

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CECÍLIA LEÃO PEREIRA RESENDE

**EXPLORING THE PHYLOGENETIC RELATIONSHIPS WITHIN VIRAL
GENOMES: A CASE STUDY ON BEGOMOVIRUSES**

Uberlândia – Minas Gerais
2024

CECÍLIA LEÃO PEREIRA RESENDE

**EXPLORING THE PHYLOGENETIC RELATIONSHIPS WITHIN VIRAL
GENOMES: A CASE STUDY ON BEGOMOVIRUSES**

Tese apresentada ao Programa de Pós-graduação em Agronomia, do Instituto de Ciências Agrárias da Universidade Federal de Uberlândia, como requisito para obtenção do título de Doutora em Agronomia.

Orientador: Prof. Dr. José Magno Queiroz Luz

Uberlândia – Minas Gerais
2024

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

R433e
2024 Resende, Cecília Leão Pereira, 1994-
Exploring the phylogenetic relationships within viral genomes
[recurso eletrônico] : a case study on begomoviruses / Cecília Leão
Pereira Resende. - 2024.

Orientador: José Magno Queiroz Luz.
Tese (Doutorado) - Universidade Federal de Uberlândia, Programa de
Pós-graduação em Agronomia.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.te.2025.5010>
Inclui bibliografia.
Inclui ilustrações.

1. Agronomia. I. Luz, José Magno Queiroz, 1967-, (Orient.). II.
Universidade Federal de Uberlândia. Programa de Pós-graduação em
Agronomia. III. Título.

CDU: 631

André Carlos Francisco
Bibliotecário-Documentalista - CRB-6/3408

CECÍLIA LEÃO PEREIRA RESENDE

**EXPLORING THE PHYLOGENETIC RELATIONSHIPS WITHIN VIRAL
GENOMES: A CASE STUDY ON BEGOMOVIRUSES**

Tese apresentada ao Programa de Pós-graduação em Agronomia, do Instituto de Ciências Agrárias da Universidade Federal de Uberlândia, como requisito para obtenção do título de Doutora em Agronomia.

Área de concentração: Produção Vegetal

Uberlândia, 02 de agosto de 2024

Banca Examinadora:

Prof. Dr. José Magno Queiroz Luz – UFU (Orientador)

Prof. Dr. Alison Talis Martins Lima – UFU

Prof. Dr. Flávio Tetsuo Sasaki – UFU

Dr. Richard Costa Polveiro – UFV

Dr. Lucas Antônio Stempkowski – UFV



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Secretaria da Coordenação do Programa de Pós-Graduação em
Agronomia
Rodovia BR 050, Km 78, Bloco 1CCG, Sala 206 - Bairro Glória, Uberlândia-MG, CEP
38400-902
Telefone: (34) 2512-6715/6716 - www.ppgagro.iclag.ufu.br - posagro@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Agronomia				
Defesa de:	Tese 011/2024, PPGAGRO				
Data:	dois de agosto de dois mil e vinte e quatro	Hora de início:	14:00	Hora de encerramento:	18:30
Matrícula do Discente:	12013AGR0008				
Nome do Discente:	Cecília Leão Pereira Resende				
Título do Trabalho:	Exploring the phylogenetic relationships within viral genomes: a case study on begomoviruses				
Área de concentração:	Produção Vegetal				
Linha de pesquisa:	Produção Vegetal em Áreas de Cerrado				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agronomia, assim composta: Professores Doutores: Alison Talis Martins Lima - UFU; Flávio Tetsuo Sasaki - UFU; Lucas Antônio Stempkowski - Syngenta; Richard Costa Polveiro - UFV; José Magno Queiroz Luz - UFU orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Dr. José Magno Queiroz Luz - UFU, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

Aprovado.

Essa tese possui mérito para ser indicada como destaque do Programa? (X) SIM
NÃO ()

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **José Magno Queiroz Luz, Professor(a) do Magistério Superior**, em 05/08/2024, às 10:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Richard Costa Polveiro, Usuário Externo**, em 05/08/2024, às 18:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Allison Talis Martins Lima, Professor(a) do Magistério Superior**, em 06/08/2024, às 11:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Flávio Tetsuo Sasaki, Professor(a) do Magistério Superior**, em 08/08/2024, às 16:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Lucas Antonio Stempkowski, Usuário Externo**, em 20/08/2024, às 09:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5578388** e o código CRC **AE809EAB**.

Referência: Processo nº 23117.049241/2024-21

SEI nº 5578388

AGRADECIMENTOS

A Deus, pela vida, saúde, proteção e bênçãos, e por estar sempre me guiando em todos os momentos.

Aos meus pais, Eugênio e Cida, e aos meus irmãos, Marília, Letícia e Thiago, pelo incentivo constante nos estudos, pelas orações e por apoiarem cada uma de minhas decisões. Aos meus avós, Oliveira, Maria Benedita e Maria Lourdes, pelo amor e carinho que sempre me dedicaram, e a toda minha família.

Ao meu namorado, Fabrício, pelo amor, carinho, compreensão e companheirismo em todos os momentos.

Ao professor e orientador Dr. José Magno Queiroz Luz, pela oportunidade e confiança, e ao professor Dr. Alison Talis Martins Lima, pelos ensinamentos, compreensão, orientação, disponibilidade e confiança.

Aos colegas do grupo de pesquisa PROVIRUS pelo apoio nas atividades e pelo companheirismo ao longo dessa trajetória, em especial ao Gabriel e a Lilian.

À CAPES, pela concessão da bolsa de doutorado.

À Universidade Federal de Uberlândia, ao seu corpo docente, direção, administração e demais funcionários, por todo o suporte.

ABSTRACT

Begomoviruses (genus *Begomovirus*, family *Geminiviridae*) cause severe losses in various crops worldwide. They possess genomes composed of one or two single-stranded DNA (ssDNA) molecules and infect a wide range of dicotyledonous plants. These viruses are persistently and circulatively transmitted by cryptic species of whiteflies within the *Bemisia tabaci* complex. In recent decades, the number of available begomovirus genomes in public databases has increased exponentially. This extensive genetic information has enabled an in-depth characterization of the genetic variability of *Begomovirus* populations from different regions worldwide. A notable feature of these populations is their high genetic variability generated by evolutionary mechanisms such as mutation, recombination, and pseudorecombination (reassortment). In this context, begomovirus genomes serve as valuable models for understanding the evolutionary dynamics of ssDNA viruses. Additionally, the presence of overlapping genes in begomovirus genomes enables comparative studies with non-overlapping coding regions. In this study, begomoviruses were used as model viruses to examine the phylogenetic relationships between coding and non-coding regions (intergenic region, IR) of the DNA-A segment from 24 distinct viral species. Gene- and IR-based phylogenies were reconstructed, and pairwise weighted Robinson-Foulds (wRF) distances were calculated between all Maximum Likelihood (ML) trees. Furthermore, similar analyses were conducted using a sliding window approach to elucidate phylogenetic relationships within genes and the IR. The rarefaction technique was applied to compare datasets of begomovirus species with heterogeneous sample sizes, enabling direct comparisons of topological incongruence among ML trees. The results demonstrate a wide variation in the degree of topological incongruence between coding and non-coding regions of the DNA-A segment. Moreover, closely positioned genes on the DNA-A segment exhibited more similar phylogenies than those farther apart. Exceptions were observed for the movement protein (mp) gene and IR, whose phylogenies exhibited divergent topologies from proximate genes. These findings were consistent with the existence of recombination hotspots often segregating these regions, resulting in topologies reflecting distinct evolutionary histories. Interestingly, beyond recombination hotspots, topological distances between trees correlated positively with the physical distance between the midpoints of each gene, indicating that phylogenies in these regions are affected by their physical distances on DNA-A. Collectively, these results demonstrate that the phylogeny of DNA-A genes in begomoviruses is strongly influenced by recombination. This observation suggests that bifurcating trees do not adequately represent the evolutionary relationships among viral isolates and highlights the need to adopt evolutionary history reconstruction methods that explicitly account for recombination.

Keywords: *Begomovirus*, Recombination, Molecular Evolution, Phylogeny

RESUMO

Os vírus do gênero *Begomovirus* (família *Geminiviridae*) causam severas perdas em diversas culturas agrícolas em todo o mundo. Eles possuem genomas compostos por uma ou duas moléculas de DNA de fita simples (ssDNA) e infectam uma ampla gama de plantas dicotiledôneas. Esses vírus são transmitidos de forma persistente e circulativa por espécies crípticas de moscas-brancas do complexo *Bemisia tabaci*. Nas últimas décadas, o número de genomas de begomovírus disponíveis em bancos de dados públicos aumentou exponencialmente. Essas extensivas informações genéticas têm permitido uma caracterização profunda da variabilidade genética de populações de begomovírus de diferentes regiões do mundo. Uma característica notável dessas populações é a alta variabilidade genética gerada por mecanismos evolutivos, como mutação, recombinação e pseudorrecombinação (rearranjo). Nesse contexto, genomas de begomovírus servem como modelos valiosos para entender a dinâmica evolutiva dos vírus de ssDNA. Além disso, a presença de genes sobrepostos nestes genomas permite estudos comparativos com regiões codificadoras não sobrepostas. Neste estudo, várias espécies de Begomovírus foram usadas como modelos para examinar as relações filogenéticas entre regiões codificadoras e não codificadoras (região intergênica, IR) do segmento de DNA-A de 24 espécies virais distintas. Foram reconstruídas filogenias baseadas em genes e IR e calculadas as distâncias par a par de Robinson-Foulds ponderada (wRF) entre todas as árvores de Máxima Verossimilhança (ML). Além disso, foram realizadas análises semelhantes utilizando uma abordagem de janelas deslizantes para elucidar as relações filogenéticas dentro dos genes e da IR. A técnica de rarefação foi aplicada para comparar conjuntos de dados de espécies de begomovírus com tamanhos de amostras heterogêneas, permitindo comparações diretas dos níveis de incongruência topológica entre árvores de ML. Os resultados indicaram uma ampla variação no grau de incongruência topológica entre regiões codificadoras e não codificadoras do segmento DNA-A. Ademais, genes fisicamente próximos no segmento de DNA-A apresentaram filogenias mais semelhantes em comparação a genes mais distantes entre si. Exceções foram o gene da proteína de movimento (mp) e a IR, cujas filogenias apresentaram topologias divergentes em relação a genes próximos. Esses resultados foram consistentes com a existência de hotspots de recombinação que frequentemente segregaram essas regiões, resultando em topologias que refletem diferentes histórias evolutivas. Curiosamente, fora de hotspots de recombinação, as distâncias topológicas entre árvores correlacionaram-se positivamente com a distância física entre os pontos médios de cada gene, indicando que as filogenias nessas regiões são moldadas por suas distâncias físicas no DNA-A. Conjuntamente, esses resultados demonstram claramente que a filogenia de genes do DNA-A em espécies de Begomovírus é fortemente influenciada por recombinação. Essa observação sugere que árvores bifurcadas não representam adequadamente as relações evolutivas entre isolados virais e destaca a necessidade de adoção de métodos de reconstrução da história evolutiva que considerem explicitamente a recombinação.

Palavras-chave: *Begomovirus*, Recombinação, Evolução Molecular, Filogenia

CONTENTS

1 GENERAL INTRODUCTION	11
2 LITERATURE REVIEW	13
2.1 Family <i>Geminiviridae</i>	13
2.2 Genus <i>Begomovirus</i>	15
2.3 Genetic diversity in viral populations and its implications for evolution and host interaction	16
2.4 Recombination in viral populations and its detection in nucleotide sequences.....	18
REFERENCES	20
CHAPTER 1	34
Abstract.....	35
Introduction	36
Material and Methods.....	37
Begomovirus species datasets.....	37
Sequence alignments.....	38
Recombination analysis	38
Phylogenetic analysis.....	38
Calculation of dissimilarities between ML phylogenetic trees.....	38
Rarefaction and comparisons between median wRF distances	39
Multidimensional scaling and ML tree clustering analysis	39
Results	39
Conflicting evolutionary histories within DNA-A of Begomovirus	39
IR- and mp-based trees are highly divergent.....	41
Pairwise wRF distances positively correlate with physical distances between genes on DNA-A	43
Incongruence across coding and non-coding regions for individual species.....	44
Long branches characterize trees with recombinant sequences.....	46

Mapping divergent evolutionary histories across genes	46
Discussion.....	49
References	52
Appendices	58

1 GENERAL INTRODUCTION

The genus *Begomovirus* comprises 445 distinct species, making it the largest genus in the family *Geminiviridae* (Fiallo-Olivé et al., 2021). Begomoviruses are transmitted by cryptic species of whiteflies within the *Bemisia tabaci* complex and infect a wide range of agricultural, medicinal, aromatic, ornamental, and wild plants (Alberter et al., 2005; Andrade et al., 2006; Avise et al., 2008; Awadalla, 2003; Bisaro, 2006; Bollyky et al., 1996; Boni et al., 2007). Infections caused by these viruses induce various symptoms, including mosaic, spots, yellowing, leaf curling, and stunting (Leke et al., 2015; Mansoor et al., 2003; Mondal et al., 2019). These symptoms strongly impact crop yield and global food security.

Begomoviruses contain genomes composed of one or two circular single-stranded DNA (ssDNA) molecules, known as DNA-A and DNA-B components, each approximately 2,600 nucleotides in length. In bipartite begomoviruses, the DNA-A component encodes essential proteins involved in replication (Laufs et al., 1995), suppression of gene silencing (Bisaro, 2006), and encapsidation of viral progeny (Harrison et al., 2002). The DNA-B component encodes essential proteins for viral movement within host plants (Sanderfoot and Lazarowitz, 1996). Thus, in bipartite begomoviruses, both DNA components are required for systemic infection and symptom induction (Fiallo-Olivé et al., 2021). Monopartite genomes consist of a single component homologous to the DNA-A of bipartite genomes and are often associated with satellite ssDNA molecules, which may significantly contribute to disease symptom induction (Zhou, 2013).

The frequent emergence of new viral diseases is mainly attributed to international trade, climate change, and the rapid evolutionary dynamics of viruses (Rubio et al., 2020). Begomoviruses are notable for their high potential for genetic variation, resulting from the accumulation of mutations, pseudorecombination, and frequent recombination (Lefeuvre and Moriones, 2015; Padidam et al., 1999; Sahu et al., 2018; Seal et al., 2006a). Recombination is a rapid mechanism for generating new genomes, providing adaptive advantages that promote the expansion of host range and, consequently, the emergence of new diseases and epidemics (Lefeuvre and Moriones, 2015). One of the most significant examples of epidemics caused by begomoviruses is the cassava mosaic disease (CMD), which involves a set of newly emerged species whose origin is attributed to inter- and intra-specific recombination events (Chikoti et al., 2019). A notable case was the emergence of the variant of the *East African cassava mosaic virus*-Uganda (EACMV-UG), an interspecific recombinant of EACMV and ACMV (Crespo-Bellido et al., 2021), which caused severe disease outbreaks in East and Central Africa in the

1990s (Pita et al., 2001; X. Zhou et al., 1997). Recombination provides an advantage to these viruses by allowing them to adapt to different environmental conditions, thus favoring the generation and emergence of new recombinants (Fiallo-Olivé and Navas-Castillo, 2023).

The high frequency of recombination in begomoviruses is associated with the rolling-circle replication (RCR) mechanism and the recombination-dependent replication (RDR) strategy, which involves the recovery of incomplete or damaged DNA molecules and their conversion into standard-length genomes through homologous recombination (Jeske et al., 2001). Frequent recombination can be explained as a necessary byproduct, since the RDR mechanism is a general process in the replication cycle of these viruses (Seal et al., 2006b). Thus, begomoviruses have the potential to generate many recombinant viruses (Lefeuvre and Moriones, 2015).

Considering their importance in the evolutionary history of begomoviruses, recombination events can be investigated using various computational tools that accurately detect their occurrence and provide valuable information on recombination's influence on the inference of phylogenetic relationships (Lewis-Rogers et al., 2004). However, detecting the occurrence of recombination is not a simple procedure and depends on various factors, such as the amount of divergence between analyzed sequences, the frequency and distribution of recombination events along the sequence, the effective population size, the mutation rate, and selection pressures (Posada, 2002; Posada and Crandall, 2001).

The wide range of available computational methods for detecting recombination events makes the choice of the most suitable method challenging (Martin et al., 2011b), as each method may perform differently depending on the type of dataset. Some methods tend to be less sensitive, providing only a general indication of the occurrence of recombination in a dataset, while more sensitive methods can identify recombinant sequences and provide detailed information on the positions of recombination breakpoints. Although accuracy is a crucial concern, there is no one-size-fits-all method for all datasets (Lemey and Posada, 2009).

Recombination Detection Program (RDP) is one of the most used tools to detect and characterize recombination events in aligned DNA sequence datasets (Martin et al., 2020). It offers an approach that includes seven different methods for detecting recombination events, allowing users to select the methods for their analysis. However, many researchers typically deem events valid when detected by at least four analytical methods, referred to here as the majority rule. Therefore, investigating whether this criterion is adequate to identify the most significant recombination events in a dataset is necessary.

2 LITERATURE REVIEW

2.1 Family *Geminiviridae*

Geminiviruses (family *Geminiviridae*) are among the most destructive plant viruses, causing diseases in major crops worldwide (Mansoor et al., 2003; Rojas et al., 2018). These viruses are transmitted by insect vectors, including various species within the families Cicadellidae and Membracidae, as well as cryptic species within the *Bemisia tabaci* complex. The family *Geminiviridae* is morphologically characterized by geminate icosahedral particles and has circular single-stranded DNA (ssDNA) genomes, each approximately 2,500 to 3,000 nucleotides in length (Böttcher et al., 2004; Krupovic et al., 2009; Zhang et al., 2001). These viruses infect monocotyledonous and dicotyledonous plants, causing significant productivity losses in agricultural crops (Gutierrez, 2000), compromising human food security (Gupta et al., 2021; Harrison and Robinson, 1999). Symptoms following infection typically range from mild or asymptomatic to severe manifestations, including plant stunting, inability of reproductive organs to develop normally, leaf wrinkling, yellowing, distortion, dwarfism, curled and deformed leaves with mosaic or streak patterns (Inoue-Nagata et al., 2016; Mansoor et al., 2003; Rojas et al., 2018).

This family encompasses approximately 520 species, as listed in the International Committee on Taxonomy of Viruses (ICTV) (Fiallo-Olivé et al., 2021). These species are classified into 14 distinct genera: *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Maldovirus*, *Mastrevirus*, *Mulcrilevirus*, *Opunvirus*, *Topilevirus*, *Topocuvirus*, and *Turncurtovirus*. This classification is based on aspects such as genome organization, host range, phylogenetic relationships, insect vectors, and pairwise genomic sequence identities (Fauquet and Stanley, 2003; Gutierrez, 2000; Mansoor et al., 2003; Roumagnac et al., 2022; Varsani et al., 2014; Zerbini et al., 2017). The development of resistance to insecticides and the emergence of new vector genotypes, especially whiteflies, have facilitated the spread of geminiviruses to other regions and the formation of emergent viral combinations in disease complexes. These characteristics have enabled these viruses to rapidly adapt to different hosts and environments (Navas-Castillo et al., 2011), resulting in the global spread of geminiviruses and posing a significant threat to food security in agricultural-producing countries (Gilbertson et al., 2015).

Geminiviruses replicate their compact genomes through double-stranded (ds)DNA intermediates within the nuclei of infected plant cells via a rolling-circle mechanism (Saunders et al., 1991; Stenger et al., 1991). These features differentiate them from most plant viruses,

which typically have RNA genomes and/or replication intermediates. These viruses encode a limited number of proteins for replication and rely entirely on the plant's DNA replication machinery (Wu et al., 2022). The infection cycle begins when viruliferous insect vectors feed on plants and introduce viral particles (Gutierrez, 1999; Hanley-Bowdoin et al., 2013; Yadava et al., 2010). The viral ssDNA is spontaneously released from the virion upon entry into the plant cell nucleus, undergoing decapsidation (Hanley-Bowdoin et al., 2013). Replication and transcription occur within the nucleus through host-encoded DNA/RNA polymerases, with DNA polymerase α generating a double-stranded (dsDNA) intermediate and DNA polymerase δ mediating the production of new viral genome copies (Hanley-Bowdoin et al., 1999; Saunders et al., 1992; Wu et al., 2021). The involvement of these DNA polymerases aligns with previous findings showing that treatments with aphidicolin, an inhibitor of DNA polymerases α , δ , and ϵ , hinder geminivirus accumulation in plants (Wu et al., 2021). Additionally, geminiviruses utilize an alternative replication mechanism known as recombination-dependent replication (RDR) (Alberter et al., 2005; Jeske et al., 2001; Jovel et al., 2007; Morilla et al., 2006; Preiss and Jeske, 2003).

The dsDNA replicative intermediate is recognized by the virus-encoded replication-associated protein (rep), which binds to a specific sequence characterized by a hairpin-like structure containing an invariant nine-nucleotide sequence (nonanucleotide, TAATATT//AC) located in the intergenic region (IR) (Fontes et al., 1992; Lazarowitz et al., 1992). Subsequently, rep recruits the cellular DNA replication machinery and initiates strand cleavage and rolling-circle replication. Replication is carried out by host DNA polymerases and, then, rep reconnects the displaced strand, releasing a new copy of the viral single-stranded DNA (ssDNA) genome (Bonnamy et al., 2023). The ssDNA can serve as a template for additional rounds of replication, spread to adjacent cells, or be encapsidated into new particles, becoming available for acquisition by an insect vector (Hanley-Bowdoin et al., 1999; Zhao et al., 2019). Rep is the only virus-encoded protein essential for replication. In addition to functioning as a helicase and possessing DNA cleavage and ligation activities, rep participates in reprogramming the cell cycle to induce the expression of DNA-dependent DNA polymerase (Hanley-Bowdoin et al., 1999; Ruhel and Chakraborty, 2019).

Inside the plant, the infection spreads as viral DNA moves out of the nucleus toward neighboring cells or the phloem, facilitated by two viral movement proteins: nuclear shuttle protein (NSP) and movement protein (mp) (Hanley-Bowdoin et al., 2013; Noueiry et al., 1994; Sanderfoot and Lazarowitz, 1996).

2.2 Genus *Begomovirus*

The genus *Begomovirus* comprises the most diverse group within the family *Geminiviridae*, encompassing a total of 445 distinct species (Fiallo-Olivé et al., 2021). Begomoviruses are classified into two groups based on geographical distribution and phylogenetic relationships. The first group includes viruses native to the Old World (OW: Europe, Africa, Asia, and Oceania), while the second group includes viruses native to the New World (NW: the Americas) (Harrison and Robinson, 1999; Mondal et al., 2019; Nawaz-ul-Rehman and Fauquet, 2009; Padidam et al., 1999). Furthermore, begomoviruses are classified based on genome organization into monopartite genomes, comprising a single-stranded DNA (ssDNA) molecule, and bipartite genomes, composed of two ssDNA molecules (referred to as DNA-A and DNA-B). Genomes encode four to eight bidirectional, partially overlapping Open Reading Frames (ORFs) (Medina-Puche et al., 2021; Roshan et al., 2017). NW begomoviruses predominantly possess bipartite genomes (Zerbini et al., 2017), while OW begomoviruses may exhibit either monopartite or bipartite genomes. Monopartite genomes consist of a single component homologous to the DNA-A of bipartite genomes and are often associated with virus-like ssDNA satellite molecules, known as alpha, beta, and gamma satellites, which significantly intensify *Begomovirus*-induced symptoms (Ferro et al., 2017; Gupta et al., 2021; Paprotka et al., 2010).

The DNA-A or DNA-like segment of begomoviruses typically encodes five to seven proteins. The rep protein performs several replication-essential functions, while the coat protein (CP) is a multifunctional protein required for viral encapsidation, self-association, vector transmission, and nuclear-cytoplasmic movement in monopartite viruses (Guerrero et al., 2020; Harrison et al., 2002; Mondal et al., 2019; Padidam et al., 1996). The V2 protein (or AV2 in bipartite begomoviruses) serves as a movement protein (mp) and a suppressor of post-transcriptional gene silencing (Chowda-Reddy et al., 2008). The transcriptional activator protein (TrAP or trap) interferes with transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS), functioning as a transcription factor essential CP and NSP protein expression in bipartite begomoviruses (Chowda-Reddy et al., 2009; Guerrero et al., 2020; Haley et al., 1992; Hanley-Bowdoin et al., 1999; Sunter and Bisaro, 1997; Trinks et al., 2005; Wang et al., 2005). The replication enhancer protein (ren, also known as C3) stimulates replication and enhances viral DNA accumulation by selectively recruiting DNA polymerase δ for the synthesis of new copies of the geminiviral ssDNA genome (Sunter and Bisaro, 1997; Wu et al., 2021). The C4 protein (or ac4) functions as an RNA silencing

suppressor (Fondong et al., 2007; Vanitharani et al., 2004). The DNA-B segment of bipartite begomoviruses encodes two proteins mediating cell-to-cell and long-distance movement, namely NSP and MP (Hanley-Bowdoin et al., 2013; Rojas et al., 2005).

In bipartite begomoviruses, the DNA-A and DNA-B segments share a common region (CR) of approximately 200 highly conserved nucleotides (> 90% identity) (Hanley-Bowdoin et al., 1999; Rojas et al., 2005). This region contains sequence elements involved in viral genome replication and transcription. The nonanucleotide sequence ('TAATATTAC') is mapped within the CR, functioning as a DNA cleavage site and an initiation point for the replication process (Bonnamy et al., 2023; Fiallo-Olivé et al., 2021).

Begomoviruses are transmitted persistent, non-propagative, and circulative manner by cryptic species within the *Bemisia tabaci* complex. Notably, recent studies suggest that *Tomato yellow leaf curl virus* (TYLCV) may replicate within the insect vector, although it remains unclear whether this phenomenon is exclusive to TYLCV or applicable to all begomoviruses (He et al., 2020).

The *Bemisia tabaci* complex is widely distributed worldwide and has significantly contributed to the efficient dissemination of begomoviruses and the emergence of epidemics. The first reports of infections by begomoviruses coincided with the global spread of *B. tabaci* species, such as Middle East-Asia Minor 1 (MEAM1) and Mediterranean (MED); these species introduced begomoviruses to previously unaffected regions, including Brazil, now recognized as a hotspot of *Begomovirus* diversity (Jones, 2003; Navas-Castillo et al., 2011; Ribeiro et al., 1998). These viruses can influence the feeding preference and behavior of whiteflies. Non-viruliferous whiteflies tend to prefer virus-infected plants, whereas viruliferous whiteflies tend to feed predominantly on non-infected plants (Zhao et al., 2022).

2.3 Genetic diversity in viral populations and its implications for evolution and host interaction

Plant viruses are infectious microscopic entities that hijack the cellular machinery of their host plants to replicate and disseminate their genomes. This intricate virus-host interaction drives genetic diversity (García-Arenal and Zerbini, 2019), influencing the evolutionary trajectories of both organisms. Understanding the mechanisms behind the genetic diversification is crucial for unraveling the complex dynamics of plant virus-host interactions (García-Andrés et al., 2007) and developing effective disease management strategies.

Mutation, a random alteration of nucleotides within the viral genome, is at the forefront of this genetic diversification. These mutations, often resulting from errors during DNA replication, generate new genetic variants in viral populations (Acosta-Leal et al., 2011; Duffy and Holmes, 2009; Lima et al., 2017). These variants serve as the raw material for natural selection, driving the evolution and adaptation of the virus. Recombination involves the exchange of sequence fragments between different viral strains or species. This evolutionary mechanism reshuffles the viral genome, producing novel gene combinations. Genetic mosaics can confer new traits on the virus, potentially enhancing its virulence, expanding its host range, or enabling it to evade host defenses. These exchanges contribute to the expansion of genetic diversity (Posada et al., 2002).

Pseudorecombination (also known as reassortment) consists of the exchange of complete genome segments between viruses with segmented genomes. This mechanism generates hybrid viruses, especially when high genetic compatibility exists within replication-mediating regions of each genomic segment (Andrade et al., 2006; Gilbertson et al., 1993; Lima et al., 2013; Paplomatas et al., 1994; Sung and Coutts, 1995). Natural selection functions on the broad spectrum of genetic variants resulting from mutation, recombination, and pseudorecombination, favoring those that enhance viral fitness. Consequently, natural selection gradually influences the genetic composition of a viral population, enabling viruses to adapt to their environment. Furthermore, genetic drift is a stochastic force that influences viral genetic diversity. Random fluctuations in allele frequencies can lead to the fixation or loss of genetic variants, regardless of their selective value. Migration, the movement of viruses between populations, introduces new genetic variants into the local gene pool. This influx of genetic diversity can influence virus-host interactions and drive new evolutionary pathways.

Understanding the intricacies of genetic diversity in plant viruses is crucial for effective disease management and agricultural productivity. Insights into these mechanisms can guide the development of control strategies, such as selecting resistant crop varieties. A deeper investigation of plant virus genetics enhances the understanding of the dynamic nature of these microscopic entities and facilitates strategies to mitigate their impact and protect crops and ecosystems.

2.4 Recombination in viral populations and its detection in nucleotide sequences

The classification of different types of recombination is based on the structure of crossover sites (Galli and Bukh, 2014). Homologous recombination occurs between genetically similar sequences sharing a common evolutionary origin, while non-homologous recombination, also known as illegitimate recombination, involves the joining of non-homologous genetic fragments. Non-homologous recombination can lead to duplications, deletions, or insertions of long sequences within viral genomes (Galli and Bukh, 2014; Lai, 1992; Martin et al., 2011b; Worobey and Holmes, 1999).

Research on genetic information exchange through recombination has revealed its potential to enhance adaptability in viral populations and other pathogens. Recombination is linked to the emergence of new viruses, expansion of host ranges, increased virulence, modified vector transmission specificity, the emergence of drug-resistant variants, and adaptive responses to environmental selective pressures (Awadalla, 2003; Cui et al., 2022; Gibbs and Weiller, 1999; Lv et al., 2019; Martin et al., 2011b; Moya et al., 2004; Nora et al., 2007; Pasick et al., 2005; Simon-Loriere and Holmes, 2011; Suarez et al., 2004). The relevance of recombination in genome evolution has led researchers to focus on detecting recombination events in viral genomes. Their primary objectives are to address three fundamental questions: Does recombination occur throughout the genome? Where are the recombination breakpoints located? Can the phylogenetic history be reconstructed to identify parental viruses? Over the past decades, various computational methods have been developed to address these and related questions, enabling precise detection, quantification, and characterization of recombination events from aligned DNA and RNA sequences.

Posada (2002) classified recombination detection methods into four categories: distance, phylogenetic, compatibility, and substitution distribution methods. Phylogenetic methods infer recombination events when phylogenies of different genome fragments result in distinct tree topologies. Additionally, these methods identify clusters of orthologous genes, which are genes from different species sharing a common ancestor. Phylogenetic methods are widely used for detecting recombination events and vary in detection capability (Awadalla, 2003; Posada, 2002; Posada et al., 2002; Posada and Crandall, 2001).

Notably, despite the variety of methods available for detecting recombination events, their performance varies depending on the specific characteristics of the dataset being analyzed (Pérez-Losada et al., 2015). Various non-parametric methods have been evaluated for recombination detection accuracy and sensitivity (e.g., false positives) (Posada and Crandall,

2001a; Awadalla, 2003). However, their performance varies with the extent of recombination, genetic diversity, and site variation rate; consequently, no single method has proven superior for detecting recombination under all conditions (Posada, 2002; Posada and Crandall, 2001).

One of the main challenges of traditional phylogenetic tree reconstruction methods is the assumption that all analyzed sequences share a single evolutionary history, represented by a single bifurcating phylogenetic tree (Martin et al., 2011b; Posada et al., 2002; Posada and Crandall, 2001). However, the presence of recombination events often indicates the existence of distinct, alternative evolutionary histories. For instance, a sequence exhibiting recombination signals in specific alignment segments may yield multiple phylogenetic trees, with different regions reflecting distinct evolutionary histories, thus affecting branching patterns and branch lengths. Consequently, failing to detect recombination events can lead to inaccurate phylogenetic tree construction and erroneous conclusions (Posada et al., 2002; Mikkelsen and Hein, 2000; Wiuf et al., 2001).

Martin and Rybicki (2000) developed the Recombination Detection Program (RDP), an accessible and user-friendly tool that facilitates interactive recombination investigation. Currently in its fifth version (RDP5), this program detects and characterizes recombination events throughout the evolutionary history of the analyzed sequences. In addition to the original RDP method (Martin and Rybicki, 2000), RDP5 includes six other methods: BootScan/RecScan (Martin et al., 2005; Salminen et al., 1995), GENECONV (Padidam et al., 1999; Sawyer, 1989), MaxChi (Posada and Crandall, 2001; Smith, 1992), Chimaera (Posada and Crandall, 2001), SiScan (Gibbs et al., 2000), and 3Seq (Boni et al., 2007).

REFERENCES

- 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*, 1136–1148.
- Alberter, B., Ali Rezaian, M., Jeske, H. (2005). Replicative intermediates of *Tomato leaf curl virus* and its satellite DNAs. *Virology*, 441–448. <https://doi.org/10.1016/j.virol.2004.10.043>
- 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*, 3687–3696. <https://doi.org/10.1099/vir.0.82279-0>
- Avise, J.C., Robinson, T.J. (2008). Hemiplasy: A New Term in the Lexicon of Phylogenetics. *Systematic Biology*, 503–507. <https://doi.org/10.1080/10635150802164587>
- Awadalla, P. (2003). The evolutionary genomics of pathogen recombination. *Nature Reviews Genetics*, 50–60. <https://doi.org/10.1038/nrg964>
- Bisaro, D.M. (2006). Silencing suppression by geminivirus proteins. *Virology*, 344, 158–168. <https://doi.org/10.1016/j.virol.2005.09.041>
- Bollyky, P.L., Rambaut, A., Harvey, P.H., Holmes, E.C. (1996). Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution*, 42, 97–102. <https://doi.org/10.1007/BF02198834>
- Boni, M.F., Posada, D., Feldman, M.W. (2007). An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. *Genetics*, 176, 1035–1047. <https://doi.org/10.1534/genetics.106.068874>
- 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, 1–21. <https://doi.org/10.1128/mbio.01692-23>
- Böttcher, B., Unseld, S., Ceulemans, H., Russell, R.B., Jeske, H. (2004). Geminate Structures of *African Cassava Mosaic Virus*. *Journal of Virology*, 78, 6758–6765. <https://doi.org/10.1128/jvi.78.13.6758-6765.2004>
- Brandes, N., Linial, M. (2016). Gene overlapping and size constraints in the viral world. *Biology Direct*, 11, 26. <https://doi.org/10.1186/s13062-016-0128-3>
- Brumme, Z.L., Brumme, C.J., Heckerman, D., Korber, B.T., Daniels, M., Carlson, J., Kadie, C., Bhattacharya, T., Chui, C., Szinger, J., Mo, T., Hogg, R.S., Montaner, J.S.G., Frahm, N., Brander, C., Walker, B.D., Harrigan, P.R. (2007). Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. *PLoS Pathogens*, 3. <https://doi.org/10.1371/journal.ppat.0030094>

- Canty A., Ripley, B.D. (2021). boot: Bootstrap R (S-Plus) Functions.
- Chikoti, P.C., Mulenga, R.M., Tembo, M., Sseruwagi, P. (2019). Correction to: Cassava mosaic disease: a review of a threat to cassava production in Zambia. *Journal of Plant Pathology*, 101, 467–477. <https://doi.org/10.1007/s42161-019-00267-w>
- 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, 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, 270–278. <https://doi.org/10.1016/j.virusres.2009.07.022>
- Crespo-Bellido, A., Hoyer, J.S., Dubey, D., Jeannot, R.B., Duffy, S. (2021). Interspecies Recombination Has Driven the Macroevolution of Cassava Mosaic Begomoviruses. *Journal of Virology*, 95. <https://doi.org/10.1128/jvi.00541-21>
- Cui, X.Y., Xia, D.S., Huang, X.Y., Tian, X.X., Wang, T., Yang, Y.B., Wang, G., Wang, H.W., Sun, Y., Xiao, Y.H., Tian, Z.J., Cai, X.H., An, T.Q. (2022). Recombinant characteristics, pathogenicity, and viral shedding of a novel PRRSV variant derived from twice inter-lineage recombination. *Veterinary Microbiology*, 271, 1–12. <https://doi.org/10.1016/j.vetmic.2022.109476>
- 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, 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. <https://doi.org/10.1038/s41467-022-34630-w>
- Edwards, K.F., Steward, G.F., Schvarcz, C.R. (2021). Making sense of virus size and the tradeoffs shaping viral fitness. *Ecology Letters*, 24, 363–373. <https://doi.org/10.1111/ele.13630>
- Fauquet, C.M., Stanley, J. (2003). Geminivirus classification and nomenclature: progress and problems. *Annals of Applied Biology*, 142, 165–189.
- Fernandes, F.R., Albuquerque, L.C., de Oliveira, C.L., Cruz, A.R.R., da Rocha, W.B., Pereira, T.G., Naito, F.Y.B., De Dias, N.M., Nagata, T., Faria, J.C., Zerbini, F.M., Aragão, F.J.L., Inoue-Nagata, A.K. (2011). Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting *Euphorbia heterophylla* plants. *Archives of Virology*, 156. <https://doi.org/10.1007/s00705-011-1070-4>
- 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

- alphasatellites. *Annals of Applied Biology*, 170, 204–218. <https://doi.org/10.1111/aab.12329>
- 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. <https://doi.org/10.1099/jgv.0.001696>
- Fiallo-Olivé, E., Navas-Castillo, J. (2023). Begomoviruses: what is the secret(s) of their success?. *Trends in Plant Science*, 28, 715–727. <https://doi.org/10.1016/j.tplants.2023.01.012>
- Fiallo-Olivé, E., Pan, L.L., Liu, S.S., Navas-Castillo, J. (2020). Transmission of begomoviruses and other whitefly-borne viruses: Dependence on the vector species. *Phytopathology*, 110, 10–17. <https://doi.org/10.1094/PHYTO-07-19-0273-FI>
- 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, 380–391. <https://doi.org/10.1094/MPMI-20-4-0380>
- Fontes, E.P.B., Luckow, V.A., Hanley-Bowdoin, L. (1992). A Geminivirus Replication Protein is a Sequence-Specific DNA Binding Protein. *Plant Cell*, 4, 597–608.
- Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., Michalakakis, Y. (2005). Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biology*, 3. <https://doi.org/10.1371/journal.pbio.0030089>
- Galli, A., Bukh, J. (2014). Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol*, 22, 354–364. <https://doi.org/10.1016/j.tim.2014.02.005>
- García-Andrés, S., Accotto, G.P., Navas-Castillo, J., Moriones, E. (2007). Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology*, 359, 302–312. <https://doi.org/10.1016/j.virol.2006.09.030>
- García-Andrés, S., Monci, F., Navas-Castillo, J., Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology*, 350. <https://doi.org/10.1016/j.virol.2006.02.028>
- García-Arenal, F., Zerbini, F.M. (2019). Life on the Edge: Geminiviruses at the Interface between Crops and Wild Plant Hosts. *Annual Review of Virology*, 6, 411–433. <https://doi.org/10.1146/annurev-virology-092818-015536>
- Gibbs, M.J., Armstrong, J.S., Gibbs, A.J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, 16, 573–582. <https://doi.org/10.1093/bioinformatics/16.7.573>

- Gibbs, M.J., Weiller, G.F. (1999). Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proceedings of the National Academy of Sciences (PNAS)*, 96(14), 8022–8027. <https://doi.org/10.1073/pnas.96.14.8022>
- 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, 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, 23–31.
- Guerrero, J., Regedanz, E., Lu, L., Ruan, J., Bisaro, D.M., Sunter, G. (2020). Manipulation of the Plant Host by the Geminivirus AC2/C2 Protein, a Central Player in the Infection Cycle. *Frontiers in Plant Science*, 11, 1–18. <https://doi.org/10.3389/fpls.2020.00591>
- Gupta, N., Reddy, K., Bhattacharyya, D., Chakraborty, S. (2021). Plant responses to geminivirus infection: guardians of the plant immunity. *Virology journal*, 18 (1), 1–25. <https://doi.org/10.1186/s12985-021-01612-1>
- Gutierrez, C. (2000). Geminiviruses and the plant cell cycle. *Plant Molecular Biology*, 43, 763–772.
- Gutierrez, C. (1999). Geminivirus DNA replication. *Cellular and Molecular Life Sciences*, 56, 313–329.
- 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, 905–909.
- Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology*, 11, 777–788. <https://doi.org/10.1038/nrmicro3117>
- Hanley-Bowdoin, L., Settlege, 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., Robinson, D.J. (1999). Natural Genomic and Antigenic Variation in Whitefly-Transmitted Geminiviruses (Begomoviruses). *Annual Review of Phytopathology*, 37, 369–398.
- Harrison, B.D., Swanson, M.M., Fargette, D. (2002). Begomovirus coat protein: serology, variation and functions. *Physiological and Molecular Plant Pathology*, 60, 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, 16928–16937. <https://doi.org/10.1073/pnas.1820132117>
- Inoue-Nagata, A.K., Lima, M.F., Gilbertson, R.L. (2016). A review of geminivirus (begomovirus) diseases in vegetables and other crops in Brazil: current status and approaches for management. *Horticultura Brasileira*, 34, 8–18. <https://doi.org/10.1590/S0102-053620160000100002>
- Jeske, H., Lütgemeier, M., Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of *Abutilon* mosaic virus. *The EMBO journal*, 20 (21), 6158–6167.
- Jombart, T., Kendall, M., Almagro-Garcia, J., Colijn, C. (2017). Treespace: Statistical exploration of landscapes of phylogenetic trees. *Molecular Ecology Resources*, 17, 1385–1392. <https://doi.org/10.1111/1755-0998.12676>
- Jones, D.R. (2003). Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology*, 109, 195–219.
- Jovel, J., Preiß, W., Jeske, H. (2007). Characterization of DNA intermediates of an arising geminivirus. *Virus Research*, 130, 63–70. <https://doi.org/10.1016/j.virusres.2007.05.018>
- 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, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kistler, K.E., Bedford, T. (2021). Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. *eLife*, 10. <https://doi.org/10.7554/eLife.64509>
- Krupovic, M., Ravanti, J.J., Bamford, D.H. (2009). Geminiviruses: A tale of a plasmid becoming a virus. *BMC Ecology and Evolution*, 9, 1–11. <https://doi.org/10.1186/1471-2148-9-112>
- Lai, M.M. (1992). RNA recombination in animal and plant viruses. *Microbiological reviews*, 56, 61–79. <https://doi.org/10.1128/mr.56.1.61-79.1992>
- Lam, H.M., Ratmann, O., Boni, M.F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Molecular Biology and Evolution*, 35. <https://doi.org/10.1093/molbev/msx263>
- Langedijk, A.C., Vrancken, B., Lebbink, R.J., Wilkins, D., Kelly, E.J., ... Madhi, S.A. (2024). The genomic evolutionary dynamics and global circulation patterns of respiratory syncytial virus. *Nature Communications*, 15(3083), 1-12. <https://doi.org/10.1038/s41467-024-47118-6>
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30, 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>

- Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., Gronenbom, B., 1995. Geminivirus replication: Genetic and biochemical characterization of Rep protein function, a review. *Biochimie*, 77, 765–773.
- Lazarowitz, S.G., Wu, L.C., Rogerqb, S.G., Scott Elmerb, J. (1992). Sequence-Specific Interaction with the Viral AL1 Protein Identifies a Geminivirus DNA Replication Origin. *Plant Cell*, 4, 799–809.
- Lefeuvre, P., Martin, D.P., Elena, S.F., Shepherd, D.N., Roumagnac, P., Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology*, 17, 632–644. <https://doi.org/10.1038/s41579-019-0232-3>
- 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>
- Leke, W.N., Mignouna, D.B., Brown, J.K., Kvarnheden, A. (2015). Begomovirus disease complex: emerging threat to vegetable production systems of West and Central Africa. *Agriculture & Food Security*, 4, 1–14. <https://doi.org/10.1186/s40066-014-0020-2>
- Lemey, P., Posada, D. (2009). Introduction to recombination detection, in: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, pp. 362–380. <https://doi.org/10.1017/CBO9780511819049>
- Lewis-Rogers, N., Crandall, K.A., Posada, D. (2004). Evolutionary analyses of genetic recombination. *Dynamical Genetics*, pp. 49–78.
- 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, 1–14. <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, 418–431. <https://doi.org/10.1099/vir.0.047241-0>
- Lozano, G., Trenado, H.P., Valverde, R.A., Navas-Castillo, J. (2009). Novel begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and *Ipomoea indica*: Taxonomic and phylogenetic implications. *Journal of General Virology*, 90. <https://doi.org/10.1099/vir.0.012542-0>
- Lv, L., Li, T., Hu, M., Deng, J., Liu, Y., Xie, Q., Shao, H., Ye, J., Qin, A. (2019). A recombination efficiently increases the pathogenesis of the novel K subgroup of avian leukosis virus. *Veterinary Microbiology*, 231, 214–217. <https://doi.org/10.1016/j.vetmic.2019.03.021>

- Maddison, W.P., Knowles, L.L. (2006). Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, 55, 21–30. <https://doi.org/10.1080/10635150500354928>
- Mahmoudabadi, G., Phillips, R. (2018). A comprehensive and quantitative exploration of thousands of viral genomes. *eLife*, 7. <https://doi.org/10.7554/eLife.31955>
- Mansoor, S., Briddon, R.W., Zafar, Y., Stanley, J. (2003). Geminivirus disease complexes: an emerging threat. *Trends in Plant Science*, 8, 128–134. [https://doi.org/10.1016/S1360-1385\(03\)00007-4](https://doi.org/10.1016/S1360-1385(03)00007-4)
- Mar, T.B., Mendes, I.R., Lau, D., Fiallo-Olivé, E., Navas-Castillo, J., Alves, M.S., Zerbini, F.M. (2017). Interaction between the new world begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite: Effects on infection and transmission by the whitefly *Bemisia tabaci*. *Journal of General Virology*, 98, 1552–1562. <https://doi.org/10.1099/jgv.0.000814>
- Martin, D., Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 16, 562–563.
- Martin, D.P., Biagini, P., Lefevre, P., Golden, M., Roumagnac, P., Varsani, A. (2011a). Recombination in eukaryotic single stranded DNA viruses. *Viruses*, 3, 1699–1738. <https://doi.org/10.3390/v3091699>
- Martin, D.P., Lefevre, P., Varsani, A., Hoareau, M., Semegni, J.Y., Dijoux, B., Vincent, C., Reynaud, B., Lett, J.M. (2011b). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathogens*, 7. <https://doi.org/10.1371/journal.ppat.1002203>
- Martin, D.P., Lemey, P., Posada, D. (2011c). Analysing recombination in nucleotide sequences. *Molecular Ecology Resources*, 11, 943–955. <https://doi.org/10.1111/j.1755-0998.2011.03026.x>
- Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1. <https://doi.org/10.1093/ve/vev003>
- Martin, D.P., Van der Walt, E., Posada, D., Rybicki, E.P. (2005). The evolutionary value of recombination is constrained by genome modularity. *PLoS genetics*, 1(4). <https://doi.org/10.1371/journal.pgen.0010051>
- Martin D.P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T., ... Murrell, B. (2020). RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*, 7, 1–3. <https://doi.org/10.1093/ve/veaa087>
- Maxmen, A. (2021). One million coronavirus sequences: popular genome site hits mega milestone. *Nature*, 593, 21–21. <https://doi.org/10.1038/d41586-021-01069-w>

- Medina-Puche, L., Orílio, A. F., Zerbini, F. M., Lozano-Durán, R. (2021). Small but mighty: Functional landscape of the versatile geminivirus-encoded C4 protein. *PLoS Pathogens*, 17, 1–15. <https://doi.org/10.1371/journal.ppat.1009915>
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30, 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- 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, 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, 74–83. <https://doi.org/10.1007/s13337-019-00524-7>
- Morilla, G., Castillo, A.G., Preiss, W., Jeske, H., Bejarano, E.R. (2006). A Versatile Transreplication-Based System To Identify Cellular Proteins Involved in Geminivirus Replication. *Journal of Virology*, 80, 3624–3633. <https://doi.org/10.1128/jvi.80.7.3624-3633.2006>
- Moya, A., Holmes, E.C., González-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology*, 2, 279–288. <https://doi.org/10.1038/nrmicro863>
- Munoz, C., Zhou, X., Otim-Nape, G.W., Liu, Y., Robinson, D.J., Calvert, L., Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, 78, 2101–2111. <https://doi.org/10.1099/0022-1317-78-8-2101>
- Navas-Castillo, J., Fiallo-Olivé, E., Sánchez-Campos, S. (2011). Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology*, 49, 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, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., Hance, A.J. (2007). Contribution of Recombination to the Evolution of Human Immunodeficiency Viruses Expressing Resistance to Antiretroviral Treatment. *Journal of Virology*, 81, 7620–7628. <https://doi.org/10.1128/jvi.00083-07>

- Noueiry, A.O., Lucas, W.J., Gilbertson, R.L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell*, 76, 925–932. [https://doi.org/10.1016/0092-8674\(94\)90366-2](https://doi.org/10.1016/0092-8674(94)90366-2)
- Nouri, S., Arevalo, R., Falk, B.W., Groves, R.L. (2014). Genetic Structure and Molecular Variability of Cucumber mosaic virus Isolates in the United States. *PLoS One*, 9(5). <https://doi.org/10.1371/journal.pone.0096582>
- Opiyo, S.A., Ateka, E.M., Owuor, P.O., Manguro, L.O.A., Karuri, H.W. (2010). Survey of sweet potato viruses in Western Kenya and detection of Cucumber mosaic virus. *Journal of Plant Pathology*, 92, 797-801.
- 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, 390–404.
- Padidam, M., Sawyer, S., Fauquet, C.M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, 265, 218–225.
- Paplomatas, E.J., Patel, V.P., Hou, Y.M., Noueiry, A.O., Gilbertson, R. L. (1994). Molecular characterization of a new sap-transmissible bipartite genome geminivirus infecting tomatoes in Mexico. *Phytopathology*, 84, 1215–1223.
- 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, 148–157. <https://doi.org/10.1016/j.virol.2010.05.003>
- Pasick, J., Handel, K., Robinson, J., Copps, J., Ridd, D., Hills, K., Kehler, H., Cottam-Birt, C., Neufeld, J., Berhane, Y., Czub, S. (2005). Intersegmental recombination between the haemagglutinin and matrix genes was responsible for the emergence of a highly pathogenic H7N3 avian influenza virus in British Columbia. *Journal of General Virology*, 86, 727–731. <https://doi.org/10.1099/vir.0.80478-0>
- 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>
- Pita, J.S., Fondong, V.N., Sangare, 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, 655–665.
- Posada, D., (2002). Evaluation of Methods for Detecting Recombination from DNA Sequences: Empirical Data. *Molecular Biology and Evolution*, 19, 708–717.
- Posada, D., Crandall, K.A., (2002). The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of Molecular Evolution*, 54, 396–402. <https://doi.org/10.1007/s00239-001-0034-9>

- Posada, D., Crandall, K.A., (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences* 98, 13757–13762.
- Posada, D., Crandall, K.A., Holmes, E.C. (2002). Recombination in evolutionary genomics. *Annu Rev Genet* 36, 75–97. <https://doi.org/10.1146/annurev.genet.36.040202.111115>
- Preiss, W., Jeske, H. (2003). Multitasking in Replication Is Common among Geminiviruses. *Journal of Virology*, 77, 2972–2980. <https://doi.org/10.1128/jvi.77.5.2972-2980.2003>
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rambaut, A., Holmes, E.C., O’Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5, 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Revell, L.J. (2024). phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, 12. <https://doi.org/10.7717/peerj.16505>
- Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *Journal of General Virology*, 77, 1953–1965. <https://doi.org/10.1099/0022-1317-77-8-1953>
- Ribeiro S.G., De Ávila, A.C., Bezerra, I.C., Fernandes, J.J., Faria, J.C., Lima, M., 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. <https://doi.org/10.1094/PDIS.1998.82.7.830C>
- Robinson, D.F., Foulds, L.R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53, 131–147. [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2)
- 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, 361–394. <https://doi.org/10.1146/annurev.phyto.43.040204.135939>
- Rojas, M.R., Macedo, M.A., Maliano, M.R., Soto-Aguilar, M., Souza, J.O., ... Gilbertson, R.L. (2018). World Management of Geminiviruses. *Annual Review of Phytopathology*, 56, 637–677. <https://doi.org/10.1146/annurev-phyto>
- Roshan, P., Kulshreshtha, A., Hallan, V. (2017). Genome Organization of Begomoviruses, in: *Begomoviruses: Occurrence and Management in Asia and Africa*. Springer Singapore, 11–32. https://doi.org/10.1007/978-981-10-5984-1_2
- Roumagnac, P., Lett, J.M., Fiallo-Olivé, E., Navas-Castillo, J., Zerbini, F.M., Martin, D.P., Varsani, A. (2022). Establishment of five new genera in the family *Geminiviridae*: *Citlodavirus*, *Maldovirus*, *Mulcrilevirus*, *Opunvirus*, and *Topilevirus*. *Archives of Virology*, 167, 695–710. <https://doi.org/10.1007/s00705-021-05309-2>

- Rubio, L., Galipienso, L., Ferriol, I. (2020). Detection of Plant Viruses and Disease Management: Relevance of Genetic Diversity and Evolution. *Frontiers in Plant Science*, 11, 1–23. <https://doi.org/10.3389/fpls.2020.01092>
- Ruhel, R., Chakraborty, S. (2019). Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease*, 30, 66–73. <https://doi.org/10.1007/s13337-018-0458-0>
- Saeed, F., Sattar, M.N., Hameed, U., Ilyas, M., Haider, M.S., Hamza, M., Mansoor, S., Amin, I. (2018). Infectivity of okra enation leaf curl virus and the role of its V2 protein in pathogenicity. *Virus Research*, 255. <https://doi.org/10.1016/j.virusres.2018.07.007>
- Sahu, A.K., Verma, R.K., Gaur, R.K., Sanan-Mishra, N. (2018). Complexity and recombination analysis of novel begomovirus associated with Spinach yellow vein disease in India. *Plant Gene*, 13, 42–49. <https://doi.org/10.1016/j.plgene.2018.01.001>
- Salminen, M.O., Carr, J.K., Burke, D.S., Mccutchan, F.E. (1995). Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Research and Human Retroviruses*, 11, 1423–1425. <https://doi.org/10.1089/aid.1995.11.1423>
- Sanderfoot, A.A., Lazarowitz, S. G. (1996). Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. *Trends in cell biology*, 6 (9), 353–358.
- Sanjuán, R., Domingo-Calap, P. (2021). Genetic Diversity and Evolution of Viral Populations. *Encyclopedia of Virology*, 53–61. <https://doi.org/10.1016/B978-0-12-809633-8.20958-8>
- Saunders, K., Lucy, A., Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Research*, 20, 6311–6315.
- Saunders, K., Lucy, A., Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Research*, 19, 2325–2330.
- Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Molecular Biology and Evolution*, 6, 526–538. <https://doi.org/10.1093/oxfordjournals.molbev.a040567>
- Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 48. <https://doi.org/10.1093/nar/gkz956>
- Schierup, M. H., Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics*, 156(2), 879–891. <https://doi.org/10.1093/genetics/156.2.879>
- Scrucca, L., Fraley, C., Murphy, T.B., Adrian E., R. (2023). Model-Based Clustering, Classification, and Density Estimation Using mclust in R. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003277965>
- Seal, S.E., Jeger, M.J., Van den Bosch, F. (2006a). Begomovirus Evolution and Disease Management. *Adv Virus Research*. [https://doi.org/10.1016/S0065-3527\(06\)67008-5](https://doi.org/10.1016/S0065-3527(06)67008-5)

- Seal, S.E., VandenBosch, F., Jeger, M.J. (2006b). Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Critical Reviews in Plant Sciences*, 25, 23–46. <https://doi.org/10.1080/07352680500365257>
- Simon-Loriere, E., Holmes, E.C. (2011). Why do RNA viruses recombine? *Nature Reviews Microbiology*, 9, 617–626. <https://doi.org/10.1038/nrmicro2614>
- Smith, J.M. (1992). Analyzing the Mosaic Structure of Genes. *Journal of Molecular Evolution*, 34, 126–129. <https://doi.org/10.1007/BF00182389>
- 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). <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, 8029–8033.
- Suarez, D.L., Senne, D.A., Banks, J., Brown, I.H., Essen, S.C., Lee, C.W., ... Alexander, D.J. (2004). Recombination Resulting in Virulence Shift in Avian Influenza Outbreak, Chile. *Emerging Infectious Diseases*, 10, 693–699. <https://doi.org/10.3201/eid1004.030396>
- 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, 2809–2815.
- 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, 269–280.
- Szöllősi, G.J., Tannier, E., Daubin, V., Boussau, B. (2015). The Inference of Gene Trees with Species Trees. *Systematic Biology*, 64, 42–62. <https://doi.org/10.1093/sysbio/syu048>
- Tahir, M., Amin, I., Haider, M.S., Mansoor, S., Briddon, R.W. (2015). Ageratum enation virus- A Begomovirus of weeds with the potential to infect crops. *Viruses*, 7. <https://doi.org/10.3390/v7020647>
- Trinks, D., Rajeswaran, R., Shivaprasad, P. V., Akbergenov, R., Oakeley, E.J., Veluthambi, K., Hohn, T., Pooggin, M.M. (2005). Suppression of RNA Silencing by a Geminivirus Nuclear Protein, AC2, Correlates with Transactivation of Host Genes. *Journal of Virology*, 79, 2517–2527. <https://doi.org/10.1128/jvi.79.4.2517-2527.2005>
- 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, 9487–9498. <https://doi.org/10.1128/jvi.78.17.9487-9498.2004>
- Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J.K., Murilo Zerbini, F., Martin, D.P. (2014). Establishment of three new genera in the family

- Geminiviridae: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. Archives of Virology, 159, 2193–2203. <https://doi.org/10.1007/s00705-014-2050-2>
- Wang, H., Buckley, K.J., Yang, X., Buchmann, R.C., Bisaro, D.M. (2005). Adenosine Kinase Inhibition and Suppression of RNA Silencing by Geminivirus AL2 and L2 Proteins. Journal of Virology, 79, 7410–7418. <https://doi.org/10.1128/jvi.79.12.7410-7418.2005>
- Wu, C., Christensen, T., Hein, J., (2001). A simulation study of the reliability of recombination detection methods. Molecular Biology and Evolution, 18, 1929–1939. <https://doi.org/10.1093/oxfordjournals.molbev.a003733>
- Worobey, M., Holmes, E.C. (1999). Evolutionary aspects of recombination in RNA viruses. Journal of General Virology, 80, 2535–2543. <https://doi.org/10.1099/0022-1317-80-10-2535>
- Wu, M., Bejarano, E.R., Castillo, A.G., Lozano-Durán, R. (2022). Geminivirus DNA replication in plants, in: Geminivirus: Detection, Diagnosis and Management. Elsevier, 323–346. <https://doi.org/10.1016/B978-0-323-90587-9.00038-9>
- 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, 1–10. <https://doi.org/10.1038/s41467-021-23013-2>
- Yadava, P., Suyal, G., Mukherjee, S.K. (2010). Begomovirus DNA replication and pathogenicity. Current Science, 98, 360–368.
- 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, 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, 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, 1–13. <https://doi.org/10.3390/microorganisms10122410>
- Zhao, L., Rosario, K., Breitbart, M., Duffy, S. (2019). Chapter Three - Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) viruses: ubiquitous viruses with small genomes and a diverse host range, in: Advances in Virus Research, 71–133.
- Zhou, X. (2013). Advances in Understanding Begomovirus Satellites. Annual Review of Phytopathology, 51, 357–381. <https://doi.org/10.1146/annurev-phyto-082712-102234>
- Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G.W., Robinson, D.J., Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. Journal of General Virology, 78, 2101–2111. <https://doi.org/10.1099/0022-1317-78-8-2101>

Zhou, Xueping, Liu, Y., Calvert, L., Munoz, C., William Otim-Nape, G., Robinson, D.J., Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, 78, 2101–2111.

CHAPTER 1

RECOMBINATION SHAPES THE GENE-BASED PHYLOGENIES IN BEGOMOVIRUSES

Resende C. L. P, Souza G. M., Verdolin L. G., Santos M. M. C., Ribeiro J. P. F, Luz J. M. Q.,
Lima A.T.M. (2024). **Recombination shapes the gene-based phylogenies in begomoviruses.**
Manuscript in preparation for publication

Recombination shapes the gene-based phylogenies in begomoviruses

C. L. P. Resende, G. M. Souza, J. M. Q. Luz, A.T.M. Lima

Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia – MG, 38405-302, Brazil.

Corresponding Author: A.T.M. Lima

E-mail: atmlima@ufu.br

Abstract

Early viral phylogenetic reconstructions relied on partial sequences from a single locus across species, assuming they represented the genome's evolutionary history. However, the availability of more complete genome data has revealed conflicting evolutionary histories among different viral genes, primarily due to incomplete lineage sorting, recombination, and differential selection pressures. This study aimed to construct high-resolution phylogenetic maps across complete viral segments to assess phylogenetic signal consistency and quantify incongruence caused by recombination. Begomoviruses (genus *Begomovirus*, family *Geminiviridae*) were used as model viruses, as their genomes are highly prone to recombination. Phylogenetic trees were reconstructed separately for gene and intergenic regions (IRs), as well as from 200-nucleotide sliding windows obtained by slicing the complete DNA-A of viral isolates representing 24 distinct species. Weighted Robinson-Foulds (wRF) distances were calculated between all pairs of phylogenetic trees derived from the same DNA-A. Sequence fragments were mapped within the coat protein (cp), replication enhancer (ren), transcriptional activator protein (trap), and replication-associated protein (rep) genes that significantly contributed to overall incongruence. The results suggest that a significant fraction of phylogenetic incongruence between gene- and IR-based trees is attributed to recombination, with recombination hotspots frequently segregating these regions and generating topologies that reflect distinct evolutionary histories. These findings have important implications for interpreting phylogenetic relationships in *Begomovirus* sequence-based bifurcating trees.

Keywords: *Begomovirus*, Recombination, Molecular Evolution, Phylogeny

Introduction

Viral genomes are typically small compared to those of other biological entities (Brandes and Linial, 2016; Edwards et al., 2021). This compact nature combined with high-throughput sequencing technologies have generated vast genetic datasets for viral full-genome studies. Over recent decades, phylogenetic and population genetics studies have utilized complete genomes to assess viral diversification dynamics (Langedijk et al., 2024; Mahmoudabadi and Phillips, 2018; Maxmen, 2021), often disregarding potential alternative evolutionary histories within viral genomes. A practical example is the widespread application of traditional phylogenetic methods, such as maximum likelihood (ML), for classifying severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolates into lineages during the COVID-19 pandemic (Rambaut et al., 2020). These methods assume a single evolutionary history across the entire sequence (Posada and Crandall, 2002), thus overlooking conflicting evolutionary histories between and within genes and non-coding regions.

Analyzing genes or non-coding regions separately from the genomic context often reveals distinct evolutionary histories within the same viral genome, resulting in markedly different tree topologies (Kistler and Bedford, 2021). Significant factors contributing to these conflicting evolutionary histories include recombination (Simon-Loriere and Holmes, 2011), incomplete lineage sorting (ILS) (Maddison and Knowles, 2006), and differential selection (Sanjuán and Domingo-Calap, 2021). Recombination involves the exchange of genome fragments between distinct viruses (Lai, 1992) and varies widely among viral populations, from rare (e.g., hepatitis B virus) (Bollyky et al., 1996) to highly frequent (e.g., *Tomato yellow leaf curl virus*, TYLCV) (Padidam et al., 1999). Notably, plant-infecting viruses from the families *Caulimoviridae* (Froissart et al., 2005), *Potyviridae* (Revers et al., 1996), and *Geminiviridae* (Padidam et al., 1999) are highly prone to recombination.

Incomplete lineage sorting occurs most frequently during rapid speciation episodes, where multiple new lineages emerge within a short timeframe from a common ancestor (Avice and Robinson, 2008; Maddison and Knowles, 2006). Consequently, ancestral genetic variations persist across successive speciation events, even after the lineages have diverged (Szöllösi et al., 2015). Differential selection within a genome occurs when different genes or genomic regions are under varying selective pressures due to their distinct functions, leading to different evolutionary trajectories (Brumme et al., 2007; Nouri et al., 2014). These selective pressures generate a mosaic of evolutionary rates and patterns across genomes, which reflects the complex interplay between genetic architecture and environmental factors.

The genus *Begomovirus* comprises geminiviruses that are transmitted by cryptic species within the *Bemisia tabaci* complex to dicotyledonous plants (Fiallo-Olivé et al., 2021; Fiallo-Olivé and Navas-Castillo, 2023; Navas-Castillo et al., 2011) and includes numerous single-stranded DNA (ssDNA) viruses whose propensity for recombination has been well documented for decades (Crespo-Bellido et al., 2021; Munoz et al., 1997; Padidam et al., 1999; Pita et al., 2001). Begomovirus genomes exhibit high recombination frequencies at the gene boundaries and within intergenic regions (IRs) (Lefeuvre and Moriones, 2015; Martin et al., 2011a, 2011b). Additionally, differential selective pressures on structural and non-structural genes (Martin et al., 2005), as well as on non-coding regions, make begomoviruses valuable models for studying the evolutionary phylogenetic history across their genomic segments (Lefeuvre et al., 2019; Seal et al., 2006a, 2006b).

In this context, the present study aimed to analyze diversification by calculating dissimilarity metrics between pairs of gene- and IR-based ML phylogenetic trees. Additionally, a sliding window approach was applied to investigate diversification patterns within coding and non-coding regions along the DNA-A segment. The results confirmed that recombination and differential selection are key factors influencing sequence-based phylogenies in begomoviruses.

Material and Methods

Begomovirus species datasets

The sequence dataset used in this study consisted of complete DNA-A sequences from monopartite and bipartite isolates of 24 distinct begomovirus species (Supplementary Table 1). These viral sequences were obtained on March 7, 2022, from the GenBank database (Sayers et al., 2019) using the Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/>). Only species with at least 50 available complete DNA-A sequences were included. The data were manually curated to include only sequences with intact genes, i.e., those without ORF frameshifts due to nucleotide indels (insertions or deletions). This information was verified from the GenBank webpage or the associated scientific literature. Sequences not meeting these criteria were excluded, resulting in a final dataset of 3,136 complete DNA-A sequences from begomovirus isolates. All sequences were organized to begin at the cut site of the conserved nonanucleotide at the origin of replication (5'-TAATATT//AC-3').

Sequence alignments

Multiple sequence alignments were constructed separately for complete DNA-A sequences, genes, and intergenic regions (IRs) using the MUSCLE5 program (Edgar, 2022). All alignments were manually inspected using the AliView alignment editor (Larsson, 2014). Furthermore, gene-based alignments were curated to ensure correct alignment of codons. A custom Python3 script was used to slice DNA-A-based alignments from isolates within the same begomovirus species into 200-nucleotide sliding windows with a step size of 20 nucleotides. DNA-A-based alignments were subjected to a prior refinement to minimize gaps caused by long nucleotide insertions in individual sequences, using the 'smart-gap' function of the ClipKIT software (Steenwyk et al., 2020).

Recombination analysis

Recombination events in each of the 24 DNA-A-based alignments were detected using the Recombination Detection Program version 4 (RDP4) (Martin et al., 2015). The analysis was conducted employing seven detection methods with default settings: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), Chimaera (Posada and Crandall, 2001), MaxChi (Smith, 1992), 3Seq (Lam et al., 2018), SiScan (Gibbs et al., 2000), and BootScan (Martin et al., 2005). DNA-A sequences in which recombination events detected by at least four of these methods were removed, creating new datasets.

Phylogenetic analysis

Maximum likelihood (ML) phylogenetic trees were constructed using IQ-TREE (Nguyen et al., 2015). The best-fit nucleotide substitution models were selected using the ModelFinder method (Kalyaanamoorthy et al., 2017) within IQ-TREE. Branch support was assessed with 5,000 bootstrap replicates for DNA-A-based alignments and 3,000 bootstrap replicates for gene-, IR-, and sliding window-based alignments (Minh et al., 2013). Topological comparisons between ML trees were visualized using tanglegrams, generated with the R package phytools (Revell, 2024), where branches were automatically rotated to maximize the number of tree tip correspondences.

Calculation of dissimilarities between ML phylogenetic trees

In order to compare the topologies and branch lengths among all phylogenetic ML trees we calculated the weighted Robinson-Foulds (wRF) metric (Robinson and Foulds, 1981) using

the R package treespace (Jombart et al., 2017). The pairwise tree comparisons were automated by means of a custom script in the R programming language.

Rarefaction and comparisons between median wRF distances

The number of complete DNA-A sequences varied widely across begomovirus species datasets, challenging direct comparisons of median wRF distances between pairs of ML phylogenetic trees. This issue was addressed by employing a rarefaction approach, with the smallest dataset size as the reference. A Python 3 script was developed to generate new sequence datasets from the original ones with pre-determined sample sizes (equal to the smallest dataset size) by performing 1,000 random samplings without replacement (jackknife method). ML trees were constructed and their pairwise wRF distances calculated for each of the 1,000 new sequence datasets. For each sequence dataset, median wRF distances along with their 95% non-parametric bootstrap confidence intervals were estimated based on 5,000 replicates using the R package boot (Canty and Ripley, 2021).

Multidimensional scaling and ML tree clustering analysis

Multidimensional scaling (MDS) was used to reduce dimensionality and visualize relationships among ML trees based pairwise wRF distance matrices. MDS was calculated using the "cmdscale" function within R (R Core Team, 2021). For ML tree clustering, principal component analysis (PCA) was first performed to minimize noise in wRF distance matrices. The minimum number of principal components explaining 85% of data variance was selected. Subsequently, the best-fit Gaussian mixture models were identified using the Bayesian Information Criterion (BIC) in the mclust package (Scrucca et al., 2023). ML tree clusters were then visualized within the MDS plots.

Results

Conflicting evolutionary histories within DNA-A of Begomovirus

Datasets containing 32 to 452 complete DNA-A sequences from isolates of 24 distinct begomovirus species were analyzed (Figure 1a). The largest datasets corresponded to TYLCV, ToLCNDV, and ACMV species ($N = 472, 391$, and 317 sequences, respectively), while the smallest were from SLCCNV, PaLCuCNV, and ToLCTV species ($33, 33$, and 53 sequences, respectively; Figure 1a). Initially, all DNA-A sequences containing ORFs with frameshifts due to indels were removed to estimate the degree of phylogenetic incongruence among viral genes,

hereafter referred to as intergenic incongruence. The highest number of frameshift variants were in SLCuV, SPLCV, and CLCuKoV datasets, with 67 (37%), 41 (33%), and 34 (11%) sequences removed, respectively. In contrast, for 14 begomovirus species datasets, less than 10% of sequences were removed due to frameshifts (Figure 1a).

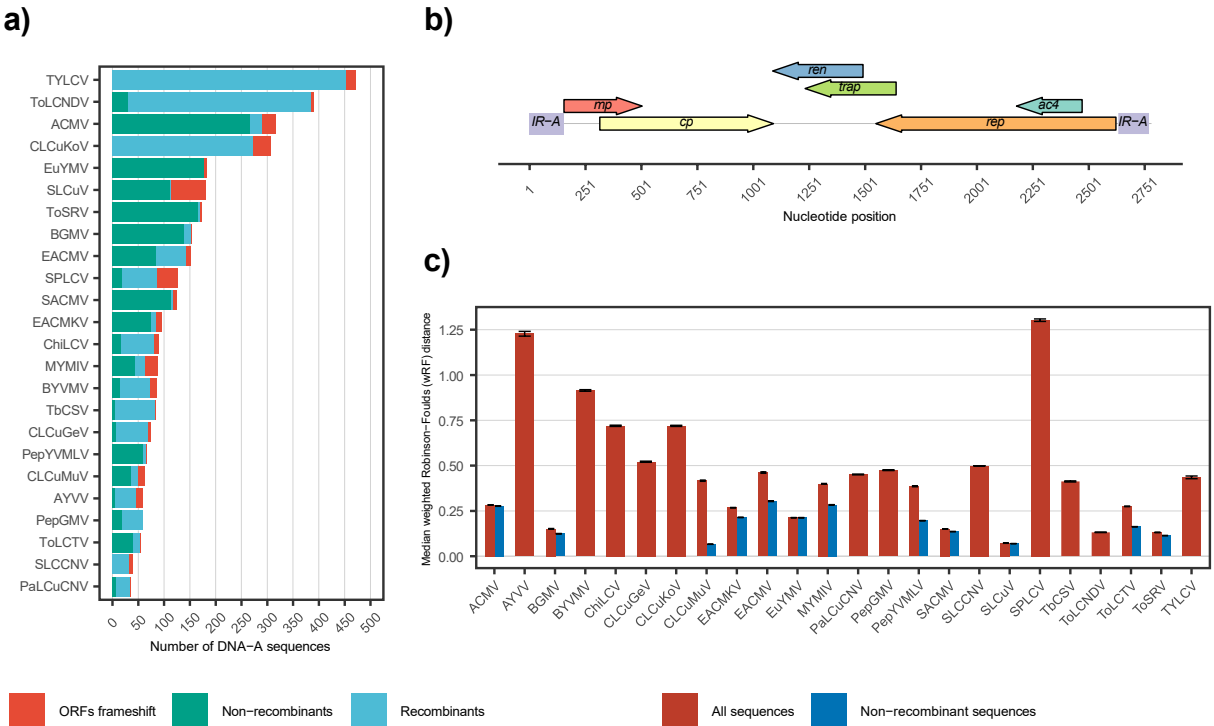


Figure 1. Number of DNA-A sequences analyzed, categorized into three groups: sequences with reading frameshift in at least one viral gene; sequences identified as recombinants by at least four RDP methods; and sequences with no detectable evidence of recombination by at least four RDP methods (a). Schematic representation of the genomic organization of a begomovirus isolate of the TYLCV species (GenBank accession: AJ489258) (b). Median weighted Robinson-Foulds (wRF) distances from pairwise comparisons of all gene- and intergenic region-based trees within the same begomovirus species dataset (c). The overlap of 95% non-parametric bootstrap confidence intervals indicates no statistically significant differences between median wRF distances.

DNA-A sequences with intact ORFs were screened for recombination events. Species datasets for AYVV, CLCuGeV, CLCuKoV, SLCCNV, TbCSV, ToLCNDV, and TYLCV contained 80% to 100% recombinant sequences. In contrast, seven species datasets (ACMV, BGMV, EuYMV, PepYVMLV, SACMV, SLCuV, and ToSRV) had less than 10% recombinant sequences, with no recombination events detected in EuYMV (Figure 1a).

All median wRF distances between ML tree pairs differed significantly, shown by their 95% non-parametric bootstrap confidence intervals (Figure 1c, Supplementary Figure 1). The highest median wRF distances occurred in SPLCV (wRF = 1.302) and AYVV (wRF = 1.228), suggesting a high degree of intergenic incongruence. Additionally, the BYVMV, ChiLCV, and

CLCuKoV datasets also showed considerable intergenic incongruence ($wRF = 0.914, 0.719,$ and 0.719 , respectively). Conversely, the ToLCNDV, ToSRV, and SLCuV datasets exhibited the lowest median distances ($wRF = 0.131, 0.131,$ and 0.071 , respectively), suggesting highly similar phylogenetic topologies across their viral genomes (Figure 1c).

After removing recombinant sequences, the EACMV dataset had the highest median distance ($wRF = 0.303$), followed by MYMIV ($wRF = 0.283$) and ACMV ($wRF = 0.276$). Contrastingly, SLCuV and CLCuMuV datasets showed highly similar phylogenetic, with wRF distances of 0.069 and 0.065 , respectively. Median wRF distances could not be calculated for AYVV, BYVMV, ChiLCV, CLCuGeV, CLCuKoV, PaLCuCNV, PepGMV, SLCCNV, SPLCV, TbCSV, ToLCNDV, and TYLCV datasets, as all sequences were identified as recombinant by RDP or had sample sizes smaller than 32 sequences (the smallest dataset size; Figure 1c).

Comparing datasets with and without recombinant sequences revealed the greatest reductions in median wRF distances for CLCuMuV (84%), PepYVMLV (49%), and ToLCTV (41%), indicating increased intergenic congruence and highlighting the impact of recombination. In contrast, ACMV, BGMV, EACMKV, EuYMV, SACMV, SLCuV, and ToSRV datasets showed less than 20% reductions in their median wRF distances, maintaining consistence in incongruence levels. Notably, these datasets already had low wRF distance values with the presence of recombinant sequences (Figure 1c). Overall, median wRF distances were higher with recombinant sequences than without, in most datasets. Distinct evolutionary dynamics were observed across different viral genomic regions, suggesting recombination's influence on these patterns.

IR- and mp-based trees are highly divergent

The global median wRF distances were calculated from pairwise comparison of ML trees based on viral genes, intergenic region (IR-A), and complete DNA-A sequences for all databases, with and without recombinant sequences (Figure 2). Consistently, medians wRF distances differed significantly, as shown by 95% non-parametric bootstrap confidence intervals. Pairwise comparisons involving IR-based ML trees showed the greatest distances; the highest wRF distance values ($0.634, 0.548,$ and 0.523) were observed when comparing IR-based ML trees with those based on movement protein (mp), replication enhancer (ren), and transcriptional activator protein (trap) genes, respectively, indicating a high degree of intergenic incongruence. Most similar ML trees were those based on closely positioned genes, such as ren

and trap (wRF = 0.178), as well as replication-associated protein (rep) and C4 protein (ac4) genes (wRF = 0.254; Figure 2a). Interestingly, the DNA-A-based tree was most similar to the rep-based tree (wRF = 0.190).

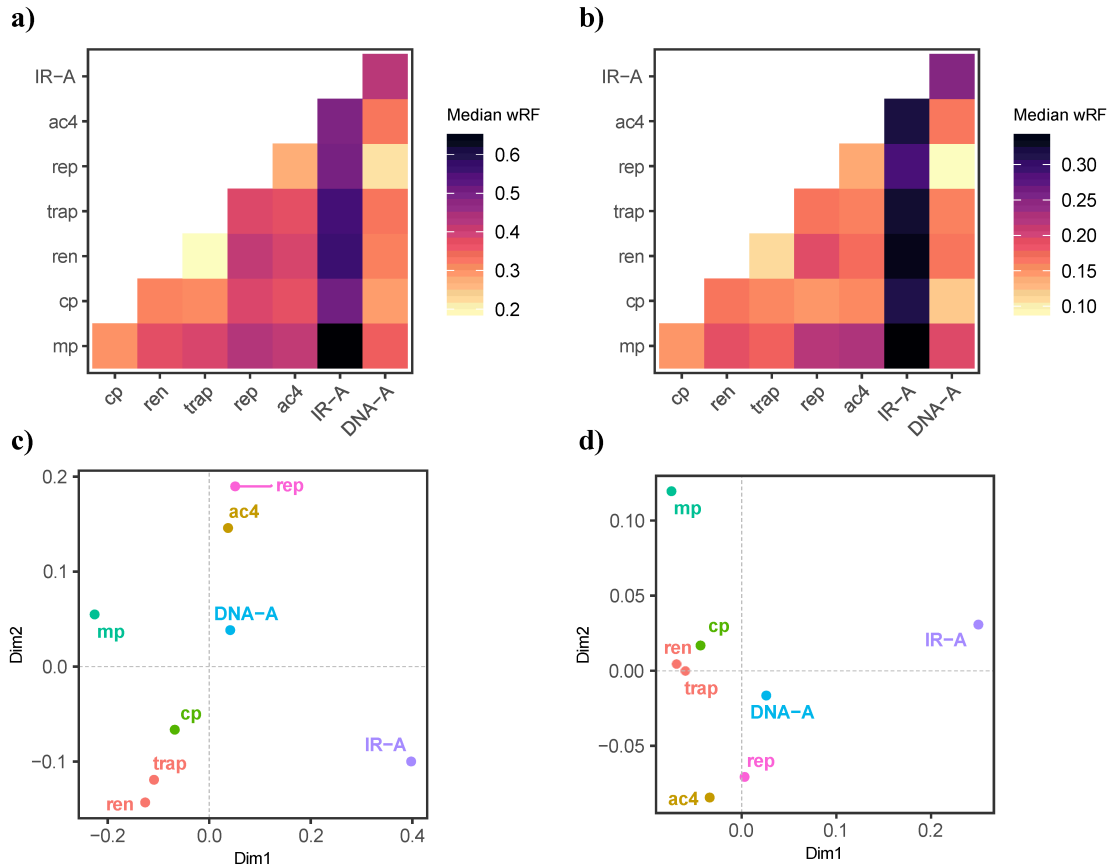


Figure 2. Heatmaps of median wRF distance matrices comparing pairs of gene-based and intergenic region (IR)-based trees (**a**, **b**). Median distances were calculated using combined datasets from all begomovirus species, including recombinant sequences (**a**) and excluding sequences identified as recombinant by at least four RDP analytical methods (**b**). Multidimensional scaling (MDS) plots of median weighted Robinson-Foulds (wRF) distances between gene- and IR-based trees (**c**, **d**). Projections of phylogenetic trees were generated from median wRF distances using combined datasets with (**c**) and without (**d**) recombinant sequences. Points of the same color represent clusters of phylogenetic trees that are comparatively more similar to each other.

Comparing ML trees from datasets after removing recombinant sequences enabled assessment of recombination's contribution to incongruence across DNA-A segments. The greatest wRF distances occurred between IR-based trees and mp- (0.343), ren- (0.331), and trap-based trees, with wRF of 0.343, 0.331, 0.320, respectively. In contrast, the smallest distances occurred between trees of physically closer genes, such as ren and rep (wRF = 0.106) and trap and ren (wRF = 0.106). The DNA-A-based tree was most similar to the rep-based tree (wRF = 0.087), even without detectable recombination. The wRF distances between IR- and

gene-based trees were up to fourfold higher than those between pairs of closely positioned genes (Figure 2b).

Overall, median wRF distances between gene- and IR-based trees decreased after removing recombinant sequences, indicating that the evolutionary mechanism of recombination accounts for a significant fraction of phylogenetic differences (Figure 2a, 2b). However, the IR-based tree remained to diverge significantly from all gene-based trees (Figure 2c, 2d). The largest distance reduction (57%) occurred between coat protein (cp)- and rep-based trees, while the smallest (31%) was between IR- and ac4-based trees. Multidimensional scaling (MDS) of wRF distance matrices revealed that cp- and mp-based trees were relatively disparate, despite their partial overlap within the DNA-A segment. However, ren- and trap-based trees (which share a long overlapping region) and rep- and ac4-based trees (with ac4 entirely nested within rep) clustered more closely in the MDS plot (Figure 2c, 2d). These results suggest that the incongruence between IR- and mp-based trees may be attributed to a recombination hotspot between these regions and their neighboring regions (Figure 2c). Thus, the opposite sides of this hotspot could originate from different parental viruses, leading to pronounced discrepancies with neighboring regions. Furthermore, MDS revealed that cp-, ren-, and trap-based trees clustered at the opposite side of rep- and ac4-based trees (Figure 2c), which is consistent with the presence of a hotspot separating these genomic regions. The removal of recombinant sequences increased congruence among gene-based trees, except for the mp-based tree. This suggests that the divergence observed in mp-based trees may not be entirely attributed to recombination, unless a persistent recombination hotspot influenced their evolutionary history.

Pairwise wRF distances positively correlate with physical distances between genes on DNA-A

Median wRF distances correlated with the physical distances between viral gene midpoints across combined datasets from all viral species, with and without recombinant sequences (Figure 3). Excluding IR-based tree comparisons (due to possible recombination hotspots), median wRF distances tended to increase as the physical distance between genes on the DNA-A segment increases. Specifically, overlapping or proximate genes on the DNA-A segment tend to exhibit more similar tree topologies than those located farther apart. However, this pattern did not occur for comparisons involving IR-based trees, which showed large wRF distances relative to trees from other regions regardless of physical separation on DNA-A (Figure 3a).

After removing recombinant sequences, wRF distances between gene-based trees showed a less pronounced positive correlation with their physical separation (Figure 3b), while distances between IR- and gene-based trees remained high.

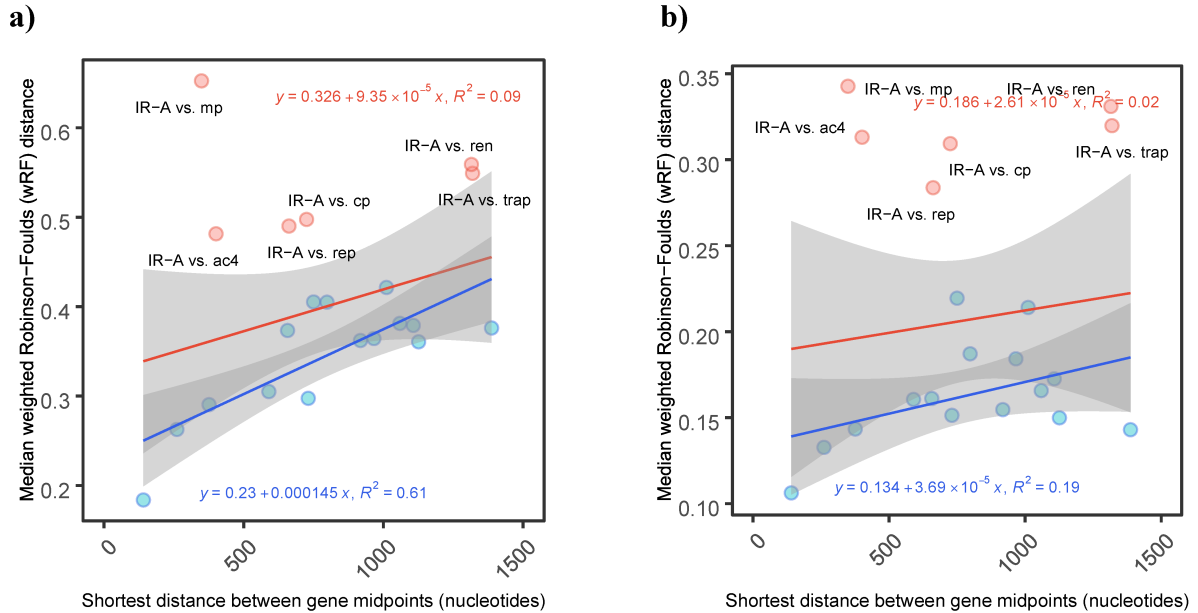


Figure 3. Linear regressions between median wRF distances and physical distances separating pairs of gene and intergenic region (IR) midpoints. Given the circular structure of the DNA-A segment, only the shortest distances between midpoint pairs were considered. Points in red represent comparisons involving IR-based trees, while points in blue represent comparisons excluding IR-based trees. Regressions represented by red lines include all pairwise tree comparisons, while regressions represented by blue lines exclude IR-based tree comparisons. Regressions were calculated from combined datasets, including (a) and excluding (b) recombinant sequences identified by at least four RDP methods.

Incongruence across coding and non-coding regions for individual species

Additionally, median wRF distances between gene- and IR-based trees were analyzed across individual species datasets (Figure 4, Supplementary Figures 2 and 3). IR-based trees showed greater divergence in the AYVV, BGMV, CLCuMuV, EACMV, SPLCV, and TYLCV datasets. AYVV and TYLCV datasets, ren-, trap-, cp-, and mp-based trees showed smaller wRF distances, indicating greater congruence among them, although these ML trees did not cluster together. Furthermore, rep- and ac4-based trees in AYVV and TYLCV showed reduced congruence (Figures 4a, 4c). In the SPLCV dataset, gene-based trees were evenly distributed in the MDS plot, indicating consistent incongruence among all pairs of trees (Figure 4b). However, cp- and mp-based trees clustered together. ML trees from overlapping rep and ac4, as well as trap and ren genes, did not cluster.

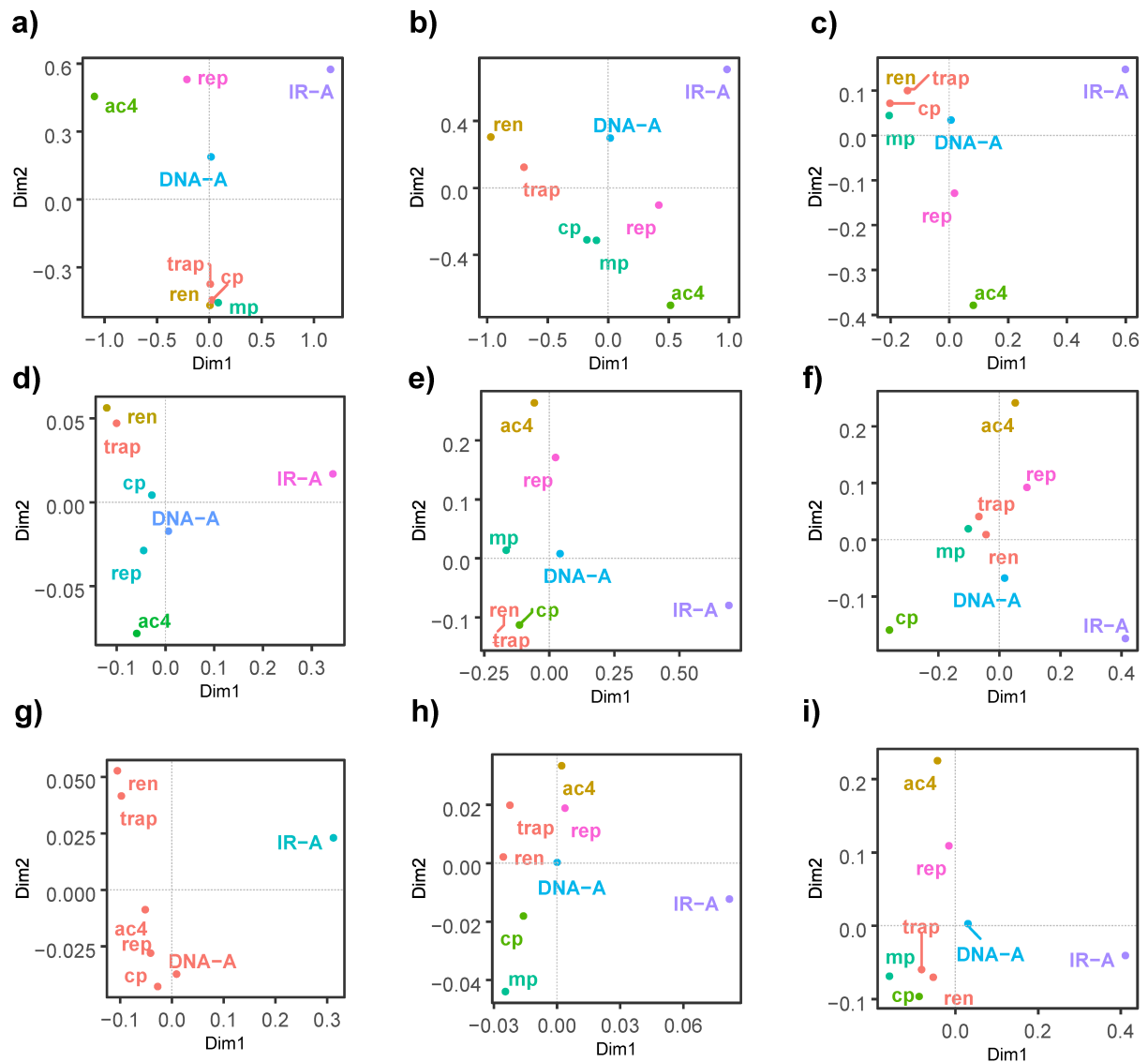


Figure 4. Multidimensional scaling (MDS) plots based on weighted Robinson-Foulds (wRF) distance matrices for gene-based and intergenic region (IR)-based trees in individual begomovirus species datasets: AYVV (a), SPLCV (b), TYLCV (c), BGMV (d, g), CLCuMuV (e, h), and EACMV (f, i). MDS plots were generated from wRF distances calculated using datasets that included recombinant sequences (a–f) and datasets that exclude sequences identified as recombinant by at least four RDP methods (g–i). Points of the same color represent clusters of maximum likelihood trees that are comparatively more similar to each other.

In the BGMV dataset, ren-, trap-, and cp-based trees were relatively close to each other, while rep- and especially ac4-based trees exhibited greater separation (Figure 4d). Removing recombinant sequences increased similarity among gene-based trees, which clustered together and consistently remained distant from the IR-based tree (Figure 4g). In the CLCuMuV dataset, the cluster composed of ren- and trap-based trees was positioned near the cp-based tree but remained distant from rep- and ac4-based trees (Figure 4e). Post-recombination removal, the proximity between ac4- and rep-based trees, as well as between cp- and mp-based trees, became

more evident, whereas the IR-based tree remained distant from the others (Figure 4h). In EACMV, ren-, trap-, and mp-based trees showed greater similarity, while cp- and mp-based trees exhibited more contrasting topologies (Figure 4f). Excluding recombinant sequences reduced distances between cp- and mp-based trees. The IR-based tree remained highly divergent from other trees (Figure 4i), consistent with patterns in other viruses.

The divergence patterns between gene- and IR-based trees revealed by MDS for individual species datasets showed some similarities to those observed in the combined datasets (Figures 2c, 2d). Notably, IR-based trees from individual species were often the most divergent. In contrast, some individual species datasets showed significantly divergent tree topologies for closely positioned genes. These results suggest no consistent pattern of similarity between pairs of phylogenetic trees, particularly for gene-based trees analyzed across individual species datasets. Thus, factors influencing incongruence among ML trees may vary across begomovirus species datasets. Divergence among trees from physically closer regions can be attributed to recombination hotspots. However, recombination frequency across these regions varies significantly among species datasets.

Long branches characterize trees with recombinant sequences

ML tree pairs from genomic regions were compared across individual species datasets to identify branching patterns and branch lengths distinguishing trees with and without recombinant sequences. In the CLCuMuV dataset, ren- and cp-based ML trees (Figure 5a, 5c) were directly compared with IR- and ren-based ML trees (Figure 5b, 5d). Incongruences between these tree pairs are typically associated with long basal branches (Figures 5a, 5b). Recombination drives this pattern, as well documented in the literature. Post-recombination removal, the presence of shorter branches increased congruence between ren- and cp-based trees (Figure 5c); however, IR- and ren-based trees remained divergent, characterized by the consistent presence of long branches in the IR-based tree (Figure 5d). This could explain the IR-based tree's divergence from other trees, even after removing recombinant sequences.

Mapping divergent evolutionary histories across genes

Gene-based analyses revealed conflicting evolutionary histories within an intragenomic context. However, cp and rep genes account for more than 70% of a typical DNA-A segment's length. Thus, a more detailed analysis was conducted using 200-nucleotide sliding windows across the DNA-A segment (Figure 6, Supplementary Figures 4 and 5). Figure 6 displays tree

topologies from sliding windows spanning the entire DNA-A length for BGMV, EACMV, and MYMIV isolates, with and without recombinant sequences.

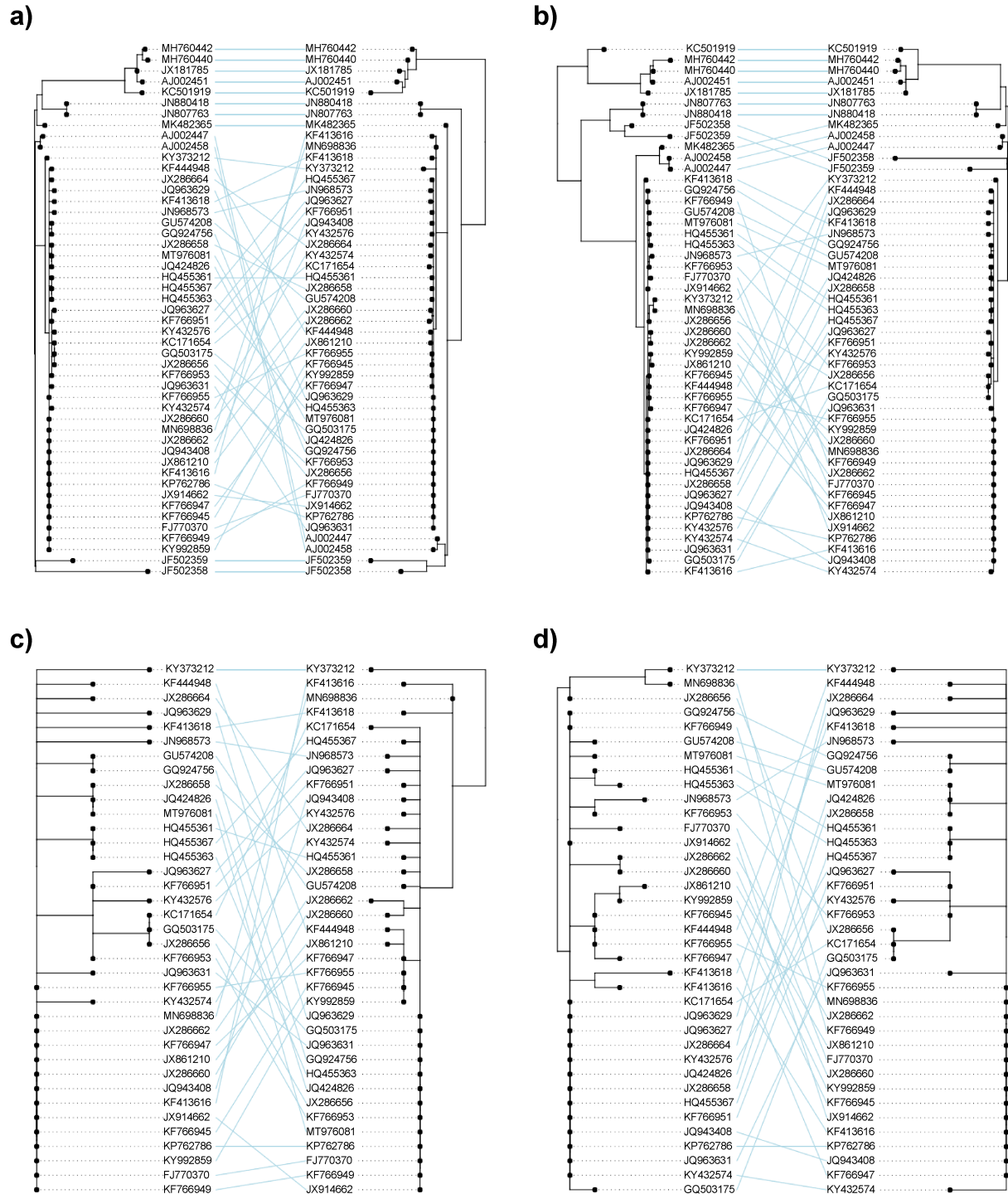


Figure 5. Tanglegrams of maximum likelihood trees based on replication enhancer (ren) and coat protein (cp) genes (a, c) and the intergenic region (IR) and ren gene (b, d) for the CLCuMuV species, including recombinant sequences (a, b) and excluding sequences identified as recombination by at least four RDP methods (c, d). Tree tips are labeled with GenBank accession numbers corresponding to the complete DNA-A sequences from which gene and IR sequences were derived. Branches were rotated to maximize tree tip correspondences.

In the BGMV dataset, tree topologies based on IR sliding windows were highly divergent from those of other regions and showed significant wRF distances among themselves. This indicates non-linear divergence among complete IR-based trees, reflecting high variability across IR sub-regions. Importantly, trees from N-terminal cp and rep coding regions exhibit substantial divergence from those of other viral gene windows (Figure 6a). Divergence patterns among EACMV sliding-window trees varied among species datasets (Figure 6b).

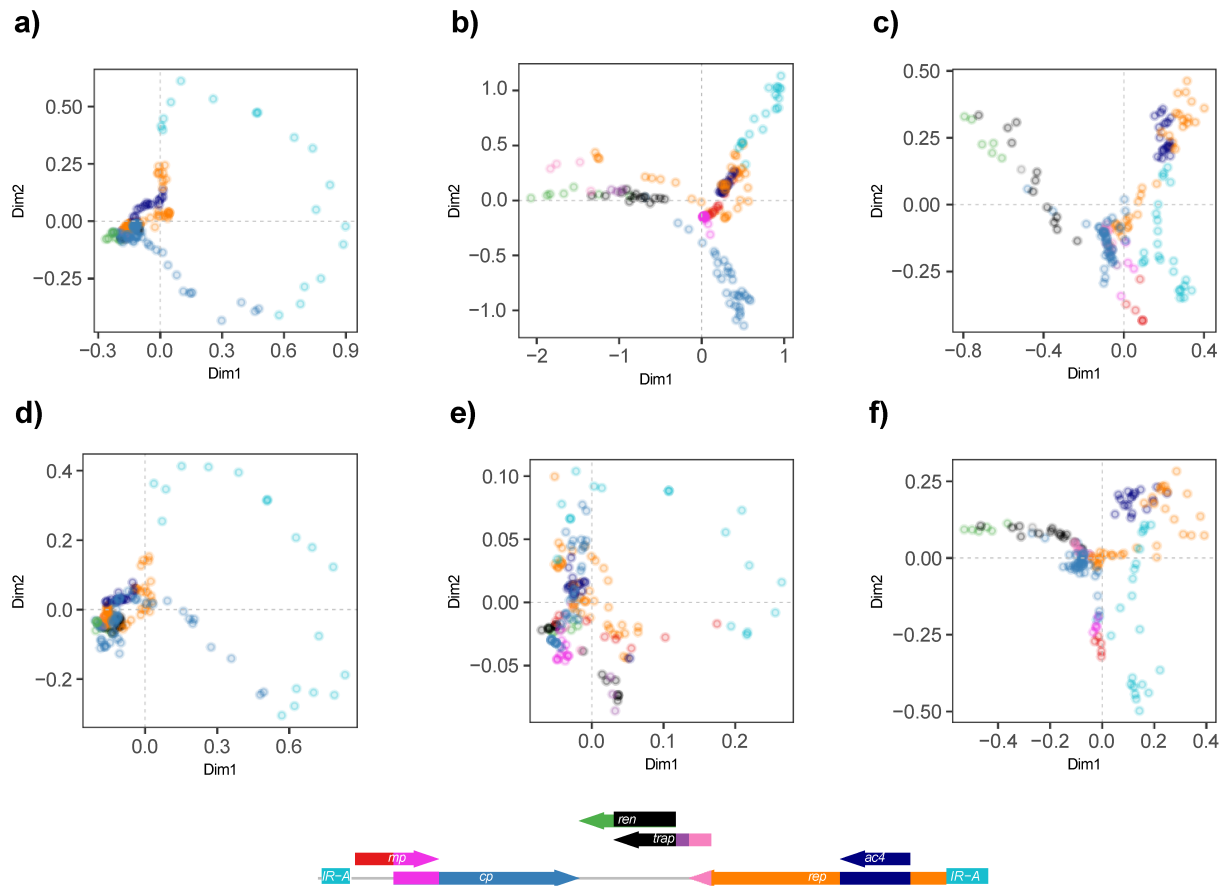


Figure 6. Multidimensional scaling (MDS) plots from weighted Robinson-Foulds (wRF) distance matrices between sliding window-based trees for individual begomovirus species datasets: BGMV (**a**, **d**), EACMV (**b**, **e**), and MYMIV (**c**, **f**). MDS plots were based on wRF distances calculated from datasets including including recombinant sequences (**a**, **c**) and excluding sequences identified as recombinant by at least four RDP methods (**d**, **e**). Points were colored according to the schematic representation of the begomovirus DNA-A segment shown at the bottom of the figure.

MDS analysis of divergence between sliding window-based trees revealed a three-pronged star pattern, with trees at the extremities showing the greatest divergence in pairwise comparisons. The prongs correspond to IR sliding-window trees, trees from N-terminal cp coding sequences, and trees from the trap-rep overlapping region and the non-overlapping C-terminal ren region.

A similar pattern was observed in MYMIV sliding-window trees (Figure 6c), with projections' extremities featured trees based on IR, N-terminal rep, non-overlapping ren, and trap-rep overlapping regions. After excluding recombinant sequences, tree projections in MDS plots (Figures 6d, 6f) exhibited no significant pattern differences for BGMV and MYMIV datasets (Figures 6d, 6f); however, the IR sliding-window tree showed significant divergence in the EACMV dataset (Figure 6e). These results confirm that divergence patterns vary among phylogenetic trees from different genomic regions across the DNA-A segment depending on the begomovirus species.

Discussion

Recombination is essential for the emergence and evolution of begomoviruses, enhancing species diversity and intraspecific variation (Crespo-Bellido et al., 2021; Martin et al., 2011a; Munoz et al., 1997; Padidam et al., 1999; Pita et al., 2001). In the present study, complete DNA-A sequences from 24 begomovirus species *were analyzed for recombination events. The results revealed detectable recombination in most species, consistent with prior reports of frequent recombination in Begomovirus genomes.* Recombination shuffles mutation-induced genetic variation, accelerating viral adaptation to new plant hosts and vectors. A key example is the recombinant EACMV-Uganda (EACMV-UG), which caused severe disease outbreaks in East and Central Africa during the 1990s. Its parents are EACMV and ACMV isolates (Crespo-Bellido et al., 2021; Pita et al., 2001; Zhou et al., 1997). Therefore, recombinant begomoviruses threaten global agriculture, causing substantial socioeconomic losses (Monci et al., 2002; Pita et al., 2001; Zhou et al., 1997).

Another critical aspect of recombination is its impact on phylogenetic reconstruction using traditional methods, as it can obscure relationships among isolates of the same species and even between different species (Posada and Crandall, 2002; Schierup and Hein, 2000; Wiuf et al., 2001). Hybrid viruses resulting from recombination contain regions of different origins, leading to conflicting evolutionary histories inherited from distinct parental viruses. Traditional phylogenetic methods assume that all analyzed sequences share a single common ancestor, an assumption that contradicts the mosaic structure of *Begomovirus* genomes, which frequently undergo recombination. Consequently, phylogenies based on recombinant sequences do not accurately reflect the evolutionary history of any parent (Posada and Crandall, 2002). Thus, recombination can lead to markedly divergent phylogenetic tree topologies, hindering accurate understanding of viral evolutionary dynamics.

EuYMV, SLCuV, and ToSRV isolates exhibited low levels of recombination in their genomes. These results are consistent with those reported by Monci et al. (2002), indicating that despite selective pressures, host-vector interactions and biological-ecological factors confronting viral populations, the genomic stability of SLCuV is maintained over time, preserving its genetic and structural functionality. These findings reinforce that recombination occurs at varying frequencies across begomovirus species.

More similar phylogenetic tree topologies were observed for proximate genes (cp and trap) and overlapping genes (ren-trap and rep-ac4 overlapping regions) in viral genomes. Conversely, marked incongruence was observed in comparisons between IR- and mp-based trees. This divergence persisted even in analyses based on datasets excluding recombinant sequences. This study proposed that congruence between trees from pairs of proximate genes (cp and mp), IR, and rep/ac4 genes results from the existence of recombination hotspots. Beyond these regions, incongruence levels appear to correlate with the physical distance between genes. Between recombination hotspots, gene segregation depends on physical distance. Beyond these hotspots, greater distances between regions increase their recombination-induced segregation. Sequences on either side of hotspots often derive from different parental viruses, resulting in highly divergent tree topologies. Excluding trees covering opposite hotspot sides, 60% of wRF distance variation between regions is attributed to their physical distance. This indicates uniform recombination across these regions, with physical distance determining segregation between sub-regions.

Removing recombinant sequences did not increase the similarity of IR- or mp gene-based tree topologies with other genomic regions. This may indicate that the study's criterion was not stringent enough to fully eliminate recombinant sequences. The Recombination Detection Program (RDP) includes seven methods, but most studies validate recombination only when detected by at least four (majority rule) methods (Wiuf et al., 2001). Thus, sequences with recombination detected by fewer than four methods persisted in the analyzed datasets. Moreover, recombination detection in a dataset requires the presence of a mosaic structure and at least one parental virus. Therefore, analyses often underestimate recombination in datasets with small sample sizes or biased sampling that inadequately captures viral diversity due to limited geographic or temporal collections.

The presence of recombination hotspots and coldspots in single-stranded DNA virus genomes, including begomoviruses, is well documented in the literature (Martin et al., 2011a; Crespo-Bellido et al., 2021; García-Andrés et al., 2007; Padidam et al., 1999). The consistent

IR-based tree divergence across all begomovirus species datasets confirms that this genomic region is the primary recombination target in the DNA-A segment. Recombination hotspots have been identified at the cp/ren and trap/rep interfaces (Lefevre and Moriones, 2015; Martin et al., 2011a). These hotspots accounted for the divergence between cp/ren/trap- and rep/ac4-based trees, evidenced by multidimensional scaling across all species datasets. In this context, this interface is the second most frequent recombination targeted. The individual dataset analysis revealed more frequent recombination at the cp/ren interface, resulting in greater divergence in cp- and ren-based trees. Other datasets showed greater divergence in rep- and trap-based trees, indicating more frequent recombination at this interface. These results confirm that recombination influences *Begomovirus* gene phylogenies. This study suggests frequent recombination at the interface between cp and mp genes.

Phylogenetic reconstruction using complete *Begomovirus* DNA-A segments obscures the relationships among viral genes. Divergence pattern variation among genes within a viral species reflects may be explained by frequent recombination at specific hotspots, highlighting distinct recombination dynamics in viral populations. Variations in recombination patterns at each hotspot may be due to selective constraints varying across viral species.

Recombination potential is a crucial aspect of *Begomovirus* populations (Seal et al., 2006a, 2006b). A notable example is the sweet potato leaf curl virus (SPLCV), which severely damages sweet potato (*Ipomoea batatas*) crops by reducing tuber yield and quality (Lozano et al., 2009; Opiyo et al., 2010). Research indicates that high genomic diversity and frequent recombination enhance viral dissemination. Recombination hotspots in genomes of sweet potato-infecting begomoviruses occur at the IR and rep gene, as well as at trap/ac4, cp/ren, trap/rep, and cp/trap interfaces (Fernandes et al., 2011; Mar et al., 2017).

Sweet potato plants are vegetatively propagated and transported globally, thus the accumulation of begomoviruses from different geographic regions is common (Lozano et al., 2009; Opiyo et al., 2010), promoting interspecific recombination. This information aligns with the results of the present study, in which SPLCV DNA-A genes exhibited highly divergent evolutionary histories. However, this contrasts with the reduced divergence observed in phylogenies of cassava-infecting begomoviruses. Limited cassava trade and cultivation may restrict mixed infections to a narrow range of begomoviruses, reducing parental virus diversity.

Several factors, such as viral replication, mixed infections, host range expansion, and polyphagous vectors, significantly influence recombination (Fiallo-Olivé et al., 2020; García-Andrés et al., 2007, 2006). Cotton-infecting begomoviruses, including CLCuMuV and

CLCuKoV, exhibit high recombination frequencies (Saeed et al., 2018; Tahir et al., 2015). In the present study, CLCuMuV phylogenetic trees were directly compared for ren and cp genes, as well as for the IR and ren gene. The presence of long basal branches was associated with incongruences between trees due to recombinant sequences. Removing recombinant sequences increased congruence between ren- and cp-based trees, characterized by shorter branches. These results are consistent with those found by Schierup and Hein (2000), who reported that recombinant sequences in phylogenetic reconstruction is characterized by long branches.

The findings of this study revealed marked variation in topological incongruence between coding and non-coding regions of the DNA-A segment in begomovirus species. Closely positioned genes generally exhibited more similar phylogenies than those farther apart, except for the mp gene and IR, which showed distinct topologies from proximate genes. These results indicate the presence of recombination hotspots often separating these regions, leading to topologies that reflect divergent evolutionary histories. Overall, the phylogeny of DNA-A genes in begomoviruses is strongly influenced by recombination, highlighting the need for evolutionary reconstruction methods that explicitly account for recombination.

References

- Avise, J.C., Robinson, T.J. (2008). Hemiplasy: A New Term in the Lexicon of Phylogenetics. *Systematic Biology*, 57, 503–507. <https://doi.org/10.1080/10635150802164587>
- Bollyky, P.L., Rambaut, A., Harvey, P.H., Holmes, E.C. (1996). Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution*, 42, 97–102. <https://doi.org/10.1007/BF02198834>
- Brandes, N., Linial, M. (2016). Gene overlapping and size constraints in the viral world. *Biology direct*, 11, 1-15. <https://doi.org/10.1186/s13062-016-0128-3>
- Brumme, Z.L., Brumme, C.J., Heckerman, D., Korber, B.T., Daniels, M., ..., Harrigan, P.R. (2007). Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. *PLoS Pathogens*, 3, 94. <https://doi.org/10.1371/journal.ppat.0030094>
- Canty, A., Ripley, B.D. (2021). Boot: Bootstrap R (S-Plus) Functions.
- Crespo-Bellido, A., Hoyer, J.S., Dubey, D., Jeannot, R.B., Duffy, S. (2021). Interspecies Recombination Has Driven the Macroevolution of Cassava Mosaic Begomoviruses. *Journal of Virology*, 95. <https://doi.org/10.1128/jvi.00541-21>
- Edgar, R.C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications*, 13. <https://doi.org/10.1038/s41467-022-34630-w>

- Edwards, K.F., Steward, G.F., Schvarcz, C.R. (2021). Making sense of virus size and the tradeoffs shaping viral fitness. *Ecology Letters*, 24, 363–373. <https://doi.org/10.1111/ele.13630>
- Fernandes, F.R., Albuquerque, L.C., de Oliveira, C.L., Cruz, A.R.R., da Rocha, W.B., ... Inoue-Nagata, A.K. (2011). Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting *Euphorbia heterophylla* plants. *Archives of Virology*, 156. <https://doi.org/10.1007/s00705-011-1070-4>
- 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. <https://doi.org/10.1099/jgv.0.001696>
- Fiallo-Olivé, E., Navas-Castillo, J. (2023). Begomoviruses: what is the secret(s) of their success? *Trends in Plant Science*, 28, 715–727. <https://doi.org/10.1016/j.tplants.2023.01.012>
- Fiallo-Olivé, E., Pan, L.L., Liu, S.S., Navas-Castillo, J. (2020). Transmission of begomoviruses and other whitefly-borne viruses: Dependence on the vector species. *Phytopathology*, 110, 10–17. <https://doi.org/10.1094/PHYTO-07-19-0273-FI>
- Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., Michalakakis, Y. (2005). Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection, *PLoS Biology*, 3(3). <https://doi.org/10.1371/journal.pbio.0030089>
- García-Andrés, S., Accotto, G.P., Navas-Castillo, J., Moriones, E., 2007. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology*, 359, 302–312. <https://doi.org/10.1016/j.virol.2006.09.030>
- García-Andrés, S., Monci, F., Navas-Castillo, J., Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology*, 350. <https://doi.org/10.1016/j.virol.2006.02.028>
- Jombart, T., Kendall, M., Almagro-Garcia, J., Colijn, C. (2017). Treespace: Statistical exploration of landscapes of phylogenetic trees. *Molecular Ecology Resources*, 17, 1385–1392. <https://doi.org/10.1111/1755-0998.12676>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kistler, K.E., Bedford, T. (2021). Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. *eLife*, 10. <https://doi.org/10.7554/eLife.64509>
- Lai, M.M. (1992). RNA recombination in animal and plant viruses. *Microbiological reviews*, 56, 61–79. <https://doi.org/10.1128/mr.56.1.61-79.1992>

- Lam, H.M., Ratmann, O., Boni, M.F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Molecular Biology and Evolution*, 35. <https://doi.org/10.1093/molbev/msx263>
- Langedijk, A.C., Vrancken, B., Lebbink, R.J., Wilkins, D., Kelly, E.J ... Madhi, S.A. (2024). The genomic evolutionary dynamics and global circulation patterns of respiratory syncytial virus. *Nature Communications*, 15, 3083. <https://doi.org/10.1038/s41467-024-47118-6>
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>
- Lefevre, P., Martin, D.P., Elena, S.F., Shepherd, D.N., Roumagnac, P., Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology*, 17, 632–644. <https://doi.org/10.1038/s41579-019-0232-3>
- Lefevre, 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>
- Lozano, G., Trenado, H.P., Valverde, R.A., Navas-Castillo, J. (2009). Novel begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and *Ipomoea indica*: Taxonomic and phylogenetic implications. *Journal of General Virology*, 90. <https://doi.org/10.1099/vir.0.012542-0>
- Maddison, W.P., Knowles, L.L. (2006). Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, 55, 21–30. <https://doi.org/10.1080/10635150500354928>
- Mahmoudabadi, G., Phillips, R. (2018). A comprehensive and quantitative exploration of thousands of viral genomes. *eLife*, 7. <https://doi.org/10.7554/eLife.31955>
- Mar, T.B., Mendes, I.R., Lau, D., Fiallo-Olivé, E., Navas-Castillo, J., Alves, M.S., Zerbini, F.M. (2017). Interaction between the new world begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite: Effects on infection and transmission by the whitefly *Bemisia tabaci*. *Journal of General Virology*, 98, 1552–1562. <https://doi.org/10.1099/jgv.0.000814>
- Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 16(6), 562–563.
- Martin, D.P., Biagini, P., Lefevre, P., Golden, M., Roumagnac, P., Varsani, A. (2011a). Recombination in eukaryotic single stranded DNA viruses. *Viruses*, 3, 1699–1738. <https://doi.org/10.3390/v3091699>
- Martin, D.P., Lefevre, P., Varsani, A., Hoareau, M., Semegni, J.Y., Dijoux, B., Vincent, C., Reynaud, B., Lett, J.M. (2011b). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathogens*, 7. <https://doi.org/10.1371/journal.ppat.1002203>

- Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1. <https://doi.org/10.1093/ve/vev003>
- Martin, D.P., Posada, D., Crandall, K.A., Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses*, 21, 98–102. <https://doi.org/10.1089/aid.2005.21.98>
- Maxmen, A. (2021). One million coronavirus sequences: popular genome site hits mega milestone. *Nature*, 593, 21–21. <https://doi.org/10.1038/d41586-021-01069-w>
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30, 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- 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, 317–326. <https://doi.org/10.1006/viro.2002.1633>
- Munoz, C., Zhou, X., Otim-Nape, G.W., Liu, Y., Robinson, D.J., Calvert, L., Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, 78, 2101–2111. <https://doi.org/10.1099/0022-1317-78-8-2101>
- Navas-Castillo, J., Fiallo-Olivé, E., Sánchez-Campos, S. (2011). Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology*, 49, 219–248. <https://doi.org/10.1146/annurev-phyto-072910-095235>
- 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, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nouri, S., Arevalo, R., Falk, B.W., Groves, R.L. (2014). Genetic Structure and Molecular Variability of Cucumber mosaic virus Isolates in the United States. *PLoS One*, 9(5). <https://doi.org/10.1371/journal.pone.0096582>
- Opiyo, S.A., Ateka, E.M., Owuor, P.O., Manguro, L.O.A., Karuri, H.W. (2010). Survey of sweet potato viruses in Western Kenya and detection of Cucumber mosaic virus. *Journal of Plant Pathology*, 92.
- Padidam, M., Sawyer, S., Fauquet, C.M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, 265, 218–225.
- Pita, J.S., Fondong, V.N., Sangare, 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, 655–665.

- Posada, D., Crandall, K.A., 2002. The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of Molecular Evolution*, 54, 396–402. <https://doi.org/10.1007/s00239-001-0034-9>
- Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences*, 98, 13757–13762.
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rambaut, A., Holmes, E.C., O’Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5, 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Revell, L.J. (2024). Phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, 12. <https://doi.org/10.7717/peerj.16505>
- Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *Journal of General Virology*, 77, 1953–1965. <https://doi.org/10.1099/0022-1317-77-8-1953>
- Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. *Mathematical Biosciences*, 53, 131–147. [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2)
- Saeed, F., Sattar, M.N., Hameed, U., Ilyas, M., Haider, M.S., Hamza, M., Mansoor, S., Amin, I. (2018). Infectivity of okra enation leaf curl virus and the role of its V2 protein in pathogenicity. *Virus Research*, 255. <https://doi.org/10.1016/j.virusres.2018.07.007>
- Sanjuán, R., Domingo-Calap, P. (2021). Genetic Diversity and Evolution of Viral Populations. *Encyclopedia of Virology*, 53–61. <https://doi.org/10.1016/B978-0-12-809633-8.20958-8>
- Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I. (2019). GenBank. *Nucleic acids research*, 48. <https://doi.org/10.1093/nar/gkz956>
- Schierup, M.H., Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics*, 156, 879–891. <https://doi.org/10.1093/genetics/156.2.879>
- Scrucca, L., Fraley, C., Murphy, T.B., Adrian E., R. (2023). Model-Based Clustering, Classification, and Density Estimation Using mclust in R. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003277965>
- Seal, S.E., Jeger, M.J., Van den Bosch, F. (2006a). Begomovirus Evolution and Disease Management. *Advances in Virus Research*, 67, 297-316. [https://doi.org/10.1016/S0065-3527\(06\)67008-5](https://doi.org/10.1016/S0065-3527(06)67008-5)
- Seal, S.E., VandenBosch, F., Jeger, M.J. (2006b). Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Critical Reviews in Plant Sciences*, 25, 23–46. <https://doi.org/10.1080/07352680500365257>

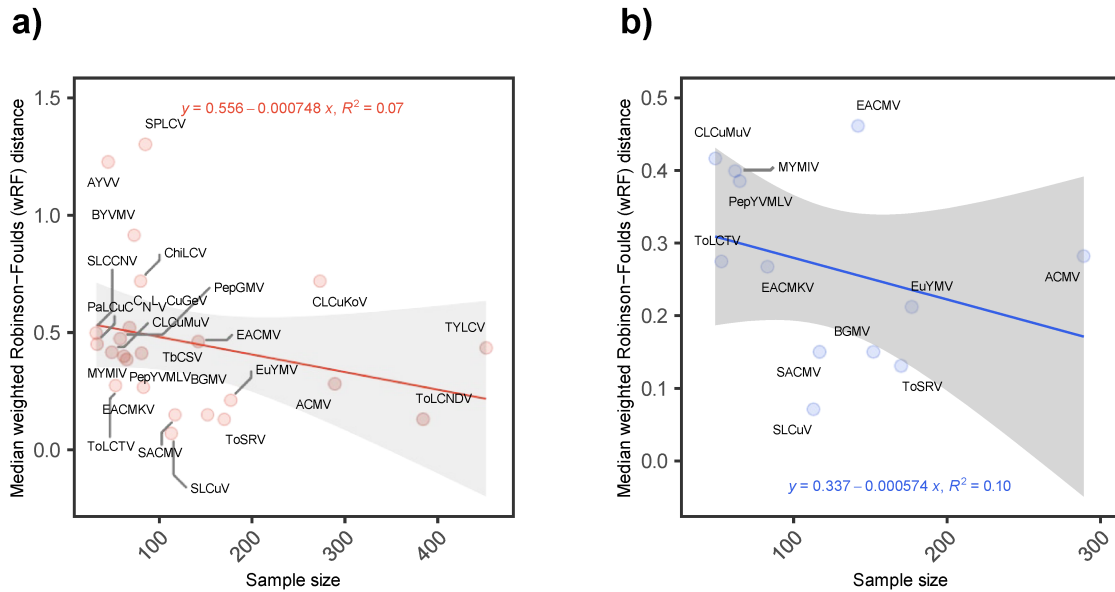
- Simon-Loriere, E., Holmes, E.C. (2011). Why do RNA viruses recombine? *Nature Reviews Microbiology*, 9, 617–626. <https://doi.org/10.1038/nrmicro2614>
- Smith, J.M. (1992). Analyzing the Mosaic Structure of Genes. *Journal of Molecular Evolution*, 34, 126–129. <https://doi.org/10.1007/BF00182389>
- 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. <https://doi.org/10.1371/journal.pbio.3001007>
- Szöllősi, G.J., Tannier, E., Daubin, V., Boussau, B. (2015). The Inference of Gene Trees with Species Trees. *Systematic Biology*, 64, 42–62. <https://doi.org/10.1093/sysbio/syu048>
- Tahir, M., Amin, I., Haider, M.S., Mansoor, S., Briddon, R.W. (2015). Ageratum enation virus- A Begomovirus of weeds with the potential to infect crops. *Viruses*, 7. <https://doi.org/10.3390/v7020647>
- Wiuf, C., Christensen, T., Hein, J. (2001). A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution*, 18(10), 1929–1939. <https://doi.org/10.1093/oxfordjournals.molbev.a003733>
- Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G.W., Robinson, D.J., Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, 78, 2101–2111. <https://doi.org/10.1099/0022-1317-78-8-2101>

Appendices

Supplementary Table 1. Begomovirus species used in this work.

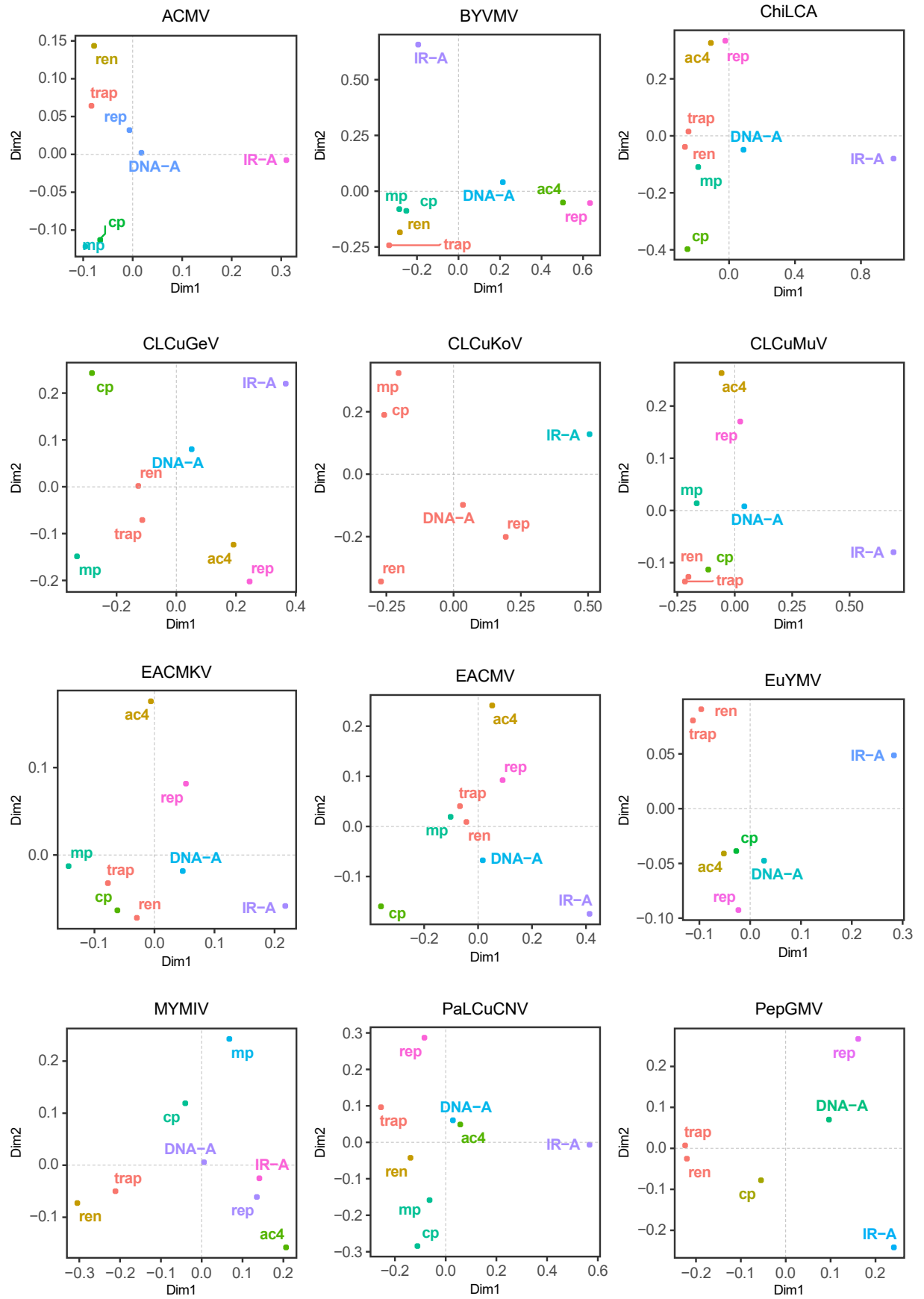
Species	Acronym
<i>African cassava mosaic virus</i>	ACMV
<i>Ageratum yellow vein virus</i>	AYVV
<i>Bean golden mosaic virus</i>	BGMV
<i>Bhendi yellow vein mosaic virus</i>	BYVMV
<i>Chilli leaf curl virus</i>	ChiLCV
<i>Cotton leaf curl Gezira virus</i>	CLCuGeV
<i>Cotton leaf curl Kokhran virus</i>	CLCuKoV
<i>Cotton leaf curl Multan virus</i>	CLCuMuV
<i>East African cassava mosaic Kenya virus</i>	EACMKV
<i>East African cassava mosaic virus</i>	EACMV
<i>Euphorbia yellow mosaic virus</i>	EuYMV
<i>Mungbean yellow mosaic India virus</i>	MYMIV
<i>Papaya leaf curl China virus</i>	PaLCuCNV
<i>Pepper golden mosaic virus</i>	PepGMV
<i>Pepper yellow vein Mali virus</i>	PepYVMLV
<i>South African cassava mosaic virus</i>	SACMV
<i>Squash leaf curl China virus</i>	SLCCNV
<i>Squash leaf curl virus</i>	SLCuV
<i>Sweet potato leaf curl virus</i>	SPLCV
<i>Tobacco curly shoot virus</i>	TbCSV
<i>Tomato leaf curl New Delhi virus</i>	ToLCNDV
<i>Tomato leaf curl Taiwan virus</i>	ToLCTV
<i>Tomato severe rugose virus</i>	ToSRV
<i>Tomato yellow leaf curl virus</i>	TYLCV

Supplementary Figure 1

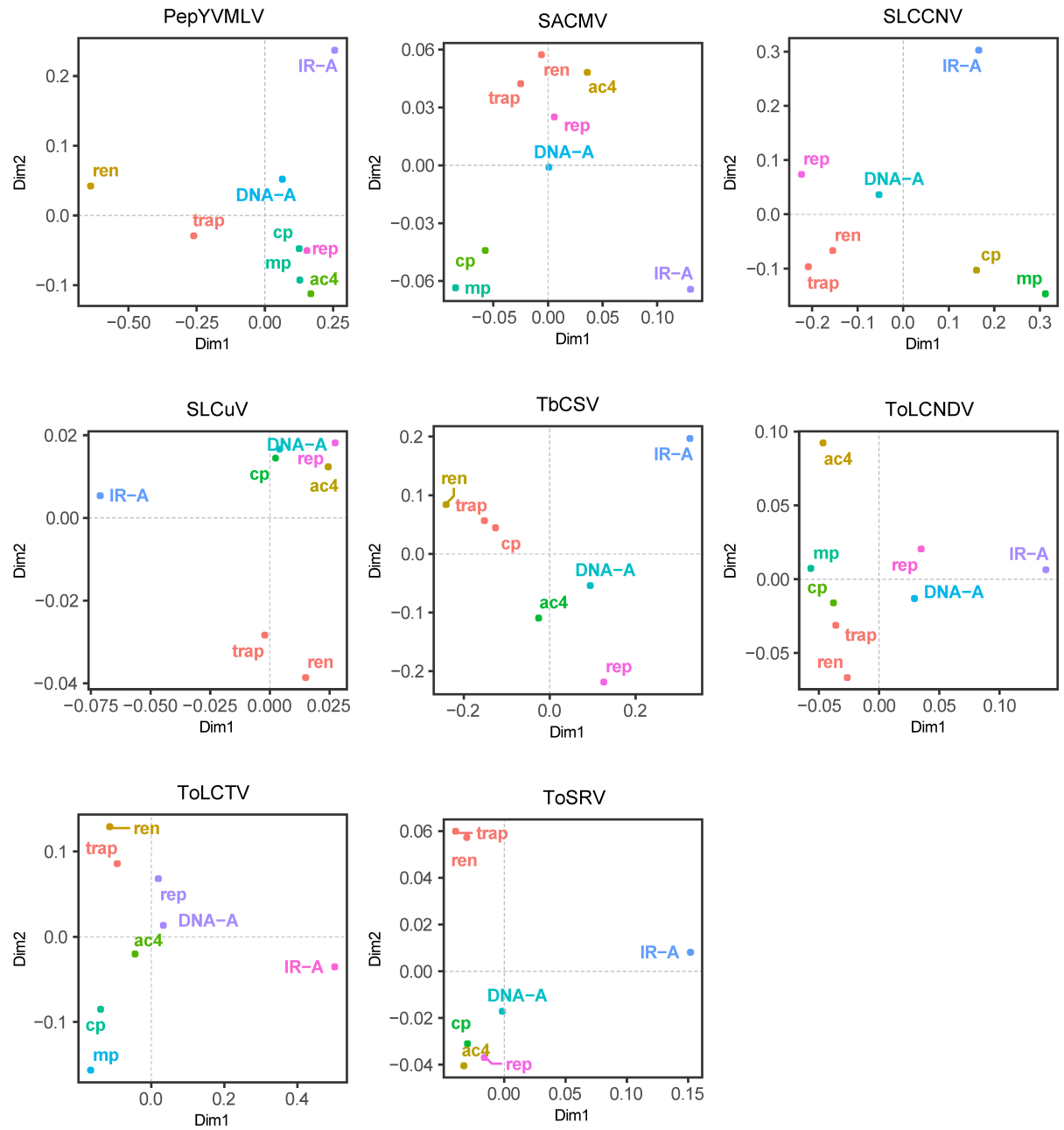


Supplementary Figure 1. Linear regression analysis of sample sizes versus median weighted Robinson-Foulds (wRF) distances for each begomovirus species dataset using the rarefaction technique. Analyses were conducted based on the original sample sizes of the datasets, including recombinant sequences (**a**) and excluding sequences identified as recombinant by at least four RDP methods (**b**).

Supplementary Figure 2

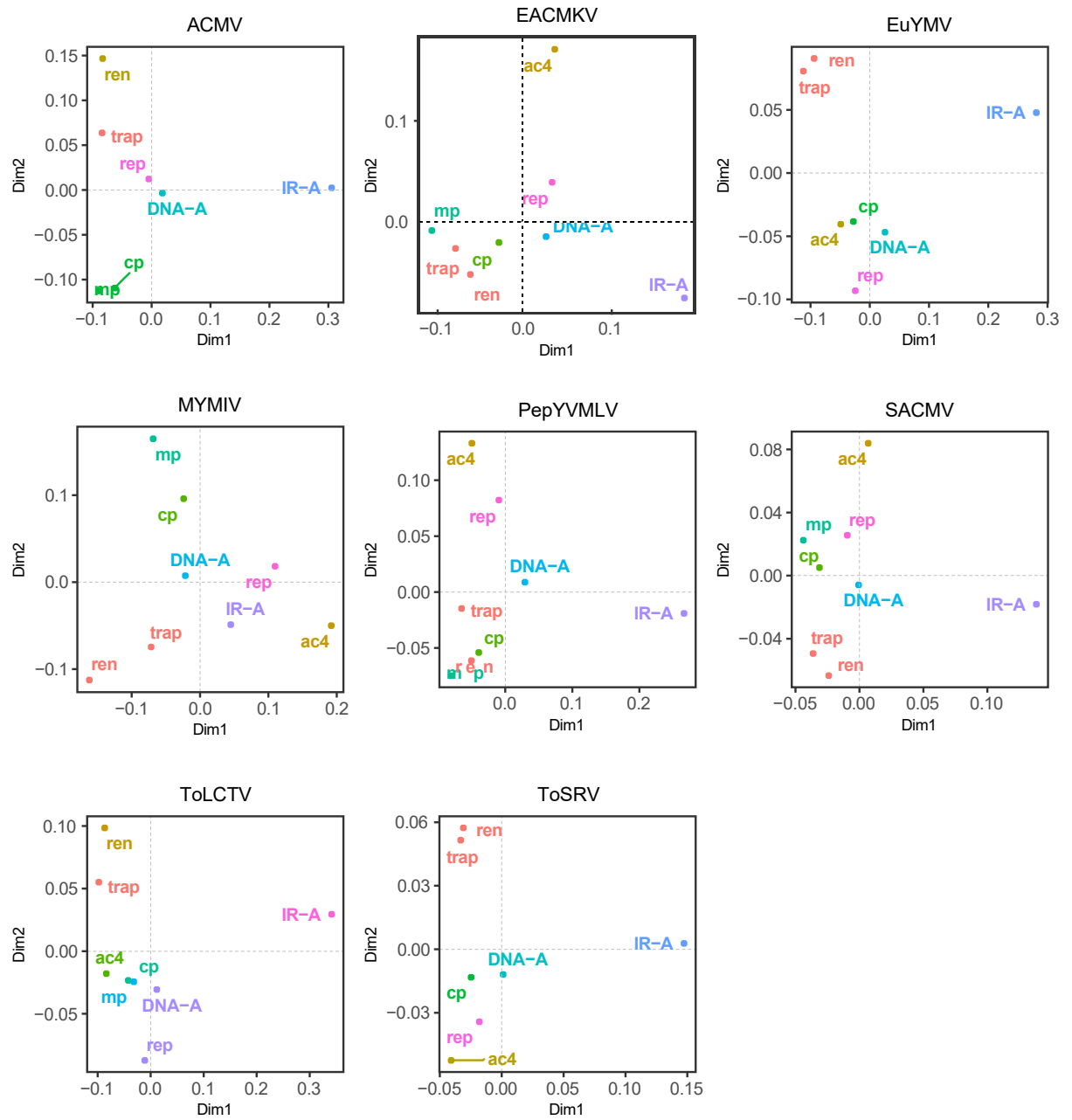


Supplementary Figure 2 (cont.)



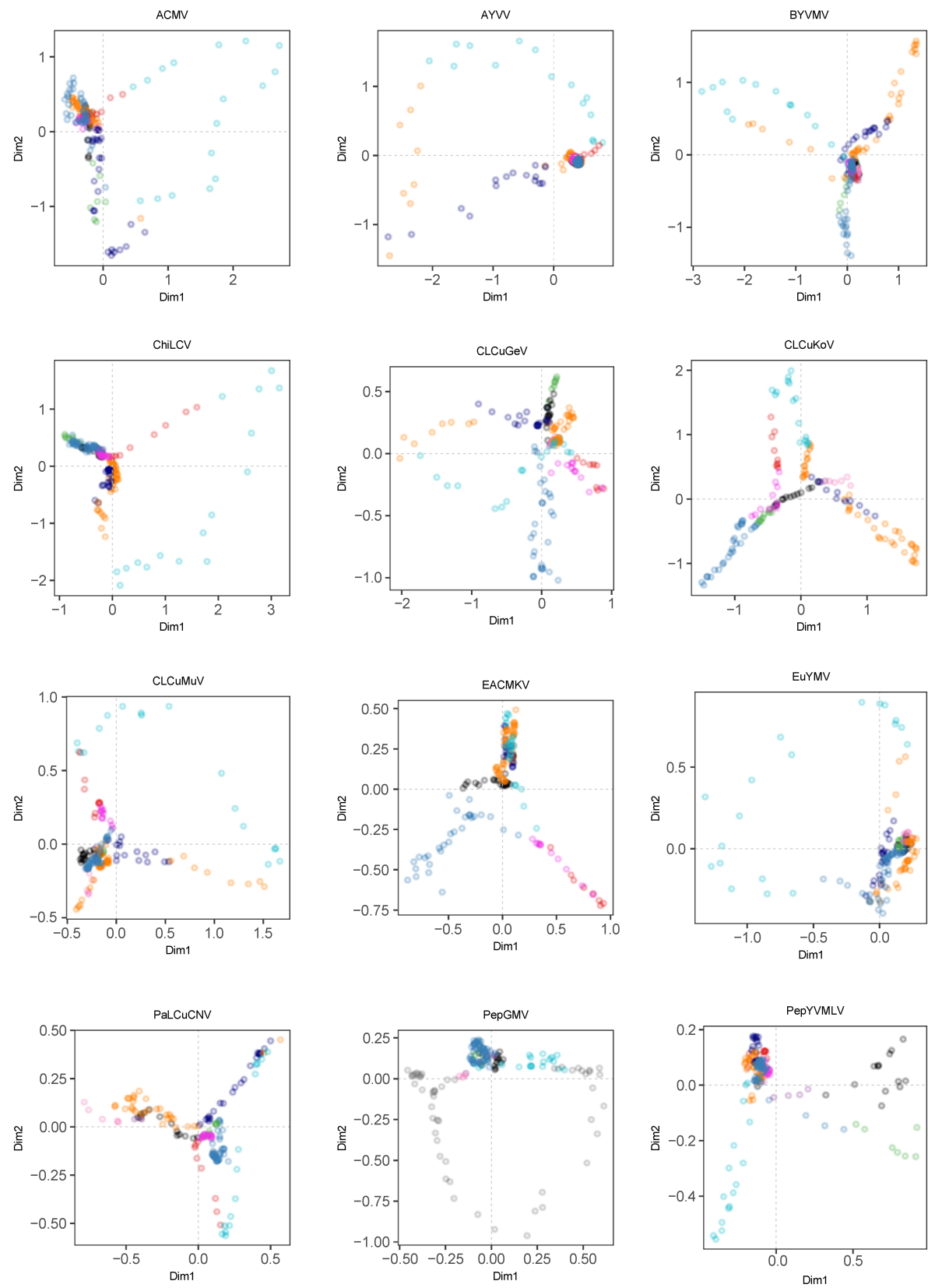
Supplementary Figure 2. Multidimensional scaling (MDS) plots based on weighted Robinson-Foulds (wRF) distance matrices between gene- and intergenic region (IR)-based trees for individual begomovirus species datasets, including recombinant sequences. Points of the same color represent clusters of maximum likelihood (ML) trees with greater similarity.

Supplementary Figure 3

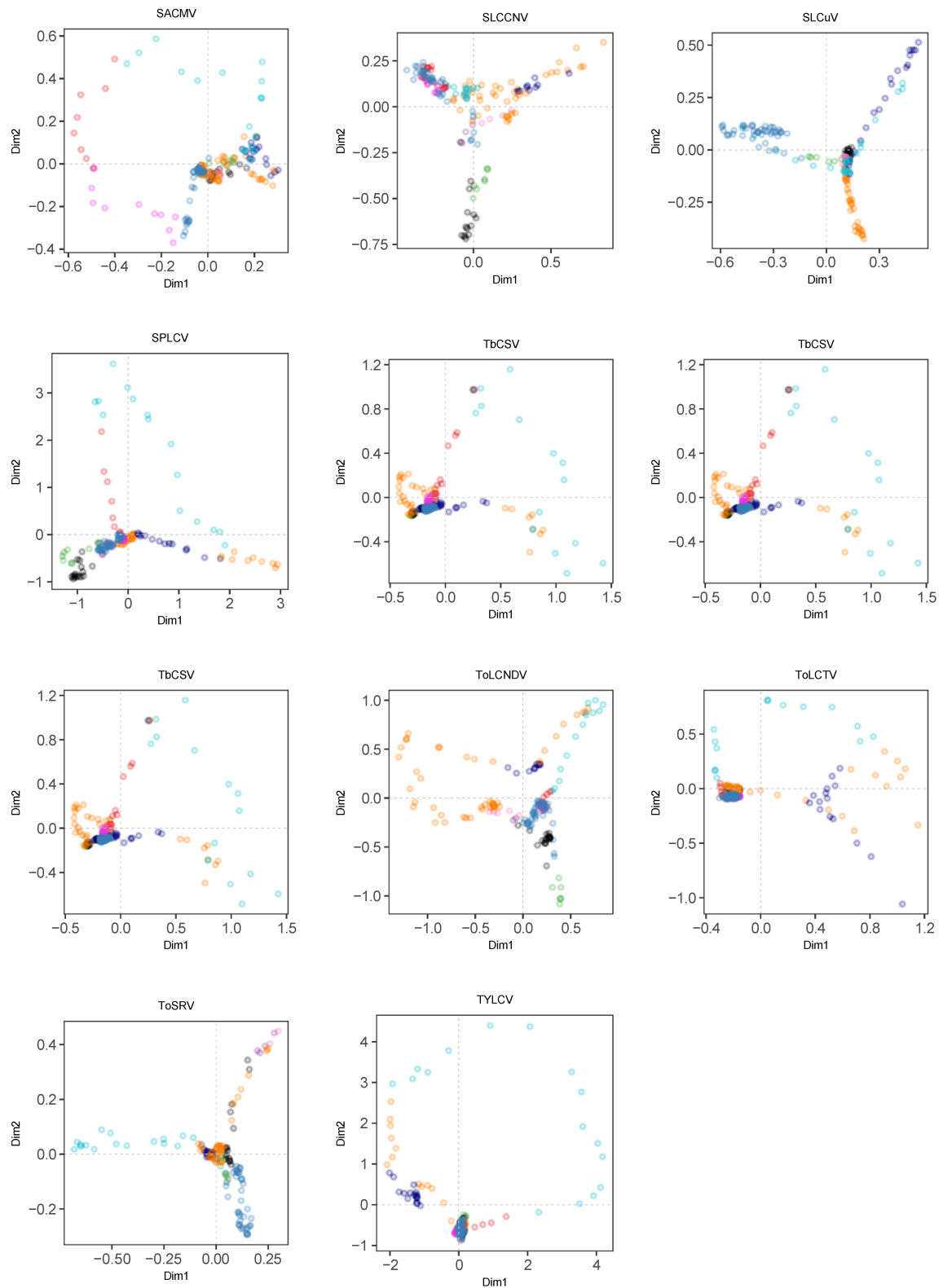


Supplementary Figure 3. Multidimensional scaling (MDS) plots based on weighted Robinson-Foulds (wRF) distance matrices between gene- and intergenic region (IR)-based trees for individual begomovirus species datasets, excluding sequences identified as recombinant by at least four RDP methods. Points of the same color represent clusters of maximum likelihood (ML) trees with greater similarity.

Supplementary Figure 4

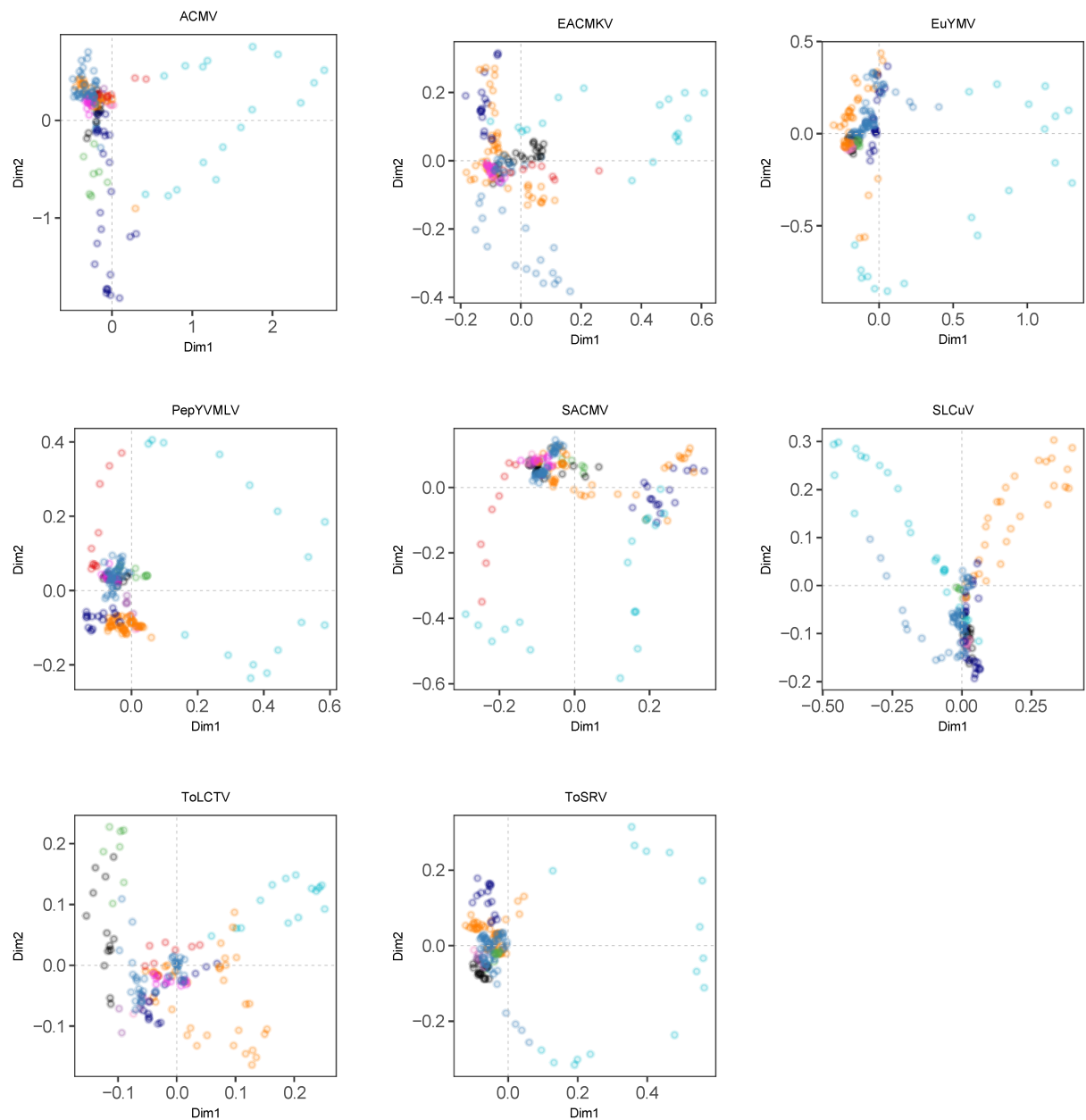


Supplementary Figure 4 (cont.)



Supplementary Figure 4. Multidimensional scaling (MDS) plots based on weighted Robinson-Foulds (wRF) distance matrices between sliding window-based trees for individual begomovirus species datasets, including recombinant sequences. Points are colored according to the schematic representation of the *Begomovirus* DNA-A segment shown at the bottom of Figure 6.

Supplementary Figure 5



Supplementary Figure 5. Multidimensional scaling (MDS) plots constructed from weighted Robinson-Foulds (wRF) distance matrices between sliding window-based trees for individual begomovirus species datasets, excluding sequences identified as recombinant by at least four RDP methods. Points are colored according to the schematic representation of the *Begomovirus* DNA-A segment shown at the bottom of Figure 6.