UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências de Botucatu - Unesp

ESTUDO DAS POPULAÇÕES DE VÍRUS PRESENTES EM PLANTAS DE CITROS CULTIVADAS EM UMA REGIÃO AFETADA PELA MORTE SÚBITA DOS CITROS

EMILYN EMY MATSUMURA

BOTUCATU – SP
2016

Instituto de Biociências – Seção Técnica de Pós-Graduação
Distrito de Rubião Júnior s/n CEP 18618-970 Botucatu-SP Brasil
posgraduacao@ibb.unesp.br

UNIVERSIDADE ESTADUAL PAULISTA
"Julio de Mesquita Filho"

INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

ESTUDO DAS POPULAÇÕES DE VÍRUS PRESENTES EM PLANTAS DE CITROS CULTIVADAS EM UMA REGIÃO AFETADA PELA MORTE SÚBITA DOS CITROS

NOME DA CANDIDATA: EMILYN EMY MATSUMURA
ORIENTADOR: MARCOS ANTONIO MACHADO

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

BOTUCATU – SP
2016

Instituto de Biociências – Seção Técnica de Pós-Graduação
Distrito de Rubião Júnior s/n CEP 18618-970 Botucatu-SP Brasil
posgraduacao@ibb.unesp.br

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências de Botucatu - Unesp

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Matsumura, Emilyn Emy.
Estudo das populações de vírus presentes em plantas de citros cultivadas em uma região afetada pela morte súbita dos citros / Emilyn Emy Matsumura. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Marcos Antonio Machado
Capes: 20202008

1. Cítricos. 2. Vírus. 3. Morte súbita dos citros. 4. Viroses das plantas. 5. Fitopatologia - Diagnóstico.

Palavras-chave: CSDaV; Morte súbita dos citros; Sequenciamento de alta performance; Vírus de citros.

Instituto de Biociências – Seção Técnica de Pós-Graduação
Distrito de Rubião Júnior s/n CEP 18618-970 Botucatu-SP Brasil
posgraduacao@ibb.unesp.br

À minha mãe, Maria B. Dognani, que nunca mediu esforços para apoiar minha vida acadêmica, e por ser meu maior exemplo de dedicação e minha maior incentivadora.

DEDICO

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de doutorado concedida, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro para realização do meu estágio no exterior através do Programa de Doutorado Sanduíche no Exterior (PDSE);

Ao Programa de Pós-Graduação em Ciências Biológicas (Genética) da UNESP-Botucatu e a todos os funcionários, em especial ao coordenador Prof. Dr. Cláudio de Oliveira e Luciana por estarem sempre dispostos a me ajudar quando precisei; a todos os professores do Departamento de Genética por todos os ensinamentos;

Ao meu orientador, Dr. Marcos A. Machado, por ter confiado este trabalho a mim, por ter me dado total liberdade para desenvolvê-lo, e por todo o investimento, apoio e bons conselhos ao longo desses quatro anos;

Ao Dr. Helvécio D. Coletta Filho, por ter participado ativamente do planejamento, desenvolvimento e conclusão deste trabalho, e por toda a ajuda e incentivo no decorrer desses anos;

Ao Dr. Bryce W. Falk, meu supervisor durante meu estágio no exterior na Universidade da Califórnia (UC Davis, CA, EUA), por ter me recebido de forma atenciosa em seu laboratório, por todas as contribuições, paciência, ensinamentos, conselhos e bons momentos durante os 12 meses do meu doutorado sanduíche;

À Dra. Shahideh Nouri, com quem trabalhei efetivamente durante meu estágio no exterior, por todas as contribuições, por me incluir em outros projetos, por compartilhar os seus conhecimentos e pelo bom convívio e amizade ao longo desses 12 meses que estive na UC Davis;

Aos amigos Silvia de Oliveira Dorta e Tiago Silva Oliveira, do Centro de Citricultura Sylvio Moreira, e Luca Nerva, do Instituto para a Proteção Sustentável das Plantas (Itália), pela amizade e por terem me ajudado com determinados experimentos durante o desenvolvimento do projeto;

Aos membros da banca examinadora, Dr. Tatsuya Nagata (UnB – Brasília, DF), Dr. Ricardo Harakava (Instituto Biológico – São Paulo, SP) e Dr. Ivan de Godoy Maia (Unesp-IBB, SP) por todas as sugestões e pela valiosa discussão durante a arguição;

Aos colegas que convivi durante meu estágio no laboratório do Dr. Falk, Inma, Maria, AJ, Ola, Mark, Jared, Jeremy, Wenjie, Yen-Wen, Ana, Tera, Sandra, Eddie, Nida e Máximo, pela amizade, bom convívio e por todas as ajudas e sugestões; à minha querida amiga Cíntia, por toda a ajuda, amizade, carinho, e todos os momentos compartilhados tanto no Brasil quanto em Davis; e a todos os amigos que fiz durante meu estágio no exterior;

A todos os técnicos do Laboratório de Biotecnologia do Centro de Citricultura Sylvio Moreira (CCSM), Rose, Anita, Kleber e Amélia, e todos os outros técnicos do CCSM, por toda a ajuda e paciência durante o tempo que estive no laboratório;

A todos os pesquisadores do CCSM, em especial ao Dr. Pedro L. R. Gonzalez (IB-SP), Dra. Alessandra A. de Souza e Dr. Marco Takita, pela ajuda, contribuições e incentivo sempre que precisei, e a todos os pós-docs, em especial a Dra. Michèle C. Breton e Dra. Carolina Munari por toda a ajuda no início do projeto;

A todos os alunos de pós-graduação e iniciação científica do CCSM, pela boa convivência e aos meus amigos Simone, Carol Sardinha, César, Gabi, Diogo, Bárbara, Willian e Camila, por todos os bons momentos compartilhados dentro e fora do laboratório;

Aos meus amigos mais que especiais, Inaiara, Francisco, Eduardo, Diogo, Tatiane, Laura, Tiago, Silvia, Paulo e Laís, por terem deixado essa vida de doutorando muito mais leve, por terem compartilhado comigo a rotina de laboratório, os cafés diários, os almoços, os jantares, os choros e as gargalhadas, e por terem sido minha família ao longo desses quatro anos, nossa família Estábulo;

Aos meus pais, Maria B. Dognani e Hiroaki Matsumura, pelo amor incondicional, por serem minha inspiração e por sempre estarem dispostos a me apoiarem em todas as minhas decisões; a todos os meus familiares e amigos de infância, que mesmo distantes estão sempre torcendo e me apoiando com muito amor; Ao Leonardo Jo, por todo apoio e todo amor nesses últimos meses de doutorado;

E por fim, agradeço a todos aqueles que contribuíram de alguma forma para o desenvolvimento deste trabalho e a todos os amigos que fiz durante essa caminhada. Muito obrigada!

Resumo

A morte súbita dos citros (MSC) causou a morte ou erradicação de aproximadamente quatro milhões de plantas de laranja doce na principal região citrícola do Brasil. Embora sua etiologia ainda não esteja completamente resolvida, seus sintomas e distribuição (especial e temporal) indicam uma provável doença viral. Trabalhos anteriores associaram a MSC ao vírus da tristeza dos citros (CTV) e ao CSDaV (*Citrus sudden death-associated virus*), no entanto, os resultados obtidos destes trabalhos não são conclusivos. Afim de estudar as populações de vírus presentes em plantas de citros afetadas pela MSC, este trabalho realizou uma análise comparativa, através do sequenciamento de alta performance do transcriptoma e dos pequenos RNAs de plantas sintomáticas e assintomáticas para MSC. Os dados revelaram uma infecção viral mista, incluindo o CTV como vírus mais predominante, seguido do CSDaV, um pararetrovírus endógeno de citros (CitPRV) e dois possíveis novos vírus, putativamente denominados de *Citrus jingmen-like virus* (CJLV) e *Citrus virga-like virus* (CVLV). As análises de correlação com a MSC indicaram uma provável associação de plantas sintomáticas com o CitPRV, enquanto que os dois novos vírus mostraram estar mais associados com plantas assintomáticas. A associação mais evidente foi observada entre plantas sintomáticas e um genótipo específico de CSDaV, o que nos conduziu a um estudo mais específico de variabilidade genética de 31 isolados de CSDaV, obtidos de plantas MSC-sintomáticas e assintomáticas. As análises de cinco regiões genômicas parciais do CSDaV, que correspondem aos domínios da metiltransferase, região de múltiplos domínios (MDR), helicase, RNA-dependente de RNA polimerase e capa proteica, mostraram que a região MDR apresentou maior diversidade genética. A diversidade nucleotídica (π) foi baixa para os isolados de CSDaV, e as análises filogenéticas revelaram a predominância de dois grupos principais, sendo que um deles se mostrou mais associado às plantas sintomáticas, o que foi coerente com os resultados relatados anteriormente. Além disso, os isolados de plantas sintomáticas mostraram maior diversidade nucleotídica, maiores taxas da relação dN/dS, e maior número de aminoácidos alterados, principalmente nas regiões mais próximas da região 5' terminal. Este trabalho gerou informações novas e importantes sobre o patossistema MSC. Tais informações foram consideradas para construção de um clone de cDNA do CSDaV, que poderá ser utilizado em outros estudos de etiologia, e também devem ser consideradas em futuros estudos epidemiológicos.

Palavras-Chave: Morte súbita dos citros, sequenciamento de alta performance, vírus de citros, CSDaV, CTV.

Abstract

Citrus sudden death (CSD) is a disease that caused death or eradication of approximately four million orange trees in a very important citrus region in Brazil. Although its etiology is still not completely clear, symptoms and distribution of affected plants indicate a viral disease. Previous works have associated *Citrus trsiteza vírus* (CTV) and *Citrus sudden death-associated virus* (CSDaV) with CSD, but all attempts to prove these association have failed so far. In attempting to study the virome of citrus plants affected by CSD, and compare the frequency and diversity of viruses between CSD-symptomatic and asymptomatic plants, we have done a comparative high-throughput sequencing analysis of the transcriptome and small RNAs from those plants by using Illumina platform. The data revealed a mixed infection that included CTV as the most predominant virus, followed by the CSDaV, *Citrus endogenous pararetrovirus* (CitPRV) and two putative novel virus tentatively named *Citrus jingmen-like virus* (CJLV) and *Citrus virga-like virus* (CVLV) in this study. We demonstrated a likely association of the CSD-symptomatic plants with CitPRV, whereas the two putative novel viruses showed to be more associated with CSD-asymptomatic plants. The strongest association was observed between CSD-symptomatic plants and a specific CSDaV genotype, which led us to study more specificaly the genetic variability among 31 CSDaV isolates obtained from both CSD-symptomatic and asymptomatic plants. Analyses of partial nucleotide sequences of five domains of the CSDaV genomic RNA, including those encoding for the methyltransferase, the multi-domain region (MDR), the helicase, the RNA-dependent RNA polymerase and the coat protein, showed that the MDR coding region was the most diverse region assessed here. The nucleotide diversity (π) was low for CSDaV isolates, but the phylogenetic analyses revealed the predominance of two main groups, one of which showed a strong association with CSD-symptomatic plants, supporting the previous results obtained from the high-throughput analyses. Isolates obtained from CSD-symptomatic plants, compared to those obtained from asymptomatic plants, showed higher nucleotide diversity, nonsynonymous and synonymous substitution rates and number of amino acid changes on the coding regions located closer to the 5' end region of the genomic RNA. This work generated new and valuable information that were used to construct an infectious clone of the CSDaV, which will be used in further etiology studies, and that must be considered in further epidemiological studies.

Keywords: Citrus sudden death, hight- throughput sequencing, citrus viruses, CSDaV, CTV

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

1.1. Aspectos gerais da citricultura e sua importância econômica para o Brasil

Considerando o valor comercial e nutricional, o gênero *Citrus* representa uma das mais importantes culturas frutíferas do mundo (SINGH; RAJAM, 2009). Diversos registros indicam que essa cultura teve origem na China e no Sudeste Asiático, onde foi cultivada por mais de 4.000 anos (NEVES et al., 2010; SINGH; RAJAM, 2009), e de onde se dispersou para outras regiões do mundo, atingindo mais de 140 países (LIU; HEYING; TANUMIHARDJO, 2012). Entretanto, as melhores condições para o desenvolvimento das espécies do gênero *Citrus* foram encontradas em regiões tropicais e subtropicais, sendo as laranjas, os limões, as limas, as toranjas e as tangerinas, os principais exemplos de citros cultivados nestas regiões (ARTUR; FRANCISCO, 2014; LIU; HEYING; TANUMIHARDJO, 2012).

Introduzidas no Brasil logo no início da colonização, as plantas de citros expandiram-se por grande parte do território nacional, sendo cultivadas, principalmente, nos estados do Pará, Sergipe, Bahia, Minas Gerais, São Paulo, Paraná e Rio Grande do Sul (ARTUR; FRANCISCO, 2014; IBGE, 2016; NEVES et al., 2010), fato que coloca o Brasil entre os maiores produtores de citros do mundo, juntamente com China, Estados Unidos, Índia, México e Espanha (LIU; HEYING; TANUMIHARDJO, 2012; USDA, 2016). No entanto, foi com a laranja doce que o Brasil conseguiu reconhecimento mundial, atingindo a primeira colocação no *ranking* de produção, e se destacando como maior produtor e exportador de suco de laranja concentrado do mundo (ARTUR; FRANCISCO, 2014; USDA, 2016). Com, aproximadamente, 700 mil hectares atuais de área plantada, o Brasil produziu cerca de 16 milhões de toneladas de laranjas na última safra, o que representa mais de 30% de toda a produção mundial (IBGE, 2016), além de ter sido responsável por mais da metade de toda a produção de suco de laranja do mundo (≈ 54,6%), o qual foi praticamente todo destinado a exportação (USDA, 2016), gerando lucros indiscutíveis para o país.

Embora diferentes Estados brasileiros tenham contribuído significativamente com o rendimento total na produção de laranja, a maior contribuição vem do Estado de São Paulo, que conta com quase 450 mil hectares de área plantada e foi responsável por, aproximadamente, 73% da produção nacional na última safra (IBGE, 2016). Dada a

importância desse Estado para a citricultura brasileira, e a alta suscetibilidade das plantas de laranja ao ataque de diferentes pragas e doenças (SINGH; RAJAM, 2009), a persistência de diversos problemas fitossanitários nesta região ainda é considerada uma grande ameaça para a citricultura do país. Doenças como a clorose variegada dos citros (CVC), pinta-preta, leprose, cancro cítrico e o *huanglongbing* (HLB) representam problemas ainda não completamente resolvidos, gerando preocupação e aumento de custo de produção para os citricultores (SINGH; RAJAM, 2009).

1.2. Da tristeza dos citros à morte súbita dos citros

Além das doenças citadas acima, uma das mais devastadoras doenças que atingiu a cultura de citros no Brasil e no mundo foi a denominada tristeza dos citros, causada pelo vírus de mesmo nome, CTV (*Citrus tristeza virus*) (SINGH; RAJAM, 2009). O CTV é um vírus limitado às células do floema, membro da família *Closteroviridae* e gênero *Closterovirus*. Possui um genoma de RNA fita simples e senso positivo, com aproximadamente 19,3 Kb de comprimento. Seu genoma é organizado em 12 *Open Reading Frames* (ORFs) e duas regiões não-transcritas, localizadas nas regiões 5’ e 3’ terminais (DAVINO et al., 2013). A ORF1 (1a e 1b) é traduzida diretamente do RNA genômico e expressa uma poliproteína relacionada com a replicação do genoma viral (DAVINO et al., 2013). As demais ORFs são expressas através dos RNAs subgenômicos 3’ coterminais e codificam as proteínas p6, p65, p61, p27, p25, p18, p13, p20, p23 e p33, requeridas para diferentes atividades, como montagem do virion, movimento célula-a-célula, supressão do silenciamento de genes e invasão ao hospedeiro (FOLIMONOVA, 2013; MORENO et al., 2008).

Dependendo da variedade copa, do porta-enxerto e do haplótipo ou mistura de haplótipos de CTV, a tristeza dos citros pode se manifestar através de diferentes sintomas, como amarelecimento das plântulas (típico de plântulas de laranja Azeda, toranjas e limões), caneluras no lenho, baixa produtividade e qualidade dos frutos (independente do porta-enxerto utilizado) e declínio rápido (atingindo, principalmente, laranjas doces enxertadas sobre o porta-enxerto laranja Azeda) (DAVINO et al., 2013). No século passado, a tristeza dos citros foi responsável pela perda (morte ou improdutividade) de mais de 100 milhões de pés de laranjas ao redor do mundo (BOVÉ; AYRES, 2007). No Brasil, entre os anos de 1940 e 1950, nove entre onze milhões de plantas de laranjas doces (*Citrus sinensis* L. Osb.) enxertadas sobre o porta-enxerto laranja Azeda (*Citrus aurantium*), principal porta-enxerto

utilizado naquela época em pomares no Estado de São Paulo, foram afetadas pela doença; evento este que quase destruiu a citricultura paulista (ROMÁN et al., 2004). Acredita-se que o CTV tenha se originado na China e atingido outros países citrícolas através da propagação de borbulhas infectadas (MORENO et al., 2008), enquanto que a disseminação local entre plantas ocorre por afídeos vetores, sendo o pulgão preto (*Toxoptera citricida*) considerado o principal e mais eficiente vetor de CTV no Brasil (MORENO et al., 2008).

Diante de todo o prejuízo causado pela tristeza dos citros e afim de retomar a citricultura no Brasil, a principal região citrícola do país que compreende o Estado de São Paulo e a região do Triângulo Mineiro, Minas Gerais, foi submetida à algumas mudanças (BOVÉ; AYRES, 2007), como o uso da proteção cruzada ou pré-imunização com complexos fracos e protetivos de CTV, e a substituição do porta-enxerto laranja Azeda, largamente cultivada nesta região e considerada suscetível a tristeza dos citros (BOVÉ; AYRES, 2007; COSTA et al., 2010). A combinação laranja doce enxertada sobre o porta-enxerto laranja Azeda foi substituída por combinações mais tolerantes à doença, utilizando-se de porta-enxertos alternativos, como o limão Cravo (*Citrus limonia* Osb), limão Volkameriano (*Citrus volkameriana*), Cleopatra mandarin (*Citrus reshni*), Sunki mandarin (*Citrus sunki*), limão Rugoso (*Citrus jambhiri* Lush) e *Poncirus trifoliata* e seus híbridos (citranges e citrumelos) (BOVÉ; AYRES, 2007). No entanto, foi com o limão Cravo que a região citrícola foi reestabelecida. Este porta-enxerto tornou-se largamente utilizado, principalmente devido suas características agronômicas importantes, como compatibilidade com todas as variedades copa, precocidade, alta produção na copa e tolerância à seca (POMPEU JUNIOR, 2001). Até o ano de 2000, 85% das laranjas doces na principal região citrícola do país estavam enxertados sobre o porta-enxerto limão Cravo (BOVÉ; AYRES, 2007).

Tendo em vista a importância do limão Cravo no Brasil, a identificação, em 1999, de plantas de laranja doce sobre este porta-enxerto com uma nova sintomatologia de declínio rápido seguido de morte da planta, tornou-se uma nova ameaça à citricultura brasileira (CANTÚ et al., 2008; ROMÁN et al., 2004). A nova doença foi denominada de morte súbita dos citros (MSC) (MÜLLER et al., 2002) e foi responsável, entre os anos de 2001 e 2010, pela morte ou erradicação de cerca de quatro milhões de pés de laranja na principal região citrícola do Brasil (MACHADO et al., 2004; GOMES et al., 2008; NEVES et al., 2010).

1.3. Aspectos gerais da morte súbita dos citros, sintomatologia e etiologia da doença

As primeiras plantas sintomáticas para MSC foram identificadas em alguns municípios do sudoeste do Estado de Minas Gerais (Comendador Gomes, Frutal e Uberlândia) e norte do Estado de São Paulo (Altair, Barretos, Colômbia e Guaraci) (BASSANEZI et al., 2003). Plantas com MSC apresentam sintomas de declínio generalizado, caracterizado por folhas de coloração verde-pálida em toda a copa, diferentes níveis de desfolha, ausência de novos brotos e morte do sistema radicular, resultantes, principalmente, da degeneração do floema (MACCHERONI et al., 2005; ROMÁN et al., 2004). Porém, o sintoma mais característico da doença, e que tem sido utilizado como diagnóstico, é a presença de uma coloração amarelada na região cambial do porta-enxerto (CANTÚ et al., 2008; MÜLLER et al., 2002). A gravidade dos sintomas aumenta quando a demanda de água é alta, e a morte das plantas pode ocorrer entre 1 a 12 meses após o aparecimento dos primeiros sintomas, dependendo da estação do ano e da variedade da planta (MACCHERONI et al., 2005). A princípio, foram observados sintomas de MSC em plantas de laranja Westin, Hamlin, Natal, Valência, Pera e Rubi, todas enxertadas sobre limão Cravo (BASSANEZI et al., 2003). No entanto, observações subsequentes revelaram que laranjas doces sobre o porta-enxerto limão Volkameriano também são suscetíveis à doença (ROMÁN et al., 2004). Outros porta-enxertos, como tangerinas Cleópatra e Sunki e *Poncirus trifoliata* e seus híbridos (citranges e citrumelos) tem se mostrado tolerantes a MSC (MACHADO et al., 2004).

O principal desafio no controle e estudo da MSC tem sido a identificação do seu agente causal. Várias hipóteses foram levantadas a respeito da sua etiologia, no entanto, essa questão ainda não está completamente resolvida (GOMES et al., 2008). Fatores abióticos, como estresse ambiental e/ou nutricional e manejo inadequado do pomar, foram desconsiderados quando pomares bem conduzidos começaram a apresentar plantas com sintomas da doença (MACHADO et al., 2004). Além disso, o fato da MSC ter sido transmitida por união de tecidos reforçava a hipótese de que um agente biótico estaria associado à doença (YAMAMOTO et al., 2011). Alguns patógenos conhecidos de solo, como *Phytophthora*, *Fusarium* e nematóides, foram colocados em segundo plano, já que suas populações eram baixas em talhões com alta incidência de plantas com MSC (MACHADO et al., 2004). Bactérias e fitoplasmas também foram excluídas em razão dos sintomas e da ausência desses patógenos nas plantas afetadas (MÜLLER et al., 2002). A hipótese de serem viróides também foi desconsiderada, principalmente, pela forma de transmissão desses

patógenos não ser consistente com a observada em plantas com MSC (MÜLLER et al., 2002). Diante dessas observações e baseado nos sintomas e distribuição espacial e temporal da doença, a hipótese de que a MSC seja uma doença viral, possivelmente transmitida por um inseto vetor, é considerada a mais provável (BASSANEZI et al., 2003).

Na busca por um agente viral causador dos sintomas de MSC, estudos demonstraram que o padrão espacial e temporal de plantas afetadas pela doença eram notavelmente semelhantes à disseminação do CTV (BASSANEZI et al., 2003). Além disso, verificou-se que a MSC se assemelha a tristeza dos citros quanto aos sintomas, características epidemiológicas (BASSANEZI et al., 2003), e características anatômicas (ROMÁN et al., 2004), o que levou os pesquisadores a sugerirem que a nova doença pudesse ser causada por um novo variante de CTV (BASSANEZI et al., 2003; BOVÉ; AYRES, 2007). No Brasil, o vírus da tristeza pode ser encontrado como populações mistas de haplótipos em todas as espécies e variedades de citros devido ao seu caráter endêmico e à transmissão generalizada pelo pulgão preto (MÜLLER et al., 2002; SOUZA, 2002). No entanto, a possibilidade de novos variantes de CTV desenvolverem sintomas em plantas de citros previamente consideradas tolerantes ao vírus, sempre foi preocupante. TARGON et al. (2003) e RIVAS-VALENCIA et al. (2008) caracterizaram a estrutura populacional de isolados de CTV, em plantas com e sem sintomas de MSC, através da análise do gene da capa proteica do vírus. A estrutura populacional de CTV se mostrou heterogênea para os dois grupos de plantas, e não foi identificado nenhum padrão dominante de CTV em plantas com MSC. A variabilidade genética do CTV, em plantas com e sem sintomas de MSC, também foi estudada a partir de análises dos genes RdRP e HSP70h do vírus (GOMES et al., 2008). Apesar da sintomatologia, as plantas apresentaram-se co-infectadas com variantes de CTV similares (GOMES et al., 2008). Esses resultados não conclusivos tem dificultado a identificação de uma associação clara entre isolados de CTV e MSC.

Em 2005, na busca por novos variantes de CTV em plantas afetadas pela MSC, um novo vírus, membro do gênero *Marafivirus* (família *Tymoviridae*) foi descrito como associado à doença, e denominado de CSDaV (*Citrus sudden death-associated vírus*) (MACCHERONI et al., 2005). MACCHERONI et al. (2005) verificaram a presença desse vírus em 99,7% das plantas analisadas com sintomas da doença, não detectando-o em plantas crescidas em regiões geográficas mais afastadas do local de incidência. O CSDaV é constituído por uma fita simples de RNA, senso positivo e com um tamanho de aproximadamente 6,8 Kb (MACCHERONI et al., 2005). Assim como o CTV, é um vírus limitado ao floema, e possui

um genoma organizado por duas ORFs. A ORF1 compreende quase todo o genoma do vírus (p240) e é gerada a partir do RNA genômico, apresentando domínios proteicos conservados para metiltransferase, protease, helicase, RdRp e a maior das duas proteínas do capsídeo (p22,5), sendo a menor delas (p21) codificada pelo RNA sub-genômico na região 3' coterminal. A ORF2, na região 3' terminal, é uma ORF pequena (p16) que parece codificar para uma proteína semelhante a proteína de movimento de um membro do gênero *Maculavirus*, também da família *Tymoviridae* (MACCHERONI et al., 2005). A presença do CSDaV também foi detectada em três espécies de pulgões vetores de CTV (*Toxoptera citricida*, *Aphis gossypii* e *A. spiraecola*) (MACCHERONI et al., 2005; SANTOS, 2011). No entanto, devido a diversas dificuldades em se estudar esse patossistema, o postulado de Koch para o CSDaV ainda não foi cumprido. Tentativas de purificação e transmissão somente do CSDaV para plantas de citros, assim como tentativas de construção de um clone infeccioso de CSDaV, não geraram resultados satisfatórios (SANTOS, 2011; NUNES, 2009). Desta forma, a associação do CSDaV com a MSC continua não completamente elucidada (CANTÚ et al., 2008).

1.4. Situação atual da morte súbita dos citros e novas estratégias para o estudo da doença

Como relatado até aqui, a etiologia da MSC não está completamente elucidada, principalmente em função das dificuldades de purificação, transmissão e caracterização dos vírus potencialmente associados à doença. A perda de aproximadamente quatro milhões de plantas de laranja entre os anos de 2001 e 2010 e a forma rápida com que a doença se espalhou entre os municípios dos Estados de Minas Gerais e São Paulo neste mesmo período (GOMES et al., 2008), levaram os citricultores da principal região citrícola do país a tomarem algumas medidas de contenção para o controle da doença. A comercialização e o trânsito de materiais vegetativos (mudas, borbulhas, porta-enxertos e sementes) produzidos a céu aberto e procedentes das áreas de ocorrência da doença foram proibidos (PORTARIA CDA-9, 2002). Entretanto, a medida mais eficiente para a superação do problema foi a substituição do porta-enxerto limão Cravo por porta-enxertos considerados tolerantes à MSC, como, principalmente, tangerine Sunki e citrumelo Swingle (POMPEU JÚNIOR; BLUMER, 2008). Embora a MSC tenha se tornado um problema controlável, todos os porta-enxertos utilizados em substituição ao limão Cravo exigem irrigação, o que acabou aumentando o custo de produção para o citricultor.

Essas sucessivas trocas de porta-enxertos, afim de impedir o desenvolvimento de determinadas doenças, o não conhecimento do agente etiológico definitivo da MSC, e a detecção de sintomas de MSC em plantas de laranjas doces enxertadas sobre diferentes porta-enxertos, até então não considerados suscetíveis à doença, como limão Rugoso (*Citrus jambiri*), Sunki da China (*Citrus sunki*) e *Citrus pennivisiculata* Lush (comunicação pessoal - não publicado), geram preocupações com relação ao futuro da citricultura no país. Poderia o agente causal da MSC começar a desenvolver sintomas em plantas consideradas tolerantes à doença? Desta forma, é clara ainda a necessidade de se confirmar a etiologia viral e entender a real função dos vírus atualmente associados a doença.

Grande parte dos estudos envolvendo plantas afetadas pela MSC utilizaram estratégias convencionais, como ELISA (*enzyme-linked immnosorbet assay*) e RT-PCR (*reverse transcription polymerase chain reaction*), para detecção dos vírus putativamente associados à doença (TARGON et al., 2003; GOMES et al., 2008; RIVAS-VALENCIA et al., 2008; SANTOS, 2011). No entanto, esses métodos tradicionais de detecção dependem de um conhecimento prévio do genoma viral, sendo altamente específicos e excluindo a identificação de outros vírus conhecidos e não conhecidos (COETZEE et al., 2010; LI et al., 2012), o que pode, possivelmente, gerar resultados não precisos quanto ao estado etiológico da planta estudada (LI et al., 2012). A estratégia de sequenciamento por *shotgun*, realizada por MACCHERONI et al. (2005), de plantas com sintomas de MSC e que resultou na identificação do CSDaV, gerou um baixo número de *reads*, o que dificultou o estudo da frequência dos vírus detectados, a diferenciação dos virus com alta diversidade genética e a identificação de outros novos vírus. Diante disso, o uso de novas metodologias eficientes na identificação e caracterização das populações de vírus que possam estar presentes e associados à plantas afetadas pela MSC, assim como o estudo da diversidade genética de populações de vírus já associados à doença (CTV e CSDaV) e dos fatores envolvidos na diversidade e na estrutura populacional desses vírus, são fundamentais para um melhor entendimento desse patossistema e, consequentemente, para a escolha de estratégias mais eficazes para o controle da doença.

A tecnologia de sequenciamento de nova geração (NGS - *Next Generation Sequencing*) tem sido amplamente utilizada, com sucesso, na detecção e caracterização de vírus já conhecidos e não conhecidos em plantas hospedeiras (JO et al., 2015; LI et al., 2012; PRABHA; BARANWAL; JAIN, 2013). O sequenciamento de alta performance do transcriptoma total (RNA-seq) e dos pequenos RNAs (sRNAs) de plantas, especialmente com

uma etiologia viral desconhecida, tem demonstrado ser uma abordagem poderosa e promissora na detecção de ambos os vírus de RNA e DNA (LI et al., 2012; NOURI et al., 2016), principalmente, pelo fato do sequenciamento de transcriptoma e de sRNAs permitirem o acesso à maioria dos RNAs expressos e aos sRNAs derivados do mecanismo de defesa das plantas contra vírus, respectivamente (LI et al., 2012; NOURI et al., 2016). O presente trabalho utilizou da tecnologia de sequenciamento de nova geração para um estudo geral das populações de vírus presentes em plantas de citros crescidas numa região afetada pela MSC, através do sequenciamento de trasncriptoma e de sRNAs. Os dados gerados permitiram o acesso às populações de vírus mais e menos predominantes e a comparação da frequência dos vírus identificados entre plantas sintomáticas e assintomáticas para a MSC, o que também serviu de base para um estudo mais específico de diversidade genética e construção de um clone infeccioso de CSDaV.

2. OBJETIVOS

2.1. *Objetivos gerais*

Acessar as populações de vírus presentes em plantas de citros crescidas em uma região afetada pela morte súbita dos citros e proceder uma análise comparativa entre plantas com e sem sintomas da doença, visando um estudo da frequência e da diversidade genética dos vírus atualmente associados à MSC e a identificação de outros vírus, conhecidos ou não, que possam ter alguma função no desenvolvimento dos sintomas.

2.2. *Objetivos específicos*

- Proceder o sequenciamento de alta performance do transcriptoma total (RNA-seq) e dos pequenos RNAs obtidos de diferentes plantas de citros crescidas em uma região afetada pela morte súbita dos citros;
- Utilizar os dados de sequenciamento para a montagem dos *contigs* e selecionar aqueles que mostraram alguma similaridade com proteínas virais, disponíveis no banco em dados;
- Analisar a riqueza populacional de vírus, visando identificar as espécies mais e menos predominantes e estimar a frequência dos vírus detectados, a partir do número de *reads* e da média de cobertura dos *contigs* derivados dos respectivos vírus;
- Validar a presença dos vírus candidatos por RT-PCR e sequenciamento de Sanger em amostras de RNA extraídas de plantas com e sem sintomas de MSC e comparar, filogeneticamente, as sequências virais obtidas com outros membros já descritos das respectivas famílias virais;
- Comparar os genótipos de CTV e CSDaV presentes em plantas sintomáticas e assintomáticas;
- Analisar e comparar mais especificamente a diversidade do CSDaV em plantas com sintomas e sem sintomas de MSC a partir da amplificação, clonagem e sequenciamento de cinco regiões do genoma do vírus; e contruir um clone infeccioso de CSDaV para futuros estudos sobre a função desse vírus em plantas de citros hospedeiras.

3. CAPÍTULO I:

Deep Sequencing Analysis of RNAs from Citrus Plants Grown in Citrus Sudden Death-affected Regions Reveals Diverse Known and Putative Novel Viruses

Deep Sequencing Analysis of RNAs from Citrus Plants Grown in Citrus Sudden Death-affected Regions Reveals Diverse Known and Putative Novel Viruses

Emilyn Emy Matsumura[1,2]; Helvecio Della Coletta Filho[2]; Shahideh Nouri[3]; Bryce W. Falk[3]; Luca Nerva[4]; Tiago Silva Oliveira[2]; Marcos Antonio Machado[2,*]

[1]Instituto de Biociências de Botucatu, Universidade Estadual Paulista, Botucatu, SP, Brazil; [2]Laboratório de Biotecnologia, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Cordeirópolis, SP, Brazil; [3]Department of Plant Pathology, University of California, Davis, California, USA; [4]Institute for Sustainable Plant Protection (IPSP), CNR, Turin, Italy.

Abstract

Citrus sudden death (CSD) is a disease that caused death or eradication of approximately four million orange trees in a very important citrus region in Brazil. Although its etiology is still not completely clear, symptoms and distribution of affected plants indicate a viral disease. In a search for viruses associated with CSD, we have done a comparative high-throughput sequencing analysis of the transcriptome and small RNAs from CSD-symptomatic and asymptomatic citrus plants by using Illumina platform. The data revealed a mixed infection that included *Citrus tristeza virus* (CTV) as the most predominant virus, followed by the *Citrus sudden death-associated virus* (CSDaV), *Citrus endogenous pararetrovirus* (CitPRV) and two putative novel virus tentatively named *Citrus jingmen-like virus* (CJLV) and *Citrus virga-like virus* (CVLV) in this study. We demonstrated a strong association of the CSD-symptomatic plants with a specific CSDaV genotype and a likely association with CitPRV as well, whereas the two putative novel viruses showed to be more associated with CSD-asymptomatic plants. This is the first high-throughput sequencing-based study of the viral sequences present in CSD-affected citrus plants, which generated valuable information for further CSD etiology studies.

Keywords: Citrus sudden death; CSDaV; CTV; plant viruses; high-throughput sequencing.

3.1. Introduction

Citrus sudden death (CSD) is a disease that was first detected in 1999 in citrus groves located in the municipality of Comendador Gomes (southwestern Minas Gerais State), Brazil (MÜLLER et al., 2002). In that time, CSD was found to affect only plants of sweet orange (*Citrus sinensis* L. Osb.) grafted on Rangpur lime rootstock (*Citrus limonia* L. Osb.), a very important non-irrigated rootstock used in Brazil (MÜLLER et al., 2002). However, CSD quickly spread into northern part of São Paulo State and, since then, has caused death or eradication of four million orange trees in those mentioned regions (GOMES et al., 2008; ROMÁN et al., 2004). The symptoms of CSD are characterized by a general decline, including pale green coloration of the leaves, overall defoliation, death of the roots, and presence of a characteristic yellow stain in the rootstock bark (BASSANEZI et al., 2007). Later, CSD-symptoms were also detected in sweet oranges grafted on the other rootstocks, such as *Citrus volkameriana* (ROMÁN et al., 2004), *Citrus jambiri* and *Citrus pennivisiculata* Lush (personal communication - data not published).

The main challenge in studying CSD is that the etiology has not been definitively determined, even after seventeen years from its first detection. Based on the symptoms and distribution of the CSD-affected plants, previous works have hypothesized that a new variant of *Citrus tristeza virus* (CTV), a member of the family *Closteroviridae* and one of the most economically important citrus viruses, might be associated in developing CSD symptoms (BASSANEZI et al., 2003; GOMES et al., 2008; RIVAS-VALENCIA et al., 2008; ROMÁN et al., 2004). However, several attempts in trying to identify an isolate or a new variant of CTV associated with CSD have failed (TARGON et al., 2004; GOMES et al., 2008; MACCHERONI et al., 2005; RIVAS-VALENCIA et al., 2008). In 2005, Maccheroni et al., by doing a shotgun sequencing of cDNAs synthesized from total double-strand RNAs (dsRNAs) obtained from citrus plants showing CSD symptoms, identified a new virus from the family *Tymoviridae*. They suggest that this was likely to be associated with CSD and was named *Citrus sudden death-associated virus* (CSDaV). However, the role of this virus in CSD-affected plants is still not completely clear. The shotgun sequencing approach generated a low number of valid reads (MACCHERONI et al., 2005), which made it difficult to study the frequency of detected viruses, to differentiate virus isolates and to discover novel viruses that might be involved with CSD disease. In addition, conventional approaches for virus detection requires prior knowledge of genome sequences (NOURI et al., 2016), thereby

allowing for identification only of specific known viruses, and thus is not suitable to study the virome within plants (JO et al., 2015; NOURI et al., 2016).

Next-generation high-throughput sequencing (NGS) has been widely and successfully used for improved detection and characterization of known and novel viruses in infected plant hosts (JO et al., 2015; PRABHA; BARANWAL; JAIN, 2013). Deep sequencing of the transcriptome (RNA-seq) and the small RNAs (sRNA) has been shown to be a promising and powerful approach in detecting both RNA and DNA viruses (LI et al., 2012; NOURI et al., 2016). Consequently, this approach can be used to better understand plant diseases, especially when the viral etiology is unknown, as well as to explore plant virus-host interactions (COETZEE et al., 2010; PRABHA; BARANWAL; JAIN, 2013). In order to identify putative viruses associated with citrus plants affected by CSD, and compare the frequency and diversity of viruses between CSD-symptomatic and asymptomatic plants, we have done a high-throughput sequencing analysis of the transcriptomes and sRNAs from CSD-symptomatic and asymptomatic citrus plants, all grown in CSD-affected regions. Our work was able to effectively identify both CSDaV and CTV in multiple virus infections and to differentiate two predominant CTV genotypes. We demonstrated a putative association of CSD-symptomatic plants with a specific CSDaV genotype and with *Citrus endogenous pararetrovirus* (CitPRV) as well. We were also able to identify two putative novel viruses that showed to be more associated with CSD-asymptomatic plants.

3.2. Methods

3.2.1. Plant collection

The citrus plants used in this study were collected at two different time points, in 2007 and 2014, to construct the transcriptome (RNA-seq) and small RNA libraries, respectively. All plant material was sourced from the same citrus regions located in the municipalities of Colombia (northern São Paulo State) and Comendador Gomes (southwestern Minas Gerais State), Brazil, which were severely affected by CSD in 2007, but slightly affected by the same disease in 2014. A total of 15 plants was sampled: six trees showed clear CSD symptoms (i.e., occurrence of yellow stain in the rootstock bark) and nine trees were asymptomatic. Genotypes and symptom information are summarized in Table 3.1. Collected samples were frozen in liquid nitrogen and stored at -80°C prior to analysis.

Table 3.1. Citrus plants used to assess the viral sequences by using high-throughput sequencing. All used canopies were from *Citrus sinensis* variety 'Valencia'. Citrus rootstock varieties, type of plant tissue and type of library constructed are shown.

	Rootstock	Collected tissue	Type of library constructed	Library ID
Asymptomatic plants	Rough lemon	Roots	sRNA	SN453
	Citrandarin Cleopatra x Rubidoux	Leaves	sRNA	SN468
	Rough lemon	Leaves	sRNA	SN470
	Trifoliata Tristeno	Leaves	sRNA	SN473
	Rangpur lime x Swingle A	Leaves	sRNA	SN476
	Sunki mandarin	Leaves	sRNA	SN483
	Sunki x Cleopatra	Leaves	sRNA	SN486
	Swingle	Leaves	sRNA	SN488
	Rangpur lime	Leaves	RNA-seq	C1-960
	Rangpur lime	Roots	RNA-seq	C4-964
	Sunki of China	Leaves	RNA-seq	C1-963
Symptomatic plants	Rough lemon	Roots	sRNA	SN464
	Rough lemon	Leaves	sRNA	SN456
	Rangpur lime x Swingle A	Leaves	sRNA	SN459
	Citrus pennivesiculata	Leaves	sRNA	SN462
	Rangpur lime	Leaves	sRNA	SN479
	Rangpur lime	Leaves	RNA-seq	C1-961
	Rangpur lime	Roots	RNA-seq	C4-965
	Sunki of China	Leaves	RNA-seq	C1-962
Total of plants: 15; total of samples: 19				

3.2.2. RNA extraction and sequencing

To construct the RNA-seq libraries, total RNA was extracted using the RNeasy Plant Mini kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. To construct the small RNA libraries, a high-quality total RNA was obtained by using the CTAB extraction protocol adapted from Bekesiova et al. (1999), where the LiCl was replaced by isopropanol (1 vol) in the precipitation phase. The quantity and quality of total RNAs were estimated using a Nanodrop ND-1000 (Uniscience, Sao Paulo, BR) and 1% agarose gel electrophoresis, respectively. Deep sequencing of both libraries was performed on the Illumina HiSeq 2000 platform by Macrogen, Inc., South Korea (www.macrogen.com).

3.2.3. RNA-seq and Small RNA bioinformatics analysis

Bioinformatics analyses of RNA-seq and sRNA data were performed on the CLC Genomic Workbench software package (CLC Bio-Qiagen, Boston, MA). Trimming of the

sRNA data set was done first by removing the adapter sequences. The low-quality reads (limit of 0.05) and the reads shorter than 15 nucleotides (nt) were discarded from the all libraries. Reads were *de novo* assembled using the CLC Assembly Cell and Trinity 2.1.1 (GRABHERR et al., 2011). Parameters for optimal assembly were selected based on number and length of the contigs (contiguous sequences) obtained. We used word size/k-mer values ranging between 15 and 19 for sRNA and 45 and 65 for RNA-seq. Generated contigs were mapped to the available *Citrus sinensis* genome (BioProject Accession no. PRJNA225998) to remove contigs related to the host and the unmapped contigs were compared against the non-redundant viral protein database available in NCBI using BLASTx (default parameters and expect value of 10^{-5} were used) (COETZEE et al., 2010). Potential viral sequences were checked one by one to confirm the Blast results and all contigs were classified according to the size and sequence with the highest bit score. Contigs that shared high identity with the same virus species were compared against nucleotide database available in NCBI using the BLASTn algorithm to identify the respective potential virus isolates. Based on the largest assembled contigs, predominant viral sequences were screened and selected as candidates for validation. Genome coverages were estimated by mapping the reads against the consensus sequences and viral contigs of the predominant viruses obtained in this study. Open reading frames (ORFs) were predicted using the ORF finder function of the SnapGene software (http://www.snapgene.com/). Re-assembly on the candidate viral sequences was also done separately with reads from the asymptomatic and symptomatic plants for comparative analysis.

3.2.4. Validation of candidate viruses

To confirm the presence of the viral sequences identified in the RNA-seq and sRNA libraries, primers designed based on *de novo*-assembled contigs that showed similarities to viral sequences were used for RT-PCR assays. The sequences of all designed primers are shown in Table S9.1. RNAs extracted from selected citrus plants were used as templates and PCR products were analyzed on 0.8 % agarose gel and sequenced by Sanger sequencing.

3.2.5. Phylogenetic analysis

Amino acid sequences of the RNA-dependent RNA polymerase (RdRp) and Helicase (He) protein, in the case of viral sequences which did not have a conserved domain for RdRP,

from each candidate virus were used to compare phylogenetic relationships with other members of the respective viral family, which showed the highest bit score in the BLAST searches. Multiple alignments of amino acid sequences were made by using Clustal X program with the default settings (LARKIN et al., 2007). Phylogenetic trees were constructed using the neighbor-joining (NJ) method in MEGA6 (TAMURA et al., 2013) with 1,000 bootstraps. GenBank accession numbers of the reference sequences used in the phylogenetic analysis are shown in Table S9.2.

3.3. Results

3.3.1. General analysis of the RNA-seq and Small RNA libraries

From the RNA-seq data, approximately 30 to 37.8 million paired-end reads of 100 bp in length were obtained from each library after removing the low-quality reads, yielding assembled viral contigs that varied between 100 and 6,109 nt in length (Tables 3.2 and 3.3). Although the RNA-seq analysis showed that the majority of reads were derived from CTV and CSDaV, these libraries have suggested the presence of viral sequences from other several distinct taxa as well. Considering all libraries, we were able to find viral sequences similar to representatives of 20 distinct virus families (Table 3.3). A great number of the identified viral sequences showed less than 50% amino acid identity to their homologs in the viral database, suggesting that they might represent novel viral sequences. Table S9.3 provides a list of all viruses from the viral database that showed hits in BLASTx analysis with the assembled sequences obtained from this work. High-throughput sequencing of the sRNA libraries generated approximately 6.8 to 14.2 million usable reads per library after trimming, with a length ranging of 16 to 30 nt. The majority of the assembled viral contigs from these libraries (>90%) was short in length (≤ 200 bp) (Table 3.2) and the BLASTx searches showed the presence of only CTV and CSDaV as viral sequences in different citrus plants accessed in this study.

Table 3.2. Read and contig count information of each RNA-seq and sRNA library. The number of confirmed assembled viral contigs was organized according to their size.

				**No. of confirmed viral contigs		
Library ID	**No. of reads after trimming**	**No. of exogenous reads**	***No. of putative viral contigs**	**≤ 200 bp**	**Between 201 to 999 bp**	**≥ 1000 bp**
C1-960	37,811,400	3,826,317	40,187	76	41	7
C1-961	37,380,448	3,613,441	29,483	74	37	9
C1-962	36,452,005	3,528,784	30,834	72	32	6
C1-963	29,942,484	3,027,793	38,028	54	25	2
C4-964	35,511,705	3,741,954	33,719	10	4	4
C4-965	35,386,080	3,617,222	39,012	24	13	0
SN453	8,091,654	756,302	776	323	19	0
SN456	8,949,837	1,117,451	395	263	5	0
SN459	9,899,316	1,065,849	571	334	6	0
SN462	6,233,982	836,217	267	198	5	0
SN464	9,042,291	782,698	821	178	29	0
SN468	8,843,660	897,286	626	384	11	0
SN470	6,866,756	1,041,001	410	329	4	0
SN473	11,892,105	1,597,924	533	311	41	0
SN476	9,076,165	917,605	631	408	14	0
SN479	11,649,840	1,232,264	733	163	24	0
SN483	8,765,391	949,445	471	326	13	0
SN486	11,001,152	1,127,439	555	242	13	0
SN488	14,231,970	1,446,481	783	507	20	0
Total	337,028,241	35,123,473	218,835	4,276	356	28

*Number of contigs that showed BLASTx hits to any viral proteins in NCBI database.

** Number of viral contigs individually confirmed by BLASTx algorithm in NCBI database.

Table 3.3. Contig counts and contig length for each viral specie identified in the BLASTx analysis of the total data set.

Closely related viruses	**Family**	**No. of contigs**	**Contigs length (nt)**	**From RNA-seq libraries**	**From sRNA libraries**	**Maximum % amino acid identity**
Citrus tristeza virus	*Closteroviridae*	4.556	50-3,180	560	3,996	92
Citrus sudden death-associated virus	*Tymoviridae*	61	50-6,109	20	41	98
Marine RNA virus SF-2	*Marnaviridae*	1	1400	1	0	22
Rice stripe necrosis virus	*Benyviridae*	1	250	1	0	39
Rhizoctonia solani negative-stranded virus 4	*Bunyaviridae*	1	126	1	0	56
Norovirus cat	*Caliciviridae*	1	144	1	0	48

Dioscorea bacilliform AL virus	*Caulimoviridae*	1	280	1	0	43
Po-Circo-like virus 51	*Circoviridae*	1	305	1	0	43
Aphid lethal paralysis virus	*Dicistroviridae*	6	115-343	6	0	97
Nakiwogo virus	*Flaviviridae*	1	2512	1	0	27
Sclerotinia sclerotiorum deltaflexivirus 1	*Flexiviridae*	5	101-329	5	0	62
Soybean leaf-associated mycoflexivirus 1	*Gammaflexiviridae*	1	196	1	0	37
Deformed wing virus	*Iflaviridae*	1	173	1	0	52
Nilaparvata lugens honeydew virus-3	*Iflaviridae*	1	118	1	0	47
Raphanus sativus cryptic virus 1	*Partitiviridae*	1	183	1	0	41
Passerivirus A1	*Picornaviridae*	1	134	1	0	46
Chilli veinal mottle virus	*Potyviridae*	1	109	1	0	48
Citrus endogenous pararetrovirus	*Caulimoviridae*	8	339-3,339	8	0	72
Lettuce necrotic leaf curl virus	*Secoviridae*	1	141	1	0	42
Rice tungro spherical virus	*Secoviridae*	1	189	1	0	43
Fusarium graminearum deltaflexivirus 1	Putative *Deltaflexiviridae*	2	153-262	2	0	71
Boutonnet virus	unclassified viruses	2	423-434	2	0	36
Bufivirus UC1	unclassified viruses	1	203	1	0	43
Fisavirus 1	unclassified viruses	1	101	1	0	56
Twyford virus	unclassified viruses	1	186	1	0	44
Beet virus Q	*Virgaviridae*	1	4,097	1	0	33
Chinese wheat mosaic virus	*Virgaviridae*	1	2,626	1	0	28

3.3.2. Contigs derived from Citrus tristeza virus

Assembled contigs derived from CTV were found in all RNA-seq and sRNA libraries constructed in this study. CTV was represented by 560 contigs from the RNA-seq libraries that varied between 100 and 3,180 nt in length, and by 3,996 contigs from the sRNA libraries, with a length ranging of 50 to 539 nt, representing the largest count for any other virus detected in the citrus plants accessed here. Based on the BLASTx and BLASTn searches and number and size of the contigs, we identified predominant assembled sequences that showed high identity (>95%) to three different CTV isolates previously identified as A18 (GenBank accession No. JQ798289), SG29 (GenBank accession No. KC748392) and Taiwan-Pum/SP/T1 (GenBank accession No. JX266712). The complete genomes of these three CTV isolates were downloaded from the NCBI database and used simultaneously as reference

sequences in further re-assembly studies using CLC mapping tool to determine the dominant genotype in these plants based on the read count. The mapping of reads from the RNA-seq and sRNA libraries along the corresponding CTV genomes showed a total of 19,121 and 4,492,130 reads aligned to the reference sequences, respectively (Table 3.4). Compared to the SG29 and Taiwan-Pum/SP/T1 CTV isolates, we noticed a lower distribution of reads on the A18 CTV isolate genome with higher read counts observed in areas with high sequence identity among the three different CTV isolates (Figure 3.1). Therefore, only the SG29 and Taiwan-Pum/SP/T1 CTV isolates were considered in the further analyses. The read counts mapped on each CTV isolate were used to calculate the average coverage of the respective genomes (Table 3.4). Although we were able to assemble only short contigs in length (<540 nt) from the sRNA libraries, a greater average coverage of the CTV genomes was obtained using reads from these libraries compared to the RNA-seq libraries (Table 3.4). CTV has a positive-sense single strand genomic RNA of about 19.3 kb in length, which encodes 12 ORFs (AMBRÓS et al., 2013). The large ORF1 (1a and 1b) encodes replicase-related proteins that are translated from the genomic RNA, and the 10 other ORFs, expressed through subgenomic RNAs, encode proteins p33, p6, p65, p61, p27, p25, p18, p13, p20, and p23, which have different roles in the virus assembly, virus movement and virus infection (HARPER, 2013). The density of CTV reads from both RNA-seq and sRNA libraries along the two assumed most predominant CTV genomes (SG29 and Taiwan-Pum/SP/T1) revealed an asymmetric distribution, with a preferential distribution at the 3' terminal region (Figure 3.1). CTV reads from RNA-seq libraries gradually increased from the p25 gene toward p20, where we were able to detect a hotspot, and then declined at the p23 gene and 3'UTR. Although CTV reads from sRNA libraries also showed to be more abundant at the 3' terminal region, where hotspots were detected along the p13 and p20 for both CTV isolates, the frequency and distribution of sRNAs over the both references was not identical. For SG29 CTV isolate, hotspots were found at the p61, p33 genes and at the 3' end of the replicase, whereas for Taiwan-Pum/SP/T1 isolate, we detected hotspots at the p33 gene and at the beginning of the replicase polyprotein. The CTV consensus sequences were reconstructed under names CTV_SPBR_01 and CTV_SPBR_02 using Taiwan-Pum/SP/T1 and SG29 as reference guide genomes, respectively. The nucleotide sequences of the CTV isolates from this study were deposited in the GenBank database under accession numbers KY110737 and KY110738.

Table 3.4. Comparison of the re-assembly data for the two dominant virus species identified in this study (CTV and CSDaV). Read counts from the simultaneous re-assembly analysis are shown for the three assumed predominant CTV isolates and for the two CSDaV isolate.

Virus	Reference Isolate	sRNA simultaneous re-assembly			RNA-seq simultaneous re-assembly		
		Read count	Percentage read count	Average coverage	Read count	Percentage read count	Average coverage
CTV	A18	711,217	15.8%	≈740x	2,450	12.8%	≈13x
	TawainPum	1,800,699	40.1%	≈1,870x	6,789	35.5%	≈35x
	SG29	1,980,214	44.1%	≈2,060x	9,882	51.7%	≈50x
	Total	4,492,130	100%	-	19.121	100%	-
CSDaV	AY884005	3,944	69.6%	≈12x	59,916	73.3%	≈810x
	DQ185573	1,723	30.4%	≈5x	21,784	26.7%	≈295x
	Total	5,667	100%	-	81,700	100%	-

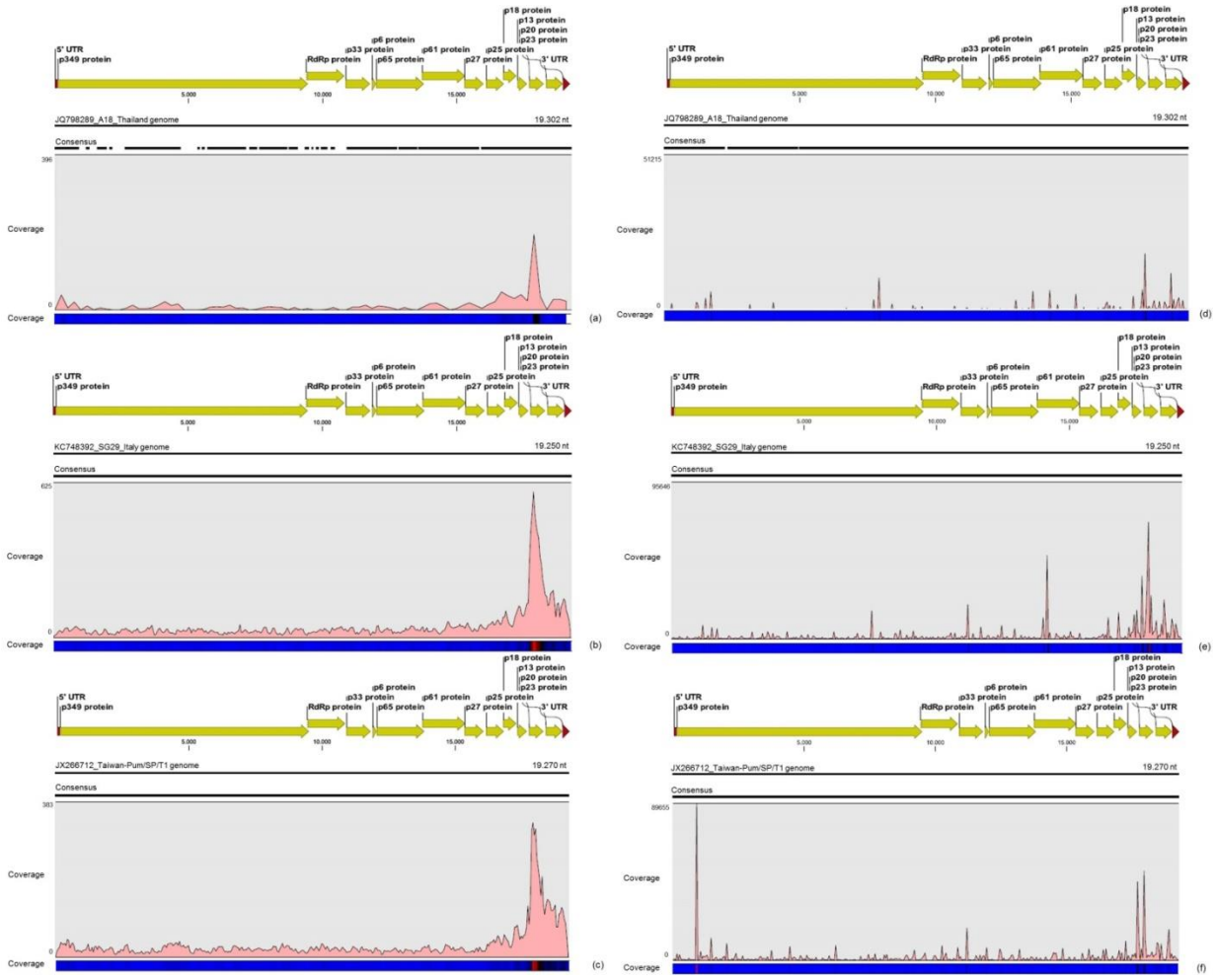

Figure 3.1. Profile distribution of reads from the RNA-seq (a, b and c) and sRNA (c, d and e) libraries along the three different *Citrus tristeza virus* isolates: A18 (a and d), SG29 (b and e) and TawainPum/SP/1 (c and f). Genome organization of the CTV references is shown above the respective graphic. Color scale varies from 0 (light blue color) to 100% (red color) of coverage.

3.3.3. Contigs derived from Citrus sudden death-associated virus

Among all libraries analyzed here, CSDaV was represented by 20 contigs from the RNA-seq libraries that varied between 100 and 6,109 nt in length, and by 41 contigs from sRNA libraries, all of them with less than 400 nt in length. The largest CSDaV assembled contigs, obtained from the RNA-seq libraries, showed different BLASTn results. The CSDaV contigs identified as CSDaV-1 (5,756 nt) and CSDaV-2 (6,109 nt) showed high identity (>97%) to one of the CSDaV isolates (P15) under accession number DQ185573 in the GenBank. On the other hand, the CSDaV contig identified as CSDaV-3 (5,265 nt) showed higher identity (92%) to the CSDaV isolate available in the GenBank under accession number AY884005. These two CSDaV reference isolates available in GenBank show 11% nucleotide diversity, and are the only CSDaV isolates fully described. Although we were able to obtain good enough numbers of contigs and the genome coverage via sequencing of transcriptomes, we had some difficulties in sRNA assembly and obtaining full-genome coverage of the CSDaV genome. The majority of the sRNA libraries (10 out of 13) did not show any CSDaV assembled contigs and, for those that showed the presence of the CSDaV, we identified a few number of short contigs. The complete genome of both of the CSDaV reference isolates were downloaded from the NCBI database and used simultaneously in re-assembly analysis. A total of 81,700 reads from the RNA-seq libraries and 5,667 from the sRNA libraries were aligned to these reference sequences (Table 3.4). The read counts mapped on each CSDaV isolate were used to calculate the average coverage of the respective genomes (Table 3.4). Different from the results obtained for the CTV sequences, a greater average coverage of the CSDaV genomes was obtained using reads from the RNA-seq libraries, compared to the sRNA libraries (Table 3.4). CSDaV has a positive-sense single-stranded RNA genome of about 6.8 kb in length, encompassing a large ORF (240 kDa) that encodes a polyprotein involved with the viral replication and virion structure, and a putative small ORF (p16) associated with virus movement (MACCHERONI et al., 2005). The density of reads from the RNA-seq libraries along the genome of both CSDaV isolates showed a preferential reads distribution at the 3' terminal region of the CSDaV polyprotein, where the prevalence of the reads was found over the CP domains (Figure 3.2). Examination of the sRNA profiles revealed a hotspot at the beginning of the CSDaV polyprotein in the 5' terminal region for both CSDaV isolates and also revealed a notable hotspot in a region close to the beginning of the peptidase domain only for AY884005 CSDaV isolate (Figure 3.2). The CSDaV consensus sequences obtained were

extracted and deposited in the GenBank database under names CSDaV_SPBR_01 and CSDaV_SPBR_02 and accession numbers KY110735 and KY110736, respectively.

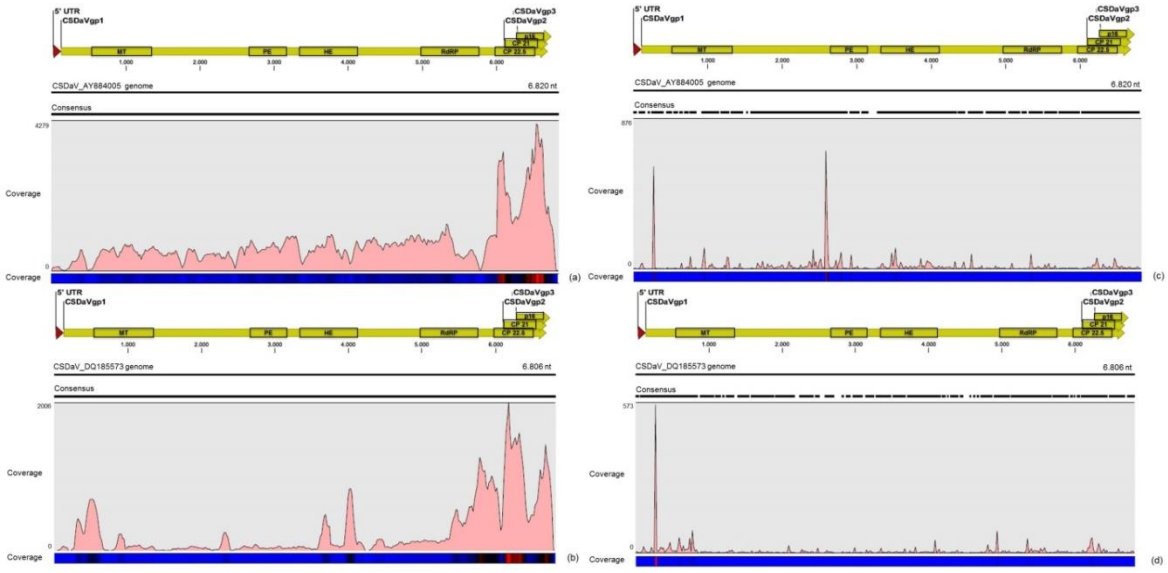

Figure 3.2. Profile distribution of reads from the RNA-seq (a and b) and sRNA (c and d) libraries along the two CSDaV isolates under accession numbers AY884005 (a and c) and DQ185573 (b and d). Genome organization of the CSDaV references is shown above the respective graphic. Color scale varies from 0 (light blue color) to 100% (red color) of coverage.

3.3.4. Description of the distinct viral sequences detected in the RNA-seq libraries

Since the majority of citrus plants grown in a CSD-affected region has shown to be infected by CSDaV and different CTV isolates, we were expecting to find these viruses in both sRNA and RNA-seq libraries constructed in this study, as we showed above. However, in the RNA-seq libraries we were also able to find viral sequences that showed some amino acid identity to several distinct viral families (Table 3.3). A few of the short assembled sequences (< 300 nt) were identified sharing between 36 and 71% amino acid identity to representatives of the following families: *Benyviridae, Bunyaviridae, Caliciviridae, Caulimoviridae, Gammaflexiviridae, Iflaviridae, Partitiviridae, Picornaviridae, Potyviridae, Secoviridae, Deltaflexiviridae* and also to unclassified viruses (Tables 3.3 and S3.3). But because of the low number and small size of these viral sequences, they were excluded from further analysis. The other assembled viral contigs (> 300 nt) showed to share between 27 and 100% amino acid identity to representatives of the families *Caulimoviridae, Dicistroviridae, Virgaviridae, Flexividae, Circoviridae, Flaviviridae* and also to unclassified viruses. A few number of short contigs (< 400 nt) showed high amino acid identity to *Aphid lethal paralysis*

virus (ALPV - Accession No. NC_004365) and *Sclerotinia sclerotiorum deltaflexivirus 1* (SsDFV1 - Accession No. KT581451), viruses from the families *Dicistroviridae* and *Flexiviridae*, respectively. But the re-assembly analyses using these virus genomes as references showed low average coverages in both RNA-seq and sRNA libraries, suggesting a low expression level of these viruses in the plants studied here (Table 3.5). Eight assembled sequences with a length ranging between 339 and 3,339 nt showed 67 to 95% amino acid identity to *Citrus endogenous pararetrovirus* (CitPRV), a virus from the family *Caulimoviridae*. The genome sequence of CitPRV (Accession No. NC_023153) was downloaded from GenBank and used as reference sequence in the re-assembly analysis, which resulted in a good average coverage either for RNA-seq (≈29x) and sRNA (≈68x) libraries. CitPRV is an endogenous pararetrovirus from the family *Caulimoviridae* that was first reported in roots of Carrizo rootstocks (a hybrid of a Washington navel sweet orange and *Poncirus trifoliate*) in a commercial orange grove in Florida, US (ROY et al., 2014). CitPRV seems to have a double-stranded (ds) DNA genome between 6,663 to 6,996 nt in length and encompassing two ORFs. ORF1 encodes a polyprotein that shows signature domains for the movement protein (MP), zinc finger (ZnF), reverse transcriptase (RT) and RNAse_H, and ORF2 seems to encode for a hypothetical protein that does not show any matches with the known viral protein available in the GenBank (ROY et al., 2014). Comparative re-assembly analysis between mapped reads from the RNa-seq and sRNA libraries on the CitPRV genome has shown different profiles of read distribution. Reads from the RNA-seq libraries showed an asymmetric distribution with accumulation of reads over the polyprotein, where the hotspot was found among the RT and RNase_H regions. Reads from the sRNA libraries showed a better coverage along the full CitPRV genome with several hotspots, which the highest one was found in the ZnF region (Figures 3.3a and 3.3c). The CitPRV consensus sequence was obtained and deposited in the GenBank database under name CitPRV_SPBR_01 (GenBank accession number not available yet). BLASTx results for the other assembled viral sequences showed a low percentage (22-60%) of amino acid identity to the known virus proteins available in the GenBank, suggesting that these assembled contigs might represent novel viral sequences. Six assembled contigs identified as CtgMarna-1 (1,400 nt and 22% amino acid identity to *Marine RNA virus SF-2, Marnaviridae*), CtgCirco-1 (305 nt and 43% amino acid identity to *Po-Circo-like virus 51, Circoviridae*), CtgFlavi-1 (2,512 nt and 27% amino acid identity to *Nakiwogo virus, Flaviviridae*), CtgUnclass-1 (423 nt and 36% amino acid identity to *Boutonnet virus*, unclassified ssRNA virus), CtgVirga-1 (4,097 nt and 33% amino acid

identity to *Beet virus Q, Virgaviridae*) and CtgVirga-2 (2,626 nt and 28% amino acid identity to *Chinese wheat mosaic virus, Virgaviridae*) were used as reference sequences in further mapping analysis. The reads from RNA-seq and sRNA libraries were mapped on these six assembled contigs to calculate the read count and the average coverage of the respective contigs. Overall, a low number of reads (less than 400) from the sRNA libraries were found mapping on the viral sequences used as references, which hence resulted in a low average coverage as well (<1x) (Table 3.5). On the other hand, mapping analysis using reads from the RNA-seq libraries resulted in a better coverage along the viral sequences used as references. Viral contigs identified as CtgFlavi-1 had the highest average coverage (≈114x), represented by 3,144 reads, followed by CtgVirga-2 (≈62x) with 1,723 reads mapped along its contig and CtgVirga-1 (≈52x) with 2,297 reads. The other viral sequences showed an average coverage less than 7x (Table 3.5).

Table 3.5. Comparison of re-assembly data among different viral sequences identified in this study. Read count and average coverage from the simultaneous re-assembly analysis are shown for each viral sequence.

Reference viral/contig sequence	**sRNA simultaneous re-assembly**		**RNA-seq simultaneous re-assembly**	
	Read count	Average coverage	Read count	Average coverage
ALPV	387	0.52x	113	1.1x
CitPRV	21,693	68.22x	2,196	28.88x
SsDFV1	227	0.49x	3	0.02x
CtgCirco-1	12	0.53x	25	6.57x
CtgFlavi-1	89	0.53x	3,144	113.95x
CtgMarna-1	68	0.64x	83	5.82x
CtgUnclass-1	103	3.37x	189	41.4x
CtgVirga-1	163	0.59x	2,297	51.4x
CtgVirga-2	105	0.57x	1,723	61.8x

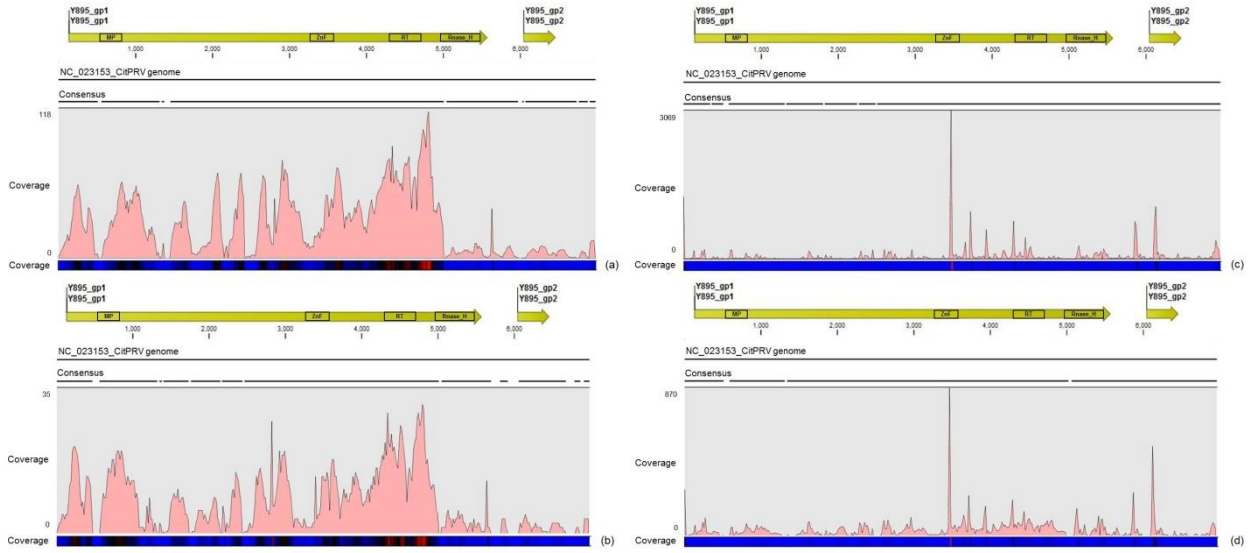

Figure 3.3. Profile distribution of total reads from the RNA-seq (a) and sRNA (c) libraries; and reads from combined asymptomatic (b) and symptomatic (d) libraries along the CitPRV genome. Genome organization of the CitPRV reference is shown above the graphics. Color scale varies from 0 (light blue color) to 100% (red color) of coverage.

3.3.5. Validation of viral sequences by RT-PCR and Sanger sequencing

RNA samples extracted from four different citrus plants (two CSD-symptomatic and two asymptomatic) were used to confirm the presence of the viruses detected in the sRNA and RNA-seq libraries by RT-PCR and Sanger sequencing. Based on the assembled viral contigs that we obtained, specific primers were designed (Table S9.1) to differentiate the two dominant CTV genotypes, to detect the CSDaV, CitPRV, ALPV and SsDFV1 viruses and to confirm the presence of the viral contigs identified as CtgMarna-1, CtgCirco-1, CtgFlavi-1, CtgUnclass-1, CtgVirga-1 and CtgVirga-2. Positive RT-PCR results were obtained for the CTV_SPBR_01 genotype, CSDaV, CitPRV, CtgVirga-1 and CtgVirga-2 in RNA samples from both symptomatic and asymptomatic plants (Figure 3.4). RT-PCR using primers to detect the CTV_SPBR_02 genotype showed an unclear band (data not shown). The presence of the contig identified as CtgFlavi-1 was confirmed only in asymptomatic plants (Figure 3.4). The RT-PCRs to detect the presence of the ALPV, SsDFV1, CtgMarna-1, CtgCirco-1 and CtgUnclass-1 viral sequences were negative for all tested plants.

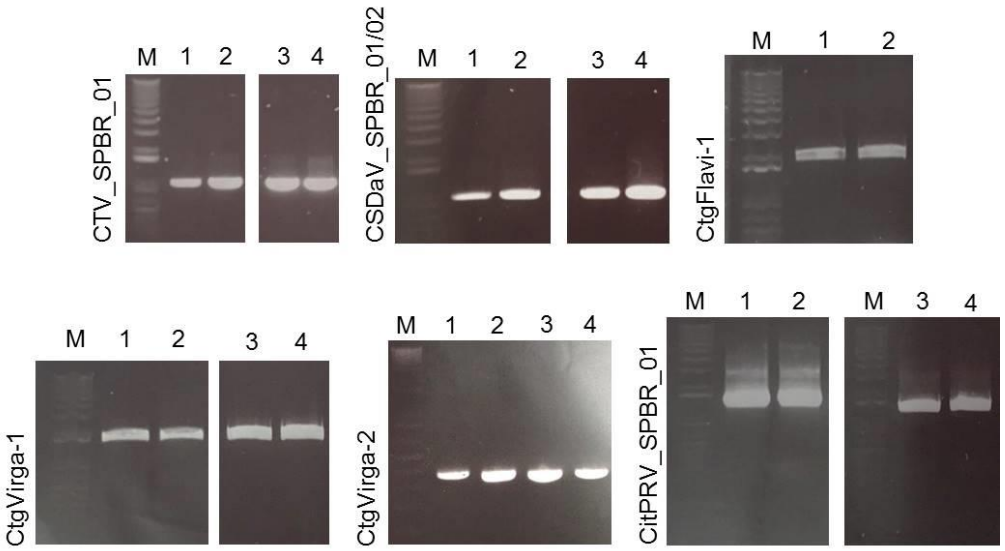

Figure 3.4. Electrophoretic analysis of virus-specific PCR products amplified from total RNAs extracted from citrus collected in a CSD-affected region. The expected of the amplified PCR products are: 1,001 nt (CTV); 974 nt (CSDaV); 1,929 nt (CtgFlavi-1); 1,936 nt (CtgVirga-1); 384 nt (CtgVirga-2) and 1,363 nt (CitPRV). 1 and 2, RNAs from CSD-asymptomatic plants; 3 and 4, RNAs from CSD-symptomatic plants; M, marker 1 kb plus DNA ladder.

3.3.6. Sequence and phylogenetic analysis of the viral sequences related to the CTV, CSDaV and CitPRV, the known viruses detected in this study

Phylogenetic analysis was done for viral sequences that were confirmed in RNA samples by RT-PCR and Sanger sequencing using specific primers. The complete CTV consensus sequences obtained in this study was found to be structurally identical to known CTV isolates, both with 12 ORFs. The complete consensus sequence derived from the CTV_SPBR_01 genotype showed to share 99% identity to TawainPum/SP/1 isolate and was found to be 19,251 nt in length, including 104 nt in the 5'UTR and 258 in the 3' UTR. The CTV_SPBR_02 complete consensus sequence showed 99% sequence identity to the SG29 isolate and was found to be 19,243 nt in length, including 102 nt in the 5'UTR and 273 in the 3' UTR. Phylogenetic analysis based on the RdRP amino acid sequences of the 31 selected previously published CTV genome sequences (Table S9.2) and the two genotypes sequenced in this study, placed CTV_SPBR_01 closer to the isolates from the RB (Resistance Breakdown) lineage, which includes NZRB isolates, an isolate from Hawaii (HA18-9), Taiwan (TawainPum/SP/1) and Puerto Rico (B301); whereas CTV_SPBR_02 was found to cluster within the VT lineage, which includes the severe isolates from Spain (T318A), Asian

(AT-1, CT11A and Nuaga), Israel (VT) and Italy (SG29) (Figure 3.5a). The complete consensus sequences of the two CSDaV genotypes obtained in this study showed similar structure to the previously reported CSDaV genome sequences, both showing a large ORF encoding for a polyprotein and a small ORF putatively representing the p16 ORF. The complete consensus sequence derived from the CSDaV_SPBR_01 genotype showed to share 93% identity to AY884005 CSDaV isolate and was found to be 6,802 nt in length, including 108 nt in the 5'UTR and 127 in the 3' UTR, excluding the poly(A) tail. The CSDaV_SPBR_02 complete consensus sequence showed 97% sequence identity to the DQ185573 CSDaV isolate and was found to be 6,803 nt in length, including 109 nt in the 5'UTR and 127 in the 3' UTR, excluding the poly(A) tail. A phylogenetic tree was constructed based on RdRP amino acid sequences from the two previously reported CSDaV genome sequences, the two CSDaV sequences from this study and from the other four members of the family *Tymoviridae* (Table S9.2). This placed the CSDaV_SPBR_01 and CSDaV_SPBR_02 in different clades, closer to AY884005 and DQ185573 CSDaV isolates, respectively (Figure 3.5b). Reverse transcriptase (RT) amino acid sequence was obtained from the CitPRV consensus sequence obtained in this study and included in a comparative phylogenetic analysis with other members of the family *Caulimoviridae* and Ty3 retrotransposon from *Saccharomyces cerevisiae* to confirm the high phylogenetic relatedness to the respective endogenous pararetrovirus, which was clustered in the same clade with high supporting bootstrap value (90%) (Figure 3.5c).

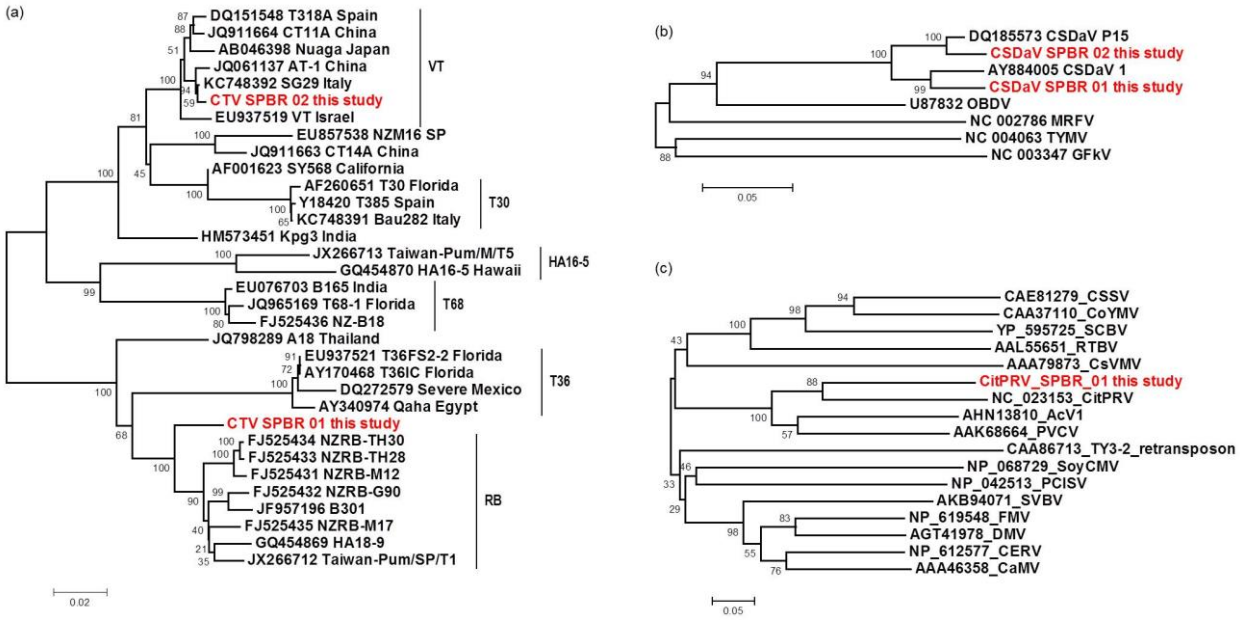

Figure 3.5. Phylogenetic relationships among RdRP (a and b) and reverse transcriptase (c) amino acid sequences from representative isolates of the CTV (a), CSDaV (b) and CitPRV (c), including the respective viral sequences identified in this study. Bootstrap values are shown as percentages and the viral sequences obtained in this study are highlighted in red color.

3.3.7. Phylogenetic analysis and preliminary genome characterization of the unknown viral sequences identified in this study

The viral contigs identified as CtgFlavi-1, CtgVirga-1 and CtgVirga-2 were assembled by the bioinformatics analysis of RNA-seq data and subsequently confirmed by RT-PCR and Sanger sequencing in the RNA samples. The contig CtgFlavi-1 was found to be 2,512 nt in length, including the 75 nt in the 5' UTR and 71 nt in the 3' UTR, excluding the poly(A) tail. Two ORFs were predicted: ORF1 (position 76 to 504) and ORF2 (position 508 to 1926). BLASTx analysis did not detect any putative conserved domains for either ORFs, but ORF2 showed a low sequence identity (27%) to a nonstructural protein NS3 of the *Nakiwogo virus*, an insect-specific flavivirus from the *Flaviviridae* family (BLITVICH; FIRTH, 2015). Viruses from the family *Flaviviridae* are positive-sense single strand RNA viruses and have shown a typical genome organization composed by a single long ORF encoding a polyprotein and flanked by UTRs (BEKAL et al., 2014). However, the CtgFlavi-1 viral sequence showed similar characteristics to the segment 3 of the *Jingmen tick virus* (JMTV), a segmented tick-borne virus, also from the family *Flaviviridae* (QIN et al., 2014). Examples of these similar characteristics are: the partial genome sequences resembling flavivirus nonstructural NS3

protein, the protein size at about 800 amino acids, the presence of UTRs and poly(A) tail, and the presence of two transmembrane regions, predicted by the TMHMM program (version 2.0; www.cbs.dtu.dk/services/TMHMM/) (Figure 3.6a). Phylogenetic analysis based on the nonstructural protein NS3 or helicase protein of the 27 selected genome sequences from the different members of the family *Flaviviridae* (Table S9.2) and the CtgFlavi-1 contig sequenced in this study, placed CtgFlavi-1 in a separate clade between Jingmenviruses and Flaviviruses, closer to the *West Nile virus* and to the *Jingmen tick virus* (Figure 3.7). This phylogenetic distance and the low protein identity obtained from the BLASTx analysis suggest that the viral sequence CtgFlavi-1 might be a genome segment belonged to a novel virus from the *Flaviviridae* family. However, all attempts to find other fragments that could be associated to CtgFlavi-1 sequence failed. The CtgFlavi-1 sequence was deposited in the GenBank as a segment 1 belonged to a putative segmented novel virus tentatively named *Citrus jingmen-like virus* (CJLV) (Accession number KY110739). BLASTx analysis using the CtgVirga-1 and CtgVirga-2 contig sequences as queries showed low sequence identity (28 to 33%) to members from the family *Virgaviridae*, which the highest bit score was found for *Beet Q virus* (genus *Pomovirus*) and *Chinese wheat mosaic virus* (genus *Furovirus*), respectively. Both contigs have shown the presence of putative conserved domains in the BLASTx analysis. The CtgVirga-1 contig seems to be almost completed with 4,097 nt in length, including 58 nt in the 3' UTR and 67 nt in the 5' UTR, and has shown two putative conserved domain encoding for methyltransferase and helicase proteins. The CtgVirga-2 contig showed a putative conserved domain encoding for RdRP protein in the BLASTx analysis and was found to be 2,626 nt in length, including 86 nt in the 3' UTR, excluding the identified poly(A) tail, but the 5' terminal region showed to be not fully completed because the sequence in this region still in ORF. Based on the genome organization of members from the *Virgaviridae* family, a positive-sense single strand RNA plant virus family, we have assumed that both CtgVirga-1 and CtgVirga-2 contigs might be part of the same virus, probably from the RNA1 of a segmented genome. However, attempts to join the two contigs have not shown a conclusive result. Attempts to find sequences related to *Virgaviridae* movement and coat protein in the RNA-seq data have failed. Comparative amino acid sequences analysis based on the helicase and the RdRP amino acid sequences of the selected genomes (34 for the helicase and 32 for the RdRP) from the different members of the family *Virgaviridae* (Table S9.2) was done including the CtgVirga-1 and CtgVirga-2 contigs, respectively. Both helicase and RdRP phylogenetic trees placed the CtgVirga-1 and CtgVirga-

2 contigs in a separated clade, phylogenetically distant to the other genus of the family *Virgaviridae*, suggesting that these contigs might be part of a novel virus that might represent a novel genus within the family *Virgaviridae*. The CtgVirga-1 and CtgVirga-2 sequences were deposited in the GenBank as un-joined fragments belonged to a putative novel virus tentatively named *Citrus virga-like virus* (CVLV) (Accession numbers KY110740 and KY110741). Schematic genome organization and predicted ORFs are shown in Figure 3.6.

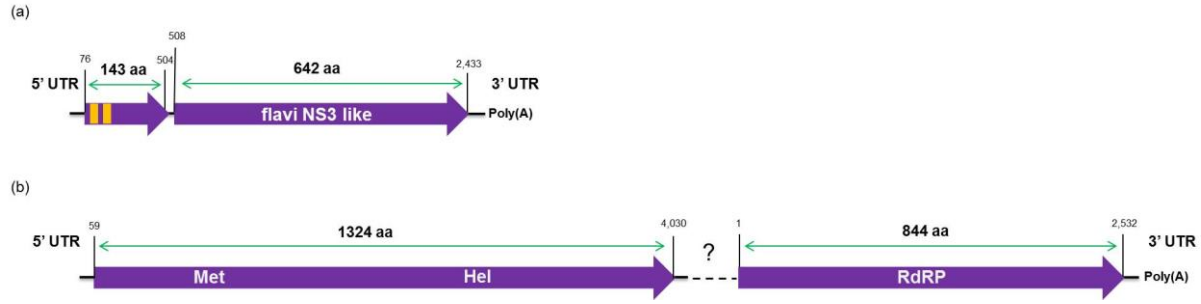

Figure 3.6. Schematic illustration of the predicted partial genome organization of the two putative novel viruses identified in this study. (a) A putative segment 1 of the CJLV genome showing two predicted ORFs (purple arrows) and two predicted transmembrane regions (orange boxes). (b) Partial CVLV predicted genome showing two predicted ORFs (purple arrows). The detected conserved domains and the amino acid length of each ORF are indicated.

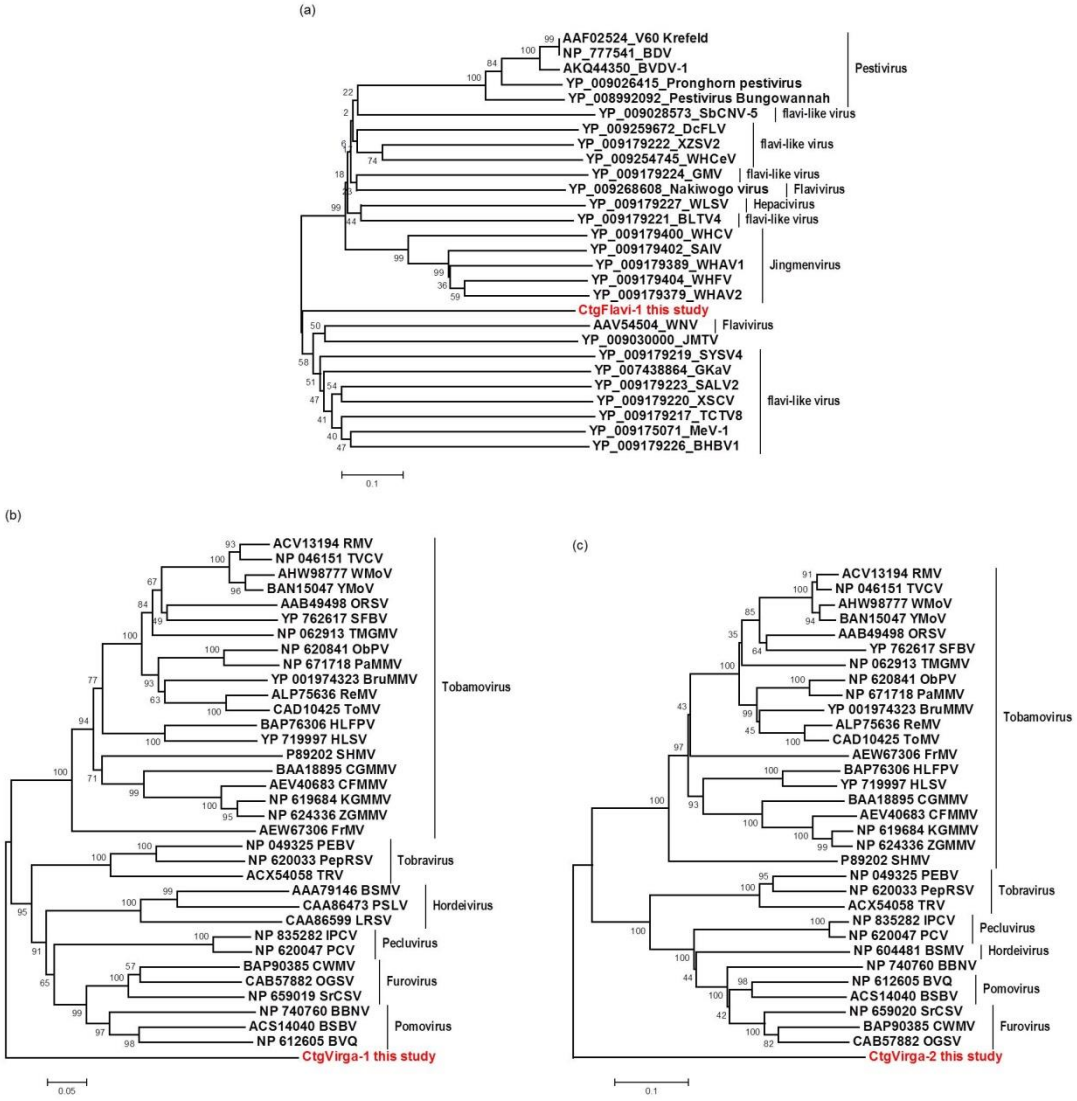

Figure 3.7. Phylogenetic relationships among helicase (a and b) and RdRP (c) amino acid sequences from representative isolates of the families *Flaviviridae* (a) and *Virgaviridae* (b and c), including the respective viral sequences identified in this study. Bootstrap values are shown as percentages and the viral sequences obtained in this study are highlighted in red color.

3.3.8. Comparison of viral sequences derived from CSD-symptomatic and asymptomatic plants

Comparative analysis between CSD-symptomatic and asymptomatic plants was done by merging reads from seven libraries constructed from asymptomatic plants (C1-960, C4-964, C1-963, SN453, SN470, SN476 and SN488) and seven libraries constructed from symptomatic plants (C1-961, C4-965, C1-962, SN464, SN456, SN459, SN479) (Table 3.1), followed by mapping of these two combined libraries to the viral sequences obtained and confirmed in this study. The full consensus sequences identified as CTV_SPBR_01,

CTV_SPBR_02, CSDaV_SPBR_01, CSDaV_SPBR_02 and CitPRV_SPBR_01, as well as the contigs identified as CtgFlavi-1, CtgVirga-1 and CtgVirga-2 were used as references in mapping analysis. The read counts mapped on each viral sequence were used to calculate the average coverage of the respective genomes (Table 3.6). For both CTV consensus sequences used as references, the average coverages were found to be at about 1.2 times higher in libraries constructed from symptomatic plants. Although this difference is not likely to be significant, we did see an asymmetric read distribution along the both CTV consensus sequences when libraries from symptomatic and asymptomatic plants were compared. From the 5' terminal to the p25 region, the read distribution along the CTV_SPBR-01 consensus sequence was similar for both asymptomatic and symptomatic libraries, where hotspots were found over the 5' terminal region of the ORF1a and over the p27 region. In the 3' terminal region, an accumulation of reads from asymptomatic libraries was found in the p13 and p20 region, which p20 showed to have the highest coverage, whereas the hotspots for symptomatic libraries were found to be in the p18, p13 and p20, which the highest coverage was over the p13 region. The mapping on the CTV_SPBR_02 consensus sequence showed that in both asymptomatic and symptomatic libraries, the read distribution was more abundant in the 3' terminal region, however, the hotspots for asymptomatic libraries were found over the p61 and p20 region, whereas notable hotspots were detected over the p13 and p23 for the symptomatic libraries (Figure 3.8).

Table 3.6. Re-assembly data among different viral sequences identified in this study by mapping reads from asymptomatic and symptomatic combined libraries for comparative analysis. Read count and average coverage are shown for each viral sequence.

Reference viral sequence	**Asymptomatic libraries re-assembly**		**Symptomatic libraries re-assembly**	
	Read count	Average coverage	Read count	Average coverage
CTV_SPBR_01	418,902	442.32	525,380	563.19
CTV_SPBR_02	553,150	584.91	633,371	693.78
CSDaV_SPBR_01	3,934	26.49	58,532	767.43
CSDaV_SPBR_02	8,844	109.53	14,350	182.44
CtgFlavi-1	3,182	114.2	28	0.15
CtgVirga-1	2,582	56.14	59	0.2
CtgVirga-2	1,791	62.18	12	0.06
CitPRV_SPBR_01	721	8.35	8,325	41.65x

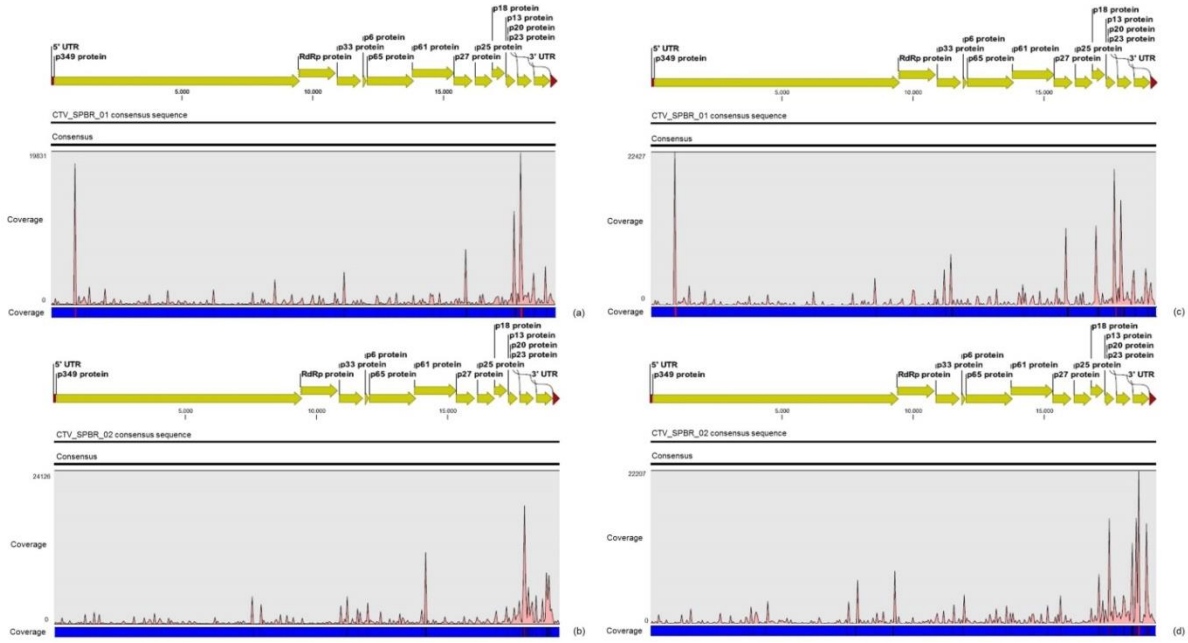

Figure 3.8. Profile distribution of reads from combined asymptomatic (a and b) and symptomatic (c and d) libraries along the two CTV consensus sequence obtained in this study: CTV_SPBR_01 (a and b) and CTV_SPBR_02 (b and d). Genome organization of the CTV references is shown above the respective graphic. Color scale varies from 0 (light blue color) to 100% of coverage (red color).

Different from the results obtained for CTV, the mapping on the CSDaV_SPBR_01 and CSDaV_SPBR_02 revealed great differences on average coverage and read distribution between them, and also between the libraries from symptomatic and asymptomatic plants. The average coverage of the CSDaV_SPBR_01 sequence using reads from the symptomatic libraries showed to be at about 29 times higher than the average coverage estimated with mapped reads from the asymptomatic libraries (Table 3.6). Accumulation of reads from asymptomatic libraries was found in several points along the CSDaV_SPBR_01 sequence, but the coverage even for these regions was low (Figure 3.9). Distribution of the reads from symptomatic libraries showed to be more abundant in the 3' terminal region of the CSDaV sequence, where hotspots were detected over the coat proteins and p16 region. The average coverage of the CSDaV_SPBR-02 consensus sequence was only about 1.6 times higher in symptomatic libraries, compared to the asymptomatic libraries. The mapping of reads from asymptomatic libraries showed a symmetric distribution of the reads along the CSDaV_SPBR_02 sequence until coming to the region encoding the CP and p16 proteins, where we detected the presence of a hotspot. On the other hand, mapped reads from

symptomatic libraries on the CSDaV_SPBR_02 sequence showed an asymmetric distribution with several points of read accumulation around the 5' terminal, helicase, RdRP and CP regions.

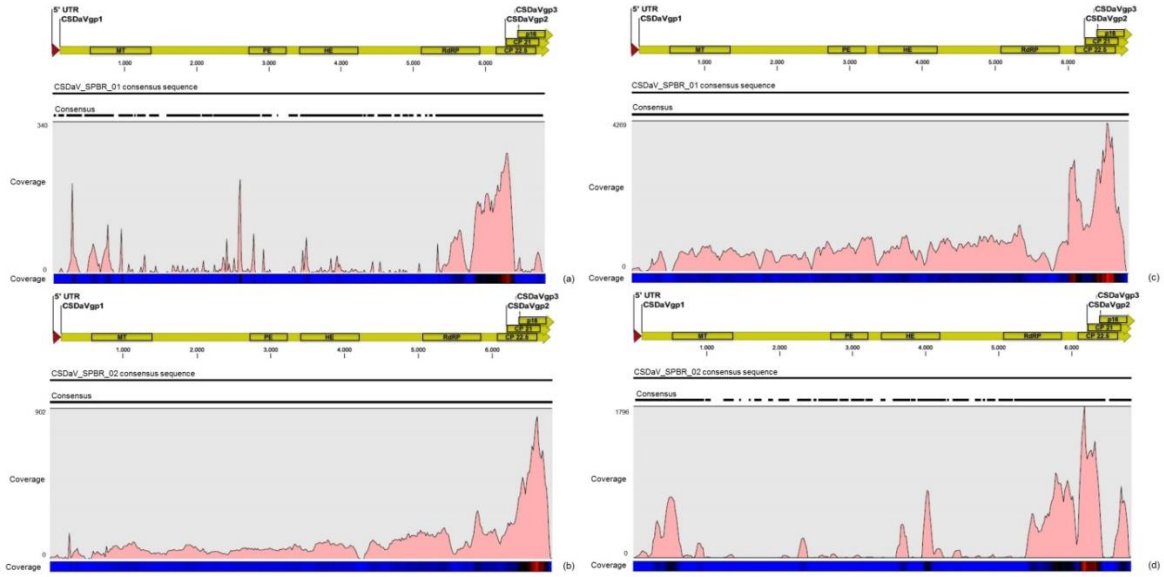

Figure 3.9. Profile distribution of reads from combined asymptomatic (a and b) and symptomatic (c and d) libraries along the two CSDaV consensus sequence obtained in this study: CSDaV_SPBR_01 (a and b) and CSDaV_SPBR_02 (b and d). Genome organization of the CSDaV references is shown above the respective graphic. Color scale varies from 0 (light blue color) to 100% (red color) of coverage.

Comparative analysis between read distribution profile of the asymptomatic libraries and symptomatic libraries on the CitPRV_SPBR_01 consensus sequence showed that symptomatic libraries have an average coverage at about ≈5x higher than in asymptomatic libraries. It has also noticed that symptomatic libraries seem to have higher expression level of small RNAs, whereas asymptomatic libraries showed more mapped reads from the RNA-seq libraries (Figure 3.3). Re-assembly analysis using mapped reads from symptomatic and asymptomatic libraries on the viral contigs CtgFlavi-1, CtgVirga-1 and CtgVirga-2 showed results opposite to what we obtained for CTV, CSDaV and CitPRV reads. Surprisingly, the average coverage of the CtgFlavi-1 contig showed to be at about 760 times higher using reads from asymptomatic libraries, compared to the symptomatic libraries (Table 3.6). Similarly, reads mapped on the contigs CtgVirga-1 and CtgVirga-2 showed to be more abundant in the

asymptomatic libraries, with an average coverage around 280 and 1,040 times higher, respectively, compared with symptomatic libraries (Table 3.6).

3.4. Discussion

In this work, Illumina high throughput sequencing of the transcriptome and small RNAs from citrus plants grown in regions affected by citrus sudden death disease has allowed us to identify and compare viral sequences presenting in these plants. The deep sequencing analyses were sensitive and sufficient to identify the predominant viruses, to obtain information about their genetic diversity, and to demonstrate the presence of putative novel and low-titer viruses. *Citrus tristeza virus* was the most predominant virus identified here, represented by 97.4% of total reads, followed by *Citrus sudden death-associated virus*, which corresponded to 1.94% of the reads, *Citrus endogenous pararetrovirus* with 0.53% of the reads, and other viruses represented by 0.13% of the reads. The presence of the two first mentioned viruses was not a surprising finding because in attempts to discover the causal agent of CSD, both of these viruses were detected and associated with CSD-affected plants. Maccheroni et al. (2005) reported a significant correlation at 99.7% between CSD symptoms and the presence of CSDaV, but the role of this virus in CSD is not yet clear. CTV is an endemic virus in Brazil (STACH-MACHADO et al., 2002), and previous published works have used different approaches to identify an isolate or new variant of CTV associated with CSD, but all attempts have failed so far (TARGON et al., 2004; GOMES et al., 2008; MACCHERONI et al., 2005; RIVAS-VALENCIA et al., 2008). Although previous works have shown that citrus plants affected by CSD are infected by a mixed population of divergent CTV variants (TARGON et al., 2004; GOMES et al., 2008; MACCHERONI et al., 2005; RIVAS-VALENCIA et al., 2008), it is still unknown which specific CTV genotypes are present in those plants. To our knowledge, this is the first high-throughput sequencing-based study of the viral sequences present in citrus plants affected by the CSD disease. Our work reveals mixed viral infections in both CSD-symptomatic and asymptomatic plants, including CTV, CSDaV, CitPRV and two putative novel viruses tentatively named in this study as *Citrus jingmen-like virus* (CJLV) and *Citrus virga-like virus* (CVLV).

For the first time, we were able to identify and obtain the full consensus sequence of the two predominant Brazilian CTV genotypes present in those CSD-affected plants, identified here as CTV_SPBR-01 and CTV_SPBR_02. Phylogenetic analysis clustered

CTV_SPBR_01 within RB-like CTV isolates, whereas CTV_SPBR_02 was clustered within VT-like CTV isolates. Both of the resistance breaking (RB) and VT known CTV strains have been characterized as severe or aggressive strains, which are associated with decline symptoms of citrus trees propagated on sour orange rootstock (*Citrus aurantium* L.) or stem pitting (SP) of the scion regardless of the rootstocks (WU et al., 2014). Although re-assembly analysis comparing CTV mapped reads between libraries from asymptomatic and symptomatic plants did not show significant differences regarding average coverage values, we did notice differences on the read distribution and hotspot regions between these two libraries. Interestingly, mapping reads from asymptomatic libraries for both CTV genotypes identified here has shown a hotspot over the silencing suppressor gene p20, besides other lower hotspots as well. Whereas mapping reads from symptomatic libraries showed an increased read coverage over the host range associated genes; the p13, p18 and p33 when CTV_SPBR_01 consensus sequence was used as reference, and the p13, when CTV_SPBR_02 was used as reference. CTV_SPBR_02 also showed a hotspot over the p23 gene, which is a multifunctional gene and is also associated with silencing suppressor activity (ALBIACH-MARTI, 2013; HARPER, 2013). Based on these results, the association of these two predominant severe-like CTV isolates with CSD-symptomatic plants is not clear, however, the results led us to think about a new question concerning these CTV isolates: Could these isolates be the helpers in mixed virus infections by using their silencing suppressor and host range genes/proteins to facilitate the systemic infection of the other virus(es)? CSDaV could be this other virus and involved with CSD. Interestingly, the CSDaV consensus sequence obtained from libraries constructed from the symptomatic plants (CSDaV_SPBR_01) showed to be phylogenetically distant from the CSDaV consensus sequence extracted from the asymptomatic libraries (CSDaV_SPBR_02), showing at about 13% nucleotide diversity between them. Furthermore, an impressive 29 times higher average coverage was found in mapping reads from symptomatic libraries on the CSDaV_SPBR_01 consensus sequence, compared to mapping reads from asymptomatic libraries. The average coverage of the CSDaV_SPBR_02 genotype using reads from the symptomatic plants was only 1.6 times higher than mapping reads from asymptomatic libraries on the same genotype. These results strongly support an association of CSDaV with CSD symptoms and suggest that there is a specific CSDaV genotype that could be more associated with this disease.

Another interesting result came from the comparative analysis between mapped reads from the asymptomatic and symptomatic libraries on the endogenous CitPRV genome.

Besides the higher average coverage of this virus in symptomatic libraries (at about 5 times), we also have noticed that symptomatic libraries have higher expression level of small RNAs, compared to the asymptomatic libraries. It has been shown that other plant pararetroviruses, such as endogenous *Petunia vein clearing virus* (PVCV) and *Tobacco vein clearing virus* (TVCV), can be in some way induced, culminating to the development of viral symptoms and sRNA accumulation (LOCKHART et al., 2000; NOREEN et al., 2007; ZAVALLO et al., 2015). Although the difference regarding the average coverage of the CitPRV between symptomatic and asymptomatic libraries was less impressive than that we obtained for CSDaV, we cannot ignore this result. As far as we know, this is the first time that CitPRV was identified in citrus plants in Brazil, and it represents the initial step in studying the possible role of CitPRV in CSD symptoms in these plants.

The high throughput sequencing approach also allowed us to identify two novel viruses infecting the plants studied here. These show low amino acid identity to viruses from the families *Flaviviridae* and *Virgaviridae*. Interestingly, although the genomes of both viruses are not completed using our data here, results obtained from the re-assembly analysis on the contigs from the *Citrus jingmen-like virus* and *Citrus virga-like virus* demonstrated an impressive higher average coverage for both viruses in asymptomatic libraries. Besides that, we also have seen a higher diversity of viral sequences in these libraries. Of 27 viral species identified in the BLASTx analysis using assembled contigs obtained in this study as queries, 21 of them were found only in asymptomatic libraries. The lower viral diversity in libraries constructed from symptomatic plants is might be attributed to a strong competition among different viruses within the host for adequate replication conditions. Our results might suggest two things: (i) in the CSD-affected plants, viruses that are associated to developing CSD symptoms (i.e. CTV, CSDaV and/or CitPRV) are the fittest viruses, eliminating or suppressing other viruses from the within-host competition and (ii) in plants not affected by CSD, other viruses (i.e. CJLV and/or CVLV) could play a role in suppressing infections by virus(es) putatively associated in developing CSD symptoms.

In summary, this work has shown that high throughput sequencing was a valid approach to identify and compare viral sequences in citrus plants grown in regions affected by CSD. The correlation of the viruses with the CSD disease indicated a strong association of the CSD-symptomatic plants with a specific CSDaV isolate/genotype and a likely association with CitPRV. We have identified two putative novel viruses that, interestingly, showed to be more associated with the CSD-asymptomatic plants. This study also contributed to describing,

for the first time, the specific predominant CTV isolates/genotypes infecting citrus plants grown in the CSD-affected region.

3.5. Acknowledgements

This work was financially supported by National Institute of Science and Technology of Genomics for Citrus Improvement (INCT Citros), National Council for Scientific and Technological Development (Proc. No. 306230/2013-5 and 2008/57909-2) and Fundação de Apoio à Pesquisa no Estado de São Paulo (Proc. No. 2008/57909-2). EEM was supported, in part, by National Council for Scientific and Technological Development (CNPq) fellowship and also by Doctoral Sandwich Program Abroad (PDSE/CAPES) fellowship. We wish to thank our lab colleagues from the 'Centro de Citricultura Sylvio Moreira' and from UC Davis Department of Plant Pathology for constructive suggestions and discussions.

4. CAPÍTULO II:

Genetic Structure and Molecular Variability Analysis of *Citrus Sudden Death-Associated Virus* Isolates from Infected Plants Grown in Brazil

Manuscrito publicado na revista *Viruses* (doi: 10.3390/v8120330).

Genetic Structure and Molecular Variability Analysis of *Citrus Sudden Death-Associated Virus* Isolates from Infected Plants Grown in Brazil

Emilyn Emy Matsumura[1,2], Helvécio Della Coletta Filho[2], Silvia de Oliveira Dorta[2], Shahideh Nouri[3], Marcos Antonio Machado[2]

[1]Instituto de Biociências de Botucatu, Universidade Estadual Paulista, Botucatu, SP, Brazil; [2]Laboratório de Biotecnologia, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Cordeiropolis, SP, Brazil; [3]Department of Plant Pathology, University of California Davis, Davis, California, United States of America

Abstract: Citrus sudden death-associated virus (CSDaV) is a monopartite positive-sense single-stranded RNA virus that was suggested to be associated with citrus sudden death (CSD) disease in Brazil. Here, we report the first study of the genetic structure and molecular variability among 31 CSDaV isolates collected from both symptomatic and asymptomatic trees in CSD-affected areas. Analyses of partial nucleotide sequences of five domains of the CSDaV genomic RNA, including those encoding for the methyltransferase, the multi-domain region (MDR), the helicase, the RNA-dependent RNA polymerase and the coat protein, showed that the MDR coding region was the most diverse region assessed here and a possible association between this region and virus adaption to differente host or plant tissues is considered. Overall, the nucleotide diversity (π) was low for CSDaV isolates, but the phylogenetic analyses revealed the predominance of two main groups, one of which showed a higher association with CSD-symptomatic plants. Isolates obtained from CSD-symptomatic plants, compared to those obtained from asymptomatic plants, showed higher nucleotide diversity, nonsynonymous and synonymous substitution rates and number of amino acid changes on the coding regions located closer to the 5' end region of the genomic RNA. This work provides new insights into the genetic diversity of the CSDaV, giving support for further epidemiological studies.

Keywords: Citrus sudden death; CSDaV; plant virus; *Marafivirus*; diversity.

4.1. Introduction

Citrus sudden death-associated virus (CSDaV) is a member of the genus *Marafivirus* in the family *Tymoviridae*, and has shown a strong association with citrus sudden death (CSD), an important citrus disease in Brazil (MACCHERONI et al., 2005). CSDaV virions are isometric particles of ≈ 30 nm in diameter, composed of a monopartite, positive-sense, single-stranded RNA genome of approximately 6.8 kb with a high cytosine content (37.4%) and encompassing two ORFs (MACCHERONI et al., 2005; MAHY et al., 2009). A large ORF (ORF1) encodes a 240 kDa polyprotein (p240) which contains conserved signatures of domains involved with replication and virion structure, including the methyltransferase (MT), the papain-like protease (PRO), the helicase (He), the RNA-dependent RNA polymerase (RdRP) domains and two subunits of the coat protein (CP) of 21 and 22 kDa in size, respectively (MACCHERONI et al., 2005). Moreover, a multi-domain region that contains numerous predicted single domains is detected in ORF1 (between the MT and PRO domains), but the function of this region in CSDaV is unknown. The small ORF (ORF 2) at the 3' end region encodes a 16 kDa putative protein (p16) that has shown 42% identity with the N-terminal portion of a putative movement protein (p31) from *Grapevine fleck virus* (GFkV), a member of the genus *Maculavirus* in the family *Tymoviridae* (MACCHERONI et al., 2005).

The first report of CSD was in 1999, affecting sweet oranges (*Citrus sinensis* L. Osb.) grafted on Rangpur lime rootstock (*Citrus limonia* L. Osb.), the main non-irrigated rootstock used in Brazil (MÜLLER et al., 2002). Since then, CSD has caused death or eradication of four million orange trees in Minas Gerais and São Paulo States (GOMES et al., 2008). Recently, CSD-symptoms have been also detected in sweet oranges grafted on other rootstocks (e.g, *Citrus volkameriana*, *Citrus jambiri* and *Citrus pennivisiculata* Lush) (personal communication - data not published). Citrus plants affected by CSD show general decline symptoms characterized by pale green coloration of leaves, different levels of defoliation, death of the root system, and a characteristic development of yellow stain in the phloem of the rootstock, which is the main diagnostic symptom of this disease (MÜLLER et al., 2002; BASSANEZI et al., 2007). However, these affected plants had an incubation period of at least 2 years before symptoms were detected (BASSANEZI et al., 2007; MACCHERONI et al., 2005), which may result in delay of management of the disease. Although the etiology of CSD has not been definitively determined, MACCHERONI et al. (2005) reported a significant correlation at 99.7% between CSD symptoms and the presence

of CSDaV, and suggested that it is probably spread by an aphid vector. The presence of CSDaV as a part of a multiple virus infections or co-infections has been reported in other hosts as well, such as in Pinot Noir grapevine (PANTALEO et al., 2010), in Nectarine (VILLAMOR et al., 2016) and in grapevine Syrah showing decline symptoms (AL RWAHNIH et al., 2009). Such co-infections are also considered for plants showing CSD symptoms (MACCHERONI et al., 2005; GOMES et al., 2008; ROMÁN et al., 2004).

Only two CSDaV isolates have been characterized so far, and their complete genome sequences are available in GenBank (accession No. AY884005 and DQ185573). However, the structure of CSDaV populations has not been studied and the evolutionary forces that may shape this structure are still unknown. To better understand the relationship between CSDaV and CSD, we studied the genetic structure and molecular variability among CSDaV isolates obtained from CSD-affected areas, and compared them with reference isolates by analyzing the partial nucleotide sequences of five coding regions including those for MT, the multi-domain region (called here as MDR), the He, the RdRP and the CP. As a result, we showed that the MDR region was the most diverse region assessed here. We identified the predominance of two main phylogenetic groups, one of which showed a strong association with CSD-symptomatic plants. CSDaV isolates from CSD-symptomatic plants showed higher nucleotide diversity, nonsynonymous and synonymous substitution rates and number of amino acid changes on the coding regions located closer to the 5' end region of the genomic RNA. These results provide relevant information for further epidemiological studies.

4.2. Materials and Methods

4.2.1. Plant collection

The CSDaV population was assessed from different citrus plants: different cultivars of sweet orange grafted on different rootstocks, susceptible and tolerant to CSD. A total of 31 plants was sampled: fifteen trees were asymptomatic and sixteen trees had clear CSD symptoms (i.e., occurrence of yellow stain in the rootstock bark), including a tree grafted on Sunki mandarim of China, which is supposed to be tolerant to CSD, and trees grafted on CSD-susceptible rootstock (Rangpur lime), but intergrafted with tolerant rootstocks (Trifoliate orange and Cleopatra mandarin). Genotypes and symptom information are summarized in Table 4.1. All selected trees were monitored since 2003 in CSD-affected areas located in the municipalities of Colombia (northern Sao Paulo State) and Comendador Gomes

(southwestern Minas Gerais State), Brazil. CSD-symptomatic plants showed the first symptoms in 2006. All citrus plants were approximately five years old at the time of collection in 2007. Collected samples were frozen in liquid nitrogen and stored at -80 °C prior to analysis.

Table 4.1. Citrus plants used to assess the population of CSDaV. Canopy and rootstock genotypes of each plant are shown. Type of plant tissue and number of collected plants are indicated.

	Canopy (*C. sinensis*)	**Rootstock**	**Collected Tissue**	**Number of Plants**
Asymptomatic plants	Natal	Rangpur lime (*C. limonia*)	Leaves	1
	Valencia	Swingle citrumelo (*P. trifoliate* x C. paradisi)	Leaves	3
	Hamlin	Rangpur lime (*C. limonia*)	Leaves	3
	Pera Rio	Goutou (unidentified *Citrus* hybrid)	Leaves	1
	Valencia	Cleopatra mandarin (*C. reshni*)	Leaves	2
	Valencia	Trifoliate orange (*P. trifoliata*)	Leaves	3
	Hamlin	Cleopatra mandarin (*C. reshni*)	Leaves	1
	Hamlin	Cleopatra mandarin (*C. reshni*)	Roots	1
Symptomatic plants	Valencia	Volkameriano lemon (*C. volkameriana*)	Leaves	1
	Natal	Rangpur lime (*C. limonia*)	Leaves	2
	Hamlin	Rangpur lime (*C. limonia*)	Leaves	2
	Hamlin	Volkameriano lemon (*C. volkameriana*)	Leaves	1
	Valencia	Rangpur lime (*C. limonia*) and Trifoliate orange (*P. trifoliata*) as interstock	Leaves	2
	Pera Rio	Rangpur lime (*C. limonia*)	Leaves	3
	Hamlin	Rangpur lime (*C. limonia*) and Cleopatra mandarin (*C. reshni*) as interstock	Leaves	2
	Hamlin	Rangpur lime (*C. limonia*)	Roots	2
	Valencia	Sunki mandarin of China (*C. sunki*)	Leaves	1
Total = 31 plants				

4.2.2. RNA extraction and RT-PCR amplification

Total RNA was extracted from all samples using the RNeasy Plant Mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. A set of primers (Table 4.2) was designed to amplify nucleotide sequences, which corresponded partially to the five domains: the methyltransferase (MT), the multi-domain region (MDR), the helicase (He), the RNA-dependent RNA polymerase (RdRP) and the coat protein (CP) coding regions based on the CSDaV reference genomes (GenBank accession No. AY884005 and DQ185573) (Figure 4.1). cDNAs were synthesized in a 20 µl volume of 1 X Reaction Buffer, containing 0.5 mM

dNTPs mix, 200 U of RevertAid H Minus M-MuLV Reverse Transcriptase (Thermo Scientific, Waltham, MA, USA), and 5 µM of a random hexamer primer. PCR reactions were performed in 25 µl volume, containing 1X High Fidelity PCR Buffer (Invitrogen, Carlsbad, CA, USA), 0.2 mM dNTP mix, 2 mM MgSO4, 0.02 U of Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and 10 mM of each forward and reverse primers. The following PCR conditions were used: 94°C for 2 min; 35 cycles each of 94°C for 15 s, 55°C (for all pair of primers) for 30 s and 68°C for 1 min. The resulted PCR products were separated by electrophoresis in a 1% agarose gel and detected by ethidium bromide staining. Bands were cut from the gel and the PCR products were purified using a QIAquick gel extraction kit (Qiagen, Valencia, CA, USA).

Table 4.2. Primer sequences designed based on five genomic regions of CSDaV genome.

Genomic region	Primer sequences (5'-3')	Annealing nucleotide position
MT	Forward- CGTCAAACTCCCNCTGAC	351-368
	Reverse- GATCANNAGAGAGTGGACTG	1094-1113
MDR	Forward - CTCCCTCTCCATCTGCAAGC	1566-1585
	Reverse - ATANTCNNTGGAGGGGTCA	2375-2393
He	Forward - AGATNTTGGCNCTNGANTC	3305-3323
	Reverse - ANTCNGAGAACATTCNGTTG	4092-4111
RdRP	Forward - CATCAAGAGAANCANGANCC	4636-4355
	Reverse - TGAGACCATAGTGGGAGTGT	5414-5433
CP	Forward - GCCATCTACACCACACTCTC	5857-5876
	Reverse - TTGGANTAGACGGAGTAGGA	6568-6587

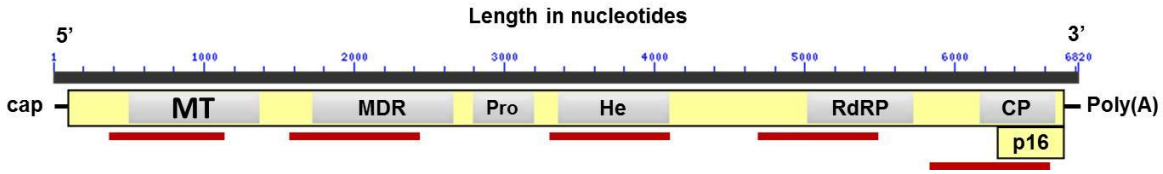

Figure 4.1: Genome organization of CSDaV. The two ORFs are represented by yellow boxes and the conserved domains are represented by grey boxes. Red bars indicate the genomic regions analyzed in this study.

4.2.3. Cloning and sequencing

The purified PCR products were cloned into pGEM-T vector (Promega, Madison, WI) using T4 DNA ligase (Promega, Madison, WI) according to the manufacturer's instructions, followed by transformation into *Escherichia coli* DH5α competent cells (SAMBROOK et al., 1989). Ten recombinant colonies were selected on screening media and confirmed by colony PCRs. Plasmid DNAs were extracted using the PureYield plasmid miniprep system kit (Promega, Madison, WI) following the manufacturer's instructions and were bi-directionally sequenced using an Applied Biosystems 3730 DNA Analyzer.

4.2.4. Nucleotide sequence analysis

CSDaV reference sequences, identified as AY884005 (CSDaV) and DQ185573 (CSDaV strain p15) were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/) and included in this analysis as representatives of CSDaV. Multiple nucleotide sequence alignments for each genomic region were obtained using the CLUSTAL W (LARKIN et al., 2007), and manually edited in the program MEGA 6.06 (TAMURA et al., 2013). Neighbor joining phylogenetic trees were inferred with 1,000 bootstraps in the MEGA 6.06 program and the generated trees were edited using FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). A set of sequences for each genomic region of CSDaV were assessed using DnaSP software version 5.1 (LIBRADO & ROZAS, 2009) to estimate genetic diversity and other population genetic parameters.

Recombination events among CSDaV isolates were examined using phylogenetic analysis and the boot-scan method in the SimPlot program (LOLE et al., 1999). Evidence of recombination was considered when 70% of permuted trees supported a particular grouping of sequences. The window width and the step size were set to 200 and 20 bp, respectively. The degree of selective constraints imposed on different regions of CSDaV genome was estimated with MEGA 6.06 program by analyzing the nonsynonymous and synonymous substitutions ratios ($dN/dS = \omega$) using the Kumar method and bootstrap with 500 replicates (KUMAR et al., 2004). Fixed effects likelihood (FEL), random effects likelihood (REL), and single likelihood ancestor counting (SLAC) tests, all included in the Hyphy package (http://www.datamonkey.org/), were performed to determine the site specific selection pressure for the coding regions. For SLAC and FEL, the cut-off *P*-value was defined as 0.1

and for REL, a Bayes factor of 50 was selected as the cut-off value. Only positive selections determined by at least two methods were accepted (NOURI et al., 2014).

4.3. Results

4.3.1. Genetic diversity of CSDaV population

The presence of CSDaV was confirmed immediately after plants collection in both symptomatic and asymptomatic plants, including trees grafted on the CSD tolerant rootstocks, such as Cleopatra and Sunki mandarins, Swingle citrumelo and *Poncirus trifoliata*. Two step RT-PCRs with specific degenerate primers sets (Table 4.2) generated amplicons with 762, 827, 806, 797, 730 bp in length for the five regions of CSDaV genomic RNA including those encoding for the MT, the MDR, the He, the RdRp and the CP, respectively (data not shown). Figures S3 and S4 and Tables S1 and S2 in the supplemental material show all conserved domains detected from conserved domain search using the CSDaV reference genomes as queries in the NCBI Conserved Domain Database (CDD). A total of 31 CSDaV isolates were obtained (Table 4.3) and the number of sequences for each region is illustrated in Table 4.4. Nucleotide diversities were estimated based on the number of segregating sites (θw) and the average number of nucleotide differences per site between sequences (π). Overall, the genetic diversity for CSDaV isolates evaluated in this study was low ranging from 0.01013 (the CP fragment) to 0.04185 (the He fragment) (Table 4.4) with a mean genetic diversity of 0.026118.

Table 4.3. List of CSDaV sequences obtained in this work. Accession numbers in the GenBank database for the different genomic regions of each isolates are indicated.

Isolate identification*	Viral genomic region	GenBank accession No.
VAVK1D	MT	KX753236
	MDR	KX753259
	CP	KX753328
CR2D	MDR	KX753263
HAVK11D	MT	KX753252
	MDR	KX753260
	CP	KX753326
NACR12D	MT	KX753233
	MDR	KX753261
	RdRP	KX753309

	CP	KX753327
PRCR19D	MT	KX753245
	MDR	KX753262
	RdRP	KX753306
	CP	KX753330
PRGTC20S	MT	KX753254
	MDR	KX753264
	CP	KX753321
VASW23S	MT	KX753248
	MDR	KX753265
	He	KX753296
	RdRP	KX753316
	CP	KX753340
PRCR24D	MT	KX753234
	MDR	KX753266
	He	KX753292
	RdRP	KX753313
	CP	KX753336
HACL26D	MT	KX753243
	MDR	KX753267
	RdRP	KX753307
	CP	KX753323
CR8D[1]	MT	KX753257
	MDR	KX753268
	He	KX753297
	RdRP	KX753318
	CP	KX753342
VASW30S	MT	KX753256
	MDR	KX753269
	He	KX753293
	RdRP	KX753299
	CP	KX753324
VASW31S	MT	KX753242
	MDR	KX753270
	RdRP	KX753298
	CP	KX753319
HACL38S	MT	KX753244
	MDR	KX753271
	RdRP	KX753301
	CP	KX753320
VATR39S	MT	KX753255
	MDR	KX753272
	RdRP	KX753305

	CP	KX753322
HACR42S	MT	KX753241
	MDR	KX753273
	RdRP	KX753302
	CP	KX753333
CLBR43S[2]	MT	KX753251
	MDR	KX753274
	He	KX753294
	RdRP	KX753317
	CP	KX753334
VACL44S	MT	KX753247
	MDR	KX753275
	CP	KX753341
VATR45S	MT	KX753239
	MDR	KX753276
	CP	KX753332
VATR47D	MT	KX753253
	MDR	KX753277
	He	KX753291
	RdRP	KX753314
	CP	KX753343
SKCH5D[3]	MT	KX753237
	MDR	KX753278
	RdRP	KX753310
VATR50S	MT	KX753249
	MDR	KX753279
	RdRP	KX753312
	CP	KX753329
VATR51D	MT	KX753258
	MDR	KX753280
	RdRP	KX753304
	CP	KX753325
HACL52D	MDR	KX753281
	RdRP	KX753315
HACR55S	MDR	KX753282
	CP	KX753331
HACR56D	MT	KX753235
	MDR	KX753283
HACR58D	MT	KX753250
	MDR	KX753284
NACR6D	MT	KX753232
	MDR	KX753285
VACL25S	MT	KX753238

	MDR	KX753286
	He	KX753289
	RdRP	KX753300
	CP	KX753337
NACR22S	MT	KX753246
	MDR	KX753287
	He	KX753295
	RdRP	KX753308
	CP	KX753339
HACR41S	MT	KX753240
	MDR	KX753288
	He	KX753290
	RdRP	KX753303
	CP	KX753338
PRCR16D	MT	KX753231
	RdRP	KX753311
	CP	KX753335

* Isolates were designated based on the citrus genotype from where the CSDaV isolates were obtained. First two letters identify the type of canopy (VA: Valencia; HA: Hamlin; PR: Pera Rio and NA: Natal), followed by the type of rootstock or interstock (VK: Volkameriano lemon; CR: Rangpur lime; GTC: Goutou; SW: Swingle citrumelo; CL: Cleopatra mandarin; TR: Trifoliate orange), the sample number and the symptom information (S: asymptomatic plant and D: symptomatic plant). [1] Isolate from Rangpur lime rootstock tissues; [2] Isolate from Cleopatra mandarin rootstock tissues; [3] Isolate from leaves of Valencia grafted on Sunki mandarin of China.

Table 4.4. Population genetic parameters estimated for five coding regions of CSDaV isolates using the DnaSP and MEGA programs.

Genomic regions	**Number of final sequences**	**S**	**η**	**π**	**θw**	**dN**	**dS**	**ω* (dN/dS)**
MT	28	82	84	0.01815	0.0346	0.005 ± 0.002	0.054 ± 0.010	0.093
MDR	30	180	214	0.04091	0.07212	0.023 ± 0.004	0.097 ± 0.012	0.237
He	9	81	83	0.04185	0.05613	0.006 ± 0.002	0.153 ± 0.020	0.039
RdRP	21	70	72	0.01955	0.02895	0.001 ± 0.001	0.068 ± 0.009	0.015
CP	25	26	27	0.01013	0.01897	0.003 ± 0.001	0.026 ± 0.007	0.115

* S: Total number of segregating sites. η: Total number of mutations. π: Nucleotide diversity, average pairwise nucleotide difference per site. θw: Mutation rate estimated from S. dN: The average number of pairwise differences per synonymous site. dS: The average number of pairwise differences per non-synonymous site. *dS and dN were estimated by the Kumar method.

4.3.2. Phylogenetic relationships of CSDaV isolates

The sequences of four representative viruses from the genera *Tymovirus* (*Turnip yellow mosaic vírus* – TYMV, NC_004063), *Maculavirus* (*Grapevine fleck virus* – GFkV, NC_003347) and *Marafivirus* (*Maize rayado fino vírus* – MRFV, NC_002786; and *Oat blue dwarf vírus* – OBDV, NC_001793) of the family *Tymoviridae* were obtained from GenBank and used as outgroups in the phylogenetic analysis of all regions except the MDR segment because this region of CSDaV did not show any homology with any genomic region of four *Tymoviridae* representatives. Because six isolates were detected as possible recombinants based on the topology of phylogenetic trees (see details in the recombination analysis section), the final trees were constructed after removing these recombinants. In general, the topology of the MT (Figure 4.2a), the MDR (Figure 4.2b), the He (Figure 4.2c) and the RdRP (Figure 4.2d) trees was similar and showed the presence of two main groups of CSDaV isolates assessed in this study with high supporting bootstrap values equal or higher than 83%. The intra-group diversity was best illustrated in the MDR tree (Figure 4.2b). The topology of the CP tree was different, in which all CSDaV isolates formed a single un-resolved polytomy clade with a supporting bootstrap value of 99% (Figure 4.2e). For all phylogenetic trees, with the exception of the CP, the main groups were clustered separately from the two CSDaV reference sequences. Divergence between CSDaV reference sequences (AY884005 and DQ185573 isolates) was observed in the MDR, He and RdRP trees (Figures 4.2b, 4.2c and 4.2d).

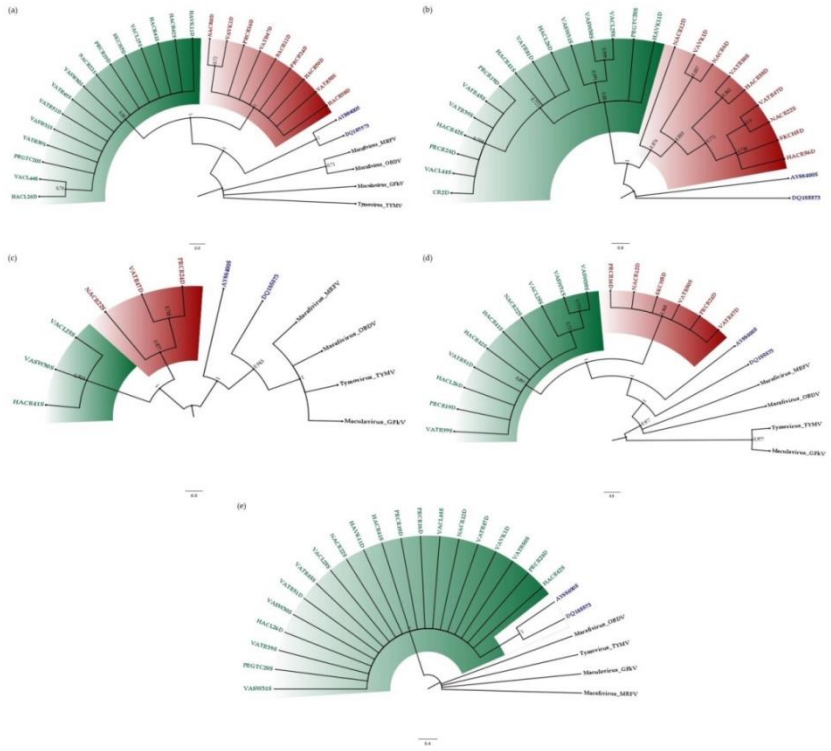

Figure 4.2. Bootstrap majority rule (70%) consensus trees reconstructed by the neighbor joining method for five genomic regions of CSDaV isolates including field collected and reference sequences. Bootstrap support values (1,000 iterations) of main branches are indicated. (a): MT segment; (b): MDR segment; (c): He segment; (d): RdRP segment; (e): CP segment. CSDaV groups are highlighted by different colors: Group I = green; Group II = red. The CSDaV reference isolates are represented in blue. The outgroups are represented in black. Isolates from asymptomatic plants are identified by letter S at the end of their identification names. Isolates from symptomatic plants are identified by letter D at the end of their identification names.

4.3.3. Comparison of genetic diversity between isolates from asymptomatic and symptomatic plants

Based on the MT, the MDR, the He and the RdRP phylogenetic trees, group I of the CSDaV isolates was formed by the majority of isolates from asymptomatic plants, whereas group II contained the majority of isolates from symptomatic plants (Figure 4.2 and Table 4.5). To further strengthen these results, CSDaV consensus sequences were obtained from transcriptome sequencing, conducted for both symptomatic and asymptomatic plants by using Illumina next generation sequencing (NGS) technology (unpublished data). The coding regions studied here were accessed in these consensus sequences and included in the phylogenetic analysis. Based on the MT, the MDR, the He and the RdRP, consensus sequence obtained from asymptomatic library clustered close to the reference isolates, whereas

consensus sequence obtained from symptomatic library grouped in the group II (Figure S1), which strongly support the results presented above.

Compared to isolates from asymptomatic plants, the nucleotide diversities estimated only for isolates obtained from symptomatic plants were higher at about 2.2, 1.5, 1.1, 1.1 and 0.9 times for the MT, MDR, He, RdRP and CP regions, respectively (Table 4.6). The dN/dS ratio values were higher for the MDR region for isolates from both symptomatic and asymptomatic plants. However, this estimated value for the isolates from symptomatic plants was 1.4 times higher than the ratio estimated for the isolates from asymptomatic plants (Table 4.6). The deduced amino acid sequences from each CSDaV genomic region showed silent mutations between isolates from symptomatic and asymptomatic plants (Table 4.7).

Table 4.5. Number of CSDaV sequences from symptomatic and asymptomatic plants between the two main phylogenetic groups assessed in this study.

	Number of isolates from symptomatic plants/number of isolates from asymptomatic plants	
	Group I*	**Group II****
MT	5/10	8/1
MDR	6/9	7/2
He	0/3	2/1
RdRP	3/7	5/1

*Group I is highlighted in green in the phylogenetic trees (Figure 4.2). ** Group II is highlighted in red in the phylogenetic trees (Figure 4.2).

Table 4.6: Comparison of the population genetic parameters estimated for five coding regions of CSDaV isolates from symptomatic (Symp.) and asymptomatic (Asymp.) plants using the DnaSP and MEGA programs.

	Symptoms	Number of Sequences	π	θw	dN	dS	ω
MT	**Symp.**	13	0.01726	0.02117	0.007 ± 0.002	0.044 ± 0.011	0.159091
	Asymp.	11	0.00770	0.01402	0.002 ± 0.001	0.020 ± 0.005	0.100000
MDR	**Symp.**	13	0.03441	0.04143	0.026 ± 0.005	0.057 ± 0.011	0.456140
	Asymp.	11	0.02268	0.02601	0.014 ± 0.004	0.042 ± 0.009	0.333333
He	**Symp.**	2	0.00942	0.00942	0.002 ± 0.002	0.023 ± 0.012	0.086957
	Asymp.	4	0.00879	0.00924	0.003 ± 0.002	0.023 ± 0.008	0.130435
RdRP	**Symp.**	8	0.00856	0.00861	0.001 ± 0.001	0.026 ± 0.008	0.038462
	Asymp.	8	0.00787	0.00918	0.001 ± 0.001	0.024 ± 0.007	0.041667
CP	**Symp.**	9	0.00781	0.00912	0.003 ± 0.001	0.019 ± 0.008	0.157895
	Asymp.	11	0.00862	0.01317	0.004 ± 0.002	0.018 ± 0.006	0.222222

* π: Nucleotide diversity, average pairwise nucleotide difference per site. θw: Mutation rate estimated from the total number of segregating sites. dN: The average number of pairwise differences per synonymous site. dS: The average number of pairwise differences per non-synonymous site. *dS and dN were estimated by the Kumar method.

Table 4.7. Amino acid changes in five CSDaV genomic regions in isolates obtained from symptomatic citrus plants compared to isolates obtained from asymptomatic citrus plants.

Domain	Number of amino acid changes	Total number of amino acid	Position of amino acid changes (Asymp → Symp)
MT	8	202	13 (I → T); 57 (P → Q); 113 (Q → R); 144 (L → V); 152 (S → *); 157 (R → K); 171 (A → V) and 199 (T → I)
MDR	10	209	13 (G → D); 14 (P → R); 22 (L → A); 29 (I → T); 103 (F → S); 109 (F → S); 110 (Q → P; 187 (S → L); 197 (H → R) and 209 (Q → R)
He	4	176	28 (V → A); 62 (L → P); 72 (T → I) and 120 (M → I)
RdRP	4	223	37 (P → L); 136 (A → V); 155 (N→ S) and 200 (L → P)
CP	1	120	55 (Q → R)

4.3.4. Recombination analysis

Based on the phylogenetic trees constructed with all, including the possible recombinants CSDaV isolates (Figure S9.2), six isolates, named CR8D, CLBR43S, VASW23S, HACL38S, HACL52D and HACR55S showed some phylogenetic incompatibilities and evidence of recombination events. The two root isolates (CLBR43S and CR8D) clustered in the same clade according to the RdRP and the CP phylogenetic trees (Figures S9.2d and S9.2e), while they were placed separately based on the MT, the MDR and the He trees (Figures S9.2a, S9.2b and S9.2c). Isolate VASW23S grouped separately in MT and the RdRP phylogenetic trees (Figures S9.2a and S9.2d), but it clustered with the main groups in the MDR, the He and the CP trees (Figures S9.2b, S9.2c and S9.2e). Isolates HACL38S, HACL52D and HACR55S were not compared in all regions of the genome analyzed here because we were not able to obtain PCR products for all segments, and they were excluded from the recombination analysis. We selected nine representative isolates from this study including three suggested recombinants and six isolates representing the two main groups of CSDaV, and two CSDaV reference sequences to concatenate their nucleotide sequences of the MT, the MDR, the He, the RdRP and the CP segments. Concatenated sequences were further analyzed using SimPlot. Both phylogenetic and Bootscan methods included in Simplot identified recombination signals as well as their possible parental sequences when VASW23S, CR8D and CLBR43S isolates were used as queries (Figures 4.3 and 4.4). Phylogenetic analysis of the concatenated sequences detected several recombination hotspots when different isolates were used as queries: positions 600 and 1322 when VASW23S isolate was used as the query (Figure 4.3a), positions 603, 1203 and 2458 when CR8D isolate was used as the query (Figure 4.3b) and positions 170 and 609 when CLBR43S isolate was used as the query (Figure 4.3c). On the other hand, Bootscan analyses demonstrated that the MDR and He segments of VASW23S isolate come from PRCR24D-like and VASW30S-like isolates, respectively (Figure 4.4a). For the CR8D isolate, a recombination event was detected by the Bootscan algorithm in which the MDR and He segments of CR8D were generated from two different origins: AY884005 and DQ185573 reference isolates, respectively (Figure 4.4b). When Bootscan analysis was performed using CLBR43S isolate as the query, we detected four recombination hotspots in which two of them were placed close to positions 170 and 619 (already shown by phylogenetic analysis), and two other hotspots were detected at positions 1271 and 1850 (Figures 4.3c and 4.4c). Furthermore, Bootscan results confirmed that the MDR and He segments of CLBR43S were generated from two CSDaV reference isolates

(Figure 4.3c) and the region from the MT segment was likely driven by recombination events between a DQ185573 reference-like isolate and CR8D-like isolate. The RdRP and the CP segments from CLBR43S isolate showed phylogenetically inconsistent regions with some similarity with the CR8D isolate and the AY884005 reference isolate (Figure 4.4c).

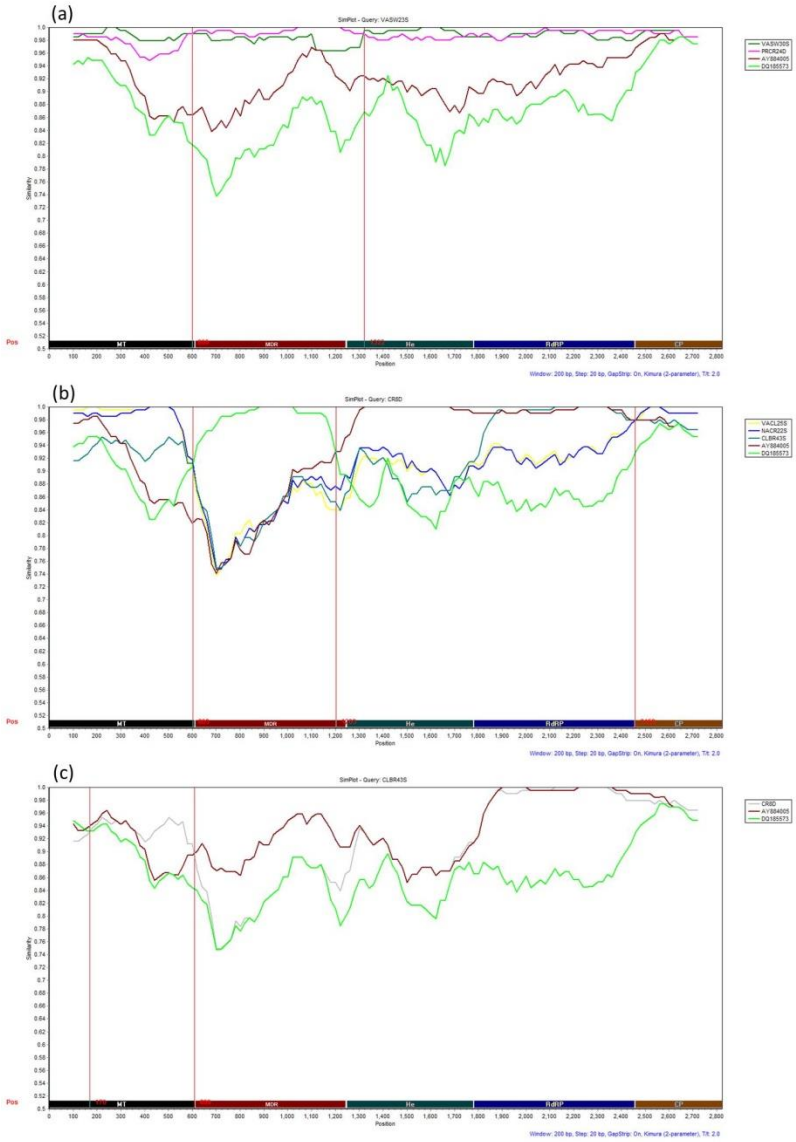

Figure 4.3. Phylogenetic relationship with potential recombinant CSDaV isolates as the query sequences based on concatenated nucleotide sequences of the MT, MDR, He, RdRP and CP genomic regions using Simplot. Three CSDaV isolates, VASW23S (a), CR8D (b) and CLBR43S (c), were used as query sequences and two CSDaV isolates were used as reference sequences. The Y-axis illustrates variation in identity percentage. Analyses were done using a sliding window of 200 bp and a step size of 20 bp. Red vertical dashed line shows the proposed recombination break point. Sequences compared with the query sequence are indicated in the legend.

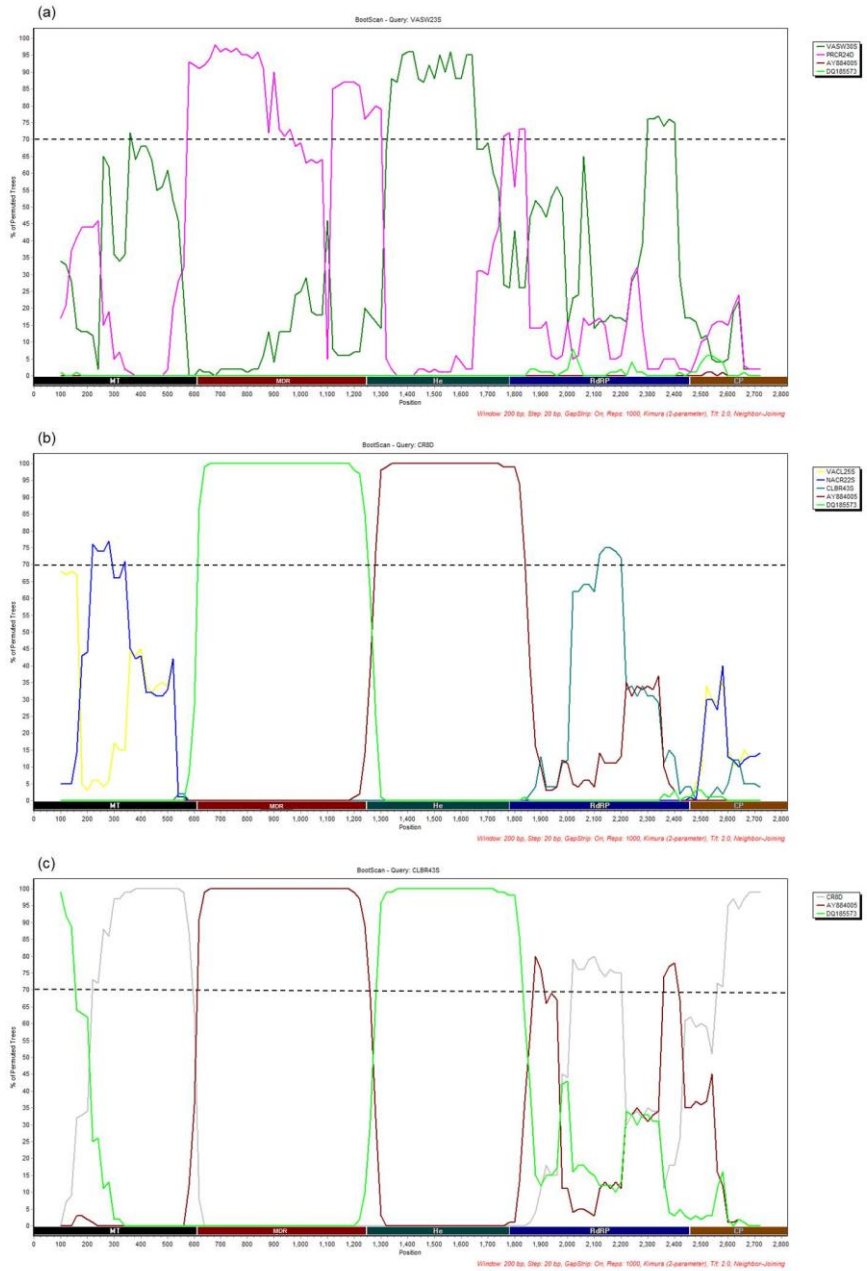

Figure 4.4. Bootscan analyses with potential recombinant CSDaV isolates as the query sequences based on concatenated nucleotide sequences of the MT, MDR, He, RdRP and CP genomic regions using Simplot. Three CSDaV isolates, VASW23S (a), CR8D (b) and CLBR43S (c), were used as query sequences and two CSDaV isolates were used as reference sequences. The Y-axis illustrates variation in percentage of permuted trees in which each selected isolate clustered with the query sequence. Analyses were done using a sliding window of 200 bp and a step size of 20 bp. Black dashed line shows the 70% cutoff level, representing possible recombination. Sequences compared with the query sequence are indicated in the legend.

Only the two root isolates detected as recombinants showed a close phylogenetic relationship to one of the CSDaV reference isolates: CR8D isolate in the MDR, the He and the RdRP trees and CLBR43S isolate in the RdRP tree (Figure S9.2). However, these isolates were phylogenetically distant from the main groups of CSDaV isolates assessed in this study. According to the MT, MDR, He and RdRP phylogenetic trees, isolate CLBR43S, obtained from tissues of Cleopatra mandarin rootstock, was found in a separate clade which was phylogenetically distant from the two main groups of CSDaV isolates (Figure S9.2). Similar results were found for isolate CR8D, obtained from a Rangpur lime rootstock, according to MDR, He and RdRP phylogenetic trees (Figure S9.2). Both of those separations were well-supported.

4.3.5. Selective pressure for different genomic regions of CSDaV

Evidence for positive selection was not found in any region of the genome for the CSDaV isolates included in this study. The mean ω (dN/dS) value for all genomic regions analyzed here was less than 1.0, indicating that all segments were subjected to negative or purifying selection. Among regions, the MT, He, RdRP and CP regions showed low ω ratios, while this ratio was higher for the MDR region (Table 4.4). Moreover, complementary maximum-likelihood methods (SLAC, FEL and REL) detected positively selected sites only for the MDR segment. Site 20 in the MDR segment was considered under positive selection significantly by two methods: FEL (dN – dS = 19.1017 and *P*-value = 0.042) and REL methods (dN – dS = 5.2268 and Bayes Factor = 23853.7) (A → T).

4.4. Discussion

We provided for the first time a snapshot of the genetic structure and variability among Brazilian CSDaV isolates collected from both symptomatic and asymptomatic citrus trees grown in fields affected by CSD disease. To date, only two CSDaV genome sequences were fully described, showing 11% nucleotide diversity between them (BARROS, 2006) (GenBank accession No. AY884005 and DQ185573). Both of these well-described CSDaV isolates were obtained from Rangpur lime tissues as rootstock of sweet orange trees collected from the same citrus region assessed in this study (MACCHERONI et al., 2005; BARROS, 2006), which is relevant since we can study the evolution of this virus in this CSD-affected area. In the current study, sequence analyses of five regions of the CSDaV genome representing almost 42% of the whole genome of 31 isolates, sampled from different hosts/plant tissues,

showed a low genetic diversity. It is not a surprising finding because genetic stability has been considered as a rule in natural plant virus populations (GARCIA-ARENAL et al., 2001) and similar low genetic diversity were previously reported for many other RNA plant viruses (NOURI et al., 2014; KONG et al., 2000; RUBIO et al., 2000; ALE-AGHA & RAKHSHANDEHROO, 2013; OGAWA et al., 2008; SEO et al., 2009; MORENO et al 2004; WALIA et al., 2013). It has been shown that systemic infections and other events such as host change and transmission can impose bottlenecks, the most common effects of genetic drift, which have been inferred from the low genetic diversity of plant virus populations (ALI & ROOSSINCK, 2008; LI & ROOSSINCK, 2004) and which might be the reason of the low genetic diversity among the CSDaV isolates.

The low nucleotide variability observed for the CP, the MT and the RdRP regions of CSDaV genome included in this study suggests that selective pressures in these segments are high to maintain nucleotide and amino acid conservation probably for biological functions (HAMMOND et al., 1997). It has been shown that the coat protein (CP) plays critical roles in virus packaging and stability, and interactions with plant host (HAMMOND et al., 1997). Similarly, the MT and the RdRP domains play key roles in viral replication, involving mRNA capping and enhanced stability of viral genomes (methyltransferase) (AHOLA et al., 2000) and transcription and replication of RNA virus genomes (RdRP) (ARNOLD & CAMERON, 2008). On the other hand, the MDR and He regions demonstrated higher nucleotide variability. The MDR segment showed the highest genetic diversity among all studied regions here, and it was the only region that had one site detected as being under positive selection. The MDR segment represents a multi-domain region that contains numerous predicted single domains related to different activities. Interestingly, the MDR was the unique multi-domain region found along the CSDaV genome and was the single region that we could not align with other reference members of the family *Tymoviridae*. It seems that this region is unique and associated with CSDaV isolates and could be related with some processes of virus adaption (MOURY, 2004) to a different host or plant tissues. However, at this time there is no information about the function(s) of this multi-domain, and then further studies are needed to evaluate the real role of this region in CSDaV. Probably because the pair of primers designed for the He region was highly degenerate, we were not able to amplify the He segment in several samples assessed here and it may be possible that the low number of isolates (nine) has influenced the results. From this work, it is clear that there is a genetic diversity between

the CSDaV isolates assessed here and the CSDaV references previous reported. The only isolates that showed close phylogenetic relationships with the CSDaV reference isolates were those isolated from the citrus roots, which were also detected as recombinants in this work, pointing the CSDaV reference sequences as the possible parents. Since we know that the CSDaV reference isolates were isolated from rootstock tissues of citrus trees as well (MACCHERONI et al., 2005; BARROS, 2006), this results also provide some evidence about the heterogeneous distribution of virus variants at different locations (leaves and roots) within hosts. Other study already reported that the diversity of virus population is different between old and young tissues, suggesting the tree could reflect the chronology of the appearance of virus diversity (JRIDI et al., 2006).

Phylogenetic analyses showed two new genetic clades for the CSDaV isolates included in this investigation, and one of them showed higher association with symptomatic trees. Higher nucleotide diversity, dN/dS ratio values and number of amino acid changes were found for isolates from symptomatic plants in coding regions located closer to the 5' end region of the CSDaV genome (MT and MDR), whereas coding regions located closer to the 3' end region showed more conservation. It is important to say that these isolates belonging to these two new genetic clades were all isolated from the citrus leaves, which have shown to have CSDaV variants, compared to the CSDaV isolates from the roots (this work and the references isolates). It is possible that the CSDaV isolates infecting rootstock tissues were subjected to some positive selection pressures, mainly on the coding regions closer to the 5' end region, to be able to infect tissues in the citrus canopy, culminating with two different variants of CSDaV, where one of them might be more efficient in infecting CSD-susceptible plants and/or more severe in developing CSD symptoms. Other factors, such as the susceptibility of the citrus rootstock and climate (drought and higher temperature) seem to contribute to the development of the CSD. The confirmed presence of CSDaV in trees grafted on symptomatic and asymptomatic susceptible rootstocks, and symptomatic and asymptomatic tolerant rootstocks, suggest that CSDaV is able to infect a wide host range in CSD-affected region, but the symptoms are not always developed. The results obtained here do not discard the possibility of a mixed or co-infection of the CSDaV and other virus(es), which was already proposed as a cause of CSD (MACCHERRONI et al., 2005; unpublished manuscript). CSDaV and other members of the genus *Marafivirus* have been frequently associated in mixed or co-infections in other pathosystems. CSDaV was found to be part of a multiple virus infection in Pinot Noir grapevine (PANTALEO et al., 2010) and in grapevine Syrah showing

decline symptoms (AL RWAHNIH et al., 2009). Recently, Villamor et al. (2016) found CSDaV infecting California nectarines showing stem-pitting symptoms and also revealed the presence of a new virus of the genus *Marafivirus*, which shared 70% of nucleotide sequence similarities to CSDaV, co-infecting these plants. All these results obtained in this investigation could together provide new insights into the role of CSDaV in symptom development in plants affected by CSD and contribute for further epidemiological studies.

4.5. Acknowledgments

This work was financially supported by National Institute of Science and Technology of Genomics for Citrus Improvement (INCT Citros), National Council for Scientific and Technological Development (Proc. No. 306230/2013-5 and 2008/57909-2) and Fundação de Apoio à Pesquisa no Estado de São Paulo (Proc. No. 2008/57909-2). EEM was supported, in part, by National Council for Scientific and Technological Development (CNPq) fellowship and also by Doctoral Sandwich Program Abroad (PDSE/CAPES) fellowship. We wish to thank our lab colleagues from 'Centro de Citricultura Sylvio Moreira' for constructive suggestions and discussions and we are grateful to Dr. Bryce W. Falk for critical reading of the manuscript and helpful suggestions.

5. CAPÍTULO III:

Brief Report of the Construction of Full-Length cDNA Clone of *Citrus Sudden Death-Associated Virus*

Brief Report of the Construction of Full-Length cDNA Clone of *Citrus Sudden Death-Associated Virus*

Emilyn Emy Matsumura[1,2], Helvécio Della Coletta Filho[2], Shahideh Nouri[3], Bryce W. Falk[3] Marcos Antonio Machado[2,*]

[1] Instituto de Biociências de Botucatu, Universidade Estadual Paulista, Botucatu, SP, Brazil; [2] Laboratório de Biotecnologia, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Cordeiropolis, SP, Brazil; [3] Department of Plant Pathology, University of California Davis, Davis, California, United States of America.

* Correspondence: marcos@centrodecitricultura.br; Tel.: +55 19 3546-1399

Abstract

Citrus sudden death-associated virus (CSDaV) is a RNA virus that was suggested to be associated with citrus sudden death (CSD) in Brazil. Here, we constructed a full-length of CSDaV cDNA clone using a citrus plant showing CSD symptoms as virus source. The infectivity of the clones was tested on *Nicotiana benthamiana* and citrus plants by agroinfiltration assay. Both *N. benthamiana* and citrus plants showed a very mild symptom of pale coloration of the leaves after 10 and 15 days of post-infiltration, respectively. However, the symptom in those plants seemed to disappear after a couple of days. By RT-PCR, we could detect the presence of CSDaV on non-agroinfiltrated upper leaves from *N. benthamiana* after 7 dpi, but after 14 dpi the same leaves were RT-PCR negative for CSDaV or showed a very weak PCR product. All tested non-agroinfiltrated upper leaves from citrus plants showed to be negative in detecting CSDaV. Work is in progress on improving the infectivity of the CSDaV clones. This tool will represent a complementary approach to study the CSDaV role in citrus plants and to investigate the etiology of CSD.

Keywords: CSDaV; infectious clone; In-Fusion cloning; CSD.

5.1. Introduction

A previous work has associated *Citrus sudden death-associated virus* (CSDaV) with Citrus sudden death (CSD), an important citrus disease in Brazil (MACCHERONI et al., 2005), which causes general decline symptoms, such as pale green coloration of leaves, defoliation and death of the roots, in plants of sweet orange grafted mainly in Rangpur lime rootstock (BASSANEZI et al., 2007; MÜLLER et al., 2002; ROMÁN et al., 2004). CSD-symptomatic plants also show an yellow stain in the rootstock bark as a characteristic symptom of this disease (BASSANEZI et al., 2007; MÜLLER et al., 2002; ROMÁN et al., 2004). Although the association of CSDaV and CSD has not been proved so far, in our previous works, we also have shown evidences of this association and suggested that there is a specific CSDaV genotype more associated with CSD-symptomatic plants (unpublished manuscript).

CSDaV was classified as a monopartite, positive-sense, single-stranded RNA virus belonged to genus *Marafivirus* in the family *Tymoviridae* (MACCHERONI et al., 2005). The CSDaV virions are isometric particles of ≈ 30 nm in diameter and its RNA genome has approximately 6.8 kb in length encompassing two ORFs (MACCHERONI et al., 2005). The ORF1 encodes for a 240 kDa polyprotein (p240), which contains conserved signatures of the methyltransferase (MT), the papain-like protease (PRO), the helicase (He), the RNA-dependent RNA polymerase (RdRP) domains and two subunits of the coat protein (CP) of 21 and 22 kDa in size, respectively (MACCHERONI et al., 2005). The ORF 2 at the 3' end region seems to encode for a putative movement protein of 16 kDa (p16) (MACCHERONI et al., 2005).

Difficulties in purifying CSDaV particle and generate CSD symptoms by trying to use conventional methods of transmission (SANTOS, 2011) have prevented progress in studying this disease. The constructions of full-length infectious cDNA clones have showed to be a powerful tool to better understand RNA virus-plant interactions (PAKNIAT et al., 2010; TUO et al., 2015). In order to study the role of the CSDaV in plant host, this work have constructed a full-length CSDaV cDNA clones and tested their infectivity on *Nicotiana benthamiana* and citrus plants.

5.2. Methods

5.2.1. Virus source and RNA extraction

CSDaV was obtained from leaves of sweet oranges grafted on Rangpur lime rootstock showing CSD symptoms (i.e., occurrence of yellow stain in the rootstock bark), collected in a grove located in the municipality of Comendador Gomes (southwestern Minas Gerais State), Brazil. The total RNA was extracted from 300 mg of leaves using the CTAB extraction protocol adapted from BEKESIOVA et al. (1999), where the LiCl was replaced by isopropanol (1 vol) in the precipitation phase.

5.2.2. In-Fusion construction of a full-length cDNA clone of CSDaV

The In-Fusion construction of a full-length cDNA clone of CSDaV was first simulated on the SnapGene software (http://www.snapgene.com/), using the full sequence of a previously reported CSDaV genome (GenBank accession number AY884005) and a linearized pJL89 binary vector under control of the 35S promoter of *Cauliflower mosaic virus*. A set of primers (Table 5.1) was automatically designed in planning to insert two CSDaV fragments (fragment I of 3320 bp and fragment II of 3360 bp) into pJL89 vector by 20-base complementary overlapping ends (Figure 5.1). The cDNA was synthesized from 500 ng of total RNA using the Superscript™ III reverse transcriptase (Invitrogen, Carlsbad, CA, US) and oligo(dT) primer, according to the manufacturer's instructions. To proceed the In-Fusion cloning, the full sequence of the pJL89 vector (4675 bp) and the CSDaV fragments I and II were amplified with their respective primers, using the CloneAmp HiFi PCR premix (Clontech), following the manufacturer's protocol. The amplicons were analyzed on 0.8% agarose gel and purified using Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, California). The In-Fusion reaction was performed in 10 ul of total volume, containing 2 ul of 5X In-Fusion HD Enzyme Premix, 100 ng of each purified fragment and dH_2O from the In-Fusion HD PCR Cloning Kit (Clontech). The reaction mix was incubated at 50 °C for 15 min, and then placed on ice for transformation using *E. coli* Stellar™ competent Cells (Invitrogen). The transformants were screened by colony PCR using specific primers for the CSDaV fragment I and II: CSDaV-midIF/CSDaV-midIR and CSDaV-midIIF/CSDaV-midIIR, respectively. The PCR-positive colonies were selected for plasmid purification using the Maxiprep Purification kit (Qiagen). The full-length cDNA

clone of CSDaV was confirmed by restriction enzyme digestion and Sanger sequencing using several primers designed along the CSDaV genome.

Table 5.1. Primers used in the construction of full-length cDNA clones of CSDaV. The purpose of each primer is presented.

Primer	Sequence (5'-3')	Purpose
CSDaV-VF	GGGTCGGCATGGCATCTC	Inverse PCR of the pJL89 vector
CSDaV-VR	CCTCTCCAAATGAAATGAACTTCCTTAT ATAGAGGA	
CSDaV-FIF	GTTCATTTCATTTGGAGAGGGTCCCCTG TGATCGTCTCTCC	Amplification of fragment I of the CSDaV
CSDaV-FIR	TGGAGATGCCATGCCGACCCGGGAGAC CAGTAATGGTTTTCCACT	
CSDaV-VIR	GGGAGACCAGTAATGGTTTTCCACTC	Inverse PCR of the pJL89 vector after insertion of fragment I
CSDaV-FIIF	AAAACCATTACTGGTCTCCCGGCCTCAG AAGCCTGGCG	Amplification of fragment II of the CSDaV
CSDaV-FIIR	TGGAGATGCCATGCCGACCCTTTTTTTT TTTTTTTATTAAATAATAAAGAAAAACG GTCTTTGGATCGAC	
CSDaV-midIF	TGGACAGATCTGTGACCTCTTCCTCT	Detection of CSDaV (fragment I)
CSDaV-midIR	TCAGATGATGGGGAGGAGAGCTGAT	
CSDaV-midIIF	TGGTTCCACAATGAGTTCCCAAAGGC	Detection of CSDaV (fragment II)
CSDaV-midIIR	CAATTCACCTTGTAGCAGAGTGGTGTC	
PJL89F	AAGGGATGACGCACAATCCCACTATC	Sanger sequencing
PJL89R	ATCGGGGAAATTCGAGCTCTCCCTTA	
CSDaV-5UTRF	CCCTCCAGCCGGAAAGATATTTTTGC	Sanger sequencing
CSDaV-3UTRR	AGAAAAACGGTCTTTGGATCGACCGG	
CSDaV-IR	GGGAGACCAGTAATGGTTTTCCACTC	Sanger sequencing
CSDaV-IIF	AGGAGATCCGTCCGTCTGATCCATAT	
CSDaV-CPF	GCCATCTACACCACACTCTC	Detection of CSDaV
CSDaV-CPR	TTGGAGTAGACGGAGTAGGA	

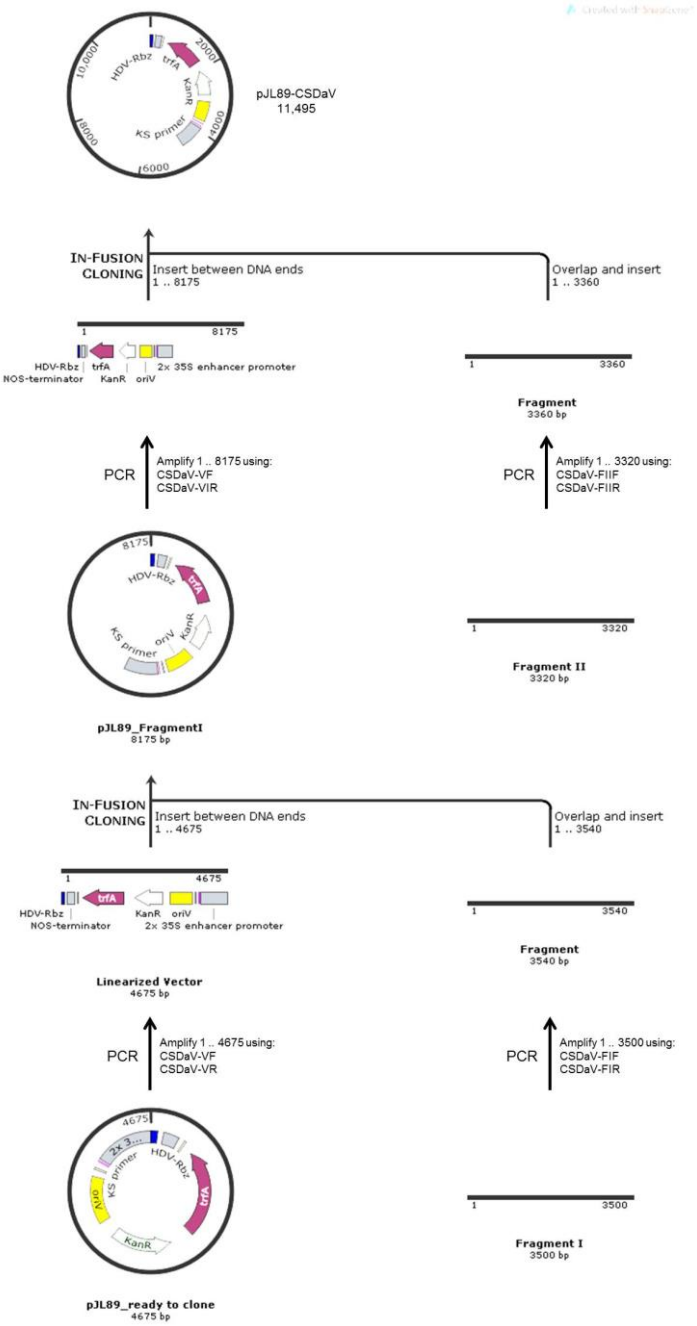

Figure 5.1. Schematic diagram showing the strategy used to construct the full-length cDNA clone of CSDaV. Two genomic fragments overlapping the complete genome (fragments I and II) and the linearized pJL89 vector containing a 35S promoter were fused to generate the pJL89-CSDaV by In-Fusion cloning. Primers used to amplify the fragments are indicated. The simulation was done on the SnapGene software.

5.2.3. Agroinfiltration assay

The full-length cDNA clones of CSDaV were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation using a Gene Pulser apparatus (Bio-Rad, Germany) according to the manufacturer's specifications. Pre-inoculum cultures of selected individual clones of *A. tumefaciens* were grown overnight at 28 °C in LB medium containing 10 μg/ml of rifampicin, 20 μg/ml of gentamicin and 50 μg/ml of kanamycin. An aliquot of 500 ul from the pre-inoculum cultures was inoculated in 25 ml of L-MESA medium (LB media supplemented with 10 mM MES and 20 μM acetosyringone), containing the same antibiotics used in the pre-inoculum, and grown overnight at 28 °C to an OD_{600} nm of between 0.8 and 1.2. The cells were centrifuged for 10 min at 3,500 x g, resuspended in agroinduction medium (10 mM MgCl2, 0.5 l M acetosyringone and 10 mM MES, pH 5.7) to an OD_{600} nm of 1.0 and incubated at room temperature for 5 h in the dark. The same procedure was done for *A. tumefaciens* colony previously transformed with P19 silencing suppressor gene from *Tomato bushy stunt virus* (TBSV). Cultures of *A. tumefaciens* containing the CSDaV constructions and P19 silencing suppressor were combined at 1:1 ratios and infiltrated on the abaxial surface of young expanded leaves of the transgenic *Nicotiana benthamiana*, expressing the helper component-proteinase (HC-Pro) derived from a potyviral genome, and leaves of *Citrus sinensis* seedlings (grafted on Rangpur lime rootstock), using a 5 ml syringe without a needle. We infiltrated four plants with each CSDaV obtained clones and four leaves per plant. A plant infiltrated with *A. tumefaciens* containing only the P19 silencing suppressor was used as negative control. The agroinfiltrated plants were maintained in a greenhouse and the symptoms were monitored during 21 days (for *N. benthamiana*) and during 45 days (for citrus plants).

5.2.4. RT-PCR for CSDaV detection

To detect CSDaV in the agroinfiltrated plants, total RNA was extracted from the agroinfiltrated and non-agroinfiltrated upper leaves of the *N. benthamiana* plants after 7 and 14 days of post-infiltration (dpi), using TRIzol reagent according to the manufacturer's protocol (Life Science Research, Carlsbad, CA). The RNA extraction from the agroinfiltrated and upper leaves of the citrus plants was done after 15 and 30 dpi using the TRIzol reagent as well. The cDNAs were synthesized by the Superscript™ III reverse transcriptase (Invitrogen, Carlsbad, CA, US), according to the manufacturer's instructions and the PCR reactions were

performed using the CloneAmp HiFi PCR premix (Clontech) by using the CSDaV specific primers. All the PCR products obtained were confirmed by Sanger sequencing.

5.3. Results

Among 40 *E. coli* colonies screened by colony PCR using CSDaV primers for fragments I and II, only three of them were positive for both fragments. The other colonies were negative or positive only for one of the two fragments. These three PCR-positive colonies were selected to plasmid extraction and checked by the digestion with a single restriction enzyme (RsrII-NEB). The restricted plasmid obtained from the three selected clones showed a single fragment with the expected size on the 0.8% agarose gel (Figure 5.2). The three CSDaV cDNA clones were named as IC-CSDaV-7, IC-CSDaV-9 and IC-CSDaV-20 and used in *A. tumefaciens* transformation in further agroinfiltration assay.

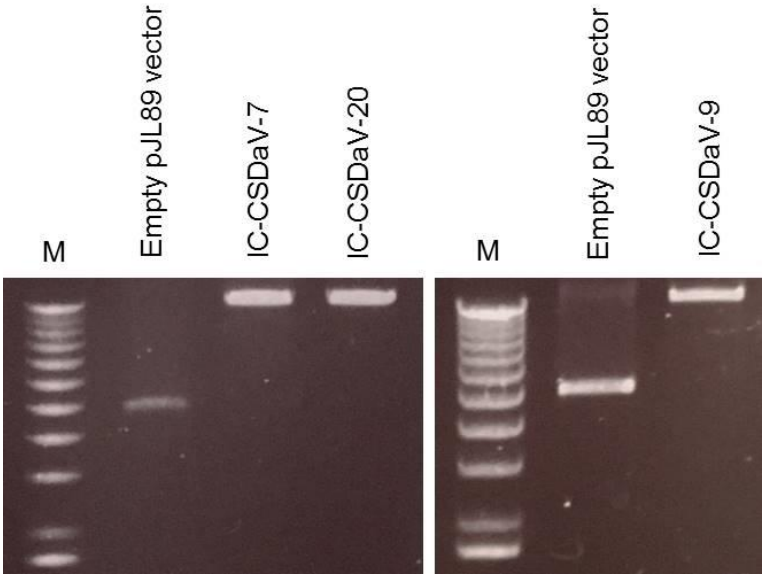

Figure 5.2. Electrophoretic pattern from the digestion of the IC-CSDaV-7, IC-CSDaV-9 and IC-CSDaV-20 clones by a single restriction enzyme (RsrII) to confirm the insertion of full-length CSDaV sequence into pJL89 vector. M: 1 Kb plus ladder.

Of 12 agroinfiltrated *N. benthamiana* plants (four of each CSDaV cDNA clone), only three of them (one agroinfiltrated with IC-CSDaV-7 and two agroinfiltrated with IC-CSDaV-9) exhibited, after 10 dpi, a very mild symptom of pale coloration on the upper new leaves (Figure 5.3b), but this symptom seemed to disappear in the following days. Total RNA was extracted from the agroinfiltrated and upper non-agroinfiltrated leaves from all plants after 7 dpi and from only upper non-agroinfiltrated leaves after 14 dpi. The reverse transcription for all RNA samples was done as described in methods. A pair of specific primers (CSDaV-

midIIF/CSDaV-midIIR) was used to amplify a 2 Kb fragment for CSDaV detection. RT-PCRs were positive for all tested agroinfiltrated leaves and for 7 non-agroinfiltrated upper leaves after 7 dpi (Figure 5.3c). After 14 days, very weak bands were obtained for only three non-agroinfiltrated upper leaves (Figure 5.3c). All the PCR products were purified from the gel and confirmed by Sanger sequencing.

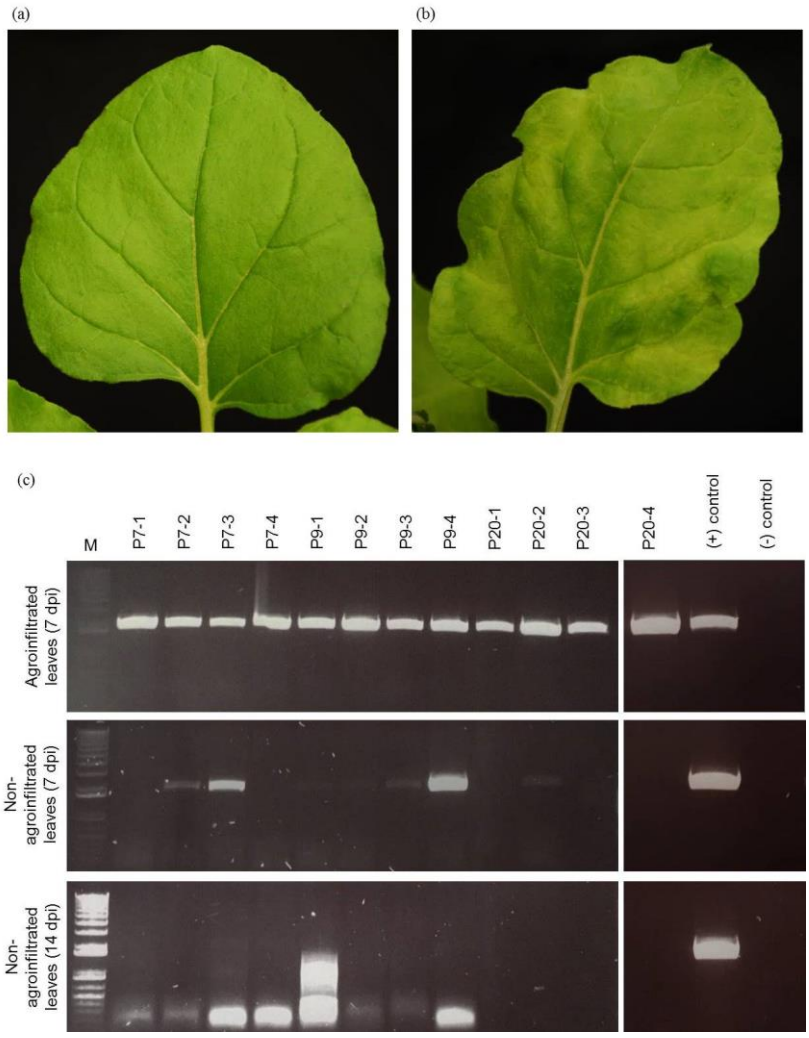

Figure 5.3. Mild symptom detected on non-agroinfiltrated upper leaf of *N. benthamiana* plant after 10 dpi with IC-CSDaV-9 clone (b), compared with non-agroinfiltrated leaf from *N. benthamiana* inoculated only with P19 silencing suppressor (negative control) (a). RT-PCR detection of CSDaV on agroinfiltrated and non-agroinfiltrated upper leaves from *N. benthamiana* inoculated with IC-CSDaV-7 (P7), IC-CSDaV-9 (P9) and IC-CSDaV-20 (P20) clones (c). M: 1 Kb plus DNA ladder; dpi: days post-inoculated.

From the agroinfiltration assay using citrus plants, three plants (two agroinfiltrated with IC-CSDaV-7 and one agroinfiltrated with IC-CSDaV-9) showed a mild symptom of pale coloration on the upper new leaves (Figure 5.4) after 15 dpi, but as we observed on *N.*

benthamiana, the citrus plants also recovered from the symptom in the following days. Leaf curling symptom was observed only in the agroinfiltrated leaves (data not shown). As soon as the symptoms were detected, the agroinfiltrated and non-agroinfiltrated upper leaves from all citrus plants were collected and used for total RNA extraction and for RT-PCR detection, using a specific pair of primers to amplify a 700 bp CSDaV fragment (CSDaV-CPF/CSDaV-CPR). RT-PCRs were positive only for agroinfiltrated leaves (Figure 5.4). Non-agroinfiltrated leaves were also collected after 30 dpi for CSDaV detection, but were negative for all tested plants. Citrus agroinfiltrated plants were monitored for 45 dpi, but any progress of the symptoms was detected.

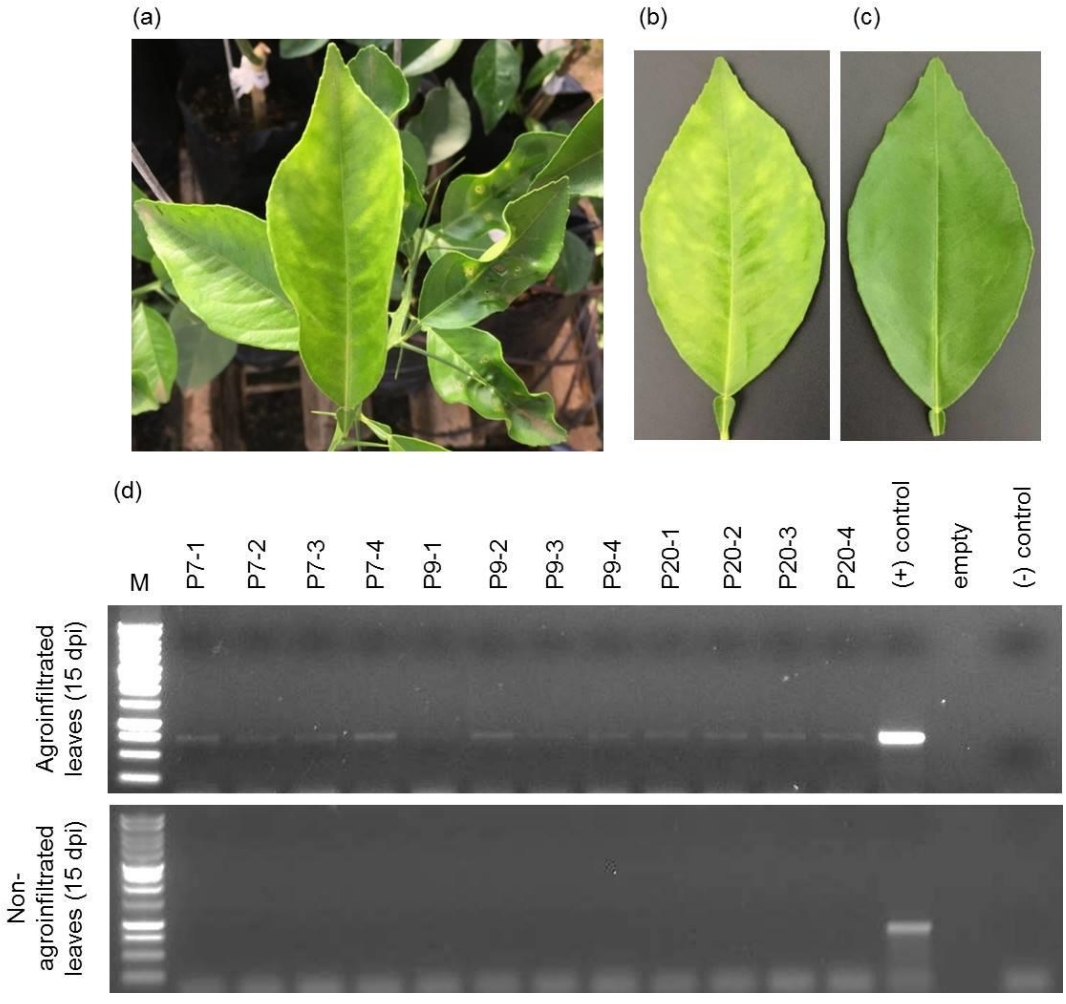

Figure 5.4. Mild symptoms detected on non-agroinfiltrated upper leaf of citrus plant after 15 dpi of inoculation with IC-CSDaV-9 clone (a and b), compared with non-agroinfiltrated leaf from citrus plant inoculated only with P19 silencing suppressor (negative control) (c). RT-PCR detection of CSDaV on agroinfiltrated and non-agroinfiltrated upper leaves from citrus plant inoculated with IC-CSDaV-7 (P7), IC-CSDaV-9 (P9) and IC-CSDaV-20 (P20) clones (d). M: HighRanger 1 Kb DNA ladder; dpi: days post-inoculated.

5.4. Discussion

In this work, we have used an In-Fusion cloning strategy to construct a full-length cDNA clone of CSDaV. This strategy allowed the cloning of two CSDaV fragments spanning the complete genome of this virus into a specific location in pJL89 vector under control of the 35S promoter. However, of 40 tested *E. coli* colonies, only three of them showed to have the full-length of CSDaV genome. Most of the other tested colonies showed the presence of only one or part of one CSDaV fragment. This low efficiency of cloning might be attributed to the instability of clones, probably, due to spontaneous mutations, deletions, or rearrangements of viral fragments during the cloning procedures (TUO et al., 2015) and due to the toxicity of the viral gene products as well (GAO et al., 2012; LÓPEZ-MOYA; GARCÍA, 2000; TUO et al., 2015)

The infectivity of CSDaV cDNA clones was tested by agroinfiltration assays on *N. benthamiana* and citrus plants. In field grown citrus, plants affected by CSD disease show symptoms characterized by pale green coloration of the leaves, defoliation, death of the roots, and presence of a characteristic yellow stain in the rootstock bark (MÜLLER et al., 2002; ROMÁN et al., 2004). From our agroinfiltration experiments, a mild symptom of pale green coloration was detected in non-agroinfiltrated upper leaves of both *N. benthamiana* and citrus plants after 10 and 15 dpi, respectively. However, the symptom was observed in a few number of plants and, interestingly, for both *N. benthamiana* and citrus plants, the symptom seemed to disappear after a couple of days. RT-PCR for CSDaV detection on upper new leaves of *N. benthamiana* plants showed consistent results with the symptom observations. We were able to detect the presence of CSDaV on the upper leaves in seven plants after 7 dpi, but after 14 days, very weak bands were obtained from only three plants. Initially, two reasons for these results were suggested, (i) the *N. benthamiana* plants are avoiding the CSDaV replication by natural plant defense mechanisms and (ii) *N. benthamiana* is not a host for CSDaV.

The combination of sweet orange grafted on Rangpur lime rootstock was the original host of CSDaV (MACCHERONI et al., 2005), but this virus has also been found in sweet orange grafted in different varieties of rootstocks (unpublished manuscript). However, the agroinfiltration experiments conducted in this work on citrus plants were not able to prove the CSDaV infectivity, because the RT-PCR assays were negative for all tested non-agroinfiltrated upper leaves, although the agroinfiltrated leaves were all RT-PCR positives. It

has been shown that the infectivity efficiency of viral cDNA clones depends on the amount of DNA inoculated per leaf, the host and the method of inoculation (LIN; ROGER; YEH, 2002). Citrus plants affected by CSD has shown an incubation period of at least 2 years before symptoms appear (BASSANEZI et al., 2007) and thus we are not sure how many days after infiltration are needed to detect CSDaV by RT-PCR. Moreover, mutations in the cDNA clones may influence the viral infectivity. Sanger sequencing of the full-length of CSDaV cDNA clones revealed nucleotide diversity between these clones and the CSDaV genome sequence available in GenBank, which might be affecting the performance of CSDaV as infectious clones.

To our knowledge, this is the first report of the construction of a full-length cDNA clone of the CSDaV. Work is in progress on improving the infectivity of the CSDaV clones in citrus plants, which will represent a complementary tool to study the CSDaV role in those plants and to investigate the etiology of CSD.

5.5. Acknowledgements

This work was financially supported by National Institute of Science and Technology of Genomics for Citrus Improvement (INCT Citros), National Council for Scientific and Technological Development (Proc. No. 306230/2013-5 and 2008/57909-2) and Fundação de Apoio à Pesquisa no Estado de São Paulo (Proc. No. 2008/57909-2). EEM was supported, in part, by National Council for Scientific and Technological Development (CNPq) fellowship and also by Doctoral Sandwich Program Abroad (PDSE/CAPES) fellowship. We wish to thank our lab colleagues from the ‘Centro de Citricultura Sylvio Moreira’ and from UC Davis Department of Plant Pathology for constructive suggestions and discussions.

6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

As abordagens utilizadas neste trabalho permitiram um estudo das populações de vírus presentes em plantas de citros cultivadas numa região afetada pela morte súbita dos citros (MSC). O sequenciamento de alta performance, combinando dados de RNA-seq e de pequenos RNAs, foi eficiente para revelar a ocorrência de uma infecção viral mista, incluindo, predominantemente, os dois vírus previamente associados à MSC, *Citrus tristeza virus* (CTV) e *Citrus sudden death-associated virus* (CSDaV), e um vírus endógeno de citros, *Citrus endogenous pararetrovirus* (CitPRV), até então não identificado em plantas de citros no Brasil. Também foi possível identificar vírus menos predominantes (0,13% dos *reads* totais), os quais permitiram a caracterização parcial do genoma de dois possíveis novos vírus, denominados nesse estudo como *Citrus jingmen-like virus* (CJLV) e *Citrus virga-like virus* (CVLV). A estratégia de sequenciamento também contribuiu com as análises de diversidade genética, permitindo a identificação de dois genótipos diferentes de CTV, filogenéticamente associados aos grupos severos RB e VT, e dois genótipos diferentes de CSDaV. Assim como em trabalhos anteriores, não foi possível fazer uma associação entre os genótipos de CTV encontrados e a MSC. No entanto, a distribuição dos *reads* sobre o genoma de referência de CTV indicam que *reads* derivados de plantas assintomáticas parecem se concentrar na região do gene p20, relacionado com a supressão de silenciamento dos genes, já *reads* derivados de plantas sintomáticas mostraram ser mais abundantes na região dos genes associados às atividades de interação com o hospedeiro (p13, p18 e p33). Embora não conclusivos, tais resultados geram novas questões sobre uma possível função(ões) desses genes no desenvolvimento da MSC. As análises comparativas, correlacionando a frequência dos vírus com a presença ou ausência de sintomas de MSC, indicaram uma possível associação entre CitPRV e plantas sintomáticas e uma forte associação entre os novos vírus identificados neste estudo (CJLV e CVLV) e plantas assintomáticas, resultados estes que geram novas questões sobre as funções desses vírus nas plantas de citros estudadas. No entanto, a abordagem deste estudo indicou, assim como em outros trabalhos, o CSDaV como o vírus mais associado a doença, e permitiu identificar, pela primeira vez, um genótipo específico de CSDaV mais associado às plantas sintomáticas. Tais resultados foram verificados através de um estudo mais específico da diversidade do CSDaV. Análises de cinco regiões genômicas de isolados de CSDaV, obtidos de plantas sintomáticas e assintomáticas para MSC, revelaram a

predominância de dois grupos filogeneticamente diferentes, e indicaram, de forma coerente com os resultados obtidos previamente, que um grupo específico está mais associado às plantas sintomáticas. Além disso, essa segunda abordagem demonstrou que o grupo de isolados mais associado às plantas com sintomas apresentou uma maior diversidade genética, comparado com os isolados de plantas assintomáticas, e que a maior diversidade foi verificada para uma região específica, e de função não conhecida, do genoma do vírus. Poderia esta região estar associada a função(ões) de adapatação do vírus tornando-o mais infectivo para a planta? Os resultados obtidos conduziram este trabalho para uma construção de um clone de cDNA do CSDaV, a ser utilizado em futuros estudos de função do vírus e etiologia da doença. Embora ainda não seja possível definir de forma conclusiva o agente causal da MSC, as abordagens utilizadas neste trabalho geraram importantes e novas informações com relação a esse patossistema e vírus envolvidos, as quais devem ser levadas em consideração em estudos futuros.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. AHOLA, T.; DEN BOON, J.A.; AHLQUIST, P. Helicase and Capping Enzyme Active Site Mutations in Brome Mosaic Virus Protein 1a Cause Defects in Template Recruitment, Negative-Strand RNA Synthesis, and Viral RNA Capping. Journal of **Virology,** v. 74, p. 19, 8803-8811, 2000.
2. ALBIACH-MARTI, M. R. The Complex Genetics of Citrus tristeza virus. **Current Issues in Molecular Virology - Viral Genetics and Biotechnological Applications**, p. 1–26, 2013.
3. ALE-AGHA, G.N.; RAKHSHANDEHROO, F. Detection and Molecular Variability of Fig Fleck-Associated Virus and Fig Cryptic Virus in Iran. **J Phytopathol**, v. 162, p. 417-425, 2013.
4. ALI, A.; ROOSSINCK, M.J. Genetic bottlenecks. In **Plant Virus Evolution**., Editor Roossinck, M.J., Eds.; Publisher: Springer Berlin Heidelberg, Germany; p. 123–131, 2008.
5. AL RWAHNIH, M.; DAUBERT, S.; GOLINO.; D AND ROWHANI, A. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms

reveals a multiple virus infection that includes a novel virus. **Virology**, v. 387, p. 395-401, 2009.

6. AMBRÓS, S.; RUIZ-RUIZ, S.; PEÑA, L.; MORENO. A genetic system for citrus tristeza virus using the non-natural host nicotiana benthamiana: An update. **Frontiers in Microbiology**, v. 4, n. JUL, p. 1–14, 2013.
7. ARNOLD, J. J.; CAMERON, C.E. Structure-Function Relationships Among RNA-Dependent RNA Polymerases. **Curr Top Microbiol Immunol**, v. 320, p. 137–156, 2008.
8. BARROS, C.C.P. **Sequenciamento do genoma completo e expressão heteróloga da capa protéica do Marafivirus associado a morte súbita dos citros.** 2006. Dissertation, Universidade Católica de Brasilia; Brasília-BR, 2006.
9. BASSANEZI, R. B.; BERGAMIN FILHO, A.; AMORIM, L.; GIMENES-FERNANDES, N.; GOTTWALD, T. R. et al. Spatial and temporal analyses of citrus sudden death as a tool to generate hypotheses concerning its etiology. **Phytopathology**, v. 93, n. 16, p. 502–512, 2003.
10. BASSANEZI, R. B.; MONTESINO, L. H.; SANCHES, A. L.; SPÓSITO, M. B.; STUCHI, E. S. et al. Effect of Citrus Sudden Death on Yield and Quality of Sweet Orange Cultivars in Brazil. **Plant Disease**, v. 91, n. 11, p. 1407–1412, 2007.
11. BEKAL, S.; DOMIER, L. L.; GONFA, B.; MCCOPPIN, N. K.; LAMBERT, K. N.; BHALERAO, K. A novel flavivirus in the soybean cyst nematode. **Journal of General Virology**, v. 95, n. PART 6, p. 1272–1280, 2014.
12. BEKESIOVA, I.; NAP, J. P AND MLYNAROVA, L. Isolation of high quality DNA and RNA from leaves of the carnivorous plant Drosera rotundifolia. **Plant Mol. Biol**. Reptr, v. 17, p. 269-277, 1999.
13. BOVÉ, J. M.; AYRES, A. J. Etiology of three recent diseases of citrus in São Paulo State: sudden death, variegated chlorosis and huanglongbing. **IUBMB life**, v. 59, n. 4–5, p. 346–354, 2007.
14. BLITVICH, B. J.; FIRTH, A. E. **Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization**. [s.l: s.n.]. v. 7
15. CANTÚ, M. D.; MARIANO A. G.; PALMA, M. S.; CARRILHO, E.; WULFF, N. A. Proteomic analysis reveals suppression of bark chitinases and proteinase inhibitors in citrus plants affected by the citrus sudden death disease. **Phytopathology**, v. 98, n. 10,

p. 1084–1092, 2008.

16. COETZEE, B.; FREEBOROUGH, M. J.; MAREE, H. J.; CELTON, J. M.; REES, D. J. G.; BURGER, J. T. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. **Virology**, v. 400, n. 2, p. 157–163, 2010.
17. COSTA, A. T.; NUNES, W. M. C.; ZANUTTO, C. A.; MÜLLER, G. A. Stability of Citrus tristeza virus protective isolates in field conditions. **Program**, n. 1, p. 693–700, 2010.
18. DAVINO, S. WILLEMSEN, A.; PANNO, S.; DAVINO, M.; CATARA, A.; ELENA, S. F.; RUBIO, R. Emergence and Phylodynamics of Citrus tristeza virus in Sicily, Italy. **PLoS ONE**, v. 8, n. 6, 2013.
19. FOLIMONOVA, S. Y. Developing an understanding of cross-protection by citrus tristeza virus. **Frontiers in Microbiology**, v. 4, n. APR, p. 1–9, 2013.
20. GAO, R.; TIANB, Y. -P.; WANGA, J.; YINA, X.; LIA, X. -D. et al. Construction of an infectious cDNA clone and gene expression vector of Tobacco vein banding mosaic virus (genus Potyvirus). **Virus Research**, v. 169, n. 1, p. 276–281, 2012.
21. GARCIA-ARENAL, F.; FRAILE, A.; MALPICA, J.M. Variability and genetic structure of plant virus populations. **Annu. Rev. Phytopathol**, v. 39, p. 157-186, 2001.
22. GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nat Biotech**, v. 29, p. 644-652, 2011.
23. GOMES, C. P. C.; NAGATA, T.; JESUS JR, W. C.; BORGES NETO, C. R.; PAPPAS JR, G. J. et al. Genetic variation and recombination of RdRp and HSP 70h genes of Citrus tristeza virus isolates from orange trees showing symptoms of citrus sudden death disease. **Virology journal**, v. 5, p. 9, 2008.
24. HAMMOND, R.W.; KOGEL, R.; RAMIREZ, P. Variability of geographically distinct isolates of maize rayado fino virus in Latin America. **Journal of General Virology**, v. 78, p. 3153-3159, 1997.
25. HARPER, S. J. Citrus tristeza virus: Evolution of complex and varied genotypic groups. **Frontiers in Microbiology**, v. 4, n. APR, 2013.
26. **IBGE**. Indicadores IBGE - Estatística de produção agrícola, 2016.
27. JO, Y.; CHOI, H.; CHO, J. K.; YOON, J. Y.; CHOI, S. K.; CHO, W. K. In silico approach to reveal viral populations in grapevine cultivar Tannat using transcriptome

data. **Scientific reports**, v. 5, p. 15841, 2015.

28. JRIDI, C.; MARTIN, J.F.; MAREIE-JEANNE, V.; LABONNE, G.; BLANC, S. Distinct viral populations differentiate and evolve independently in a single perennial host plant. **J Virol**, v. 80, p. 2349-2357, 2006.
29. KONG, P.; RUBIO, L.; POLEK, M.; FALK, W.B. Population Structure and Genetic Diversity within California Citrus tristeza virus (CTV) Isolates. **Virus Genes**, v. 21, p. 139-145, 2000.
30. KUMAR, S.; TAMURA, K.; NEI, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. **Bioinform**, v. 5, p. 150-163, 2004.
31. LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P.; CHENNA, R.; MCGETTIGAN, P. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947–2948, 2007.
32. LIBRADO, P.; ROZAS, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, p. 1451–1452, 2009.
33. LI, R.; GAO, S.; HERNANDEZ, A. G.; WECHTER, W. P.; FEI, Z.; LING, K. S. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. **PLoS ONE**, v. 7, n. 5, p. 1–10, 2012.
34. LI, H., ROOSSINCK, M.J. Genetic bottlenecks reduce population variation in an experimental RNA virus population. **J. Virol**, v. 78, p. 10582-10587, 2004.
35. LIN, S. S.; ROGER, F. H.; YEH, S. D. Construction of in vivo and in vitro infectious transcripts of a Taiwan strain of Zucchini yellow mosaic virus. Bot. Bull. **Acad. Sin**, v. 43, p. 261-269, 2002.
36. LIU, Y.; HEYING, E.; TANUMIHARDJO, S. A. History, Global Distribution, and Nutritional Importance of Citrus Fruits. **Comprehensive Reviews in Food Science and Food Safety**, v. 11, n. 6, p. 530–545, 2012.
37. LOCKHART, B. E., MENKE, J., DAHAL, G. AND OLSZEWSKI, N.E. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. **J. Genet. Virol**, v. 81, p. 1579–1585, 2000.
38. LOLE, S.; BOLLINGER, R.C.; PARANJAPE, R.S.; GADKARI, D.; KULKARNI, S.S.; NOVAK, N.G. Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Fulllength human

immunodeficiencey virus Type 1 genomes from subtype Cinfected seroconverters in India, with evidence of intersubtype recombination. **J. Virol**, v. 73, p. 152-160, 2009.

39. LÓPEZ-MOYA, J. J.; GARCÍA, J. A. Construction of a stable and highly infectious intron-containing cDNA clone of plum pox potyvirus and its use to infect plants by particle bombardment. **Virus Research**, v. 68, n. 2, p. 99–107, 2000.
40. MACCHERONI, W.; ALEGRIA, M. C.; GREGGIO, C. C.; PIAZZA, J. P.; KAMLA, R. F., et al. Identification and Genomic Characterization of a New Virus (Tymoviridae Family) Associated with Citrus Sudden Death Disease Identification and Genomic. **Journal of Virology**, v. 79, n. 5, p. 3028–3037, 2005.
41. MAHY, B.W.J.; VAN REGENMORTEL, M.H.V. Emerging and Reemerging Virus Diseases of Plants. In Desk encyclopedia of plant and fungal virology, 1st ed.; Mahy, B.W.J., Van Regenmortel, M.H.V, eds. Publisher: **Ed. Elsevier**, p. 412-417, 2009.
42. MOURY, B. Differential Selection of Genes of Cucumber Mosaic Virus Subgroups. **Mol. Biol. Evol,** v. 21, p. 1602-1611, 2004.
43. MORENO, I.M.; MALPICA, J.M.; DI´AZ-PENDON, J.A.; MORIONES, E.; FRAILE, A.; GARCIA-ARENALA, F. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. **Virology**, v. 318, p. 451-460, 2004.
44. MORENO, P.; AMBROS, S.; ALBIACH-MARTI M, R.; GUERRI, J.; PENA, L. Citrus tristeza virus: A pathogen that changed the course of the citrus industry. **Molecular Plant Pathology**, v. 9, n. 2, p. 251–268, 2008.
45. MÜLLER, G.W.; DE NEGRI, J.D.; AGUILAR-VILDOSO, C.I.; MATTOS JÚNIOR, D.; POMPEU JÚNIOR, J. et al. Citrus sudden death: a new citrus disease in Brazil. Proceedings of Conference of IOCV, Riverside, US, p. 405-407, 2002.
46. NEVES, M. F.; TROMBIN, V. G.; MILAN, P.; LOPES, F. F.; CRESSONI, F. et al. O retrato da citricultura brasileira. **CitrusBR,** p. 138, 2010.
47. NOREEN, F.; AKBERGENOV, R.; HOHN, T.; RICHERT-POGGELER, K. R. Distinct expression of endogenous Petunia vein clearing virus and the DNA transposon dTph1 in two Petunia hybrida lines is correlated with differences in histone modification and siRNA production. **Plant Journal**, v. 50, n. 2, p. 219–229, 2007.
48. NOURI, S.; AREVALO, R.; FALK, B.W.; GROVES, R.L. Genetic Structure and Molecular Variability of Cucumber mosaic virus Isolates in the United States. **Plos One**, v. 9, 2014.

49. NOURI, S.; SALEM, N.; NIGG, J. C.; FALK, B. W. Diverse Array of New Viral Sequences Identified in Worldwide Populations of the Asian Citrus Psyllid (Diaphorina citri) Using Viral Metagenomics. **Journal of Virology**, v. 90, n. 5, p. 2434–2445, 2016.
50. NUNES, F. A. C. Construção de cDNA infeccioso de *Citrus sudden death-associated virus*. 2009. Dissertação de Mestrado. Universidade católica de Brasília. 2009.
51. OGAWA, T.; TOMITAKA, Y.; NAKAGAWA, A.; OHSHIMAB, K. Genetic structure of a population of Potato virus Y inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. **Virus Research**, v. 131, p. 199-212, 2008.
52. OLIVEIRA, P. N; DOVIS, V. L; MATTOS JÚNIOR, D. Nitrogênio na cultura de citros. In **Informações agronômicas**., Editor CASARIN, V., n. 19, p. 6-14, 2014.
53. PANTALEO, V.; MOXON, S.; MIOZZI, L.; MOULTON, V.; DALMAY, T.; BURGYAN, J. Identification of grapevine microRNAs and their targets using high throughput sequencing and degradome analysis. **Plant J**, v. 62, p. 960-976, 2010.
54. PRABHA, K.; BARANWAL, V. K.; JAIN, R. K. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. **Indian Journal of Virology**, v. 24, n. 2, p. 157–165, 2013.
55. QIN, X.-C.; SHIA, M.; TIANC, J.-H.; LIND, X.-D.; GAOA, D.-Y. et al. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 18, p. 6744–9, 2014.
56. RIVAS-VALENCIA, P.; LOEZA-KUK, E.; MORA-AGUILERA, G.; FEBRES, V.; OCHOA-MARTÍNEZ, D. et al. Population Structure of Citrus Tristeza Virus Isolates and Its Association. **Agrociancia**, v. 42, n. 1, p. 85–93, 2008.
57. ROMÁN, M. P.; CAMBRA, M.; JUÁREZ, J.; MORENO, P.; DURAN-VILA, N.; TANAKA, F. A. O. et al. Sudden Death of Citrus in Brazil: A Graft-Transmissible Bud Union Disease. **Plant Disease**, v. 88, n. 5, p. 453–467, 2004.
58. ROY, A.; SHAO, J.; SCHNEIDER, W. L.; HARTUNG, J. S.; BRLANSKY, R. H. Population of endogenous pararetrovirus genomes in carrizo citrange. **Genome announcements**, v. 2, n. 1, p. e01063-13, 2014.
59. RUBIO, L.; AYLLÓN, M.A.; KONG, P.; FERNÁNDEZ, A.; POLEK, M. et al.

Genetic Variation of Citrus Tristeza Virus Isolates from California and Spain: Evidence for Mixed Infections and Recombination. **Virol,** v. 75, p. 8054-8062, 2001.

60. SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. Molecular Cloning. A Laboratory Manual, 2nd ed.; Publisher: Cold Spring Harbor Laboratory Press, 1989.
61. SANTOS, M. A. **Tentativa de purificação e produção de antissoro contra o vírus da morte súbita dos citros e isolamento de CSDaV em plantas hebáceas.** 2011. Dissertação de mestrado. Universidade de São Paulo - Escola superior de Agricultura "Luiz de Queiroz", 2011.
62. SEO, J.K.; OHSHIMA, K.; LEE, H.G.; SON, M.; CHOI, H.S. et al. Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. **Virology**, v. 393, p. 91-103, 2009.
63. SINGH, S.; RAJAM, M. V. Citrus biotechnology: Achievements, limitations and future directions. **Physiology and Molecular Biology of Plants**, v. 15, n. 1, p. 3–22, 2009.
64. SOUZA, A. A.; MÜLLER, G. W.; TARGON, M. L.; TAKITA, M. A.; MACHADO, M. A. Stability of the Mild Protective "PIAC" Isolate of Citrus Tristeza Virus. In: Proc. **15th Conference of IOCV**. Riverside, CA. pp: 131-135, 2002.
65. STACH-MACHADO, D. R., PERONI, L. A., DIAS, L. C. F., CAPORRINO, M. C., MÜLLER, G. W., TARGON, M. L. P. N., MACHADO, M. A. Characterization of monoclonal antibodies for identification of the severe strain of 'Capão Bonito' Citrus tristeza virus. In: Proc. **15th Conf. IOCV**, p. 165-171, 2002.
66. TUO, D.; SHEN, W.; YAN, P.; LI, X.; ZHOU, P. Rapid Construction of Stable Infectious Full-Length cDNA Clone of Papaya Leaf Distortion Mosaic Virus Using In-Fusion Cloning. **Viruses**, v. 7, n. 12, p. 6241–6250, 2015.
67. **USDA**. Citrus: World Markets and Trade, p. 1–9, 2016.
68. VILLAMOR, D.E.V.; MEKURIA, T.A.; PILLAI, S.S.; EASTWELL, K.C. High throughput sequencing identifies novel viruses in nectarine: Insights to the etiology of stem pitting disease. **Phytopathology**, v. 106, p. 519-527, 2016.
69. WALIA, J.J.; WILLEMSEN, A.; ELCI, E.; CAGLAYAN, K.; FALK, B.F.et al. Genetic Variation and Possible Mechanisms Driving the Evolution of Worldwide Fig mosaic virus Isolates. **Phytopathology**, v. 104, p. 108-114, 2013.
70. WU G. W, TANG, M., WANG, G. P., WANG, C. X., LIU, Y. et al. The epitope structure of Citrus tristeza virus coat protein mapped by recombinant proteins and

monoclonal antibodies. **Virology**, v. 448, p. 238–246, 2014.

71. YAMAMOTO, P. T., BASSANEZI, R. B., WULFF, N. A., SANTOS, M. A., SANCHES, A. L. et al. Citrus Sudden Death Is Transmitted by Graft-Inoculation and Natural Transmission Is Prevented by Individual Insect-Proof Cages. **Plant Disease**, v. 95, n. 2, p. 104–112, 2011.
72. ZAVALLO, D.; DEBAT, H. J.; CONTI, G.; MANACORDA, C. A.; RODRIGUEZ, M. C. et al. Differential mRNA accumulation upon early Arabidopsis thaliana infection with ORMV and TMV-Cg is associated with distinct endogenous small RNAs level. **PLoS ONE**, v. 10, n. 8, p. 1–24, 2015.

9. APÊNDICES E MATERIAL SUPLEMENTAR

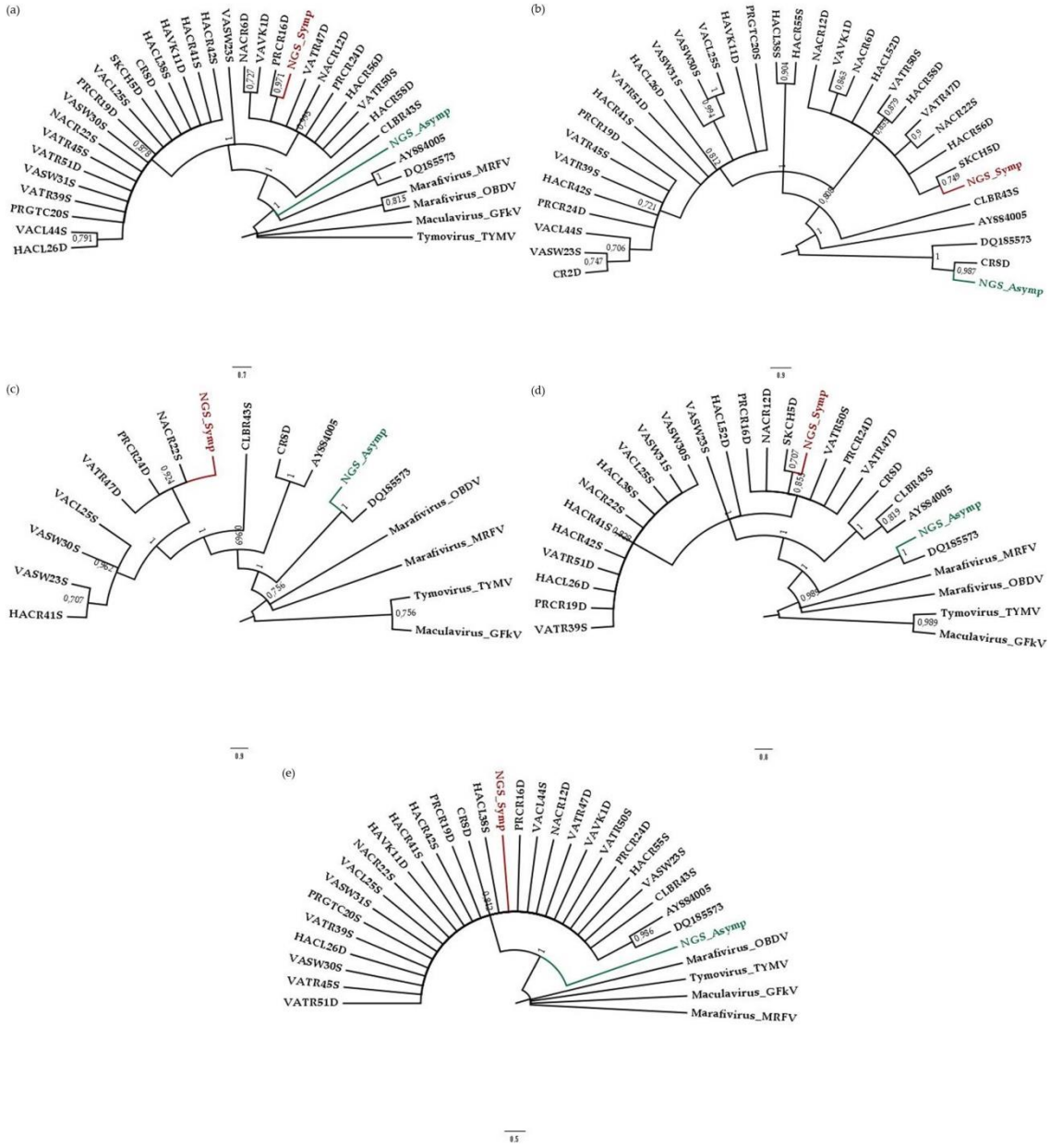

Figure S9.1. Bootstrap majority rule (70%) consensus trees reconstructed by the neighbor joining method for five genomic regions of CSDaV isolates including the consensus sequences from transcriptome sequencing of the symptomatic and asymptomatic plants by using Illumina platform. Bootstrap values are given above branches. (a): MT segment; (b): MDR segment; (c): He segment; (d): RdRP segment; (e): CP segment. CSDaV consensus sequences are differentiated by colors: From asymptomatic plants = green; from symptomatic plants = red.

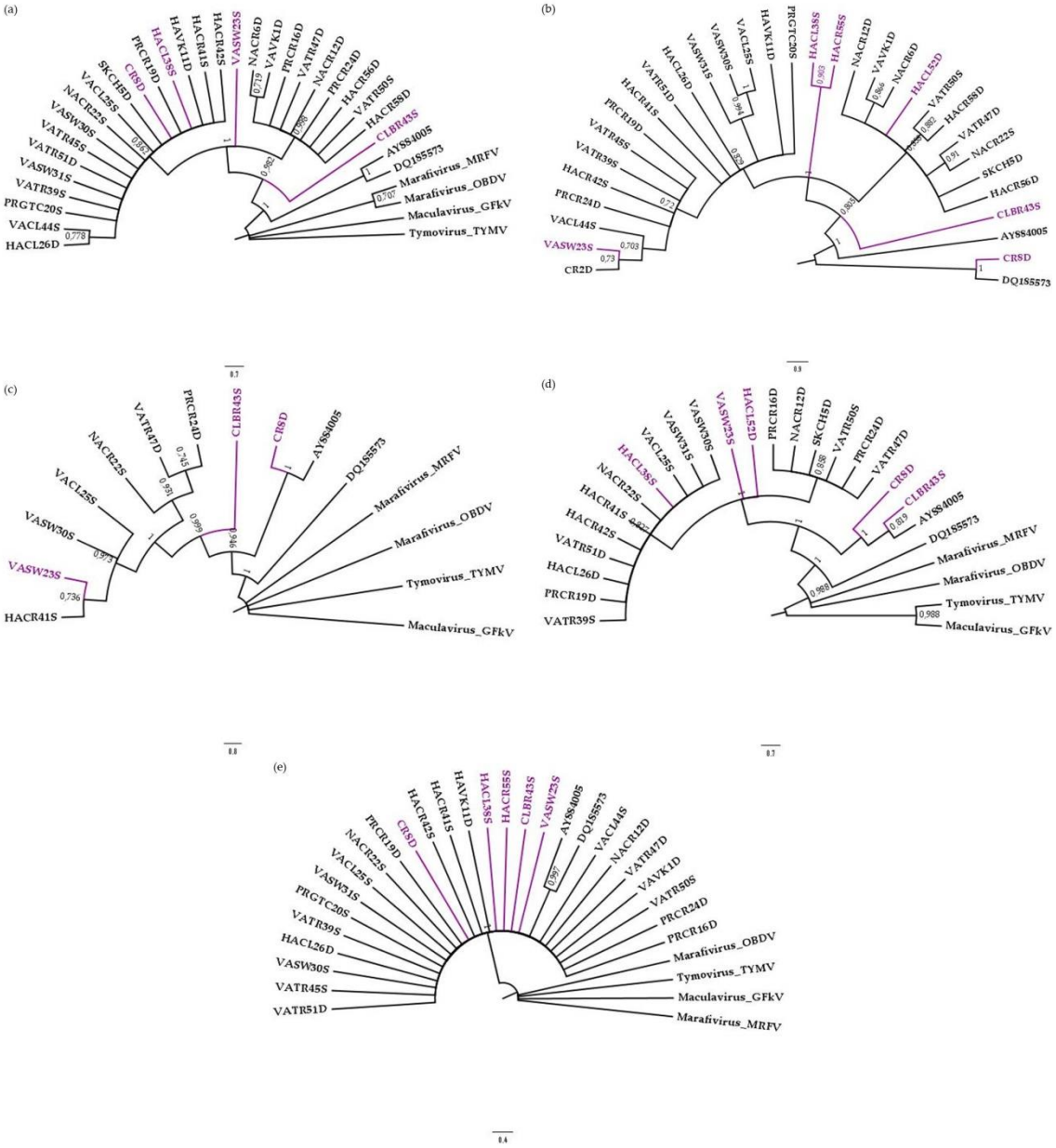

Figure S9.2. Bootstrap majority rule (70%) consensus trees reconstructed by the neighbor joining method for five genomic regions of CSDaV isolates including possible recombinant isolates, field collected and reference sequences. Bootstrap values are given above branches. (a): MT segment; (b): MDR segment; (c): He segment; (d): RdRP segment; (e): CP segment. The possible recombinant isolates are represented by purple color.

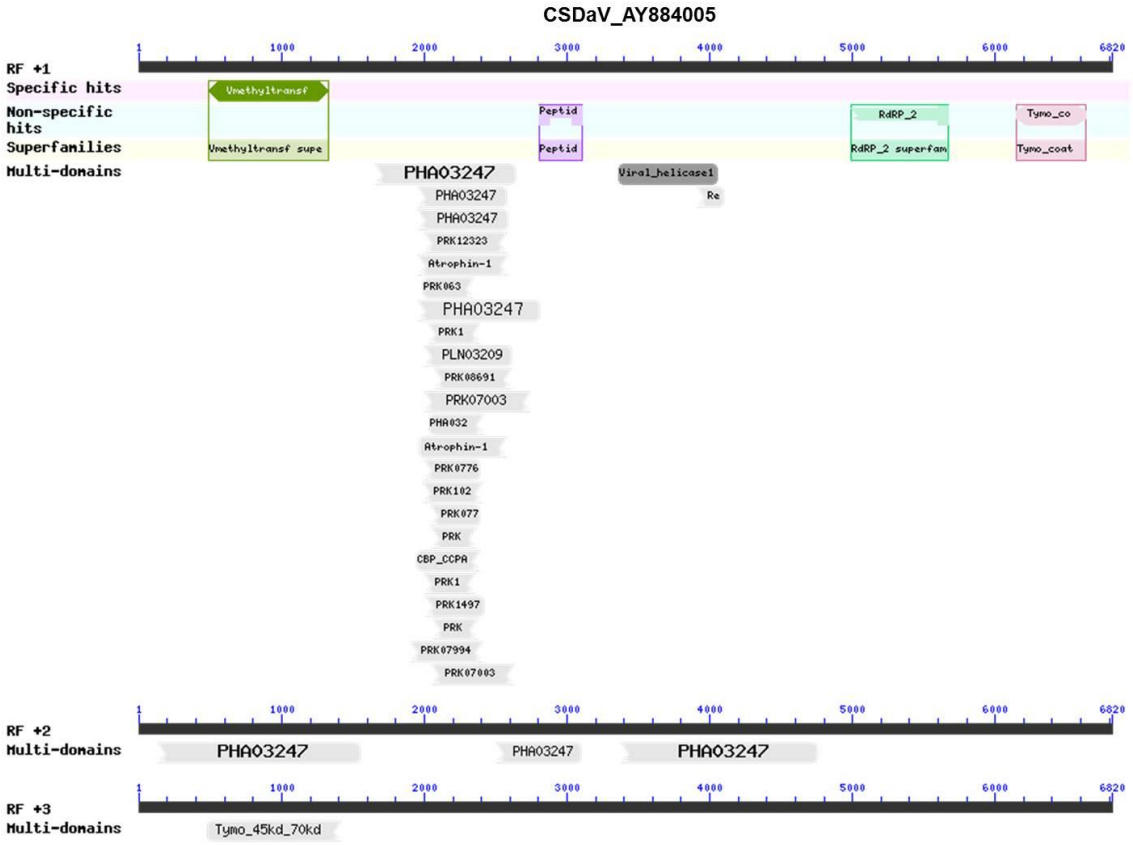

Figure S9.3. Graphical summary showing the conserved domains detected from conserved domain search using the CSDaV AY884005 reference sequence as query in the NCBI Conserved Domain Database (CDD).

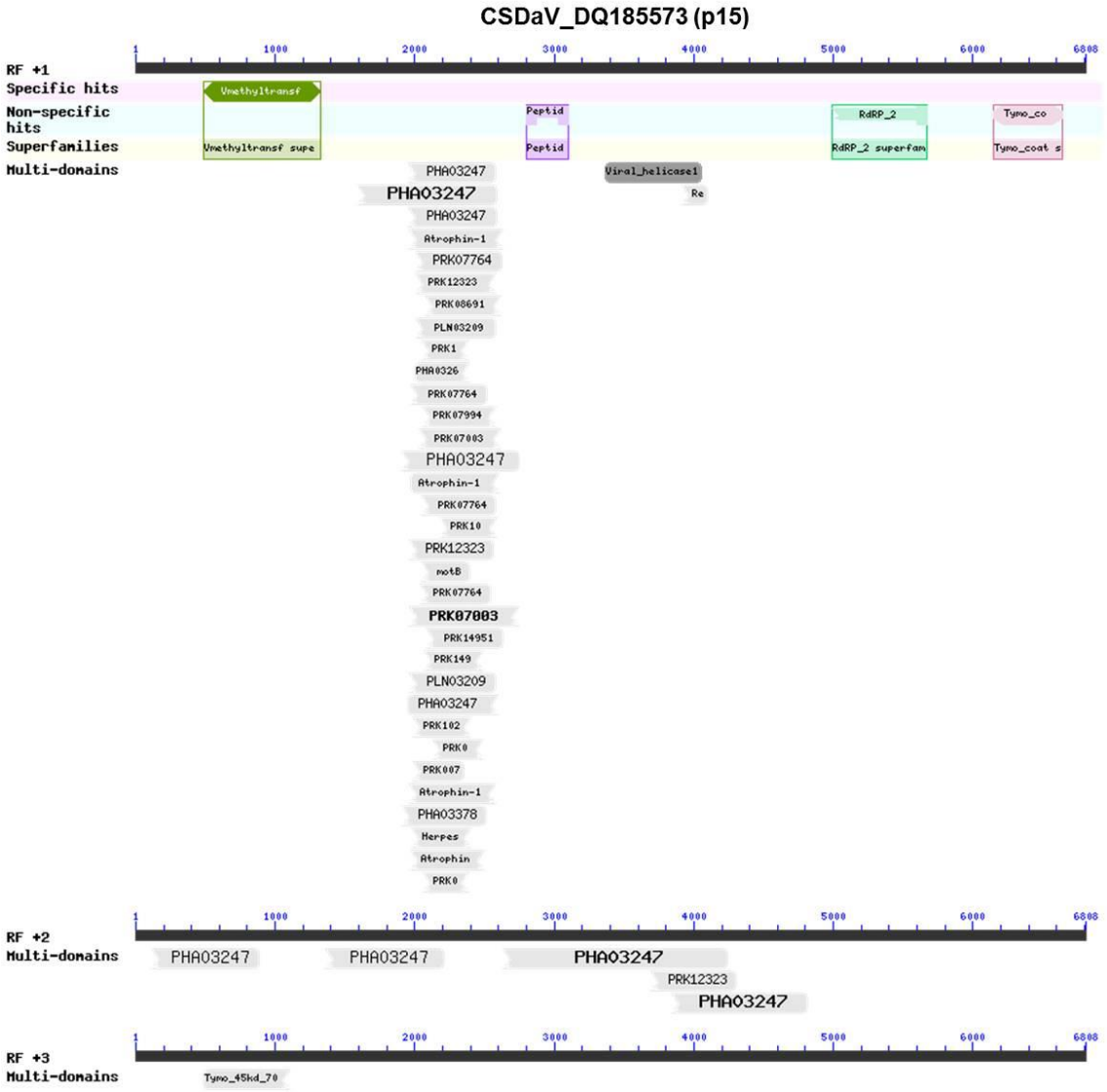

Figure S9.4. Graphical summary showing the conserved domains detected from conserved domain search using the CSDaV DQ185573 reference sequence as query in the NCBI Conserved Domain Database (CDD).

Table S9.1: Primer sequences designed based on de novo-assembled contigs to validation assays.

Reference viruses or viral contigs for primer designing	*Primer sequences (5'-3')	PCR product expected size
CTV_SPBR_01	F - CCGAAGAAGTGACACCAGTCTGTAAG R - AGAAGCCGCACCAGTAACGTACTTAG	1001 nt
CTV_SPBR_02	F - GGACTTGTAGAAGACGGCAAGAAGTC R - CCGGCGGTTTTAACAACGTGTTTAG	995 nt
CSDaV_SPBR_01/02	F - TCCAACCAGTTCCCACACATGGACAA R - GCGTACCGACCCCTTTCTTTCTTGAA	974 nt
ALPV	F - AACAAGACTTGGCCCGTGTATCTGAC R - CTACAGAGATGCCTTTGTTCATGTGCG	320 nt
CitPRV	F - TGACATCCTGCACCAGATCAAACACC R - CCATTCGGATGAGAAAATGGTGTCC	1363 nt
SsDFV1	F - CCCCAGATGGAAATTCAGCCTGATC R - TCAACATTGAATGCATGCGCATGCC	300 nt
CtgCirco-1	F - CGTATGTATCAGGCACTTTGGGAGC R - GGATATTCATCGGAGCCATAATGGC	200 nt
CtgFlavi-1	F - TAATCTGGTCTTTAGCCCTCTCTGGG R - TATGCCTCCAGTTCCCAGGTGGTACTA	1929 nt
CtgMarna-1	F - CGGACGCCATGTCAAATTTGCATTC R - CCGTTTTCAGCCTTGAAGAGAGGAGA	1300 nt
CtgUnclass-1	F - ATTCGGTCGCTTCACCACTATCGATG R - GAAGAAAGGAACCGGGAGCCATGAAA	250 nt
CtgVirga-1	F - CTGTAAACGGCGATAGGATATACAGGC R - GCTTCAGCGTAATCCTGTCAAATGGG	1936 nt
CtgVirga-2	F - AAATCTGGTGCACCCGACACGATGTT R - AAGTTGATGGTGCCTCGTGTACGAC	384 nt

*F, forward primer; R, reverse primer.

Table S9.2: Accession numbers of the reference sequences used in the phylogenetic analysis.

Name	Abbreviation	Family	Accession no.	*Additional information
Citrus tristeza virus	CTV	*Closteroviridae*	GQ454869	HA18-9 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JX266712	Taiwan-Pum/SP/T1 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525435	NZRB-M17 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525432	NZRB-G90 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JF957196	B301 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525434	NZRB-TH30 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525433	NZRB-TH28 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525431	NZRB-M12 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	DQ272579	Severe_Mexico isolate
Citrus tristeza virus	CTV	*Closteroviridae*	EU937521	T36FS2-2_Florida isolate
Citrus tristeza virus	CTV	*Closteroviridae*	AY170468	T36IC_Florida isolate
Citrus tristeza virus	CTV	*Closteroviridae*	AY340974	Qaha_Egypt isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JQ798289	A18_Thailand isolate
Citrus tristeza virus	CTV	*Closteroviridae*	DQ151548	T318A_Spain isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JQ911664	CT11A_China isolate
Citrus tristeza virus	CTV	*Closteroviridae*	AB046398	Nuaga_Japan isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JQ061137	AT-1_China isolate
Citrus tristeza virus	CTV	*Closteroviridae*	KC748392	SG29_Italy isolate
Citrus tristeza virus	CTV	*Closteroviridae*	EU937519	VT_Israel isolate
Citrus tristeza virus	CTV	*Closteroviridae*	HM573451	Kpg3_India isolate
Citrus tristeza virus	CTV	*Closteroviridae*	EU857538	NZM16_SP isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JQ911663	CT14A_China isolate
Citrus tristeza virus	CTV	*Closteroviridae*	AF260651	T30_Florida isolate
Citrus tristeza virus	CTV	*Closteroviridae*	Y18420	T385_Spain isolate
Citrus tristeza virus	CTV	*Closteroviridae*	KC748391	Bau282_Italy isolate
Citrus tristeza virus	CTV	*Closteroviridae*	AF001623	SY568_California isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JQ965169	T68-1_Florida isolate
Citrus tristeza virus	CTV	*Closteroviridae*	FJ525436	NZ-B18 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	EU076703	B165_India isolate
Citrus tristeza virus	CTV	*Closteroviridae*	JX266713	Taiwan-Pum/M/T5 isolate
Citrus tristeza virus	CTV	*Closteroviridae*	GQ454870	HA16-5_Hawaii isolate

Citrus tristeza virus	CTV	*Closteroviridae*	this study	CTV_SPBR_01 _this study
Citrus tristeza virus	CTV	*Closteroviridae*	this study	CTV_SPBR_02 _this study
Citrus sudden death-associated virus	CSDaV	*Tymoviridae*	DQ185573	P15 isolate
Citrus sudden death-associated virus	CSDaV	*Tymoviridae*	AY884005	CSDaV_1
Turnip yellow mosaic virus	TYMV	*Tymoviridae*	NC_004063	NAI
Grapevine fleck virus	GFkV	*Tymoviridae*	NC_003347	NAI
Maize rayado fino virus	MRFV	*Tymoviridae*	NC_002786	NAI
Oat blue dwarf virus	OBDV	*Tymoviridae*	U87832	NAI
Soybean chlorotic mottle virus	SoyCMV	*Caulimoviridae*	NP_068729	NAI
Peanut chlorotic streak virus	PCISV	*Caulimoviridae*	NP_042513	NAI
Figwort mosaic virus	FMV	*Caulimoviridae*	NP_619548	NAI
Dahlia mosaic virus	DMV	*Caulimoviridae*	AGT41978	NAI
Carnation etched ring virus	CERV	*Caulimoviridae*	NP_612577	NAI
Cauliflower mosaic virus	CaMV	*Caulimoviridae*	AAA46358	NAI
Strawberry vein banding virus	SVBV	*Caulimoviridae*	AKB94071	NAI
Citrus endogenous pararetrovirus	CitPRV	*Caulimoviridae*	NC_023153	NAI
Aristotelia chilensis virus 1	AcV1	*Caulimoviridae*	AHN13810	NAI
Petunia vein clearing virus	PVCV	*Caulimoviridae*	AAK68664	NAI
Cassava vein mosaic virus	CsVMV	*Caulimoviridae*	AAA79873	NAI
Cacao swollen shoot virus	CSSV	*Caulimoviridae*	CAE81279	NAI
Commelina yellow mottle virus	CoYMV	*Caulimoviridae*	CAA37110	NAI
Sugarcane bacilliform MO virus	SCBV	*Caulimoviridae*	YP_595725	NAI
Rice tungro bacilliform virus	RTBV	*Caulimoviridae*	AAL55651	NAI
Saccharomyces cerevisiae (TY3-2)		*Saccharomycetaceae*	CAA86713	NAI
Pestivirus reindeer-1 V60-Krefeld	V60-Krefeld	*Flaviviridae*	AAF02524	NAI
Border disease virus	BDV	*Flaviviridae*	NP_777541	NAI
Bovine viral diarrhea	BVDV-1	*Flaviviridae*	AKQ44350	NAI

virus 1				
Pronghorn antelope pestivirus	Pronghorn pestivirus	*Flaviviridae*	YP_009026415	NAI
Porcine pestivirus	Pestivirus Bungowannah	*Flaviviridae*	YP_008992092	NAI
Xinzhou spider virus 2	XZSV2	*Flaviviridae*	YP_009179222	NAI
Wuhan centipede virus	WHCeV	*Flaviviridae*	YP_009254745	NAI
Wenling shark virus	WLSV	*Flaviviridae*	YP_009179227	NAI
Wuhan flea virus	WHFV	*Flaviviridae*	YP_009179404	NAI
Wuhan aphid virus 2	WHAV2	*Flaviviridae*	YP_009179379	NAI
Wuhan aphid virus 1	WHAV1	*Flaviviridae*	YP_009179389	NAI
Shuangao insect virus 7	SAIV7	*Flaviviridae*	YP_009179402	NAI
Wuhan cricket virus	WHCV	*Flaviviridae*	YP_009179400	NAI
Bole tick virus 4	BLTV4	*Flaviviridae*	YP_009179221	NAI
Diaphorina citri flavi-like virus	DcFLV	*Flaviviridae*	YP_009259672	NAI
Gamboa mosquito virus	GMV	*Flaviviridae*	YP_009179224	NAI
Soybean cyst nematode virus 5	SbCNV-5	*Flaviviridae*	YP_009028573	NAI
Nakiwogo virus		*Flaviviridae*	YP_009268608	NAI
Tacheng tick virus 8	TCTV8	*Flaviviridae*	YP_009179217	NAI
Macrosiphum euphorbiae virus 1	MeV-1	*Flaviviridae*	YP_009175071	NAI
Shuangao lacewing virus 2	SALV2	*Flaviviridae*	YP_009179223	NAI
Xingshan cricket virus	XSCV	*Flaviviridae*	YP_009179220	NAI
Gentian Kobu-sho-associated virus	GKaV	*Flaviviridae*	YP_007438864	NAI
Beihai barnacle viurs 1	BHBV1	*Flaviviridae*	YP_009179226	NAI
Shayang spider virus 4	SYSV4	*Flaviviridae*	YP_009179219	NAI
West Nile virus	WNV	*Flaviviridae*	AAV54504	NAI
Jingmen tick virus	JMTV	*putative Flaviviridae*	YP_009030000	NAI
Citrus jingmen-like virus	CJLV	*Flaviviridae*	this study	CtgFlavi-1 contig
Chinese wheat mosaic virus	CWMV	*Virgaviridae*	BAP90385	NAI
Oat golden stripe virus	OGSV	*Virgaviridae*	CAB57882	NAI
Sorghum chlorotic spot virus	SrCSV	*Virgaviridae*	NP_659019	NAI

Beet soil-borne virus	BSBV	*Virgaviridae*	ACS14040	NAI
Beet virus Q	BVQ	*Virgaviridae*	NP_612605	NAI
Broad bean necrosis virus	BBNV	*Virgaviridae*	NP_740760	NAI
Barley stripe mosaic virus	BSMV	*Virgaviridae*	AAA79146	NAI
Poa semilatent virus	PSLV	*Virgaviridae*	CAA86473	NAI
Lychnis ringspot virus	LRSV	*Virgaviridae*	CAA86599	NAI
Indian peanut clump virus	IPCV	*Virgaviridae*	NP_835282	NAI
Peanut clump virus	PCV	*Virgaviridae*	NP_620047	NAI
Pea early-browning virus	PEBV	*Virgaviridae*	NP_049325	NAI
Pepper ringspot virus	PepRSV	*Virgaviridae*	NP_620033	NAI
Tobacco rattle virus	TRV	*Virgaviridae*	ACX54058	NAI
Ribgrass mosaic virus	RMV	*Virgaviridae*	ACV13194	NAI
Turnip vein-clearing virus	TVCV	*Virgaviridae*	NP_046151	NAI
Wasabi mottle virus	WMoV	*Virgaviridae*	AHW98777	NAI
Youcai mosaic virus	YMoV	*Virgaviridae*	BAN15047	NAI
Odontoglossum ringspot virus	ORSV	*Virgaviridae*	AAB49498	NAI
Streptocarpus flower break virus	SFBV	*Virgaviridae*	YP_762617	NAI
Tobacco mild green mosaic virus	TMGMV	*Virgaviridae*	NP_062913	NAI
Rehmannia mosaic virus	ReMV	*Virgaviridae*	ALP75636	NAI
Tomato mosaic virus	ToMV	*Virgaviridae*	CAD10425	NAI
Brugmansia mild mottle virus	BruMMV	*Virgaviridae*	YP_001974323	NAI
Obuda pepper virus	ObPV	*Virgaviridae*	NP_620841	NAI
Paprika mild mottle virus	PaMMV	*Virgaviridae*	NP_671718	NAI
Hibiscus latent Fort Pierce virus	HLFPV	*Virgaviridae*	BAP76306	NAI
Hibiscus latent Singapore virus	HLSV	*Virgaviridae*	YP_719997	NAI
Kyuri green mottle mosaic virus	KGMMV	*Virgaviridae*	NP_619684	NAI
Zucchini green mottle mosaic virus	ZGMMV	*Virgaviridae*	NP_624336	NAI
Cucumber fruit mottle mosaic virus	CFMMV	*Virgaviridae*	AEV40683	NAI
Cucumber green	CGMMV	*Virgaviridae*	BAA18895	NAI

mottle mosaic virus				
Sunn-hemp mosaic virus	SHMV	*Virgaviridae*	P89202	NAI
Frangipani mosaic virus	FrMV	*Virgaviridae*	AEW67306	NAI
Citrus virga-like virus	CVLV	*Virgaviridae*	this study	CtgVirga-1 contig
Citrus virga-like virus	CVLV	*Virgaviridae*	this study	CtgVirga-2 contig

*NAI, no additional information.

Table S9.3: Query coverage and maximum amino acid identity of the viruses found in this work obtained from the BLASTx analysis against to the viral database using assembled contigs as query sequences. Putative virus encoded protein and E value are shown.

Closely related species	Contig lenght used as query (nt)	Query % coverage	Maximum % identity	E value	Putative virus encoded protein
Citrus tristeza virus	3180	99	92	0.0	Polyprotein replicase
Citrus sudden death-associated virus	6109	97	98	0.0	Polyprotein
Marine RNA virus SF-2	1400	50	22	0.019	Coat protein
Rice stripe necrosis virus	250	79	39	1,00E-06	Replication protein
Rhizoctonia solani negative-stranded virus 4	126	97	56	3,00E-10	RdRP
Norovirus cat	144	64	48	0.026	RdRP
Dioscorea bacilliform AL virus	140	24	43	3.3	RT-like
Po-Circo-like virus 51	305	54	43	0.046	Replicase
Aphid lethal paralysis virus	343	99	97	3,00E-70	Nonstructural protein
Nakiwogo virus	2512	21	27	0.001	Nonstructural protein NS3
Sclerotinia sclerotiorum deltaflexivirus 1	329	88	62	3,00E-34	Polyprotein
Soybean leaf-associated mycoflexivirus 1	196	85	37	0.051	RdRP
Deformed wing virus	173	53	52	0.010	Polyprotein
Nilaparvata lugens honeydew virus-3	118	96	47	7,00E-04	RT-like
Raphanus sativus cryptic virus 1	183	67	41	0.001	Coat protein
Passerivirus A1	134	85	46	0.42	Polyprotein
Chilli veinal mottle virus	109	96	48	0.016	Polyprotein
Citrus endogenous pararetrovirus	3339	53	72	0.0	Polyprotein
Lettuce necrotic leaf curl virus	141	80	42	0.11	Polyprotein
Rice tungro spherical virus	189	95	43	0.028	RdRP
Fusarium graminearum deltaflexivirus 1	262	98	71	4,00E-36	Replication protein
Boutonnet virus	423	60	36	2,00E-06	Polyprotein
Bufivirus UC1	203	93	43	5,00E-10	RT-like
Fisavirus 1	101	95	56	0.020	Polyprotein
Twyford virus	186	72	44	0.054	Helicase
Beet virus Q	4097	18	33	6,00E-25	Replication

					protein
Chinese wheat mosaic virus	2626	52	28	9,00E-29	Replicase readthrough

Table S9.4. Description of domains detected from conserved domain search using the CSDaV AY884005 reference sequence as query. The interval and E-value of each identified domain are shown.

Name	Accession	Description	Interval	E-value
Vmethyltransf	pfam01660	Viral methyltransferase; This RNA methyltransferase domain is found in a wide range of ssRNA ...	484-1329	1.59e-77
Peptidase_C21	pfam05381	Tymovirus endopeptidase; Corresponds to Merops family C21. The best-studied plant alpha-like ...	2803-3102	7.52e-20
RdRP_2	pfam00978	RNA dependent RNA polymerase; This family may represent an RNA dependent RNA polymerase. The ...	4990-5670	4.65e-11
Tymo_coat	pfam00983	Tymovirus coat protein;	6148-6639	6.90e-11
Viral_helicase1	pfam01443	Viral (Superfamily 1) RNA helicase; Helicase activity for this family has been demonstrated ...	3361-4053	3.24e-50
PHA03247	PHA03247	large tegument protein UL36; Provisional	1651-2640	2.37e-08
PHA03247	PHA03247	large tegument protein UL36; Provisional	1960-2577	9.15e-08
PHA03247	PHA03247	large tegument protein UL36; Provisional	1954-2586	1.21e-06
PRK12323	PRK12323	DNA polymerase III subunits gamma and tau; Provisional	2002-2577	2.39e-06
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1957-2583	3.61e-06
PRK06347	PRK06347	autolysin; Reviewed	1987-2349	8.60e-06
PHA03247	PHA03247	large tegument protein UL36; Provisional	1954-2805	1.56e-05
PRK14971	PRK14971	DNA polymerase III subunits gamma and tau; Provisional	2029-2385	2.14e-05
PLN03209	PLN03209	translocon at the inner envelope of chloroplast subunit 62; Provisional	2002-2613	9.59e-05
PRK08691	PRK08691	DNA polymerase III subunits gamma and tau; Validated	2074-2613	1.17e-04
PRK07003	PRK07003	DNA polymerase III subunits gamma and tau; Validated	2002-2745	1.30e-04
PHA03269	PHA03269	envelope glycoprotein C; Provisional	2032-2415	1.31e-04
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1969-2574	1.42e-04
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2002-2385	2.14e-04
PRK10263	PRK10263	DNA translocase FtsK; Provisional	1990-2406	3.64e-04
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2041-2385	1.26e-03
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2050-2361	2.32e-03
CBP_CCPA	pfam17040	Cellulose-complementing protein A;	1945-2391	2.65e-03

		CBP_CCPA is a family of bacterial cellulose-complementing ...		
PRK14959	PRK14959	DNA polymerase III subunits gamma and tau; Provisional	2002-2358	3.48e-03
PRK14971	PRK14971	DNA polymerase III subunits gamma and tau; Provisional	2005-2415	4.20e-03
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2059-2370	4.31e-03
PRK07994	PRK07994	DNA polymerase III subunits gamma and tau; Validated	1900-2409	4.31e-03
PRK07003	PRK07003	DNA polymerase III subunits gamma and tau; Validated	2056-2628	4.41e-03
RecD	COG0507	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I [Replication, ...	3910-4098	7.98e-03
PHA03247	PHA03247	large tegument protein UL36; Provisional	134-1546	2.47e-05
PHA03247	PHA03247	large tegument protein UL36; Provisional	3377-4753	3.42e-05
PHA03247	PHA03247	large tegument protein UL36; Provisional	2504-3097	6.19e-04
Tymo_45kd_70kd	pfam03251	Tymovirus 45/70Kd protein; Tymoviruses are single stranded RNA viruses. This family includes a ...	483-1418	4.49e-12

Table S9.5. Description of domains detected from conserved domain search using the CSDaV DQ185573 reference sequence as query. The interval and E-value of each identified domain are shown.

Name	Accession	Description	Interval	E-value
Vmethyltransf	pfam01660	Viral methyltransferase; This RNA methyltransferase domain is found in a wide range of ssRNA ...	484-1329	1.16e-76
Peptidase_C21	pfam05381	Tymovirus endopeptidase; Corresponds to Merops family C21. The best-studied plant alpha-like ...	2803-3102	3.70e-19
RdRP_2	pfam00978	RNA dependent RNA polymerase; This family may represent an RNA dependent RNA polymerase. The ...	4993-5673	4.40e-11
Tymo_coat	pfam00983	Tymovirus coat protein;	6151-6642	6.88e-11
Viral_helicase1	pfam01443	Viral (Superfamily 1) RNA helicase; Helicase activity for this family has been demonstrated ...	3361-4053	2.10e-49
PHA03247	PHA03247	large tegument protein UL36; Provisional	1951-2577	3.29e-09
PHA03247	PHA03247	large tegument protein UL36; Provisional	1594-2592	1.03e-07
PHA03247	PHA03247	large tegument protein UL36; Provisional	1954-2586	2.15e-07
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1957-2628	4.51e-07
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2002-2622	8.40e-07
PRK12323	PRK12323	DNA polymerase III subunits gamma and tau; Provisional	2002-2577	2.01e-06
PRK08691	PRK08691	DNA polymerase III subunits gamma and tau; Validated	2074-2613	1.61e-05
PLN03209	PLN03209	translocon at the inner envelope of chloroplast subunit 62; Provisional	2002-2577	2.41e-05
PRK14971	PRK14971	DNA polymerase III subunits gamma and tau; Provisional	2047-2385	2.55e-05
PHA03269	PHA03269	envelope glycoprotein C; Provisional	2008-2397	4.43e-05
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	1969-2514	5.84e-05
PRK07994	PRK07994	DNA polymerase III subunits gamma and tau; Validated	2056-2580	7.54e-05
PRK07003	PRK07003	DNA polymerase III subunits gamma and tau; Validated	2035-2613	9.69e-05
PHA03247	PHA03247	large tegument protein UL36; Provisional	1906-2745	1.41e-04
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1969-2604	1.89e-04
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2050-2577	4.09e-04

PRK10263	PRK10263	DNA translocase FtsK; Provisional	2185-2589	5.53e-04
PRK12323	PRK12323	DNA polymerase III subunits gamma and tau; Provisional	1957-2559	6.46e-04
motB	PRK12799	flagellar motor protein MotB; Reviewed	2050-2391	7.16e-04
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2041-2547	7.39e-04
PRK07003	PRK07003	DNA polymerase III subunits gamma and tau; Validated	1957-2745	1.00e-03
RecD	COG0507	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I [Replication, ...	3910-4098	1.81e-03
PRK14951	PRK14951	DNA polymerase III subunits gamma and tau; Provisional	2083-2634	2.52e-03
PRK14959	PRK14959	DNA polymerase III subunits gamma and tau; Provisional	2071-2490	2.79e-03
PLN03209	PLN03209	translocon at the inner envelope of chloroplast subunit 62; Provisional	1957-2583	3.79e-03
PHA03247	PHA03247	large tegument protein UL36; Provisional	1951-2595	4.08e-03
PRK10263	PRK10263	DNA translocase FtsK; Provisional	1990-2406	4.31e-03
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2125-2490	4.88e-03
PRK00708	PRK00708	sec-independent translocase; Provisional	1990-2349	5.09e-03
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1960-2583	5.98e-03
PHA03378	PHA03378	EBNA-3B; Provisional	1906-2505	7.08e-03
Herpes_BLLF1	pfam05109	Herpes virus major outer envelope glycoprotein (BLLF1); This family consists of the BLLF1 ...	1975-2400	7.60e-03
Atrophin-1	pfam03154	Atrophin-1 family; Atrophin-1 is the protein product of the dentatorubral-pallidoluysian ...	1969-2499	8.38e-03
PRK07764	PRK07764	DNA polymerase III subunits gamma and tau; Validated	2062-2388	9.60e-03
PHA03247	PHA03247	large tegument protein UL36; Provisional	2630-4237	5.22e-04
PRK12323	PRK12323	DNA polymerase III subunits gamma and tau; Provisional	3695-4297	9.15e-04
PHA03247	PHA03247	large tegument protein UL36; Provisional	3839-4813	1.17e-03
PHA03247	PHA03247	large tegument protein UL36; Provisional	116-886	4.63e-03
PHA03247	PHA03247	large tegument protein UL36; Provisional	1358-2206	7.60e-03
Tymo_45kd_70kd	pfam03251	Tymovirus 45/70Kd protein; Tymoviruses are single stranded RNA viruses. This family includes a ...	483-1106	3.79e-07