

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Genômica comparativa de dois isolados de *S. scitamineum* com
níveis diferentes de agressividade na doença do carvão da cana-de-
açúcar**

Pedro Fernando Vilanova Ferreira

Trabalho de conclusão de curso apresentado como
parte dos requisitos para obtenção do título de
Bacharel em Ciências Biológicas.

**Piracicaba
2023**

Pedro Fernando Vilanova Ferreira

Genômica comparativa de dois isolados de *S. scitamineum* com níveis diferentes de agressividade na doença do carvão da cana-de-açúcar

Orientadora:

Prof^a Dr^a **CLAUDIA BARROS MONTEIRO-VITORELLO**

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciências Biológicas.

**Piracicaba
2023**

Aos meus amados avós Pedro e Maria, assim como à minha mãe Katy e ao meu padrasto Diego, a resiliência de vocês é minha fonte inesgotável de inspiração. Que este trabalho seja uma homenagem ao seu legado.

AGRADECIMENTOS

À Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) pela disponibilização dos espaços educacionais que foram essenciais para minha formação e por proporcionar um ambiente acadêmico estimulante e propício ao desenvolvimento científico.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processos 135233/2020-87 e 143879/2021-9) e à Fundação de Amparo à Pesquisa do estado de São Paulo (FAPESP) (Processos 2017/13268-2 e 2022/03962-7) pela concessão das bolsas que foram responsáveis por financiar e dar subsídio para o desenvolvimento deste projeto.

Aos meus avós Pedro e Maria, que lutaram todos os dias para me fornecer as condições e o fomento necessário para minha formação humana e acadêmica, sendo constantes em minha vida, oferecendo amor, apoio e sabedoria ao longo dos anos.

À minha mãe Katy, uma fonte de inspiração que me ensinou a resiliência em face dos desafios e da adversidade, ensinando-me a persistir e buscar constantemente o conhecimento.

Ao meu padrasto Diego, cujo cuidado, afeto e suporte foram fundamentais ao longo de minha jornada.

À minha orientadora Profa. Dra. Claudia Barros Monteiro-Vitorello, cuja expertise, dedicação, comprometimento, disponibilidade e incentivo foram imperativos para meu desenvolvimento humanístico e profissional com justiça e paz. Agradeço por abrir as portas para tantas oportunidades, serei eternamente grato.

Aos estimados membros do Grupo de Genômica, com quem convivi ao longo dos anos, pela contribuição única e individual de cada um através de ideias, conhecimentos e perspectivas que foram essenciais para o aprimoramento de minhas pesquisas. Agradeço por compartilharem tempo e experiência comigo, sou grato pela atmosfera colaborativa e acolhedora que sempre me circundou.

Ao meu eu do passado, pela perseverança, determinação, disciplina e coragem, mesmo nos momentos mais desafiadores e impossíveis.

“Se pude enxergar mais longe, foi porque me apoiei nos ombros de gigantes”.

(Isaac Newton)

SUMÁRIO

1. INTRODUÇÃO.....	17
2. REFERENCIAL TEÓRICO	18
2.1. A cana-de-açúcar	18
2.2. O carvão da cana-de-açúcar	21
2.3. Genômica de microrganismos	27
2.4. A interação planta-patógeno	31
3. OBJETIVO GERAL	34
4. OBJETIVOS ESPECÍFICOS.....	35
5. MATERIAIS E MÉTODOS	35
5.1. Material biológico	35
5.2. Obtenção de células leveduriformes haploides.....	35
5.3. Isolamento dos <i>mating-types</i>	36
5.4. Teste de cruzamento <i>in vitro</i>	36
5.5. Microscopia de luz	37
5.6. Obtenção de DNA de alto peso molecular e preparo de biblioteca genômica	37
5.7. Análise de discriminação molecular de <i>mating-types</i> opostos	38
5.8. Confirmação molecular da identidade celular dos isolados de <i>Sporisorium scitamineum</i>	39
5.9. Sequenciamento e montagem dos genomas.....	40
5.10. Análises de sintenia	40
5.11. Predição gênica.....	40
5.12. Predição de órtologos	41
5.13. Análise de SNVs (<i>Single Nucleotide Variants</i>)	41
6. RESULTADOS	42
6.1. Cruzamento <i>in vitro</i> das células leveduriformes haplóides de <i>mating-type</i> opostos.....	42
6.2. Extração de DNA total para Reação em Cadeia da Polimerase (PCR)	43
6.3. Discriminação molecular de <i>mating-types</i> opostos	45
6.4. Identificação de haplótipos	51
6.5. Sequenciamento de Nova Geração (NGS) pela plataforma Oxford Nanopore MinION	53
6.5.1. Isolado SSC04B (<i>MAT-1</i>).....	53

6.5.2. Isolado SSC39A (MAT-2)	54
6.6. Montagem <i>de novo</i> e análises de genômica comparativa.....	55
6.7. Análise de variantes estruturais	67
6.8. Análise de Variantes de Nucleotídeo Único (SNVs).....	72
7. DISCUSSÃO	85
8. CONSIDERAÇÕES FINAIS	92
REFERÊNCIAS	94
APÊNDICES	103
APÊNDICE A – SEQUENCIAMENTO DO ISOLADO SSC04 MAT-1.....	103
APÊNDICE B – SEQUENCIAMENTO DO ISOLADO SSC39 MAT-2.....	106

RESUMO

Genômica comparativa de dois isolados de *S. scitamineum* com níveis diferentes de agressividade na doença do carvão da cana-de-açúcar

A doença do carvão da cana-de-açúcar é resultado da interação biotrófica entre o fungo basidiomiceto *Sporisorium scitamineum* e a planta hospedeira, a cana-de-açúcar. A doença é caracterizada por profundas mudanças no metabolismo do hospedeiro e a emissão de uma longa estrutura a partir do meristema da planta chamada de chicote, onde estão localizados os teliósporos do patógeno que serão dispersos pelo vento. Distintos genótipos de cana possuem níveis diferentes de resistência quanto à colonização do fungo. Estudos anteriores avaliaram a natureza da agressividade de dois haplótipos brasileiros do patógeno infectando plantas resistentes e suscetíveis. Resultados demonstraram que o isolado SSC04 é capaz de promover desenvolvimento precoce da doença quando comparado ao isolado SSC39, além de induzir um maior número de chicotes nas plantas infectadas. Neste projeto, foram sequenciados os genomas completos de células sexualmente compatíveis dos dois isolados, buscando elucidar o grau de polimorfismo e/ou heteroziguidade entre estes indivíduos. A análise comparativa revelou uma alta porcentagem de sintenia entre os genomas e sequências gênicas, entretanto, foram encontradas alterações estruturais associadas à região entre os *loci* de *mating-type*, onde estão localizados os genes responsáveis pela determinação sexual, essenciais para formação da hifa dicariótica, única fase infectiva do patógeno. Estas variações estruturais provavelmente foram mediadas por remanescentes de sequências relacionadas a elementos transponíveis. Além disso, o isolado mais agressivo SSC04 apresentou maior heteroziguidade em relação aos genomas do isolado SSC39, sendo que a variação encontrada entre os haploides do isolado SSC39 se restringe, majoritariamente, aos genes de *mating*. Variantes de nucleotídeo único (SNVs) foram identificadas dentro da sequência codante de genes candidatos a efetores entre os dois isolados, corroborando dados da literatura.

Palavras-chave: heteroziguidade, efetores, sequenciamento, isolados, agressividade, planta-patógeno

ABSTRACT

Comparative genomics of two isolates of *S. scitamineum* with different levels of aggressiveness in sugarcane smut disease

Sugarcane smut disease is the result of the biotrophic interaction between the basidiomycete fungus *Sporisorium scitamineum* and its host plant, sugarcane. The disease is characterized by profound changes in the host's metabolism and the emission of a long structure called a whip, which originates from the plant's meristem and contains the pathogen's teliospores that are dispersed by the wind. Different sugarcane genotypes exhibit varying levels of resistance to fungal colonization. Previous studies have assessed the aggressiveness of two Brazilian haplotypes of the pathogen by infecting resistant and susceptible plants. The results demonstrated that the SSC04 isolate is capable of inducing early disease development compared to the SSC39 isolate, and it also induces a higher number of whips in infected plants. In this project, the complete genomes of sexually compatible cells from both isolates were sequenced to elucidate the degree of polymorphism and/or heterozygosity between these individuals. Comparative analysis revealed a high percentage of synteny between the genomes and gene sequences; however, structural alterations were found associated with the region between the mating-type loci, where genes responsible for sexual determination are located. These genes are essential for the formation of the pathogen's infective dicaryotic hyphae. These structural variations were likely mediated by remnants of transposable element-related sequences. Furthermore, the more aggressive SSC04 isolate exhibited higher heterozygosity compared to the genomes of the SSC39 isolate, and the variation found among the haploids of the SSC39 isolate is mainly restricted to the mating genes. Single nucleotide variants (SNVs) were identified within the coding sequence of candidate effector genes between the two isolates, which supports existing literature data.

Keywords: heterozygosity, effectors, sequencing, isolates, aggressiveness, plant-pathogen

LISTA DE FIGURAS

Figura 1 – Chicote, principal sintoma da doença do carvão na cana-de-açúcar causada pelo fungo biotrófico <i>Sporisorium scitamineum</i> . Fonte: Autoria própria	24
Figura 2 – Fases do ciclo de vida do fungo biotrófico <i>Sporisorium scitamineum</i> . Fonte: MONTEIRO-VITORELLO et al., 2018	25
Figura 3 – Representação esquemática do processo de cruzamento sexual em <i>S. scitamineum</i> . Fonte: Autoria própria.....	26
Figura 4 — Teste de compatibilidade entre os isolados 04 e 39 e seus respectivos <i>mating-types</i> . A denominação “A” refere-se ao alelo <i>MAT-2</i> , enquanto a denominação “B” refere-se ao alelo <i>MAT-1</i> . Reações sexuais positivas apresentaram uma colônia de morfologia branca e cotonosa com um halo de hifas (1,3,6,8,9,11,14,16), enquanto reações negativas apresentaram colônias lisas e leveduriformes (2,4,5,7,10,12,13,15). Cruzamentos ocorreram em meio YM sólido a 28°C por 3 dias. Isolados são capazes de hibridização.....	42
Figura 5 — Identificação visual cruzamento <i>in vitro</i> dos isolados SSC04A/SSC04B e SSC39A/SSC39B. A – Reação positiva de cruzamento entre dois <i>mating-types</i> diferentes sob microscopia de luz, aumento de 40x. É possível observar a formação de hifas, gerando aspecto cotonoso em placa. B – Células leveduriformes haplóides sob microscopia de luz, aumento de 40x. É possível observar reação ausente de cruzamento, quando dois <i>mating-types</i> idênticos são cruzados. Crescimento foi realizado com meio YM sólido e inóculo crescido a 28°C. Fotos realizadas em microscópio de luz OLYMPUS – BX40F4	43
Figura 6 — Quantificação das amostras de extração de DNA genômico total medido pelo espectrofotômetro NanoDrop 2000/2000c. A — Espectro de absorbância e comprimento de onda da amostra SSC04A. B — Espectro de absorbância e comprimento de onda da amostra SSC04B. De acordo com a ThermoFisher Scientific, os ácidos nucleicos possuem absorbância máxima em 260 nm	44
Figura 7 — Quantificação das amostras de extração de DNA genômico total medido pelo espectrofotômetro NanoDrop 2000/2000c. A — Espectro de absorbância e comprimento de onda da amostra SSC39A. B — Espectro de absorbância e comprimento de onda da amostra SSC39B. De acordo com a ThermoFisher Scientific, os ácidos nucleicos possuem absorbância máxima em 260 nm	45

Figura 8 — Teste de discriminação molecular de *mating-types* opostos do isolado SSC04 de *Sporisorium scitamineum*. Produtos de PCR dos *primers* *bE1F1/bE1-2R1* (*MAT-1*) e *bE2F2/bE1-2R1* (*MAT-2*) foram colocados em gel de eletroforese (agarose 1%) com *ladder* de 1 kpb (ThermoFisher Scientific) e registrados em transiluminador. Lanes 3 e 4 são referentes ao DNA de um dos haplóides, e as lanes 6 e 7 do haplóide oposto. Lanes 2, 5 e 7 são referentes ao controle negativo 46

Figura 9 — Teste de discriminação molecular de *mating-types* opostos do isolado SSC39 de *Sporisorium scitamineum*. Produtos de PCR dos *primers* *bE1F1/bE1-2R1* (*MAT-1*) e *bE2F2/bE1-2R1* (*MAT-2*) foram colocados em gel de eletroforese (agarose 1%) com *ladder* de 1 kpb (ThermoFisher Scientific) e registrados em transiluminador. Lanes 2 e 3 são referentes ao DNA de um dos haplóides, e as lanes 4 e 5 do haplóide oposto.. Lanes 2, 5 e 7 são referentes ao controle negativo 46

Figura 10 — Comprimento da sequência genômica do alelo *bE1* do gene *bE* no *locus* de *mating-type* de *Sporisorium scitamineum*. Os *primers* *bE1F1* e *bE1-2R1*, indicados pelos triângulos invertidos, amplificam a região N-terminal..... 47

Figura 11 — Região genômica de anelamento dos *primers* *bE1F1* (5'-ATGCGTGAATTTGCGC-3') e *bE1-2R1* (5'-TTGATGAACCAGAGCGTGAG-3') na região N-terminal do alelo *bE1* do gene *bE*. A região amplificada corresponde a 668 pares de base. Alinhamento de *primers* e tradução da sequência foi realizada pelo Benchling..... 47

Figura 12 — Sequenciamento pelo método de Sanger (SS) do produto de PCR gerou uma sequência consenso de 639 pb, referente ao par de *primers* de *mating-type* *bE1F1/bE1-2R1* para identificação do alelo *MAT-1* de *S. scitamineum*. Setas em verde são referentes a sequência dos *primers* e sua direção. A primeira linha representa a referência de sequência codante parcial do alelo *bE1* como depositado no *NCBI* pelo acesso MZ926872.1. A segunda linha representa a sequência consenso gerada pela sequência *forward* e *reverse*. Terceira e quarta linha representam, respectivamente, a sequência *reverse* e *forward*. Cromatogramas são referentes à discriminação e qualidade das bases. Cromatogramas e alinhamento das sequências foram processados pelo CLC Genomics Workbench v. 20 (QIAGEN) 48

Figura 13 — Comprimento da sequência genômica do alelo *bE2* do gene *bE* no *locus* de *mating-type* de *Sporisorium scitamineum*. Os *primers* *bE2F2* e *bE1-2R1*, indicados pelos triângulos invertidos, amplificam a região N-terminal..... 49

Figura 14 — Região genômica de anelamento dos *primers* *bE2F2* (5'-CTTTCCTACAATCCAAACCAATA-3') e *bE1-2R1* (5'-TTGATGAACCAGAGCGTGAG-3') na região N-terminal do alelo *bE2* do gene *bE*. A região amplificada corresponde a 458 pares de base49

Figura 15 — Sequenciamento pelo método de Sanger (SS) do produto de PCR gerou uma sequência consenso de 437 pb, referente ao par de *primers* de *mating-type* *bE2F2/bE1-2R1* para identificação do alelo *MAT-2* de *S. scitamineum*. Setas em verde são referentes a sequência dos *primers* e sua direção. A primeira linha representa a referência de sequência codante parcial do alelo *bE2* como depositado no *NCBI* pelo acesso MZ926873.1. A segunda linha representa a sequência consenso gerada pela sequência *forward* e *reverse*. Terceira e quarta linha representam, respectivamente, a sequência *reverse* e *forward*. Cromatogramas são referentes à discriminação e qualidade das bases. Cromatogramas e alinhamento das sequências foram processados pelo CLC Genomics Workbench v. 20 (QIAGEN).....50

Figura 16 — Amplificação do gene candidato a efetor *g5159* visualizada em gel de agarose (1%). Lane 1 mostra ladder 1 kpb, enquanto as lanes 2,4 e 6 mostram um produto de 2500 pb, resultante da amplificação dos *primers forward* e *reverse* do início ao final do gene. As lanes 3, 5 e 7 mostram um produto de, aproximadamente, 1500 pb, resultante da amplificação a partir do *primer forward* interno ao gene até o seu *reverse*51

Figura 17 — Região genômica de anelamento dos *primers* do *g5159 forward* (5'-GAAAAGGGGAGGGATCTA-3'), *forward* interno (5'-CTGAGCCTGAACCAAAACGC-3') e *reverse* (5'-AGTCTCAAACACACAGGCAA-3'). A representação mostra a sequência reversa complementar. Quadrados coloridos representam os aminoácidos que compõem a proteína putativa do *g5159*. O asterisco representa o códon de parada da cadeia polipeptídica.....52

Figura 18 — Polimorfismos nas posições 689 e 690 do gene candidato a efetor *g5159* nos dois isolados. Fonte: Crestana (2015).53

Figura 19 — Gráfico de barras demonstrando o comprimento da sequência genômica de cada *contig* correspondente aos cromossomos da referência. Em cinza escuro, o tamanho dos cromossomos de SSC04A e, em cinza claro, os contigs da montagem *de novo* de SSC04B58

Figura 20 — Gráfico de barras demonstrando o comprimento da sequência genômica de cada <i>contig</i> correspondente aos cromossomos da referência. Em cinza escuro, os <i>contigs</i> da montagem <i>de novo</i> de SSC39A e, em cinza claro, a referência SSC39B (Taniguti et al., 2015).	60
Figura 21 — <i>Dot plot</i> das sequências genômicas de <i>mating-types</i> complementares e opostos do isolado SSC04 de <i>S. scitamineum</i> . O gráfico ilustra o alinhamento de cada <i>contig</i> da montagem <i>de novo</i> em relação à referência e demonstra sintonia conservada expressa pela linha diagonal contínua	61
Figura 22 — <i>Dot plot</i> das sequências genômicas de <i>mating-types</i> complementares e opostos do isolado SSC39 de <i>S. scitamineum</i> . O gráfico ilustra o alinhamento de cada <i>contig</i> da montagem <i>de novo</i> em relação à referência e demonstra sintonia conservada expressa pela linha diagonal contínua	61
Figura 23 — Gráfico comparando a completude genômica das diferentes montagens <i>de novo</i> com suas referências.	62
Figura 24 — Densidade de sequências codantes distribuídas ao longo do genoma SSC04 MAT-1. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5	63
Figura 25 — Densidade de sequências codantes distribuídas ao longo do genoma SSC04 MAT-2. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5	64
Figura 26 — Densidade de sequências codantes distribuídas ao longo do genoma SSC39 MAT-1. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5	65
Figura 27 — Densidade de sequências codantes distribuídas ao longo do genoma SSC39 MAT-2. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5	66
Figura 28 — <i>Dot plot</i> das sequências genômicas de <i>mating-types</i> complementares e opostos dos isolados SSC04 (esquerda) e SSC39 (direita) de <i>S. scitamineum</i> . <i>Dot plot</i> gerado pelo CLC Genomics Workbench (QIAGEN). O gráfico ilustra o alinhamento do cromossomo 2 nos dois isolados, demonstrando a presença de inversões que são apresentadas como linhas invertidas, repetições, expressas pelas linhas adjacentes e lacunas.	68
Figura 29 — <i>Dot plot</i> das sequências genômicas de <i>mating-types</i> iguais dos isolados SSC04A e SSC39A (esquerda) e SSC04B e SSC39B (direita) de <i>S. scitamineum</i> .	

<i>Dot plot</i> gerado pelo CLC Genomics Workbench (QIAGEN). O gráfico ilustra o alinhamento do cromossomo 2 nos dois isolados. Entre os genomas de alelo <i>MAT-2</i> (representado pelo “A”), há presença de uma inversão, expressada pela linha invertida, e nos genomas de alelo <i>MAT-1</i> , há sintonia contínua	68
Figura 30 — <i>Dot plot</i> das sequências gênicas de <i>mating-types</i> complementares e opostos do isolado SSC04 de <i>S. scitamineum</i> , gerados pelo MCScanX	69
Figura 31 — Profundidade sintênica entre os genomas SSC04A (<i>MAT-2</i>) e SSC04B (<i>MAT-1</i>) gerada pelo MCScanX	69
Figura 32 — <i>Dot plot</i> das sequências gênicas de <i>mating-types</i> complementares e opostos do isolado SSC39 de <i>S. scitamineum</i> , gerados pelo MCScanX	70
Figura 33 — Profundidade sintênica entre os genomas SSC04A (<i>MAT-2</i>) e SSC04B (<i>MAT-1</i>) gerada pelo MCScanX	71
Figura 34 — Alterações estruturais na região entre os <i>loci</i> de <i>mating-type</i> em diferentes isolados de <i>S. scitamineum</i> . Sequências genômicas anotadas foram alinhadas pelo CLC Genomics Workbench (QIAGEN) e ajustadas manualmente	72
Figura 35 — Número de variantes encontradas entre os <i>mating-types</i> de dois isolados de <i>S. scitamineum</i> através dos <i>softwares</i> Medaka e Clair3. O eixo x representa a cobertura mínima de <i>reads</i> consideradas. O eixo y representa o número de variantes.....	73
Figura 36 — Variantes entre os <i>mating-types</i> opostos dos isolados SSC04 e SSC39 suportadas pelos <i>softwares</i> Medaka e Clair3.....	74
Figura 37 — Variantes encontradas entre os genomas quando comparados com a referência SSC39 <i>MAT-1</i> (Taniguti et al., 2015).....	75
Figura 38 — Distribuição do número de variantes ao longo do genoma de dois isolados diferentes e seus respectivos <i>mating-types</i>	76
Figura 39 — Distribuição das variantes entre os alelos <i>MAT-1</i> e <i>MAT-2</i> do isolado SSC04. Escala de 10^6	77
Figura 40 — Número de variantes por qualidade de <i>phred score</i> entre os genomas SSC04 <i>MAT-1</i> e <i>MAT-2</i>	78
Figura 41 — Distribuição do efeito das variantes entre os genomas SSC04 <i>MAT-1</i> e <i>MAT-2</i> por tipo e região do genoma.....	79
Figura 42 — Distribuição das variantes entre os alelos <i>MAT-1</i> e <i>MAT-2</i> do isolado SSC39. Escala de 10^6	79

Figura 43 — Número de variantes por qualidade de <i>phred score</i> entre os genomas SSC39 <i>MAT-1</i> e <i>MAT-2</i>	80
Figura 44 — Distribuição do efeito das variantes entre os genomas SSC39 <i>MAT-1</i> e <i>MAT-2</i> por tipo e região do genoma	81
Figura 45 — Distribuição das variantes entre os alelos <i>MAT-2</i> do isolado SSC04 e <i>MAT-1</i> do isolado SSC39. Escala de 10^6	81
Figura 46 — Número de variantes por qualidade de <i>phred score</i> entre os genomas SSC04 <i>MAT-2</i> e SSC39 <i>MAT-1</i>	83
Figura 47 — Distribuição do efeito das variantes entre os genomas SSC04 <i>MAT-2</i> e SSC39 <i>MAT-1</i> por tipo e região do genoma	84
Figura 48 — Panorama geral da distribuição de variantes entre genomas de <i>S. scitamineum</i> . De fora para dentro: cromossomos, densidade gênica, genes candidatos a efetores (em vermelho: polimórficos) e <i>loci</i> de <i>mating-type</i> em azul, SNVs do isolado SSC04 <i>MAT-2</i> , SNVs do isolado SSC04 <i>MAT-1</i> e SNVs do isolado SSC39 <i>MAT-2</i> . Em cinza claro, destacada região de <i>mating-type</i> . Escala da densidade gênica é de 10^4 . Imagem gerada através do Circos v. 0.69	85

LISTA DE TABELAS

Tabela 1 — <i>Primers</i> utilizado por Agisha et al. (2022) para amplificação dos alelos referentes ao gene <i>bE</i> , presente no <i>locus b</i> de <i>mating-type</i>	38
Tabela 2 — <i>Primers</i> utilizado para amplificação da sequência completa do gene candidato a efector <i>g5159</i> de <i>S. scitamineum</i>	39
Tabela 3 — Quantificação de extração de DNA genômico total das amostras SSC04 medido pelo espectrofotômetro NanoDrop 2000 e razões	44
Tabela 4 — Quantificação de extração de DNA genômico total das amostras SSC39 medido pelo espectrofotômetro NanoDrop 2000 e razões	45
Tabela 5 — Dados referentes às montagens <i>de novo</i> dos genomas das células haploides SSC04B e SSC39A de <i>S. scitamineum</i>	56
Tabela 6 — Dados referentes às montagens <i>de novo</i> dos genomas dos haploides SSC04B e SSC39A de <i>S. scitamineum</i> quando comparados com suas referências, SSC04A e SSC39B, respectivamente.....	56
Tabela 7 — Comprimento dos <i>contigs</i> gerados pela montagem <i>de novo</i> do genoma do haplóide SSC04B de <i>S. scitamineum</i> após três rodadas de polimento em comparação com a referência SSC04A (dados não publicados).	57
Tabela 8 — Número de bases dos <i>contigs</i> gerados pela montagem <i>de novo</i> do genoma do haplóide SSC39A de <i>S. scitamineum</i> após três rodadas de polimento em comparação com a referência SSC39B (Taniguti et al., 2015).....	59
Tabela 9 — Panorama geral comparativo entre os genomas de dois isolados diferentes de <i>S. scitamineum</i> com seus <i>mating-types</i> complementares.	67
Tabela 10 — Distribuição de variantes em cada cromossomo entre os genomas SSC04 <i>MAT-1</i> e <i>MAT-2</i>	77
Tabela 11 — Número de efeitos por tipo em relação às variantes entre os genomas SSC04 <i>MAT-1</i> e <i>MAT-2</i>	78
Tabela 12 — Distribuição de variantes em cada cromossomo entre os genomas SSC39 <i>MAT-1</i> e <i>MAT-2</i>	80
Tabela 13 — Número de efeitos por tipo em relação às variantes entre os genomas SSC39 <i>MAT-1</i> e <i>MAT-2</i>	80
Tabela 14 — Distribuição de variantes em cada cromossomo entre os genomas SSC04 <i>MAT-2</i> e SSC39 e <i>MAT-1</i>	82

Tabela 15 — Número de efeitos por tipo em relação às variantes entre os genomas SSC04 <i>MAT-2</i> e SSC39 <i>MAT-1</i>	83
--	----

1. INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp.) é uma planta pertencente ao grupo economicamente importante das gramíneas (CARDOSO-SILVA et al., 2014). A espécie possui uma capacidade superior de crescimento em relação a outras culturas, devido, principalmente, à sua alta eficiência fotossintética como uma planta C4, em condições de altas temperaturas e luminosidade (RICAUD et al., 2012). A cana-de-açúcar passou por grandes transformações induzidas de forma antropogênica e se converteu em uma importante fonte de alimento e energia (BABU et al., 2022). A utilização de seus subprodutos, dentre eles, o álcool e a biomassa, tornou-se essencial para vários setores da economia, como o setor de combustíveis e bioeletricidade, sendo uma alternativa de fonte renovável e limpa (GOES et al., 2011). Além disso, a espécie é responsável por contribuir com grande parte da produção de açúcar no mundo (BABU et al., 2022). Dentro deste cenário, é irrefutável o papel protagonista do Brasil, visto que, segundo a União da Indústria de Cana-de-Açúcar e Bioenergia (UNICA), foi contabilizado um volume de 27 milhões de toneladas de açúcar exportado pelo país em 2022, tendo a China como principal destino. Entretanto, tanto fatores ambientais quanto doenças sistêmicas podem limitar a produção de cana-de-açúcar (RICAUD et al., 2012; SUNDAR et al., 2012).

Dentre estas doenças sistêmicas está a doença do carvão, causada pelo fungo biotrófico *Sporisorium scitamineum* (Syd.) M. Piepenbr., M. Stoll & Oberw (PIEPENBRING et al., 2002), capaz de causar perdas significativas de produção (SUNDAR et al., 2012). A colonização sistêmica do fungo induz o desenvolvimento de uma longa estrutura no meristema (tanto apical quanto lateral) da planta denominada de chicote, onde ocorre a esporogênese de teliosporos responsáveis pela disseminação da doença (MARQUES et al., 2016). Estas mudanças são marcadas por profundas alterações transcricionais capazes de direcionar o metabolismo do hospedeiro em direção a um aumento na produção de energia, envolvendo processos como a glicólise e o ciclo do ácido cítrico (SCHAKER et al., 2016). Assim, o desenvolvimento da doença ocasiona reduções no conteúdo de sacarose e, portanto, na produção de açúcar, além de outras características como a redução do crescimento da planta, a quantidade e qualidade do suco (RAJPUT et al., 2021).

Estudos anteriores demonstraram a presença de dois haplótipos distintos de *S. scitamineum* dentre os isolados brasileiros (BENEVENUTO et al.,

2016). Os isolados SSC04 e SSC39 pertencentes a diferentes haplótipos foram avaliados quanto ao nível de agressividade da doença e demonstraram ser distintos em relação a severidade do sintoma, indicando a possibilidade de uma natureza genética diferencial (OLIVEIRA, 2021). Na literatura, abordagens moleculares têm sido empregadas na identificação dos genes relacionados à agressividade (XU et al., 2006). Estes componentes genéticos quantitativos abrem as portas para a adaptação do patógeno, e a busca por compreender esta adaptação inclui a procura por polimorfismos no DNA (PARIAUD et al., 2009). Pesquisas comparativas anteriores, realizadas entre diferentes isolados de *S. scitamineum*, foram responsáveis por identificar a existência de polimorfismos entre diferentes genes candidatos a efetores, dentre eles, polimorfismos conferindo substituições não-sinônimas (TEIXEIRA-SILVA et al., 2020). Além disso, segundo Stukenbrock & Dutheil (2012), genomas fúngicos possuem uma alta plasticidade, sendo que variações cromossômicas estruturais podem ocorrer. Por esta razão, o presente projeto visou sequenciar e comparar os genomas haplóides isolados de células sexualmente compatíveis de dois indivíduos 2n (teliósporos) do fungo *Sporisorium scitamineum* contrastantes quanto à agressividade em cana-de-açúcar: as células do isolado mais agressivo SSC04A (MAT-2), SSC04B (MAT-1) e do isolado menos agressivo SSC039A (MAT -2), SSC39B (MAT-1). Buscando, desta forma, elucidar se o nível de polimorfismos e/ou heterozigossidade são determinantes para a agressividade do patógeno.

2. REFERENCIAL TEÓRICO

2.1. A cana-de-açúcar

A cana-de-açúcar é uma planta híbrida perene de alta estatura, propagada vegetativamente, do gênero *Saccharum* spp. pertencente à família Poaceae, com centro de origem no sudeste asiático (DANIELS; ROACH, 1987), na região da Nova Guiné e Indonésia (MARIN, 2014). Segundo Bakker (1999), o nome *Saccharum* foi dado por Linnaeus em 1753, visto que, ao longo da história, os romanos transformaram a palavra Sakchar ou Sakcharon, do grego antigo “açúcar”, para “Saccharum”, no latim. Estudos indicam que a espécie *S. officinarum* se derivou a partir da domesticação de *S. robustum* (BAKKER, 1999; KAFFKA e GRANTZ,

2014). Variedades modernas são resultado da hibridização de diversas espécies do gênero, sendo que as comerciais são cruzamentos de *S. officinarum* x *S. spontaneum*, apesar de existirem programas de hibridização que envolvem outras espécies como *S. barberi* e *S. sinense* para obtenção de características de interesse, como maior vigor e resistência à doenças (KAFFKA e GRANTZ, 2014; THIRUGNANASAMBANDAM et al., 2018). Segundo Marin (2014), a espécie *S. officinarum* possui uma alta capacidade de acumular sacarose, apesar de ser facilmente afetada por doenças (BAKKER, 1999), enquanto *S. spontaneum* atua como repositório de genes de resistência e tolerância ambiental (KAFFKA e GRANTZ, 2014). O genoma nuclear das espécies modernas, segundo Thirugnanasambandam et al. (2018), possui contribuições dos genomas destas duas espécies. De acordo com Kaffka & Grantz (2014), as espécies do gênero *Saccharum* são autoploiploides e seu nível de ploidia, assim como tamanho do genoma, pode variar consideravelmente entre as espécies devido à sua natureza híbrida.

A cana-de-açúcar é cultivada em mais de 100 países, em regiões tropicais e semitropicais, principalmente por causa de sua habilidade em armazenar altas concentrações de sacarose nos colmos (TEW; COBILL, 2008). É responsável por atender 80% da demanda de açúcar do mundo, rendendo uma variedade de produtos que vão desde produtos químicos a industriais (CHAMI et al., 2020; QAMAR et al., 2021). Para os anos de 2022/23, a produção global de açúcar deve apresentar um aumento de 1.7 milhões de toneladas, totalizando 182.9 milhões de toneladas (USDA, 2022). Segundo a Organização das Nações Unidas para a Alimentação e Agricultura (FAO), para o ano de 2020, a área colhida relativa à cana-de-açúcar no mundo foi na ordem aproximada de 26.5 milhões de hectares. Dentro deste cenário, o Brasil é considerado o maior produtor e exportador (USDA, 2022). Segundo Cursi et al. (2022), durante a safra 2019-2020, a produção de açúcar atingiu 29.8 milhões de toneladas sendo que destas, 18.9 milhões foram exportadas, o que gerou, aproximadamente, 5.8 bilhões de dólares para o país. Séries históricas comprovam que o cultivo de cana-de-açúcar no país sofreu rápida expansão desde sua introdução durante a colonização (GOES et al., 2014). Segundo o Instituto Brasileiro de Geografia e Estatística (2020), a área colhida de cana-de-açúcar correspondeu, em 2020, a 10 milhões de hectares, com uma produção aproximada de 757 milhões de toneladas, sendo o Estado de São Paulo o maior produtor.

A rápida expansão do cultivo da cana vem acompanhada de uma crescente demanda da indústria alimentícia, mas, principalmente, pelo bioetanol (CHAMI et al., 2020). O bioetanol é amplamente utilizado em todo o mundo como um biocombustível (BORDONAL et al., 2018), sendo que, em 2021, o Brasil foi responsável por 27% da produção global, atrás apenas dos Estados Unidos (RFA, 2021). Para Cherubin et al. (2021), a maioria das alternativas no impedimento da crise climática depende da bioenergia para redução da emissão dos gases do efeito estufa e, neste cenário, a cana-de-açúcar se constitui como uma cultura bioenergética protagonista devido a seu alto potencial de produção.

Segundo Jaiswal et al. (2017), a expansão do bioetanol no país representa uma possível solução para redução das emissões de gases, visto a baixa emissão de carbono durante o ciclo de vida da cana-de-açúcar quando devidamente manejado (CHERUBIN et al., 2021). Além dos usos que contemplam a indústria alimentícia, a cana-de-açúcar tem sido utilizada como uma fonte de energia, pois é considerada uma das culturas mais eficientes na conversão de energia do sol em energia química (TEW; COBILL, 2008), fato evidenciado pela produção em larga-escala do etanol no Brasil, por exemplo. Além disso, é sabido que o país possui um potencial técnico para cogeração de energia de 146.000 GWh mediante uso de toda biomassa disponível nos campos (CURSI et al., 2022). Para Souza et al. (2018), “o Brasil é o único país em que os produtos da cana-de-açúcar possuem uma contribuição expressiva no setor energético”.

Entretanto, apesar de fatores abióticos já terem sido descritos na literatura (TEW; COBILL, 2008), fatores bióticos podem impactar negativamente o cultivo desta cultura, como doenças causadas por patógenos como fungos e bactérias (CURSI et al., 2022). A infecção por estes microrganismos leva a ativação do sistema de defesa da planta ocasionando em mudanças no metabolismo que podem, ocasionalmente, afetar o crescimento e desenvolvimento normal (BERGER et al., 2007). O contato com estes patógenos pode resultar em alterações na taxa fotossintética, causando alterações nos níveis de açúcares (BERGER et al., 2007). A resposta de defesa contra patógenos por parte das plantas está envolvida com respostas inatas chamadas de PTI (*pattern-triggered immunity*) e ETI (*effector-triggered immunity*), sendo que a primeira é iniciada a partir de proteínas receptoras (PRRs, do inglês, *pattern-recognition receptors*) localizadas na superfície da célula capazes de reconhecer padrões moleculares, enquanto a segunda é iniciada a partir

de receptores intracelulares citoplasmáticos (YUAN et al., 2021). Os receptores localizados na membrana plasmática envolvidos na resposta de PTI são responsáveis por reconhecer os chamados *PAMPs*, ou padrões moleculares associados a patógenos, sendo que a detecção destas moléculas é a primeira camada envolvida na ativação da resposta imune antes da invasão celular (FAULKNER; ROBATZEK, 2012). Fitopatógenos são capazes de suprimir a resposta de PTI através da liberação de moléculas efetoras codificadas por genes efetores do patógeno que, quando reconhecidas pela planta, ativam respostas de hipersensibilidade (HR) (CUNNAC et al., 2009; YUAN et al., 2021). Dentre as respostas de defesa da planta, presentes tanto em PTI quanto em ETI, estão a produção de espécies reativas de oxigênio (*ROS*, do inglês, *reactive oxygen species*), influxo de íons de cálcio, ativação da cascata MAPK e reprogramação transcricional (YUAN et al., 2021). Mais recentemente, Chang et al. (2022), reforçando os achados de pesquisas anteriores, lembraram que para que as respostas de PTI e ETI ocorram, os componentes principais na ativação de PTI também são necessários na ativação de ETI, e vice-versa. Além de que, ambos sistemas de defesa atuam potencializando um ao outro (CHANG et al., 2022) indo ao encontro de Tena (2021) quando diz que ambos os sistemas funcionam como uma única via, necessitando que todos seus componentes estejam completamente funcionais.

Dentre os diversos microrganismos que ocasionam doenças, alguns são: *Xanthomonas albilineans*, causadora da escaudadura das folhas (BLANCH et al., 2015), *Puccinia kuehnii*, causadora da ferrugem alaranjada (MOREIRA et al., 2018), *Leifsonia xyli* subsp. *xyli*, causadora do raquitismo-da-soqueira (MONTEIRO-VITORELLO et al., 2004) e *Sporisorium scitamineum* (SUNDAR, 2012), fungo biotrófico causador da doença do carvão. Patógenos biotróficos como o fungo *S. scitamineum* devem ser capazes de manipular o metabolismo do hospedeiro, de tal forma que os nutrientes necessários para sua sobrevivência sejam entregues (FAULKNER; ROBATZEK, 2012).

2.2. O carvão da cana-de-açúcar

Dentre os patógenos da cana-de-açúcar, um dos mais importantes é o fungo *Sporisorium scitamineum*, responsável por perdas significativas de produção e lucratividade (QUE et al., 2014; BHUIYAN et al., 2021). O patógeno pertence a um

grupo de fungos causadores da doença do carvão, responsável por afetar diversas gramíneas economicamente importantes, como a cevada e outros cereais e, por essa razão, estes fungos constituem um dos grupos de fitopatógenos mais relevantes (BEGEROW et al., 2014). Os fungos do carvão verdadeiros estão classificados dentro da classe dos Ustilaginomycetos no filo Basidiomycota (MONTEIRO-VITORELLO, 2018). Estes microrganismos são capazes de infectar uma única espécie ou alguns hospedeiros suscetíveis, contudo, análises moleculares confirmaram uma estreita distribuição no número de espécies por parasita, confirmando uma íntima interação hospedeiro-específica, tornando estes fungos patógenos especializados (BEGEROW et al., 2014). Dentre os patógenos mais importantes do grupo estão *Ustilago hordei*, agente causal da doença na cevada e aveia, *Ustilago maydis* e *Sporisorium reilianum*, no milho, e *Sporisorium scitamineum*, na cana-de-açúcar (FELDBRÜGGE et al., 2013). Uma característica comum a estes fungos é a necessidade de cruzamento sexual para o início do desenvolvimento patogênico e produção de uma hifa dicariótica infectiva (YAN et al., 2016).

Segundo Amorim (1989), a doença do carvão na cana-de-açúcar foi primeiro descrita em 1877 na África do Sul, com relatos surgindo, inicialmente, em países do continente africano e asiático. Desde então, severos problemas com a doença foram relatados em países como a Índia por volta da década de 1930 e, posteriormente, identificados em países da América do Sul como a Argentina, em 1940, e o Brasil, em 1946. Já em 1988, a doença foi reportada na Austrália. (revisado em MONTEIRO-VITORELLO et al, 2018). Papua Nova Guiné, considerada centro de origem da cana-de-açúcar, teve o primeiro caso reportado da doença do carvão apenas em agosto de 2016 (TOM et al., 2017). Ainda segundo Tom et al. (2017), até 2017, a doença ocorria em todos os países do mundo com exceção de Fiji.

O carvão da cana foi descrito no Estado de São Paulo, segundo relatório de Carvalho (1949), como sendo motivo da destruição de muitas lavouras, visando o controle da doença, em meio ao temor dos elevados danos causados em outros países produtores de cana-de-açúcar do mundo. De acordo com Amorim (1989), a presença do fungo foi notada no Estado de São Paulo ainda em 1946, época em que variedades suscetíveis foram erradicadas para evitar a propagação do patógeno. Ainda segundo a autora, embora a transição no uso de variedades

suscetíveis para variedades resistentes tenha mantido a doença em níveis controlados, a partir de 1979, a doença do carvão atingiu rapidamente proporções epidêmicas, principalmente pelo aumento do uso da variedade NA56-79 que, em 1982, correspondia a 50% da área plantada do estado. A doença já se encontrava presente em todo território nacional antes do final da década de 80, sendo muito provável que a dispersão entre os países da América do Sul e dentro do próprio território brasileiro tenha se dado pela própria atividade humana através do transporte de plantas de cana-de-açúcar assintomáticas quanto à doença (BENEVENUTO et al., 2016).

Segundo Borrás-Hidalgo (2005), a progressão da infecção do fungo em variedades suscetíveis do hospedeiro está ligada amplamente às condições ambientais, além de que a resistência para a doença não segue um padrão gene a gene como é possível observar em outras interações planta-patógeno, visto que as variedades modernas de cana-de-açúcar são híbridos poliploides (SCHENCK, 2003) e, portanto, uma variedade resistente em um país pode ser suscetível em outro (THUSHARI; DE COSTA, 2023)

Wu et al. (1988) discorreram sobre como estudos de genética quantitativa reportaram uma baixa herdabilidade no que diz respeito à resistência às raças do fungo. Além disso, de acordo com Schenck (2003), há dificuldade em determinar a presença de raças na espécie causadora do carvão da cana-de-açúcar, e até mesmo em outras espécies de carvão, como *U. maydis*, visto que hibridização é comum entre isolados e entre espécies. Esta facilidade na hibridização, segundo Bakkeren & Kronstad (1996), ocorre porque os produtos gênicos do *locus b* de diferentes espécies podem interagir, além de que os feromônios de *S. scitamineum* podem ser reconhecidos pelos receptores de outras espécies.

A característica mais evidenciada como sintoma da infecção do fungo do carvão na cana-de-açúcar é o aparecimento de um longa estrutura em forma de chicote (Figura 1) no meristema apical ou botões laterais. Essa estrutura pode assumir variações quanto à cor, número de ocorrências e tamanho (SUNDAR et al., 2012), sendo que a estrutura pode ir de alguns centímetros a 1.5 m como reportado por Comstock (2005). A emergência desta estrutura no ápice do caule é impulsionada por severas mudanças no metabolismo do hospedeiro (MARQUES et al., 2017). Sua forma estrutural é constituída por um núcleo central composto por tecido vegetal cercado por esporos, estes, por sua vez, cobertos com uma fina

membrana (COMSTOCK, 2005). Através desta estrutura, os esporos se desenvolvem e são capazes de disseminar a doença (MARQUES et al., 2017). A fase final de desenvolvimento dos chamados teliósporos ocorre a partir de processos celulares como a cariogamia e a fragmentação das hifas (BENEVENUTO et al., 2016).



Figura 1 – Chicote, principal sintoma da doença do carvão na cana-de-açúcar causada pelo fungo biotrófico *Sporisorium scitamineum*. Fonte: Autoria própria.

S. scitamineum é um fungo biotrófico e apresenta diversas fases em seu ciclo de vida. Teliósporos diplóides germinam formando um pró-basídio na superfície das gemas da cana-de-açúcar (SUNDAR et al., 2012) quando, por meiose, quatro esporídios haplóides são formados. Esses esporídios se dividem por mitose formando células leveduriformes haplóides com tipos sexuais compatíveis que devem se fundir para produzir uma hifa dicariótica e infectiva (Figura 2) (revisado em LONGATTO et al., 2015). É importante ressaltar que apesar de as fases de vida do fungo passarem pela germinação, formação do apressório, penetração e colonização do meristema, esta última não necessariamente se

traduzirá na doença, visto a influência exercida pelo grau de resistência do hospedeiro (LATIZA et al., 1980; TEIXEIRA-SILVA, 2020).

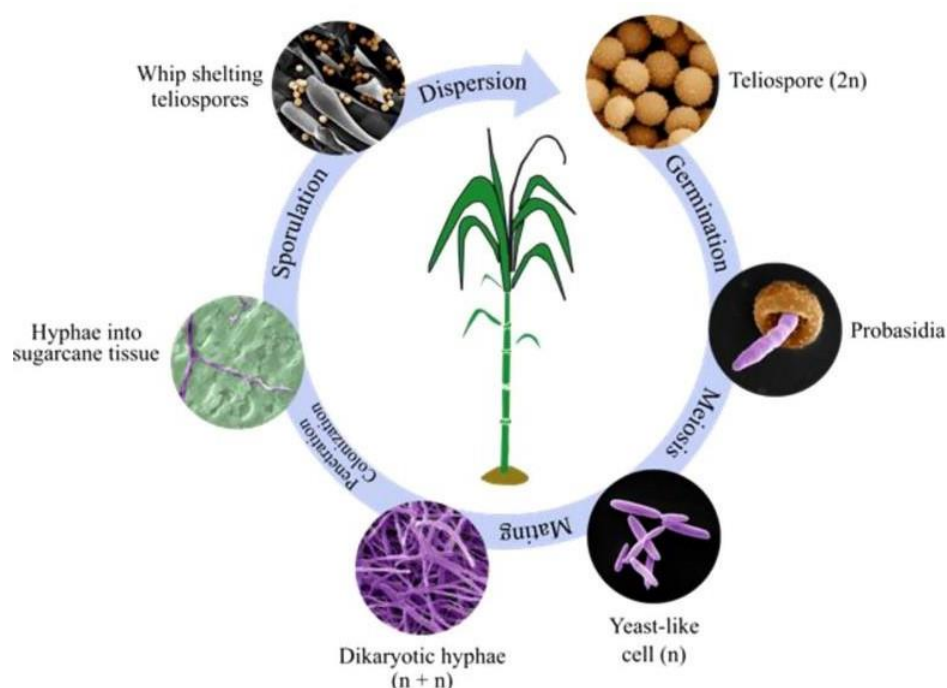


Figura 2 – Fases do ciclo de vida do fungo biotrófico *Sporisorium scitamineum*. Fonte: MONTEIRO-VITORELLO et al., 2018.

A compatibilidade sexual é dependente do sistema de *mating* que, em *S. scitamineum*, é bipolar e, portanto, produz esporídios de dois *mating-types* opostos nomeados de *MAT-1* e *MAT-2* (ZHU et al., 2019), sendo este sistema controlado pelos *loci* fisicamente ligados, *a* e *b*, nos quais estão localizados genes conservados, separados por 59 kpb, no cromossomo 2 (ALBERT; SCHENCK, 1996; TANIGUTI et al., 2015; BENEVENUTO, 2016). O *locus* bialélico *a* é responsável pelo reconhecimento e controle da fusão inicial entre células haplóides compatíveis e codifica um feromônio (*mfa*) e um receptor de feromônio (*pra*) (KRONSTAD; STABEN, 1997; YAN et al., 2016). Segundo Kronstad & Staben (1997), “a reação de cruzamento ocorre a partir da formação de tubos de conjugação das extremidades das células haplóides mediante o reconhecimento e resposta ao feromônio e posterior fusão para formação do dicário”. Após a fusão das células complementares haplóides, a manutenção da hifa dicariótica infectiva e a penetração no hospedeiro são controlados pelo *locus b*, onde estão localizados os genes homeóticos *bE* e *bW* responsáveis por codificar subunidades de um fator de transcrição heterodimérico essencial para regulação da expressão de genes que resultarão no crescimento

filamentoso e na patogenicidade, atuando através de uma cascata regulatória que modula diversas vias, dentre elas, a do ciclo celular (MARTÍNEZ-SOTO et al., 2020) (Figura 3). Entretanto, o fator de transcrição é ativado apenas quando as proteínas que o formam são advindas de diferentes alelos, sendo que diferenças nos domínios são responsáveis pelo autorreconhecimento (KRONSTAD; STABEN, 1997; YAN et al., 2016; MONTEIRO-VITORELLO, 2018).

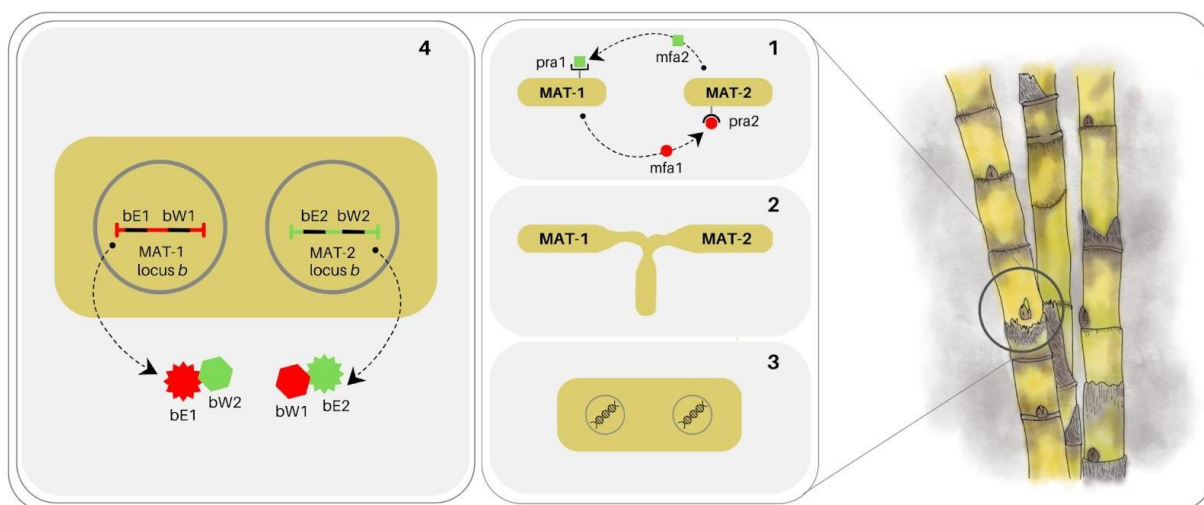


Figura 3 – Representação esquemática do processo de cruzamento sexual em *S. scitamineum*. Fonte: Autoria própria. 1 – Sinalização e reconhecimento molecular, 2 – Plasmogamia, 3 – Produção da hifa dicariótica infectiva, 4 – Heterodimerização de subunidades dos fatores de transcrição.

Em *Ustilago maydis*, agente causal da doença do carvão no milho, a dimerização deste fator de transcrição só ocorreu quando regiões variáveis da extremidade N-terminal se combinaram com o produto de um alelo diferente (KRONSTAD; STABEN, 1997). A hifa dicariótica, resultante da reação sexual entre *mating-types* compatíveis, diferencia-se em um apressório para penetração nos tecidos do hospedeiro, de forma inter- e intracelular, principalmente em tecidos epidérmicos indiferenciados como os meristemas durante a germinação, principal local de entrada do *S. scitamineum* no hospedeiro (SCHAKER et al., 2016; PETERS et al., 2017; MONTEIRO-VITORELLO, 2018, TEIXEIRA-SILVA, 2020). A formação do apressório é dependente da heterodimerização do fator de transcrição codificado pelo *locus b* de *mating-type*, apesar de evidências demonstrarem que a compatibilidade do *locus a* desempenha importante papel no desenvolvimento da patogenicidade (SCHAKER et al., 2016). O sistema bipolar de *mating*, como ocorre em *S. scitamineum*, favorece um nível elevado de homozigossidade na população, de forma a reduzir a variabilidade genética (BENEVENUTO et al., 2016). Estudos

anteriores utilizando marcadores determinaram que a estrutura da população é extremamente homogênea, informação corroborada por Benevenuto et al. (2016) quando demonstraram pouca variação nos isolados brasileiros. Apesar disso, ainda segundo os autores, foi identificada presença de variação dentro das próprias células germinadas a partir do mesmo teliósporo.

Segundo Rajput et al. (2021), a cana-de-açúcar, uma planta C4 com uma natureza poliplóide complexa, quando saudável, possui uma taxa fotossintética normal, entretanto, quando atacada pelo fungo, sofre alterações fisiológicas que resultam em taxas fotossintéticas alteradas, assim como um conteúdo reduzido de sacarose. Ainda, foram documentadas perdas na qualidade do suco e, apesar das perdas estarem relacionadas também aos diferentes cultivares e às variações climáticas, uma perda total da colheita pode ocorrer mediante condições favoráveis para a infecção, sendo que *S. scitamineum* afeta todos os parâmetros qualitativos e quantitativos em determinadas cultivares comerciais (RAJPUT et al., 2012). Danos médios em produtividade causados pelo fungo, reportados na Austrália por Magarey et al. (2010), já chegaram a 26%, demonstrando o potencial destrutivo do patógeno nesta cultura global e economicamente importante.

2.3. Genômica de microrganismos

O advento das tecnologias de sequenciamento e a disponibilidade de sequências genômicas alterou a percepção dos estudos envolvendo a interação molecular planta-patógeno e como os fungos do carvão estabelecem uma relação com seu hospedeiro (DUTHEIL et al., 2016). Segundo a Nature Portfolio, o sequenciamento de DNA é qualquer “procedimento químico, enzimático ou tecnológico para determinar a ordem linear das bases nucleotídicas do DNA”. Além disso, as tecnologias de sequenciamento de DNA foram responsáveis por fundamentar a área da genômica, sendo que o método de terminação de cadeia de Sanger permitiu o sequenciamento do genoma humano. Apesar de estudos descreverem as respostas de diferentes cultivares do hospedeiro em resposta ao ataque do patógeno, tanto a nível molecular quanto celular (MARQUES et al., 2018; HIDAYAH et al., 2021;), poucos isolados de *S. scitamineum* tiveram o genoma completamente sequenciados (TANIGUTI et al., 2015), sendo a maioria a nível de *contig* ou *scaffold* (QUE et al., 2014; DUTHEIL et al., 2016; ULLMANN et al., 2022).

Além disso, dentre as sequências depositadas no *National Center for Biotechnology Information (NCBI)*, a maioria é advinda de sequenciamento de *short-reads* (Acesso: PRJNA240344, PRJNA816553, PRJNA240344, PRJEB6265, PRJEB5169). Apesar do avanço significativo suscitado pelo surgimento de técnicas de sequenciamento por síntese e estratégias de paralelização, o tamanho pequeno dos fragmentos sequenciados representa um desafio (ZHANG et al., 2022) Segundo Moolhuijzen et al. (2021), a montagem de genomas fúngicos a partir de *short-reads* representa, de fato, uma dificuldade, devido às características regiões repetitivas presentes, enquanto que tecnologias de sequenciamento de *long-reads* representam uma oportunidade de superar alguns problemas, como a alta fragmentação das montagens *short-reads*. Apesar de as tecnologias de *short-reads* serem apoiadas por uma variedade de análises, protocolos e *pipelines*, além de alta acurácia, o tamanho dos fragmentos sequenciados dificulta a reconstrução dos genomas. Segundo Amarasinghe (2020):

A tecnologia de *long-reads* pode aprimorar a montagem *de novo*, a certeza do mapeamento, a identificação de isoformas de transcritos e a detecção de variações estruturais. [...] Essas capacidades, em conjunto com o contínuo progresso da acurácia e redução do custo, têm tornado o sequenciamento de *long-reads* uma opção para uma ampla gama de aplicações na genômica de organismos modelo e não-modelo.

As tecnologias de sequenciamento de *long-reads* já foram utilizadas, sozinhas ou em conjunto (montagem híbridas), em uma variedade de estudos envolvendo fungos, principalmente no que diz respeito à montagem de genomas fúngicos e de oomicetos (FAURE et al., 2020; MCKENZIE et al., 2020; SHAN et al., 2021; ZHANG et al., 2021). Kovaka et al. (2023) consideram que o ano de 2022 foi um ponto de inflexão em relação às tecnologias de sequenciamento *long-read* que agora representam o padrão-ouro no que diz respeito à acurácia, velocidade e preços competitivos, além disso, os autores vão ao encontro de Moolhuijzen et al. (2021) ao afirmar que as *reads* longas permitem com que regiões complexas e repetitivas do genoma sejam resolvidas. Ademais, os autores relembram que a publicação do primeiro genoma humano completo de telômero a telômero (SNURK et al., 2022) teve grande influência das *long-reads*. Snurk et al. (2022) confirmam que as *reads* “ultra-longas” geradas pela Oxford Nanopore foram responsáveis por permitir a montagem completa do centrômero do cromossomo humano Y e, mais

tarde, a montagem completa do cromossomo X. Apesar de os dados iniciais produzidos por tecnologias de sequenciamento de *reads* longas tenham tido grande taxas de erro, atualmente modos de alta acurácia foram responsáveis por melhorar significativamente a qualidade das *reads*, tanto na plataforma PacBio, quanto na plataforma Oxford Nanopore Technologies (KOVAKA et al., 2023).

O sequenciamento completo de genomas lançou as bases para o campo da genômica comparativa, uma abordagem holística que objetiva comparar duas ou mais sequências, buscando encontrar suas diferenças e similaridades (WEI et al., 2022). Para as análises de genômica comparativa, as sequências são comparadas ao serem alinhadas umas contra as outras e, desta forma, é possível identificar regiões homólogas baseadas na identidade das sequências comparadas. Segundo Stukenbrock & Dutheil (2012), “a sintenia descreve a extensão da colinearidade entre dois genomas e, portanto, refere-se ao grau de correspondência estrutural entre os cromossomos” sendo que a caracterização de tal sintenia depende inteiramente na qualidade das montagens. Genomas fúngicos possuem alta plasticidade genômica, o que indica que rearranjos cromossômicos podem ocorrer e diferenciar as sequências, dentre estes rearranjos, podem estar envolvidas duplicações, inversões e translocações de sequências, por exemplo (STUKENBROCK; DUTHEIL, 2012). Análises genômicas comparativas entre diferentes fungos do carvão foram conduzidas por Benevenuto et al. (2018) e possibilitaram detectar a presença de repertórios gênicos distintos entre as espécies, além de novos *insights* relacionados à especificidade do hospedeiro dentro do grupo, como, por exemplo, o fato de a especificidade ser provavelmente determinada, principalmente, pelas perdas e ganhos de genes efetores. Em *U. hordei*, há evidências de que transposons ativos foram responsáveis pela translocação de genes candidatos a efetores, visto que estes elementos transponíveis estão associados, em muitos casos, à plasticidade genômica e evolução rápida (ALI et al., 2014; BENEVENUTO et al., 2018; LAURIE et al., 2012). Ainda, estas análises comparativas reportaram, brevemente, a presença de translocações e inversões nos *scaffolds* que carregam os genes de *mating-type* em *S. reilianum* e *S. scitamineum*, além da presença de candidatos a efetores ortólogos entre várias espécies de carvão. Variações nas regiões de *mating* em espécies de fungo são importantes pois a reprodução sexual das leveduras só é iniciada a partir do reconhecimento de sua célula complementar para a fusão do citoplasma e,

posteriormente, fusão do núcleo para formação de células diplóides e, por esta razão, a comunicação dos feromônios no processo de cruzamento é imperativa (MAEKAWA; KANEKO, 2014). Maekawa & Kaneko (2014) já reportaram a existência de plasticidade envolvendo o locus de *mating* na espécie da levedura *Hansenula polymorpha* na forma de uma conversão de *mating-types*. Durante o cruzamento, os pesquisadores notaram que a região cromossômica entre os dois *loci* de *mating* foram invertidas, ocasionando em uma conversão no *locus* que estava sendo expresso. De acordo com os autores do estudo, este fenômeno já foi observado em outras espécies de fungo, como *Saccharomyces cerevisiae*, *Kluyveromyces lactis* e *Schizosaccharomyces pombe*. Além disso, autores como Bakkeren et al. (2008) já reportaram que a orientação dos genes do *locus b* no fungo do carvão *Ustilago hordei* parece estar invertida, além de sugerirem um papel preponderante dos elementos transponíveis no processo de evolução destes *loci* envolvendo o *mating*. Outras análises realizadas entre espécies de carvão, como com *Ustilago maydis* e *Sporisorium reilianum*, demonstraram que os genomas são altamente similares, e lançaram luz sobre a evolução deste grupo (WOLLENBERG; SCHIRAWSKI, 2014).

A publicação da primeira sequência genômica completa de telômero a telômero de um isolado brasileiro de *S. scitamineum* por Taniguti et al. (2015), combinando *reads* curtas (Illumina) e longas (PacBio), representou um ponto de inflexão para análises genômicas relacionadas à espécie. A montagem final do estudo apresentou por volta de 20 Mpb distribuídos em 26 *contigs*, sendo os últimos 3 *contigs* compostos majoritariamente por elementos repetitivos, levantando dúvidas sobre a possibilidade de serem *misassemblies*. O genoma do isolado SSC39B apresentou 6.677 genes codantes e 1.24% da sequência caracterizada como elementos repetitivos. Análises de genômica comparativa confirmaram *S. reilianum* como sendo a espécie mais próxima de *S. scitamineum*, demonstrando diferentes arranjos entre os cromossomos das duas espécies. Dentre estes arranjos, um dos mais notáveis, é um ponto de ruptura, enriquecido com elementos transponíveis, presente na região que liga fisicamente os *loci* de *mating-type* de *S. scitamineum* no cromossomo 2, já que este cromossomo é homólogo aos cromossomos 1 e 20, onde estão presentes os genes de *mating* em *S. reilianum*. As análises conduzidas por Taniguti et al. (2015) reportaram poucos polimorfismos entre isolados, corroborando

com achados posteriores sobre a baixa variabilidade genética da espécie como em Benevenuto et al. (2016).

2.4. A interação planta-patógeno

Um dos fatores críticos para a colonização de fitopatógenos em seus hospedeiros é a secreção de proteínas ou outras moléculas, chamadas de efetoras, que são capazes de modular a resposta de defesa da planta e, portanto, permitir a progressão da infecção (HOGENHOUT et al., 2009). Esse arsenal de moléculas irá trabalhar de forma a aumentar o fitness do fungo em frente ao sistema imune do hospedeiro (TEIXEIRA-SILVA, 2020). Segundo Hogenhout et al. (2009), “métodos computacionais robustos aplicados a dados de sequências genômicas de fitopatógenos resultaram em vastos catálogos de genes efetores putativos”. Em *S. scitamineum*, estudos recentes demonstraram que os genes candidatos a efetores atuam em vários processos e em diversos compartimentos subcelulares, algo comum para patógenos filamentosos, além de serem afetados pelo genótipo do hospedeiro (TEIXEIRA-SILVA, 2020). Em trabalhos anteriores, como de Benevenuto et al. (2016), polimorfismos foram identificados em vários genes candidatos a efetores. A presença de raças pode estar envolvida com estes polimorfismos nestes genes candidatos, mas resultados anteriores não permitem inferir a emergência de novas raças, sendo que o próprio conceito ainda é um tópico controverso para essa espécie (BENEVENUTO et al., 2016) como discutido em Schenck (2003). Apesar disso, quando analisando populações de isolados brasileiros de *S. scitamineum*, Benevenuto et al. (2016) observaram a formação de dois grandes grupos geneticamente distintos no país.

Patógenos biotróficos precisam ser capazes de ultrapassar as defesas do hospedeiro ao mesmo tempo em que causam danos fisiológicos mínimos para estabelecer uma colonização bem-sucedida (MAIA et al., 2022). Moléculas efetoras, como descritas anteriormente, são capazes de modular a resposta metabólica do hospedeiro e por isso são de extrema importância. Estudo realizado por Maia et al. (2022) demonstrou que, dentre uma seleção de genes candidatos a efetores, há evidência da presença de parálogos no genoma de *S. scitamineum*. Além disso, estes parálogos de genes candidatos a efetores estão organizados em *tandem* no genoma, de forma semelhante ao genoma de *S. reilianum* (MAIA et al., 2020). Estes achados hipotetizaram que os candidatos a efetores de *S. scitamineum* passaram

por rodadas de duplicações em pequena escala e posterior retenção, provavelmente orquestrada por elementos transponíveis, indo ao encontro de Benevenuto et al. (2018) quando discorre a respeito da associação destes elementos com a plasticidade genômica. Visto que patógenos biotróficos lançam mão da utilização de uma variedade de efetores em momentos distintos do desenvolvimento da doença para permitir sua colonização, a caracterização destas sequências pode fornecer informações valiosas sobre a biologia do fungo (JASWAL et al., 2020).

Estudo anterior de Oliveira (2021) (dados não publicados) reportou que dois isolados brasileiros de *S. scitamineum* pertencentes a dois haplótipos diferentes (BENEVENUTO et al., 2016) diferem quanto à sua agressividade. No estudo, o isolado SSC04, coletado na Bahia em 2009, foi capaz de induzir o desenvolvimento precoce do chicote em plantas hospedeiras da variedade RB925345, considerada de resistência intermediária, quando comparado ao isolado SSC39, coletado em Ribeirão Preto. Além disso, segundo Oliveira (2021), o número de chicotes em plantas inoculadas com o isolado SSC04 foi quatro vezes maior do que naquelas inoculadas com o isolado SSC39, corroborando hipóteses levantadas por Crestana (2017) sobre a natureza distinta da agressividade dos haplótipos. De acordo com Castiblanco et al. (2020), apesar de a agressividade ser caracterizada por herança genética complexa, ela é frequentemente analisada ao avaliar os sintomas em tempos diferentes.

Para Nayaleni et al. (2021), “a severidade da doença depende de dois grandes fatores, isto é, o grau de resistência do genótipo do hospedeiro e o nível de virulência do patógeno”. A patogenicidade pode ser definida como a capacidade do patógeno em causar doença, enquanto a agressividade está ligada a aspectos quantitativos da interação planta-patógeno (VIEIRA et al., 2019). Já o conceito de grau de resistência está relacionado à capacidade do hospedeiro em superar, de forma parcial ou integral, o efeito do patógeno. De acordo com Pariaud et al. (2009), na literatura fitopatológica, a agressividade é frequentemente utilizada como um sinônimo de traços quantitativos relacionados à patogenicidade, expostos em características como, a eficiência da infecção, período latente e a taxa de esporulação, por exemplo. De acordo com Oliveira (2021), o isolado SSC04 proporcionou um período de latência menor na mesma variedade de hospedeiro em comparação ao isolado SSC39, o que, de acordo com Rago et al. (2009), pode ser considerado como uma diferença na agressividade. Apesar de variações da

agressividade em populações do patógeno já terem sido identificadas no Estado de São Paulo, os estudos acerca do tópico são de tortuosa execução, visto a ausência de metodologias claras e difícil reprodução dos resultados, ocasionando na impossibilidade de comparação entre diferentes estudos (RAGO et al., 2009). Além disso, a interação entre o hospedeiro e o fungo sofre influências ambientais, tornando complexa a tarefa de avaliar o nível de agressividade de isolados (BHUIYAN et al., 2021).

Ainda que o estudo das bases genéticas da agressividade em fungos seja escasso, e que seus componentes tenham sido pouco estudados em fungos fitopatogênicos, a natureza genética da resistência no hospedeiro é provavelmente complexa e, conseqüentemente, o controle genético da agressividade do patógeno envolverá diversos fatores, sendo que estudos sugerem que seus componentes são determinados de forma poligênica (PARIAUD et al., 2009). Segundo Rago et al. (2009), em patossistemas controlados por fatores quantitativos, é comum que as variedades se comportem de maneiras distintas, mediante influências ambientais, no que diz respeito às doenças. É sabido que o estado fisiológico do hospedeiro afeta a progressão e severidade da doença, assim como o estado fisiológico do patógeno pode resultar em variações na agressividade (PARIAUD et al., 2009; SUNDAR et al., 2012). Dentre os fungos causadores da doença do carvão, as bases genéticas da agressividade foram examinadas em espécies como *Ustilago hordei*, ocasião em que se sugeriu um sistema poligênico, além de diferenças em *mating-types* segregantes que podem indicar a existência de um fator ligado ao *mating* que afeta a agressividade (CATEN et al., 1984). Ainda, evidências para determinação poligênica da agressividade em *Ustilago maydis* remonta décadas (BLANCH et al., 1981).

Ademais, segundo Tian et al. (2017), a investigação da heterozigosidade é um processo importante, visto que podem ocorrer mudanças rápidas de virulência em isolados que são heterozigotos para essa característica. De acordo com Halisky (1965), o comportamento dos fenótipos patogênicos dentre os fungos do carvão é baseado na constituição genotípica. Ainda segundo o autor, o número de genes que governa esta característica terá consequência direta na variação patogênica.

Apesar de estudo anterior de Benevenuto et al. (2016) não ter encontrado associação entre a distância genética e geográfica entre os isolados

brasileiros e nenhuma presença de polimorfismo na região ITS, foi reportada a existência de variabilidade entre os isolados. Mais recentemente, Thushari & Costa (2023) encontraram variabilidade entre os isolados do Sri Lanka, apesar de os autores admitirem que a razão dessa variabilidade pode ter sido causada pelo uso de teliósporos. Populações de outros fungos biotróficos patogênicos, como *Puccinia striiformis*, demonstraram possuir baixa diversidade genética, entretanto, os autores afirmam que um melhor entendimento sobre a variação de virulência será essencial no desenvolvimento de cultivares resistentes (ZHENG et al., 2013).

Anteriormente, genes candidatos a efetores selecionados por Teixeira-Silva (2018) demonstraram polimorfismos entre isolados de *S. scitamineum* de diferentes países como África do Sul, Austrália, Brasil e China. Dentre estes genes candidatos, dois candidatos a efetores selecionados por Oliveira (2021), *g1052* e *g5159*, demonstraram ser polimórficos entre os isolados SSC04 e SSC39. Além disso, uma árvore de haplótipos construída a partir de isolados brasileiros considerando o gene candidato a efector *g4257*, em estudo realizado por Ferreira (2019), colocou os isolados SSC04 e SSC39 como pertencentes a haplótipos distintos indo ao encontro dos achados de Benevenuto et al. (2016). Estes achados são importantes visto que proteínas efectoras são, provavelmente, componentes essenciais envolvidos na agressividade de um patógeno (FABRE et al., 2019).

Segundo Monteiro-Vitorello et al. (2018), apesar de diversos estudos terem tentado acessar a variabilidade de *S. scitamineum*, suas bases genéticas ainda são pouco entendidas e, por esta razão, pesquisas envolvendo a natureza dos polimorfismos e variações estruturais serão essenciais na discussão a respeito de raças patogênicas.

3. OBJETIVO GERAL

O presente projeto visou sequenciar e comparar os genomas haploides isolados de células sexualmente compatíveis de dois indivíduos 2n (teliósporos) do fungo *Sporisorium scitamineum* contrastantes quanto à agressividade em cana-de-açúcar: as células do isolado menos agressivo SSC39A (*MAT-2*), SSC39B (*MAT-1*) e do isolado mais agressivo SSC04A (*MAT-2*), SSC04B (*MAT-1*). Buscando, desta forma, elucidar se o nível de polimorfismos e/ou heterozigosidade são determinantes para a agressividade do patógeno.

4. OBJETIVOS ESPECÍFICOS

Este projeto teve como objetivos específicos:

- 1) Isolar o DNA dos isolados SSC04A (*MAT-2*), SSC04B (*MAT-1*), SSC39A (*MAT-2*), SSC39B (*MAT-1*);
- 2) Utilizar a plataforma Oxford/Nanopore de sequenciamento de *reads* longas;
- 3) Utilizar ferramentas computacionais para montagem dos cromossomos do fungo, predição dos genes e análises comparativas;
- 4) Identificar genes polimórficos;
- 5) Avaliar o nível de heterozigosidade;

5. MATERIAIS E MÉTODOS

5.1. Material biológico

Teliósporos foram coletados a partir de chicotes da variedade de cana-de-açúcar RB925345 referentes ao isolado 04 e 39 e disponibilizados no Laboratório de Genética de Microrganismos João Lúcio de Azevedo” no campus da Escola de Agricultura Luiz de Queiroz (ESALQ) da Universidade de São Paulo. Após coleta, a estrutura dos chicotes foi colocada em papel pardo e disposta em estufa a 37°C para secagem e posterior armazenamento em microtubos plásticos de 1,5 mL selados com plástico filme. O armazenamento dos microtubos foi realizado em recipiente de vidro contendo sílica e selado com vaselina, como disposto em Longatto (2014).

Teliósporos foram colocados em 1 mL de água miliQ e tratados com sulfato de estreptomicina (5 mg/mL) e tetraciclina (2,5 mg/mL) durante 1 hora. Posteriormente, microtubos contendo a solução foram centrifugados a 8000 rpm por 5 minutos. A solução foi ressuspensa em 1 mL de solução NaCl 0,85 Mol L⁻¹ para posterior inoculação em meio de cultura.

5.2. Obtenção de células leveduriformes haploides

As células leveduriformes haplóides do isolado 04 referentes aos *mating-types MAT-1* e *MAT-2* foram retiradas de estoque em glicerol 60% a -80°C. 50µL da solução contendo as células foi pingada em placa de Petri contendo meio

YM (*yeast universal medium*) sólido (3 g L⁻¹ de extrato de levedura, 3 g L⁻¹ de extrato de malte, 5 g L⁻¹ de peptona, 10 g L⁻¹ de glicose e 15 g L⁻¹ de ágar) e estriada com alça Drigalski. Placas foram posteriormente incubadas em BOD a 28°C.

Solução de teliósporos tratados anteriormente com antibióticos foram utilizados para obtenção das células leveduriformes haplóides do isolado 39. A solução foi inoculada em 50 mL de meio líquido YM (3 g L⁻¹ de extrato de levedura, 3 g L⁻¹ de extrato de malte, 5 g L⁻¹ de peptona, 10 g L⁻¹ de glicose) contendo 25µL de sulfato de estreptomicina (100mg/mL) e 250µL tetraciclina (5mg/mL). O meio com os teliósporos foi incubado a 28°C e 200rpm durante 24 horas.

5.3. Isolamento dos *mating-types*

Após germinação e proliferação das células haploides leveduriformes provenientes dos teliósporos, foram realizadas diluições seriadas para separação das colônias e posterior isolamento dos diferentes *mating-types* para realização de testes de cruzamento. Diluições seriadas de 10⁻¹ a 10⁻⁶ foram realizadas e inoculadas, novamente, em meio YM sólido. A partir destas diluições, colônias foram isoladas e repicadas em placas contendo meio YM sólido contendo 25µL de sulfato de estreptomicina (100mg/mL) e 250µL tetraciclina (5mg/mL).

5.4. Teste de cruzamento *in vitro*

A avaliação quanto ao *mating-type* de cada colônia isolada foi realizada em placas de Petri em testes de cruzamentos para confirmação do tipo sexual. Uma porção de cada colônia isolada foi coletada com palito estéril e diluída em microtubos (um para cada colônia) contendo 900µL de solução NaCl 0,85 Mol L⁻¹. Para realização dos testes de cruzamento *in vitro* foi realizada adaptação de Carvalho et al. (2016), 10µL da suspensão de determinada colônia foram misturados com mais 10µL da suspensão de outra colônia em microtubo separado, de forma a cobrir todas as possibilidades de cruzamento. Esta solução foi inoculada em meio YM sólido a 28°C. Como descrito em Albert & Schenck (1996), a fusão de hifas a partir de células leveduriformes haploides compatíveis resulta na formação de uma colônia cotonosa, indicando reação sexual positiva, enquanto que células não-compatíveis (portanto, de mesmo *mating-type*) produzem colônias lisas sem a

presença de hifas quando observadas em microscopia de luz, mas sim com a aparência de células leveduriformes.

5.5. Microscopia de luz

A coleta de dados sobre as células leveduriformes haplóides e o cruzamento compatível entre os tipos de reações sexuais com formação de hifa dicariótica foi feita através de análise em microscopia óptica pelo OLYMPUS – BX40F4. As imagens foram capturadas com câmera de vídeo acoplada ao microscópio de luz.

5.6. Obtenção de DNA de alto peso molecular e preparo de biblioteca genômica

Para obter uma quantidade significativa de células para extração de DNA genômico e posterior sequenciamento, foi realizado um pré-inóculo dos isolados 04 (*MAT-1* e *MAT-2*) e 39 (*MAT-1* e *MAT-2*) do fungo *Sporisorium scitamineum* em 50mL de meio YM líquido submetido a 28°C e agitação de 200 rpm durante 24 horas. Após 24 horas, 100µL do pré-inóculo foram adicionados em 50mL de meio YM líquido divididos entre 5 erlenmeyers para cada isolado, totalizando 250mL de meio líquido tanto para o isolado 04 quanto para o 39. Após o crescimento do inóculo, o meio de cultura de cada erlenmeyer foi vertido em tubo Falcon de 50mL e centrifugado a 8000g por 15 minutos. O sobrenadante foi descartado do tubo e o *pellet* formado pelas células centrifugadas foi lavado com solução estéril de Tris 10mM e EDTA 1mM. Células foram então estocadas a -20°C até utilização para extração de DNA genômico.

Para extração de DNA genômico, o material fúngico foi transferido para microtubos de 1,5µL para seguimento de protocolo de extração de DNA genômico total para obtenção de moléculas de alto peso molecular objetivando o sequenciamento de *reads* longas (SCHWESSINGER; RATHJEN, 2017) na plataforma Oxford Nanopore MinION. Análises de integridade das moléculas de DNA foram realizadas a partir de eletroforese em gel convencional (0.8% agarose) com tampão TBE 0,5X. A quantificação foi realizada a partir do fluorômetro Qubit 2.0 (Invitrogen, ThermoFisher Scientific, Waltham, MA, USA) e NanoDrop (ThermoFisher Scientific, Wilmington, Delaware, USA). A biblioteca de sequenciamento MinION foi

preparada de acordo com o kit Genomic DNA by Ligation (SQK-LSK109) (Oxford Nanopore Technologies), NEBNext Ultra II End-prep reaction/FFPE DNA Repair (New England Biolabs), Ligation Buffer (LNB), NEBNext Quick T4 DNA Ligase, e Adapter Mix (AMX), de acordo com as instruções da fabricante. A biblioteca final foi quantificada e 5-50 fmol utilizados para carregar a *flow cell* (MinION Flow Cells (R9.4.1) (Oxford Nanopore Technologies; FLO-MIN106_SQK).

5.7. Análise de discriminação molecular de *mating-types* opostos

Células leveduriformes haploides pertencentes a cada *mating-type* e a cada isolado, 04 ou 39, crescidas da forma como descrito na seção anterior foram utilizadas para extração de DNA seguindo protocolo Doyle & Doyle (1987). Foram utilizados três *primers* (Tabela 1), como descritos por Agisha et al. (2022), para amplificação de alelos *bE1* e *bE2* do gene *bE*, gene responsável por codificar a subunidade de um fator de transcrição heterodimérico envolvido na reação de *mating-type* de *Sporisorium scitamineum*. A visualização dos *primers* foi realizada pela plataforma Benchling.

Tabela 1 — *Primers* utilizado por Agisha et al. (2022) para amplificação dos alelos referentes ao gene *bE*, presente no locus *b* de *mating-type*.

Nome	Orientação	Sequência	Tamanho do amplicon (pb)	Alelo
<i>bE1F1</i>	<i>Forward</i>	5'-ATGCGTGAATTTGCGC-3'	668	MAT-1
<i>bE1-2R1</i>	<i>Reverse</i>	5'-TTGATGAACCAGAGCGTGAG-3'	-	
<i>bE2F2</i>	<i>Forward</i>	5'-CTTTCCTACAATCCAAACCAATA-3'	458	MAT-2

Fonte: AGISHA et al. (2022).

Para amplificação dos *primers*, foi utilizada reação em termocicladora seguindo os parâmetros de desnaturação inicial a 94°C por 5 minutos, 35 ciclos de desnaturação a 94°C por 1 minuto, anelamento a 56°C por 30 segundos e polimerização a 72°C por 1 minuto, seguido de alongação final a 72°C por 5 minutos. Os produtos de PCR provenientes desta amplificação foram visualizados em gel de agarose convencional 1% com tampão TAE 1X. A purificação destes produtos foi realizada seguindo protocolo do kit Wizard® SV Gel and PCR Clean-Up System (Promega) e encaminhada para sequenciamento Sanger. Sequências advindas de

sequenciamento Sanger foram analisadas pelo *software* Sequencher v5.0.0 e pelo CLC Genomics Workbench v. 20 (QIAGEN).

5.8. Confirmação molecular da identidade celular dos isolados de *Sporisorium scitamineum*

Estudo anterior relacionado aos genes candidatos efetores de *S. scitamineum* identificou polimorfismos no gene *g5159* nas posições 689 e 690, responsáveis por causar uma mutação não-sinônima de uma tirosina para um triptofano (TEIXEIRA-SILVA, 2018). Estas modificações demonstraram ser isolado-específicas e por isso foram utilizados *primers* para amplificar a região (Tabela 2) e confirmar a identidade das células leveduriformes haploides do isolado 04 e 39 para o sequenciamento. A visualização dos *primers* foi realizada pela plataforma Benchling.

Tabela 2 — *Primers* utilizado para amplificação da sequência completa do gene candidato a efetor *g5159* de *S. scitamineum*.

Nome	Orientação	Sequência	Tamanho do amplicon (bp)
g5159_F	Forward	5'-GAAAAAGGGGAGGGATCTA-3'	1511
g5159_F2	Forward	5'-CTGAGCCTGAACCAAAACGC-3'	991
g5159_R	Reverse	5'-AGTCTCAAACACACAGGCAA-3'	-

Para amplificação dos primers, foi utilizada reação em termocicladora seguindo os parâmetros de desnaturação inicial a 94°C por 3 minutos, 35 ciclos de desnaturação a 94°C por 45 segundos, anelamento a 55°C por 30 segundos e polimerização a 72°C por 1 minuto e 16 segundos, seguido de elongação final a 72°C por 10 minutos. Os produtos de PCR provenientes desta amplificação foram visualizados em gel de agarose convencional 1% com tampão TBE 0,5X. A purificação destes produtos foi realizada seguindo protocolo do kit Wizard® SV Gel and PCR Clean-Up System (Promega) e encaminhada para sequenciamento Sanger. Sequências advindas de sequenciamento Sanger foram analisadas pelo *software* Sequencher v5.0.0 e pelo CLC Genomics Workbench v. 20. (QIAGEN).

5.9. Sequenciamento e montagem dos genomas

O acompanhamento em tempo real do sequenciamento e estatísticas foram obtidas através do *software* MinKNOW 22.12.7 e visualizadas com o pacote MinIONQC v1.4.2. Guppy v6.4.6 (Oxford Nanopore Technologies) foi utilizado para o *basecalling* das *reads* através do modo *super high accuracy (sup)*. Após *basecalling*, o *software* Porechop v0.2.4 realizou a excisão dos adaptadores presentes nas *reads* provenientes do processo de preparo da biblioteca. Foram utilizados dois *softwares* de montagem de genomas *de novo* em busca dos melhores resultados para o conjunto de dados, sendo eles, Flye v. 2.9 (KOLMOGOROV, 2019) e, Canu v.1.8 (KOREN et al., 2017). Para o Canu, foram utilizados os parâmetros *genomeSize=20m*, *correctedErrorRate=0.144* (14% de erro), e *-nanopore-raw* para especificar as *reads* da plataforma Oxford Nanopore Technologies (ONT). Para o Flye, foram utilizados parâmetros padrões.

Sequências resultantes dos *softwares* descritos foram polidas com os *softwares* RACON (<https://github.com/isovic/racon>) e o Medaka v.1.2.2 (Oxford Nanopore Technologies) (<https://github.com/nanoporetech/medaka>). Para analisar a completude dos genomas foi utilizado o *software* BUSCO v.5.4.3 (MANNI, 2021) com banco de dados “*basidiomycota_odb10*”. O controle de qualidade foi realizado pelo QUAST (GUREVICH et al., 2013).

5.10. Análises de sintenia

Alinhamentos de sequências genômicas e gênicas foram realizados a partir do *software* MCScanX v.1.3.6 (WANG et al., 2012) para avaliação da sintenia, em conjunto com o CLC Genomics Workbench v. 20 (QIAGEN) (parâmetros padrões) para resolução dos *loci* de *mating-type*.

5.11. Predição gênica

Genes foram preditos para os quatro genomas dos isolados SSC04 e SSC39 utilizando o *pipeline* FunGAP v. 1.1.1 (MIN et al., 2017). Antes da predição, os genomas foram mascarados utilizando os *softwares* RepeatModeler v1.0.11 (<http://www.repeatmasker.org>) e RepeatMasker v4.1.0 (<http://www.repeatmasker.org>) para resolução de regiões repetitivas.

O modelo de *Ustilago maydis* foi utilizado como *input* para o Augustus dentro do *pipeline* FunGAP, enquanto o proteoma irmão foi recuperado de Taniguti et al. (2015). *Reads* de *RNAseq* do fungo *S. scitamineum in vitro* (Taniguti et al., 2015) foram usadas como evidência extrínseca. Por fim, o conjunto de dados “basidiomycota_odb10” pelo BUSCO avaliou a completude da predição. A ferramenta DensityMap (<https://github.com/sguizard/DensityMap>) foi utilizada para plotar a visualização da densidade gênica ao longo dos cromossomos. A predição funcional dos genes relacionados aos *loci* de *mating-type* foi realizada a partir do banco de dados do UniProt, Interpro e NCBI.

5.12. Predição de órtólogos

Em busca de encontrar a correspondência dos órtólogos para cada um dos genomas, *best hits* dos genes órtólogos foram avaliados utilizando BLASTp *all-against-all* com um cutoff e^{-05} , mínimo de 40% de identidade e 80% de cobertura, seguido por ranqueamento por *bitscore*.

5.13. Análise de SNVs (Single Nucleotide Variants)

A descoberta de variantes entre as montagens dos genomas foi realizada por ambos os *softwares* Medaka v1.2.2 e Clair3 v1.0.2 (ZHENG et al., 2022). Antes da descoberta de variantes os genomas foram mascarados pelos *softwares* RepeatModeler v1.0.11 (<http://www.repeatmasker.org>) e RepeatMasker v4.1.0 (<http://www.repeatmasker.org>).

Para uma variante ser considerada, em ambos os programas, foi aplicado um filtro de *phred score* maior que 20 e uma cobertura mínima de 20X através do *software* BCFtools v.1.8 (LI, 2011). A ferramenta *bcftools isec* foi utilizada para encontrar as variantes em comum, ou seja, presentes em ambas as análises. Para análise de variantes entre as células complementares do isolado SSC04, o genoma SSC04A MAT-2 (de maior cobertura) foi utilizado como referência, enquanto para o isolado SSC39, o genoma SSC39B MAT-1 (Taniguti et al., 2015) foi a referência. Para comparação das variantes entre todos os genomas, o genoma SSC39 MAT-1 foi utilizado como referência. O pacote CMap do R v4.3.0 (<https://www.bioconductor.org/packages/release/bioc/html/cmapR.html>) e o Circos v0.69-9 (KRZYWINSKI et al., 2009) realizaram a projeção das variantes ao longo

dos cromossomos, além disso o VCFtoolz 1.2.0 (DAVIS, 2019) foi utilizado para o *plot* comparativo de variantes e diagramas de Venn. A anotação funcional das variantes foi realizada pelo *software* SnpEff v.4.3 (CINGOLANI et al., 2012).

6. RESULTADOS

6.1. Cruzamento *in vitro* das células leveduriformes haplóides de *mating-type* opostos

Quando inoculadas em meio de cultura YM, as colônias referentes a *mating-types* opostos e complementares apresentaram uma morfologia branca e cotonosa com um halo de hifas ao redor, indicando cruzamento positivo, evidenciado pela transição do estado unicelular para um estado multicelular. Colônias cruzadas com o mesmo *mating-type* apresentaram uma morfologia lisa e sem presença de hifas, indicando ausência de cruzamento. Células leveduriformes haplóides de isolados diferentes foram capazes de cruzar entre si, indicando não haver nenhuma barreira para o cruzamento (Figura 4).

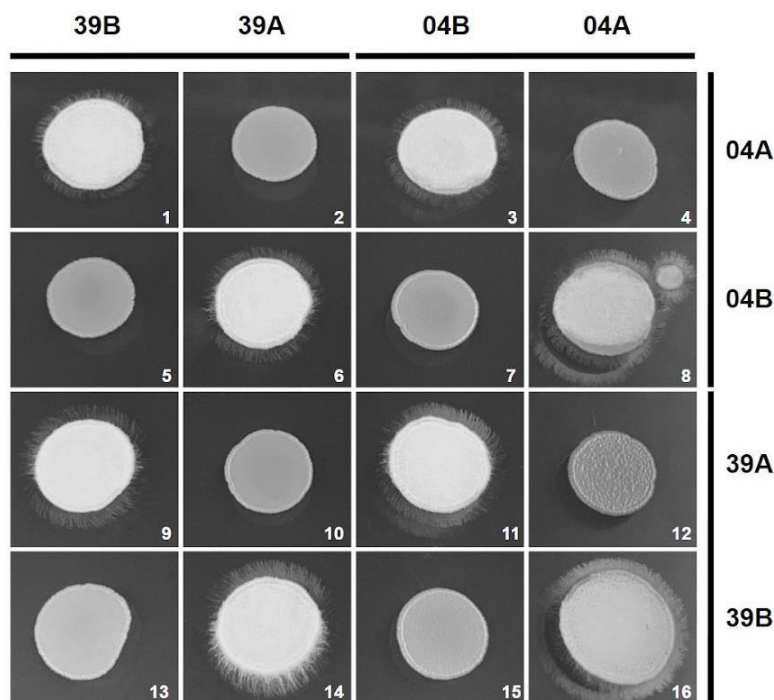


Figura 4 — Teste de compatibilidade entre os isolados 04 e 39 e seus respectivos *mating-types*. A denominação “A” refere-se ao alelo *MAT-2*, enquanto a denominação “B” refere-se ao alelo *MAT-1*. Reações sexuais positivas apresentaram uma colônia de morfologia branca e cotonosa com um halo de hifas (1,3,6,8,9,11,14,16), enquanto reações negativas apresentaram colônias lisas e leveduriformes (2,4,5,7,10,12,13,15). Cruzamentos ocorreram em meio YM sólido a 28°C por 3 dias. Isolados são capazes de hibridização.

Colônias com cruzamento positivo analisadas em microscopia óptica confirmaram a transição do estado celular, apresentando longas hifas emergindo das extremidades, resultado da plasmogamia entre leveduras complementares (Figura 5A). As colônias com cruzamento negativo apresentaram células com morfologia de levedura unicelular ao longo de toda a extensão, permanecendo juntas em uma colônia coesa (Figura 5B).

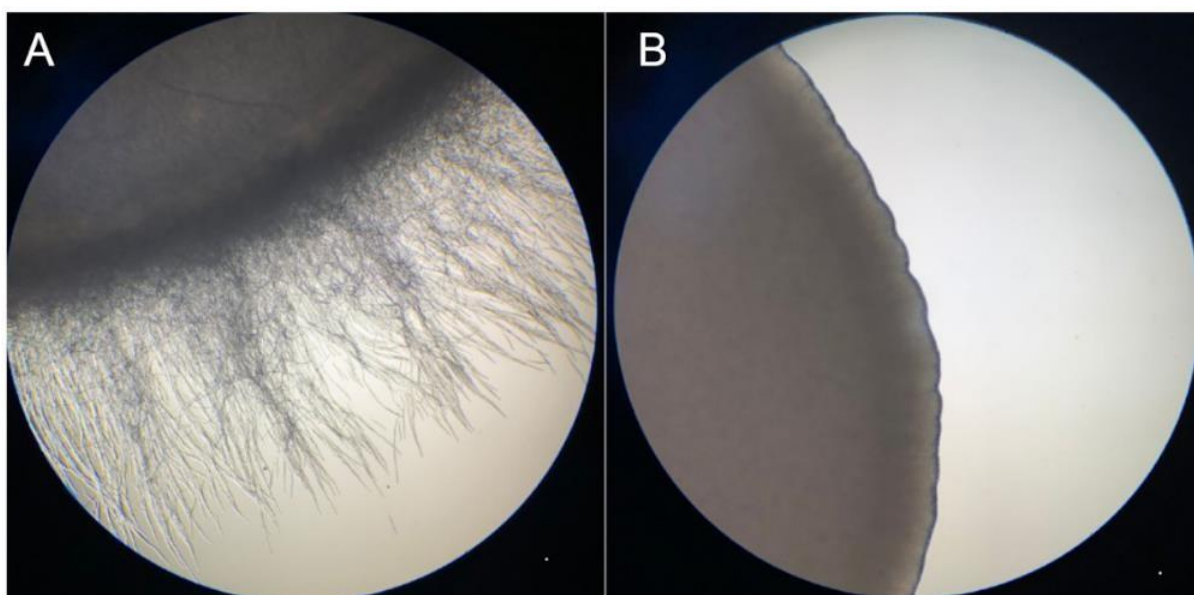


Figura 5 — Identificação visual cruzamento *in vitro* dos isolados SSC04A/SSC04B e SSC39A/SSC39B. A – Reação positiva de cruzamento entre dois *mating-types* diferentes sob microscopia de luz, aumento de 40x. É possível observar a formação de hifas, gerando aspecto cottonoso em placa. B – Células leveduriformes haplóides sob microscopia de luz, aumento de 40x. É possível observar reação ausente de cruzamento, quando dois *mating-types* idênticos são cruzados. Crescimento foi realizado com meio YM sólido e inóculo crescido a 28°C. Fotos realizadas em microscópio de luz OLYMPUS – BX40F4.

6.2. Extração de DNA total para Reação em Cadeia da Polimerase (PCR)

A extração de DNA genômico total com protocolo Doyle & Doyle (1987) do isolado SSC04A (*MAT-2*) resultou em 5820,5 ng/μL, enquanto SSC04B (*MAT-1*) resultou em 5430,8 ng/μL (Tabela 3). A amostra do isolado SSC39A (*MAT-2*) resultou em 2611,2 ng/μL, enquanto SSC39B (*MAT-1*) resultou em 2081,4 ng/μL (Tabela 4). Ambos isolados apresentaram espectro de absorbância esperado para leitura entre 220 a 350 nm (Figura 6 e Figura 7). A razão 260/280 de todas as amostras apresentaram o valor acima do recomendado de, aproximadamente, 1,8, esperado para uma amostra de DNA “puro”, segundo o boletim técnico (T123) da

ThermoFisher Scientific (Tabelas 3 e 4). Apesar disso, segundo o boletim, altas razões não indicam necessariamente um problema.

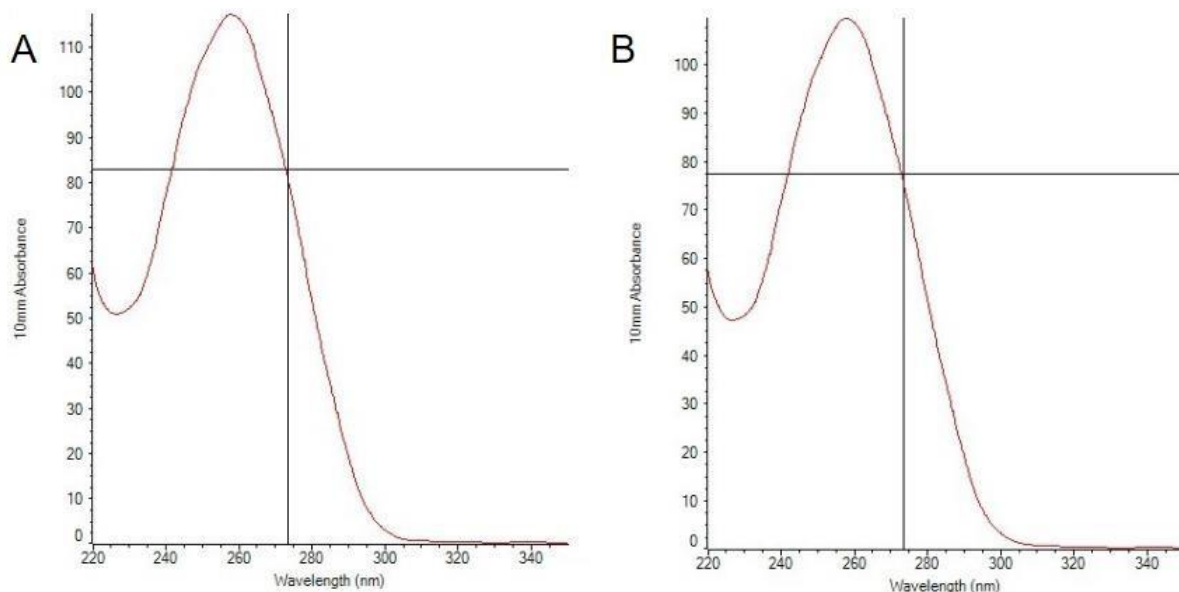


Figura 6 — Quantificação das amostras de extração de DNA genômico total medido pelo espectrofotômetro NanoDrop 2000/2000c. A — Espectro de absorbância e comprimento de onda da amostra SSC04A. B — Espectro de absorbância e comprimento de onda da amostra SSC04B. De acordo com a ThermoFisher Scientific, os ácidos nucleicos possuem absorbância máxima em 260 nm.

Tabela 3 — Quantificação de extração de DNA genômico total das amostras SSC04 medido pelo espectrofotômetro NanoDrop 2000 e razões.

	SSC04A (MAT-2)	SSC04B (MAT-1)
Concentração (ng/μL)	5820,5	5430,8
260/280	2,15	2,10
260/230	2,24	2,27

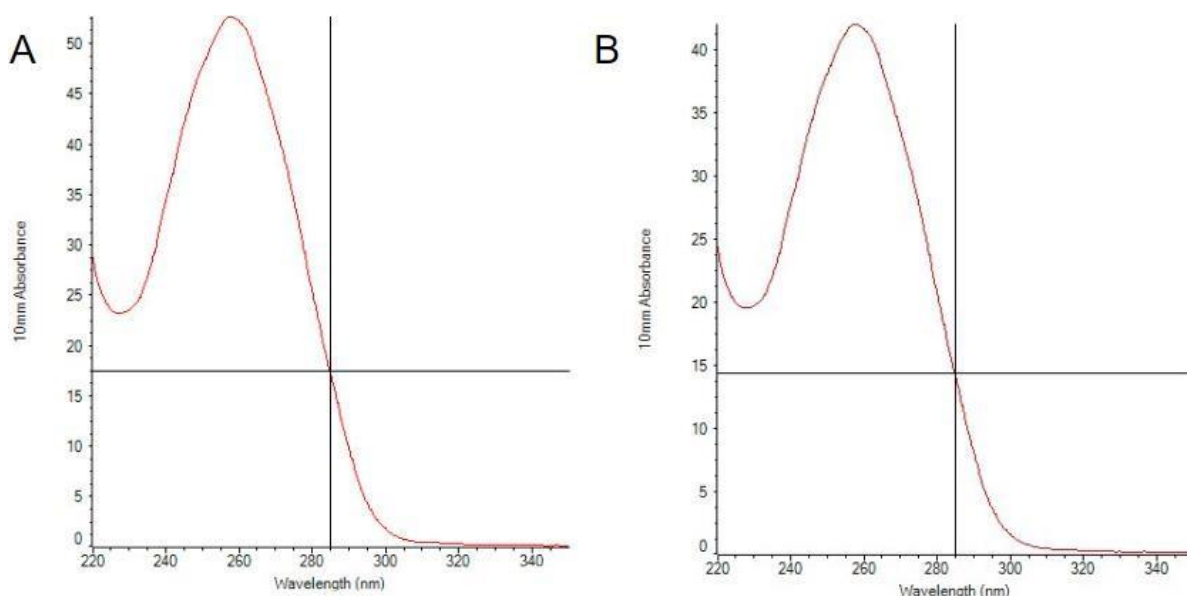


Figura 7 — Quantificação das amostras de extração de DNA genômico total medido pelo espectrofotômetro NanoDrop 2000/2000c. A — Espectro de absorvância e comprimento de onda da amostra SSC39A. B — Espectro de absorvância e comprimento de onda da amostra SSC39B. De acordo com a ThermoFisher Scientific, os ácidos nucleicos possuem absorvância máxima em 260 nm.

Tabela 4 — Quantificação de extração de DNA genômico total das amostras SSC39 medido pelo espectrofotômetro NanoDrop 2000 e razões.

	SSC39A (MAT-2)	SSC39B (MAT-1)
Concentração (ng/μL)	2611,2	2081,4
260/280	2,04	2,00
260/230	2,23	2,13

6.3. Discriminação molecular de *mating-types* opostos

A discriminação molecular de *mating-types* opostos foi conduzida a partir da amplificação da região N-terminal variável do gene *bE*, presente no *locus b* de *mating-type* no cromossomo 2. No isolado SSC04, os *primers forward* *bE1F1* e o *reverse* *bE1-2R1*, utilizados para identificação do *MAT-1*, produziram um *amplicon* entre 750 e 1000 pb (Figura 8). Enquanto os *primers forward* *bE2F2* e o *reverse* *bE1-2R1*, utilizados para identificação do *MAT-2*, produziram um *amplicon* entre 500 e 750 pb (Figura 9). No isolado SSC39, os *primers bE1F1/bE1-2R1* produziram um *amplicon* entre 750 e 1000 pb, enquanto os *primers bE2F2/bE1-2R1* produziram um *amplicon* entre 500 e 750 pb. Desta forma, os *primers* utilizados por Agisha et al. (2022) foram capazes de distinguir *mating-types* diferentes.

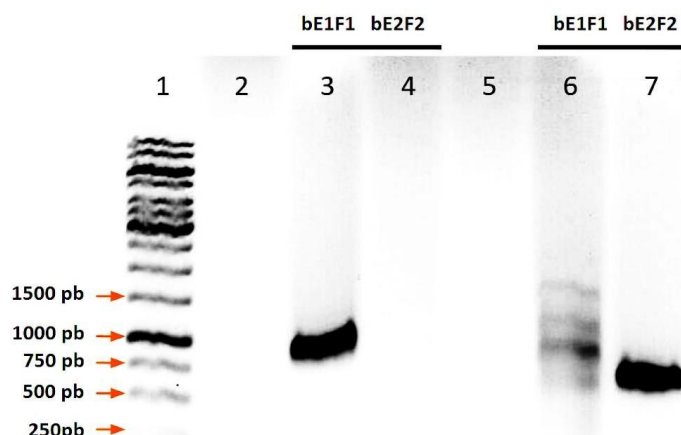


Figura 8 — Teste de discriminação molecular de *mating-types* opostos do isolado SSC04 de *Sporisorium scitamineum*. Produtos de PCR dos *primers* *bE1F1/bE1-2R1* (MAT-1) e *bE2F2/bE1-2R1* (MAT-2) foram colocados em gel de eletroforese (agarose 1%) com *ladder* de 1 kpb (ThermoFisher Scientific) e registrados em transiluminador. Lanes 3 e 4 são referentes ao DNA de um dos haplóides, e as lanes 6 e 7 do haplóide oposto. Lanes 2, 5 e 7 são referentes ao controle negativo.

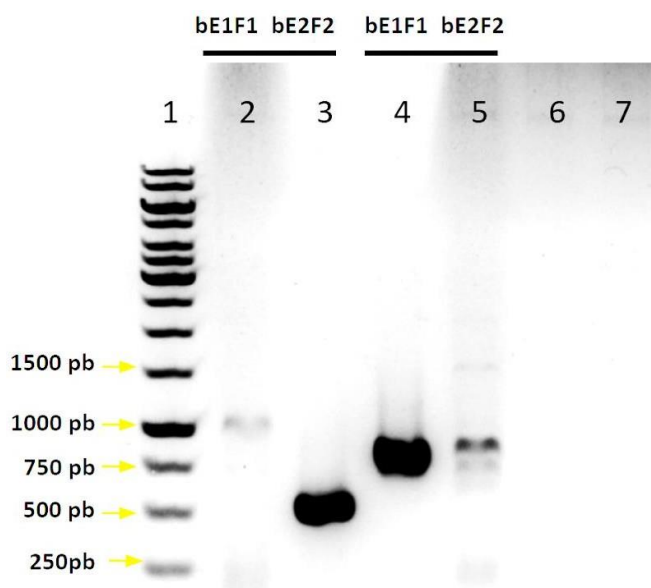


Figura 9 — Teste de discriminação molecular de *mating-types* opostos do isolado SSC39 de *Sporisorium scitamineum*. Produtos de PCR dos *primers* *bE1F1/bE1-2R1* (MAT-1) e *bE2F2/bE1-2R1* (MAT-2) foram colocados em gel de eletroforese (agarose 1%) com *ladder* de 1 kpb (ThermoFisher Scientific) e registrados em transiluminador. Lanes 2 e 3 são referentes ao DNA de um dos haplóides, e as lanes 4 e 5 do haplóide oposto.. Lanes 2, 5 e 7 são referentes ao controle negativo.

Os *primers* utilizados para discriminação molecular do alelo *MAT-1* (*bE1F1/ bE1-2R1*) se anelaram na região N-terminal variável, como observado nas Figuras 10 e 11 e confirmado por sequenciamento Sanger na Figura 12.

bE1 (1392 bp)

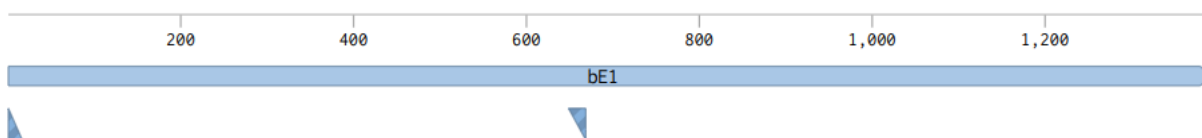


Figura 10 — Comprimento da sequência genômica do alelo *bE1* do gene *bE* no locus de mating-type de *Sporisorium scitamineum*. Os primers *bE1F1* e *bE1-2R1*, indicados pelos triângulos invertidos, amplificam a região N-terminal.



Figura 11 — Região genômica de anelamento dos primers *bE1F1* (5'-ATGCGTGAATTTGCGC-3') e *bE1-2R1* (5'-TTGATGAACCAGAGCGTGAG-3') na região N-terminal do alelo *bE1* do gene *bE*. A região amplificada corresponde a 668 pares de base. Alinhamento de primers e tradução da sequência foi realizada pelo Benchling.

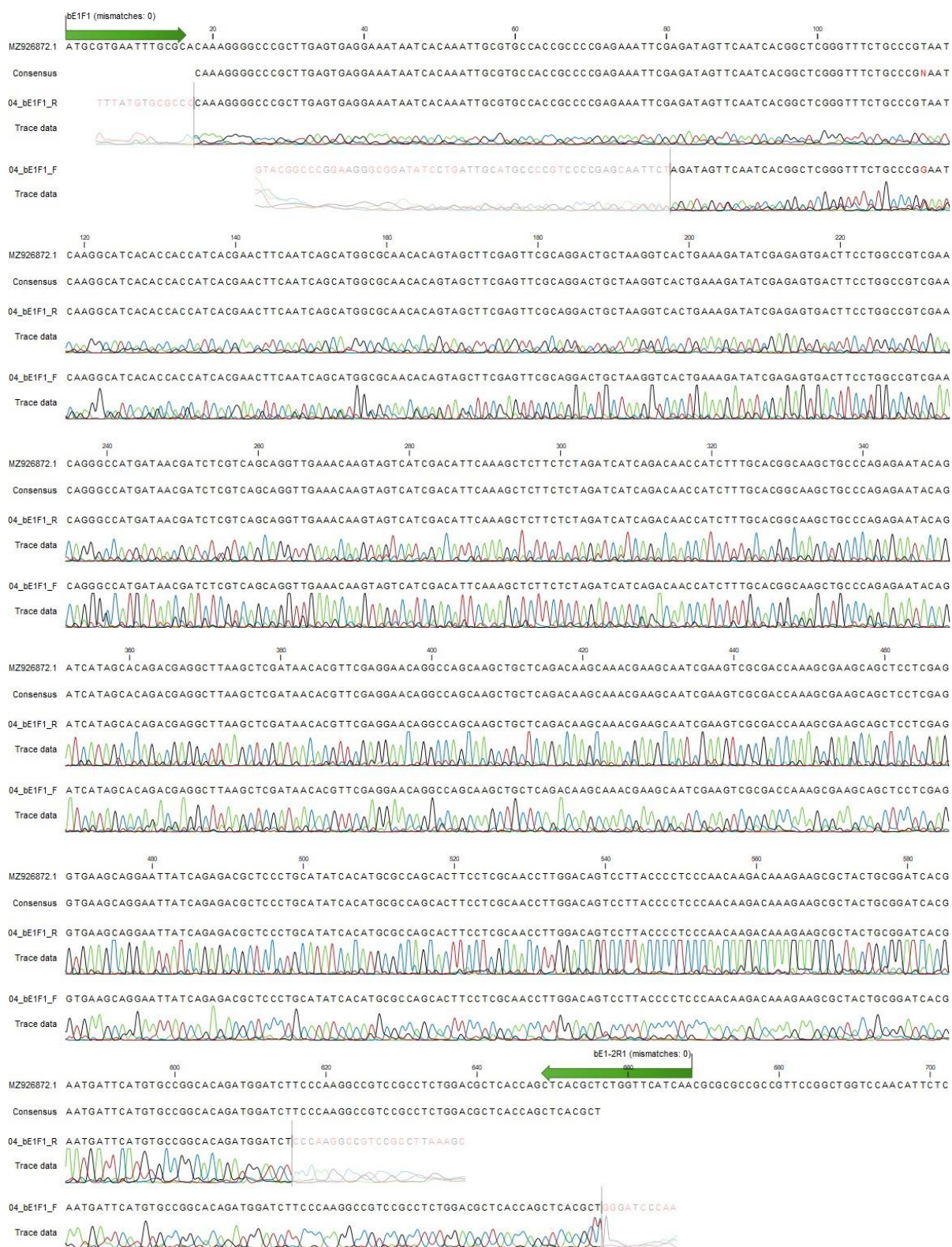


Figura 12 — Sequenciamento pelo método de Sanger (SS) do produto de PCR gerou uma sequência consenso de 639 pb, referente ao par de *primers* de *mating-type* *bE1F1/bE1-2R1* para identificação do alelo *MAT-1* de *S. scitamineum*. Setas em verde são referentes a sequência dos *primers* e sua direção. A primeira linha representa a referência de sequência codante parcial do alelo *bE1* como depositado no *NCBI* pelo acesso MZ926872.1. A segunda linha representa a sequência consenso gerada pela sequência *forward* e *reverse*. Terceira e quarta linha representam, respectivamente, a sequência *reverse* e *forward*. Cromatogramas são referentes à discriminação e qualidade das bases.

Cromatogramas e alinhamento das sequências foram processados pelo CLC Genomics Workbench v. 20 (QIAGEN).

Os *primers* utilizados para discriminação molecular do alelo *MAT-2* (*bE2F2/ bE1-2R1*), assim como no outro alelo, também se anelaram na região N-terminal variável, como observado nas Figuras 13 e 14 e confirmado por sequenciamento Sanger na Figura 15.

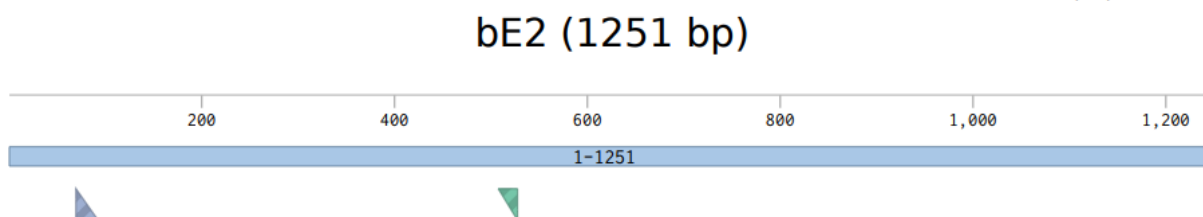


Figura 13 — Comprimento da sequência genômica do alelo *bE2* do gene *bE* no *locus* de *mating-type* de *Sporisorium scitamineum*. Os *primers* *bE2F2* e *bE1-2R1*, indicados pelos triângulos invertidos, amplificam a região N-terminal.



Figura 14 — Região genômica de anelamento dos *primers* *bE2F2* (5'-CTTTCCTACAATCCAAACCAATA-3') e *bE1-2R1* (5'-TTGATGAACCAGAGCGTGAG-3') na região N-terminal do alelo *bE2* do gene *bE*. A região amplificada corresponde a 458 pares de base.

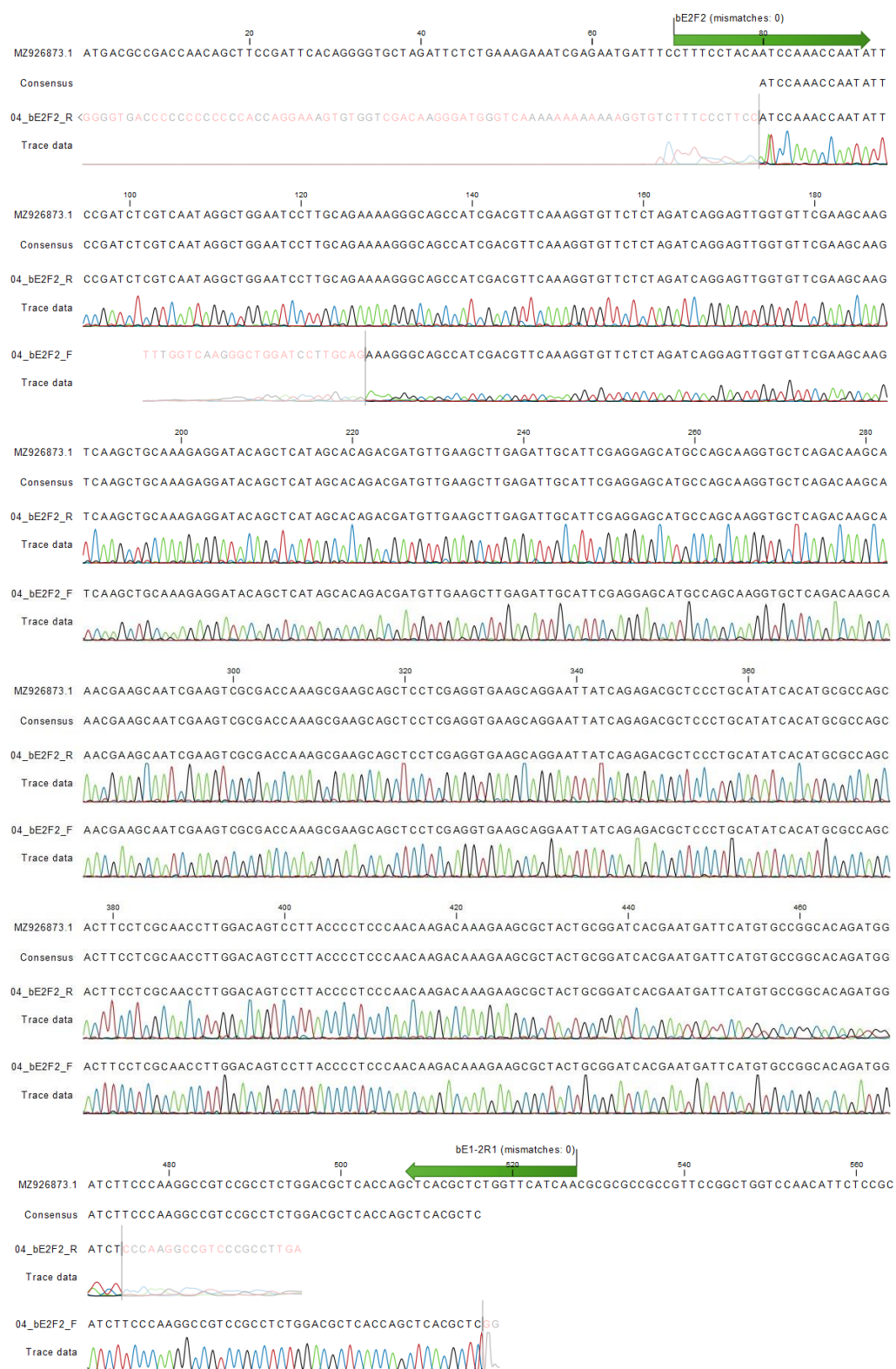


Figura 15 — Sequenciamento pelo método de Sanger (SS) do produto de PCR gerou uma sequência consenso de 437 pb, referente ao par de *primers* de *mating-type* *bE2F2/bE1-2R1* para identificação do alelo *MAT-2* de *S. scitamineum*. Setas em verde são referentes a sequência dos *primers* e sua direção. A primeira linha representa a referência de sequência codante parcial do alelo *bE2* como depositado no *NCBI* pelo acesso MZ926873.1. A segunda linha representa a sequência consenso gerada pela sequência *forward* e *reverse*. Terceira e quarta linha representam, respectivamente, a sequência *reverse* e *forward*. Cromatogramas são referentes à discriminação e qualidade das bases.

Cromatogramas e alinhamento das sequências foram processados pelo CLC Genomics Workbench v. 20 (QIAGEN).

6.4. Identificação de haplótipos

Após os *mating-types* complementares e opostos terem sido identificados corretamente, fez-se necessário a confirmação dos haplótipos utilizados. A confirmação ocorreu a partir da amplificação do gene candidato a efetor *g5159*, visto que o gene carrega um *SNV* na posição 689 e outro na posição 690, como descrito por Silva (2018). Os isolados SSC04 e SSC39 foram agrupados em diferentes haplótipos em pesquisa de Benevenuto et al (2016), além de que estudos anteriores (dados não publicados) identificaram *SNVs* diferentes entre os dois isolados. Os *primers* utilizados para amplificação do gene corretamente identificaram a região como visto na Figura 16.

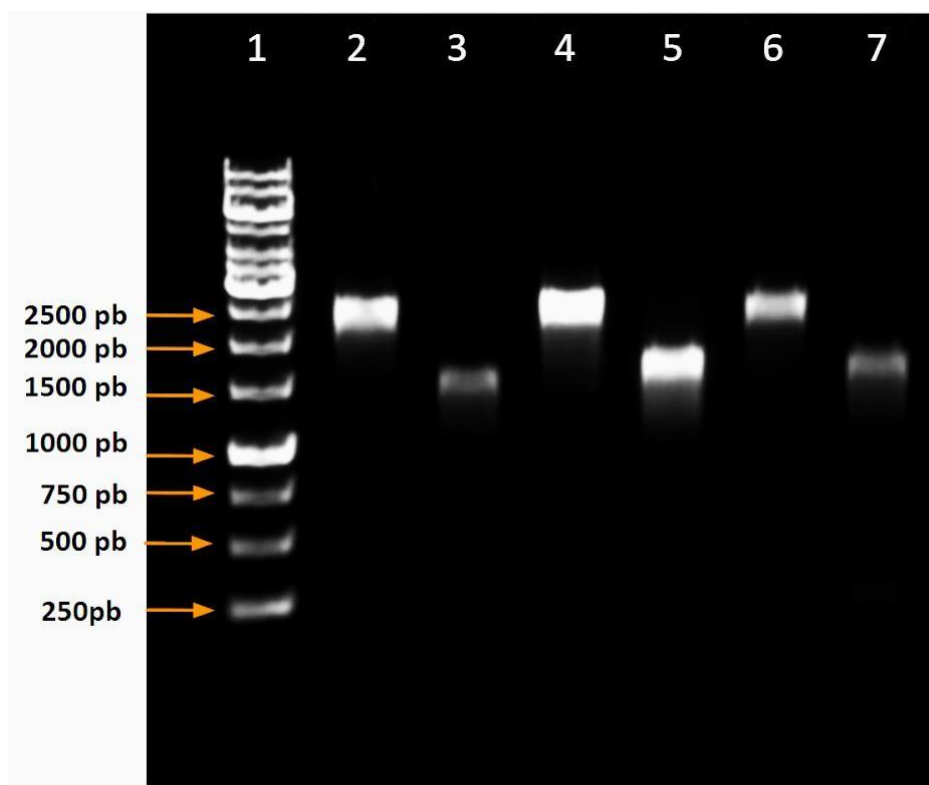


Figura 16 — Amplificação do gene candidato a efetor *g5159* visualizada em gel de agarose (1%). Lane 1 mostra ladder 1 kpb, enquanto as lanes 2,4 e 6 mostram um produto de 2500 pb, resultante da amplificação dos *primers forward* e *reverse* do início ao final do gene. As lanes 3, 5 e 7 mostram um produto de, aproximadamente, 1500 pb, resultante da amplificação a partir do *primer forward* interno ao gene até o seu *reverse*.

Os *primers* utilizados para amplificação anelaram-se alguns pares de bases antes do início do gene e alguns depois de seu fim, enquanto outro *primer* possuía a região interna do gene como alvo (Figura 17). Esta estratégia foi utilizada para otimizar o sequenciamento Sanger na obtenção de todo comprimento da sequência.

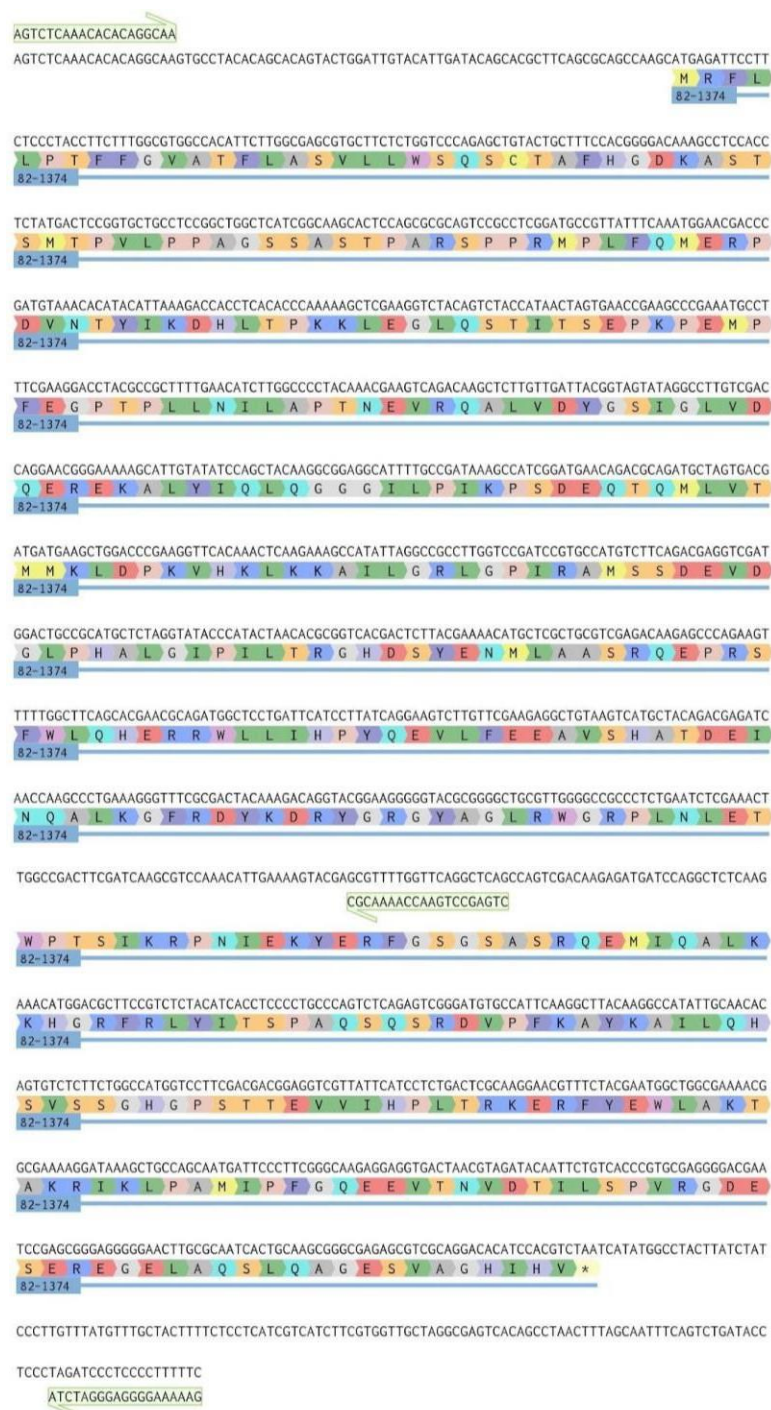


Figura 17 — Região genômica de anelamento dos *primers* do *g519 forward* (5'-GAAAAAGGGGAGGGATCTA-3'), *forward* interno (5'-CTGAGCCTGAACCAAAACGC-3') e *reverse* (5'-AGTCTCAAACACACAGGCAA-3'). A representação mostra a sequência reversa complementar. Quadrados coloridos representam os aminoácidos que compõem a proteína putativa do *g519*. O asterisco representa o códon de parada da cadeia polipeptídica.

Os polimorfismos no gene *g5159* utilizados para identificação dos haplótipos estão na posição 689 e 690 da sequência codante. O primeiro polimorfismo se dá pela troca de uma base guanina (G) por uma citosina (C), enquanto o segundo polimorfismo se dá pela troca de uma timina (T) por uma citosina (C) (Figura 18). Estas alterações nucleotídicas ocasionam a alteração do códon do aminoácido e, portanto, a troca de um Triptofano (W) por um aminoácido Tirosina (Y).

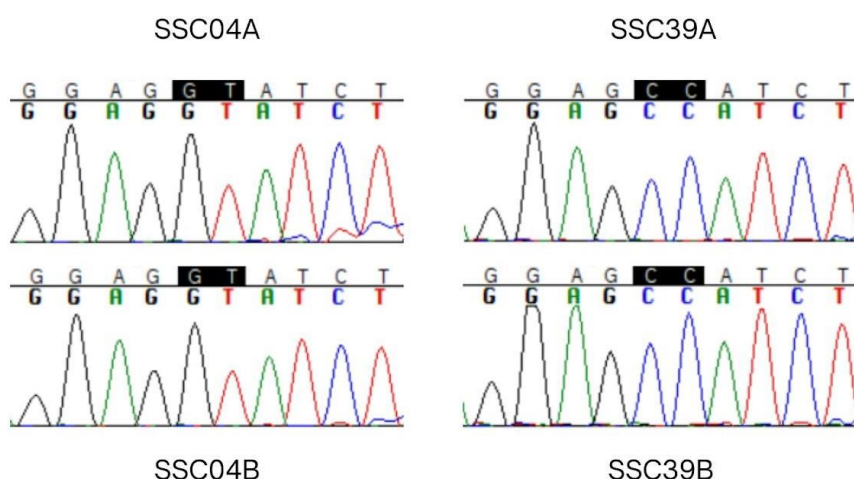


Figura 18 — Polimorfismos nas posições 689 e 690 do gene candidato a efetor *g5159* nos dois isolados. Fonte: Crestana (2015).

6.5. Sequenciamento de Nova Geração (NGS) pela plataforma Oxford Nanopore MinION

6.5.1. Isolado SSC04B (*MAT-1*)

O sequenciamento do isolado SSC04B (*MAT-1*) (APÊNDICE A) pela plataforma *Oxford Nanopore MinION* produziu um total de 5,71 Gb de dados. Foram geradas, aproximadamente, 65.720 *reads*, com um número estimado de 656,93 Mb de bases totais. O N50 estimado foi de 17,38 kb. O comprimento médio das *reads* foi de 10.214,5 pb. 30.120 *reads* tiveram comprimento maior que 10 mil pares de bases (pb), 10.247 tiveram comprimento maior que 20 mil pb, e apenas 91 *reads* ultrapassaram comprimento maior que 50 mil pb (Figura A-1).

A corrida se iniciou com 299 poros ativos na *flow cell* para o sequenciamento. A atividade dos poros decaiu progressivamente ao longo da duração da corrida, chegando a poucos poros ativos por volta de 22 horas e 45

minutos após o início, momento em que o sequenciamento foi encerrado (Figura A-2).

Por conta do decaimento progressivo natural dos nanoporos presentes na *flow cell*, ocorreu decréscimo no número total de *reads* sequenciadas por hora, resultado da diminuição dos poros ativos e operantes. Por volta da vigésima hora da corrida, o número de sequências produzidas já era quase nulo. Foi possível observar que as *reads* com *score* maior ou igual a 7 foram sequenciadas em menor número quando comparadas às *reads* totais (Figura A-3).

Entretanto, apesar da performance reduzida dos poros ao longo da duração da corrida, não foram observados impactos no comprimento médio das *reads* sequenciadas. O comprimento das *reads* se manteve constante e concentrado na faixa dos 10 kpb (Figura A-4).

Além do comprimento médio das *reads*, a qualidade média total esteve concentrada em, aproximadamente, 9 ao longo de toda corrida, enquanto *reads* com um *score* maior ou igual a 7 estiveram concentradas em uma qualidade média de, aproximadamente, 12 (Figura A-5).

O número total de *reads* com um *score* de qualidade maior ou igual a 7 foi de 39.963 do conjunto inicial de 65.720 *reads* (Figura A-6). Dentro deste conjunto de dados, o valor estimado do N50 foi de, aproximadamente, 18,33 kb, enquanto o comprimento médio das *reads* que passaram este filtro foi de 10.821,1 pb. A qualidade das *reads* com $Q \geq 7$ esteve concentrada em, aproximadamente, 11,2. Ainda neste conjunto, 19.391 *reads* tiveram comprimento maior que 10 mil kpb, 6.928 foram maiores que 20 kpb e apenas 69 *reads* ultrapassaram 50 kpb. *Reads* que não passaram no filtro $Q \geq 7$ foram descartadas e não utilizadas nas montagens finais.

6.5.2. Isolado SSC39A (MAT-2)

O sequenciamento do isolado SSC39A (APÊNDICE B) pela plataforma *Oxford Nanopore MinION* produziu um total de 7,52 Gb de dados. Foram geradas, aproximadamente, 76,29 mil *reads*, com um número estimado de 846,19 Mb de bases totais. O N50 estimado foi de 19,53 kb. O comprimento médio das *reads* foi de 11.475,1 pb. 35.917 *reads* tiveram comprimento maior que 10 kpb, 15.226 tiveram comprimento maior que 20 kpb, e 179 *reads* ultrapassaram comprimento maior que 50 kpb (Figura B-1).

A corrida se iniciou com 280 poros ativos na *flow cell* para o sequenciamento. A corrida teve uma duração de 5 horas e 42 minutos, momento em que o sequenciamento foi encerrado (Figura B-2).

Assim como ocorreu durante o sequenciamento do isolado SSC04B, as *reads* com qualidade igual ou maior que 7 foram sequenciadas em menor quantidade em relação ao total de *reads*, como pode ser observado na Figura B3.

Durante a corrida, foi possível observar um aumento do comprimento médio das *reads* sequenciadas (Figura B-4). A corrida se iniciou com o sequenciamento de fragmentos de comprimento médio aproximado de 10 kpb.

O *score* médio de qualidade do total das *reads* sequenciadas do isolado SSC39A manteve-se constante durante toda a corrida próximo ao valor de 8,7 (Figura B5). As *reads* sequenciadas com *score* igual ou maior a 7 possuíram um *score* de qualidade médio maior do que o total das *reads* no experimento, 10,9.

O número total de *reads* com um *score* de qualidade igual ou maior que 7 foi de 46.836 do conjunto inicial de 76.288 *reads* (Figura B-6). Dentro deste conjunto de dados, o valor estimado do N50 foi de, aproximadamente, 20,35 kb. O comprimento médio das *reads* dentro deste conjunto foi de 12.349,1 pb. Das *reads* $Q \geq 7$, 23.789 tiveram comprimento maior que 10 kpb, 10.568 maior que 20 kpb e apenas 139 ultrapassaram os 50 kpb. Assim como no sequenciamento do isolado SSC04, as *reads* que não passaram pelo filtro foram descartadas e não participaram da montagem final do genoma.

6.6. Montagem *de novo* e análises de genômica comparativa

Apenas 26 *contigs* fizeram parte da montagem final dos genomas dos dois haploides, seguindo o estabelecido pela referência SSC39B de Taniguti et al. (2015). Em ambas montagens, o maior *contig* teve um comprimento de >2Mb. O comprimento total do genoma SSC04B foi de 19.978.807 pb, e o genoma SSC39A compreendeu 19.845.973 pb (Tabela 5).

Tabela 5 — Dados referentes às montagens *de novo* dos genomas das células haploides SSC04B e SSC39A de *S. scitamineum*.

Dados da Montagem	SSC04B (MAT-1)	SSC39A (MAT-2)
Contigs	26	26
Maior contig	2.012.307	2.010.856
Comprimento total	19.978.807	19.845.973
GC (%)	55.14	55.07
N50	871.046	859.510

Quando comparada com sua referência SSC04A (dados não publicados), a montagem SSC04B apresentou um conteúdo de, aproximadamente, 14 *mismatches* e 92 *indels* a cada 100 kpb. O total de sequências alinhadas entre a montagem *de novo* e sua referência foi de 19.936.950, valor muito próximo do comprimento total da sequência genômica. Em comparação com sua referência SSC39B (Taniguti et al., 2015), a montagem SSC39A apresentou um conteúdo de, aproximadamente, 5 *mismatches* e 45 *indels* a cada 100 kpb. O total de sequências alinhadas entre a montagem *de novo* e a referência foi de 19.806.782, assim como ocorreu para o haploide do isolado SSC04. Outros dados referentes às análises comparativas com os genomas referências estão descritos na Tabela 6.

Tabela 6 — Dados referentes às montagens *de novo* dos genomas dos haploides SSC04B e SSC39A de *S. scitamineum* quando comparados com suas referências, SSC04A e SSC39B, respectivamente.

Estatísticas Genômicas com Referência	SSC04B (MAT-1)	SSC39A (MAT-2)
Maior alinhamento	2.012.146	2.010.685
Total alinhado	19.936.950	19.806.782
<i>Misassemblies</i>	18	18
Relocações	12	10
Translocações	4	8
Inversões	2	0
<i>Mismatches</i> por 100 kpb	14.83	5.44
<i>Mismatches</i>	2956	1078
<i>Indels</i> por 100 kpb	92.46	45.96
<i>Indels</i> (<= 5 bp)	17480	8969
<i>Indels</i> (> 5 bp)	954	134

Quando comparados individualmente com sua referência, os *contigs* gerados pela montagem SSC04B (MAT-1) apresentaram grande similaridade no que diz respeito ao comprimento das sequências, sendo que pequenas variações no

tamanho foram encontradas, principalmente, pela complexidade na montagem de sequências teloméricas. A única discrepância foi encontrada no *contig* 23 correspondente ao cromossomo 23 da referência SSC04A (MAT-2) onde encontram-se o *cluster* de genes rDNA (TANIGUTI et al., 2015) (Tabela 7).

Tabela 7 — Comprimento dos *contigs* gerados pela montagem *de novo* do genoma do haplóide SSC04B de *S. scitamineum* após três rodadas de polimento em comparação com a referência SSC04A (dados não publicados).

SSC04A (referência)			Montagem SSC04B		
ID	Cromossomo	Tamanho (pb)	ID	Cromossomo	Tamanho (pb)
Sag_01_pilon	1	2.010.843	Chr1	1	2.012.307
Sag_02_pilon	2	1.682.132	Chr2	2	1.677.552
Sag_03_pilon	3	1.571.048	Chr3	3	1.572.524
Sag_04_pilon	4	1.271.557	Chr4	4	1.272.781
Sag_05_pilon	5	1.054.858	Chr5	5	1.040.170
Sag_06_pilon	6	1.026.702	Chr6	6	1.027.827
Sag_07_pilon	7	944.452	Chr7	7	922.846
Sag_08_pilon	8	827.561	Chr8	8	828.504
Sag_09_pilon	9	814.175	Chr9	9	815.192
Sag_10_pilon	10	884.414	Chr10	10	871.046
Sag_11_pilon	11	772.128	Chr11	11	779.534
Sag_12_pilon	12	714.214	Chr12	12	715.037
Sag_13_pilon	13	686.470	Chr13	13	668.952
Sag_14_pilon	14	657.027	Chr14	14	657.876
Sag_15_pilon	15	609.286	Chr15	15	611.033
Sag_16_pilon	16	652.924	Chr16	16	653.618
Sag_17_pilon	17	591.589	Chr17	17	592.400
Sag_18_pilon	18	612.409	Chr18	18	613.002
Sag_19_pilon	19	583.410	Chr19	19	584.117
Sag_20_pilon	20	553.972	Chr20	20	554.576
Sag_21_pilon	21	556.426	Chr21	21	557.130
Sag_22_pilon	22	465.144	Chr22	22	466.966
Sag_23_pilon	23	141.908	Chr23	23	207.241
Sag_24_pilon	24	73.322	Chr24	24	73.644
Sag_25_pilon	25	94.517	Chr25	25	94.904
Sag_26_pilon	26	107.681	Chr26	26	108.028

A representação gráfica dos comprimentos dos *contigs* em relação aos cromossomos de sua referência pode ser observada na Figura 19.

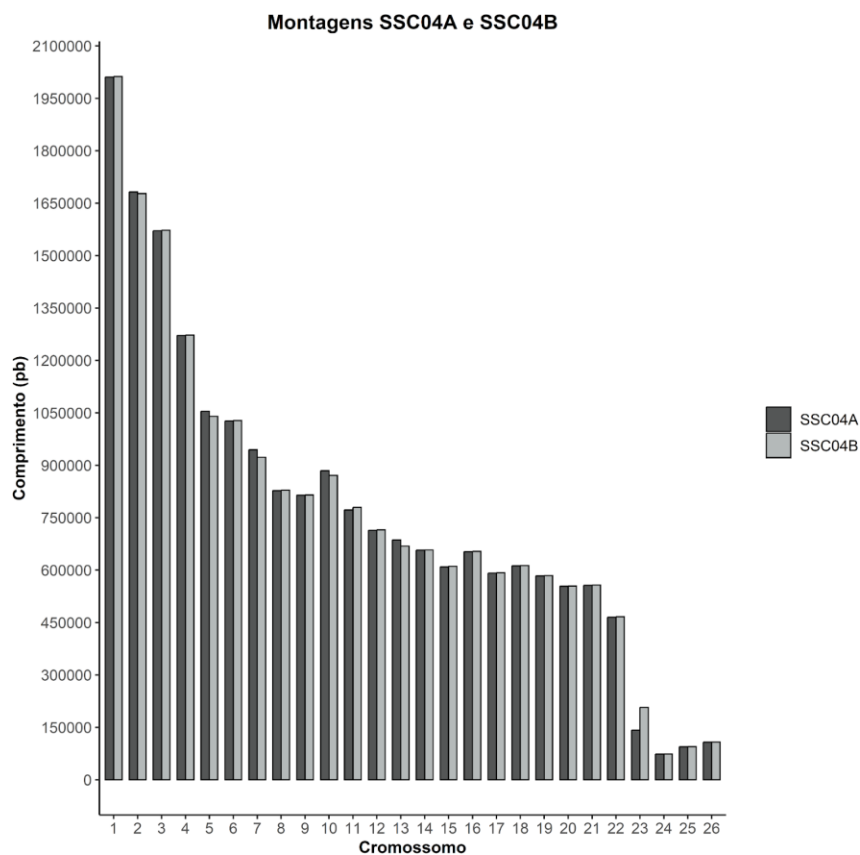


Figura 19 — Gráfico de barras demonstrando o comprimento da sequência genômica de cada *contig* correspondente aos cromossomos da referência. Em cinza escuro, o tamanho dos cromossomos de SSC04A e, em cinza claro, os contigs da montagem *de novo* de SSC04B.

Da mesma forma que a montagem anterior, a montagem SS39A foi comparada com sua referência SSC39B (Taniguti et al., 2015) e demonstrou a alta similaridade existente entre os comprimentos dos *contigs* da montagem *de novo* e a referência (Tabela 8). Apesar de uma diferença de, aproximadamente, 32 kpb no *contig* 23 dos dois genomas, ela foi menor que a diferença de, aproximadamente, 65 kpb neste mesmo *contig* entre os genomas do isolado SSC04.

Tabela 8 — Número de bases dos *contigs* gerados pela montagem *de novo* do genoma do haplóide SSC39A de *S. scitamineum* após três rodadas de polimento em comparação com a referência SSC39B (Taniguti et al., 2015).

Montagem SSC39A			SSC39B (referência)		
ID	Cromossomo	Tamanho (pb)	ID	Cromossomo	Tamanho (pb)
Chr1	1	2.010.856	CP010913.1	1	2.009.762
Chr2	2	1.671.832	CP010914.1	2	1.671.606
Chr3	3	1.568.132	CP010915.1	3	1.566.833
Chr4	4	1.272.646	CP010916.1	4	1.271.980
Chr5	5	1.053.466	CP010917.1	5	1.055.339
Chr6	6	1.022.517	CP010918.1	6	1.032.954
Chr7	7	939.058	CP010919.1	7	938.453
Chr8	8	824.091	CP010920.1	8	837.966
Chr9	9	802.280	CP010921.1	9	808.176
Chr10	10	859.510	CP010922.1	10	875.830
Chr11	11	764.289	CP010923.1	11	790.822
Chr12	12	705.453	CP010924.1	12	716.888
Chr13	13	681.909	CP010925.1	13	681.008
Chr14	14	658.775	CP010926.1	14	652.314
Chr15	15	619.076	CP010927.1	15	634.981
Chr16	16	654.104	CP010928.1	16	653.278
Chr17	17	589.514	CP010929.1	17	605.135
Chr18	18	594.350	CP010930.1	18	598.089
Chr19	19	574.890	CP010931.1	19	580.969
Chr20	20	544.644	CP010932.1	20	573.042
Chr21	21	526.395	CP010933.1	21	539.152
Chr22	22	464.207	CP010934.1	22	475.902
Chr23	23	118.345	CP010935.1	23	140.422
Chr24	24	63.509	CP010936.1	24	87.293
Chr25	25	83.824	CP010937.1	25	82.225
Chr26	26	89.722	CP010938.1	26	99.152

A representação gráfica dos comprimentos dos *contigs* em relação aos cromossomos de sua referência pode ser observada na Figura 20.

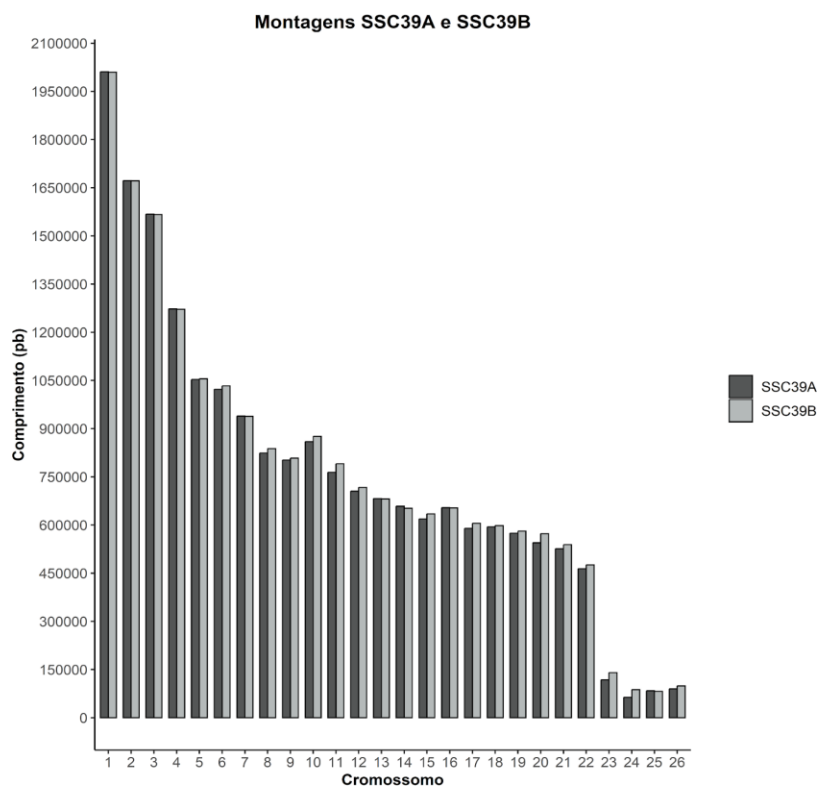


Figura 20 — Gráfico de barras demonstrando o comprimento da sequência genômica de cada *contig* correspondente aos cromossomos da referência. Em cinza escuro, os *contigs* da montagem *de novo* de SSC39A e, em cinza claro, a referência SSC39B (Taniguti et al., 2015).

Apesar da alta similaridade entre o número de bases dos *contigs* das montagens *de novo* e suas referências, foi necessário realizar uma comparação global entre as sequências genômicas para identificação de regiões sintênicas. Desta forma, foi gerado um gráfico de pontos (ou *dot plot*) bidimensional em que cada eixo representa uma sequência e um ponto é desenhado a partir de um *score* de similaridade calculado. Entre as montagens *de novo* e suas referências, não foram encontradas grandes diferenças macro sintênicas, como aquelas relacionadas à rearranjos cromossômicos extensos (Figuras 21 e 22).

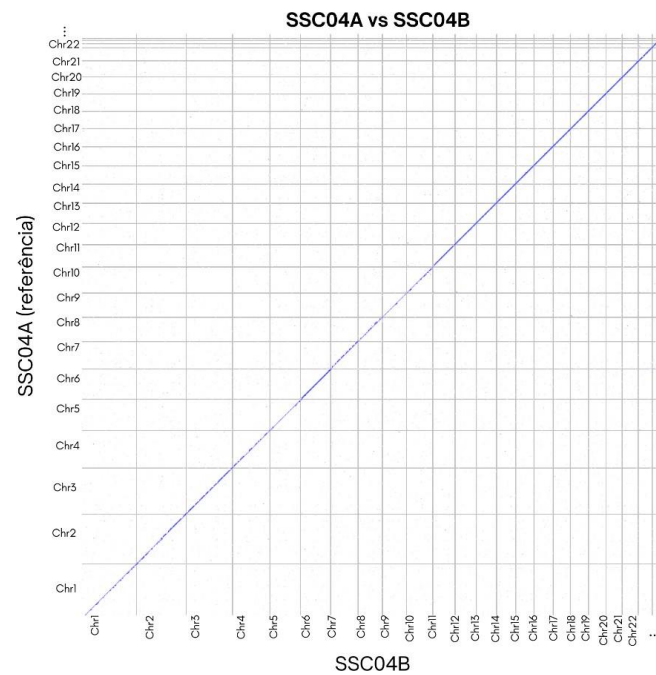


Figura 21 — Dot plot das sequências genômicas de *mating-types* complementares e opostos do isolado SSC04 de *S. scitamineum*. O gráfico ilustra o alinhamento de cada *contig* da montagem *de novo* em relação à referência e demonstra sintenia conservada expressa pela linha diagonal contínua.

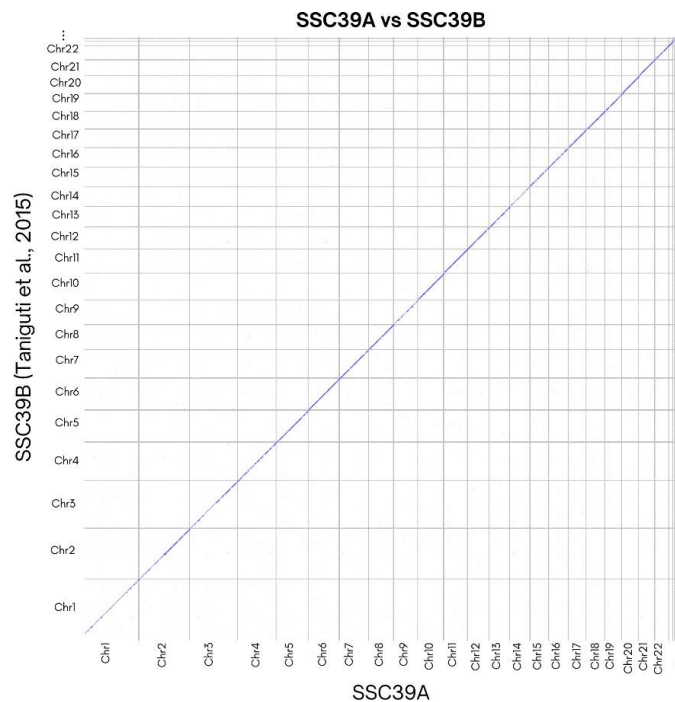


Figura 22 — Dot plot das sequências genômicas de *mating-types* complementares e opostos do isolado SSC39 de *S. scitamineum*. O gráfico ilustra o alinhamento de cada *contig* da montagem *de novo* em relação à referência e demonstra sintenia conservada expressa pela linha diagonal contínua.

A avaliação da completude das montagens *de novo* foi realizada utilizando a ferramenta BUSCO para investigar a presença de ortólogos altamente conservados entre os basidiomicetos. A sequência referência completa telômero a telômero do genoma do isolado SSC39B (*MAT-1*) publicado em 2015 por Taniguti e colaboradores apresentou 94.3% de completude. O genoma de maior completude foi a sequência do isolado SSC04A, apresentando 98.1% ortólogos completos e de cópias únicas. As montagens *de novo* referentes aos genomas SSC04B e SSC39B ambas apresentaram completude maior do que a referência de Taniguti e colabores, como observado na Figura 23.

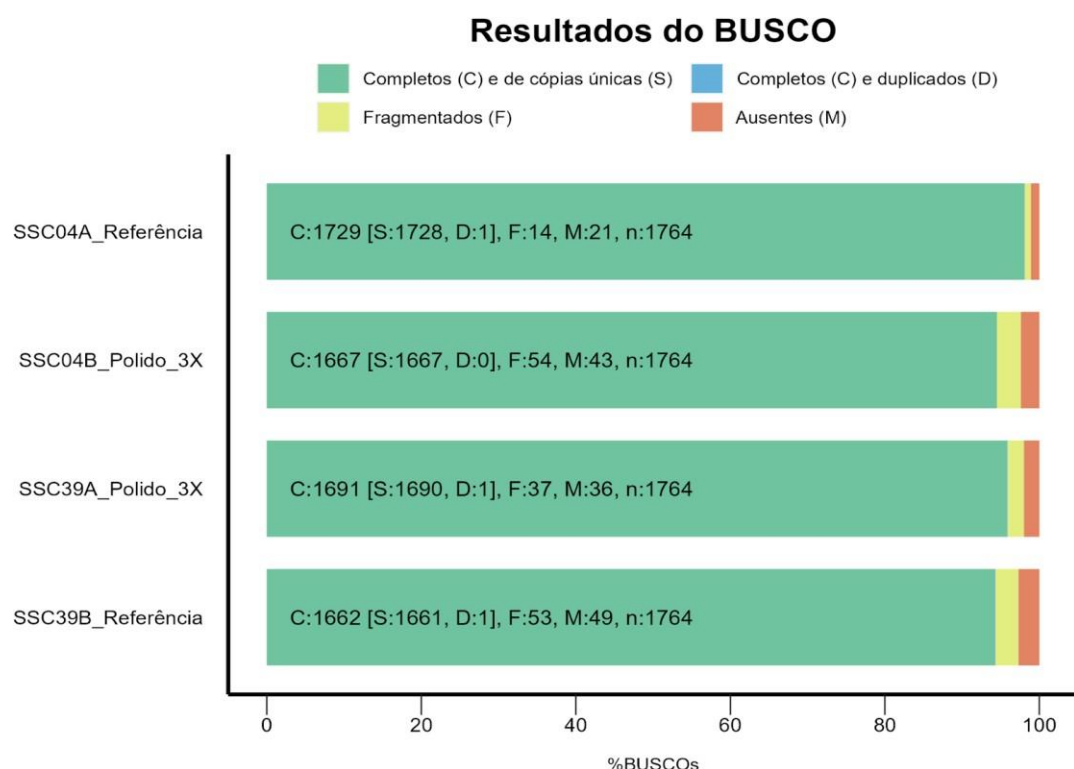


Figura 23 — Gráfico comparando a completude genômica das diferentes montagens *de novo* com suas referências.

A predição gênica realizada pelo *FunGAP* foi realizada, de forma independente, para cada um dos genomas. Para o genoma do isolado SSC04 *MAT-1*, foram preditas 6.988 sequências codantes que informam a produção de proteínas. O número de bases médio das *CDS* (*coding sequence*) foi de 1.728,9 pb, enquanto a densidade gênica foi de 349,77 genes por megabase (Figura 24). 37,59% dos genes foram preditos com *splicing*. O número médio de íntrons por gene, assim como o número médio de éxons por gene foi 1. O número de bases médio de

transcritos predito foi de 1.830,6 pb. O número total de íntrons preditos para este genoma foi de 4.694 e o total de éxons foi de 11.682.

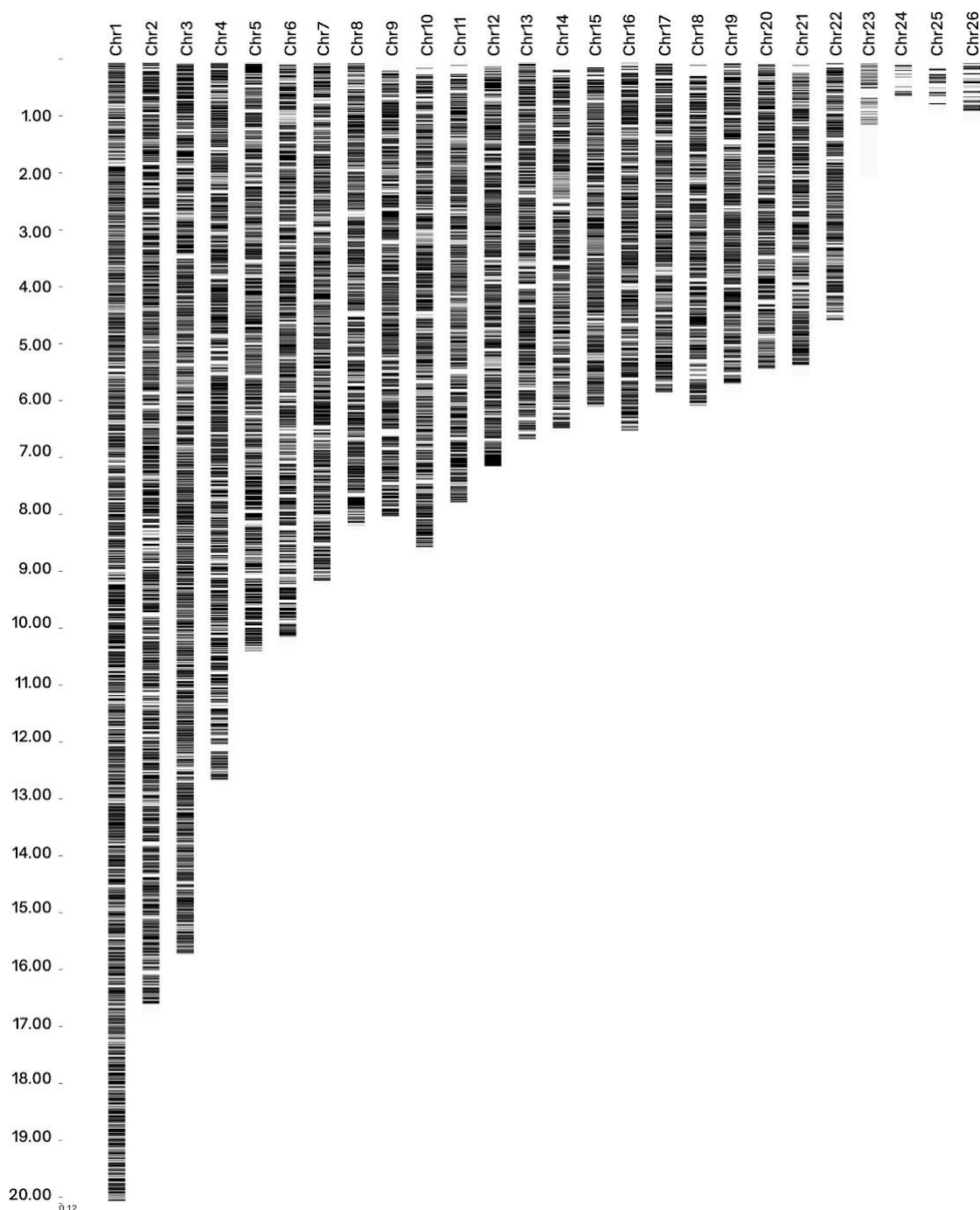


Figura 24 — Densidade de seqüências codantes distribuídas ao longo do genoma SSC04 *MAT-1*. Linhas pretas representam seqüências codantes. A escala utilizada é de 10^5 .

Para o genoma do isolado SSC04 *MAT-2*, foram preditas 6.933 seqüências codantes que informam a produção de proteínas. O comprimento médio das *CDS* (*coding sequence*) foi de 1.750,6 pb, e a densidade gênica foi de 347,34

genes por megabase (Figura 25). 34,57% dos genes foram preditos com *splicing*. O número médio de íntrons por gene, assim como o número médio de éxons por gene foi 1. O comprimento médio de transcritos predito foi de 1.846,6 pb. O número total de íntrons preditos para este genoma foi de 4.284 e o total de éxons foi de 11.217.

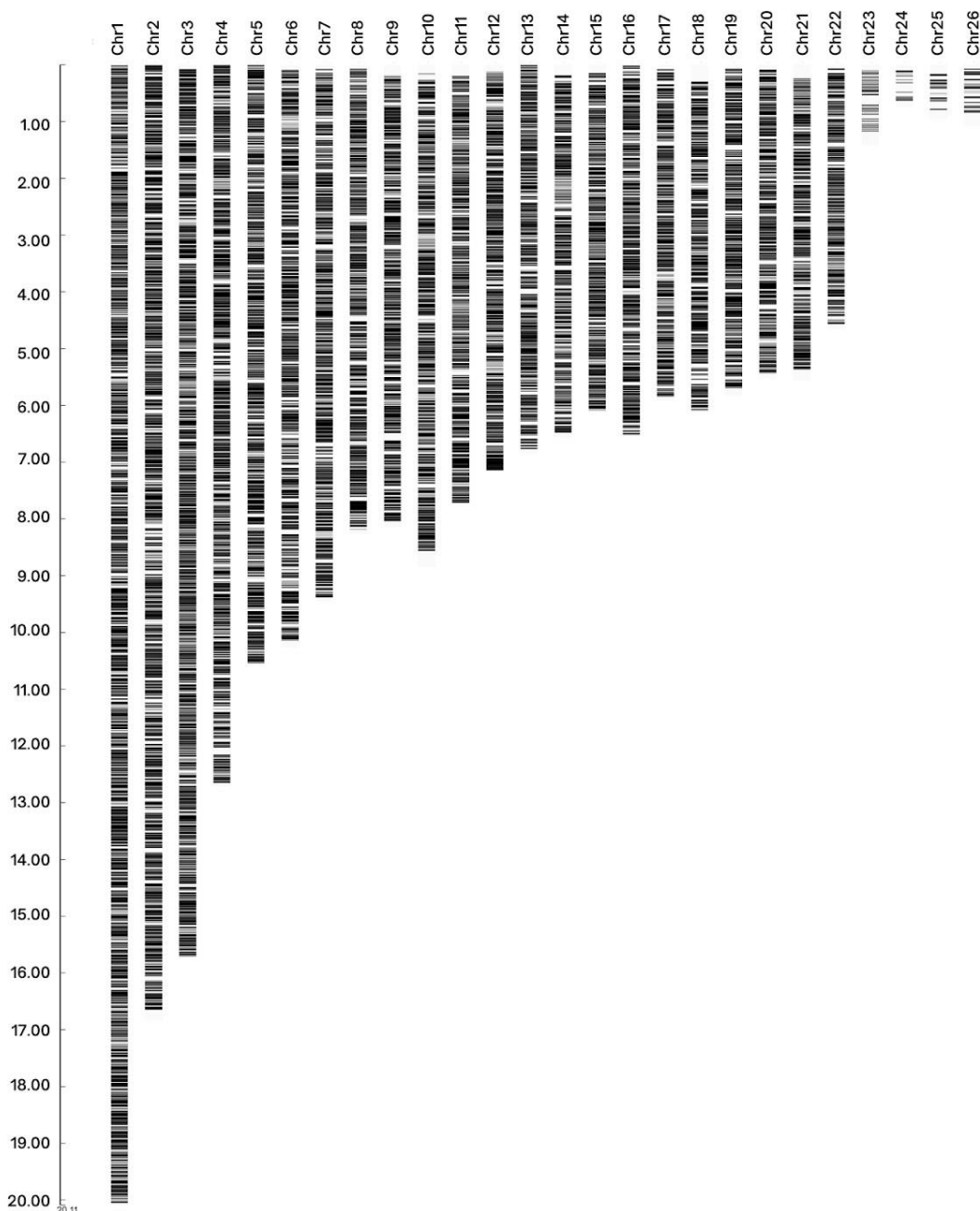


Figura 25 — Densidade de seqüências codantes distribuídas ao longo do genoma SSC04 MAT-2. Linhas pretas representam seqüências codantes. A escala utilizada é de 10^5 .

Para o genoma do isolado SSC39 MAT-1, foram preditas 6.896 seqüências codantes que informam a produção de proteínas. O comprimento médio

das CDS (*coding sequence*) foi de 1.754,2 pb, e a densidade gênica foi de 343,64 genes por megabase (Figura 26). 42,2 % dos genes foram preditos com *splicing*. O número médio de íntrons por gene, assim como o número médio de éxons por gene foi 1. O número de bases médio de transcritos predito foi de 1.869,1 pb. O número total de íntrons preditos para este genoma foi de 5.196 e o total de éxons foi de 12.092.

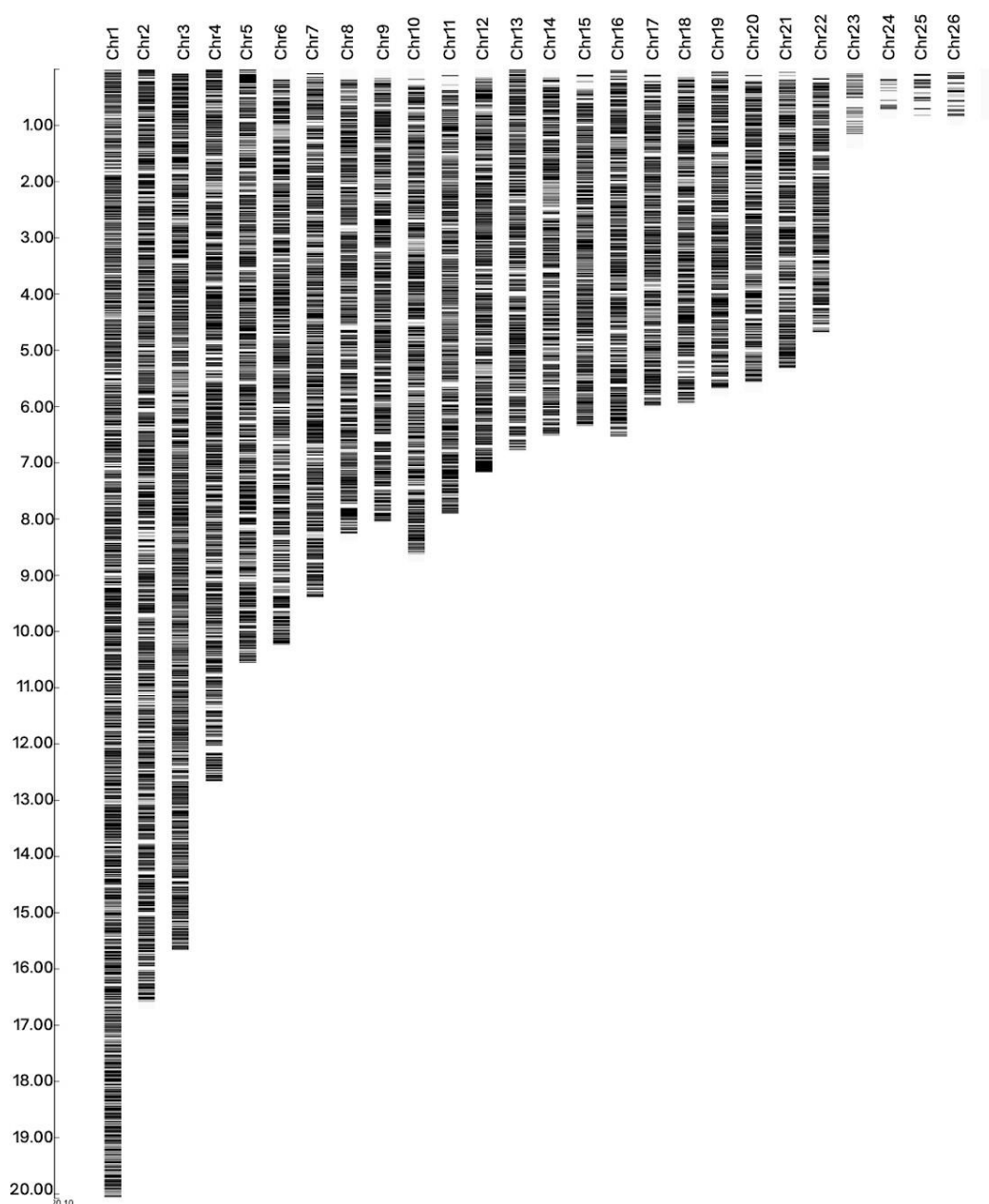


Figura 26 — Densidade de sequências codantes distribuídas ao longo do genoma SSC39 *MAT-1*. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5 .

Para o genoma do isolado SSC39 *MAT-2*, foram preditas 6.956 sequências codantes que informam a produção de proteínas. O comprimento médio das *CDS* (*coding sequence*) foi de 1.749,4 pb, a densidade gênica foi de 352,07 genes por megabase (Figura 27). 36,44 % dos genes foram preditos com *splicing*. O número médio de íntrons por gene, assim como o número médio de éxons por gene foi 1. O número de bases médio de transcritos predito foi de 1.847,2 pb. O número total de íntrons preditos para este genoma foi de 4.444 e o total de éxons foi de 11.400.

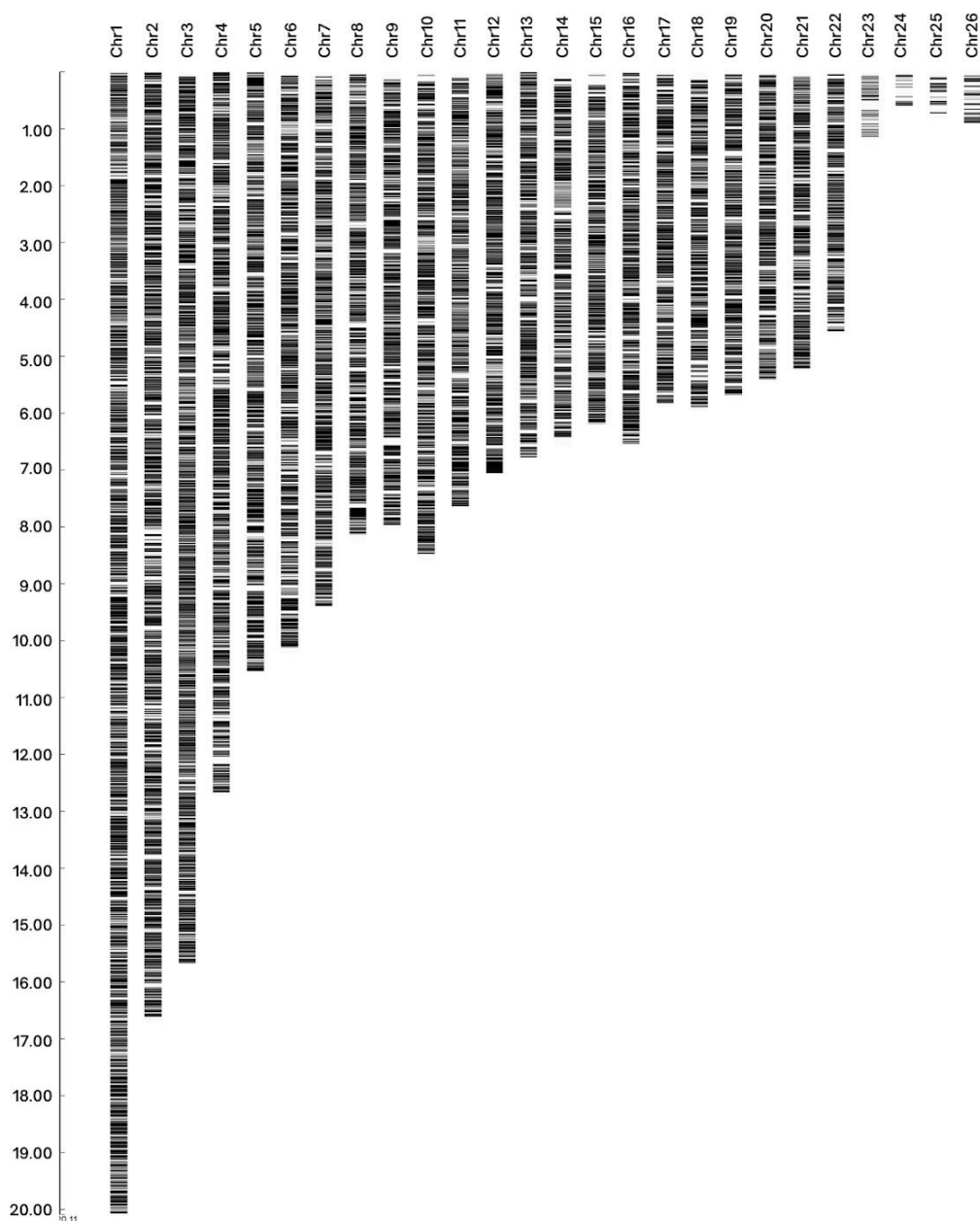


Figura 27 — Densidade de sequências codantes distribuídas ao longo do genoma SSC39 *MAT-2*. Linhas pretas representam sequências codantes. A escala utilizada é de 10^5 .

A montagem do genoma SSC04B (ou *MAT-1*) apresentou uma cobertura de 19,61X, e a montagem do genoma SSC39A (ou *MAT-2*) teve, aproximadamente, 25X de cobertura (Tabela 9). Dados gerais comparativos sobre os genomas podem ser visualizados na Tabela 9.

Tabela 9 — Panorama geral comparativo entre os genomas de dois isolados diferentes de *S. scitamineum* com seus *mating-types* complementares.

Característica	SSC04A	SSC04B	SSC39A	SSC39B
Plataforma	ONT	ONT	ONT	PacBio e Illumina
Software de Montagem	Canu	Canu	Flye	HGAP.2 e CLC Workbench
Tamanho do Genoma (Mb)	20,42	19,97	19,84	19,97
Cobertura	195,59X	19,61X	~25X	>100X
Número de <i>contigs</i>	26	26	26	26
Maior cromossomo (bp)	2.010.843	2.012.146	2.010.685	2.009.762
Menor cromossomo (bp)	73.322	73.471	63.409	82.225
Conteúdo GC (%)	55,5	55,14	55,07	55
BUSCO (%)	98,1	94,5	95,9	94,3
Número de genes preditos	6.933	6.988	6.956	6.896

6.7. Análise de variantes estruturais

Apesar de uma extensa colinearidade entre as sequências genômicas haplóides de dois isolados distintos de *S. scitamineum*, quando comparados *mating-types* opostos e complementares de mesmo isolado, foram encontradas inversões, repetições e lacunas no *locus* de *mating* (Figura 28) que, segundo Taniguti et al. (2015) está localizado entre as regiões 792,295 bp a 863,606 bp no cromossomo 2. Além disso, quando comparados os genomas de *mating-types* iguais, os genomas haplóides SSC04A e SSC39A referentes ao *MAT-2*, demonstraram não ser idênticos por apresentarem uma inversão no *locus* de *mating* no cromossomo 2. Os genomas haplóides SSC04B e SSC39B referentes ao *MAT-1*, apresentaram sintenia completa (Figura 29).

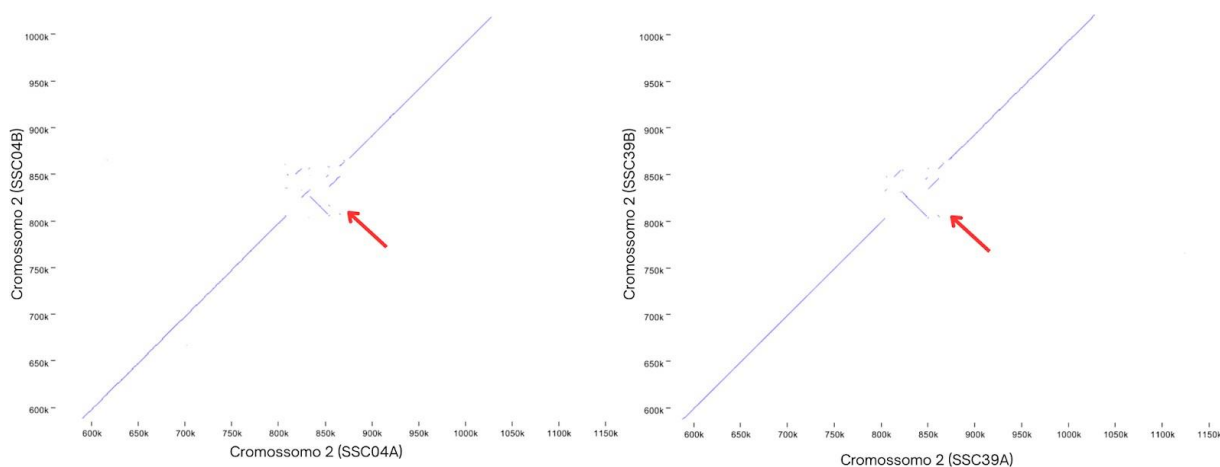


Figura 28 — Dot plot das sequências genômicas de *mating-types* complementares e opostos dos isolados SSC04 (esquerda) e SSC39 (direita) de *S. scitamineum*. Dot plot gerado pelo CLC Genomics Workbench (QIAGEN). O gráfico ilustra o alinhamento do cromossomo 2 nos dois isolados, demonstrando a presença de inversões que são apresentadas como linhas invertidas, repetições, expressas pelas linhas adjacentes e lacunas.

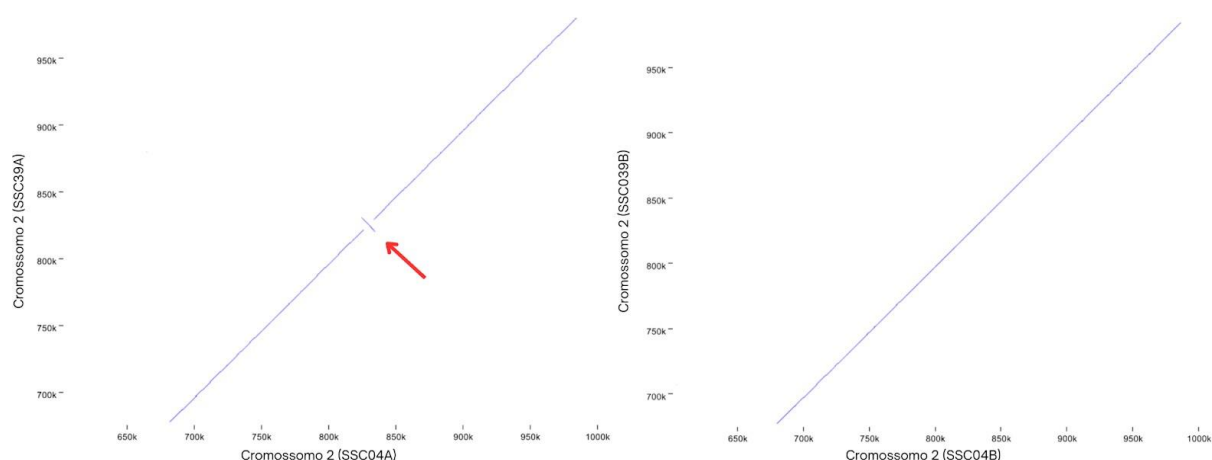


Figura 29 — Dot plot das sequências genômicas de *mating-types* iguais dos isolados SSC04A e SSC39A (esquerda) e SSC04B e SSC39B (direita) de *S. scitamineum*. Dot plot gerado pelo CLC Genomics Workbench (QIAGEN). O gráfico ilustra o alinhamento do cromossomo 2 nos dois isolados. Entre os genomas de alelo MAT-2 (representado pelo “A”), há presença de uma inversão, expressada pela linha invertida, e nos genomas de alelo MAT-1, há sintonia contínua.

Além das sequências genômicas, as sequências dos genes preditos foram avaliadas quanto ao grau de colinearidade entre os genomas dos isolados SSC04A (MAT-2) e SSC04B (MAT-1). Não foi possível observar nenhuma variante estrutural a nível global em nenhum dos cromossomos (Figura 30), a não ser pelas variações estruturais locais presentes no cromossomo 2.

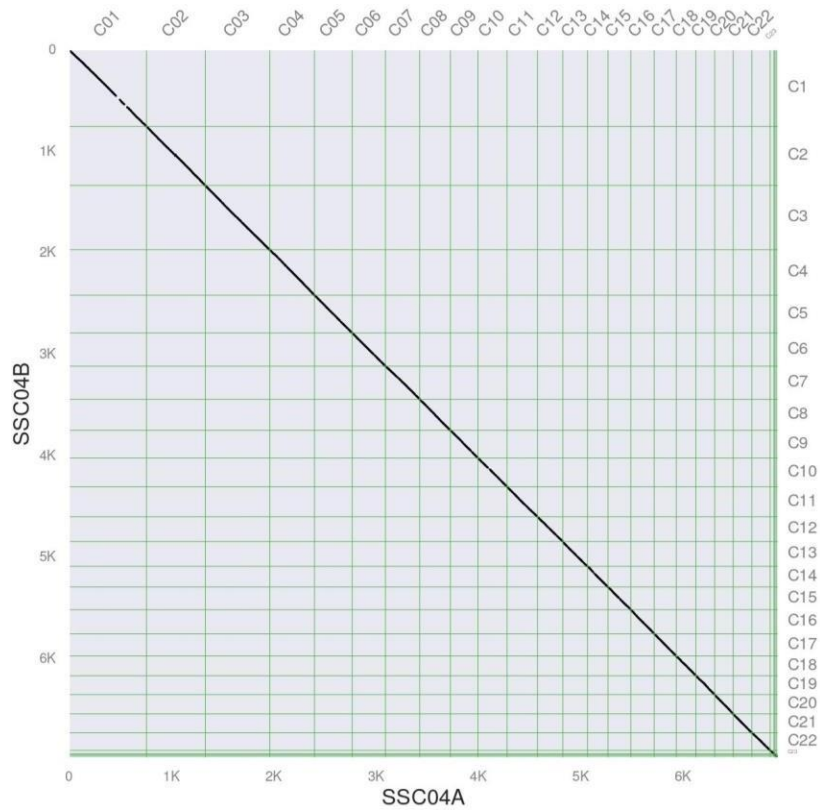


Figura 30 — Dot plot das sequências gênicas de *mating-types* complementares e opostos do isolado SSC04 de *S. scitamineum*, gerados pelo MCScanX.

A profundidade sintênica que, segundo Tang et al. (2015), “refere-se ao número de regiões sintênicas identificadas em um genoma alvo para uma determinada posição de consulta”, foi analisada comparando as sequências gênicas pelo MCScanX. Entre os genomas do isolado SSC04, a profundidade sintênica representou 97% de todo genoma (Figura 31).



Figura 31 — Profundidade sintênica entre os genomas SSC04A (MAT-2) e SSC04B (MAT-1) gerada pelo MCScanX.

As sequências dos genes preditos entre os genomas SSC39A (*MAT-1*) e SSC39B (*MAT-2*) também foram avaliados quanto a colinearidade e seguiram o padrão apresentado pelo isolado mais agressivo, visto que não foram encontradas variantes estruturais globais, com exceção de variantes locais no cromossomo 2 no *locus* de *mating-type* (Figura 32).

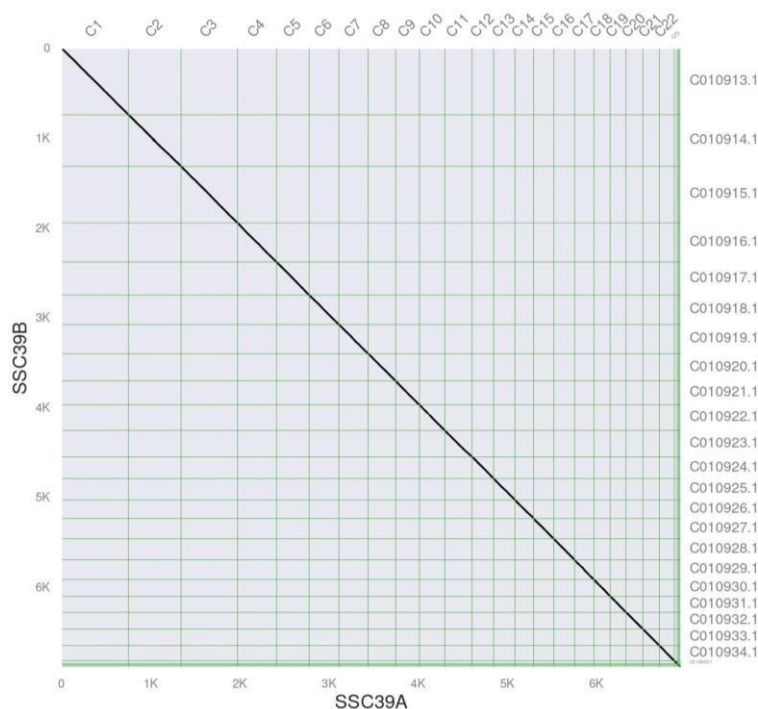


Figura 32 — Dot plot das sequências gênicas de *mating-types* complementares e opostos do isolado SSC39 de *S. scitamineum*, gerados pelo MCScanX.

Entre os genomas do isolado SSC39, a profundidade sintênica representou 98% do genoma, frente a 97% do isolado mais agressivo (Figura 33).

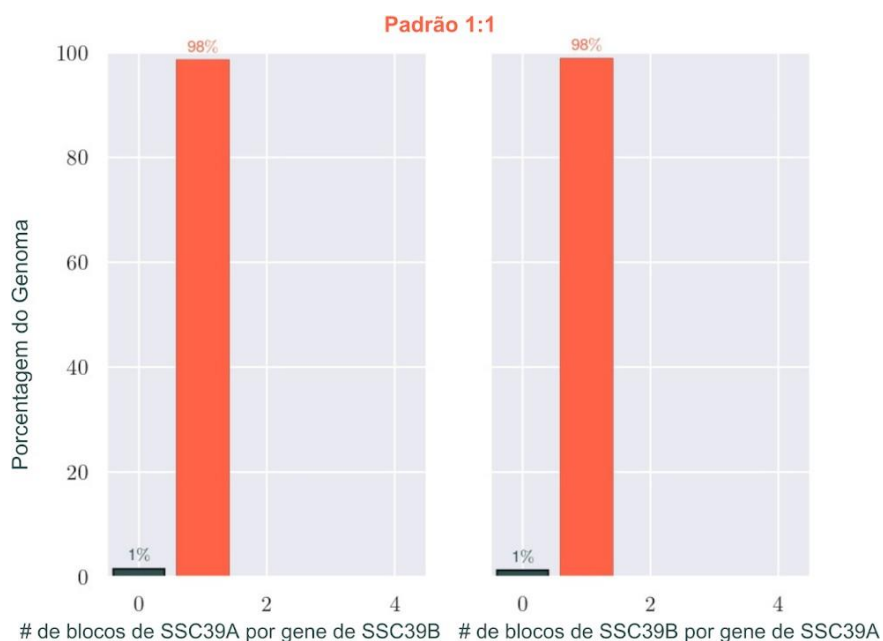


Figura 33 — Profundidade sintênica entre os genomas SSC04A (*MAT-2*) e SSC04B (*MAT-1*) gerada pelo MCScanX.

As sequências entre os genomas de *mating-types* opostos que sofreram variações estruturais estão localizadas entre os *loci* de *mating-type* (Figura 34). Os genes *c1d1* (*g969*), *bE* (*g970*) e *bW* (*g971*), pertencentes ao *locus b* de *mating-type*, não sofreram nenhum tipo de alteração em nenhum dos quatro genomas. Entretanto, o gene *nat-1* (*g972*), responsável por codificar uma acetiltransferase, foi invertido entre os alelos *MAT-1* e *MAT-2*, em ambos isolados. Além disso, um conjunto de genes foi invertido junto com o segmento, dentre eles, pode-se citar um gene (*g973*) contendo um domínio de transporte de íons (IPR027359), outro gene (*g974*) com um domínio *Pal1* (IPR009571), um gene (*g976*) cujo produto é uma proteína de controle de montagem do citoesqueleto (UniProt: A0A0D1E8M) e, por fim, uma subunidade 26S reguladora do proteassomo (*g977*) (UniProt: A0A0D1CGU5). Além deste segmento invertido entre os *mating-types*, dois genes foram realocados, sendo um deles, um provável regulador de replicação do DNA (*g986*) (IPR021110, IPR040203). Entre as sequências gênicas presentes entre os *loci* de *mating-type*, é possível observar a presença de sequências relacionadas a transposons, flanqueando as regiões que sofreram alterações estruturais.

Apesar dos mesmos padrões de alterações terem sido observados nos *mating-types* opostos de ambos isolados, o isolado SSC39 possui uma inversão não presente no isolado SSC04 e, por esta razão, a sequência referente a esta região dos genomas SSC04 (*MAT-2*) e SSC39 (*MAT-2*) não é igual, como demonstrado

anteriormente pela Figura 41. A sequência invertida possui um gene predito como possível caseína quinase *hhp1* (*g979*) (Uniprot: A0A0D1EA76).

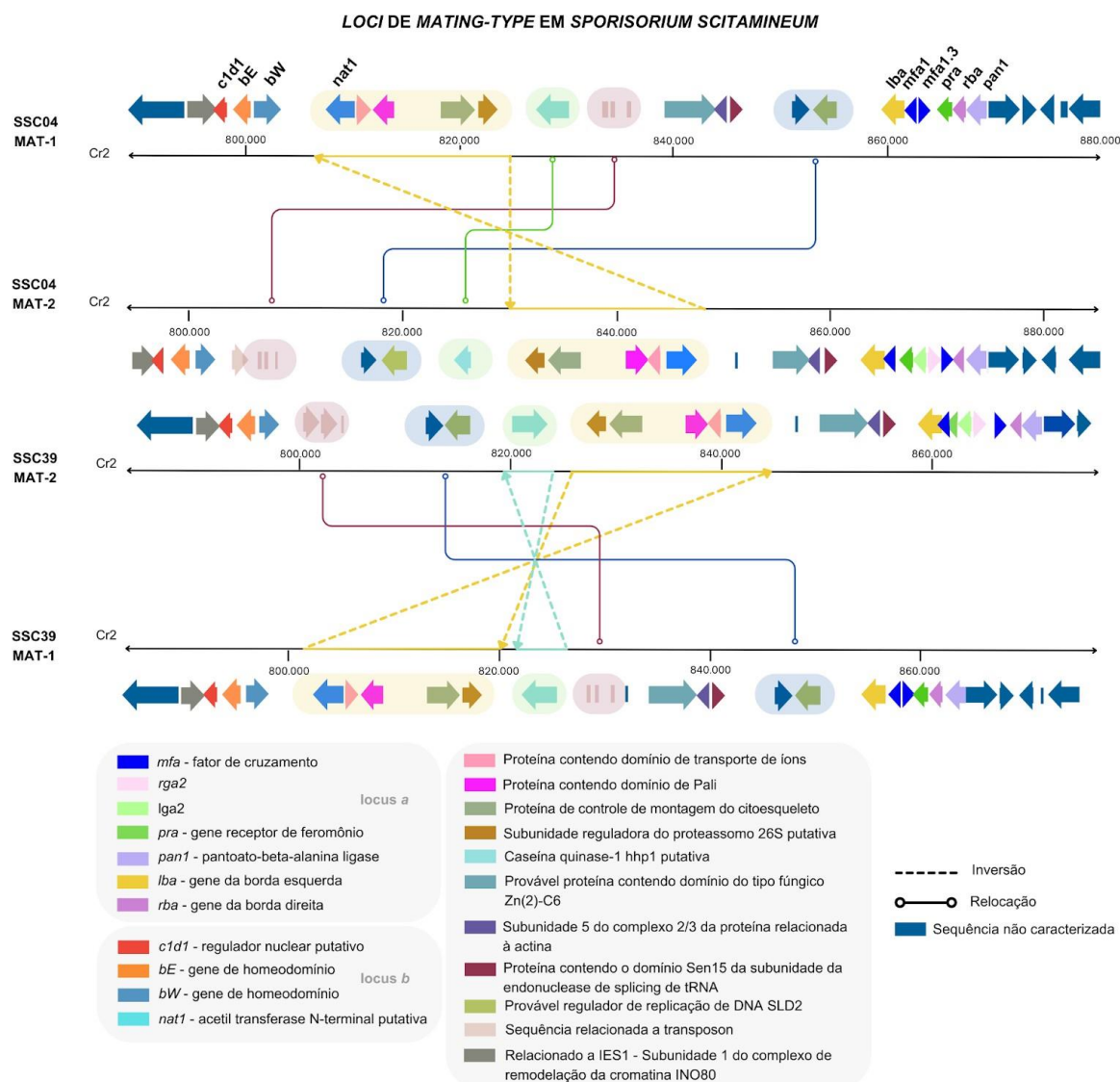


Figura 34 — Alterações estruturais na região entre os *loci* de *mating-type* em diferentes isolados de *S. scitamineum*. Sequências genômicas anotadas foram alinhadas pelo CLC Genomics Workbench (QIAGEN) e ajustadas manualmente.

6.8. Análise de Variantes de Nucleotídeo Único (SNVs)

Variantes de nucleotídeo único (SNVs) foram chamadas a partir dos *softwares* Medaka e Clair3 (ZHENG, Z. et al, 2021). Entre os *mating-types* opostos do isolado SSC04, foram encontradas 469 variantes que apresentaram *phred score* maior que 20 e uma cobertura de no mínimo 20X, através do *software* Medaka. Entretanto, o *software* Clair3 encontrou apenas 205 variantes que cumpriam os

mesmos requisitos estabelecidos. Em busca de analisar o efeito da cobertura genômica sobre a presença destas variantes, foram chamadas variantes até uma cobertura de 10X (Figura 35). Em todas as coberturas analisadas, o *software* Medaka encontrou um número maior de variantes do que o *software* Clair3 para o isolado SSC04. Além disso, mesmo em coberturas diferentes, a variação do número de variantes encontradas pelo Clair3 não foi significativa. Em relação aos *mating-types* do isolado SSC39, o Medaka seguiu sua tendência de encontrar mais variantes do que o Clair3, visto que o *software* encontrou 164 variantes, enquanto o último encontrou apenas 118 variantes. Apesar da flutuação no número de variantes encontradas pelos *softwares* distintos, o isolado SSC04 apresentou um número maior de variantes do isolado SSC39 em todas as coberturas independente do *software* utilizado.

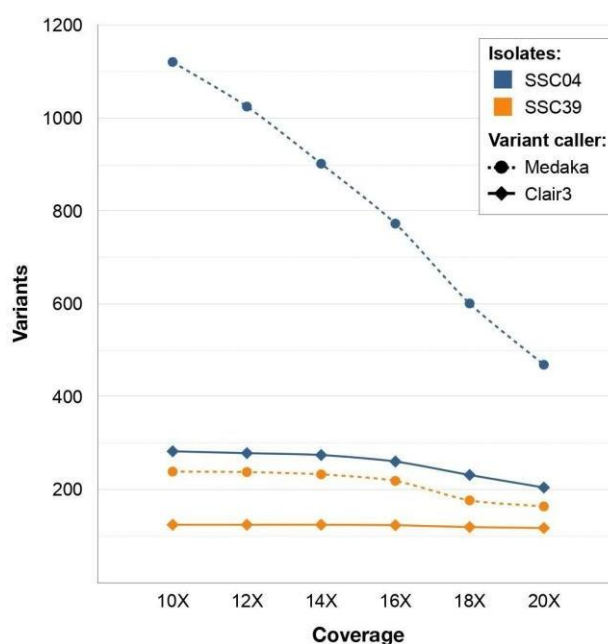


Figura 35 — Número de variantes encontradas entre os *mating-types* de dois isolados de *S. scitamineum* através dos *softwares* Medaka e Clair3. O eixo x representa a cobertura mínima de *reads* consideradas. O eixo y representa o número de variantes.

Visto a variação no número de variantes encontradas por ambos *softwares*, foram consideradas apenas as variantes suportadas por ambos os programas (Figura 36). Desta forma, 145 variantes entre as células haplóides *MAT-1* e *MAT-2* do isolado SSC04 foram suportadas tanto pelo Medaka quanto pelo Clair3,

e, para o isolado SSC39, foram consideradas 88 variantes, evidenciando a diferença de variabilidade entre os isolados.

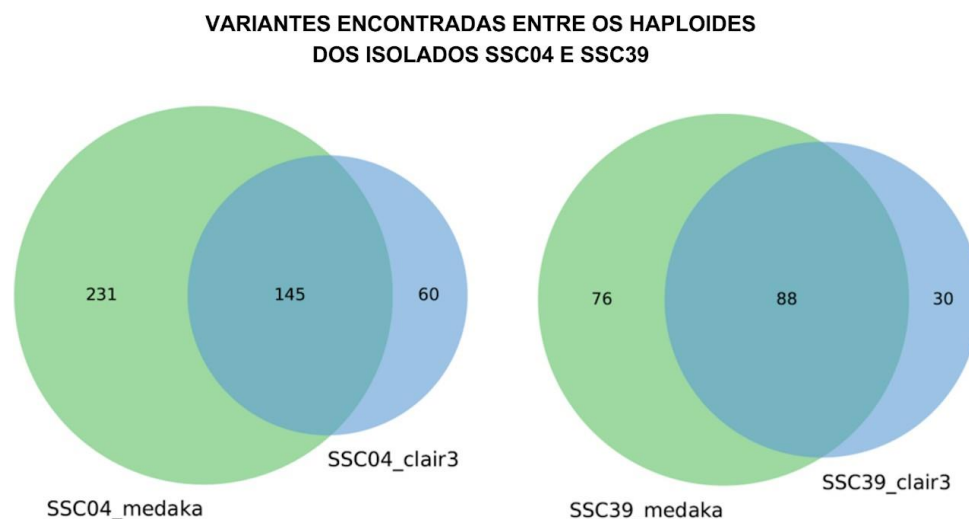


Figura 36 — Variantes entre os *mating-types* opostos dos isolados SSC04 e SSC39 suportadas pelos softwares Medaka e Clair3.

A sequência mais bem montada de *S. scitamineum*, pertencente ao isolado SSC39 *MAT-1*, publicado por Taniguti et al. (2015), foi tomada como sequência de referência. *Reads* provenientes de outras células haplóides foram utilizadas para encontrar variantes em relação à referência (Figura 37). Apenas variantes suportadas pelos softwares Medaka e Clair3 foram consideradas. O isolado SSC04 *MAT-1* apresentou 96 variantes em relação ao genoma SSC39 *MAT-1*, sendo que apenas 3 variantes foram únicas deste genoma. Quando comparado a um diferente isolado e *mating-type* oposto, o isolado SSC04 *MAT-2* apresentou 860 variantes em relação à referência, sendo que 686 foram encontradas apenas neste genoma. Quando comparadas sequências do mesmo isolado, mas de *mating-type* complementares, o isolado SSC39 *MAT-2* apresentou 88 variantes, sendo apenas 7 únicas a ele.

Dentre as 96 variantes encontradas entre o isolado SSC04 *MAT-1* e a referência, 93 destas variantes são compartilhadas com o isolado SSC04 *MAT-2*. Enquanto 81 das 88 variantes encontradas entre o isolado SSC39 *MAT-2* e a referência são compartilhadas com o isolado SSC04 *MAT-2*.

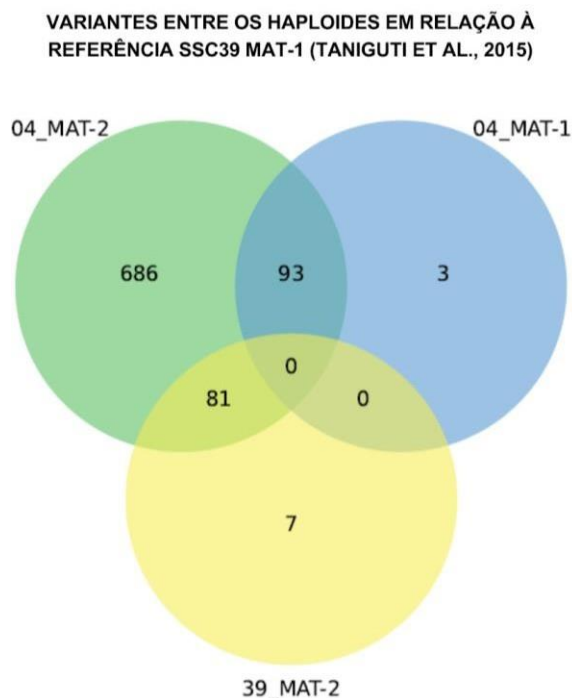


Figura 37 — Variantes encontradas entre os genomas quando comparados com a referência SSC39 MAT-1 (Taniguti et al., 2015).

As 145 variantes encontradas entre os isolados SSC04 MAT-1 e MAT-2 estão distribuídas por todo o genoma, não restritas aos *loci* de *mating-type* presente no cromossomo 2. Por outro lado, as 88 variantes encontradas entre os isolados SSC39 MAT-1 e MAT-2 estão concentradas, em sua maioria, nos *loci* de *mating-type*, sendo que apenas uma variante se encontra fora da região. Desta forma, o isolado SSC04 possui maior variabilidade genética em relação ao isolado SSC39 quando são comparadas as células haploides opostas e complementares. Quando comparados *mating-types* iguais, entretanto, de isolados diferentes, foram encontradas 96 variantes entre os isolados SSC04 MAT-1 e SSC39 MAT-1, sendo que as variantes não estão concentradas no *locus* de *mating-type*. Ao comparar-se os isolados SSC04 MAT-2 e SSC39 MAT-2 foram encontradas 390 variantes distribuídas por todo o genoma.

Desta forma, foi possível observar variação genética dentro dos próprios isolados e a presença de variação até mesmo entre *mating-types* iguais, porém em isolados diferentes, sendo que o alelo MAT-2 é mais variável (Figura 38).

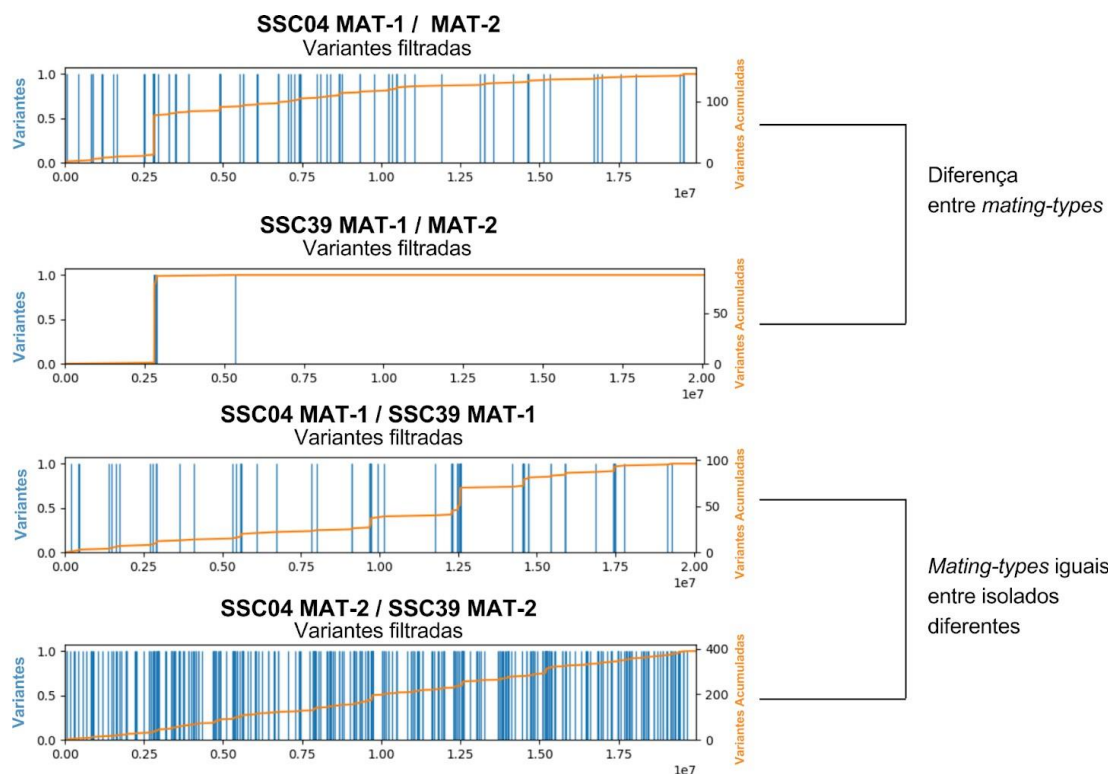


Figura 38 — Distribuição do número de variantes ao longo do genoma de dois isolados diferentes e seus respectivos *mating-types*.

Entre as células haplóides SSC04 *MAT-1* e *MAT-2*, das 145 variantes encontradas, apenas 48 variantes estão presentes em regiões codantes e apenas 22 destas não estão relacionadas aos genes de *mating-type* (Figura 39). Portanto, a maioria destas variantes estão concentradas no cromossomo 2 (Tabela 10).

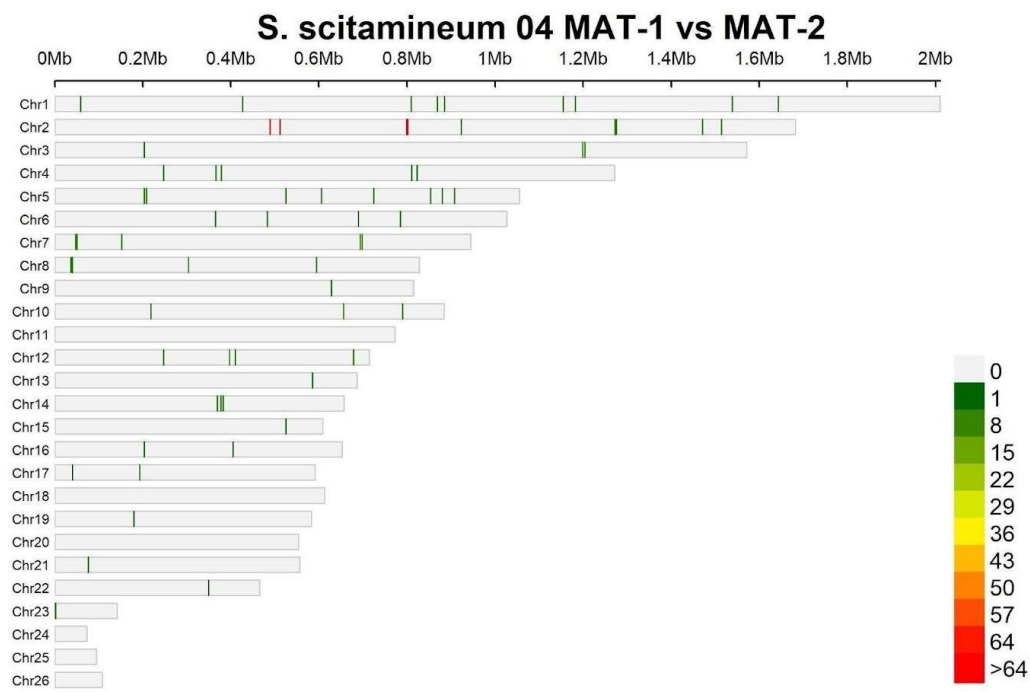


Figura 39 — Distribuição das variantes entre os alelos *MAT-1* e *MAT-2* do isolado SSC04. Escala de 10^6 .

Tabela 10 — Distribuição de variantes em cada cromossomo entre os genomas SSC04 *MAT-1* e *MAT-2*.

<u>Cromossomo</u>	<u>Comprimento</u>	<u>Variantes</u>
1	2,010,843	10
2	1,682,132	72
3	1,571,048	9
4	1,271,557	5
5	1,054,858	9
6	1,026,702	4
7	944,452	7
8	827,561	5
9	814,175	1
10	884,414	4
12	714,214	4
13	686,470	1
14	657,027	3
15	609,286	1
16	652,924	2
17	591,589	2
19	583,410	1
21	556,426	1
22	465,144	1
23	141,908	3
Total	17,746,140	145

A média da qualidade de cada variante baseada no *phred score* foi de 22, sendo que o limiar para ser considerada uma variante foi de 20 (Figura 40).

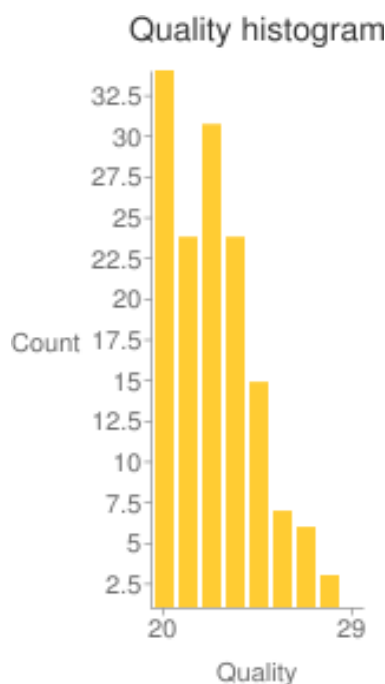


Figura 40 — Número de variantes por qualidade de *phred score* entre os genomas SSC04 MAT-1 e MAT-2.

Dentre as variantes, ocorreram 31 eventos responsáveis por mutações não-sinônimas ou *missense*, ou seja, responsáveis pela substituição de um aminoácido por outro, em decorrência da alteração do códon. 17 mutações sinônimas foram encontradas em decorrência das variações (Tabela 11).

Tabela 11 — Número de efeitos por tipo em relação às variantes entre os genomas SSC04 MAT-1 e MAT-2.

Tipo	Variantes	Percentual (%)
Variantes não-sinônimas	31	3,97
Variantes sinônimas	17	2,18

A maioria das variantes esteve concentrada em regiões à montante (*upstream*) e à jusante (*downstream*) a região codante dos genes (Figura 41). Além disso, das 145 variantes, 59 foram transições, ou seja, substituições de bases pirimídicas por outra pirimídica, e 86 foram transversões, bases pirimídicas trocadas por bases púricas e vice-versa.

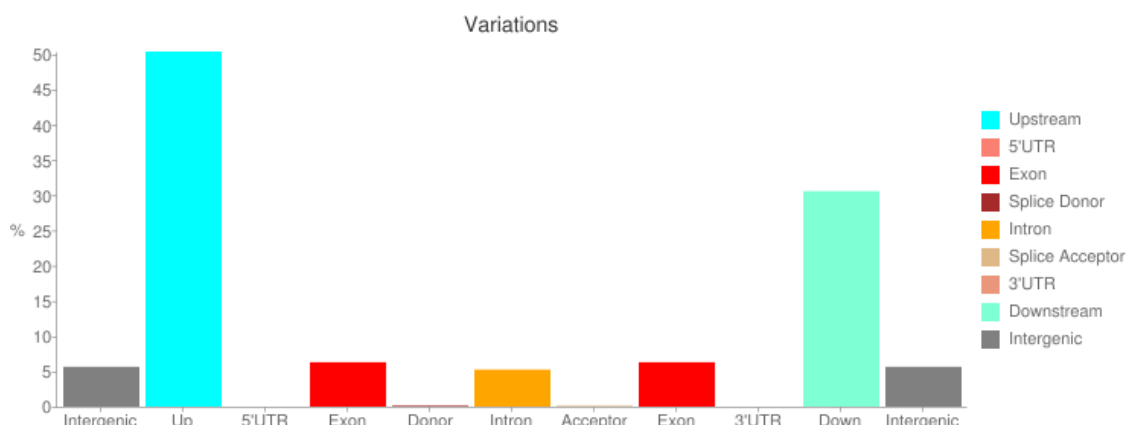


Figura 41 — Distribuição do efeito das variantes entre os genomas SSC04 *MAT-1* e *MAT-2* por tipo e região do genoma.

Já entre as células haplóides SSC39 *MAT-1* e *MAT-2*, das 88 variantes encontradas, apenas 77 caíram em regiões codantes, sendo que 74 estão presentes nos genes de *mating-type* (Figura 42). As 3 variantes em regiões *CDS* restantes estão presentes em um gene responsável pela produção de uma provável proteína contendo domínio do tipo fúngico Zn(2)-C6 (IPR007219, IPR001138, IPR036864) na região entre os *loci a* e *b* de *mating*. Assim como no isolado SSC04, a maioria destas variantes estão concentradas no cromossomo 2, como pode ser visto na Tabela 12.

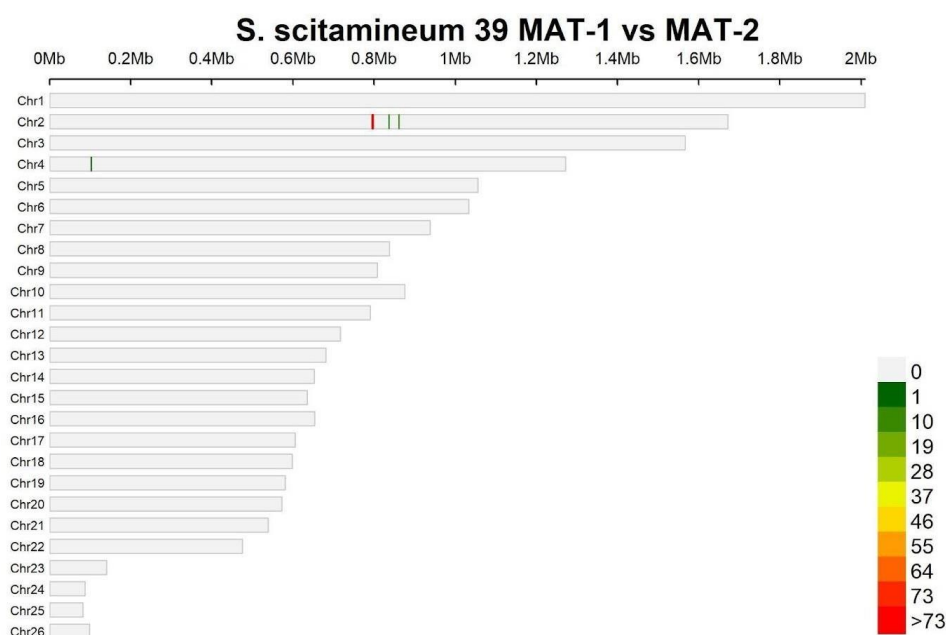
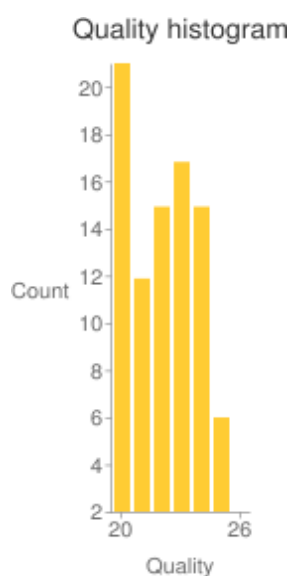


Figura 42 — Distribuição das variantes entre os alelos *MAT-1* e *MAT-2* do isolado SSC39. Escala de 10⁶.

Tabela 12 — Distribuição de variantes em cada cromossomo entre os genomas SSC39 *MAT-1* e *MAT-2*.

Cromossomo	Comprimento	Variantes
2	1,671,606	87
4	1,271,980	1
Total	2,943,586	88

A média da qualidade de cada variante baseada no *phred score* foi de 22, sendo que o limiar para ser considerada uma variante foi de 20, assim como no isolado SSC04 (Figura 43).

**Figura 43** — Número de variantes por qualidade de *phred score* entre os genomas SSC39 *MAT-1* e *MAT-2*.

Dentre as variantes, ocorreram 39 eventos responsáveis por mutações não-sinônimas ou *missense*. 38 mutações sinônimas foram encontradas em decorrência das variações (Tabela 13).

Tabela 13 — Número de efeitos por tipo em relação às variantes entre os genomas SSC39 *MAT-1* e *MAT-2*.

Tipo	Variantes	Percentual (%)
Variantes não-sinônimas	39	10,83
Variantes sinônimas	38	10,55

A maioria das variantes esteve concentrada em regiões à montante (*upstream*) e à jusante (*downstream*) da região codante dos genes, assim como no isolado SSC04 (Figura 44). Das 88 variantes, 47 foram transições, enquanto 41 foram transversões.

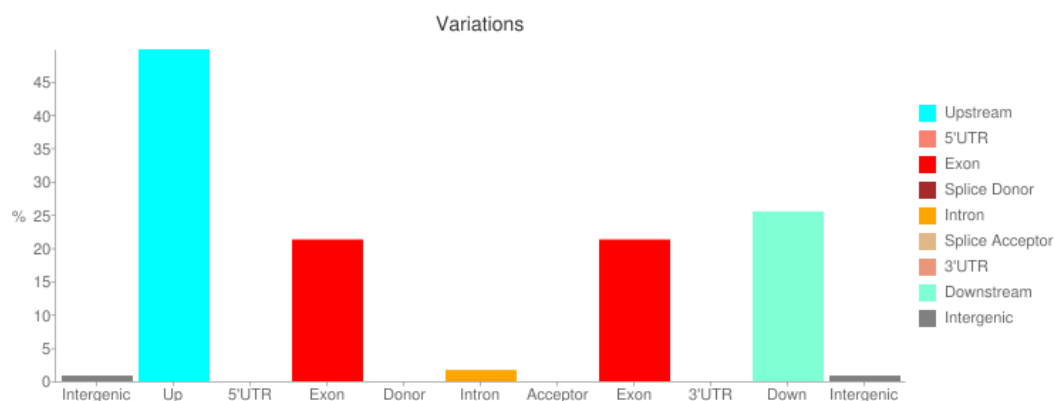


Figura 44 — Distribuição do efeito das variantes entre os genomas SSC39 *MAT-1* e *MAT-2* por tipo e região do genoma.

Para analisar as variantes existentes entre os dois isolados, SSC04 e SSC39, o genoma mais bem montado e com maior profundidade de sequenciamento foi escolhido como referência, neste caso, o genoma SSC39B (*MAT-1*) (Taniguti et al., 2015). Para a etapa de *variant calling*, foram utilizadas sequências do isolado SSC04 com maior quantidade de *reads*, portanto, SSC04A (*MAT-2*). Foram encontradas 860 variantes entre o isolado SSC04 e SSC39, sendo que o cromossomo 2 foi o cromossomo que mais carregou variantes, visto a presença dos *loci* de *mating-type* (Figura 45). Os cromossomos 2, 4, 10, 11 e 16 apresentaram maior número de variação em relação aos demais cromossomos.

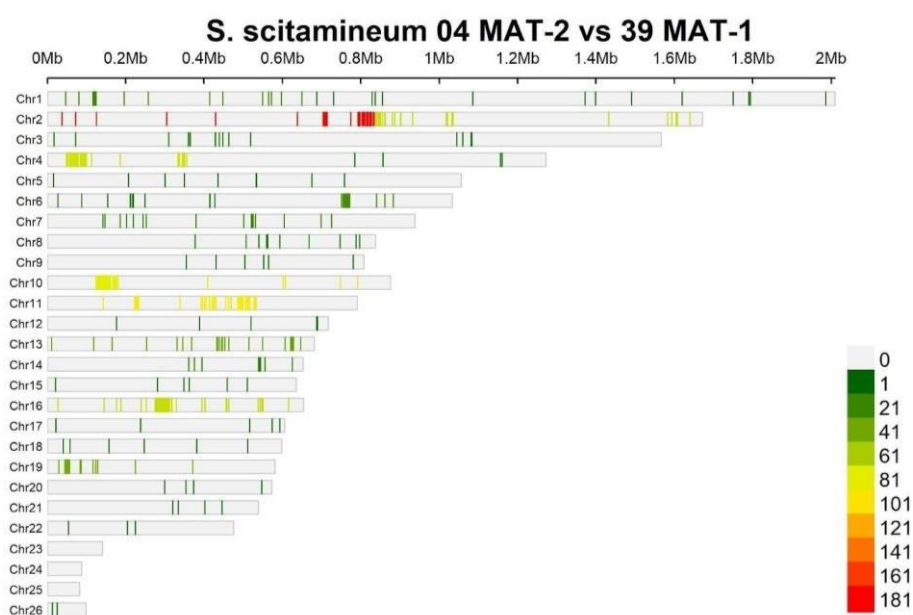


Figura 45 — Distribuição das variantes entre os alelos *MAT-2* do isolado SSC04 e *MAT-1* do isolado SSC39. Escala de 10^6 .

Tabela 14 — Distribuição de variantes em cada cromossomo entre os genomas SSC04 *MAT-2* e SSC39 e *MAT-1*.

Cromossomo	Número de bases (pb)	Número de Variantes
1	2,009,762	38
2	1,671,606	253
3	1,566,833	18
4	1,271,980	80
5	1,055,339	10
6	1,032,954	37
7	938,453	19
8	837,966	10
9	808,176	7
10	875,830	82
11	790,822	98
12	716,888	5
13	681,008	36
14	652,314	16
15	634,981	13
16	653,278	72
17	605,135	5
18	598,089	6
19	580,969	42
20	573,042	4
21	539,152	4
22	475,902	3
26	99,152	2
Total	19,669,631	860

A média da qualidade de cada variante baseada no *phred score* foi de 33, sendo que o limiar para ser considerada uma variante foi de 20, assim como nas análises anteriores, sendo que a variante com maior qualidade foi de 49 (Figura 46).

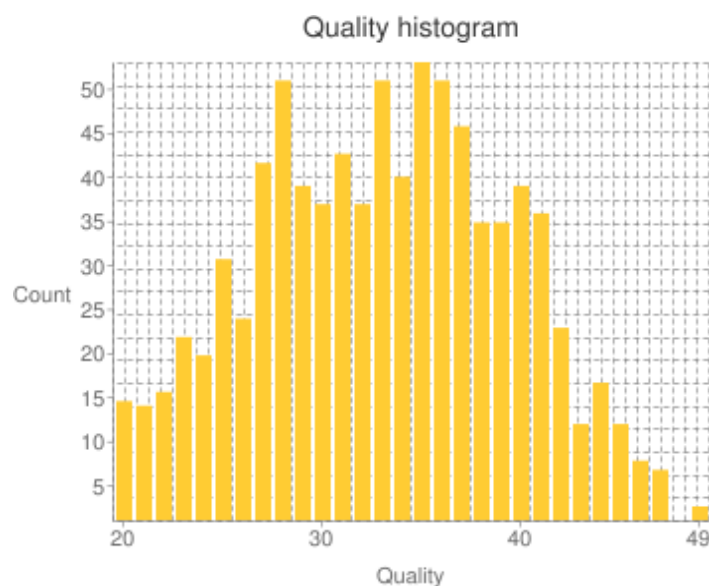


Figura 46 — Número de variantes por qualidade de *phred score* entre os genomas SSC04 *MAT-2* e SSC39 *MAT-1*.

Dentre as variantes, ocorreram 312 eventos responsáveis por mutações não-sinônimas ou *missense*. 253 mutações sinônimas foram encontradas em decorrência das variações (Tabela 15).

Tabela 15 — Número de efeitos por tipo em relação às variantes entre os genomas SSC04 *MAT-2* e SSC39 *MAT-1*.

Tipo	Variantes	Percentual (%)
Variantes não-sinônimas	312	7,94
Variantes sinônimas	253	6,44

A maioria das variantes esteve concentrada em regiões à montante (*upstream*) e à jusante (*downstream*) dos genes, assim como nas análises entre haploides do mesmo isolado (Figura 47). Das 860 variantes, 709 foram transições, enquanto 151 foram transversões.

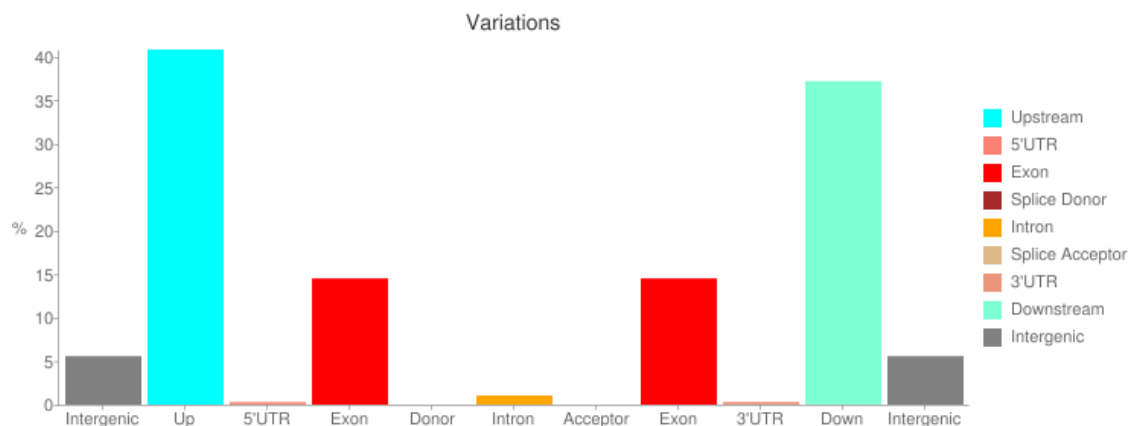


Figura 47 — Distribuição do efeito das variantes entre os genomas SSC04 *MAT-2* e SSC39 *MAT-1* por tipo e região do genoma.

Quando analisadas as variantes em relação a genes candidatos a efetores preditos por alguns estudos (BASSE et al., 2006; KEMPER et al., 2006; KHRUNYK et al., 2010; WOLLENBERG et al., 2014; TANIGUTI et al., 2015; ZHAO et al., 2015), foram encontradas variantes em três genes candidatos a efetores: *g1052*, *g5159* e o *Eff1* (*g4257* e *g4260*) (Figura 48) entre os isolados SSC04 e SSC39. Além disso, é possível observar que existem variantes ao longo do genoma que são específicas ao isolado SSC04, quando comparadas com a referência SSC39 *MAT-1*. Como descrito anteriormente, o isolado SSC04 *MAT-2* possui o maior número de variantes inferidas, e entre o isolado SSC39 *MAT-1* e *MAT-2*, a maioria das variantes está concentrada apenas nos *loci* de *mating-type*. As variantes inferidas entre os dois isolados evidenciam as diferenças alélicas nos genes responsáveis pelo controle da reação sexual, entretanto, foi possível observar que os genes presentes entre os dois *loci* também estão carreando mutações entre os diferentes *matings*, como os genes: *c1d1* (*g969*), *nat1* (*g972*), o gene que codifica uma proteína com domínio Pal1, o gene para o controle da montagem do citoesqueleto e a subunidade do complexo de actina, além da subunidade regulatório do proteassomo. Estes resultados evidenciam que a variação entre os alelos de diferentes *mating-types* não ocorre apenas nos genes envolvidos diretamente com a reação, mas que a região entre os *loci* também é variável.

testar diferentes *variant callers* para um conjunto de dados de *SNPs* atrelados aos genes *BRCA1* e *BRCA2* em seres humanos, o *software* Clair3 atingiu a maior taxa de performance, sendo que o programa ultrapassou o Medaka em acurácia quando utilizando dados de Oxford Nanopore Technologies e, por esta razão, neste estudo, foram consideradas apenas variantes suportadas pelos dois *softwares*, na busca da obtenção de um maior nível de certeza quanto a existência das variantes. Ainda de acordo com Harvey et al. (2023), a chamada de variantes baseadas em *reads* com o *software* Clair3 demonstrou a menor variabilidade entre *softwares* e abordagens, indo ao encontro dos resultados observados no presente estudo, visto que as variantes aparentaram atingir um *plateau* a partir de 16X de cobertura. A variação genética diferencial encontrada entre células do mesmo isolado vai ao encontro dos resultados obtidos por Benevenuto et al. (2016) quando evidenciou a presença de variabilidade até mesmo entre isolados provenientes do mesmo teliósporo. Além da variabilidade genética encontrada no isolado SSC04 não se limitar aos *loci* de *mating-type*, diferente do isolado SSC39, foi realizada a identificação de candidatos a efetores polimórficos entre os dois isolados, em concordância com Teixeira-Silva et al. (2020). As proteínas efetoras são capazes de atuar em uma variedade de alvos para subverter o sistema imune do hospedeiro e, portanto, culminar em um *fitness* melhorado frente ao seu hospedeiro (STERGIOPOULOS; DE WIT, 2009). Por esta razão, a futura caracterização funcional destes candidatos polimórficos será essencial na elucidação de seu papel na agressividade diferencial entre os isolados. *Sporisorium scitamineum*, fungo biotrófico responsável pela doença do carvão na cana-de-açúcar, possui um sistema de cruzamento bipolar, em que células haplóides de *mating-types* opostos e complementares se fundem para produzir uma hifa dicariótica infectiva (ALBERT; SCHENCK, 1996). O processo de cruzamento para início da infecção no hospedeiro é comum às espécies de fungos do carvão. O cruzamento é regulado por dois *loci*, *a* e *b*, sendo que no *locus a*, os genes presentes codificam feromônios e receptores, enquanto no *locus b*, genes codificam duas subunidades de um fator de transcrição heterodimérico (BAKKEREN et al., 2008). Desta forma, o *locus a* é responsável pelo reconhecimento de células haplóides opostas e complementares e sua fusão, exercendo um controle pré-cruzamento (HARTMANN et al., 2021), o *locus b* é responsável pela manutenção da hifa dicariótica infectiva e, portanto, a compatibilidade pós-cruzamento (YAN et al., 2016; HARTMANN et al., 2021). Os fatores de transcrição heterodiméricos

codificados pelo *locus b* possuem papel central na regulação dos genes necessários para a patogênese da hifa dicariótica, como ocorre em *Ustilago maydis* (KRONSTAD; STABEN, 1997). A conservação da organização dos genes *bE* e *bW* no *locus b* observado nos quatro genomas haplóides (SSC04 MAT-1, MAT-2 e SSC39 MAT-1, MAT-2) de *S. scitamineum* segue o padrão dos fungos patogênicos e filamentosos, visto que o *locus b* é altamente conservado entre estes organismos (YAN et al., 2016). Apesar disso, é possível observar a presença de uma região N-terminal variável nos genes *bE* e *bW* entre os alelos *MAT-1* e *MAT-2*, frente a uma região C-terminal conservada (YAN et al., 2016), tornando possível a discriminação molecular pela reação em cadeia da polimerase (PCR) (AGISHA et al., 2021) dos isolados utilizados neste estudo. Entretanto, ao contrário do *locus b*, foi observado que os genes presentes no *locus a* de *mating-type* possuem orientação e localização variável entre os diferentes alelos indo ao encontro dos achados de Benevenuto (2017). Além disso, assim como em *U. maydis*, o *locus a* possui dois genes adicionais *rga2* e *lga2* presentes apenas no alelo *MAT-2* (BAKKEREN et al., 2008). Os genes *rga2* e *lga2*, em *U. maydis*, estão relacionados com a herança mitocondrial uniparental, visto que as mitocôndrias na hifa dicariótica são herdadas de um único *mating*, enquanto as outras mitocôndrias serão degeneradas (HARTMANN et al., 2021).

Foi observada a presença de duas sequências referentes ao gene responsável pela codificação do feromônio envolvido no processo de reconhecimento de células haplóides, *mating factor* ou *mfa*. Estas duas sequências *mfa1* e *mfa1.3*, presentes nos dois alelos, estão lado a lado no alelo *MAT-1*, enquanto no alelo *MAT-2*, flanqueiam os genes *pra* (receptor de feromônio), *lga2* e *rga2*. Ainda, *mfa1* e *mfa1.3* possuem ortólogos em *Sporisorium reilianum*, onde também flanqueiam os genes *pra*, *lga2* e *rga2* (BAKKEREN et al., 2008). Entretanto, segundo Benevenuto (2017), a segunda cópia do gene *mfa* (*mfa1.3*) não possui a sequência que sinaliza o processamento pós-traducional necessário para a produção de um feromônio lipoproteico secretável, assim como descrito por Taniguti et al. (2015).

A região *interlocus* entre os *loci a* e *b* de *mating-type* sofreu eventos de inversão e rearranjos intracromossômicos nos distintos alelos presentes nos dois isolados descritos neste estudo. O sequenciamento genômico progressivo de novos isolados em fungos tem demonstrado a presença de rearranjos no locus de *mating-*

type em diversas espécies (HARTMANN et al., 2021). O segmento invertido no genoma dos alelos *MAT-1* e *MAT-2* em *S. scitamineum* corresponde às sequências dos genes: *nat1* (*g972*), uma N-terminal acetiltransferase, a proteína *g973* contendo um domínio de transporte de íons (IPR027359), a proteína *g974* com domínio *Pal1* (IPR009571), a proteína *g976* de controle da montagem do citoesqueleto (UniProt: A0A0D1E8M) e, por fim, uma subunidade regulatório 26S do proteassomo (*g977*) (UniProt: A0A0D1CGU5). Inversões já foram descritas no *locus* de *mating-type* em outros carvões, como em *Ustilago hordei*, visto que seu alelo *MAT-1* possui um acúmulo de sequências repetitivas remissivas de elementos transponíveis que podem estar envolvidos no rearranjo do genoma ao longo da evolução (BAKKEREN et al., 2008). Além disso, estes elementos repetitivos, segundo Bakkeren et al. (2008), são frequentemente encontrados entre regiões sintênicas. Estas sequências remissivas também foram encontradas na região entre os *loci* *a* e *b* em *S. scitamineum*, ao lado dos eventos de inversão descritos, como descrito em Taniguti et al. (2015). De acordo com Hartmann et al. (2021), “o acúmulo destes elementos móveis tendem a ocorrer, especialmente, em regiões de recombinação suprimida, já que a seleção contra inserções deletérias é menos efetiva que em regiões de recombinação livre”. O próprio fato da acumulação de mutações em regiões repetitivas resulta em um decréscimo das taxas de recombinação, visto o decréscimo da similaridade entre as regiões (HARTMANN et al., 2021).

Taniguti et al. (2015) reportaram que o cromossomo 2 de *S. scitamineum* é homólogo aos cromossomos 1 e 20 de *S. reilianum*. Os genes de *mating-type* de *S. reilianum* estão presentes nestes cromossomos e, por esta razão, o fungo possui um sistema de cruzamento tetrapolar, visto que os genes segregam de forma independente. A similaridade aos cromossomos 1 e 20 termina na região entre os *loci* *a* e *b* em *S. scitamineum*, sugerindo a transição do sistema tetrapolar para bipolar, onde os genes de *mating-type* estão ligados e segregam como uma unidade. Os elementos transponíveis podem ter sido os protagonistas no processo de fusão dos *loci* de *mating-type*, sendo que a transição de um estado tetrapolar para bipolar pode ter ocorrido de forma independente e múltiplas vezes através do fenômeno de ligação cromossômica (BENEVENUTO et al., 2017). Visto a ocorrência de um processo de ligação cromossômica em *S. scitamineum* quanto aos *loci* de *mating-type*, um indivíduo diplóide é capaz de gerar apenas dois alelos relacionados ao *mating*, permitindo com que cada célula haplóide seja capaz de se cruzar com

metade das outras células haploides e, desta forma, o fenômeno de *linkage* é vantajoso e tem ocorrido em diversos gêneros dentro do grupo dos basidiomicetos (HARTMANN et al., 2021).

Segundo Hartmann et al. (2021), a incorporação de genes nos *loci* de *mating-type* que não estão envolvidos diretamente na determinação do *mating* é comum em fungos. Neste estudo, foi possível observar a presença de genes na região entre os *loci* *a* e *b* que estão majoritariamente relacionados a uma variedade de processos celulares. Dentre estes genes, estão genes relacionados com a arquitetura celular, visto a presença de uma proteína envolvida no controle da montagem do citoesqueleto, além de uma subunidade do complexo proteico relacionado à actina. A presença destes genes nesta região evidencia o fato de que fungos causadores do carvão devem passar por transições morfológicas significativas, visto a transição do estado unicelular para o estado multicelular infectivo (MARTÍNEZ-SOTO et al., 2020). Estes achados corroboram estudos de Hartmann et al. (2021) quando descrevem que, em basidiomicetos, existem genes que estão envolvidos no processo de cruzamento, além dos genes determinantes, presentes no *locus* de *mating-type* como, por exemplo, fatores de transcrição e proteínas quinase. Segundo estes pesquisadores, os genes envolvidos na cascata de sinalização MAP quinase (MAPK) de *Cryptococcus* spp. foram ligados ao *locus* de *mating-type*.

Segundo Martínez-Soto et al. (2020), a via das MAPK é responsável pela regulação de processos fisiológicos, como a multicelularidade e o cruzamento. Em *U. maydis*, a via *MAPK PMM* (do inglês, “*pathogenesis, mating and morphogenesis*”, ou seja, patogênese, cruzamento e morfogênese) é responsável pelo crescimento pluricelular e, dentre os genes regulados por esta via, estão genes que codificam a organização do citoesqueleto de actina e genes que codificam quinases como serina/treoninas quinases (MARTÍNEZ-SOTO et al., 2020). A presença dos genes relacionados ao controle da montagem do citoesqueleto (UniProt: A0A0D1E8M) e da actina (UniProt: A0A0D1C8W4) e, a presença de uma caseína quinase (*g979*) *hhp1* putativa (uma serina/treonina quinase) (UniProt: A0A0D1EA76) na região entre os *loci* de *mating-type* em *S. scitamineum* pode indicar que estes genes estão envolvidos nas vias responsáveis pela transição do estado leveduriforme para a hifa dicariótica. Uma das sequências gênicas envolvidas no segmento invertido entre os alelos contém um domínio *Pal1* (IPR009571),

potencialmente envolvido no crescimento da extremidade da célula, permitindo um crescimento polarizado, fenômeno observado durante a formação do tubo de conjugação mediante cruzamento positivo entre células haplóides opostas e complementares. Além disso, existem genes na região que podem atuar diretamente na regulação da expressão gênica, como uma sequência que possui domínios de ligação ao DNA (IPR007219, IPR001138, IPR036864), ainda, na região, é observada a presença de genes relacionados com a regulação da replicação e remodelamento do complexo da cromatina, potencialmente desempenhando papel na divisão celular para a formação da hifa infectiva.

É interessante notar que a sequência do gene responsável pela codificação da caseína quinase *hhp1* (*g979*) tem orientações opostas nos genomas *MAT-1* e *MAT-2* do isolado SSC39, no isolado SSC04, o gene segue a mesma orientação. Isso indica que, no isolado SSC39, o gene *hhp1* foi invertido junto com o segmento descrito anteriormente e, por esta razão, adquiriu orientações opostas nos diferentes alelos. No isolado SSC04, o gene *hhp1* não foi incluído no segmento invertido, preservando sua orientação nos dois alelos. Desta forma, quando avaliada a sintenia entre genomas de mesmo *mating*, os genomas SSC04 *MAT-1* e SSC39 *MAT-1* apresentaram sintenia completa, enquanto que os genomas SSC04 *MAT-2* e SSC39 *MAT-2* apresentaram uma inversão, indicando que apesar de alelos iguais, as sequências são diferentes entre os dois isolados.

A maioria das variantes encontradas entre diferentes alelos dos isolados SSC04 e SSC39 está concentrada na região dos *loci* de *mating-type*. Segundo Hartmann (2021), os alelos que estão ligados ao *locus* de *mating* tendem a acumular mutações de forma independente ao longo do tempo, nos diferentes genomas dos *mating-types* complementares, visto a ocorrência de uma supressão da recombinação na região. Essa supressão da recombinação ocorre para que cada *mating-type* consiga manter sua combinação alélica e, conseqüentemente, um cruzamento correto, já que a compatibilidade das células será determinada por mecanismos moleculares que estão relacionados e controlados pelos *loci* de *mating-type* (HARTMANN et al., 2021). É interessante notar que, se há uma recombinação suprimida nestes *loci*, os rearranjos que ocorrem são acumulados de forma independente em cada alelo. Ainda segundo Hartmann et al. (2021), em *Schizosaccharomyces*, foi identificada a presença de metilação da histona (H3) e interferência por RNA, fenômenos associados à supressão da recombinação. Neste

sentido, é importante ressaltar a presença de um gene provavelmente relacionado com uma subunidade do complexo de remodelação da cromatina adjacente ao *locus b* em *S. scitamineum*, logo à montante da região de *mating-type*, pois pode desempenhar, também, algum papel na conformação dos *loci*. Apesar do progresso contínuo nas estratégias de sequenciamento genômico, são necessários estudos posteriores para identificar e entender as causas da supressão da recombinação nestes *loci* de *mating-type* e os efeitos causados pelas variações estruturais.

8. CONSIDERAÇÕES FINAIS

Os *primers* desenvolvidos por Agisha et al. (2022) foram bem-sucedidos na discriminação molecular de *mating-types* opostos e complementares em *S. scitamineum*. Além disso, o gene candidato a efetor *g5159* pode atuar como um marcador molecular para a identificação dos distintos haplótipos dos isolados brasileiros descritos por Benevenuto et al. (2016). A plataforma Oxford Nanopore MinION se constitui como uma alternativa viável na geração de sequências longas e de alta cobertura na montagem de genomas fúngicos, como *S. scitamineum*, a partir da obtenção de DNA de alto peso molecular.

Genes responsáveis por controlar o cruzamento sexual, presentes no *locus b* de *mating-type*, tiveram orientação e posição conservadas entre todos os genomas analisados, entretanto, os genes no *locus a* são polimórficos entre os *matings*. Os genes presentes na região entre os *loci* de *mating-type* que não estão diretamente ligados com a determinação sexual, também carregam mutações entre os *mating-types* e entre os isolados, estas sequências podem estar envolvidas com processos relacionados com a biologia da reação de cruzamento e a transição do estado unicelular leveduriforme para o estado pluricelular na formação de uma hifa dicariótica infectiva. As causas das alterações estruturais nos *loci* de *mating-type* têm sido pouco exploradas entre os fungos e estudos posteriores serão necessários para avaliar a natureza evolutiva das mudanças.

A partir dos resultados obtidos, foi possível acessar as variações genéticas entre diferentes isolados do fungo biotrófico *S. scitamineum* com níveis diferentes de agressividade na doença do carvão da cana-de-açúcar. As análises genéticas revelaram a existência de variação entre células haplóides opostas e complementares de mesmo isolado, além de que o alelo *MAT-2* do isolado SSC39 possui uma inversão em relação ao mesmo alelo do isolado SSC04. O isolado mais agressivo apresentou um número elevado de variantes nucleotídicas únicas em relação ao isolado menos agressivo, indicando uma maior heterozigosidade. Além disso, alguns genes candidatos a efetores entre os dois isolados apresentaram mutações não-sinônimas que devem ser posteriormente caracterizadas na busca de compreender seus efeitos na agressividade.

Espera-se que os resultados recentes apresentados neste estudo contribuam para um maior entendimento sobre a natureza da variabilidade genética

de *S. scitamineum* e sua relação com a agressividade dos sintomas na doença do carvão da cana-de-açúcar.

REFERÊNCIAS

- AGISHA, V. N. et al. Molecular Discrimination of Opposite Mating Type Haploids of *Sporisorium scitamineum* and Establishing Their Dimorphic Transitions During Interaction with Sugarcane. **Sugar Tech**, v. 24, n. 5, p. 1430–1440, 12 jan. 2022.
- ALBERT, H.H.; SCHENCK, S. PCR amplification from a homolog of the bE mating-type gene as a sensitive assay for the presence of *Ustilago scitaminea* DNA. **Plant Disease**, Saint Paul, v. 80, n. 10, p. 1189-1192, 1996.
- ALI, S. et al. An Immunity-Triggering Effector from the Barley Smut Fungus *Ustilago hordei* Resides in an Ustilaginaceae-Specific Cluster Bearing Signs of Transposable Element-Assisted Evolution. **PLOS Pathogens**, v. 10, n. 7, 3 jul. 2014.
- AMARASINGHE, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. **Genome Biology**, v. 21, n. 1, 7 fev. 2020.
- AMORIM, Lílian. Epidemiologia do carvão (*Ustilago scitaminea*, Syd.) da cana-de-açúcar: curvas anuais e poliéticas de progresso da doença. 1989. Tese (Doutorado em Fitopatologia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1990.
- BABU, K. S, D. et al. A short review on sugarcane: its domestication, molecular manipulations and future perspectives. **Genetic Resources and Crop Evolution**, 16 set. 2022.
- BAKKER, H. **The Origins of Sugar Cane**. Sugar Cane Cultivation and Management, p. 1–2, 1999.
- BAKKEREN, G.; KÄMPER, J.; SCHIRAWSKI, J. Sex in smut fungi: Structure, function and evolution of mating-type complexes. **Fungal Genetics and Biology**, v. 45, p. S15–S21, ago. 2008.
- BERGER, S.; SINHA, A. K.; ROITSCH, T. Plant physiology meets phytopathology: plant primary metabolism and plant pathogen interactions. **Journal of Experimental Botany**, v. 58, n. 15-16, p. 4019–4026, 26 nov. 2007.
- BENEVENUTO, J. et al. Comparative Genomics of Smut Pathogens: Insights From Orphans and Positively Selected Genes Into Host Specialization. **Frontiers in Microbiology**, v. 9, 6 abr. 2018.
- BLANCH, P. A.; ASHER, M. J. C.; BURNETT, J. H. Inheritance of pathogenicity and cultural characters in *Gaeumannomyces graminis* var. tritici. **Transactions of the British Mycological Society**, v. 77, n. 2, p. 391–399, out. 1981.
- BORRÁS-HIDALGO, O. et al. Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with *Ustilago scitaminea* or *Bipolaris sacchari*. **Plant Physiology and Biochemistry**, v. 43, n. 12, p. 1115–1121, dez. 2005.

BORDONAL, R. DE O. et al. Sustainability of sugarcane production in Brazil. A review. **Agronomy for Sustainable Development**, v. 38, n. 2, 27 fev. 2018.

BHUIYAN, S. A. et al. Sugarcane Smut, Caused by *Sporisorium scitamineum*, a Major Disease of Sugarcane: A Contemporary Review. **Phytopathology**, v. 111, n. 11, p. 1905–1917, nov. 2021.

CARDOSO-SILVA, C. B. et al. De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. **PLoS ONE**, v. 9, n. 2, p. e88462, 11 fev. 2014.

CARVALHO, R. DE S. Carvão da cana. **Anais da Escola Superior de Agricultura Luiz de Queiroz**, v. 6, p. 1–12, 1949.

CARVALHO, G. et al. *Sporisorium scitamineum* colonization of sugarcane genotypes susceptible and resistant to smut revealed by GFP-tagged strains. **Annals of Applied Biology**, v. 169, n. 3, p. 329–341, 27 jul. 2016.

CASTIBLANCO, V.; CASTILLO, H. E.; MIEDANER, T. Be flexible and adapt easily—The great role of plasticity relative to genetic variation for aggressiveness of *Fusarium culmorum* isolates. **Journal of Phytopathology**, v. 168, n. 3, p. 162–174, 6 jan. 2020.

CATEN, C. E. et al. The genetics of pathogenic aggressiveness in three dikaryons of *Ustilago hordei*. **Canadian Journal of Botany**, v. 62, n. 6, p. 1209–1219, 1 jun. 1984.

CHANG, M. et al. PTI and ETI: convergent pathways with diverse elicitors. **Trends in Plant Science**, v. 27, n. 2, p. 113–115, fev. 2022.

CHERUBIN, M. R. et al. Land Use and Management Effects on Sustainable Sugarcane-Derived Bioenergy. **Land**, v. 10, n. 1, p. 72, 15 jan. 2021.

CINGOLANI, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. **Fly**, v. 6, n. 2, p. 80–92, abr. 2012.

COMSTOCK, J. C. Smut. In: ROTT, P. A Guide to Sugarcane Diseases. p. 181-185. 2000.

CRESTANA, Gustavo Schiavone. **Caracterização de genes candidatos a efetores de *Sporisorium scitamineum*: diversidade genética e expressão gênica**. 53 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Ciências Biológicas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2017.

CRESTANA, Gustavo Schiavone. **The complete genome sequence and comparative effectorome of *Sporisorium panici-leucophaei*: the causal smut disease agent in sourgrass (*Digitaria insularis*)**. 2020. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2020.

CUNNAC, S.; LINDEBERG, M.; COLLMER, A. *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. **Current Opinion in Microbiology**, v. 12, n. 1, p. 53–60, 1 fev. 2009.

CURSI, D. E. et al. History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. **Sugar Tech**, 16 fev. 2021.

DANIELS, J.; ROACH, B. T. Taxonomy and Evolution. *In*: HEINZ, D. J. **Sugarcane improvement through breeding**. Amsterdam: Elsevier, 1987. p. 7-84.

DAVIS, S. vcftoolz: a Python package for comparing and evaluating Variant Call Format files. **Journal of Open Source Software**, v. 4, n. 35, p. 1144, 26 mar. 2019.

DNA sequencing. Nature. Disponível em: <<https://www.nature.com/subjects/dna-sequencing#:~:text=Definition>>. Acesso em: 18 jan. 2023.

DOYLE, J.L.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v. 19, n. 1. p. 11-15. 1987.

DUTHEIL, J. Y. et al. A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. **Genome Biology and Evolution**, v. 8, n. 3, p. 681–704, 12 fev. 2016.

EL CHAMI, D. et al. What are the impacts of sugarcane production on ecosystem services and human well-being? A review. **Annals of Agricultural Sciences**, nov. 2020.

FABRE, F. et al. Unbalanced Roles of Fungal Aggressiveness and Host Cultivars in the Establishment of the Fusarium Head Blight in Bread Wheat. **Frontiers in Microbiology**, v. 10, 11 dez. 2019.

FAULKNER, C.; ROBATZEK, S. Plants and pathogens: putting infection strategies and defence mechanisms on the map. **Current Opinion in Plant Biology**, v. 15, n. 6, p. 699–707, dez. 2012.

FAURE, C. et al. Long-Read Genome Sequence of the Sugar Beet Rhizosphere Mycoparasite *Pythium oligandrum*. **G3 Genes|Genomes|Genetics**, v. 10, n. 2, p. 431–436, 1 fev. 2020.

FELDBRÜGGE, M.; KELLNER, R.; SCHIPPER, K. The biotechnological use and potential of plant pathogenic smut fungi. **Applied Microbiology and Biotechnology**, v. 97, n. 8, p. 3253–3265, 2 mar. 2013.

Food and Agriculture Organization (FAO). Production/Yield quantities of Sugar cane in World + (Total). Disponível em: <<https://www.fao.org/faostat/>>. Acesso em: 19 jan. 2023

GOES, T. et al. Sugarcane in Brazil: Current technologic stage and perspectives. **Revista de Política Agrícola**, v. 20, n. 1, p. 52–65, 30 mar. 2011.

HARTMANN, F. E. et al. Recombination suppression and evolutionary strata around *mating-type* loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. **New Phytologist**, v. 229, n. 5, p. 2470–2491, dez. 2020.

HARVEY, W. T. et al. Whole-genome long-read sequencing downsampling and its effect on variant calling precision and recall. 4 maio 2023.

HELAL, A. A. et al. Evaluation of the Available Variant Calling Tools for Oxford Nanopore Sequencing in Breast Cancer. **Genes**, v. 13, n. 9, p. 1583, 3 set. 2022.

HIDAYAH, N. et al. Resistance mechanisms and expression of disease resistance-related genes in sugarcane (*Saccharum officinarum*) to *Sporisorium scitamineum* infection. **Functional Plant Biology**, v. 48, n. 12, p. 1302, 2021.

HOGENHOUT, S. A. et al. Emerging Concepts in Effector Biology of Plant-Associated Organisms. **Molecular Plant-Microbe Interactions**, v. 22, n. 2, p. 115–122, fev. 2009.

KAFFKA, S. R.; GRANTZ, D. A. Sugar Crops. In: ALFEN, N.K.V. **Encyclopedia of Agriculture and Food Systems**. Elsevier: Amsterdam, 2014.

KHELIK, K. et al. NucDiff: In-depth characterization and annotation of differences between two sets of DNA sequences. **BMC Bioinformatics**, v. 18, n. 1, p. 338, 12 jul. 2017.

KOLMOGOROV, M. et al. Assembly of long, error-prone reads using repeat graphs. **Nature Biotechnology**, v. 37, n. 5, p. 540–546, 1 abr. 2019.

KOREN, S. et al. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. **Genome Research**, v. 27, n. 5, p. 722–736, 1 maio 2017.

KOVAKA, S. et al. Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing. **Nature Methods**, v. 20, n. 1, p. 12–16, 1 jan. 2023.

KRONSTAD, J. W.; STABEN, C. Mating type in filamentous fungi. **Annual Review of Genetics**, v. 31, n. 1, p. 245–276, dez. 1997.

KRZYWINSKI, M. et al. Circos: an information aesthetic for comparative genomics. **Genome Research**, v. 19, n. 9, p. 1639–45, 2009.

KURTZ, S. et al. Versatile and open software for comparing large genomes. **Genome biology**, v. 5, n. 2, p. R12, 30 jan. 2004.

JAISWAL, D. et al. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. **Nature Climate Change**, v. 7, n. 11, p. 788–792, 23 out. 2017.

JASWAL, R. et al. Effector Biology of Biotrophic Plant Fungal Pathogens: Current Advances and Future Prospects. **Microbiological Research**, v. 241, p. 126567, dez. 2020.

LATIZA, A.S. et al. Reaction of sugarcane clones to strain B of *Sporisorium scitamineum*, **International Society of Sugarcane Technologists**, v. 2, p. 1456–1462, 1980.

LAURIE, J. D. et al. Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. **The Plant Cell**, v. 24, n. 5, p. 1733–1745, 1 maio 2012.

LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, n. 21, p. 2987–2993, 8 set. 2011.

LONGATTO, Daniel Prezotto. **Herança dos polimorfismos de restrição associados à região subtelomérica de *Sporisorium scitamineum* em análise de cruzamentos sexuais do fungo *in planta***. 2014. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2014.

LONGATTO, Daniel P et al. Carvão da cana-de-açúcar: avanços na compreensão deste patossistema. **Revisão Anual de Patologia de Plantas**. Tradução. Brasília: Sociedade Brasileira de Fitopatologia, 2015. v. 23. p. 62-89.

NALAYENI, K. et al. Comparative expression analysis of potential pathogenicity-associated genes of high- and low-virulent *Sporisorium scitamineum* isolates during interaction with sugarcane. **3 Biotech**, v. 11, n. 7, 23 jun. 2021.

MAEKAWA, H.; KANEKO, Y. Inversion of the Chromosomal Region between Two Mating Type Loci Switches the Mating Type in *Hansenula polymorpha*. **PLoS Genetics**, v. 10, n. 11, p. e1004796, 20 nov. 2014.

MAGAREY, R. C. et al. Yield losses caused by sugarcane smut in several crops in Queensland. **Proceedings of the Australian Society of Sugar Cane Technologists**, 2010.

MAIA, T. et al. A Bacterial Type Three Secretion-Based Delivery System for Functional Characterization of *Sporisorium scitamineum* Plant Immune Suppressing Effector Proteins. **Phytopathology**, v. 112, n. 7, p. 1513–1523, 1 jul. 2022.

MANNI, M. et al. BUSCO: Assessing Genomic Data Quality and Beyond. **Current Protocols**, v. 1, n. 12, dez. 2021.

MARIN, Fábio Ricardo. **Eficiência de produção da cana-de-açúcar brasileira: estado atual e cenários futuros baseados em simulações multimodelos**. 2014. Tese (Livre Docência em Agrometeorologia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2014.

MARTÍNEZ-SOTO, D. et al. Molecular Mechanisms Involved in the Multicellular Growth of Ustilaginomycetes. **Microorganisms**, v. 8, n. 7, p. 1072, 18 jul. 2020.

MCKENZIE, S. K.; WALSTON, R. F.; ALLEN, J. L. Complete, high-quality genomes from long-read metagenomic sequencing of two wolf lichen thalli reveals enigmatic genome architecture. **Genomics**, v. 112, n. 5, p. 3150–3156, set. 2020.

MARQUES, J. P. R. et al. Sugarcane smut: shedding light on the development of the whip-shaped sorus. **Annals of Botany**, v. 119, n. 5, p. 815–827, 2016.

MARQUES, J. P. R. et al. Sugarcane Cell Wall-Associated Defense Responses to Infection by *Sporisorium scitamineum*. **Frontiers in Plant Science**, v. 9, 23 maio 2018.

MENZEL, P.; NG, K. L.; KROGH, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. **Nature Communications**, v. 7, n. 1, 13 abr. 2016.

MIN, B.; GRIGORIEV, I. V.; CHOI, I.-G. FunGAP: Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. **Bioinformatics**, v. 33, n. 18, p. 2936–2937, 15 set. 2017.

MONTEIRO VITORELLO, C. B. et al. Progress in understanding fungal diseases affecting sugarcane: smut. Achieving sustainable cultivation of sugarcane Volume 2, p. 221–243, 5 mar. 2018.

MONTEIRO-VITORELLO, C. B. et al. The Genome Sequence of the Gram-Positive Sugarcane Pathogen *Leifsonia xyli* subsp. *xyli*. **Molecular Plant-Microbe Interactions**, v. 17, n. 8, p. 827–836, ago. 2004.

MOOLHUIJZEN, P.; SEE, P. T.; MOFFAT, C. S. The first genome assembly of fungal pathogen *Pyrenophora tritici-repentis* race 1 isolate using Oxford Nanopore MinION sequencing. **BMC Research Notes**, v. 14, n. 1, 28 ago. 2021.

MOREIRA, A. S. et al. Pathogenic and molecular comparison of *Puccinia kuehnii* isolates and reactions of sugarcane varieties to orange rust. **Plant Pathology**, v. 67, n. 8, p. 1687–1696, 11 maio 2018.

NALAYENI, K. et al. Comparative expression analysis of potential pathogenicity-associated genes of high- and low-virulent *Sporisorium scitamineum* isolates during interaction with sugarcane. **3 Biotech**, v. 11, n. 7, 23 jun. 2021.

OLIVEIRA, Lâina da Silva de. **Análise da agressividade do patógeno *Sporisorium scitamineum* quanto a infecção em cana-de-açúcar**. 39 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Ciências Biológicas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2021.

PARIAUD, B. et al. Aggressiveness and its role in the adaptation of plant pathogens. **Plant Pathology**, v. 58, n. 3, p. 409–424, 1 jun. 2009.

PIEPENBRING, M.; STOLL, M.; OBERWINKLER, F. The generic position of *Ustilago maydis*, *Ustilago scitaminea*, and *Ustilago esculenta* (Ustilaginales). **Mycological Progress**, v. 1, n. 1, p. 71–80, fev. 2002.

PETERS, L. P. et al. Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. **Planta**, v. 245, n. 4, p. 749–764, 21 dez. 2016.

POTGIETER, L. et al. On Variant Discovery in Genomes of Fungal Plant Pathogens. **Frontiers in Microbiology**, v. 11, 16 abr. 2020.

QAMAR, Z. et al. Novel approaches to circumvent the devastating effects of pests on sugarcane. **Scientific Reports**, v. 11, n. 1, 14 jun. 2021.

QUE, Y. et al. Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. **BMC Genomics**, v. 15, n. 1, p. 996, 2014.

RAGO et al. Variabilidade patogênica de *Ustilago scitaminea* no Estado de São Paulo. **Summa Phytopathologica**, Jaguariúna, v. 35, n. 2, p. 93-97, 2009.

RAJPUT, M. A. et al. Sugarcane Smut: Current Knowledge and the Way Forward for Management. **Journal of Fungi**, v. 7, n. 12, p. 1095, 19 dez. 2021.

RICAUD et al., **Diseases of Sugarcane: Major Diseases**. Elsevier Science, 2012.

ROBINSON, J. T. et al. Integrative genomics viewer. **Nature Biotechnology**. Nature Publishing Group, 1 jan. 2011.

SCHAKER, P. D. C. et al. RNAseq Transcriptional Profiling following Whip Development in Sugarcane Smut Disease. **PLOS ONE**, v. 11, n. 9, p. 1-21, 1 set. 2016.

SCHENCK, S. New race of sugarcane smut on Maui. **Pathology Report**, v. 69, p. 1-3, 2003.

SCHWESSINGER, B.; RATHJEN, J. P. Extraction of High Molecular Weight DNA from Fungal Rust Spores for Long Read Sequencing. **Methods in Molecular Biology (Clifton, N.J.)**, v. 1659, p. 49–57, 2017.

SNURK, S. et al. The complete sequence of a human genome. **Science**, v. 376, n. 6588, p. 44–53, abr. 2022.

SOUZA, S. P. et al. Sugarcane can afford a cleaner energy profile in Latin America & Caribbean. **Renewable Energy**, v. 121, p. 164–172, jun. 2018.

STERGIOPOULOS, I.; DE WIT, P. J. G. M. Fungal Effector Proteins. **Annual Review of Phytopathology**, v. 47, n. 1, p. 233–263, set. 2009.

STUKENBROCK, E. H.; DUTHEIL, J. Y. Comparing fungal genomes: Insight into functional and evolutionary processes. **Methods in Molecular Biology**, v. 835, p. 531–548, 2012.

Sugar: World Markets and Trade. USDA. Foreign Agricultural Service. Disponível em: <<https://www.fas.usda.gov/data/sugar-world-markets-and-trade>>. Acesso em: 3 jun. 2022

SUNDAR, A. et al. A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*. **Botany**, v. 53, p. 107–128, 2012.

TANIGUTI, L. M. et al. Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. **PLoS ONE**, v. 10, n. 6, 12 jun. 2015.

TIAN, Y. et al. Determination of heterozygosity for avirulence/virulence loci through sexual hybridization of *Puccinia striiformis* f. sp. tritici. **Frontiers of Agricultural Science and Engineering**. v. 4, n. 1, p. 48–48, 1 jan. 2017.

TENA, G. PTI and ETI are one. **Nature Plants**, v. 7, n. 12, p. 1527–1527, dez. 2021.

TEIXEIRA-SILVA, Natália de Sousa. **Functional analysis of candidate effector proteins during *Sporisorium scitamineum* x sugarcane interaction**. 2018. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2019.

TEW, T. L.; COBILL, R. M. Genetic Improvement of Sugarcane (*Saccharum* spp.) as an Energy Crop. In: VERMERRIS, W. **Genetic Improvement of Bioenergy Crops**. Nova Iorque: Springer, 2008. p. 273–294.

THIRUGNANASAMBANDAM, P. P.; HOANG, N. V.; HENRY, R. J. The Challenge of Analyzing the Sugarcane Genome. **Frontiers in Plant Science**, v. 9, p. 616, 14 maio 2018.

THUSHARI, A. N. W. S.; DE COSTA, D. M. Molecular and Genetic Variability of *Sporisorium scitamineum* (Sugarcane Smut Pathogen) in Sugarcane Plantations in Sri Lanka. **Sugar Tech**, 3 fev. 2023.

TOM et al. Incursion of sugarcane smut in commercial cane crops at Ramu, Papua New Guinea. **Proceedings of the Australian Society of Sugar Cane Technology**, v. 39, p. 377-384. 2017.

ULLMANN, L. et al. Seventeen Ustilaginaceae High-Quality Genome Sequences Allow Phylogenomic Analysis and Provide Insights into Secondary Metabolite Synthesis. **Journal of Fungi**, v. 8, n. 3, p. 269, 8 mar. 2022.

UNICA, **Observatório da Cana**. Disponível em: <<https://observatoriodacana.com.br>>. Acesso em 09 de fev. 2023.

VIEIRA, A. et al. Genome-Wide Signatures of Selection in *Colletotrichum kahawae* Reveal Candidate Genes Potentially Involved in Pathogenicity and Aggressiveness. **Frontiers in Microbiology**, v. 10, 19 jun. 2019.

WANG, Y. et al. MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. **Nucleic Acids Research**, v. 40, n. 7, p. e49–e49, 4 jan. 2012.

WEI, L. et al. Comparative genomics approaches to study organism similarities and differences. **Journal of Biomedical Informatics**, v. 35, n. 2, p. 142–150, abr. 2002.

WOLLENBERG, T.; SCHIRAWSKI, J. Comparative Genomics of Plant Fungal Pathogens: The *Ustilago-Sporisorium* Paradigm. **PLoS Pathogens**, v. 10, n. 7, p. e1004218, 3 jul. 2014.

WU, K. K.; HEINZ, D. J.; HOGARTH, D. M. Association and heritability of sugarcane smut resistance to races A and B in Hawaii. **Theoretical and Applied Genetics**, v. 75, n. 5, p. 754–760, maio 1988.

XU, J.-R. et al. The Dawn of Fungal Pathogen Genomics. **Annual Review of Phytopathology**, v. 44, n. 1, p. 337–366, 1 set. 2006.

YAN, M. et al. The mating-type locus b of the sugarcane smut *Sporisorium scitamineum* is essential for mating, filamentous growth and pathogenicity. **Fungal Genetics and Biology**, v. 86, p. 1–8, 1 jan. 2016.

YUAN, M. et al. PTI-ETI crosstalk: an integrative view of plant immunity. **Current Opinion in Plant Biology**, v. 62, p. 102030, ago. 2021.

ZHANG, T. et al. Complex genome assembly based on long-read sequencing. **Briefings in Bioinformatics**, v. 23, n. 5, p. 1-11, 20 set. 2022.

ZHENG, W. et al. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. **Nature Communications**, v. 4, n. 1, 23 out. 2013.

ZHENG, Z. et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. v. 2, n. 12, p. 797–803, 19 dez. 2022.

ZHU, G. et al. Identification and Functional Analysis of the Pheromone Response Factor Gene of *Sporisorium scitamineum*. **Frontiers in Microbiology**, v. 10, 10 set. 2019.

APÊNDICES

APÊNDICE A – SEQUENCIAMENTO DO ISOLADO SSC04 MAT-1

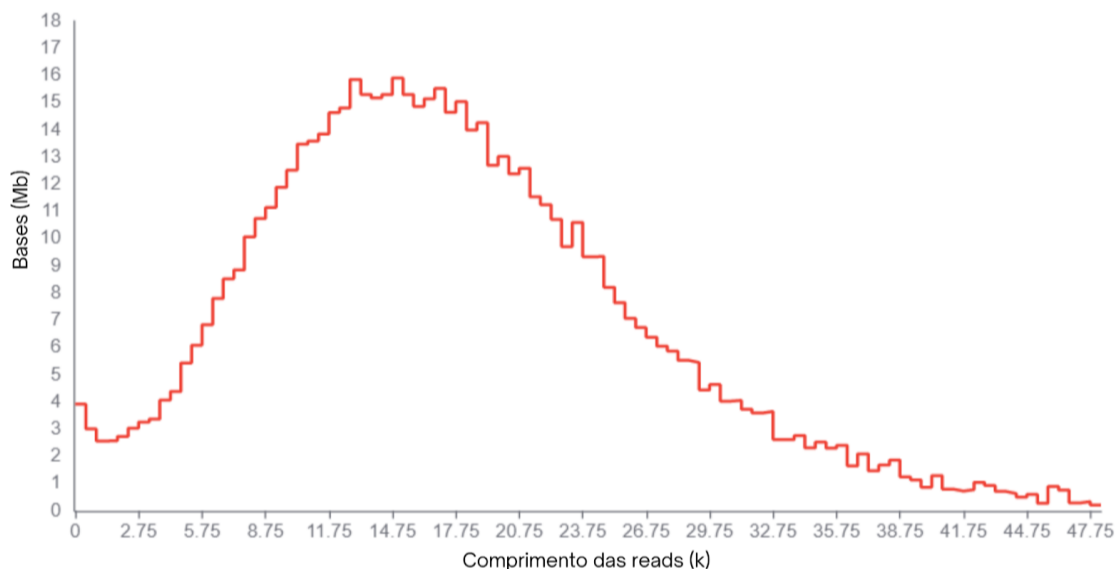


Figura A1 — Gráfico representando o comprimento das *reads* mostrando o número total de bases pelo comprimento da *read* no sequenciamento SSC04B, gerado pelo MinKNOW v22.12.7.

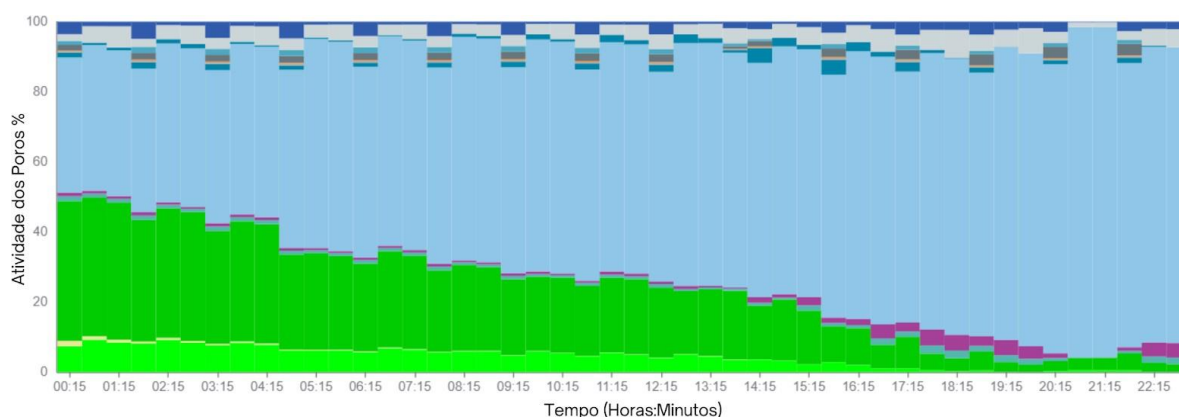


Figura A-2 — Gráfico de atividade dos poros do Oxford Nanopore MinION demonstrando a performance da amostra enquanto foi sequenciada durante a corrida do SSC04B. Verde claro representa a proporção de poros sequenciando, verde escuro são poros disponíveis para sequenciamento, azul claro representa nenhum poro, cinza claro representa canais desativados. Figura gerada pelo MinKNOW v22.12.7.

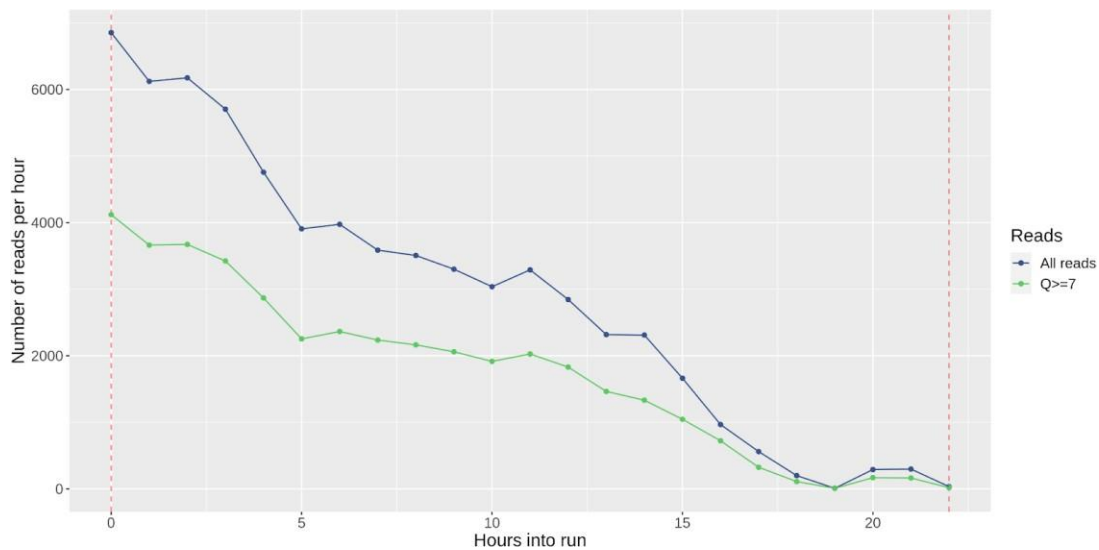


Figura A-3 — Gráfico demonstrando o número de *reads* (eixo y) sendo sequenciadas por hora durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

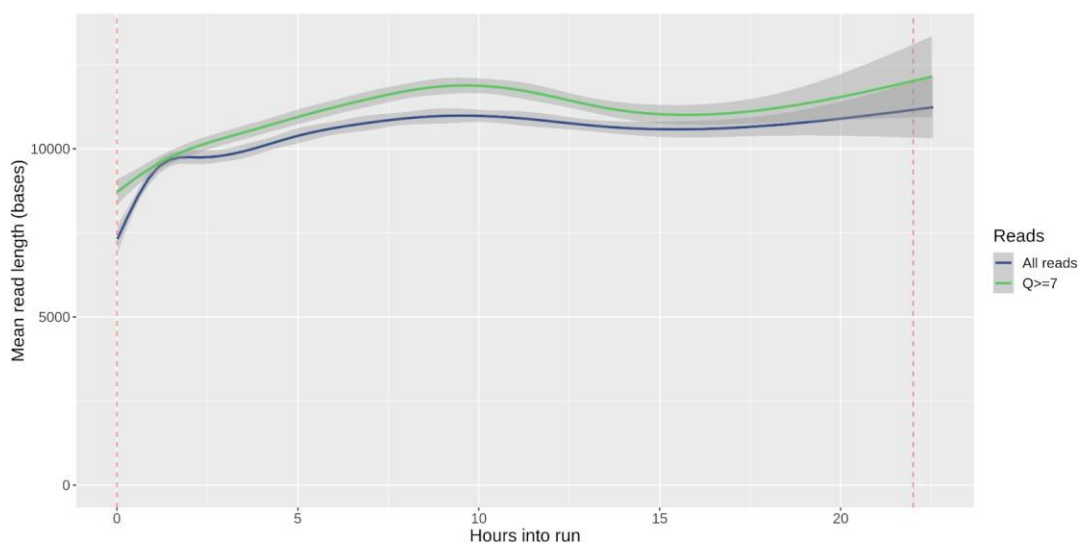


Figura A-4 — Gráfico demonstrando o comprimento médio das *reads* (eixo y) durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

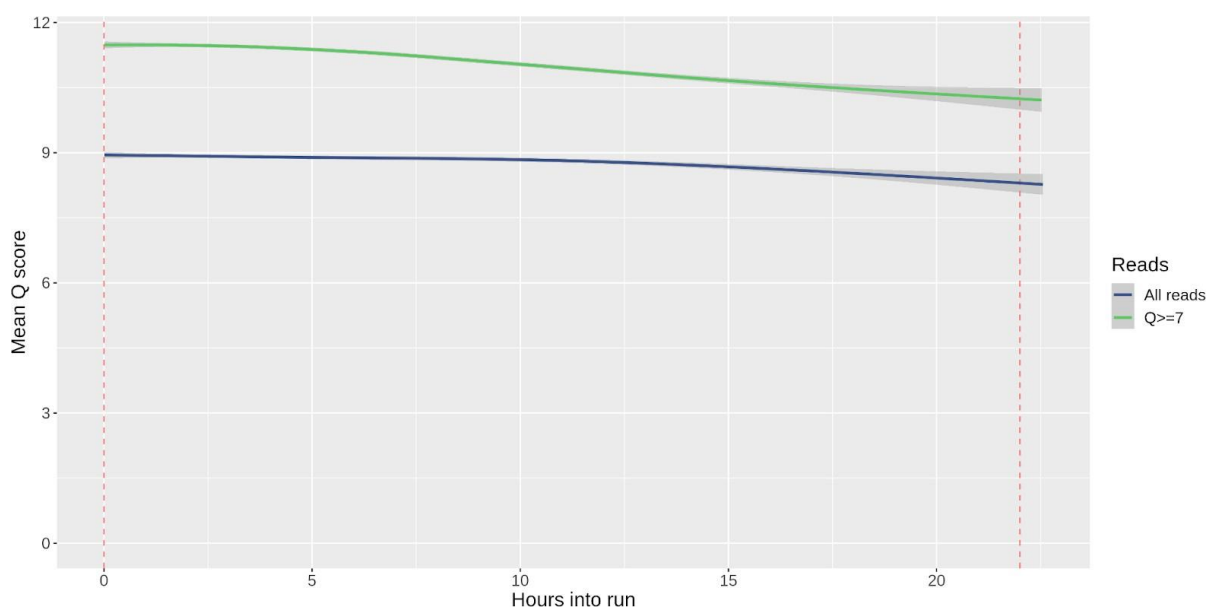


Figura A-5 — Gráfico demonstrando o score médio de qualidade das *reads* (eixo y) durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

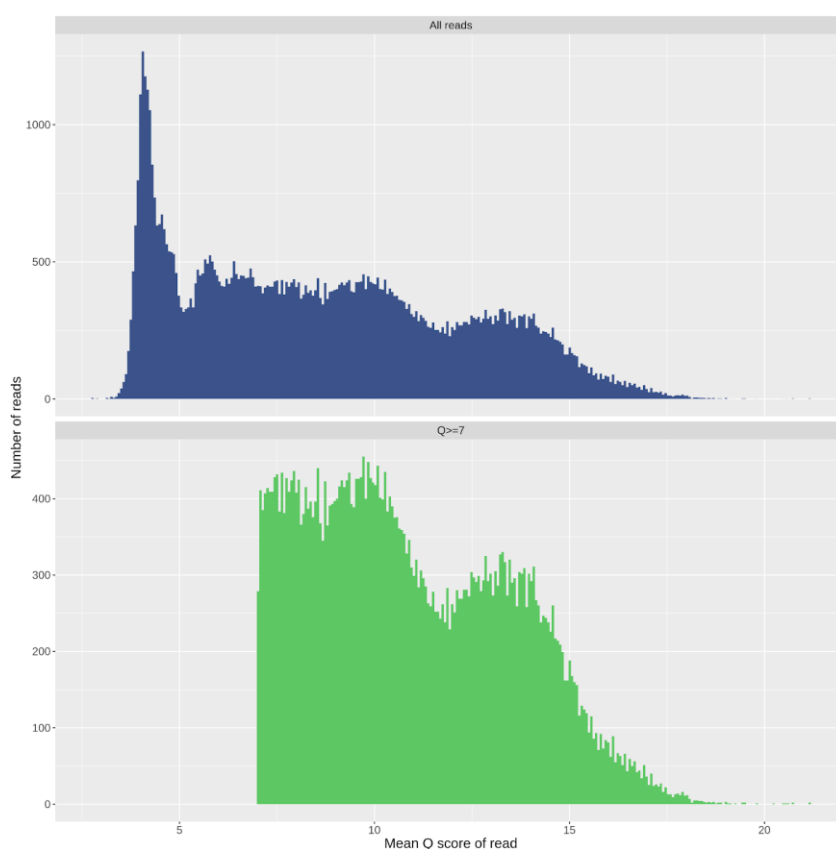


Figura A-6 — Gráfico demonstrando o número de *reads* em função do score médio de qualidade. Figura gerada pelo MinIONQC v1.4.2.

APÊNDICE B – SEQUENCIAMENTO DO ISOLADO SSC39 MAT-2

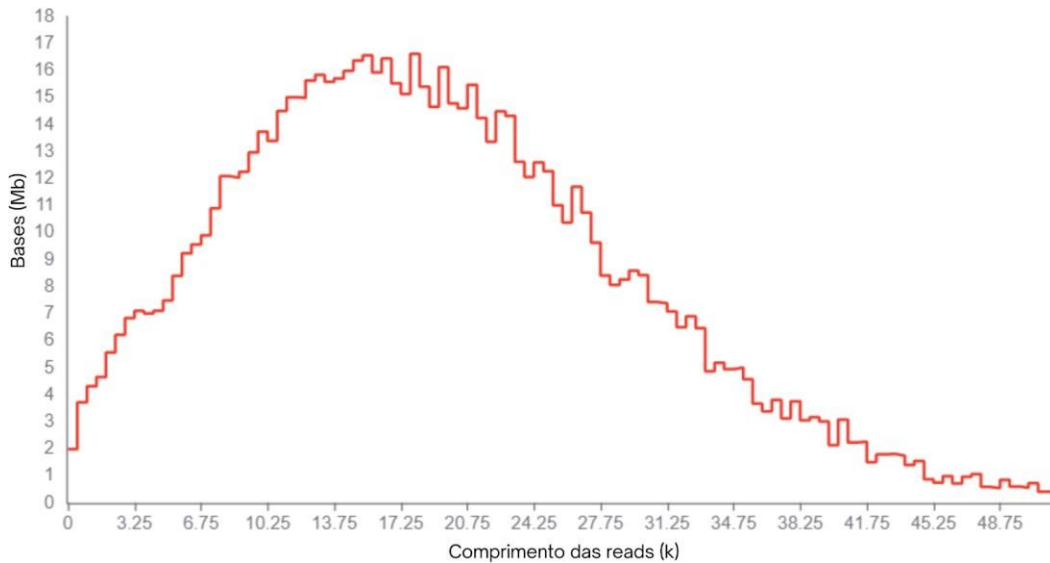


Figura B-1 — Gráfico representando o comprimento das *reads* mostrando o número total de bases pelo comprimento da read no sequenciamento SSC39A, gerado pelo MinKNOW v22.12.7.

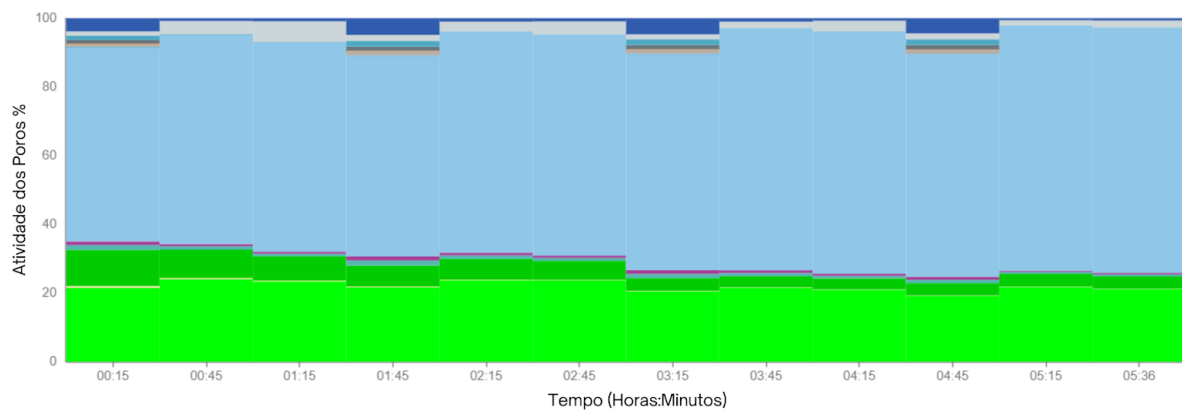


Figura B-2 — Gráfico de atividade dos poros do *Oxford Nanopore MinION* demonstrando a performance da amostra enquanto foi sequenciada durante a corrida do SSC39A. Verde claro representa a proporção de poros sequenciando, verde escuro são poros disponíveis para sequenciamento, azul claro representa nenhum poro, cinza claro representa canais desativados. Figura gerada pelo MinKNOW v22.12.7.

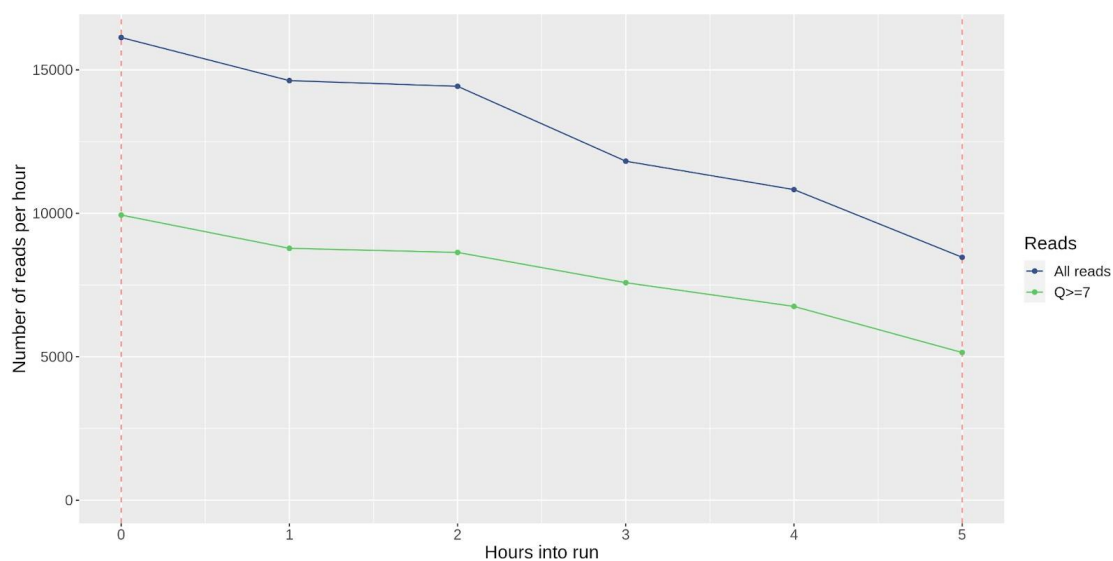


Figura B-3 — Gráfico demonstrando o número de *reads* (eixo y) sendo sequenciadas por hora durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

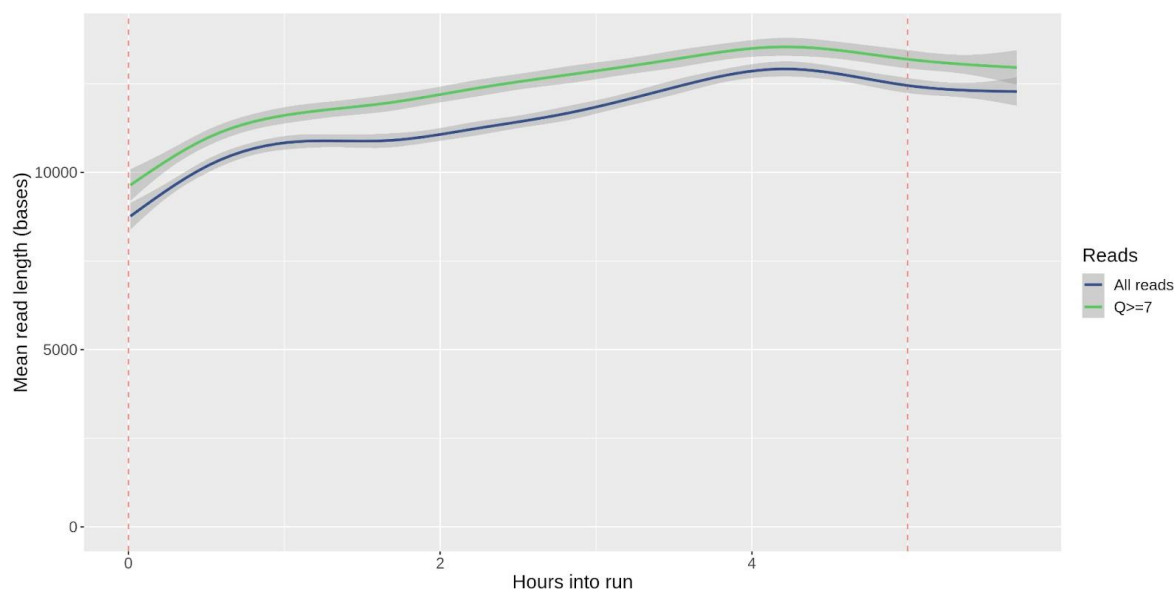


Figura B-4 — Gráfico demonstrando o comprimento médio das *reads* (eixo y) durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

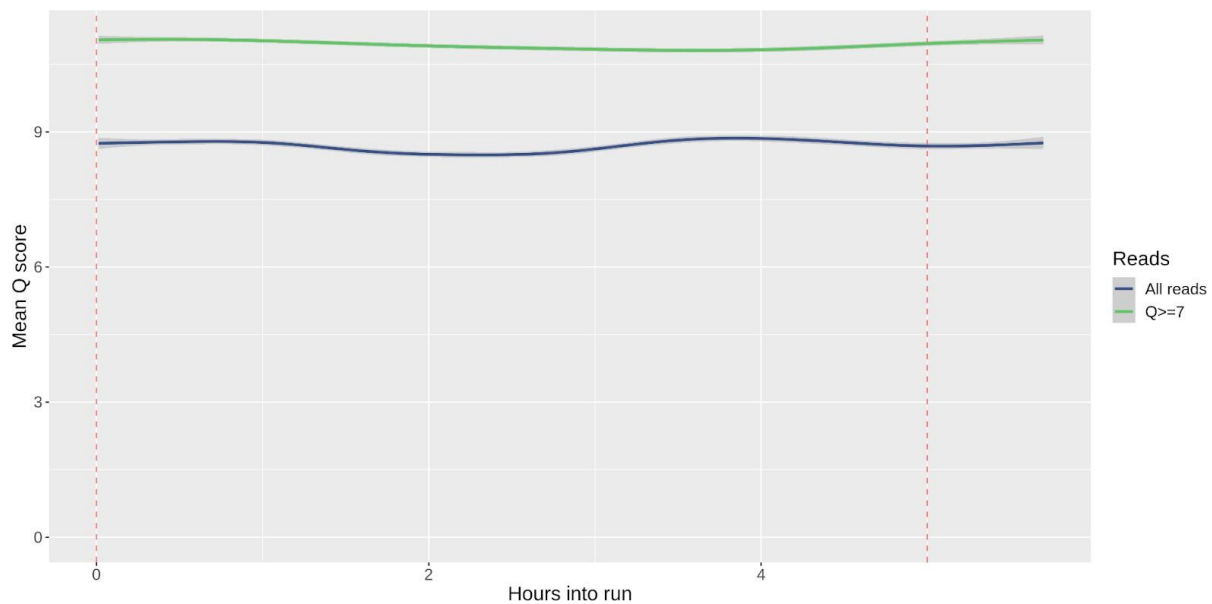


Figura B-5 — Gráfico demonstrando o score médio de qualidade das *reads* (eixo y) durante as horas da corrida (eixo x). Figura gerada pelo MinIONQC v1.4.2.

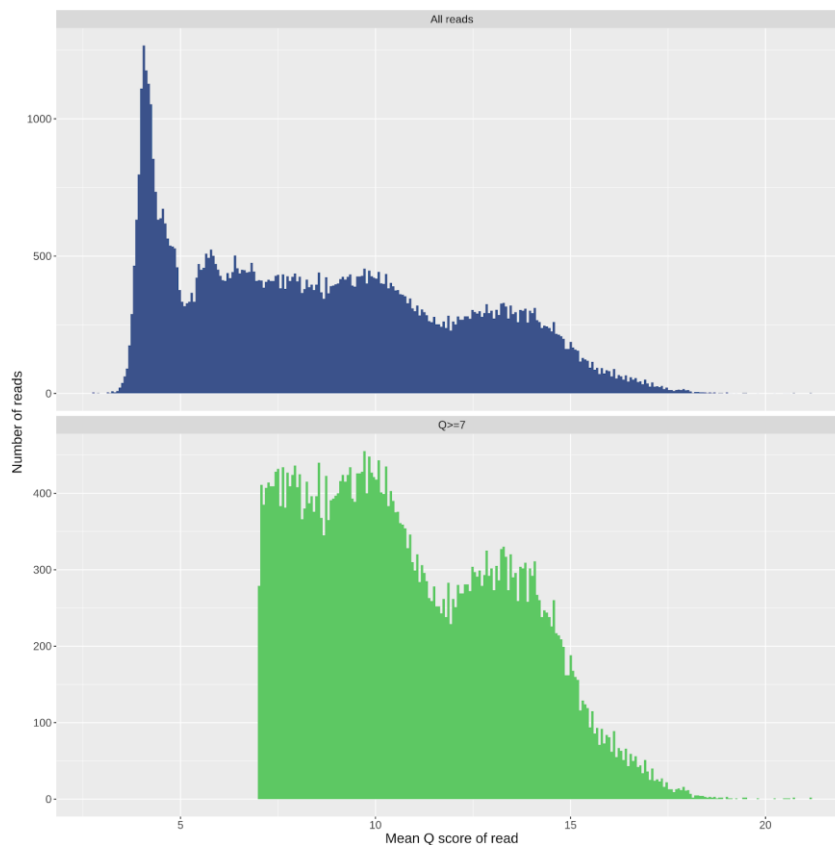


Figura B-6 — Gráfico demonstrando o número de *reads* em função do score médio de qualidade. Figura gerada pelo MinIONQC v1.4.2.