmente está relacionada ao silenciamento do gene em eucariotos (LAW; JACOBSEN, 2010), em plantas a metilação do genoma nuclear é predominante no contexto CG e isto é uma característica conservada evolutivamente (FINNEGAN; PEACOCK; DENNIS, 2000;

ROBERT A. MARTIENSSEN; VINCENT COLOT, 2001), segundo Smith (2012), o contexto CG também é comum entre as organelas de plantas, e seu maior conteúdo estando em locais bem restritos dos genomas o que pode sugerir que o viés CG seja fruto de seleção natural nessas organelas (SMITH, 2012). As maiores quantidades encontradas de CG nos cloroplastos de cana-de-açúcar foi em *S.spontaneum* (38,43%) e RB867515(38,44%). Um estudo realizado em cloroplasto de milho mostrou que a metilação neste genoma pode estar envolvidos nos mecanismos regulatórios de genes relacionados à fotossíntese (NGERNPRASIRTSIRI et al., 1989). Já em genomas de mitocôndrias a metilação estaria associada aos mecanismos regulatórios de disponibilidade de energia (ATP). Dessa forma, uma diferença nos padrões de metilação destes genomas, poderia interferir no desenvolvimento da planta (RODRIGUES et al., 2017). Isso corrobora com a pouca quantidade de variações CG encontradas no genoma mitocondrial de cana-de-açúcar (**Tabela 2**).

A anotação dos genomas cloroplastidiais exibiu um quantitativo de 129 genes em *S. spontaneum*, RB72454, Roxa, Tambo e B436269; 135 genes em RB867515, conforme já descrito (VIDIGAL et al., 2016) e 159 genes para o selvagem *S. officinarum*. O quantitativo de genes obtidos pelo selvagem é muito próximo do quantitativo descrito para *Saccharum officinarum* IJ76-514 e *Miscanthus floridulus* que é de 161 genes (EVANS; JOSHI, 2016), apenas dois genes a mais que o clone de *S. officinarum* utilizado aqui. Dentre os genes anotados, quatro genes (*ccsa*, *infA*, *cemA* e *ACRS*) foram anotados pelo software como pseudogenes. O *ACRS* só aparece em *S. officinarum*, assim como também foi relatado para *M. floridulus* e IJ76-514 (EVANS; JOSHI, 2016). Este gene foi identificado em um íntron do RNAt – Ala mitocondrial de *Citrus jambhiri*, sendo este gene de resposta à sensibilidade a toxina ACR (que causa a doença mancha marron) produzida pelo fungo *Alternaria alternata* (OHTANI; YAMAMOTO; AKIMITSU, 2002). Em *S. officinarum* ele está localizado na região IR (**figura 8A**) e pode ser proveniente de translocação com o genoma mitocondrial (EVANS; JOSHI, 2016). O gene *ccsa* também encontrado em cloroplastos de *Euglena gracilis*, parece ser um gene similar a um gene de núcleo (ch-42) encontrado em *Arabidopsis thaliana* (KONCZ et al., 1990). Este gene pode ter sido transferido para o genoma cloroplastidial, sendo sua função relacionada com a pigmentação do cloroplasto (ORSAT et al., 1992). O gene *cemA* é uma proteína de membrana de cloroplastos (PIOT et al., 2018) e o gene *infA* é um fator para o início da tradução encontrado em cloroplastos, que ao longo da evolução pode ter sido transferido

para o núcleo, embora muitas plantas ainda possuem este gene no genoma de cloroplasto (MILLEN, 2001). Em todos os cloroplastos *Saccharum* o número de RNAt foi o mesmo (39), o que caracteriza a conservação desses RNAs (**tabela 3**). O número de RNAr foi de oito para todos os genomas estudados. Até mesmo em espécies distantes filogeneticamente do gênero *Saccharum* como *Ilex* (YAO et al., 2016), *Arabidopsis thaliana* (STADERMANN; HOLTGRÄWE; WEISSHAAR, 2016) e *Capsicum annuum* (RAVEENDAR et al., 2015), também foram encontrados oito RNA ribossomais, o que mostra o alto grau de conservação destes RNAs em cloroplastos. Apenas as CDS tiveram seus números variáveis nos diferentes cultivares, o que fez com que a soma do número total de genes variasse entre os cloroplastos. No selvagem *S. officinarum* foi identificada a presença dos pseudogenes (*ycf1*, *ycf68*, *ycf73*, *ycf15* e *ycf2*) na região IR. É interessante observar que nenhum dos outros cloroplastos estudados aqui apresentou essa característica. Estudos relatam a perda, por exemplo, do pseudogene *ycf2* para angiospermas (SATO et al., 1999; SHINOZAKI et al., 1986; WOLFE, K. H., MORDEN, C. W. & PALMER, 1992).

A anotação do cromossomo 1 das mitocôndrias de todos os sete genomas *Saccharum* apresentou grande conservação, com a mesma quantidade de CDS e RNAr, sendo 20 CDS e 3 RNAr (**tabela 4**). O número de RNAt foi variável no cromossomo 1, sendo de 17 em RB72454 à 25 em *S. officinarum*. No cromossomo 2 a CDS também é conservada sendo 15 para todas as mitocôndrias estudadas. O cromossomo 2 não apresentou RNAr e o número de RNAt variou pouco, sendo de 4 para RB867515, 6 para *S. officinarum* e 5 para *S. spontaneum*, Tambo, Roxa, B4362 e RB72454 (**Figura 9 A-G**). Os modelos mitocondriais gerados neste trabalho foram coerentes com o único modelo *Saccharum* publicado, o cultivar híbrido Khon Kaen 3 (SHEARMAN et al., 2016), contendo inclusive o trnL-CAA que foi perdido em sorghum. As diferenças encontradas nas mitocôndrias montadas neste trabalho com relação a cultivar Khon Kaen 3 foram: a ausência de um íntron no meio dos genes *nad2* e *nad5* no cromossomo 2 e de trnF-RAA e trnF-GAA apresentados no cromossomo 2. Esses genes são provenientes de cloroplastos e são funcionais em mitocondriais, assim como ocorre em mitocôndrias de trigo (JOYCE; GRAY, 1989). A mitocôndria se apresentou conservada entre os selvagens e cultivares híbridos de cana-de-açúcar, mostrando poucas diferenças em RNAt.

5.2. Análise de divergência de sequência nucleotídica nas organelas

Tanto no genoma cloroplastidial quanto no mitocondrial de cana-de-açúcar, as regiões mais divergentes foram as intergênicas, conforme já descrito para análises em 15 cloroplastos de algodão (WU et al., 2018). Nos genomas de cloroplasto dos selvagens *S. officinarum* e *S. spontaneum* foram encontrados quatro principais pontos divergentes com cultivares (*matk/psbl*, *ycf2*, *ccsa* e *atpA/rps14*), o que caracteriza a similaridade do genoma desses selvagens. Encontramos também um ponto de divergência em Tambo e RB72454 (*petN/rpoB*). Tambo apresentou outra divergência no gene *rpoC2* e *S. officinarum* e B4362 uma divergência entre *rbcl/cemA* (**Figura 10**). Essas regiões que variam no genoma de cloroplasto tem sido alvo de estudos filogenéticos, principalmente no desenvolvimento de marcadores de DNA, os genes *rbcl* e *matk* são utilizados com essa finalidade (YAO et al., 2016).

Nos genomas mitocondriais analisados, o cromossomo 1 (**Figura 11**) se apresenta mais conservado que o cromossomo 2 (**Figura 12**), e nos dois cromossomos a maior divergência também foi na região intergênica. No cromossomo 1 o gene *rps2* é menos conservado nos selvagens *S. officinarum* e *S. spontaneum* e no cromossomo 2 houve picos de divergência em alguns éxons: *ccmFc* para *S. officinarum* e RB72454; *cox3* para os selvagens; *rps4* para Tambo e RB72454; *rps12* para *S. officinarum* e Tambo; *atp6* para os selvagens, Roxa e RB72454. Dessa forma, RB72454 obteve três pontos de divergências em genes mitocondriais (*ccmFc*, *atp6* e *rps4*). O gene *ccmFc* já foi descrito para *Arabidopsis thaliana* como componente do complexo de maturação do citocromo C (RAYAPURAM et al., 2008). O gene *atp6* que faz parte do complexo de síntese de ATP nas mitocôndrias, foi comparado entre milho e trigo, e também apresentou divergência de sequência. Autores relatam a possibilidade de rearranjos deste DNA em monocotiledôneas (BONEN; BIRD, 1988), o que pode ter ocorrido com RB72454 para este gene. O gene *rps4* que codifica a subunidade s4 do ribossomo, também foi descrita para *Brassica napus* e *Oryza sativa*, onde também se detectou divergência de sequência mostrando 78,1% de identidade entre ambas (HANDA; KUBO, K. KADOWAKI, 1998), o que é coerente com nossas análises para este gene. Com esses pontos variáveis identificados no genoma de cloroplasto e no genoma de mitocôndria dessas espécies selvagens e cultivares híbridos de cana-de-açúcar, podem facilitar a construção de primers para identificação destes genótipos em cruzamentos.

5.3. Análise de regiões repetitivas nos genomas organelares

As repetições nos genomas cloroplastidiais e mitocondriais também foram abordadas nesse estudo, um total de 724 repetições foram encontradas em cloroplastos (**Figura 13**) e 1354 em mitocôndrias (**Figuras 16A e B**). Dessas repetições, a mais comum nos dois tipos de genomas são as de microssatélites (SSRs). Nos genomas de cloroplastos analisados os SSRs chegam a 46% do genoma, próximo a porcentagem encontrada para algodão, que foi de 42% (WU et al., 2018). Para o genoma mitocondrial a porcentagem de SSRs aumenta ainda mais, chegando a 64% no cromossomo 1 e 67% no cromossomo 2, o que totaliza 877 repetições SSRs. Um número bem maior do que o encontrado no genoma mitocondrial de algodão, que foi de 674 SSRs (BI et al., 2016).

Dos microssatélites identificados para cloroplasto (**Figura 15**), o motivo mais comum são os tetra-nucleotídeos com 60% dos SSRs, os di-nucleotídeos com 20%, os penta-nucleotídeos com 13% e, por fim, os trinucleotídeos com 7%. Esses resultados são similares com as frequências de identidade de SSRs encontrados para nove cloroplastos de gramíneas (SASKI et al., 2007), no qual se encontrou 50% de frequência de tetra-nucleotídeos, sendo esses os mais comuns motivos. Assim como descrito neste trabalho, para di-nucleotídeos Sasaki et al (2007) encontraram uma frequência de 30%, 10% a mais em gramíneas. A quantidade de penta-nucleotídeos com tri-nucleotídeos somam um total de 20% (SASKI et al., 2007), similar as taxas de cana-de-açúcar para estes motivos. Assim como nos cloroplastos analisados por Sasaki et al (2007), todos os cloroplastos analisados aqui apresentaram também os di-nucleotídeos (AT) dentro da região codificadora do gene *rpoC2*. Somente a cultivar Tambo possui uma sequência AT a mais que os outros cloroplastos analisados (**figura 15**). *S. officinarum* e *S. spontaneum* compartilham a similaridade de possuir um penta-nucleotídeo (AAAGT) repetido três vezes nos seus genomas, o que pode explicar a similaridade encontrada também na análise de sequências divergentes. A maior frequência de motivos para mitocôndrias são os mono-nucleotídeos (A e T), os motivos C e G geralmente são raros ou estão ausentes em mitocôndrias de monocotiledôneas (KUNTAL; SHARMA, 2011). Nossos resultados são similares ao descrito acima, pois no cromossomo 1 foram identificados os mono-nucleotídeos A e T, em um total de 17 a 20 motivos entre as mitocôndrias e em repetições de 10 à 15. No cromossomo 2 os mono-nucleotídeos só aparecem na repetição 10 com motivos (A,T e G), embora o G seja em menor número (dados não mostrados). Neste cromossomo os tetra-nucleotídeos são mais frequentes. Na frequência de 4 repetições encontramos vários trinucleotídeos para o cromossomo 1 e 2, o que não ocorre no genoma mitocondrial de *Oryza*

e *Sorghum*, mas há ocorrência no genoma de *Zea Mays* (KUNTAL; SHARMA, 2011). Por fim na frequência de 5 repetições o cromossomo 1 apresentou di-nucleotídeos e tri-nucleotídeos e o cromossomo 2 tri-nucleotídeos e um penta-nucleotídeo. Segundo Kuntal e Sharma (2011) as repetições de penta e hexanucleotídeos no genoma mitocondrial também pode ser observadas em *Zea mays*, *Megaceros aenigmaticus* e *Cucurbita pepo*. Em conjunto, as informações dos SSRs e a análise de divergência de sequências em mitocôndrias, mostraram que os picos de divergência nos cromossomos 1 de *S. officinarum* e *S. spontaneum* podem ser relacionados aos SSRs encontrados (**Figura 18A**). Neste mesmo cromossomo há uma divergência entre os genes *metB* e *atp1* compartilhada com *S. officinarum*, *S. spontaneum* e RB867515 que pode ser detectada através de SSRs na frequência de cinco repetições, que mostra a ausência da sequência TC nessas mitocôndrias. No cromossomo 2 no gene *rps12* houve picos de divergência compartilhados por *S.officinarum* e Tambo (**figura 12**). E a análise de SSRs também mostrou para essas espécies que na frequência de quatro repetições no cr2 as mitocôndrias não possuem a sequência TAC diferente dos demais e possuem o tri-nucleotídeo ACT que os demais não possuem (**Figura 18B**). Os selvagens compartilham algumas divergências entre os genes *ccmFc* e *atp9*, no gene *cox3* e em regiões intergênicas (**Figura 12**), que podem estar relacionadas ao penta-nucleotídeo (TCTTT) encontrado na frequência três de repetição ou a ausência de um tetra-nucleotídeo (TGGA) na frequência três do cr2 das mitocôndrias desses selvagens (**Figura 18B**). Essas informações de divergência de SSRs podem fornecer informações relevantes, facilitando os estudos filogenéticos, como a genotipagem molecular realizada para clones de cana-de-açúcar (PAN et al., 2003), mapeamento de alelos na cultivar *Saccharum Q165A* (AITKEN; JACKSON; MCINTYRE, 2005), e até a confirmação de descendência para cultivares híbridos de cana (CAI et al., 2005).

5.4. Análise de inserção/deleção nos genomas organelares

Todos os *indels* encontrados nos cloroplastos de cana-de-açúcar estão localizados na região LSC, em sua maioria na região não codificante. Essas alterações estruturais presentes na região não codificante do cloroplasto parecem ter relevância na evolução deste genoma, pois evoluiria a partir de substituições nucleotídicas que podem se acumular no genoma cloroplastidial (BRITTEN et al., 2003; KELCHNER, 2000). Os genes afetados pelas maiores porcentagens de *indels* foram os genes *rpoC2* (*S. officinarum*, Roxa, *S. spontaneum*, RB72454 e B4362) e *ycf1* (*S. spontaneum*). O gene *rpoC2* já foi descrito

como o principal alvo dessas substituições em outras gramíneas (ASANO et al., 2004; DIEKMANN et al., 2009) e também em seis espécies de bambu (WANG et al., 2017). O *ycf1* também é alvo de indels em algumas espécies, dentre elas cinco espécies de *Quercus* (YANG et al., 2016) e em *Passiflora* (JANSEN et al., 2007). A cultivar Tambo apresentou uma menor conservação no gene *rpoC2*. A análise de indels mostrou duas regiões com porcentagens de indels (10% e 11%) e que se somadas alcançam 21%. Essas regiões estão localizadas dentro do gene *rpoC2*, o que explica a sua maior diversidade nucleotídica (**apêndice A**).

O genoma mitocondrial de angiosperma além de ser complexo, são extensos em tamanho e possuem grandes regiões intergênicas (RODRÍGUEZ-MORENO et al., 2011; WARD; ANDERSON; BENDICH, 1981). As diferenças encontradas nas substituições de nucleotídeos no genoma mitocondrial de eucariotos, que fez surgir a teoria de que as mutações são as principais responsáveis pela modificação da estrutura organelar (LYNCH; KOSKELLA; SCHAACK, 2006). As substituições detectadas em nosso estudo no cromossomo 1 do genoma mitocondrial (**Figura 20**) estavam em regiões não codificantes. Embora a análise de divergência mostre os selvagens com divergência na região do gene *rps2* (**figura 11**) não identificamos indels nessa região. O cromossomo 2 apresentou alta conservação para todas as mitocôndrias analisadas em relação a presença de indels, sendo todas as maiores porcentagens de indels localizados dentro do gene *rps4*. Indels também foram encontrados dentro do gene *rps4* em mitocôndrias de beterraba (ITCHODA et al., 2002) e SSRs foram descritos dentro do gene *rps4* para trigo e sorgo (HONMA et al., 2011). Segundo Hisano et al (2016) a diversidade de SSRs em mitocôndrias pode ser bem alta o que pode ser utilizado para buscar polimorfismos entre espécies selvagens e híbridos (HISANO et al., 2016).

5.5. Análise das fronteiras IRs de cloroplastos

Essas análises realizadas nos cloroplastos de cana-de-açúcar mostraram que os genes analisados (**Figura 21**), com exceção de *psbA*, sofrem pequenas compressões ou expansões entre os diferentes genótipos, o que corrobora com as análises realizadas nessas bordas em arroz e espécies relacionadas (ASAF et al., 2017a), como também em doze espécies de *Arabidopsis* (ASAF et al., 2017b) e em bambus herbáceos (WANG; CHEN; ZHANG, 2018). Segundo alguns autores, essa variação existente nas fronteiras IRs pode

estar diretamente relacionadas ao tamanho total do genoma apresentado pelos cloroplastos (KODE et al., 2005; RAUBESON et al., 2007; YAO et al., 2015). Estas regiões IRs são bem conservadas em cloroplastos (CURTIS; CLEGG, 1984; WOLFE; LI; SHARP, 1987), principalmente em razão da quantidade de RNA ribossomal presentes nessas regiões (PALMER, 1985). As mutações detectadas nesse estudo confirmam essa conservação, visto que as substituições nucleotídicas ocorreram na região LSC do genoma cloroplastidial. O gene *psbA* é conservado em todos cloroplastos de cana-de-açúcar, e parece ter sofrido uma seleção positiva, este gene codifica para uma subunidade central do fotossistema II, conforme relatado para sorgo (WU et al., 1999), esse fotossistema II é de fundamental importância, pois ele é um dos principais responsáveis pela liberação de oxigênio na atmosfera terrestre por parte das cianobactérias, algas e plantas (POSPÍŠIL, 2012). Essas análises comparativas contribuem para agregar informações sobre o DNA organelar no gênero *Saccharum*, suas divergências e similaridades para que essas informações possam ser utilizadas em marcadores moleculares, ajudando a compreender os polimorfismos existentes e correlaciona-los com a hibridização no gênero.

6. CONCLUSÕES

- A espécie selvagem *S. officinarum* apresentou em seu genoma cloroplastidial os genes *ycf1*, *ycf68*, *ycf73*, *ycf15* e *ycf2*, enquanto que *S. spontaneum* e os cultivares híbridos não apresentaram tais genes, que podem ter se perdido ao longo da evolução deste gênero;
- O genoma de cloroplasto do selvagem *S. officinarum* apresentou o pseudogene *ACRS* que confere sensibilidade à toxina ACR relacionado à doença da mancha marrom, sugerindo que este gene de origem mitocondrial teria se translocado para o genoma de cloroplasto, se tornando um pseudogene;
- As mitocôndrias de cana-de-açúcar possuem o conteúdo CG mais conservado do que os genomas de cloroplasto;
- Todos os genomas de mitocôndrias no cromossomo 1 e 2 mantiveram o mesmo número de CDS e RNAs ribossomais, o que caracteriza a conservação do número de genes nas mitocôndrias do gênero *Saccharum*;
- O genomas mitocondriais de cana-de-açúcar apresentaram o RNA transportador *TrnL-CAA* que foi perdido em sorgo. Além disso, apresentaram dois RNAt (*TrnF-*

RAA e TrnF- GAA) a mais do que o genoma mitocondrial publicado da cultivar Khon Kaen 3;

- Mitocôndrias e cloroplastos apresentaram menor conservação de sequência nucleotídica nas regiões intergênicas do genoma.
- As espécies selvagens compartilham similaridades de sequências genômicas, assim marcadores moleculares específicos para o genoma cloroplastidial e mitocondrial de selvagens de cana-de-açúcar podem ser desenvolvidos;
- Das repetições analisadas nos genomas o tipo microsatélites (SSRs) é o mais frequente, o genoma de cloroplasto é composto por 46% e o genoma mitocondrial 65% desses SSRs;
- Os motivos de SSRs mais comuns detectados para o genoma de cloroplasto foram os de tetra-nucleotídeos e para o genoma mitocondrial os mono-nucleotídeos no cromossomo 1 e os tetra-nucleotídeos no cromossomo 2;
- Todos os indels existentes no genoma de cloroplasto estão localizados na região LSC, os genes afetados por estas substituições foram *rpoC2* e *ycf1*, nas mitocôndrias do gênero *Saccharum* o gene *rps4* foi o principal alvo de mutações;
- A análise das fronteiras IRs de cloroplastos mostrou que todos os genes (*rpl22*, *rps19*, *ndhF*, *ndhH* e *rps15*) com exceção do *psbA*, sofreram expansão e compressão de tamanho entre as espécies selvagens e cultivares híbridos de cana-de-açúcar.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMS, K. L.; PALMER, J. D. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. **Molecular Phylogenetics and Evolution**, v. 29, n. 3, p. 380–395, 2003.
- AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. **Theoretical and Applied Genetics**, v. 110, n. 5, p. 789–801, 2005.
- ALLEN, J. F. et al. A structural phylogenetic map for chloroplast photosynthesis. **Trends in Plant Science**, v. 16, n. 12, p. 645–655, 2011.
- ALTSCHUL, S. F. et al. Basic Local Alignment Search Tool 2Department of Computer Science. **J Mol Biol**, v. 215, n. 3, p. 403–410, 1990.
- ALVERSON, A. J. et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). **Molecular Biology and Evolution**, v. 27, n. 6, p. 1436–1448, 2010.
- ALVERSON, A. J. et al. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. **PLoS ONE**, v. 6, n. 1, 2011.
- ANDERSON, S. et al. Sequence and organization of the human mitochondrial genome. **Nature**, v. 290, n. 5806, p. 457–465, 1981.
- ANDRÉS-COLÁS, N. et al. Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. **Proceedings of the National Academy of Sciences**, v. 114, n. 33, p. 201705815, 2017.
- ANDRIETTA, M. G. S.; ANDRIETTA, S. R.; STECKELBERG, C.; STUPIELLO, E. N. A. A. Bioethanol - Brazil, 30 years of Proálcool. v. 109, n. 1299, p. 195–200, 2007.
- APUYA, N. R. et al. RASPBERRY3 Gene Encodes a Novel Protein Important for Embryo Development. v. 1606, 2002.
- ASAF, S. et al. The Complete Chloroplast Genome of Wild Rice (*Oryza minuta*) and Its Comparison to Related Species. **Frontiers in Plant Science**, v. 8, n. March, p. 1–15, 2017a.
- ASAF, S. et al. Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmaifera* and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. **Scientific Reports**, v. 7, n. 1, p. 7556, 2017b.

- ASANO, T. et al. Complete Nucleotide Sequence of the Sugarcane (*Saccharum*. v. 99, p. 93–99, 2004.
- ASHRAFI, G.; SCHWARZ, T. L. The pathways of mitophagy for quality control and clearance of mitochondria. **Cell Death and Differentiation**, v. 20, n. 1, p. 31–42, 2013.
- AVERY, O. T. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation By a Desoxyribonucleic Acid Fraction Isolated From *Pneumococcus* Type Iii. **Journal of Experimental Medicine**, v. 79, n. 2, p. 137–158, 1944.
- BANKEVICH, A. et al. and Its Applications to Single-Cell Sequencing. v. 19, n. 5, p. 455–477, 2012.
- BARR, C. M.; NEIMAN, M.; TAYLOR, D. R. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. **New Phytologist**, v. 168, n. 1, p. 39–50, 2005.
- BASTOS, E. Cana-de-açúcar, o verde mar de energia. **São Paulo, Ícone**, p. 127 p., 1987.
- BEIER, S. et al. MISA-web: a web server for microsatellite prediction. **Bioinformatics (Oxford, England)**, v. 33, n. 16, p. 2583–2585, 2017.
- BENDICH, A. J. The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. **BioEssays**, v. 29, n. 5, p. 474–483, 2007.
- BENSON, D. A. et al. GenBank. **Nucleic Acids Research**, v. 41, n. D1, p. 36–42, 2012.
- BENSON, G. Tandem repeats finder : a program to analyze DNA sequences. v. 27, n. 2, p. 573–580, 1999.
- BI, C. et al. Analysis of the Complete Mitochondrial Genome Sequence of the Diploid Cotton *Gossypium raimondii* by Comparative Genomics Approaches. v. 2016, 2016.
- BIBB, M. J. et al. Sequence and gene organization of mouse mitochondrial DNA. **Cell**, v. 26, n. 2 PART 2, p. 167–180, 1981.
- BISON, O. et al. Inbreeding depression in Eucalyptus clones. **Crop Breeding and Applied Biotechnology**, v. 4, n. 4, p. 459–464, 2004.
- BLATTNER, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. **Science**, v. 277, n. 5331, p. 1453–1462, 1997.
- BOETZER, M. et al. Scaffolding pre-assembled contigs using SSPACE. **Bioinformatics**, v. 27, n. 4, p. 578–579, 2011.
- BONEN, L.; BIRD, S. Sequence analysis of the wheat mitochondrial *atp6* gene reveals

- a fused upstream reading frame and markedly divergent N termini among plant ATP6 proteins. **Gene**, v. 73, n. 1, p. 47–56, dez. 1988.
- BRITTEN, R. J. et al. Majority of divergence between closely related DNA samples is due to indels. **Proceedings of the National Academy of Sciences**, v. 100, n. 8, p. 4661–4665, 2003.
- BRYANT, N. et al. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. **Plant Physiol**, v. 155, n. 4, p. 1678–1689, 2011.
- CAI, Q. et al. Verification of the introgression of Erianthus arundinaceus germplasm into sugarcane using molecular markers. **Plant Breeding**, v. 124, n. 4, p. 322–328, ago. 2005.
- CALSA JÚNIOR, T. et al. Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. **Current Genetics**, v. 46, n. 6, p. 366–373, 2004.
- CARDOSO-SILVA, C. B. et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. **PLoS ONE**, v. 9, n. 2, 2014.
- CHANG, S. et al. The Mitochondrial Genome of Soybean Reveals Complex Genome Structures and Gene Evolution at Intercellular and Phylogenetic Levels. **PLoS ONE**, v. 8, n. 2, 2013.
- CHEAVEGATTI-GIANOTTO, A. et al. Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. **Tropical plant biology**, v. 4, n. 1, p. 62–89, mar. 2011.
- CHEN, H. et al. RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. **Journal of Integrative Plant Biology**, v. 53, n. 12, p. 961–970, 2011.
- CHEN, Z. J. Molecular mechanisms of polyploidy vigor. **Trends in plant scienceends.**, v. 15, n. 2, p. 1–28, 2010.
- CHUMLEY, T. W. et al. The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. **Molecular Biology and Evolution**, v. 23, n. 11, p. 2175–2190, 2006.
- CLIFTON, S. W. et al. Sequence and Comparative Analysis of the Maize NB Mitochondrial Genome. **Plant Physiology**, v. 136, n. 3, p. 3486–3503, 2004.
- COHEN, S. N. et al. Construction of Biologically Functional Bacterial Plasmids In

- Vitro. **Proceedings of the National Academy of Sciences**, v. 70, n. 11, p. 3240–3244, 1973.
- CONESA, A. et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674–3676, 2005.
- CURTIS, S. E.; CLEGG, M. T. Molecular evolution of chloroplast DNA sequences. **Molecular biology and evolution**, v. 1, n. 4, p. 291–301, jul. 1984.
- CUSHING, D. A. et al. Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. p. 424–436, 2005.
- DAL-BIANCO, M. et al. Sugarcane improvement: How far can we go? **Current Opinion in Biotechnology**, v. 23, n. 2, p. 265–270, 2012.
- DANIELS, J.; ROACH, B. **Taxonomy and evolution. In: Heinz DJ (ed) Sugarcane improvement through breeding.** [s.l: s.n.].
- DELUCA AND O'FARREL. Barriers to Male Transmission of Mitochondrial. **Dev. Cell**, v. 22, n. 3, p. 611–620, 2012.
- DIAS, L. A. D. S. Biofuel plant species and the contribution of genetic improvement. **Crop Breeding and Applied Biotechnology**, v. 11, n. spe, p. 16–26, 2011.
- DIEKMANN, K. et al. Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). **Dna research**, v. 16, n. 3, p. 165–176, 2009.
- DOMBROVSKA, O.; QIU, Y. L. Distribution of introns in the mitochondrial gene nad1 in land plants: Phylogenetic and molecular evolutionary implications. **Molecular Phylogenetics and Evolution**, v. 32, n. 1, p. 246–263, 2004.
- DUMOLIN-LAPEGUE, S.; PEMONGE, M.-H.; PETIT, R. J. An enlarged set of consensus primers for the study of organelle DNA in plants. **Molecular Ecology**, v. 6, n. 4, p. 393–397, 1997.
- EDME, S. J. et al. Genetic Contribution to Yield Gains in the Florida Sugarcane Industry across 33 Years. **Crop Science Society of America**, v. 45, n. 3, p. 92–97, 2001.
- EVANS, D. L.; JOSHI, S. V. Complete chloroplast genomes of *Saccharum spontaneum*, *Saccharum officinarum* and *Miscanthus floridulus* (Panicoideae: Andropogoneae) reveal the plastid view on sugarcane origins. **Systematics and Biodiversity**, v. 14, n. 6, p. 548–571, 2016a.

- EVANS, D. L.; JOSHI, S. V. Complete chloroplast genomes of *Saccharum spontaneum*, *Saccharum officinarum* and *Miscanthus floridulus* (Panicoideae: Andropogoneae) reveal the plastid view on sugarcane origins. **Systematics and Biodiversity**, v. 2000, n. June, p. 1–24, 2016b.
- FEDOROFF, N. V. The past, present and future of crop genetic modification. **New Biotechnology**, v. 27, n. 5, p. 461–465, 2010.
- FERRARI, F. Caracterização cromossômica em Cana-de-açúcar (*Saccharum* spp., Poaceae). p. 91, 2010.
- FINNEGAN, E. J.; PEACOCK, W. J.; DENNIS, E. S. DNA methylation, a key regulator of plant development and other processes. **Elsevier Science Ltd.**, p. 217–223, 2000.
- FLEISCHMANN, R. D. et al. Whole genome random sequencing and assembly of *Haemophilus influenzae* rd. **Science**, v. 269, n. 5223, p. 496–512, 1995.
- FORTUCCI, M. P.; DIVISION, T. **International Sugar Conference**. Havana, Cuba: 1999
- FRANCISCO, Q. D. S. a Agroindústria Canavieira De Campos-Rj E Os Royalties Do Petróleo a Partir Da Percepção Dos Atores. 2009.
- FRAZER, K. A. et al. VISTA: Computational tools for comparative genomics. **Nucleic Acids Research**, v. 32, n. WEB SERVER ISS., p. 273–279, 2004.
- GARSMEUR, O. et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. **Nature Communications**, v. 9, n. 1, p. 2638, 2018.
- GASPAR, C. G. **EFEITOS DA HIBRIDIZAÇÃO NO GENOMA DA CANA-DE-AÇÚCAR**. [s.l.] Universidade Federal do Rio de Janeiro, 2014.
- GLASER, P. et al. *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333deg; **Molecular Microbiology**, v. 10, n. 2, p. 371–384, 1993.
- GOCKEL, G.; HACHTEL, W.; MELKONIAN, M. Complete gene map of the plastid genome of the nonphotosynthetic Euglenoid flagellate *astasia longa*. **Protist**, v. 151, n. 4, p. 347–351, 2000.
- GRAHAM, S. W. et al. Microstructural Changes in Noncoding Chloroplast DNA: Interpretation, Evolution, and Utility of Indels and Inversions in Basal Angiosperm Phylogenetic Inference. **International Journal of Plant Sciences**, v. 161, n. S6, p. S83–S96, nov. 2000.
- GRATIVOL, C. et al. Sugarcane genome sequencing by methylation filtration provides

- tools for genomic research in the genus *Saccharum*. **The Plant Journal**, v. 79, n. 1, p. 162–172, 2014.
- GRAY, M. W. et al. Irremediable complexity? **Science**, v. 330, n. 6006, p. 920–921, 2010.
- GRAY, M. W.; BURGER, G.; LANG, B. F. Mitochondrial evolution. **Science**, v. 283, n. 5407, p. 1476–1481, 1999.
- GREEN, E. D. Strategies for the systematic sequencing of complex genomes. **Nature Reviews Genetics**, v. 2, n. 8, p. 573–583, 1 ago. 2001.
- GREINER, S. et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. **Nucleic Acids Research**, v. 36, n. 7, p. 2366–2378, 2008.
- GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current opinion in plant biology**, v. 5, n. 2, p. 122–7, abr. 2002.
- GRIVET, L.; GLASZMANN, J.; D'HONT, A. Molecular evidence of sugarcane evolution and domestication. **Darwin's Harvest, New Approaches to the Origins, Evolution and Conservation of Crops**, p. 49–66, 2006.
- GUALBERTO, J. M. et al. The plant mitochondrial genome: Dynamics and maintenance. **Biochimie**, v. 100, n. 1, p. 107–120, 2014.
- GUILLAUMOT, D. et al. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. **Proceedings of the National Academy of Sciences**, p. 201705780, 2017.
- HAHN, C.; BACHMANN, L.; CHEVREUX, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. **Nucleic Acids Research**, v. 41, n. 13, 2013.
- HANDA, H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. **Nucleic Acids Research**, v. 31, n. 20, p. 5907–5916, 2003.
- HANDA, H.; KUBO, K. KADOWAKI, N. Genes for the ribosomal S4 protein encoded in higher plant mitochondria are transcribed, edited, and translated. **Molecular and General Genetics MGG**, v. 258, n. 3, p. 199–207, 22 maio 1998.
- HARRIS, S. A.; INGRAM, R. Chloroplast DNA and Biosystematics: The Effects of Intraspecific Diversity and Plastid Transmission. **Taxon**, v. 40, n. 3, p. 393–412, 1991.

- HATTORI, N. et al. Mitochondrial DNA heteroplasmy in wheat, aegilops and their nucleus-cytoplasm hybrids. **Genetics**, v. 160, n. 4, p. 1619–1630, 2002.
- HAVEY, M. J. Predominant Paternal Transmission of the Mitochondrial Genome in Cucumber. **J Hered**, v. 88, n. 3, p. 232–235, 1997.
- HISANO, H. et al. Mitochondrial genome sequences from wild and cultivated barley (*Hordeum vulgare*). **BMC Genomics**, v. 17, n. 1, p. 1–12, 2016.
- HOANG, N. V. et al. Next generation sequencing of total DNA from sugarcane provides no evidence for chloroplast heteroplasmy. **New Negatives in Plant Science**, v. 1-2, p. 33–45, 2015.
- HOANG, N. V. et al. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. **BMC Genomics**, v. 18, n. 1, p. 1–22, 2017.
- HOFFMAN ET AL. Variedades RB de Cana-de-açúcar. **Departamento de Biotecnologia Vegetal de Melhoramento da Cana-de-açúcar**, v. 1, n. UFSCar, p. 30p, 2008.
- HONMA, Y. et al. Polymorphic minisatellites in the mitochondrial DNAs of *Oryza* and *Brassica*. **Current Genetics**, v. 57, n. 4, p. 261–270, 2011.
- IBGE. Indicadores IBGE. **Estática de produção agrícola**, p. 1–76, 2017.
- INTERNATIONAL ENERGY AGENCY. Biofuels for transport. **An Internacional Perspective**, p. 1–216, 2004.
- ITCHODA, N. et al. The rps4 gene in sugar beet mitochondria: Insertion/deletion mutations occur within the gene but do not disrupt the reading frame. **Journal of Plant Physiology**, v. 159, n. 2, p. 211–215, 2002.
- JACKSON, D. A.; SYMONS, R. H.; BERG, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 69, n. 10, p. 2904–2909, 1972.
- JANSEN, R. K. et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 49, p. 19369–19374, 2007.
- JOYCE, P. B.; GRAY, M. W. Chloroplast-like transfer RNA genes expressed in wheat mitochondria. **Nucleic acids research**, v. 17, n. 14, p. 5461–76, 25 jul. 1989.

- JUNG, J. et al. AGORA: organellar genome annotation from the amino acid and nucleotide references. **Bioinformatics**, v. 34, n. 15, p. 2661–2663, 2018.
- KELCHNER, S. A. The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics. **Annals of the Missouri Botanical Garden**, v. 87, n. 4, p. 482, 2000.
- KODE, V. et al. The tobacco plastid accD gene is essential and is required for leaf development. **Plant Journal**, v. 44, n. 2, p. 237–244, 2005.
- KONCZ, C. et al. Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. **The EMBO journal**, v. 9, n. 5, p. 1337–46, 1990.
- KULKARNI, N. H. et al. The DSC1 channel, encoded by the smi60E locus, contributes to odor-guided behavior in *Drosophila melanogaster*. **Genetics**, v. 161, n. 4, p. 1507–1516, 2002.
- KUNTAL, H.; SHARMA, V. *In Silico* Analysis of SSRs in Mitochondrial Genomes of Plants. **OMICS: A Journal of Integrative Biology**, v. 15, n. 11, p. 783–789, 2011.
- LANDER, E. S. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 2001.
- LANG, B. F. et al. **An ancestral mitochondrial DNA resembling a eubacterial genome in miniature** *Nature*, 1997.
- LANG, B. F.; GRAY, M. W.; BURGER, G. Mitochondrial Genome Evolution And The Origin Of Eukariotes. **Analysis**, v. 33, n. 1, p. 351–397, 1999.
- LAW, J. A.; JACOBSEN, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. **Nature Rev Genet**, v. 11, n. 3, p. 204–220, 2010.
- LEITE, R. C. DE C.; LEAL, M. R. L. V. O biocombustível no brasil. **Novos Estudos - CEBRAP**, v. 78, p. 15–21, 2007.
- LEVY, J. Sequencing the yeast genome: An international achievement. **Yeast**, v. 10, n. 13, p. 1689–1706, 1994.
- LI, L. Y.; LUO, X.; WANG, X. Endonuclease G is an apoptotic DNase when released from mitochondria. **Nature**, v. 412, n. 6842, p. 95–99, 2001.
- LI, Z. et al. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. **Briefings in Functional Genomics**, v. 2, n. 1, p. 25–37, 1 jan. 2011.
- LI, Z. et al. Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. **Briefings in Functional Genomics**, v. 11, n. 1,

p. 25–37, 2012.

LIN, T. P. et al. Evidence for the existence of some dissociation in an otherwise strong linkage disequilibrium between mitochondrial and chloroplastic genomes in *Cyclobalanopsis glauca*. **Molecular Ecology**, v. 12, n. 10, p. 2661–2668, 2003.

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, 2007.

LONSDALE, D. M.; HODGE, T. P.; FAURON, C. M. © IRLPresi Limited , Oxford , England . *Nucleic Acids Research*. v. 12, n. 24, p. 9249–9262, 1984.

LOWE, T. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Research**, v. 25, n. 5, p. 955–964, 1997.

LUO, S. et al. Biparental Inheritance of Mitochondrial DNA in Humans. **Proceedings of the National Academy of Sciences**, p. 201810946, 2018.

LYNCH, M.; KOSKELLA, B.; SCHAACK, S. Mutation Pressure and the Evolution of. **Science**, v. 311, n. March, p. 1727–1731, 2006.

MA, P. F.; GUO, Z. H.; LI, D. Z. Rapid sequencing of the bamboo mitochondrial genome using illumina technology and parallel episodic evolution of organelle genomes in grasses. **PLoS ONE**, v. 7, n. 1, 2012.

MACEDO, I. C.; SEABRA, J. E. A.; SILVA, J. E. A. R. Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020. **Biomass and Bioenergy**, v. 32, n. 7, p. 582–595, 2008.

MALLET, J. Hybridization as an invasion of the genome. **Trends in Ecology and Evolution**, v. 20, n. 5, p. 229–237, 2005.

MARIAC, C. et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. **Molecular Ecology Resources**, v. 14, n. 6, p. 1103–1113, 2014.

MARTIN, W. et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. **Proceedings of the National Academy of Sciences**, v. 99, n. 19, p. 12246–12251, 2002.

MATSUOKA, S. et al. The Brazilian Experience of Sugarcane Ethanol Industry. **Springer**, v. 45, n. 3, p. 372–381, 2009.

MAXAM, A M.; GILBERT, W. A new method for sequencing DNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 2, p.

560–4, 1977.

MENDEL, G. Versuche Über Pflanzen-Hybriden. **Verh. Naturf. Ver. in Brünn**, v. 10, p. Brünn 4, 3–47, 1865.

MILLAR, A. H. et al. Mitochondrial Biogenesis and Function in Arabidopsis †. **The Arabidopsis Book**, v. 6, p. e0111, 2008.

MILLEN, R. S. Many Parallel Losses of *infA* from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus. **the Plant Cell Online**, v. 13, n. 3, p. 645–658, 2001.

MINISTÉRIO DA AGRICULTURA, P. E. A. Açúcar e álcool no Brasil. **Departamento de Cana-de-açúcar e Agroenergia**, p. 1–21, 2007.

MOGENSEN, H. L. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383L... v. 83, n. 3, p. —, 1996.

MOORE, P. H. Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum. **Field Crops Research**, v. 92, n. 2-3, p. 119–135, jun. 2005.

MORSCHBACKER, A. Bio-ethanol based ethylene. **Polymer Reviews**, v. 49, n. 2, p. 79–84, 2009.

MUKHERJEE, S. K. Origin and Distribution of *Saccharum*. **Botanical Gazette**, v. 119, n. 1, p. 2473264, 1957.

MULLIGAN, R. M.; CHANG, K. L. C.; CHOU, C. C. Computational analysis of RNA editing sites in plant mitochondrial genomes reveals similar information content and a sporadic distribution of editing sites. **Molecular Biology and Evolution**, v. 24, n. 9, p. 1971–1981, 2007.

NCBI. **National Center for Biotechnology Information**. Disponível em:

<<https://www.ncbi.nlm.nih.gov/>>. Acesso em: 29 ago. 2018.

NGERNPRASIRTSIRI, J. et al. DNA methylation and the differential expression of C4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. **Journal of Biological Chemistry**, v. 264, n. 14, p. 8241–8248, 1989.

NIRENBERG, M. W. The genetic. **Sci Am**, v. 208, p. 80–94, 1963.

NOSEK, J.; TOMÁŠKA, L. Mitochondrial genome diversity: Evolution of the molecular architecture and replication strategy. **Current Genetics**, v. 44, n. 2, p. 73–84, 2003.

NOTSU, Y. et al. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering

- plants. **Molecular Genetics and Genomics**, v. 268, n. 4, p. 434–445, 2002.
- OHTANI, K.; YAMAMOTO, H.; AKIMITSU, K. Sensitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. **Proceedings of the National Academy of Sciences**, v. 99, n. 4, p. 2439–2444, 2002.
- OHYAMA ET AL. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. **Nature**, v. 322, p. 698–699, 1986.
- OLMSTEAD, R.-G.; PALMER, J.-D. Chloroplast DNA systematics: A review of methods and data analysis. **American Journal of Botany**, v. 81, n. 9, p. 1205–1224, 1994.
- ORSAT, B. et al. Mapping and sequencing of an actively transcribed *Euglena gracilis* chloroplast gene (*ccsA*) homologous to the *Arabidopsis thaliana* nuclear gene *cs(ch-42)*. **FEBS Letters**, v. 303, n. 2-3, p. 181–184, 1992.
- PALMER, J. D. Comparative Organization of Chloroplast Genomes. **Annual Review of Genetics**, v. 19, n. 1, p. 325–354, 1985.
- PALMER, J. D.; HERBON, L. A. Unicircular structure of the *Brassica hirta* mitochondrial genome. **Current Genetics**, v. 11, n. 6-7, p. 565–570, 1987.
- PALMER, J. D.; HERBON, L. A. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. **Journal of Molecular Evolution**, v. 28, n. 1-2, p. 87–97, 1988.
- PALOMINO, E. C. et al. Genetic diversity of common bean genotypes of Carioca commercial group using RAPD markers. p. 80–85, 2005.
- PAN, Y. B. et al. Molecular genotyping of sugarcane clones with microsatellite DNA markers. **Maydica**, v. 48, n. 4, p. 319–329, 2003.
- PANJE, R. R.; BABU, C. N. Studies in *Saccharum spontaneum* Distribution and Geographical Association Chromosome Numbers. **Cytologia**, v. 25, p. 152–172, 1960.
- PARKS, M.; CRONN, R.; LISTON, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. **BMC Biology**, v. 7, p. 1–17, 2009.
- PHILLIPPY, A. M.; SCHATZ, M. C.; POP, M. Genome assembly forensics: Finding the elusive mis-assembly. **Genome Biology**, v. 9, n. 3, 2008.
- PIOT, A. et al. One-third of the plastid genes evolved under positive selection in PACMAD grasses. **Planta**, v. 247, n. 1, p. 255–266, 2018.
- POP, M. et al. Comparative genome assembly. **Briefings in Bioinformatics**, v. 5, n. 3, p. 237–248, 2004.

- POSPÍŠIL, P. Molecular mechanisms of production and scavenging of reactive oxygen species by photosystem II. **Biochimica et Biophysica Acta - Bioenergetics**, v. 1817, n. 1, p. 218–231, 2012.
- PRICE, S. A. M. Cytological studies in saccharum and allied genera. viii. f2 and bc1 progenies from 112- and 136-chromosome s. officinarum x s. spontaneum hybrids1. p. 186–190, 1955.
- PROVAN, J.; MURPHY, S.; MAGGS, C. A. Universal plastid primers for Chlorophyta and Rhodophyta. **European Journal of Phycology**, v. 39, n. 1, p. 43–50, 2004.
- QIU, Y. L. et al. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. **Journal of Systematics and Evolution**, v. 48, n. 6, p. 391–425, 2010.
- RAMAN, G.; PARK, S. J. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. **PLoS ONE**, v. 10, n. 10, p. 1–21, 2015.
- RAUBESON, L. A. et al. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. **BMC Genomics**, v. 8, 2007.
- RAVEENDAR, S. et al. The Complete Chloroplast Genome of *Capsicum annuum* var. *glabriusculum* Using Illumina Sequencing. **Molecules**, v. 20, n. 7, p. 13080–13088, 2015.
- RAYAPURAM, N. et al. The three mitochondrial encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in Arabidopsis. **Journal of Biological Chemistry**, v. 283, n. 37, p. 25200–25208, 2008.
- RIAÑO-PACHÓN, D. M.; MATTIELLO, L. Draft genome sequencing of the sugarcane hybrid SP80-3280. **F1000Research**, v. 6, n. 0, p. 861, 2017.
- ROBERT A. MARTIENSSEN; VINCENT COLOT. DNA Methylation and Epigenetic Inheritance in Plants and Filamentous Fungi. **Science**, v. 293, n. August, p. 1070–1075, 2001.
- ROBERTIS, E. M. F. DE. Bases da Biologia Celular e Molecular. p. 389, 2001.
- RODRIGUES, M. C. et al. Comparison among Inbreeding Systems in Maize. **Crop Breeding and Applied Biotechnology**, v. 1, n. 2, p. 105–114, 2001.
- RODRIGUES, M. G. F. et al. Detection of differentially methylated regions of irradiated fig tree selections. **Scientia Agricola**, v. 74, n. 4, p. 285–293, 2017.
- RODRÍGUEZ-MORENO, L. et al. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome

- in plants contains a significant amount of DNA having a nuclear origin. **BMC Genomics**, v. 12, n. 1, p. 424, 2011.
- ROSA SAMBUICHI, R. H. . . [ET AL.]. **Políticas agroambientais e sustentabilidade : desafios, oportunidades e lições aprendidas**. [s.l: s.n.].
- SAKAMOTO, W.; MIYAGISHIMA, S.; JARVIS, P. Chloroplast Biogenesis: Control of Plastid Development, Protein Import, Division and Inheritance. **The Arabidopsis Book**, v. 6, p. e0110, 2008.
- SALLY, L. D. et al. Domestication to Crop Improvement : Genetic Resources for Sorghum and S accharum (Andropogoneae). **Annals of Botany**, v. 100, p. 975–989, 2007.
- SANGER, F. et al. Nucleotide sequence of bacteriophage phi X174 DNA. **Nature**, v. 265, n. 5596, p. 687–695, 1977.
- SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal of Molecular Biology**, v. 94, n. 3, p. 441–448, 1975.
- SANTA BRIGIDA, A. B. et al. Sugarcane transcriptome analysis in response to infection caused by *Acidovorax avenae* subsp. *avenae*. **PLoS ONE**, v. 11, n. 12, p. 1–30, 2016.
- SASKI, C. et al. 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.
- SATO, S. et al. Complete Structure of the Chloroplast Genome of *thaliana*. **DNA research**, v. 290, p. 283–290, 1999.
- SHAHID MASOOD, M. et al. The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: First genome wide comparative sequence analysis of wild and cultivated rice. **Gene**, v. 340, n. 1, p. 133–139, 2004.
- SHANTZ, H. . L. . The Place of Grasslands in the Earth ' s Cover. v. 35, n. 2, p. 143–145, 1954.
- SHEARMAN, J. R. et al. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. **Scientific Reports**, v. 6, n. August, p. 31533, 2016.
- SHINOZAKI, K. et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **EMBO Journal**, v. 5, n. 9, p. 2043–2049, 1986.

- SILVA, M. et al. Genetic diversity among sugarcane clones (*Saccharum spp.*). 2005.
- SILVA, O. P. 500 anos dos Campos dos Goytacazes. **Fundação Cultural Jornalista Oswaldo Lima**, v. 1, p. 400, 2004.
- SLOAN, D. B. et al. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. **BMC Evolutionary Biology**, v. 10, n. 1, 2010.
- SLOAN, D. B. et al. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. **PLoS Biology**, v. 10, n. 1, p. e1001241, 17 jan. 2012.
- SMITH, D. R. Updating our view of organelle genome nucleotide landscape. **Frontiers in Genetics**, v. 3, n. SEP, p. 1–10, 2012.
- SMITH, L. M. et al. Fluorescence detection in automated DNA sequence analysis. **Nature**, v. 321, n. 6071, p. 674–679, 1986.
- SOLTIS, P. S.; SOLTIS, D. E. The Role of Hybridization in Plant Speciation. **Annual Review of Plant Biology**, v. 60, n. 1, p. 561–588, 2009.
- SOUSA, H. **Cyclo Aureo - História do 1º Centenário de Campos**. Memórias f ed. [s.l: s.n.].
- STADERMANN, K. B.; HOLTGRÄWE, D.; WEISSHAAR, B. Chloroplast Genome Sequence of *Arabidopsis thaliana* Accession Landsberg erecta , Assembled from Single-Molecule, Real-Time Sequencing Data. **Genome Announcements**, v. 4, n. 5, p. 4–5, 27 out. 2016.
- STEFAN KURTZ, JOMUNA V. CHOUDHURI, ENNO OHLEBUSCH, CHRIS SCHLEIERMACHER, JENS STOYE, R. G. REPuter: the manifold applications of repeat analysis on a genomic scale. v. 29, n. 0, p. 1–32, 2001.
- STOEBE, B.; KOWALLIK, K. V. Gene-cluster analysis in chloroplast genomics. **Trends in Genetics**, v. 15, n. 9, p. 344–347, 1999.
- SULSTON, J. et al. The *C. elegans* genome sequencing project: a beginning. **Nature**, v. 356, n. 6364, p. 37–41, 1992.
- SZMRECSÁNYI, T.; MOREIRA, E. P. O desenvolvimento da agroindústria canavieira do Brasil desde a Segunda Guerra Mundial. **Estudos Avançados**, v. 5, n. 11, p. 57–79, 1991.
- TABERLET, P. et al. Universal primer for amplification of three non-coding regions of chloroplast DNA. **Plant Molecular Biology**, v. 17, n. 5, p. 1105–1109, 1991.
- TANG, M. et al. Multiplex sequencing of pooled mitochondrial genomes - A crucial

- step toward biodiversity analysis using mito-metagenomics. **Nucleic Acids Research**, v. 42, n. 22, 2014.
- TÁVORA, F. L. História e Economia dos Biocombustíveis no Brasil. **Economia**, 2011.
- TESTER, M.; LANGRIDGE, P. Breeding Technologies to Increase Crop Production in a Changing World. **Science**, v. 327, n. 5967, p. 818–822, 2010.
- THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, v. 408, n. 6814, p. 796–815, 2000.
- TILLICH, M. et al. GeSeq - Versatile and accurate annotation of organelle genomes. **Nucleic Acids Research**, v. 45, n. W1, p. W6–W11, 2017.
- TIMMIS, J. N. et al. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. **Nature Reviews Genetics**, v. 5, n. 2, p. 123–135, 2004.
- VAILLANCOURT, R. E.; PETTY, A.; MCKINNON, G. E. Maternal inheritance of mitochondria in *Eucalyptus globulus*. **Journal of Heredity**, v. 95, n. 4, p. 353–355, 2004.
- VAN DINGENEN, J. et al. Plants grow with a little help from their organelle friends. **Journal of Experimental Botany**, v. 67, n. 22, p. 6267–6281, 2016.
- VETTORE, A. L. et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, v. 13, n. 12, p. 2725–2735, 2003.
- VIDIGAL, P. M. P. et al. Complete Chloroplast Genome Sequence and Annotation of the *Saccharum* hybrid cultivar RB867515. **Genome Announc**, v. 4, n. 5, p. e01157–16, 2016.
- WAKASUGI, T. et al. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 21, p. 9794–9798, 1994.
- WANG, K.; KLIONSKY, D. J. Mitochondria removal by autophagy. **Autophagy**, v. 7, n. 3, p. 297–300, 2011.
- WANG, W.; CHEN, S.; ZHANG, X. Whole-Genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (*Bambusoideae: Olyreae*). **Molecules**, v. 23, n. 7, p. 1–20, 2018.
- WANG, Z. et al. High-throughput m6A-seq reveals RNA m6A methylation patterns in the chloroplast and mitochondria transcriptomes of *Arabidopsis thaliana*. **PLoS ONE**, v. 12, n. 11, p. 1–24, 2017.

- WARD, B. L.; ANDERSON, R. S.; BENDICH, A. J. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). **Cell**, v. 25, n. 3, p. 793–803, 1981.
- WATSON, J. D.; CRICK, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. **Nature**, v. 171, n. 4356, p. 737–738, 1953.
- WEISING, K.; GARDNER, R. C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. **Genome**, v. 42, n. 1, p. 9–19, 1999.
- WOLFE, K. H.; LI, W.-H.; SHARP, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. **Proc. Natl. Acad. Sci. USA**, v. 84, n. 24, p. 9054–9058, 1987.
- WOLFE, K. H., MORDEN, C. W. & PALMER, J. D. function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. **Proc. Natl. Acad. Sci. USA**, v. 89, n. 16, p. 10648–10652, 1992.
- WORLD WATCH INSTITUTE. State of the World 2006 – Special focus: China and India. 2006.
- WU, C. S. et al. Comparative chloroplast genomes of pinaceae: Insights into the mechanism of diversified genomic organizations. **Genome Biology and Evolution**, v. 3, n. 1, p. 309–319, 2011.
- WU, N. H. et al. Structural feature of sorghum chloroplast psbA gene and regulation effects of its 5'-noncoding region. **Science in China Series C-Life Sciences**, v. 42, n. 4, p. 383–394, 1999.
- WU, Y. et al. Comparative Chloroplast Genomics of Gossypium Species: Insights Into Repeat Sequence Variations and Phylogeny. **Frontiers in Plant Science**, v. 9, n. March, p. 1–14, 2018.
- XIAO-MING, Z. et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. **Scientific Reports**, v. 7, n. 1, p. 1555, 2017.
- XIU, Z. et al. EMPTY PERICARP16 is required for mitochondrial nad2 intron 4 cis-splicing, complex i assembly and seed development in maize. **Plant Journal**, v. 85, n. 4, p. 507–519, 2016.
- XU, S. et al. Transcriptomic characterization and potential marker development of contrasting sugarcane cultivars. **Scientific Reports**, v. 8, n. 1, p. 1–11, 2018.
- YANG, Y. et al. Comparative Analysis of the Complete Chloroplast Genomes of Five

- Quercus Species. **Frontiers in Plant Science**, v. 07, n. June, p. 959, 2016.
- YAO, X. et al. The first complete chloroplast genome sequences in Actinidiaceae: Genome structure and comparative analysis. **PLoS ONE**, v. 10, n. 6, p. 1–14, 2015.
- YAO, X. et al. Chloroplast genome structure in Ilex (Aquifoliaceae). **Scientific reports**, v. 6, n. April, p. 28559, 2016.
- ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v. 18, n. 5, p. 821–829, 2008.
- ZHANG, J. et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. **Nature Genetics**, v. 50, n. 11, p. 1565–1573, 2018.
- ZHOU, Q. et al. Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization. **Science (New York, N.Y.)**, v. 353, n. 6297, p. 394–9, 2016.
- ZMUDJAK, M. et al. mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in Arabidopsis. **New Phytologist**, v. 199, n. 2, p. 379–394, 2013.

APÊNDICE A – Lista de indels identificados no cloroplasto da cultivar Tambo

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
6376--6421	16	2.9	16	93	3	76	56	13	4	26	1.55
7247--7273	13	2.1	13	100	0	54	51	22	0	25	1.48
14513--14537	12	2.1	12	100	0	50	36	16	16	32	1.90
16661--16710	25	2.0	25	96	0	91	46	8	4	42	1.52
16661--16708	13	3.8	13	80	5	62	45	8	4	41	1.53
16764--16895	67	2.0	67	89	0	201	31	16	24	28	1.96
19501--19529	14	2.0	15	93	6	51	75	0	17	6	1.01
19680--19711	16	2.0	16	100	0	64	18	12	6	62	1.50
20983--21022	20	2.0	20	100	0	80	35	10	10	45	1.71
21422--21485	32	2.0	32	100	0	128	40	0	0	59	0.97
27794--27834	21	2.0	21	100	0	82	26	29	14	29	1.95
28687--28715	15	1.9	15	100	0	58	27	34	17	20	1.95
29133--29161	15	1.9	15	100	0	58	27	34	17	20	1.95
28392--29283	446	2.0	446	100	0	1784	27	18	23	30	1.98
30206--30310	21	5.0	21	85	0	138	40	14	33	12	1.83
30207--30310	42	2.5	42	95	0	181	39	14	33	12	1.84
30368--30458	21	4.3	21	73	5	85	40	14	25	19	1.89
30362--30458	42	2.3	42	87	3	133	40	13	27	18	1.88
30381--30470	21	4.3	21	72	10	85	38	14	27	18	1.90
30552--30634	33	2.5	32	79	11	96	40	10	32	15	1.82
30618--30652	12	2.9	12	100	0	70	42	17	31	8	1.79
32997--33026	11	2.7	11	100	0	60	30	0	16	53	1.44
36092--36123	16	2.0	16	93	0	55	28	0	0	71	0.86
36364--36391	14	2.0	14	100	0	56	57	7	7	28	1.52
39721--39762	21	2.0	21	100	0	84	71	4	14	9	1.28
46812--46863	26	2.0	26	100	0	104	15	38	7	38	1.76
49323--49357	17	2.1	17	100	0	70	60	11	17	11	1.59
58064--58094	15	2.1	15	100	0	62	19	32	6	41	1.77
60448--60473	13	2.0	13	100	0	52	7	30	7	53	1.57
68861--68926	25	2.7	25	81	9	82	24	0	3	72	0.98
68896--68926	10	3.2	10	95	4	55	29	0	0	70	0.87
68861--68925	8	8.1	8	76	17	64	23	0	3	73	0.97
68861--68913	15	3.5	15	86	2	79	20	0	3	75	0.96
69054--69178	21	6.0	21	85	3	180	36	26	5	31	1.80
69054--69175	63	1.9	63	91	3	201	36	27	4	31	1.78
69466--69524	21	2.8	21	75	10	66	49	22	10	18	1.77
71821--71867	17	2.7	17	86	3	58	46	8	8	36	1.65
71838--71871	17	2.0	17	100	0	68	52	5	11	29	1.61
79897--79936	18	2.2	18	86	0	53	2	20	7	70	1.24
79909--79949	18	2.3	18	86	0	55	9	14	2	73	1.19
85922--86642	360	2.0	360	100	0	1442	24	21	18	35	1.95
95835--95867	14	2.4	14	94	0	57	27	15	3	54	1.55
95860--95921	20	3.0	21	95	2	108	37	4	8	50	1.54
95861--95924	21	3.1	21	95	2	112	35	4	7	51	1.52