

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS AGRÁRIAS
CURSO DE AGRONOMIA

JOÃO PEDRO FERREIRA RIBEIRO

MAPEAMENTO DO SINAL TEMPORAL AO LONGO DE SEQUÊNCIAS DE DNA-A
DO *TOMATO SEVERE RUGOSE VIRUS*

Uberlândia - MG

2024

JOÃO PEDRO FERREIRA RIBEIRO

**MAPEAMENTO DO SINAL TEMPORAL AO LONGO DE SEQUÊNCIAS DE DNA-A
DO *TOMATO SEVERE RUGOSE VIRUS***

Trabalho de Conclusão de Curso submetido à
Universidade Federal de Uberlândia como
requisito para adquirir o Grau de Bacharel em
Engenharia Agrônômica.

Orientador: Prof. Alison Talis Martins Lima

Uberlândia - MG

2024

JOÃO PEDRO FERREIRA RIBEIRO

**MAPEAMENTO DO SINAL TEMPORAL AO LONGO DE SEQUÊNCIAS DE DNA-A
DO *TOMATO SEVERE RUGOSE VIRUS***

Trabalho de Conclusão de Curso submetido à
Universidade Federal de Uberlândia como
requisito para adquirir o Grau de Bacharel em
Engenharia Agrônômica.

Uberlândia, 24/05/2024

Banca Examinadora:

Prof. Dr. Alison Talis Martins Lima
(Orientador)

Prof. Dr. Flávio Tetsuo Sasaki
(Examinador)

Dr. Richard Costa Polveiro
(Examinador)

AGRADECIMENTOS

A Deus por minha vida e pelas inúmeras bênçãos.

Aos meus pais, Olyntho e Francielle, por todo o apoio material e emocional não só durante os anos de graduação, mas em toda a minha vida.

Aos colegas do grupo de pesquisa PROVIRUS, em especial à Meirielly pela parceria na execução deste trabalho.

Ao Professor Alison pela orientação, paciência, dedicação e ensinamentos proporcionados durante toda minha permanência no PROVIRUS.

À Universidade Federal de Uberlândia pela oportunidade.

À Lara, meu amor, pelo apoio incondicional durante os anos de graduação.

RESUMO

O gênero *Begomovirus* (família *Geminiviridae*) constitui um dos mais diversos grupos de vírus que infectam plantas. Eles são conhecidos por sua ampla gama de hospedeiros dentre plantas dicotiledôneas e são transmitidos por meio de um complexo de espécies crípticas de moscas-brancas referido como *Bemisia tabaci*. Populações de begomovírus evoluem rapidamente por meio dos mecanismos de mutação, recombinação e pseudorecombinação. O tomate severe rugose virus (ToSRV) causa grandes prejuízos na tomaticultura brasileira, sobretudo nas regiões central e sul do país. Dada a importância do ToSRV, este trabalho se propôs a analisar a estrutura temporal ao longo de suas sequências de DNA-A disponíveis no Genbank obtidas de isolados coletados entre os anos de 2000 a 2018. Além disso, objetivou-se quantificar o impacto de eventos de recombinação sobre a calibração do relógio molecular. Para isso, foi realizado um processo criterioso de cura dos dados para obtenção de informações sobre os isolados virais tais como, local de coleta, hospedeiro e data de amostragem. As análises de avaliação do sinal temporal foram conduzidas com base em dois conjuntos de dados primários: um primeiro incluindo-se todas as sequências de DNA-A, inclusive aquelas duplicadas e um segundo mantendo-se apenas um isolado representante de cada haplótipo viral. Além disso, os alinhamentos múltiplos de sequências também foram submetidos a uma etapa automatizada de refinamento. Os alinhamentos foram fatiados em janelas móveis de 100, 200 e 300 nucleotídeos (nt) de comprimento, com 10, 20 e 30 nt de passo, respectivamente. Análises de regressão linear raiz-ponta entre as distâncias genéticas e as datas de amostragem dos isolados virais foram conduzidas no programa TreeTime. No geral, observou-se pouco sinal temporal ao longo das sequências de DNA-A do ToSRV, exceto para dois segmentos dentro das regiões codificadoras da porção central da proteína CP e da porção C-terminal da proteína Rep. Concluiu-se que a submissão dos alinhamentos a uma etapa prévia de refinamento melhora a qualidade dos resultados e que a manutenção de somente um representante de cada conjunto de haplótipos invariavelmente levou aos mesmos resultados que daqueles alinhamentos com sequências duplicadas. Além disso, foi também observado que a ocorrência de eventos de recombinação afeta não somente a intensidade do sinal temporal como também o padrão de distribuição dos valores de R^2 calculados a partir das análises de regressão linear raiz-ponta ao longo de sequências do DNA-A do ToSRV.

Palavras-chave: Begomovírus, Filodinâmica, Recombinação

ABSTRACT

The genus *Begomovirus* (family *Geminiviridae*) constitutes one of the most diverse groups of plant viruses. They are known for their wide host range among dicotyledonous plants and are transmitted through a complex of cryptic species of whiteflies referred to as *Bemisia tabaci*. Populations of begomoviruses evolve rapidly through mechanisms of mutation, recombination, and pseudorecombination. The tomato severe rugose virus (ToSRV) causes significant damage to Brazilian tomato crops, especially in the central and southern regions of the country. Given the importance of ToSRV, this work aimed to analyze the temporal structure across their DNA-A sequences available from Genbank, obtained from viral isolates collected between 2000 and 2018. Additionally, the study aimed to quantify the impact of recombination events on the calibration of the molecular clock. To achieve this, a meticulous data curation process was carried out to obtain information on viral isolates such as collection location, host, and sampling date. Temporal signal evaluation analyses were conducted based on two primary datasets: the first including all DNA-A sequences, even duplicated ones, and the second maintaining only one representative isolate of each viral haplotype. Furthermore, multiple sequence alignments were also subjected to an automated refinement step. The alignments were sliced into sliding windows of 100, 200, and 300 nucleotides (nt) in length, with steps of 10, 20, and 30 nt, respectively. Root-to-tip linear regression analyses between the genetic distances and sampling dates of viral isolates were conducted using the TreeTime program. Overall, little temporal signal was observed throughout the DNA-A sequences of ToSRV, except for two segments within the coding regions of the central portion of the CP protein and the C-terminal portion of the Rep protein. It was concluded that subjecting the alignments to a preliminary refinement step improves the quality of results and that maintaining only one representative of each haplotype set invariably led to the same results as those alignments with duplicated sequences. Additionally, it was also observed that the occurrence of recombination events affects not only the intensity of the temporal signal but also the distribution pattern of R2 values calculated from root-to-tip linear regression analyses across DNA-A sequences of ToSRV.

Keywords: Begomovirus, Phylodynamics, Recombination

SUMÁRIO

1. INTRODUÇÃO	8
2. REVISÃO DE LITERATURA.....	11
3. MATERIAS E MÉTODOS	15
3.1. Elaboração dos conjuntos de dados de DNA-A de isolados da espécie <i>Tomato severe rugose virus</i>	15
3.2. Alinhamentos múltiplos de sequências	15
3.3. Análise de recombinação	16
3.4. Análise filogenética.....	17
3.5. Avaliação do sinal temporal por meio de análises de regressão linear raiz-ponta...	17
4. RESULTADOS	18
4.1. Sinal temporal em sequências completas de DNA-A de ToSRV	18
4.2. Validação do conteúdo de sinal temporal observado ao longo dos genomas virais	22
4.3. Efeito de eventos de recombinação sobre o sinal temporal	25
4.4. Validação do conteúdo de sinal temporal observado ao longo de sequências não recombinantes de DNA-A.....	29
5. DISCUSSÃO	34
6. CONCLUSÃO	37
7. REFERÊNCIAS.....	38
ANEXOS	45

1. INTRODUÇÃO

Os begomovírus (gênero *Begomovirus*, família *Geminiviridae*) constituem um dos mais diversos grupos de vírus que infectam plantas e possuem relevante impacto econômico na agricultura mundial. O begomovírus *Tomato yellow leaf curl virus* (TYLCV) é um dos principais patógenos na cultura do tomateiro (*Solanum lycopersicum* L.) em nível mundial, causando sérios impactos na produção em regiões tropicais e subtropicais (MABVAKURE et al., 2016). No Brasil, o mosaico dourado do feijoeiro causado pelo begomovírus *Bean golden mosaic virus* (BGMV) pode acarretar perdas de 40 a 100% na cultura do feijoeiro-comum (*Phaseolus vulgaris* L.), dependendo da incidência (FARIA; ARAGÃO, 2013). Os begomovírus são transmitidos por meio de um complexo de espécies crípticas de moscas-brancas, coletivamente conhecido como *Bemisia tabaci* (DE BARRO et al., 2011). Infecções por begomovírus podem induzir sintomas severos, como mosaico, amarelecimento, enrolamento foliar, manchas e nanismo. As plantas de interesse econômico afetadas pelos begomovírus pertencem ao grupo das dicotiledôneas, entretanto estudos recentes sugerem a possibilidade de que ocorram infecções por begomovírus em espécies de plantas monocotiledôneas (KIL et al., 2021).

Os begomovírus possuem genoma composto por uma ou duas moléculas de DNA de fita simples (ssDNA), sendo conhecidas como DNA-A e DNA-B, cada uma com cerca de 2600 nucleotídeos de comprimento. O DNA-A dos begomovírus bipartidos é homólogo ao genoma dos begomovírus monopartidos e codifica proteínas envolvidas na replicação (LAUFS et al., 1995), supressão de silenciamento genômico (BISARO, 2006), e encapsidação de progênie viral (HARRISON; SWANSON; FARGETTE, 2002). Em contrapartida, o DNA-B codifica proteínas essenciais para a movimentação do vírus dentro das plantas hospedeiras (SANDERFOOT; LAZAROWITZ, 1996). Uma infecção bem-sucedida por um begomovírus bipartido requer a presença de ambos os componentes genômicos (DNA-A e DNA-B) no hospedeiro (FIALLO-OLIVÉ et al., 2021).

Populações de begomovírus apresentam elevada variabilidade genética devido às altas taxas de substituição de nucleotídeos (DUFFY; HOLMES, 2008), ocorrência frequente de eventos de recombinação (LEFEUVRE; MORIONES, 2015) e pseudorecombinação (GILBERTSON et al., 1993). De acordo com estudos conduzidos por Ge et al. (2007), a espécie viral *Tomato yellow leaf curl China virus* (TYLCCNV) possui elevada frequência de mutação, similar àquelas estimadas para vírus de RNA de fita simples. Em outro estudo essa observação foi validada para o *Tomato severe rugose virus* (ToSRV) (PINTO et al., 2021).

O *Tomato severe rugose virus* (ToSRV) é um dos principais begomovírus presentes no Brasil e afeta a cultura do tomateiro (*Solanum lycopersicum* L.) causando grandes danos econômicos. De acordo com Mituti et al. (2019), o ToSRV é a principal espécie viral infectando tomateiros nos estados de Goiás, Minas Gerais e São Paulo. Os sintomas expressados pelas plantas de tomateiro infectadas por ToSRV são diversos, incluindo mosaico, encarquilhamento das folhas, amarelecimento, redução da área foliar, redução do crescimento das plantas (INOUE-NAGATA; ÁVILA; VILLAS-BOAS et al., 2009). A cultura do tomateiro é considerada de elevado investimento, visto que o custo por hectare (R\$/ha) pode ultrapassar R\$ 100.000,00, a depender da região (HORTIFRUTI BRASIL, 2017). De acordo com Inoue-Nagata et al. (2016), a alta incidência de ToSRV na cultura do tomateiro podem levar a perdas de até 100%.

Embora geralmente associado ao tomateiro, o ToSRV possui uma considerável gama de hospedeiros, tendo sido descrito infectando outras espécies cultivadas, como batata (*Solanum tuberosum* L.), berinjela (*Solanum melongena* L.), feijoeiro-comum (*Phaseolus vulgaris* L.), soja (*Glycine max* L.), pimentão (*Capsicum annum* L.); além de infectar também plantas daninhas, como *Nicandra physalodes* L., *Chenopodium album* L., e *Oxalis latifolia* Kunth. (BARBOSA et al., 2011; SILVA, 2021). Dentre esses hospedeiros secundários, o Joá-de-capote (*Nicandra physalodes* L.) se demonstrou, experimentalmente, uma ótima fonte de inóculo de ToSRV para *Bemisia tabaci* (BARBOSA et al., 2011).

Em 2017, estudo reportou infecção assintomática de soja (*Glycine max* L.) com ToSRV, fato preocupante visto à extensa presença da sojicultura no território brasileiro. Nesse trabalho, foram amostradas 301 plantas de soja assintomáticas em campos de produção próximos a lavouras de tomateiro, em seis municípios distintos nos estados de Goiás, Minas Gerais e São Paulo. Dentre as amostras, dez continham ToSRV (MACEDO et al., 2017). De acordo com Favara et al. (2023), a cultura da soja possui potencial para atuar como hospedeiro secundário para o ToSRV, e também possui a capacidade de amplificar a espécie viral, podendo adquirir importante papel na ocorrência de epidemias em lavouras de tomateiros. Um aspecto relevante apresentado no estudo é que a importância da soja como fonte de inóculo do ToSRV é dependente da cultivar em questão.

De acordo com a hipótese de Zuckerland e Pauling (1965), mudanças nas sequências de nucleotídeos dos organismos seguem uma taxa constante, que pode ser mensurada tornando possível a calibração de um relógio molecular – uma correlação entre divergência genética e tempo – com o qual se pode datar eventos evolutivos importantes de uma população como, por exemplo, o surgimento de um ancestral comum entre duas ou mais espécies. A hipótese de

Zuckerland e Pauling era de um relógio molecular estrito, que consiste em um modelo considerando um intervalo temporal constante entre a ocorrência de substituições de nucleotídeos, ou seja, uma taxa evolutiva constante para todos os ramos de uma árvore filogenética. Considerando a complexidade da dinâmica evolutiva dos organismos, Sanders (2002) propôs uma abordagem considerando a substituição de nucleotídeos como variável, em que cada linhagem possuía sua própria taxa de mutação, um relógio molecular relaxado. Desde então novos modelos estatísticos surgiram, aumentando a precisão das calibrações de relógios moleculares (FOURMENT; DARLING, 2018).

Dada a importância do ToSRV para a tomaticultura brasileira, objetivou-se neste estudo analisar a estrutura temporal do ToSRV, visando mapear a intensidade de sinal temporal ao longo de sequências de DNA-A de isolados coletados de 2000 a 2018 disponíveis no Genbank, e a quantificação da interferência de eventos de recombinação sobre a calibração do relógio molecular.

2. REVISÃO DE LITERATURA

O gênero *Begomovirus* é o mais numeroso e diversificado da família *Geminiviridae*, abrangendo um total de 445 espécies distintas, de acordo com o Comitê Internacional de Taxonomia de Vírus (ICTV) (FIALLO-OLIVÉ, et al., 2021). Begomovírus podem ser divididos em dois grupos de acordo com a origem geográfica e relações filogenéticas. O primeiro grupo é aquele composto por isolados virais nativos do ‘Velho Mundo’ (Europa, África, Ásia e Oceania), enquanto o segundo grupo inclui isolados begomovírus nativos do ‘Novo Mundo’ (Américas) (NAWAZ-UL-REHMAN; FAUQUET, 2009; MONDAL et al., 2019).

Os begomovírus podem apresentar genomas monopartidos, ou seja, compostos por uma única molécula de DNA de fita simples ou genomas bipartidos compostos por duas moléculas de DNA de fita simples, cada uma com aproximadamente 2600 nucleotídeos de comprimento (figura 1). Begomovírus do Novo Mundo apresentam predominantemente genomas bipartidos, enquanto os begomovírus do Velho Mundo apresentam genomas predominantemente monopartidos. Begomovírus monopartidos se assemelham ao componente DNA-A dos bipartidos, e estão frequentemente associados a moléculas satélites similares a vírus, conhecidas por satélites alpha e beta, que possuem papel importante na indução dos sintomas virais (GUPTA; REDDY; BHATTACHARYYA; CHAKRABORTY, 2021; PAPROTKA; METZLER; JESKE, 2010; FERRO et al., 2017).

O componente DNA-A dos begomovírus bipartidos tipicamente codifica de cinco a sete proteínas. A proteína Rep (Replication Protein) possui papel essencial na replicação viral, enquanto a CP (Coat Protein) está envolvida na montagem do capsídeo viral, facilitadora de transmissão pelo vetor e mediadora de movimento núcleo-citoplasma em vírus monopartidos (HARRISON; SWANSON; FARGETTE, 2002; PADIDAM; BEACHY; FAUQUET, 1996). A proteína V2 (ou AV2 em begomovírus bipartidos) atua como supressora de silenciamento gênico pós-transcricional (CHOWDA-REDDY et al., 2008). A proteína TrAP (Transcriptional Activator Protein) interfere com silenciamento gênico transcricional e pós-transcricional, atuando como um fator de transcrição necessário para a expressão das proteínas CP e NSP em begomovírus bipartidos (HANLEY-BOWDOIN et al., 1999; SUNTER; BISARO, 1997; HALEY et al., 1992; CHOWDA-REDDY, et al. 2009). A proteína REn (Replication Enhancer Protein) aumenta o acúmulo de DNA viral e recruta a DNA polimerase δ para a síntese de novas cópias do genoma de ssDNA dos geminivírus (WU et al., 2021; SUNTER; BISARO, 1997). A proteína C4 (ou AC4) atua como supressora do silenciamento de RNA (FONDONG et al., 2007; RAMACHANDRAN; PADMANABHAN; PITA; FAUQUET, 2004). No componente DNA-B, são codificadas proteínas responsáveis pelo movimento célula-a-célula (Nuclear Shuttle

Protein, NSP) e sistêmico (Movement Protein, MP), respectivamente (HANLEY-BOWDOIN; BEJARANO; ROBERTSON, 2013; ROJAS; HAGEN; LUCAS; GILBERTSON, 2005). Um estudo recente detectou pequenas ORFs adicionais, incluindo a V3, que atua como supressora do silenciamento gênico (GONG et al., 2021). Por último, a recém-descoberta proteína C7, codificada por isolados de tomato yellow leaf virus, possui papel relevante na infecção viral como um fator de patogenicidade, embora menos eficaz como supressora do silenciamento de RNA quando comparada a outras (LIU et al., 2023).

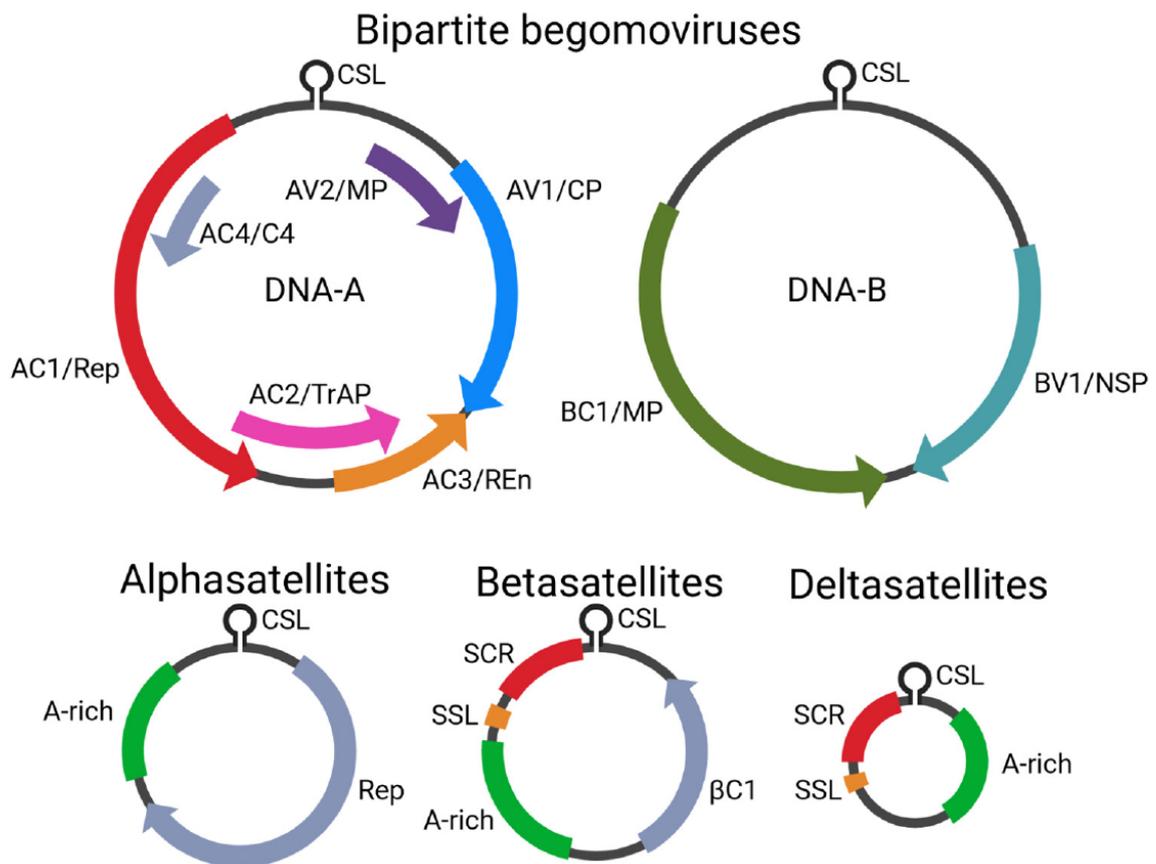


Figura 1. Representação esquemática de um genoma típico de begomovírus bipartidos e DNA satélites associados (alphasatellites, betasatellites e deltasatellites). O DNA-A de begomovírus bipartidos típicos codifica seis proteínas, sendo elas AV1/CP: capsid protein; AV2/MP: putative movement protein (ausente em begomovírus do Novo Mundo); AC1/Rep: replication-associated protein; AC2/TrAP: transcriptional activator protein; AC3/REn: replication enhancer protein; AC4/C4: C4 protein. O DNA-B codifica duas proteínas, sendo elas BV1/NSP: nuclear shuttling protein; BC1/MP: movement protein. Alphasatellites codificam uma proteína Rep, enquanto betasatellites codificam uma proteína β C1.

Fonte: Fiallo-Olivé; Navas-Castillo, 2023.

Em begomovírus bipartidos, ambos DNA-A e DNA-B compartilham segmentos similares dentro de suas regiões intergênicas, abrangendo aproximadamente 200 nucleotídeos

conhecidos por “região comum” (RC). A RC apresenta elementos de sequência envolvidos nos processos de replicação e transcrição do genoma viral. A sequência do nonanucleotídeo (‘TAATATTAC’) é mapeada dentro da RC, funcionando como ponto de clivagem para o DNA e ponto de início para o processo de replicação (FIALLO-OLIVÉ et al., 2021; BONNAMY; BLANC; MICHALAKIS, 2023).

Begomovírus são transmitidos de maneira persistente, não-propagativa, e circulativa, por um complexo de espécies crípticas de moscas-brancas, conhecido como *Bemisia tabaci* (figura 2). Begomovírus inicialmente estabelecem sua infecção no floema de plantas infectadas (NAVAS-CASTILLO; FIALLO-OLIVÉ; SÁNCHEZ-CAMPOS, 2011; GILBERTSON; BATUMAN; WEBSTER; ADKINS, 2015). Pesquisas recentes sugerem a possibilidade de que o tomato yellow leaf curl virus (TYLCV) poderia se replicar dentro do inseto-vetor, contudo permanece incerto se este fenômeno é exclusivo para o TYLCV ou se é aplicável a todos os begomovírus (HE et al., 2020).

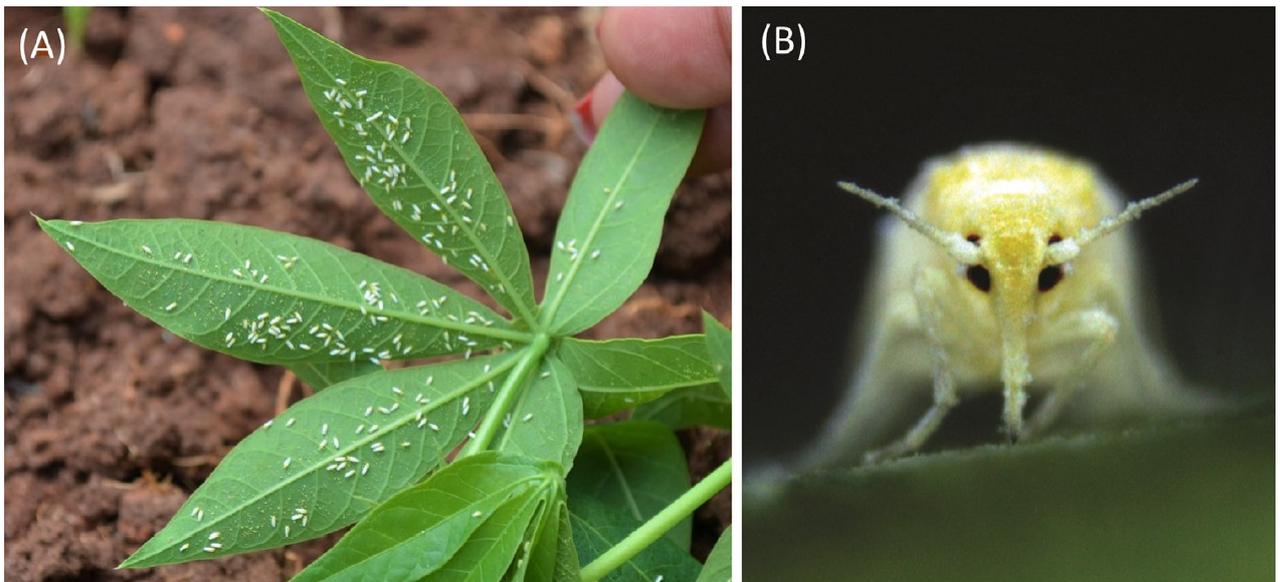


Figura 2. *Bemisia tabaci* sensu lato. (A) Indivíduos adultos de *B. tabaci* na parte abaxial de uma folha de mandioca. (B) Adulto de *B. tabaci* se alimentando em uma folha de tabaco.

Fonte: Fiallo-Olivé; Navas-Castillo, 2023.

A distribuição ampla do inseto-vetor tem apresentado papel importante no sucesso na distribuição de begomovírus em nível global. Notavelmente, os primeiros relatos de infecções por begomovírus em plantas cultivadas amostradas de alguns países coincidem com a distribuição mundial de espécies de *Bemisia tabaci* como a Middle East-Asia Minor 1 (MEAM1) e a Mediterranean (MED), que introduziram estes vírus em áreas anteriormente não

afetadas, incluindo o Brasil, que atualmente é reconhecido como um centro de diversidade de begomovírus (NAVAS-CASTILLO; FIALLO-OLIVÉ; SÁNCHEZ-CAMPOS, 2011; JONES 2003; RIBEIRO et al., 1998). Foi demonstrado que os begomovírus possuem a habilidade de manipular preferências e comportamento alimentar das moscas-brancas. Moscas-brancas avirulentas tendem a preferir plantas infectadas por vírus, enquanto moscas-brancas virulentas demonstram maior propensão a se alimentar de plantas não infectadas (ZHAO; LIU; WANG; YANG; PAN, 2022).

Begomovírus possuem uma capacidade notável para rápida evolução por meio dos mecanismos de mutação, pseudorecombinação e recombinação (LEFEUVRE; MORIONES, 2015; ROJAS; HAGEN; LUCAS; GILBERTSON, 2005; PITA et al., 2001; LIMA et al., 2013). Mutações são alterações no material genético dos organismos, incluindo em genomas virais, e podem ocorrer devido a falhas durante o processo de replicação do DNA. As elevadas taxas de substituição de nucleotídeos em genomas de begomovírus são similares àquelas observadas em vírus de RNA, e contribuem significativamente para a variabilidade genética observada em suas populações (ACOSTA-LEAL, DUFFY, XIONG, HAMMOND ELENA, 2011; LIMA et al., 2017; DUFFY; HOLMES, 2009).

O processo de recombinação envolve a troca de segmentos de DNA ou RNA entre vírus, um processo que aumenta significativamente a diversidade genética e o potencial adaptativo de uma população viral (LEFEUVRE; MORIONES, 2015; PADIDAM; SAWYER; FAUQUET, 1999). No contexto da agricultura, o estudo da recombinação é particularmente relevante, pois pode gerar novos genomas virais capazes de suplantar a resistência genética das plantas. Vírus recombinantes podem superar mecanismos de resistência e ameaçar a viabilidade agrônômica das culturas. Deste modo, compreender e monitorar eventos de recombinação em populações virais é importante no manejo de doenças virais (AWADALLA, 2003; SATTAR; KVARNHEDEN; SAEED; BRIDDON, 2013; MONCI; SÁNCHEZ-CAMPOS; NAVAS-CASTILLO; MORIONES, 2002).

Outro mecanismo que contribui significativamente para a diversidade genética em begomovírus é a pseudorecombinação, também conhecida como rearranjo. A pseudorecombinação envolve a troca de componentes genômicos completos entre vírus, tipicamente entre indivíduos de uma mesma espécie. Esse processo pode resultar na formação de vírus híbridos, especialmente quando existe um elevado grau de compatibilidade genética, particularmente na RC desses vírus (ANDRADE et al., 2006; GILBERTSON et al., 1993; PAPLOMATAS; PATEL; YU-MING HOU; NOUEIRY; GILBERTSON, 1994; PITA et al., 2001; SUNG; COUTTS, 1995).

3. MATERIAS E MÉTODOS

3.1. Elaboração dos conjuntos de dados de DNA-A de isolados da espécie *Tomato severe rugose virus*

O conjunto de dados utilizado neste trabalho foi composto por sequências de DNA-A de isolados virais pertencentes à espécie *Tomato severe rugose virus* (ToSRV), disponíveis no GenBank até a data de 7 de março de 2022. Realizou-se o processo de cura dos dados, efetuando-se a conferência manual juntamente à literatura associada às sequências e as informações disponíveis no GenBank, visando eliminar sequências com dados incompletos sobre local, data de coleta do isolado viral e hospedeiro. De acordo com esses critérios, foram mantidas no conjunto de dados final um total de 172 sequências de nucleotídeos completas de DNA-A.

3.2. Alinhamentos múltiplos de sequências

Um alinhamento múltiplo inicial foi construído utilizando-se o programa Muscle5 (EDGAR, 2022) composto por todas as sequências completas de DNA-A do ToSRV que passaram pelos filtros de cura listados no item 3.1. Esse alinhamento foi posteriormente submetido a uma etapa de identificação de sequências redundantes (cada conjunto de sequências redundantes representando um dado haplótipo viral) utilizando-se um script customizado escrito na linguagem de programação Python3. Um novo alinhamento múltiplo foi obtido a partir do alinhamento inicial por meio da remoção das sequências duplicadas, mantendo-se apenas uma sequência representante de cada haplótipo viral.

Posteriormente, os alinhamentos com e sem sequências duplicadas foram submetidos a uma etapa de refinamento para remoção de colunas compostas majoritariamente por “gaps” utilizando-se o programa Clipkit (STEENWYK et al., 2020). Foi utilizada a opção de refinamento “*smart-gap*” que aplica um limite dinâmico de porcentagem de “gaps” para remoção de colunas do alinhamento calculado com base em trechos do alinhamento original ao invés de empregar um único limite baseado no alinhamento completo. Dessa forma, o processo de refinamento envolveu principalmente a remoção de colunas em regiões problemáticas do alinhamento, o que preveniu a remoção demasiada de colunas de regiões bem alinhadas e que poderiam conter informações filogenéticas relevantes.

Em síntese, as análises subsequentes foram realizadas empregando-se quatro tipos básicos de alinhamentos múltiplos de sequências: (i) composto por todas as sequências de DNA-A do ToSRV, ou seja, com sequências duplicadas, e sem etapa posterior de refinamento (referido ao longo desse estudo como “172N”, indicando ser formado por todas as 172

sequências de DNA-A e “N” indicando não ter sido submetido a etapa de refinamento), (ii) composto por todas as sequências de DNA-A do ToSRV e submetido a uma etapa posterior de refinamento (referido como “172R”, onde “R” indica que o conjunto de dados foi submetido a uma etapa de refinamento), (iii) composto por apenas uma sequência representante de cada haplótipo, ou seja, com remoção de duplicatas, e sem etapa posterior de refinamento (referido como “154N”, onde “154” representa o número de sequências mantidas no conjunto de dados após a remoção das duplicatas), e (iv) composto por apenas uma sequência representante de cada haplótipo e submetido a uma etapa posterior de refinamento (“154R”).

Em seguida, outro script customizado escrito na linguagem de programação Python3 foi utilizado para fatiar os quatro alinhamentos múltiplos listados acima em subalinhamentos (segmentos dos alinhamentos completos que serão designados de agora em diante como janelas móveis) de 100, 200 e 300 nucleotídeos de comprimento, com respectivamente 10, 20 e 30 nucleotídeos de tamanho de passo (distância entre os sítios de início de janelas móveis adjacentes).

Foram geradas 762 janelas móveis (415 janelas móveis de 100 nucleotídeos, 208 janelas móveis de 200 nucleotídeos, e 139 janelas móveis de 300 nucleotídeos) para o alinhamento composto por todas as sequências de DNA-A e para o alinhamento no qual se manteve somente um representante para cada conjunto de haplótipos. Para o alinhamento composto por todas as sequências de DNA-A submetido a refinamento foram geradas 605 janelas móveis (330 janelas móveis de 100 nucleotídeos, 165 janelas móveis de 200 nucleotídeos, e 110 janelas móveis de 300 nucleotídeos). Para o alinhamento no qual se manteve somente um representante para cada conjunto de haplótipos e subsequentemente submetido a refinamento, foram geradas 565 janelas móveis (308 janelas móveis de 100 nucleotídeos, 154 janelas móveis de 200 nucleotídeos, e 103 janelas móveis de 300 nucleotídeos).

3.3. Análise de recombinação

Os quatro alinhamentos múltiplos baseados em sequências completas de DNA-A do ToSRV foram submetidos a uma varredura visando identificar sequências recombinantes utilizando-se o programa RDP4 (MARTIN et al., 2015). Sequências nas quais foram identificados eventos de recombinação detectados por quatro ou mais dos sete métodos analíticos disponíveis no programa (RDP, Geneconv, Bootscan, Maximum Chi Square, Chimaera, Sister Scan, e 3Seq) foram removidas dos alinhamentos originais, gerando assim um segundo alinhamento (referido ao longo desse estudo pelo sufixo “4+”, indicando a remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos). Foi

também obtido um terceiro alinhamento múltiplo a partir do inicial no qual foi aplicada uma abordagem ainda mais conservadora de remoção de recombinantes. Nesse último alinhamento, foram removidas todas as sequências contendo eventos de recombinação, independentemente do número de métodos de detecção (referido nesse estudo pelo sufixo “1+”, simbolizando a remoção de sequências identificadas como recombinantes por ao menos um método analítico). Ou seja, nesse último alinhamento, toda e qualquer sequência de DNA-A com evidência detectável de recombinação foi removida do alinhamento original.

3.4. Análise filogenética

Árvores de máxima verossimilhança foram construídas utilizando-se o programa IQ-Tree (NGUYEN et al., 2015). Os modelos de substituição de nucleotídeos que melhor se ajustam aos conjuntos de dados (alinhamentos de sequências completas de DNA e subalinhamentos) foram determinados utilizando-se o programa ModelFinder (KALYAANAMOORTHY et al., 2017), implementado no IQ-Tree.

3.5. Avaliação do sinal temporal por meio de análises de regressão linear raiz-ponta

As árvores filogenéticas geradas no IQ-Tree foram submetidas à análise de regressão linear raiz-ponta no programa TreeTime (SAGULENKO; PULLER; NEHER, 2018) com base no número de substituições de nucleotídeos e na escala temporal separando cada data de coleta dos isolados virais dos quais foram obtidas as sequências de DNA-A. Os resultados foram apresentados por meio de mapas de calor construídos utilizando-se o pacote R ComplexHeatmap (GU; EILS; SCHLESNER, 2016). A fim de se verificar a robustez do sinal temporal detectado a partir de cada subalinhamento, os alinhamentos baseados em sequências completas de DNA-A foram subdivididos em 2 novos alinhamentos; um primeiro composto por todas as colunas do alinhamento original associadas a janelas móveis que resultaram em valores de $R^2 > 0,3$ e um segundo composto por todas as colunas do alinhamento original associadas a janelas móveis que resultaram em valores de R^2 menores que 0,3. A subdivisão dos alinhamentos foi realizada através do programa trimAL (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009). Árvores filogenéticas foram construídas a partir de cada um desses novos alinhamentos e submetidas a uma segunda análise de regressão linear raiz-ponta no programa TreeTime.

4. RESULTADOS

4.1. Sinal temporal em sequências completas de DNA-A de ToSRV

Esse estudo envolveu análises conduzidas sobre um conjunto de dados contendo 172 sequências completas de DNA-A de isolados de begomovírus pertencentes à espécie *Tomato severe rugose virus* (ToSRV). Todas as sequências foram obtidas de isolados amostrados no Brasil, sendo que 14 sequências não possuíam informação sobre o estado brasileiro de amostragem, as demais foram coletadas nos seguintes estados: Espírito Santo (13), Goiás (13), Minas Gerais (117), São Paulo (7) e Distrito Federal (7). O isolado mais antigo foi coletado no ano de 2000 em Minas Gerais (#GenBank: AY029750), enquanto o mais recente foi coletado em março de 2018 no Espírito Santo (#GenBank: MN156310). Das 172 sequências, 114 foram obtidas de isolados coletados entre os anos de 2010 e 2018. Por meio do uso de um script customizado escrito na linguagem de programação Python, foi observado que o conjunto de dados é composto por 154 haplótipos de DNA-A. Um novo alinhamento contendo apenas um representante de cada haplótipo foi produzido utilizando-se o mesmo script.

As análises para avaliação do sinal temporal ao longo dos genomas do ToSRV foram baseadas em quatro tipos básicos de alinhamentos múltiplos de sequências: um primeiro composto por todas as sequências de DNA-A do ToSRV e sem refinamento (designado aqui como 172N), o segundo composto por todas as sequências de DNA-A do ToSRV e submetido a uma etapa posterior de refinamento (172R), o terceiro composto por apenas uma sequência representante de cada haplótipo e sem refinamento (154N), e o quarto composto por apenas uma sequência representante de cada haplótipo e submetido a uma etapa posterior de refinamento (154R). Os quatro alinhamentos foram fatiados em janelas móveis com comprimentos de 100, 200 e 300 nucleotídeos (nt) e tamanhos de passo de 10, 20 e 30 nt, respectivamente.

Mapas de calor foram gerados a partir dos valores de R^2 calculados a partir da regressão raiz-ponta utilizando-se árvores de máxima verossimilhança. Os mapas obtidos a partir das janelas móveis de 100 nt de comprimento (figura 3) indicaram uma ampla variação nos valores de R^2 ao longo de todo o DNA-A do ToSRV, com janelas apresentando desde valores de $R^2 = 0$, indicando ausência total de sinal temporal, até $R^2 = 1$, sugerindo elevado conteúdo de sinal temporal.

Oito janelas móveis com valores de $R^2 = 1$ foram observadas no alinhamento 172N, três no alinhamento 172R, quatro no alinhamento 154N, e quatro no alinhamento 154R. A partir da inspeção manual de cada um dos alinhamentos, observou-se que as janelas móveis com $R^2 = 1$ se sobrepunham a regiões genômicas com elevado número de colunas majoritariamente

compostas por gaps, portanto, esses resultados mais provavelmente tratam-se de artefatos devido à má qualidade dos alinhamentos nessas regiões.

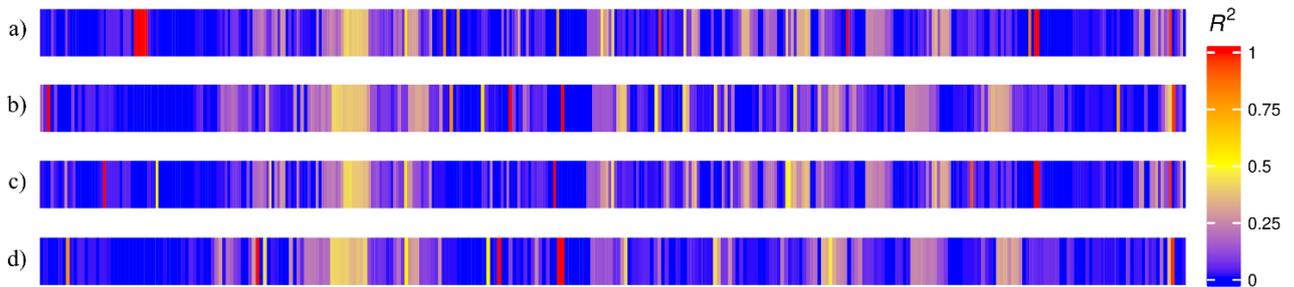


Figura 3. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de 100 nucleotídeos. **(a)** alinhamento composto por todas as sequências de DNA-A e sem etapa de refinamento (172N); **(b)** alinhamento composto por todas as sequências de DNA-A e submetido a uma etapa de refinamento (172R); **(c)** alinhamento composto por apenas uma sequência representante de cada haplótipo e sem etapa de refinamento (154N); **(d)** alinhamento composto por apenas uma sequência representante de cada haplótipo e submetido a etapa de refinamento (154R).

Por outro lado, analisando-se o heatmap obtido a partir do alinhamento 172N (figura 3a) é possível observar na região codificadora da porção central da proteína CP um total de nove janelas móveis com valores de $R^2 \sim 0,4$. Essas janelas móveis foram posicionadas ao redor dos pontos médios de 1101 a 1181 e apresentaram valores de R^2 variando de 0,36 até 0,42. Resultados similares a esses foram observados no alinhamento 172R (10 janelas; figura 3b), alinhamento 154N (9 janelas; figura 3c), e no alinhamento 154R (10 janelas; figura 3d). Foram também observadas outras janelas móveis com valores de $R^2 > 0,3$ ao longo do restante do DNA-A em todos os alinhamentos, contudo essas não se apresentaram agrupadas, distribuindo-se de forma alternada com outras janelas apresentando valores baixos de R^2 .

Os padrões de distribuição de R^2 observados a partir das janelas de 100 nt de comprimento entre os alinhamentos compostos por 172 sequências de DNA-A (figuras 3a e 3b) e aqueles compostos por 154 sequências (figuras 3c e 3d) foram similares, o que indica que os conjuntos de dados menores retiveram o mesmo conteúdo de sinal temporal quando comparados aos conjuntos de dados contendo duplicatas.

Frações proporcionalmente maiores de janelas móveis com valores $R^2 > 0,3$ foram observadas a partir dos alinhamentos refinados: 172R (37 de 330 janelas; 11,21%) versus 172N (41 de 415 janelas; 9,87%) e 154R (30 de 308 janelas; 9,74%) versus 154N (36 das 415 janelas, 8,67%). Os padrões de distribuição dos valores de R^2 entre os alinhamentos não refinados (figuras 3a e 3c) e aqueles refinados (figuras 3b e 3d) foram similares, exceto pelo fato de que

o agrupamento de janelas móveis com valores de $R^2 \sim 0,4$ na região codificadora da porção central da proteína CP incluíram um maior número de janelas nos alinhamentos refinados (9 janelas móveis nos alinhamentos 172N e 154N contra 10 janelas móveis para os alinhamentos 172R e 154N, respectivamente). Desconsiderando-se as janelas móveis com valores de $R^2 = 1$, os maiores valores de R^2 foram observados na região da sequência codificadora da proteína CP em todos os alinhamentos.

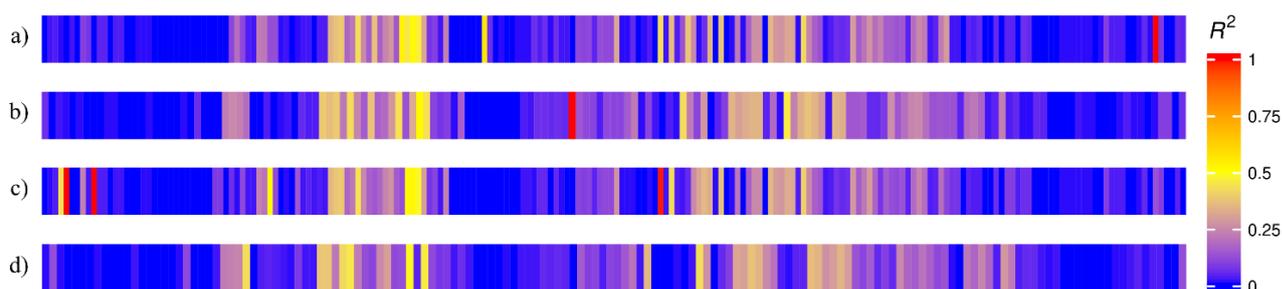


Figura 4. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de 200 nucleotídeos. **(a)** alinhamento composto por todas as sequências de DNA-A e sem etapa de refinamento (172N); **(b)** alinhamento composto por todas as sequências de DNA-A e submetido a uma etapa de refinamento (172R); **(c)** alinhamento composto por apenas uma sequência representante de cada haplótipo e sem etapa de refinamento (154N); **(d)** alinhamento composto por apenas uma sequência representante de cada haplótipo e submetido a etapa de refinamento (154R).

Similar aos mapas de calor construídos a partir das análises envolvendo janelas de 100 nt de comprimento, aqueles construídos a partir das janelas de 200 nt também indicaram ampla variação nos valores de R^2 (de $R^2 = 0$ até 1) (figura 4). Uma janela móvel com valor de $R^2 = 1$ foi observada no alinhamento 172N (figura 4a), uma no alinhamento 172R (Figura 4b) e duas no alinhamento 154N (figura 4c). No alinhamento 154R (figura 4d) não foram observadas janelas móveis com valores de $R^2 = 1$. A partir de uma inspeção manual de cada um dos alinhamentos, observou-se que as janelas móveis com $R^2 = 1$ também foram associadas a regiões genômicas com grande número de colunas majoritariamente compostas por gaps.

O maior conteúdo de sinal temporal observado em janelas adjacentes da região codificadora da porção central da proteína capsidial (CP) foi também observado nas análises baseadas em janelas móveis de 200 nt. Entretanto, o número de janelas móveis adjacentes com valores de $R^2 > 0,3$ foi menor (variando de 3 a 5 janelas móveis). Analisando o alinhamento 172N, o maior valor de R^2 nessa região foi de 0,53 (na janela com ponto médio no nt 1301), $R^2 = 0,51$ para o alinhamento 172R (ponto médio no nt 1081), $R^2 = 0,52$ para o alinhamento 154N

(ponto médio no nt 1341) e $R^2 = 0,49$ para o alinhamento 154R (ponto médio no nt 981). Similar aos alinhamentos fatiados em janelas de 100 nt, é possível se observar janelas móveis de 200 nt de comprimento com valores de R^2 acima de 0,3 dispersas ao longo da maior parte dos alinhamentos, entretanto estas alternam-se com janelas apresentando baixa evidência de sinal temporal.

Os alinhamentos contendo uma única sequência representante de cada conjunto de haplótipos (154N e 154R) renderam resultados similares àqueles dos alinhamentos contendo duplicatas em termos de distribuição de valores de R^2 . Novamente, esses resultados sugerem que a remoção das duplicatas não afeta os resultados, indicando que análises visando avaliar o sinal temporal em sequências de DNA-A do ToSRV podem ser conduzidas com base em conjuntos de dados menores, o que reduz consideravelmente a carga computacional sem qualquer prejuízo aos resultados obtidos.

A fração de janelas móveis com valores de $R^2 > 0,3$ foi maior no alinhamento 172R (21/165; 12,7%) quando comparado ao alinhamento 172N (22/208; 10,6%), reforçando os resultados obtidos previamente. Por outro lado, houve uma pequena redução no número de janelas móveis com valores de $R^2 > 0,3$ nos alinhamentos contendo 154 sequências (154R: 15/154; 9,7% versus 154N: 21/208; 10,1%). Excluindo-se as janelas móveis apresentando valores de $R^2 = 1$, os maiores valores de R^2 foram observados na região codificadora da proteína CP em todos os alinhamentos.

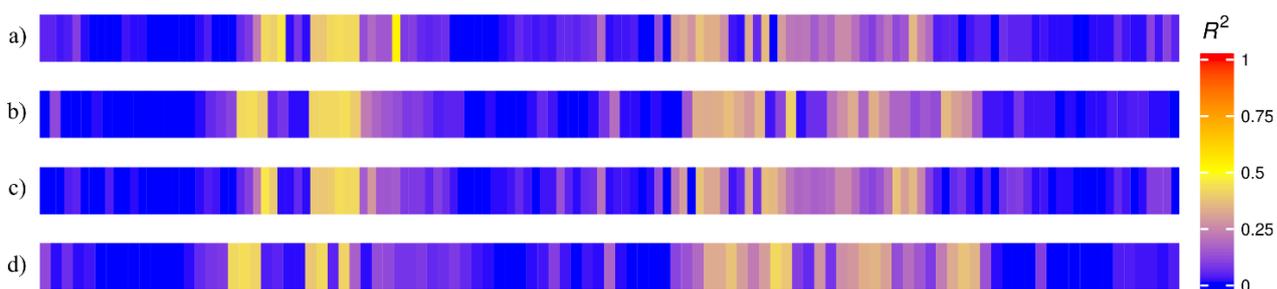


Figura 5. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de 300 nucleotídeos. **(a)** alinhamento composto por todas as sequências de DNA-A e sem etapa de refinamento (172N); **(b)** alinhamento composto por todas as sequências de DNA-A e submetido a uma etapa de refinamento (172R); **(c)** alinhamento composto por apenas uma sequência representante de cada haplótipo e sem etapa de refinamento (154N); **(d)** alinhamento composto por apenas uma sequência representante de cada haplótipo e submetido a etapa de refinamento (154R).

Divergindo-se dos resultados observados nos mapas de calor construídos para janelas de 100 e 200 nt, não foram observadas janelas com valores de $R^2 = 1$ nos mapas de calor

construídos a partir de janelas de 300 nt (figura 5). Além disso, também foi observada a presença de janelas móveis adjacentes na região codificadora da porção central da proteína CP com valores de $R^2 \sim 0,4$. Para o alinhamento 172N, seis janelas compuseram esse conjunto, cinco janelas para 172R, cinco janelas para 154N, e três janelas móveis para o alinhamento 154R. O valor de R^2 destas janelas móveis variou de 0,38 a 0,43. Mesmo se tratando de conjuntos com menor número de janelas móveis totais, os alinhamentos fatiados em janelas de 300 nt apresentaram mais janelas móveis nessa região quando comparados os alinhamentos fatiados em janelas de 200 nt.

Os resultados observados nos mapas de calor construídos para os alinhamentos não refinados fatiados em janelas de 300 nt (172N, Figura 5a; 154N, Figura 5c) se apresentaram similares entre si. O mesmo foi observado para os mapas de calor construídos para os alinhamentos refinados (172R, Figura 5b; 154R, Figura 5d), indicando que a manutenção de apenas uma sequência para os conjuntos de haplótipos gera resultados similares àqueles obtidos a partir do conjunto de dados contendo duplicatas. Como citado anteriormente, essa prática possibilita trabalhar com conjuntos de dados menores, gerando menor carga computacional e maior rapidez na obtenção de resultados.

Comparando-se a fração de janelas móveis de 300 nt com valores de $R^2 > 0,3$ entre alinhamentos refinados e não refinados, destacam-se os conjuntos refinados, como observado para os alinhamentos fatiados em janelas de 100 e 200 nucleotídeos. Para o alinhamento 172N, 17/139 janelas móveis possuíam valores de $R^2 > 0,3$ (12,23%), 18/110 para o alinhamento 172R (16,36%), 16/139 para o alinhamento 154N (11,51%) e 16/103 para o alinhamento 154R (15,53%). Observa-se, portanto, que a fração de janelas móveis apresentando valores de $R^2 > 0,3$ foi maior nos alinhamentos fatiados em janelas de 300 nt do que naquelas de 100 e 200 nt. Em todos os alinhamentos fatiados em janelas de 300 nt, as janelas apresentando os maiores valores de R^2 estão localizadas na região codificadora da proteína CP.

4.2. Validação do conteúdo de sinal temporal observado ao longo dos genomas virais

Visando validar os resultados obtidos que indicam uma distribuição desuniforme do sinal temporal ao longo de sequências de DNA-A do ToSRV, foi realizada uma etapa segregando cada alinhamento de sequências completas de DNA-A em dois alinhamentos novos, o primeiro composto por todas as colunas do alinhamento que se sobrepõem às janelas 300 nt das quais foram obtidos valores de $R^2 > 0,3$ e o segundo pelas colunas restantes do alinhamento cujas janelas de 300 nt indicaram valores de $R^2 < 0,3$. Ambos os alinhamentos foram

reanalisados utilizando-se o TreeTime. O alinhamento contendo colunas de janelas com $R^2 > 0,3$ derivado do alinhamento 172N apresentou o valor de $R^2 = 0,49$, enquanto o alinhamento contendo colunas de janelas com $R^2 < 0,3$ apresentou, inesperadamente, no valor de $R^2 = 0,34$. Quando se realizou o mesmo procedimento a partir do alinhamento 172R, o alinhamento supostamente contendo maior sinal temporal apresentou o valor de $R^2 = 0,43$ e aquele contendo supostamente o menor sinal resultou em $R^2 = 0,20$. Por sua vez, os alinhamentos derivados do 154N apresentaram os valores de $R^2 = 0,42$ e $R^2 = 0,41$, respectivamente; enquanto aqueles derivados do alinhamento 154R resultaram nos valores de $R^2 = 0,54$ e $R^2 = 0,15$. Desta forma, somente os alinhamentos que passaram por uma etapa prévia de refinamento foram validados para seus conteúdos de sinal temporal.

No geral, observa-se nas análises de regressão linear baseadas em regiões com menor conteúdo de sinal temporal (figuras 6b, 6d, 6f e 6h) um maior número de pontos correspondentes a isolados do ToSRV ignorados para a maximização dos valores de R^2 .

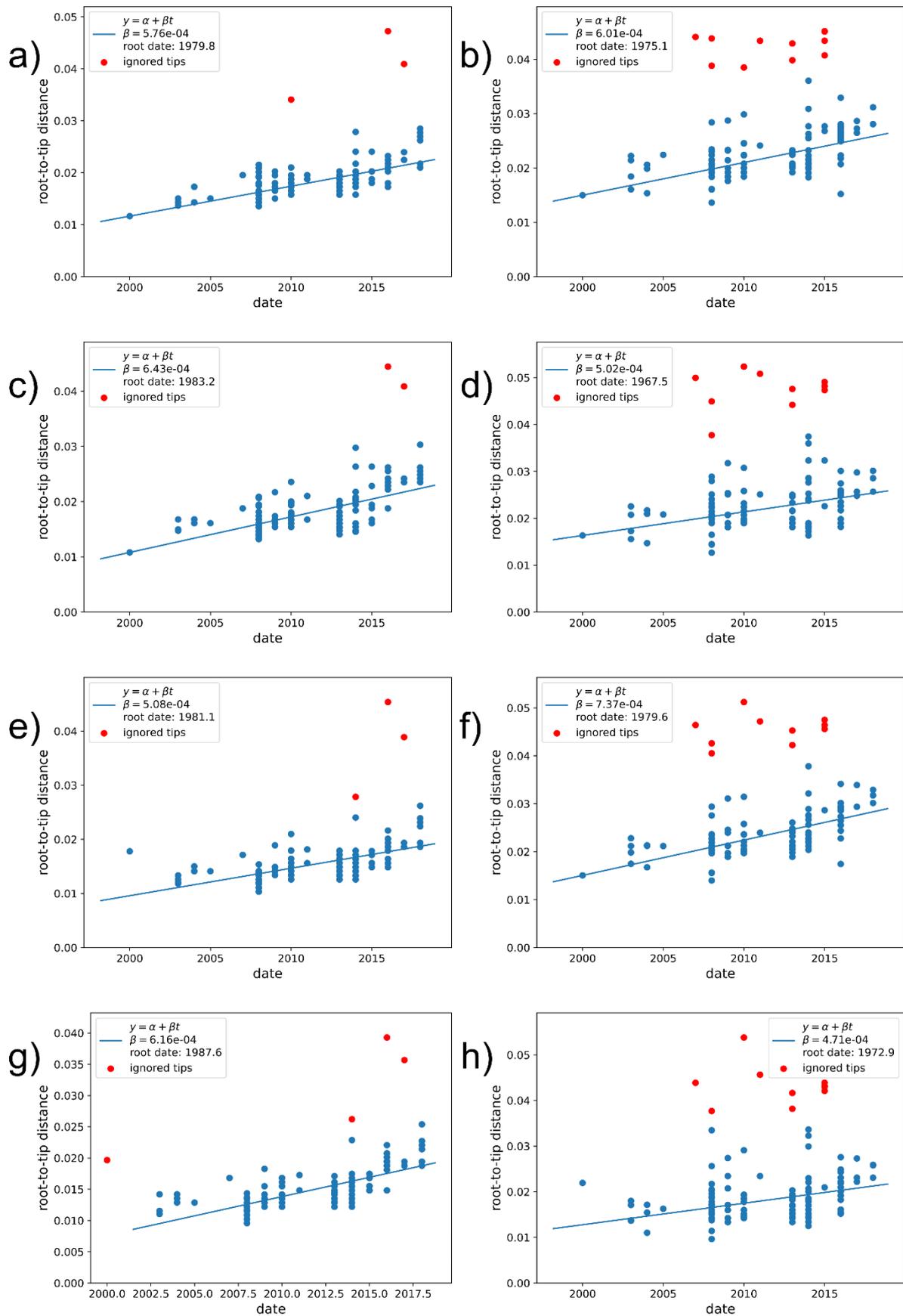


Figura 6. Gráficos de dispersão das análises de regressão linear conduzidas no TreeTime entre as distâncias raiz-ponta calculadas a partir das árvores de máxima verossimilhança e as datas de amostragem dos isolados do ToSRV.

As análises foram baseadas em alinhamentos construídos a partir da concatenação de colunas dos alinhamentos iniciais mapeadas em regiões de janelas móveis de 300 nt apresentando valores de $R^2 > 0,3$ (a, c, e, g) e $< 0,3$ (b, d, f, h) dos alinhamentos 172N (a, b); 172R (c, d); 154N (e, f) e 154R (g, h). Pontos apresentados na cor vermelha representam isolados do ToSRV que foram ignorados na análise de regressão linear.

Para os gráficos de regressão linear entre as distâncias raiz-ponta das árvores de máxima verossimilhança e das datas de coleta dos isolados virais gerados a partir das regiões com valores de $R^2 > 0,3$ e $R^2 < 0,3$, com base nos alinhamentos com sequências recombinantes (figura 6), observou-se para todos os alinhamentos a presença de maior número de isolados ignorados para análise de regressão linear para os conjuntos formados a partir das janelas com valor de $R^2 < 0,3$.

4.3. Efeito de eventos de recombinação sobre o sinal temporal

Análises de avaliação do sinal temporal também foram realizadas removendo-se dos conjuntos de dados as sequências identificadas como recombinantes no RDP, visando-se estimar a influência da recombinação sobre o conteúdo de sinal temporal presente ao longo dos genomas virais. A partir das análises no RDP, foram construídos quatro novos alinhamentos, sendo dois deles derivados dos alinhamentos de 172 sequências e os outros dois derivados dos alinhamentos de 154 sequências, construídos pela remoção de sequências de DNA-A identificadas como recombinantes por quatro ou mais dos sete métodos analíticos do programa RDP4. Esses novos conjuntos de dados foram designados ao longo do texto por meio do sufixo “+4”. Os alinhamentos 124N_4+ e 169R_4+ foram derivados dos alinhamentos 172N e 172R, respectivamente, enquanto os alinhamentos 151N_4+ e 151R_4+ foram derivados dos alinhamentos 154N e 154R, respectivamente.

Usando uma abordagem ainda mais conservadora, foram construídos outros quatro alinhamentos pela remoção de sequências de DNA-A nas quais qualquer evento de recombinação tenha sido detectado, independentemente do número de métodos analíticos do RDP envolvidos na detecção. Esses novos conjuntos de dados foram designados ao longo do texto por meio do sufixo “+1” indicando que os eventos de recombinação foram detectados por um ou mais dos sete métodos analíticos do RDP. Os alinhamentos 123N_1+ e 157R_1+ foram derivados dos alinhamentos 172N e 172R, respectivamente, enquanto os alinhamentos 133N_1+ e 115R_1+ foram derivados dos alinhamentos 154N e 154R, respectivamente.

Os novos alinhamentos foram submetidos ao processo de fatiamento em janelas móveis nos mesmos moldes dos alinhamentos prévios. Foram gerados mapas de calor com base nos

valores de R^2 calculados a partir da regressão raiz-ponta utilizando-se árvores de máxima verossimilhança (figuras 7, 8 e 9).

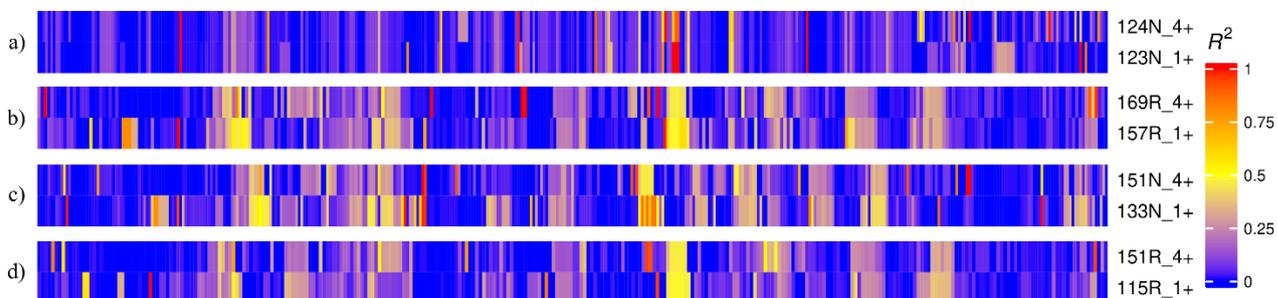


Figura 7. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de 100 nucleotídeos. **(a)** alinhamentos derivados do 172N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (124N_4+), e por ao menos um método analítico (123N_1+) do RDP; **(b)** alinhamentos derivados do 172R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (169R_4+), e por ao menos um método analítico (157R_1+) do RDP; **(c)** alinhamentos derivados do 154N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151N_4+), e por ao menos um método analítico (133N_1+) do RDP; **(d)** alinhamentos derivados do 154R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151R_4+), e por ao menos um método analítico (115R_1+) do RDP.

Analisando-se os mapas construídos a partir de janelas móveis de 100 nt (figura 7), nota-se – tal qual para os mapas construídos para os alinhamentos sem remoção de sequências recombinantes – ampla variação nos valores de R^2 ao longo do DNA-A de ToSRV, com janelas apresentando desde $R^2 = 0$ até $R^2 = 1$. Foi observada a presença de janelas com valores de $R^2 = 1$ na região codificadora da porção N-terminal da proteína associada à replicação (Rep) do alinhamento 123N_1+, e após conferência concluiu-se que estas não tiveram sinal temporal induzido pela presença de gaps advindos da etapa de alinhamento. Para os outros alinhamentos há a presença de sinal temporal nessa região, porém com menor intensidade. Observou-se a presença de janelas na região codificadora da porção central da proteína CP com valores de $R^2 \sim 0,5$ nos alinhamentos 157R_1+, 133N_1+, 169R_4+, 151N_4+, 151R_4+ e 115R_4+.

Contrastando os resultados observados para os alinhamentos fatiados em janelas de 100 nt onde não houve a remoção de sequências recombinantes (figura 3) foi bastante evidente a mudança na distribuição do sinal temporal presente ao longo do DNA-A do ToSRV. Para todos os alinhamentos sem a remoção de sequências recombinantes, observou-se janelas com maiores valores de R^2 na região codificadora da porção central da proteína CP, o que foi notado somente

em alguns dos alinhamentos após a remoção de sequências recombinantes, sobretudo nos alinhamentos em que foram removidas todas as sequências com evidência de recombinação (alinhamentos com sufixo “1+”) (figura 7). Por outro lado, em todos os alinhamentos com remoção de sequências recombinantes (figura 7) observou-se a presença de janelas com valores de $R^2 \sim 0,5$ na região codificadora da porção C-terminal da sequência codificadora da proteína Rep, o que não ocorreu para os alinhamentos fatiados em janelas de 100 nt sem remoção de sequências recombinantes (figura 3). A proporção de janelas móveis de 100 nt com valores de $R^2 > 0,3$ variou bastante. Para o alinhamento 124N_4+, 27/415 janelas móveis apresentaram valor de $R^2 > 0,3$ (6,50%), 20/415 para o alinhamento 123N_1+ (4,81%), 39/329 para o alinhamento 169R_4+ (11,85%), 54/329 para o alinhamento 157R_1+ (16,41%), 46/415 para o alinhamento 151N_4+ (11,08%), 86/415 para o alinhamento 133N_1+ (20,72%), 30/308 para o alinhamento 151R_4+ (9,74%) e 29/308 para o alinhamento 115R_1+ (9,41%). A remoção de sequências recombinantes gerou maior fração de janelas móveis com valores de R^2 maiores que 0,3 somente para os alinhamentos baseados no 172R (169R_4+ e 157R_1+, Figura 7b) e no 154N (151N_4+ e 133N_1+, Figura 7c).

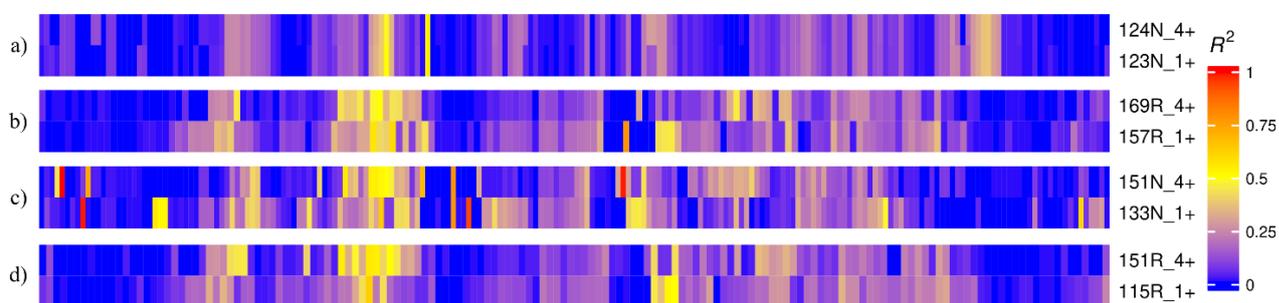


Figura 8. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de 200 nucleotídeos. **(a)** alinhamentos derivados do 172N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (124N_4+), e por ao menos um método analítico (123N_1+) do RDP; **(b)** alinhamentos derivados do 172R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (169R_4+), e por ao menos um método analítico (157R_1+) do RDP; **(c)** alinhamentos derivados do 154N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151N_4+), e por ao menos um método analítico (133N_1+) do RDP; **(d)** alinhamentos derivados do 154R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151R_4+), e por ao menos um método analítico (115R_1+) do RDP.

Nos mapas de calor gerados para os alinhamentos fatiados em janelas móveis de 200 nt, com remoção de sequências recombinantes (figura 8), observou-se a presença de somente uma

janela móvel com valor de $R^2 = 1$ no alinhamento 151N_4+, entretanto analisando-se a região do alinhamento na qual está essa janela, há a presença de várias colunas compostas majoritariamente por gaps, o que sugere que esse resultado se trata de um artefato advindo da etapa de alinhamento. Há presença de janelas com valores de $R^2 \sim 0,5$ na região codificadora da porção C-terminal da proteína Rep nos alinhamentos fatiados em janelas de 200 nt, após remoção de sequências recombinantes, como foi observado para os alinhamentos análogos fatiados em janelas de 100 nt. Na região codificadora da porção central da proteína CP foi observado um número considerável de janelas adjacentes com valores de $R^2 \sim 0,5$ para todos os alinhamentos, exceto nos baseados no alinhamento 172N (124N_4+ e 123N_1+, Figura 8a). A remoção de um número maior de sequências recombinantes nos alinhamentos do tipo “1+” não levou a valores mais altos de R^2 em comparação aos alinhamentos do tipo “4+” fatiados em janelas de 200 nt.

Comparando-se os resultados observados para os alinhamentos fatiados em janelas de 200 nt, sem remoção de sequências recombinantes (figura 4) com os resultados baseados nos alinhamentos correspondentes com remoção de sequências recombinantes (figura 8), observou-se similaridade na distribuição do sinal temporal ao longo das sequências de DNA-A e também na intensidade dos valores de R^2 (variando de $R^2 = 0$ até 1), inclusive para o padrão de distribuição de janelas na região codificadora da porção central da proteína CP. O alinhamento 124N_4+ apresentou 8/208 janelas móveis com valor de $R^2 > 0,3$ (3,80%), 8/208 para o alinhamento 123N_4+ (3,80%), 18/165 para o alinhamento 169R_4+ (10,91%), 22/165 para o alinhamento 157R_1+ (14,90%), 28/208 para o alinhamento 151N_4+ (13,46%), 31/208 para o alinhamento 133N_1+ (14,90%), 19/154 para o alinhamento 151R_4+ (12,33%) e 12/154 para o alinhamento 115R_1+ (7,79%). A fração de janelas com valores de $R^2 > 0,3$ se manteve-se similar entre os alinhamentos fatiados em janelas de 200 nt com e sem a remoção de sequências recombinantes.

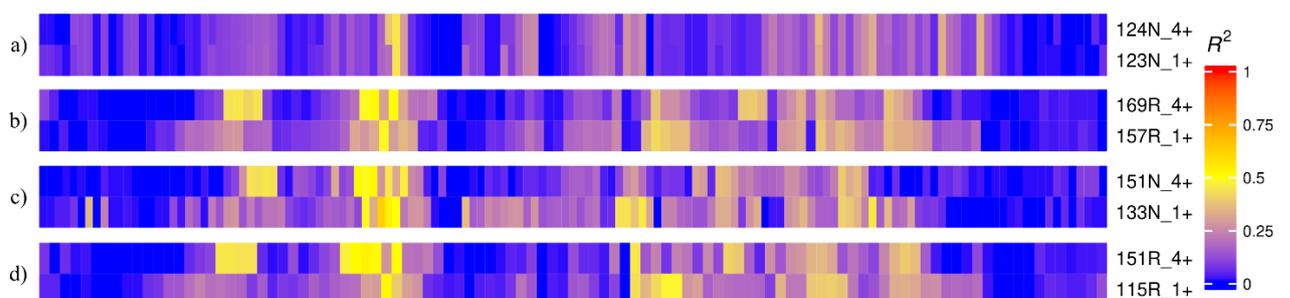


Figura 9. Mapas de calor representando valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para janelas móveis de

300 nucleotídeos. **(a)** alinhamentos derivados do 172N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (124N_4+), e por ao menos um método analítico (123N_1+) do RDP; **(b)** alinhamentos derivados do 172R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (169R_4+), e por ao menos um método analítico (157R_1+) do RDP; **(c)** alinhamentos derivados do 154N, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151N_4+), e por ao menos um método analítico (133N_1+) do RDP; **(d)** alinhamentos derivados do 154R, a partir da remoção de sequências identificadas como recombinantes por quatro ou mais métodos analíticos (151R_4+), e por ao menos um método analítico (115R_1+) do RDP.

Quanto aos resultados gerados a partir dos alinhamentos fatiados em janelas de 300 nt após remoção de sequências recombinantes (figura 9), observou-se menor intensidade de sinal temporal quando comparados aos de 100 e 200 nt (figuras 7 e 8), sendo o maior valor de R^2 de 0,63 obtido a partir do alinhamento 133N_1+. Não se notou presença significativa de sinal temporal distribuído ao longo das janelas móveis presentes na região codificadora da porção N-terminal da proteína Rep (como observado nos alinhamentos fatiados em janelas de 100 nt). Em contraponto, para a região codificadora da porção central da proteína CP, observou-se a presença de janelas com valores de $R^2 \sim 0,5$ em todos os alinhamentos. A proposta de manter-se somente uma sequência para cada conjunto de haplótipos para os alinhamentos fatiados em janelas de 300 nt, se mostrou inviável para alinhamentos não refinados (124N_4+, 123N_1+ versus 151N_4+, 133N_1+), gerando resultados discrepantes. Contudo para os alinhamentos refinados (169R_4+, 157R_1+ versus 151R_4+, 115R_1+), os resultados foram notavelmente similares.

A distribuição e intensidade de sinal temporal ao longo do DNA-A é similar para os alinhamentos fatiados em janelas de 300 nt sem (figura 5) e com remoção de sequências recombinantes (figura 9). Quanto à fração de janelas móveis de 300 nt com valores de $R^2 > 0,3$ para os alinhamentos apresentados na Figura 9, foram constatados os seguintes resultados: 124N_4+: 6/139 (4,31%), 123N_1+: 5/139 (3,59%), 169R_4+: 16/110 (14,54%), 157R_1+: 17/110 (15,45%), 151N_4+: 17/139 (12,23%), 133N_1+: 21/139 (15,10%), 151R_4+: 18/103 (17,47%), 115R_1+: 15/103 (14,56%). A fração de janelas com valor de $R^2 > 0,3$ foi similar comparando com os alinhamentos sem remoção de sequências recombinantes.

4.4. Validação do conteúdo de sinal temporal observado ao longo de sequências não recombinantes de DNA-A

De forma semelhante à efetuada para os alinhamentos contendo sequências recombinantes, a fim de se validar o sinal temporal, realizou-se uma nova etapa de segregação

dos alinhamentos completos do tipo 4+ e 1+ em dois novos alinhamentos: um primeiro supostamente com maior sinal temporal composto por todas as colunas do alinhamento original onde se mapearam janelas com $R^2 > 0,3$ e outro composto pelas colunas restantes do alinhamento original.

Alinhamento	$R^2 > 0.3$	$R^2 < 0.3$
124N_4+	0,45	0,03
123N_1+	0,44	0,04
169R_4+	-	-
157R_1+	0,48	0,43
151N_4+	-	-
133N_1+	0,35	0,25
151R_4+	0,57	0,06
115R_1+	0,43	0,25

Tabela 1. Valores de R^2 calculados no programa TreeTime a partir de análises de regressão linear raiz-ponta utilizando-se árvores de máxima verossimilhança construídas para alinhamentos gerados a partir da concatenação de colunas dos alinhamentos iniciais mapeadas em regiões de janelas móveis de 300 nt apresentando valores de $R^2 > 0,3$ e $< 0,3$.

A partir dos resultados obtidos a partir dos alinhamentos $R^2 > 0,3$ e $< 0,3$ (tabela 1), foram validados os resultados para os alinhamentos 124N_4+, 123N_1+, 133N_1+, 151R_4+ e 115R_1+. Os alinhamentos formados a partir do 169R_4+ e do 151N_4+ não obtiveram correlação para a análise, por isso não foram apresentados valor de R^2 e gráficos de regressão linear, desta forma esses alinhamentos foram considerados invalidados. Diferentemente dos resultados da validação realizada a partir dos alinhamentos contendo sequências recombinantes, nessa última, os resultados provenientes de um alinhamento refinado (157R_1+) não foram validados, já que o valor de R^2 para a concatenação de janelas móveis $R^2 < 0,3$ retornou valor acima de 0,3 ($R^2 = 0,43$).

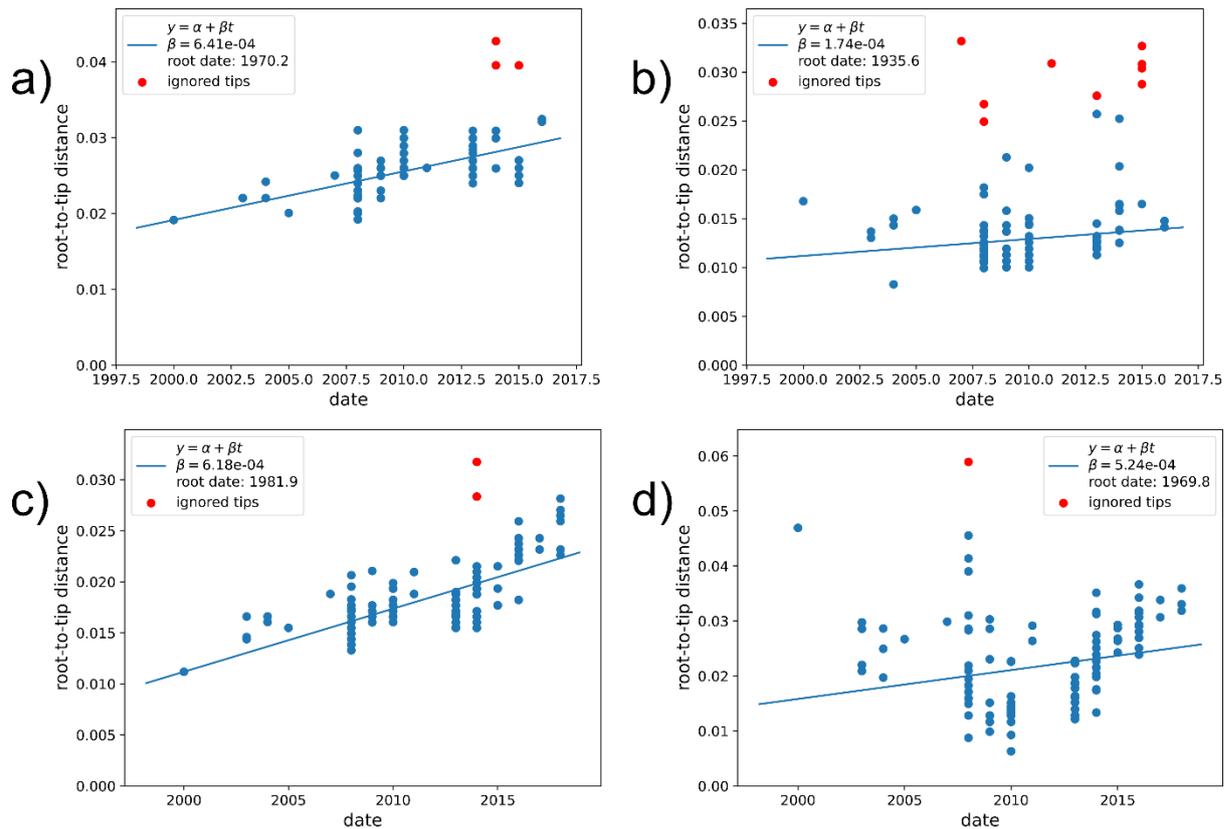


Figura 10. Gráficos de dispersão das análises de regressão linear conduzidas no TreeTime entre as distâncias raiz-ponta calculadas a partir das árvores de máxima verossimilhança e as datas de amostragem dos isolados do ToSRV. As análises foram baseadas em alinhamentos construídos a partir da concatenação de colunas dos alinhamentos iniciais mapeadas em regiões de janelas móveis de 300 nt apresentando valores de $R^2 > 0,3$ (a, c) e $R^2 < 0,3$ (b, d) dos alinhamentos 124N_4+ (a, b) e 151R_4+ (c, d). Pontos apresentados na cor vermelha representam isolados do ToSRV que foram ignorados na análise de regressão linear.

Nos gráficos de regressão linear entre as distâncias raiz-ponta das árvores de máxima verossimilhança e das datas de coleta dos isolados virais gerados a partir das regiões com valores de $R^2 > 0,3$ e $R^2 < 0,3$, com base nos alinhamentos “4+” (figura 10), observou-se para o alinhamento 124N_4+ (figura 10a) número consideravelmente maior de isolados ignorados para maximização do valor de R^2 para o conjunto derivado de regiões com janelas apresentando valores de $R^2 < 0,3$ (9 ignorados versus para $R^2 < 0,3$ versus 3 ignorados para $R^2 > 0,3$). Para o alinhamento 151R_4+ não foi possível perceber essa tendência, visto que o número de isolados ignorados para a regressão linear foi baixo e similar entre os dois conjuntos ($R^2 > 0,3$ = 2 isolados sem correlação e $R^2 < 0,3$ = 1 isolado não correlacionado). Não foram gerados gráficos para os alinhamentos 169R_4+ e 151N_4+ pois não houve correlação entre as sequências submetidas a análise.

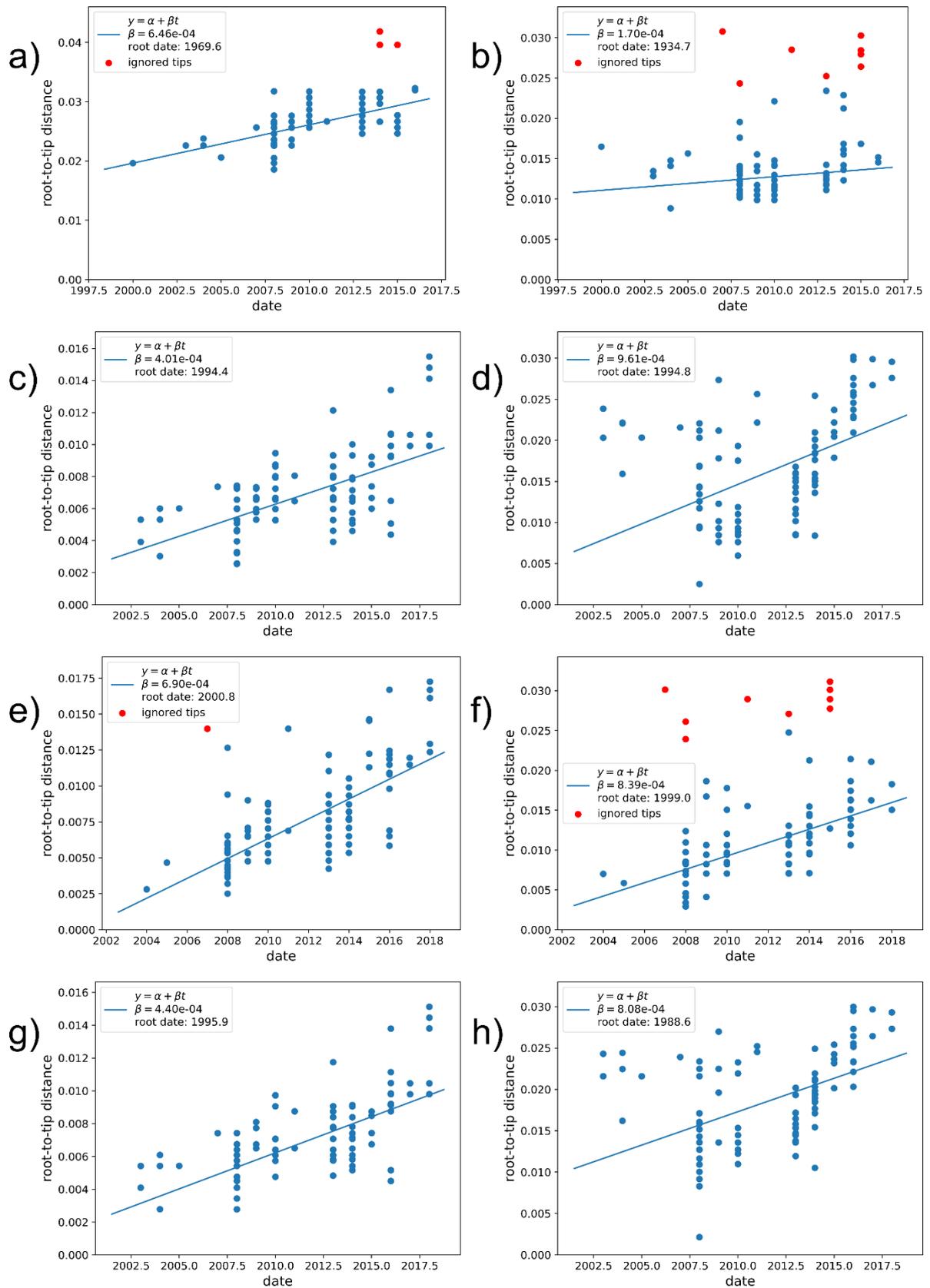


Figura 11. Gráficos de dispersão das análises de regressão linear conduzidas no TreeTime entre as distâncias raiz-ponta calculadas a partir das árvores de máxima verossimilhança e as datas de amostragem dos isolados do ToSRV. As análises foram baseadas em alinhamentos construídos a partir da concatenação de colunas dos alinhamentos

iniciais mapeadas em regiões de janelas móveis de 300 nt apresentando valores de $R^2 > 0,3$ (a, c, e, g) e $< 0,3$ (b, d, f, h) dos alinhamentos 123N_1+ (a, b); 157R_1+ (c, d); 133N_1+ (e, f) e 115R_1+ (g, h). Pontos apresentados na cor vermelha representam isolados do ToSRV que foram ignorados na análise de regressão linear.

Para os gráficos de regressão linear entre as distâncias raiz-ponta das árvores de máxima verossimilhança e das datas de coleta dos isolados virais gerados a partir das regiões com valores de $R^2 > 0,3$ e $R^2 < 0,3$, com base nos alinhamentos “1+” (figura 11), observou-se para os conjuntos $R^2 > 0,3$ e $R^2 < 0,3$ dos alinhamentos refinados (157R_1+ e 115R_1+) resultados demonstrando ausência de isolados ignorados pela análise de regressão linear, o que não ocorreu para os alinhamentos não refinados (123N_1+ e 133N_1+), ambos possuem isolados que foram ignorados pela análise, sendo maior o número de ignorados para os conjuntos $R^2 < 0,3$. Esses resultados indicam que alinhamentos que sofreram etapa de refinamento tendem a expressar maior correlação entre suas sequências de isolados.

5. DISCUSSÃO

Populações de begomovírus apresentam alto nível de variação genética gerada por meio dos mecanismos de mutação (DUFFY; HOLMES, 2008), recombinação (LEFEUVRE; MORIONES, 2015) e pseudorecombinação (GILBERTSON et al., 1993). Toda essa variação contida nos genomas virais pode ser estudada sob diferentes perspectivas por meio da correlação entre dados filogenéticos e geográficos (filogeografia) e informações temporais (filodinâmica) (VENKATARAVANAPPA et al., 2011; DUFFY; HOLMES, 2008).

Estudos sobre a evolução de espécies virais ao longo do tempo são de grande importância para a compreensão de sua dinâmica e comportamento, um estudo recente sobre o *Barley yellow dwarf virus* (BYDV), um luteovírus, inferiu por meio da calibração de um relógio molecular relaxado que o ancestral comum mais recente para a espécie existiu há cerca de 600 anos atrás. Por meio dessa informação temporal, os pesquisadores puderam traçar prováveis rotas de migração do vírus ao redor do planeta, associando a dispersão às atividades humanas (WEI et al., 2023).

O uso de janelas móveis de 200 nucleotídeos de comprimento, com 20 nucleotídeos de passo é uma prática comum em análises de bioinformática para begomovírus (KUMARI et al., 2011; SRIVASTAVA et al., 2015), neste trabalho foram utilizados três tamanhos de janelas (100, 200 e 300 nt), com três tamanhos de passo (10, 20 e 30 nt), visando observar a variação do sinal temporal ao longo do DNA-A de isolados do ToSRV sob diferentes níveis de resolução. Janelas de 100 nt apresentam uma maior resolução de detalhes, porém com um menor conteúdo de informação genética por janela; janelas de 200 nt representam um cenário intermediário e janelas de 300 nt apresentam uma menor resolução, porém com um maior conteúdo de informação genética em cada janela móvel. Essa prática se demonstrou eficaz para a proposta do trabalho, já que foi possível visualizar diferentes nuances do sinal temporal a partir de um mesmo alinhamento.

Sequências idênticas podem representar uma considerável parte do conjunto de dados a ser analisado. Neste trabalho o alinhamento composto por todas as sequências de isolados é aproximadamente 10% maior do que o alinhamento no qual manteve-se apenas um representante para cada conjunto de haplótipos (172 sequências *versus* 154 sequências). Pressupõe-se que caso os resultados da análise de regressão do tipo raiz-ponta sejam semelhantes para alinhamentos cuja única diferença seja a adoção de um único representante para cada conjunto de haplótipos, pode-se adotar o uso de alinhamentos compostos por apenas um representante para o conjunto de sequências idênticas, o que diminui significativamente o número de sequências a se analisar gerando menos carga computacional e reduzindo-se o tempo

necessário para as análises. Para os alinhamentos 172N, 172R, 154N e 154R o uso dessa abordagem não resultou em diferenças nos resultados obtidos, indicando a possibilidade de aplicação.

Durante o processo de alinhamento de sequências, são inseridos gaps para alinhar adequadamente regiões homólogas em todas as sequências do conjunto de dados em questão. Essas lacunas por vezes compõem considerável parte do alinhamento, o que diminui a informação genética (número de nucleotídeos) presente em cada uma das janelas móveis. Em teoria, a adição de uma etapa de refinamento (remoção de colunas majoritariamente compostas por lacunas no alinhamento) aumenta a concentração de informação genética nas janelas móveis. Visando-se aumentar a qualidade dos alinhamentos mantendo-se a informação temporal presente nos conjuntos de dados originais, submeteu-se os alinhamentos anteriormente criados (172 e 154) à etapa de refinamento. A etapa de refinamento alterou os resultados positivamente de forma geral. Para todos os alinhamentos, os resultados observados comparando-se não refinados versus refinados foram similares, entretanto com maior número de janelas com sinal temporal e maiores valores de R^2 notados nos alinhamentos refinados.

A fim de analisar a influência de eventos de recombinação sobre o sinal temporal ao longo da extensão do DNA-A do ToSRV, foram removidas sequências de DNA-A identificadas como recombinantes por meio do programa RDP4. Estudos prévios consideram apropriada a remoção de sequências nas quais eventos de recombinação tenham sido detectados por pelo menos três métodos analíticos do RDP (LIMA et al., 2013; MISHRA et al., 2020; MISHRA et al., 2022; SAHU et al., 2024). Neste estudo foram aplicadas duas abordagens para remoção de sequências recombinantes (4+ e 1+). De acordo com Posada e Crandall (2001), filogenias obtidas a partir de conjuntos permeados por eventos de recombinação podem ser bem diferentes das histórias evolutivas reais destes. Para este trabalho, ao remover sequências recombinantes, para ambas as abordagens (4+ e 1+), obtivemos resultados diferentes dos observados nos alinhamentos com sequências recombinantes, sendo a principal diferença um acúmulo de janelas com valor de $R^2 \sim 0,4$ na região codificadora da porção C-terminal da proteína Rep, sendo esse sinal temporal mais intenso para alinhamentos 1+. De forma geral, a remoção de sequências recombinantes trouxe resultados anteriormente ocultos devido à influência dos eventos de recombinação nos alinhamentos contendo sequências recombinantes.

O gene *cp* codifica a proteína capsidial responsável pela morfologia geminada característica dos begomovírus. A proteína CP possui função de acoplamento e proteção das moléculas de ssDNA, além de ser necessária para a transmissão viral por meio do vetor *Bemisia tabaci*, fazendo dela um exemplo de proteína viral multifuncional. Para a região codificadora

da porção central da CP considerando a espécie viral AbMV (um begomovírus bipartido), um estudo identificou presença de uma sequência de aminoácidos responsável pela transmissão via mosca-branca (HARRISON; SWANSON; FARGETTE, 2002). Unseld et al. (2001) localizaram o sinal de localização nuclear da CP nas regiões N-terminal (entre aminoácidos 1 e 54), central (entre aminoácidos 100 e 127) e C-terminal (entre aminoácidos 201 e 258) da proteína. Pitaksutheepong et al. (2007) realizaram um estudo que determinou a presença do domínio de ligação ao DNA na região N-terminal da CP, entre os aminoácidos 1 e 62. Ward e Lazarowitz (1999) demonstraram a presença de um domínio NES (nuclear export signal) na região central da CP (entre aminoácidos 177 e 198). A região N-terminal da CP apresenta diversas funções, como a auto-ligação (CP-CP), ligação entre a CP e o ssDNA viral (CP-DNA), além de ser a região imunodominante. A região C-terminal possui funções similares, entretanto não sendo imunodominante (HARRISON; SWANSON; FARGETTE, 2002). Neste trabalho notou-se presença de sinal temporal mediano na porção central da região codificadora da CP, não sendo observado sinal temporal relevante nas porções N-terminal e C-terminal.

A proteína Rep, codificada pelo gene *acI*, é indispensável para a replicação viral dentro do hospedeiro vegetal, possuindo diversas funções, como a iniciação da replicação do DNA, alongamento e terminação por meio da interação com proteínas virais e/ou do hospedeiro (RIZVI et al., 2015). A região codificadora da porção N-terminal possui domínios de oligomerização (entre aminoácidos 120 e 182), ligação a DNA (entre os aminoácidos 1 e 130) e clivagem de DNA (entre aminoácidos os 1 e 120) (RIZVI et al., 2015). Já a região codificadora da porção C-terminal inclui o domínio ATPase (BROWN, 2000; OROZCO et al., 1997), os domínios Walker A e B que são necessários para a atividade de ATPase (CHOUDHURY et al., 2006), e o domínio de ligação aos iterons (RUHEL; CHAKRABORTY, 2019). Neste trabalho observou-se presença de sinal temporal mediano na porção C-terminal da proteína Rep, após remoção de sequências recombinantes.

6. CONCLUSÃO

- O uso da abordagem de janelas móveis permite visualização mais profunda e detalhada da disposição do sinal temporal ao longo de genomas virais, facilitando a identificação de segmentos de sequências de nucleotídeos com maior correlação entre suas datas de amostragem e sua distância genética.
- O componente DNA-A do ToSRV apresenta pouco sinal temporal em sua extensão.
- Há sinal temporal moderado no DNA-A do ToSRV na região codificadora da porção central da proteína CP.
- O uso de somente uma sequência como representante para cada conjunto de haplótipos gerou resultados semelhantes aos alinhamentos com sequências recombinantes.
- A aplicação de uma etapa de refinamento aumentou a intensidade do sinal temporal detectável em todos os alinhamentos.
- Eventos de recombinação diminuem a intensidade do sinal temporal ao longo do DNA-A do ToSRV.
- A remoção de sequências recombinantes nas análises evidencia existência de sinal temporal moderado no DNA-A do ToSRV na região codificadora da porção C-terminal da proteína Rep.

7. REFERÊNCIAS

- Acosta-Leal, R., Duffy, S., Xiong, Z., Hammond, R. W., & Elena, S. F. (2011). Advances in Plant Virus Evolution: Translating Evolutionary Insights into Better Disease Management. *Phytopathology*[®], *101*(10), 1136–1148. <https://doi.org/10.1094/PHYTO-01-11-0017>
- Andrade, E. C., Manhani, G. G., Alfenas, P. F., Calegario, R. F., Fontes, E. P. B., & Zerbini, F. M. (2006). Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida* sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*. *Journal of General Virology*, *87*(12), 3687–3696. <https://doi.org/10.1099/vir.0.82279-0>
- Awadalla, P. (2003). The evolutionary genomics of pathogen recombination. *Nature Reviews Genetics*, *4*(1), 50–60. <https://doi.org/10.1038/nrg964>
- Barbosa, J. C., Barreto, S. da S., Inoue-Nagata, A. K., & Rezende, J. A. M. (2011). Characterization and Experimental Host Range of a Brazilian Tomato Isolate of Tomato severe rugose virus. *Journal of Phytopathology*, *159*(9), 644–646. <https://doi.org/10.1111/j.1439-0434.2011.01817.x>
- Bisaro, D. M. (2006). Silencing suppression by geminivirus proteins. *Virology*, *344*(1), 158–168. <https://doi.org/10.1016/j.virol.2005.09.041>
- Bonnamy, M., Blanc, S., & Michalakakis, Y. (2023). Replication mechanisms of circular ssDNA plant viruses and their potential implication in viral gene expression regulation. *MBio*, *14*(5). <https://doi.org/10.1128/mbio.01692-23>
- Böttcher, B., Unseld, S., Ceulemans, H., Russell, R. B., & Jeske, H. (2004). Geminata Structures of *African Cassava Mosaic Virus*. *Journal of Virology*, *78*(13), 6758–6765. <https://doi.org/10.1128/JVI.78.13.6758-6765.2004>
- Brown, J. K. (2000). Molecular markers for the identification and global tracking of whitefly vector–Begomovirus complexes. *Virus Research*, *71*(1–2), 233–260. [https://doi.org/10.1016/S0168-1702\(00\)00221-5](https://doi.org/10.1016/S0168-1702(00)00221-5)
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Choudhury, N. R., Malik, P. S., Singh, D. K., Islam, M. N., Kaliappan, K., & Mukherjee, S. K. (2006). The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Research*, *34*(21), 6362–6377. <https://doi.org/10.1093/nar/gkl903>
- Chowda-Reddy, R. V., Achenjang, F., Felton, C., Etarock, M. T., Anangfac, M.-T., Nugent, P., & Fondong, V. N. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. *Virus Research*, *135*(1), 115–124. <https://doi.org/10.1016/j.virusres.2008.02.014>
- Chowda-Reddy, R. V., Dong, W., Felton, C., Ryman, D., Ballard, K., & Fondong, V. N. (2009). Characterization of the cassava geminivirus transcription activation protein putative nuclear localization signal. *Virus Research*, *145*(2), 270–278. <https://doi.org/10.1016/j.virusres.2009.07.022>
- Claverie, S., Bernardo, P., Kraberger, S., Hartnady, P., Lefeuvre, P., Lett, J.-M., Galzi, S., Filloux, D., Harkins, G. W., Varsani, A., Martin, D. P., & Roumagnac, P. (2018). From

- Spatial Metagenomics to Molecular Characterization of Plant Viruses: A Geminivirus Case Study (pp. 55–83). <https://doi.org/10.1016/bs.aivir.2018.02.003>
- de Barro, P. J., Liu, S.-S., Boykin, L. M., & Dinsdale, A. B. (2011). *Bemisia tabaci*: A Statement of Species Status. *Annual Review of Entomology*, 56(1), 1–19. <https://doi.org/10.1146/annurev-ento-112408-085504>
- Duffy, S., & Holmes, E. C. (2008). Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus. *Journal of Virology*, 82(2), 957–965. <https://doi.org/10.1128/JVI.01929-07>
- Duffy, S., & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *Journal of General Virology*, 90(6), 1539–1547. <https://doi.org/10.1099/vir.0.009266-0>
- Edgar, R. C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications*, 13(1), 6968. <https://doi.org/10.1038/s41467-022-34630-w>
- FARIA, J. C. de, & ARAGAO, F. J. L. (2013). Embrapa 5.1: o feijoeiro geneticamente modificado resistente ao mosaico dourado. *Santo Antônio de Goiás: Embrapa Arroz e Feijão*.
- Favara, G. M., de Oliveira, F. F., Chinelato, G. A., Bergamin Filho, A., & Rezende, J. A. M. (2023). Characterization of Soybean, Tomato, and *Nicandra physalodes* as Sources of Inoculum of Tomato Severe Rugose Virus to Tomato Crops. *Plant Disease*, 107(4), 1087–1095. <https://doi.org/10.1094/PDIS-10-21-2160-RE>
- Ferro, C. G., Silva, J. P., Xavier, C. A. D., Godinho, M. T., Lima, A. T. M., Mar, T. B., Lau, D., & Zerbini, F. M. (2017). The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from *Sida* spp. and *Leonurus sibiricus*, plus two New World alphsatellites. *Annals of Applied Biology*, 170(2), 204–218. <https://doi.org/10.1111/aab.12329>
- Fiallo-Olivé, E., & Navas-Castillo, J. (2023). Begomoviruses: what is the secret(s) of their success? *Trends in Plant Science*, 28(6), 715–727. <https://doi.org/10.1016/j.tplants.2023.01.012>
- Fiallo-Olivé, E., Lett, J.-M., Martin, D. P., Roumagnac, P., Varsani, A., Zerbini, F. M., & Navas-Castillo, J. (2021). ICTV Virus Taxonomy Profile: Geminiviridae 2021. *Journal of General Virology*, 102(12). <https://doi.org/10.1099/jgv.0.001696>
- Fondong, V. N., Reddy, R. V. C., Lu, C., Hankoua, B., Felton, C., Czymbek, K., & Achenjang, F. (2007). The Consensus N -Myristoylation Motif of a Geminivirus AC4 Protein Is Required for Membrane Binding and Pathogenicity. *Molecular Plant-Microbe Interactions®*, 20(4), 380–391. <https://doi.org/10.1094/MPMI-20-4-0380>
- Fourment, M., & Darling, A. E. (2018). Local and relaxed clocks: the best of both worlds. *PeerJ*, 6, e5140. <https://doi.org/10.7717/peerj.5140>
- Ge, L., Zhang, J., Zhou, X., & Li, H. (2007). Genetic Structure and Population Variability of Tomato Yellow Leaf Curl China Virus. *Journal of Virology*, 81(11), 5902–5907. <https://doi.org/10.1128/JVI.02431-06>
- Gilbertson, R. L., Batuman, O., Webster, C. G., & Adkins, S. (2015). Role of the Insect Supervectors *Bemisia tabaci* and *Frankliniella occidentalis* in the Emergence and Global Spread of Plant Viruses. *Annual Review of Virology*, 2(1), 67–93. <https://doi.org/10.1146/annurev-virology-031413-085410>

- Gilbertson, R. L., Hidayat, S. H., Paplomatas, E. J., Rojas, M. R., Hou, Y.-M., & Maxwell, D. P. (1993). Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses. *Journal of General Virology*, *74*(1), 23–31. <https://doi.org/10.1099/0022-1317-74-1-23>
- Gong, P., Tan, H., Zhao, S., Li, H., Liu, H., Ma, Y., Zhang, X., Rong, J., Fu, X., Lozano-Durán, R., Li, F., & Zhou, X. (2021). Geminiviruses encode additional small proteins with specific subcellular localizations and virulence function. *Nature Communications*, *12*(1), 4278. <https://doi.org/10.1038/s41467-021-24617-4>
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>
- Gupta, N., Reddy, K., Bhattacharyya, D., & Chakraborty, S. (2021). Plant responses to geminivirus infection: guardians of the plant immunity. *Virology Journal*, *18*(1), 143. <https://doi.org/10.1186/s12985-021-01612-1>
- Haley, A., Zhan, X., Richardson, K., Head, K., & Morris, B. (1992). Regulation of the activities of African cassava mosaic virus promoters by the AC1, AC2, and AC3 gene products. *Virology*, *188*(2), 905–909. [https://doi.org/10.1016/0042-6822\(92\)90551-Y](https://doi.org/10.1016/0042-6822(92)90551-Y)
- Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., & Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology*, *11*(11), 777–788. <https://doi.org/10.1038/nrmicro3117>
- Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., & Robertson, D. (1999). Geminiviruses: Models for Plant DNA Replication, Transcription, and Cell Cycle Regulation. *Critical Reviews in Plant Sciences*, *18*(1), 71–106. <https://doi.org/10.1080/07352689991309162>
- Harrison, B. D., Swanson, M. M., & Fargette, D. (2002). Begomovirus coat protein: serology, variation and functions. *Physiological and Molecular Plant Pathology*, *60*(5), 257–271. <https://doi.org/10.1006/pmpp.2002.0404>
- He, Y.-Z., Wang, Y.-M., Yin, T.-Y., Fiallo-Olivé, E., Liu, Y.-Q., Hanley-Bowdoin, L., & Wang, X.-W. (2020). A plant DNA virus replicates in the salivary glands of its insect vector via recruitment of host DNA synthesis machinery. *Proceedings of the National Academy of Sciences*, *117*(29), 16928–16937. <https://doi.org/10.1073/pnas.1820132117>
- INOUE-NAGATA, A. K.; ÁVILA, A. C.; VILLAS-BÔAS, G. L. Os geminivírus em sistema de produção integrada de tomate indústria. Brasília: Embrapa Hortaliças, 2009. Circular Técnica, 71.
- Inoue-Nagata, A.K., Carvalho, C.M., Zerbini, F.M., Rezende, J.A.M., Krause-Sakate, R., Nagata, T. (2016). Vírus transmitidos por moscas-brancas no Brasil: vetores, principais doenças e manejo. *Revisão Anual de Patologia de Plantas* 24:7-29.
- Jones, D. R. (2003). Plant Viruses Transmitted by Whiteflies. *European Journal of Plant Pathology*, *109*(3), 195–219. <https://doi.org/10.1023/A:1022846630513>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kil, E.-J., Byun, H.-S., Hwang, H., Lee, K.-Y., Choi, H.-S., Kim, C.-S., & Lee, S. (2021). Tomato Yellow Leaf Curl Virus Infection in a Monocotyledonous Weed (*Eleusine indica*). *The Plant Pathology Journal*, *37*(6), 641–651. <https://doi.org/10.5423/PPJ.FT.11.2021.0162>

- Kumari, P., Singh, A. K., Sharma, V. K., Chattopadhyay, B., & Chakraborty, S. (2011). A novel recombinant tomato-infecting begomovirus capable of transcomplementing heterologous DNA-B components. *Archives of Virology*, 156(5), 769–783. <https://doi.org/10.1007/s00705-011-0915-1>
- LAUFS, J. (1995). Geminivirus replication: Genetic and biochemical characterization of rep protein function, a review. *Biochimie*, 77(10), 765–773. [https://doi.org/10.1016/0300-9084\(96\)88194-6](https://doi.org/10.1016/0300-9084(96)88194-6)
- Lefeuvre, P., & Moriones, E. (2015). Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Current Opinion in Virology*, 10, 14–19. <https://doi.org/10.1016/j.coviro.2014.12.005>
- Lima, A. T. M., Silva, J. C. F., Silva, F. N., Castillo-Urquiza, G. P., Silva, F. F., Seah, Y. M., Mizubuti, E. S. G., Duffy, S., & Zerbini, F. M. (2017). The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evolution*, 3(1). <https://doi.org/10.1093/ve/vex005>
- Lima, A. T. M., Sobrinho, R. R., González-Aguilera, J., Rocha, C. S., Silva, S. J. C., Xavier, C. A. D., Silva, F. N., Duffy, S., & Zerbini, F. M. (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *Journal of General Virology*, 94(2), 418–431. <https://doi.org/10.1099/vir.0.047241-0>
- Liu, H., Chang, Z., Zhao, S., Gong, P., Zhang, M., Lozano-Durán, R., Yan, H., Zhou, X., & Li, F. (2023). Functional identification of a novel C7 protein of tomato yellow leaf curl virus. *Virology*, 585, 117–126. <https://doi.org/10.1016/j.virol.2023.05.011>
- Mabvakure, B., Martin, D. P., Kraberger, S., Cloete, L., van Brunschot, S., Geering, A. D. W., Thomas, J. E., Bananej, K., Lett, J.-M., Lefeuvre, P., Varsani, A., & Harkins, G. W. (2016). Ongoing geographical spread of Tomato yellow leaf curl virus. *Virology*, 498, 257–264. <https://doi.org/10.1016/j.virol.2016.08.033>
- Macedo, M. A., Barreto, S. S., Costa, T. M., Rocha, G. A., Dianese, E. C., Gilbertson, R. L., & Inoue-Nagata, A. K. (2017). First Report of *Tomato severe rugose virus*, a Tomato-Infecting Begomovirus, in Soybean Plants in Brazil. *Plant Disease*, 101(11), 1959–1959. <https://doi.org/10.1094/PDIS-05-17-0644-PDN>
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015a). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1(1). <https://doi.org/10.1093/ve/vev003>
- Mishra, M., Verma, R. K., Marwal, A., Sharma, P., & Gaur, R. K. (2020). Biology and Interaction of the Natural Occurrence of Distinct Monopartite Begomoviruses Associated With Satellites in *Capsicum annum* From India. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.512957>
- Mishra, M., Verma, R. K., Pandey, V., Srivastava, A., Sharma, P., Gaur, R., & Ali, A. (2022). Role of Diversity and Recombination in the Emergence of Chilli Leaf Curl Virus. *Pathogens*, 11(5), 529. <https://doi.org/10.3390/pathogens11050529>
- Monci, F., Sánchez-Campos, S., Navas-Castillo, J., & Moriones, E. (2002). A Natural Recombinant between the Geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus Exhibits a Novel Pathogenic Phenotype and Is Becoming Prevalent in Spanish Populations. *Virology*, 303(2), 317–326. <https://doi.org/10.1006/viro.2002.1633>

- Mondal, D., Mandal, S., Shil, S., Sahana, N., Pandit, G. K., & Choudhury, A. (2019). Genome wide molecular evolution analysis of begomoviruses reveals unique diversification pattern in coat protein gene of Old World and New World viruses. *VirusDisease*, 30(1), 74–83. <https://doi.org/10.1007/s13337-019-00524-7>
- Navas-Castillo, J., Fiallo-Olivé, E., & Sánchez-Campos, S. (2011). Emerging Virus Diseases Transmitted by Whiteflies. *Annual Review of Phytopathology*, 49(1), 219–248. <https://doi.org/10.1146/annurev-phyto-072910-095235>
- Nawaz-ul-Rehman, M. S., & Fauquet, C. M. (2009). Evolution of geminiviruses and their satellites. *FEBS Letters*, 583(12), 1825–1832. <https://doi.org/10.1016/j.febslet.2009.05.045>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Orozco, B. M., Miller, A. B., Settlage, S. B., & Hanley-Bowdoin, L. (1997). Functional Domains of a Geminivirus Replication Protein. *Journal of Biological Chemistry*, 272(15), 9840–9846. <https://doi.org/10.1074/jbc.272.15.9840>
- Padidam, M., Beachy, R. N., & Fauquet, C. M. (1996). The Role of AV2 (“Precoat”) and Coat Protein in Viral Replication and Movement in Tomato Leaf Curl Geminivirus. *Virology*, 224(2), 390–404. <https://doi.org/10.1006/viro.1996.0546>
- Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, 265(2), 218–225. <https://doi.org/10.1006/viro.1999.0056>
- Paplomatas, E. J., Viresh, P. P., Yu-Ming Hou, Noueir, A. O., & Gilbertson, R. L. (1994). Molecular Characterization of a New Sap-Transmissible Bipartite Genome Geminivirus Infecting Tomatoes in Mexico. *Phytopathology*, 84(10), 1215–1224. <https://doi.org/10.1094/Phyto-84-1215>
- Paprotka, T., Metzler, V., & Jeske, H. (2010). The first DNA 1-like α satellites in association with New World begomoviruses in natural infections. *Virology*, 404(2), 148–157. <https://doi.org/10.1016/j.virol.2010.05.003>
- Pavan, M. A., Sakate, R. K., Moura, M. F., & Braga, R. de S. (2018). Viroses. In *Hortaliças-fruto* (pp. 241–269). EDUEM. <https://doi.org/10.7476/9786586383010.0009>
- Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, 30, 296–307. <https://doi.org/10.1016/j.meegid.2014.12.022>
- Pinto, V. B., Quadros, A. F. F., Godinho, M. T., Silva, J. C., Alfenas-Zerbini, P., & Zerbini, F. M. (2021). Intra-host evolution of the ssDNA virus tomato severe rugose virus (ToSRV). *Virus Research*, 292, 198234. <https://doi.org/10.1016/j.virusres.2020.198234>
- Pita, J. S., Fondong, V. N., Sangaré, A., Otim-Nape, G. W., Ogwal, S., & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology*, 82(3), 655–665. <https://doi.org/10.1099/0022-1317-82-3-655>
- Pitaksutheepong, C., Vimoltat, A., Nathwong, B., & Attathom, S. (2007). The N-terminal 62 amino acid residues of the coat protein of Tomato yellow leaf curl Thailand virus are responsible for DNA binding. *Journal of General Plant Pathology*, 73(1), 72–75. <https://doi.org/10.1007/s10327-006-0317-5>

- Posada, D., & Crandall, K. A. (2002). The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of Molecular Evolution*, 54(3), 396–402. <https://doi.org/10.1007/s00239-001-0034-9>
- Ribeiro, S. G., de Ávila, A. C., Bezerra, I. C., Fernandes, J. J., Faria, J. C., Lima, M. F., Gilbertson, R. L., Maciel-Zambolim, E., & Zerbini, F. M. (1998). Widespread Occurrence of Tomato Geminiviruses in Brazil, Associated with the New Biotype of the Whitefly Vector. *Plant Disease*, 82(7), 830–830. <https://doi.org/10.1094/PDIS.1998.82.7.830C>
- Rizvi, I., Choudhury, N. R., & Tuteja, N. (2015). Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. *Archives of Virology*, 160(2), 375–387. <https://doi.org/10.1007/s00705-014-2297-7>
- Rojas, M. R., Hagen, C., Lucas, W. J., & Gilbertson, R. L. (2005). Exploiting Chinks in the Plant's Armor: Evolution and Emergence of Geminiviruses. *Annual Review of Phytopathology*, 43(1), 361–394. <https://doi.org/10.1146/annurev.phyto.43.040204.135939>
- Ruhel, R., & Chakraborty, S. (2019). Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease*, 30(1), 66–73. <https://doi.org/10.1007/s13337-018-0458-0>
- Sagulenko, P., Puller, V., & Neher, R. A. (2018). TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution*, 4(1). <https://doi.org/10.1093/ve/vex042>
- Sahu, A. K., Pandey, V., Srivastava, A., Verma, R. K., Mishra, N. S., Akhtar, A., & Gaur, R. K. (2024). Molecular characterization of a bipartite begomovirus associated with leaf crinkle and curling symptoms of *Duranta erecta* in India. *Indian Phytopathology*, 77(1), 191–201. <https://doi.org/10.1007/s42360-024-00718-0>
- SANDERFOOT, A., & LAZAROWITZ, S. (1996). Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. *Trends in Cell Biology*, 6(9), 353–358. [https://doi.org/10.1016/0962-8924\(96\)10031-3](https://doi.org/10.1016/0962-8924(96)10031-3)
- Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1), 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- Sattar, M. N., Kvarnheden, A., Saeed, M., & Briddon, R. W. (2013). Cotton leaf curl disease – an emerging threat to cotton production worldwide. *Journal of General Virology*, 94(4), 695–710. <https://doi.org/10.1099/vir.0.049627-0>
- Srivastava, A., Kumar, S., Jaidi, M., & Raj, S. K. (2015). Characterization of a novel begomovirus associated with yellow mosaic disease of three ornamental species of *Jatropha* grown in India. *Virus Research*, 201, 41–49. <https://doi.org/10.1016/j.virusres.2015.02.015>
- Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X.-X., & Rokas, A. (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology*, 18(12), e3001007. <https://doi.org/10.1371/journal.pbio.3001007>
- Stenger, D. C., Revington, G. N., Stevenson, M. C., & Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proceedings of the National Academy of Sciences*, 88(18), 8029–8033. <https://doi.org/10.1073/pnas.88.18.8029>
- Sung, Y. K., & Coutts, R. H. A. (1995). Pseudorecombination and complementation between potato yellow mosaic geminivirus and tomato golden mosaic geminivirus. *Journal of General Virology*, 76(11), 2809–2815. <https://doi.org/10.1099/0022-1317-76-11-2809>

- Sunter, G., & Bisaro, D. M. (1997). Regulation of a Geminivirus Coat Protein Promoter by AL2 Protein (TrAP): Evidence for Activation and Derepression Mechanisms. *Virology*, 232(2), 269–280. <https://doi.org/10.1006/viro.1997.8549>
- Unsel, S., Höhnle, M., Ringel, M., & Frischmuth, T. (2001). Subcellular Targeting of the Coat Protein of African Cassava Mosaic Geminivirus. *Virology*, 286(2), 373–383. <https://doi.org/10.1006/viro.2001.1003>
- Vanitharani, R., Chellappan, P., Pita, J. S., & Fauquet, C. M. (2004). Differential Roles of AC2 and AC4 of Cassava Geminiviruses in Mediating Synergism and Suppression of Posttranscriptional Gene Silencing. *Journal of Virology*, 78(17), 9487–9498. <https://doi.org/10.1128/JVI.78.17.9487-9498.2004>
- Venkataravanappa, V., Lakshminarayana Reddy, C., Swaranalatha, P., Jalali, S., Briddon, R. W., & Reddy, M. K. (2011). Diversity and phylogeography of begomovirus-associated beta satellites of okra in India. *Virology Journal*, 8(1), 555. <https://doi.org/10.1186/1743-422X-8-555>
- Ward, B. M., & Lazarowitz, S. G. (1999). Nuclear Export in Plants: Use of Geminivirus Movement Proteins for a Cell-Based Export Assay. *The Plant Cell*, 11(7), 1267–1276. <https://doi.org/10.1105/tpc.11.7.1267>
- Wei, S., Chen, G., Yang, H., Huang, L., Gong, G., Luo, P., & Zhang, M. (2023). Global molecular evolution and phylogeographic analysis of barley yellow dwarf virus based on the cp and mp genes. *Virology Journal*, 20(1), 130. <https://doi.org/10.1186/s12985-023-02084-1>
- Wu, M., Wei, H., Tan, H., Pan, S., Liu, Q., Bejarano, E. R., & Lozano-Durán, R. (2021). Plant DNA polymerases α and δ mediate replication of geminiviruses. *Nature Communications*, 12(1), 2780. <https://doi.org/10.1038/s41467-021-23013-2>
- Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., & Varsani, A. (2017). ICTV Virus Taxonomy Profile: Geminiviridae. *Journal of General Virology*, 98(2), 131–133. <https://doi.org/10.1099/jgv.0.000738>
- Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W., & McKenna, R. (2001). Structure of the Maize Streak Virus Geminiate Particle. *Virology*, 279(2), 471–477. <https://doi.org/10.1006/viro.2000.0739>
- Zhao, K., Liu, S.-S., Wang, X.-W., Yang, J.-G., & Pan, L.-L. (2022). Manipulation of Whitefly Behavior by Plant Viruses. *Microorganisms*, 10(12), 2410. <https://doi.org/10.3390/microorganisms10122410>
- Zuckerkindl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2), 357–366. [https://doi.org/10.1016/0022-5193\(65\)90083-4](https://doi.org/10.1016/0022-5193(65)90083-4)

ANEXOS

Tabela suplementar 1. Sequências de ToSRV utilizadas neste estudo

#GenBank	Local de coleta	Hospedeiro	Ano de coleta
AY029750	Minas Gerais, Brasil	<i>Capsicum sp.</i>	2000
DQ207749	Brasil	<i>Capsicum sp.</i>	2003
JX415188	Goianápolis, Goiás, Brasil	<i>Nicandra physalodes</i>	2003
JF803260	Piriripau, Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2003
JF803261	Piriripau, Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2003
JF803263	Alexandre Gusmão, Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2004
MT215001	Goianápolis, Goiás, Brasil	<i>Solanum lycopersicum</i>	2004
JF803262	Goiás, Brasil	<i>Solanum lycopersicum</i>	2004
FJ824808	Brasil	<i>Capsicum annuum</i>	2005
EU086591	Sumaré, São Paulo, Brasil	<i>Nicandra physalodes</i>	2007
JX415190	Luziânia, Goiás, Brasil	<i>Crotalaria juncea</i>	2008
JX415197	Goianápolis, Goiás, Brasil	<i>Nicandra physalodes</i>	2008
KC004082	Carandaí, Minas Gerais, Brasil	<i>Sida sp.</i>	2008
KC706617	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC706618	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC706619	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC706620	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004075	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004076	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004077	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004078	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004079	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004080	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004081	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004083	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004084	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004085	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004086	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004087	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004088	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004089	Carandaí, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004073	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004070	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004071	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004072	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004074	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
JX415198	Goianápolis, Goiás, Brasil	<i>Solanum lycopersicum</i>	2008
JX415199	Goianápolis, Goiás, Brasil	<i>Solanum lycopersicum</i>	2008
KC004068	Jaíba, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008
KC004069	Jaíba, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2008

JX415193	Luziânia, Goiás, Brasil	<i>Solanum lycopersicum</i>	2008
JX415196	Luziânia, Goiás, Brasil	<i>Solanum lycopersicum</i>	2008
JX415201	Acreúna, Goiás, Brasil	<i>Solanum lycopersicum</i>	2009
JX415202	Acreúna, Goiás, Brasil	<i>Solanum lycopersicum</i>	2009
MW596530	Chã Grande, Pernambuco, Brasil	<i>Solanum lycopersicum</i>	2009
JX865626	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865627	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865629	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865615	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865616	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865617	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865619	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865634	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865635	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865639	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865641	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865618	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
JX865640	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2009
KC004090	Viçosa, Minas Gerais, Brasil	<i>Sida sp.</i>	2010
JX865620	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865621	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865622	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865623	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865624	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865625	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865628	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865630	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865631	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865632	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865633	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865636	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865637	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865638	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865642	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865643	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865644	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865645	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865646	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865647	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865648	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865649	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
JX865650	Viçosa, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2010
MW596522	Capão Bonito, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2011
MW596544	Planaltina, Goiás, Brasil	<i>Solanum lycopersicum</i>	2011
MT627045	Coimbra, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2013

MT627092	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627093	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627094	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627095	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627096	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627097	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627090	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627091	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
MT627089	Florestal, Minas Gerais, Brasil	<i>Solanum lycopersicum</i>	2014
KY524458	Gama, Distrito Federal, Brasil	<i>Solanum betaceum</i>	2015
MW596587	Capão Bonito, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2015
MW596588	Capão Bonito, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2015
MW596594	Capão Bonito, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2015
MN156291	Brasil	<i>Solanum lycopersicum</i>	2015
MN156292	Brasil	<i>Solanum lycopersicum</i>	2015
MN156306	Afonso Cláudio, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MW596580	Afonso Cláudio, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MW596560	Brazlândia, Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2016
MN156298	Domingos Martins, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MW596589	Mogi Guaçu, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2016
MW596590	Mogi Guaçu, São Paulo, Brasil	<i>Solanum lycopersicum</i>	2016
MT214084	Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2016
MN156293	Brasil	<i>Solanum lycopersicum</i>	2016
MN156294	Brasil	<i>Solanum lycopersicum</i>	2016
MN156297	Brasil	<i>Solanum lycopersicum</i>	2016
MW596570	Paranoá, Distrito Federal, Brasil	<i>Solanum lycopersicum</i>	2016
MW596553	Santa Maria de Jetibá, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MW596572	São João D'Aliança, Goiás, Brasil	<i>Solanum lycopersicum</i>	2016
MN059848	Venda Nova do Imigrante, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MN156295	Venda Nova do Imigrante, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MN156296	Venda Nova do Imigrante, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2016
MN156299	Brasil	<i>Solanum lycopersicum</i>	2017
MN156300	Santa Maria de Jetiba, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2017
MN156301	Venda Nova do Imigrante, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2017
MN156307	Domingos Martins, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2018
MN156308	Domingos Martins, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2018
MN156309	Domingos Martins, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2018
MN156310	Domingos Martins, Espírito Santo, Brasil	<i>Solanum lycopersicum</i>	2018
MN156302	Brasil	<i>Solanum lycopersicum</i>	2018
MN156303	Brasil	<i>Solanum lycopersicum</i>	2018
MN156305	Brasil	<i>Solanum lycopersicum</i>	2018
MN156304	Brasil	<i>Solanum lycopersicum</i>	2018

Tabela suplementar 2. Eventos de recombinação detectados no alinhamento 172N, para a criação do alinhamento 124N_4+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	2056	2552	MW596530	Unknown	RGBMCS ₃	4.53E-23
2	^MN156301	730	1058	MN156306	Unknown	RGBMCS ₃	7.71E-12
3	JX865633	1072	2012	JX865643	Unknown	RGBMS ₃	1.70E-11
4	^MW596530	3046	1168	KC706619	KY524458	MCS ₃	6.34E-04

^ A sequência recombinante pode ter sido identificada incorretamente.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 3. Eventos de recombinação detectados no alinhamento 172N, para a criação do alinhamento 123N_1+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	2056	2552	MW596530	Unknown	RGBMCS ₃	4.53E-23
2	^MN156301	730	1058	MN156306	Unknown	RGBMCS ₃	7.71E-12
3	JX865633	1072	2012	JX865643	Unknown	RGBMS ₃	1.70E-11
4	^MN156302	2227	2370	MW596553	Unknown	<u>G</u>	3.18E-03
5	^MW596530	3046	1168	KC706619	KY524458	MCS ₃	6.34E-04
6	^JX865619	3920	4102	MT627062	Unknown	<u>G</u>	2.36E-02

^ A sequência recombinante pode ter sido identificada incorretamente.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 4. Eventos de recombinação detectados no alinhamento 172R, para a criação do alinhamento 169R_4+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	1668	2131	MW596530	Unknown	RGMCS ₃	2.25E-23
2	MN156301	504	806	MN156306	Unknown	RGBMCS ₃	3.83E-12

3	JX865633	820	1627	JX865643	Unknown	<u>RGBMS3</u>	8.45E-12
---	----------	-----	------	----------	---------	---------------	----------

^ A sequência recombinante pode ter sido identificada incorretamente.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 5. Eventos de recombinação detectados no alinhamento 172R, para a criação do alinhamento 157R_1+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	1668	2131	MW596530	Unknown	<u>RGMCS3</u>	2.25E-23
2	^MN156301	504	806	MN156306	Unknown	<u>RGBMCS3</u>	3.83E-12
3	JX865633	820	1627	JX865643	Unknown	<u>RGBMS3</u>	8.45E-12
4	^MN156302	1834	1978	MW596553	Unknown	<u>G</u>	1.73E-03
5	KY524458	2544	913	Unknown	MW596572	<u>MCS</u>	4.46E-03
6	^MT214084	2610*	3278	KC706619	MW596572	<u>MC3</u>	1.66E-03

^ A sequência recombinante pode ter sido identificada incorretamente.

* O ponto de interrupção (*breakpoint*) atual não é determinado.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 6. Eventos de recombinação detectados no alinhamento 154N, para a criação do alinhamento 151N_4+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	2056	2552	MW596530	Unknown	<u>RGBMCS3</u>	4.53E-23
2	^MN156301	730	1058	MN156301	Unknown	<u>RGBMCS3</u>	7.71E-12
3	JX865633	1072	2012	JX865643	Unknown	<u>RGBMS3</u>	1.70E-11

^ A sequência recombinante pode ter sido identificada incorretamente.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 7. Eventos de recombinação detectados no alinhamento 154N, para a criação do alinhamento 133N_1+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	2056	2552	MW596530	Unknown	<u>RGBMCS3</u>	4.53E-23
2	^MN156301	730	1058	MN156306	Unknown	<u>RGBMCS3</u>	7.71E-12
3	JX865633	1072	2012	JX865643	Unknown	<u>RGBMS3</u>	1.70E-11
4	^MN156302	2227	2370	MW596553	Unknown	<u>G</u>	3.18E-03
5	^KX458238	3046	1168	Unknown	MW596530	<u>MC</u>	8.96E-03
6	^MT214084	3112*	4118	KC706619	MW596572	<u>MC3</u>	9.58E-03

^ A sequência recombinante pode ter sido identificada incorretamente.

* O ponto de interrupção (*breakpoint*) atual não é determinado.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 8. Eventos de recombinação detectados no alinhamento 154R, para a criação do alinhamento 151R_4+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	1558	1998	MW596530	Unknown	<u>RGBMCS3</u>	2.15E-23
2	^MN156301	454	747	MN156306	Unknown	<u>RGBMCS3</u>	3.65E-12
3	JX865633	760	1521	JX865643	Unknown	<u>RGBMS3</u>	8.04E-12

^ A sequência recombinante pode ter sido identificada incorretamente.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.

Tabela suplementar 9. Eventos de recombinação detectados no alinhamento 154R, para a criação do alinhamento 115R_1+.

Event	Recombinant	Recombination Breakpoints		Parents		Methods ^a	P-value ^b
		Initial	Final	Major	Minor		
1	MT214084	1558	1998	MW596530	Unknown	<u>RGBMCS3</u>	2.15E-23
2	^MN156301	454	747	MN156306	Unknown	<u>RGBMCS3</u>	3.65E-12
3	JX865633	760	1521	JX865643	Unknown	<u>RGBMS3</u>	8.04E-12

4	^MN156302	1713	1851	MW596553	Unknown	<u>G</u>	1.64E-03
5	^KX458238	2394	847	Unknown	KX458238	<u>MCS</u>	4.24E-03
6	^MT214084	2394*	3061	KC706619	MW596572	<u>M3</u>	8.74E-03
7	^JX865619	360	2923	Unknown	JX865637	<u>S3</u>	9.74E-09

[^] A sequência recombinante pode ter sido identificada incorretamente.

^{*} O ponto de interrupção (*breakpoint*) atual não é determinado.

^a R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

^b O valor de P (*P-value*) é relativo ao método com caractere sublinhado, sendo o menor valor de P calculado para o evento em questão.