

SECUENCIA DEL GENOMA DE UN AISLAMIENTO DEL VIRUS DE LA TRISTEZA DE LOS CÍTRICOS

GENOME SEQUENCE OF A CITRUS TRISTEZA VIRUS ISOLATE

Jesús Di Carlo Quiroz Velásquez¹, Ma. de los Ángeles Peña del Río², Ma. Antonia Cruz Hernández¹,
Susana Fernández Dávila¹, Maurilio González Paz y Alberto Mendoza Herrera*

¹ Laboratorio de Interacción Planta-Microorganismo, Biotecnología Vegetal, Centro de Biotecnología Genómica, Instituto Politécnico Nacional. Blvd. del Maestro/Elías Piña, s/n Col. Narciso Mendoza, Cd. Reynosa, Tam. 77810. Tel: (52-899) 924-36 27, Fax: (52-899)9 25-16-56. ² Campo experimental General Terán, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. AP 67400. General Terán NL.

*Autor para correspondencia (amendozah@ipn.mx)

RESUMEN

El genoma de un aislamiento severo del virus de la tristeza de los cítricos (VTC) de México (CBG-VTC1) se secuenció en su totalidad. Los 19 300 nucleótidos (nt) del genoma se dividen en 12 marcos de lectura abierta (ORF's) que codifican para 15 proteínas y dos regiones no traducibles (5' y 3'-UTR). El primer marco de lectura abierta inició en el nucleótido 108 con un tamaño de 9587 nt, y codificó para tres proteínas traslapadas, una poliproteína de 357 kDa que contiene dos proteasas contiguas (P-PRO), una metiltransferasa (MTR) y una helicasa (HEL). El segundo ORF traslapó en los últimos 55 nt de las proteasas; éste codifica para una ARN polimerasa ARN-dependiente (RdRp) con un peso molecular de 47 kDa cuya función es la replicación. Los ORF's del 2 al 10 codificaron 10 productos proteicos con un rango que va de 6 a 65 kDa. El genoma CBG-VTC1 presentó sintenia con genomas ya reportados, y difirió solamente en 2 a 74 nt. La región 5'-UTR del CBG-VTC1 mostró 55 % de identidad con T30, 57.9 % con SY568, 100 % con T36 y solamente 58.9 % con VT, mientras que la región 3'-UTR tuvo 96 % de identidad en todos los aislamientos. La identidad aminoacídica promedio fue de 87.0 % con un aislamiento débil (T30) y de 85.9 %, 93.7 % y 85.13 % con las razas severas SY568, T36 y VT, respectivamente.

Palabras clave: *Clusterovirus*, virus de la tristeza de los cítricos, genoma, secuenciación.

SUMMARY

The genome of a citrus tristeza virus severe strain from México was completely sequenced. The 19 300 nt genome has 12 open reading frames (ORFs) encoding for 15 proteins and two UTRs (5' and 3' -UTR). The first ORF starts at the nucleotide 108 position and comprises 9587 nt, and it encodes for three overlapping proteins, a large polyprotein of 357 kDa that contains two papain-like proteases (P-PRO), a methyltransferase (MT) and a helicase (HEL). The second ORF overlaps on the last 55 nt of the papain-like proteases and encodes for a 47 kDa putative RNA-dependent RNA polymerase (RdRp) with a replication function. The following ORFs encodes for 10 proteins from 6 to 65 kDa. Both genomes had 55 % identity at 5'UTR, 57.9 % at T30, 100 % at SY568, 58.9 % at T36 and VT, and 96 % identity at 3' UTR for all isolates. Average of putative amino acid sequences identities showed of 87.0 % with a mild isolate (T30) and 85.9 %, 93.7 % and 85.13 % with severe isolates T30, SY568, T36 and VT, respectively.

Key words: *Clusterovirus*, citrus tristeza virus, genome, sequence.

INTRODUCCIÓN

El virus de la tristeza de los cítricos (VTC) es el agente causal de una de las enfermedades económicamente más importantes que afectan a los cítricos. Los aislamientos del VTC difieren ampliamente en sus características biológicas, tales como transmisibilidad por áfidos y tipo de síntomas causados en varias especies de hospederos. Estos síntomas son de tres tipos: declinamiento y muerte de plantas injertadas en naranjo agrio (*Citrus aurantium* L.); achaparramiento y amarillamiento de naranjo agrio, toronja (*Citrus paradisi* Macf.) o limón [*Citrus limon* (L.) Burm. F.] y picado de tallo de diferentes cultivos de cítricos sobre su propio patrón o injertados en algún otro patrón. En las últimas dos décadas, los estudios se han dirigido hacia el desarrollo de nuevas técnicas para una rápida diferenciación de aislamientos del VTC e identificación mediante marcadores moleculares. Se han descrito variaciones en las reacciones de serología, mapas de los péptidos que conforman la cápside, patrones de ARN de cadena doble, hibridación con sondas de ADN, perfiles en los fragmentos de restricción (Mendoza *et al.*, 2003) y polimorfismos conformacionales de cadena sencilla (SSCP) con el fin de diferenciar aislamientos del VTC (Rubio *et al.*, 2001; Manjunath *et al.*, 1993).

El análisis de la secuencia nucleotídica es el procedimiento más seguro para la diferenciación del VTC y para estimar la variación molecular o genética. Actualmente se han reportado las secuencias completas del genoma de cinco aislamientos del VTC, como son T36 y T30 de Florida, EE. UU., VT de Israel, SY568 de California, EE. UU. y T385 de España (Rubio *et al.*, 2001), y ninguna en México.

Los viriones del VTC son filamentosos y flexibles con un tamaño de 2000 x 11 nm de largo (Bar-Joseph *et al.*,

1989), de una sola cadena de ARN en sentido positivo (5' a 3'), con una longitud aproximada de 19.3 kb y dos proteínas de la cápside de 25 y 27 kDa, que abarcan 95 y 5 % de la longitud del virión, respectivamente (Sambade *et al.*, 2003). El genoma del VTC está organizado en 12 marcos de lectura abierta (ORFs), que codifican por lo menos 17 productos proteicos y dos regiones no traducibles (UTRs) de aproximadamente 107 y 273 nucleótidos de la región 5' y 3' terminal, respectivamente (Karesev y Hilf, 1997). La secuenciación del ARN genómico (ARNg) de los aislamientos T36 y T30 de Florida, VT de Israel, T385 de España, SY568 de California y NuagA de Japón, muestran que están organizadas en 12 ORFs y regiones no traducidas, tanto en 5' y 3'-terminal (Sambade *et al.*, 2003).

Los ORFs del extremo 5'-proximal codifican una importante poliproteína que contiene dominios característicos de dos proteasas contiguas (P-PRO), una metiltransferasa (MT), una helicasa (HEL) y una ARN polimerasa ARN dependiente de (RdRp). Los ORF's 2 al 11 del extremo 3' codifican para potenciales productos proteicos de 33 kDa (p33), 6 kDa (p6), y una proteína celular homóloga a las de choque térmico (HSP70), 61 kDa (p61), una copia divergente de la proteína de la cápside (dCP), proteína de la cápside (CP), 18 kDa (p18), 13 kDa (p13), 20 kDa (p20) y 23 kDa (p23). El extremo 3'-terminal de 9288 nucleótidos del genoma del VTC demuestra que abarca por lo menos 11 ORFs (Karesev *et al.*, 1995; Pappu *et al.*, 1994). Según el análisis de la secuencia del gen de la ARN polimerasa dependiente de ARN (RdRp), el VTC demostró pertenecer a un linaje diferente de *Closterovirus* dentro de la superfamilia alfa de virus de ARN de cadena positiva (Karesev *et al.*, 1995; Dolja *et al.*, 1994).

En este trabajo se reporta la secuencia completa del genoma viral de un aislamiento severo del Estado de Tamaulipas, México (CBG-VTC1) y además la comparación de las secuencias de los aislamientos T36, VT, T30 y SY568.

MATERIALES Y MÉTODOS

Extracción de ARN

Se tomaron muestras de tejido a partir de corteza de varetas y hojas de árboles de limón mexicano infectado con una cepa severa del VTC, originaria de Cd. Güemez, Tamps.; la cepa había sido mantenida en condiciones de invernadero en el Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) Campo Gral. Terán, N. L. El método utilizado fue el de extracción cruda de ácido ribonucleico total basado en el TRIzol (Gibco, BRL). Éste se aplicó en 100 mg de tejido vegetal con 1 mL de reactivo TRIzol. Una vez que se extrajo el ARN,

se observó en un gel de agarosa a 1 %, conforme a la técnica descrita por Goda y Minton (1995).

Amplificación del genoma por RT-PCR

Los reactivos que se utilizaron para llevar a cabo las reacciones de transcripción reversa (RT) y reacción en cadena de la polimerasa (PCR), fueron obtenidas de las siguientes casas comerciales: las enzimas M-MLV, desoxinucleósidos trifosfatados (dNTPS) procedieron de Promega Corporation (Madison, WI, USA). La *taq* DNA polimerasa de Altaenzymes (Alberta, Canadá). Los oligonucleótidos (primers) fueron sintetizados en SIGMA GENOSYS (Woodlands, TX, USA). Para la clonación de los productos de PCR, se utilizó el estuche comercial "TA Cloning Kit" de la compañía Invitrogen (Carlsbad, CA, USA).

Para todas las reacciones de transcripción reversa se utilizaron 2 µg del extracto de ARN, 0.5 µM de los iniciadores antisentido (Cuadro 1), cada uno en reacciones separadas para un volumen final de 15 µL con agua ultrapura grado MilliQ estéril. Esta mezcla se incubó a 70 °C por 5 min y después se colocó en hielo. Posteriormente, se adicionó amortiguador de la transcriptasa reversa MMLV (Moloney Murine Leukemia Virus) (1X), dNTPs (0.5 mM), RNAsin (15.6 U) y 200 U de la enzima transcriptasa reversa MMLV, y se aforó a un volumen final de 25 µL. Finalmente se incubó a 42°C durante 50 min. Una vez obtenido el cADN de cada una de las regiones que integraban el genoma completo, se procedió a amplificar en forma independiente cada clona, por medio de la PCR. Las reacciones de PCR se realizaron en un volumen final de 25 µL, utilizando 2.5 µL de cADN procedente de la RT, amortiguador (1X), cloruro de magnesio (1.5 mM), dNTP's (0.2 mM), iniciadores (0.5 µM) y de la enzima *Taq* ADN polimerasa (1.25 U), y el volumen final se completo con agua MQ estéril. El programa de amplificación que se utilizó consistió de 1 ciclo de 5 min. a 94 °C y 30 ciclos de 30 s a 94 °C, 30 s a 50 °C y 1 min a 72 °C. Finalmente, se dio un ciclo de extensión final de 7 min a 72 °C. Las reacciones de RT-PCR se llevaron a cabo en un termociclador GeneAmp modelo 9700 de la compañía Perkin Elmer (Norwalk, Connecticut, USA). Los productos amplificados se resolvieron por electroforesis en geles de agarosa a 1 % teñidos con bromuro de etidio (0.002 mg mL⁻¹) e irradiados con luz UV.

Clonación, transformación y secuenciación

La proporción óptima inserto-vector (pCR 2.1) para la ligación fue de 3:1, con adición de 1 µL de amortiguador de ligación (10X), 2 µL del vector pCR 2.1 (25 ng µL⁻¹), 1 µL de T4 DNA ligasa (4 unidades) en un volumen final

de 10 µL. La reacción se incubó a una temperatura de 14 °C durante toda la noche.

La transformación de células *E coli* se efectuó agregando 5 µL de cada producto de ligación, a 50 µL de células

TOP10F' ultracompetentes e incubando en hielo durante 30 min. Después se dio un choque térmico a 42 °C por 30 s, e inmediatamente la muestra se colocó en hielo por 2 min.

Cuadro 1. Oligonucleótidos empleados para la amplificación por PCR de la totalidad del genoma CBG-CTV1 (Yang *et al.*, 1999).

Iniciador	Secuencias 5'-3'	Fuente de la secuencia	Tm (°C)	Clonas obtenidas	Pb
F214	ATCGATAAACTTAAAGCATCATT	Sekiya <i>et al.</i> ,1991	57.4	1	1685
R215	CCGAGTCTATGTTAGCTAGACGTC	pTJ555	62		
CTVF4	CGTCGCTTTGTTTAGCGCG	SY568(pTEM24)	68.8	2	2028
CTV3'	CTATGGGGGGCCAACATAGGTCCAT05	T36	73.3		
F1700	GCGAGTACTCAAAGAAGTACCCGAGC	T36	69.5	3	2334
P61/27R1	CCCAAATCGCGTAGGTTA	SY568(Ptem24)	61		
p3444	GTCGGTAGATGCGTTAAT	SY568 (pZY23A)	55	4	2067
PCLa	AGTCGAATTTAGCCAACCTCGTAAC	SY568(pZY19A)	63.2		
P65,5'	GAAGTGCTTCCCCTTACGGCACTC	T36	71.4	5	1732
R1920	GCGCATTCAAAGACTCTGCCTAC	SY568(pZY23A)	67.9		
p3'pol	CGCTATTCATTGTGTGCGATCGAA	T36	71.5	6	1601
p14.3'	CCAGAAGCCGCTAAATATCCGCA	SY568(pZY18A)	71.9		
pol5'	TGAGGTCGTCGTTTGGTTCTTCTGT	T36	70.6	7	1615
p16.3'	ACGTTGGCTCTCAGCGAGAATTGA	SY568(pZY15A)	71.8		
F7920	ACTTCACGGGAMKTTGRG	T36,BYV &LIYV	58.4	8	1541
R9430	GGACGCTTTTCTTGCGGAATTGC	SY568(pZY16A)	75.1		
F20	CAAATCACCCGTACCCTCCGGA	T36 & VT	73.1	9	1543
R1550	TGATGGTTAGGAAYAGATC	T36 & VT	51.1		
R3040	TCTCACCCACCCATCTTATG	T36 & VT	51.1	10	1542
F1490	CACTTTTGGCAGTGTATTTCA	SY568 (pZY5A)	60.2		
F3040	CATAAGATGGGTGGTGAAGA	SY568(pZY30A)	59.9	11	1028
R4050	GGTAAACAACACTACGAGTCTTA	T36 & VT	52.7		
F4050	TAAGACTCGTAGTTGTTTACC	SY568(pZY40A)	52.7	12	3956
R7990	TCATCGTCYAAATCCAT	SY568(pZY80A)	50.3		
FCTV1	TTCCGTTYTCWGTTTCGCGC	T36, VT & T30	60	13	1000
RCTV1	GTGAAGACCYCCRACGGAA	T36, VT & T30	58		
FCTV2	CTVGCTCGTTTGTGCRGTT	T36, VT & T30	58	14	1000
RCTV2	TAACACCCYGGTAAACAAC	T36, VT & T30	54		
FCTV3	KGGCGTGAARGGTAGGCAA	T36, VT & T30	60	15	1000
RCTV3	CACTAARCARGACTCGTC	T36, VT & T30	60		
FCTV4	CCTYCCMGTTTGTTRKTT	T36, VT & T30	54	16	1000
RCTV4	GAATGMGTCTTHCCACCAC	T36, VT & T30	56		
FCTV5	CGAYGWGTGYTTTATGGTT	T36, VT & T30	56	17	1000
RCTV5	CGCTMGGAGACATAAACTT	T36, VT & T30	56		
FCTV6	AATTTGATCTGTGAACGYG	T36, VT & T30	52	18	1000
RCTV6	GTACCGAACATATAACTCC	T36, VT & T30	54		

N=G o A o T o C, M= A o C, K= T o G, R= A o G, Y= C o T, V = A o C o G, H = A o C o T, W = A o T.

Al cultivo de células se adicionaron 250 μL de medio SOC (2.0 % Tryptona, 0.5 % extracto de levadura, 10 mM NaCl, 2.5 mM KCl, 10.0 mM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 20 mM Glucosa), y se incubó a 37 °C a 6 g por 1 h. Para la selección de colonias recombinantes se plaquearon 150 μL en medio de selección sólido Luria Bertani (LB) más ampicilina (100 $\mu\text{g mL}^{-1}$), adicionado con 50 μL de X-gal (5-bromo-4-cloro-3-indolyl- β -D-galactoside 20 mg mL^{-1}) y 20 μL de IPTG (Isopropylthio- β -D-galactosidase 100 mM). Las cajas petri se incubaron a 37 °C durante 18 h.

Las colonias que contenían el inserto de interés se multiplicaron en 4 mL de medio LB líquido conteniendo ampicilina (100 $\mu\text{g mL}^{-1}$) a 37 °C en agitación constante (6 g) durante toda la noche, para así aumentar la masa bacteriana. La extracción del ADN plasmídico se hizo con el método de lisis alcalina según Sambrook *et al.* (1989). Las clonas positivas se secuenciaron en el Centro de Biotecnología Genómica del Instituto Politécnico Nacional por el método Sequi Therm Excel II DNA Sequencing en el secuenciador DNA Analyzer (LI-COR Inc.) y en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México por el método Taq FS Dye Terminator Cycle Seq. F Base en un ABI-Prism (Foster City, CA, USA).

Análisis de la secuencia de las clonas

Una vez que se obtuvieron las secuencias de las clonas, se procedió a hacer una comparación de éstas con las del banco de genes GenBank del NCBI (National Center for Biotechnology Information) mediante el programa computacional Blast (Basic Local Alignment Search Tool www.ncbi.nih.gov/BLAST/, 2005), para verificar la similitud de las secuencias obtenidas con las previamente reportadas para el VTC.

Los fragmentos de la secuenciación obtenidos se ensamblaron empleando el programa CAP Assembler IFOM (www.bio.ifom-firc.it/ASSEMBLY/assemble.html, 2005) el cual reconoció traslapes significativos entre los fragmentos y los ensambló en una secuencia única denominada "ensamble". Cuando se obtuvo la secuencia final ensamblada se sometió al programa ORF Finder (www.ncbi.nlm.nih.gov/gorf/gorf.html, 2005) para localizar los marcos de lectura abierta.

Análisis bioinformático

Una vez que se obtuvo la estructura del genoma, se hizo un análisis tipo "dotpath" por medio del programa EMBOSS (The European Molecular Biology Open Software Suite) (Rice *et al.*, 2000). El análisis filogenético se llevó a cabo con el paquete Phylip versión 3.6, (Felsenstein,

2005), primeramente transportando las secuencias alineadas al programa DNADIST para estimar las distancias entre secuencias por medio del coeficiente "Jukes and Cantor", mientras que la construcción del árbol se llevó a cabo con el método basado en distancia "Neighbor Joining" (Saitou y Nei, 1987). La robustez del árbol generado se obtuvo mediante la técnica de remuestreo Bootstrap; para esto último se recurrió nuevamente al paquete Phylip y a los programas SEQBOOT y CONSENSE, lo que permitió realizar el remuestreo y obtener el consenso del árbol (Felsenstein, 2005).

RESULTADOS Y DISCUSIÓN

Ensamblaje de los productos secuenciados

La totalidad de genoma aquí reportado representa una estrategia de 6X; es decir, el genoma fue secuenciado seis veces para asegurar una correcta lectura. La secuencia de nucleótidos completa ensamblada se depositó en la base de datos del Centro Nacional para la Información Biotecnológica NCBI (número de acceso DQ272579). Conforme se obtuvo la secuencia de las clonas que abarcan el genoma del VTC, se compararon con secuencias reportadas en el banco de genes para asegurar que correspondieran a las regiones esperadas y la homología que se observó estuvo cercana a 90% en cada una de ellas. Posteriormente, a medida que se ubicó cada fragmento en el genoma y se observó que potencialmente podían formar fragmentos más grandes, se procedió a ensamblar las secuencias de dichos fragmentos (Claverie y Notredame, 2003).

El orden en el que se formaron los ensamblajes, la Figura 1 muestra la estrategia seguida durante la clonación y secuenciación del aislamiento CBG-VTC1, que dio como resultado una secuencia final ensamblada con una longitud de 19 300 nucleótidos más corta que el genoma de la cepa NuagA de Japón y 40 y 74 nucleótidos, 2 nucleótidos más larga que los genomas T385 y VT, respectivamente, y similar a las cepas severa T36 y débil T30.

La organización genómica del CBG-VTC1 fue idéntica a la encontrada en otros aislamientos del VTC (Karasev *et al.*, 1995), y la traducción de la secuencia nucleotídica en aminoácidos indicó productos proteínicos de tamaño similar a los reportados, lo que sugiere conservación de la organización de los genes en el VTC.

Búsqueda de marcos de lectura abierta (ORFs)

El primer codón de inicio (ATG) se localizó en el nucleótido 108 "corriente abajo" de la región 5' terminal (Figura 2). El ORF terminó en el nucleótido 9587, y potencialmente codifica para una proteína de 3159 aminoácidos

con un peso molecular calculado de 357.452 KDa, contiene dos proteasas contiguas, una metiltransferasa y una helicasa. El siguiente ORF “corriente abajo” inició en el nucleótido 9532 y finalizó en el nucleótido 10782, y se traslapó con el primer ORF en los últimos 55 nucleótidos y codificó para una proteína putativa (p23) aminoácidos con

un peso molecular calculado de 47.728 kDa, cuya posible función es la de replicación (Vives *et al.*, 1999) Los ORF “corriente abajo” del 2 al 10 (Figura 2) codifican para 10 productos proteínicos con un rango que va de los 6 a 65 kDa (Karesev *et al.*, 1994; Pappu *et al.*, 1994).

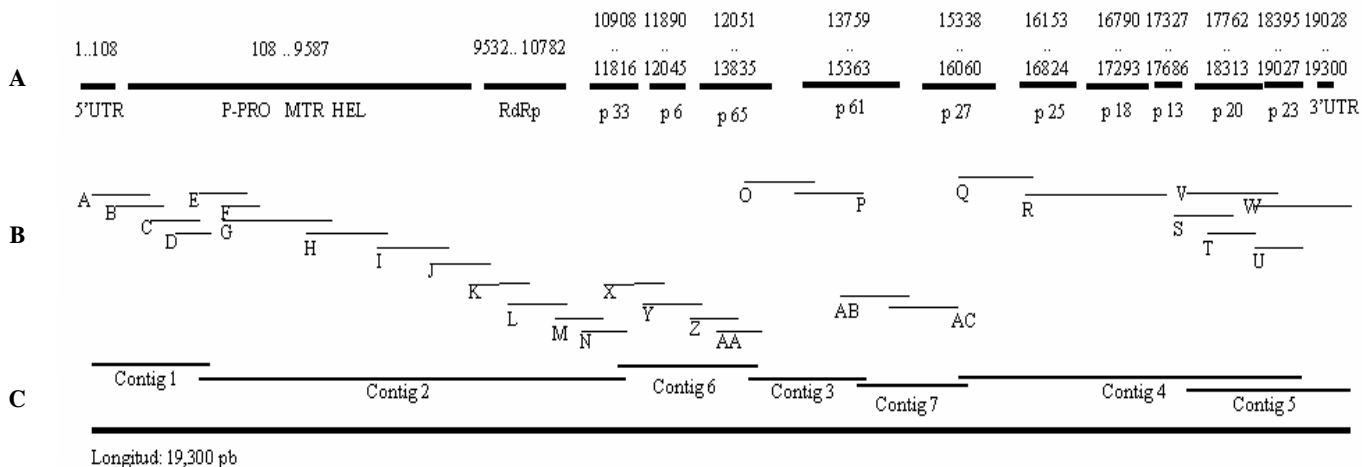


Figura 1. Representación esquemática del genoma del VTC y estrategia de clonación, secuenciación y ensamblaje del aislamiento CBG-VTC1. A) Las líneas en la parte superior representan los ORF's y su correspondiente producto proteínicos putativo. B) Inmediatamente debajo de los ORF's se muestra el ensamblaje de los 29 fragmentos. C) La formación de siete ensambles, que en conjunto dan una secuencia final de 19.3 kb, representada por la línea ubicada en la parte inferior del diagrama.

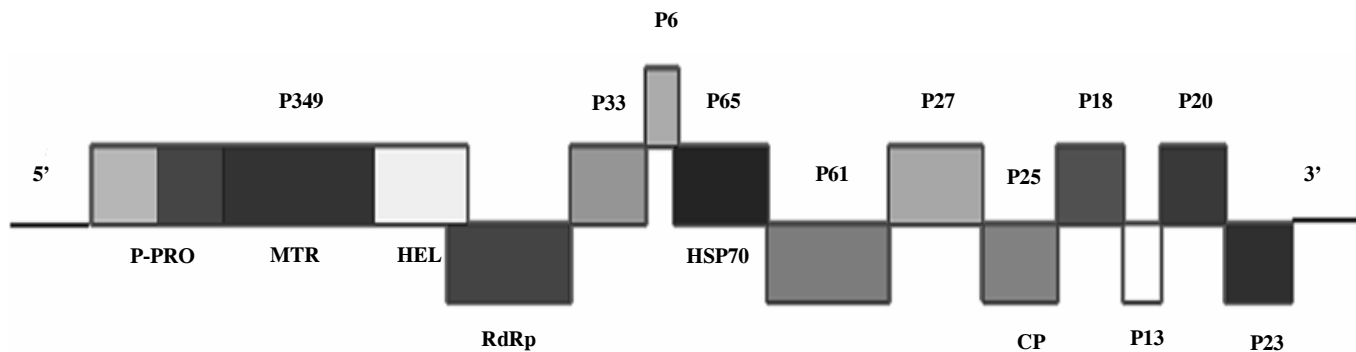


Figura 2. Representación esquemática del genoma del CBG-VTC1. Los rectángulos representan los ORF's, con su respectivo nombre en la parte superior y en la inferior el tamaño en pb.

Identidad y similitud nucleotídica y aminoacídica entre los ORFs y UTRs

El alineamiento de la secuencia aminoacídica de la poliproteína de 357 kDa del CBG-VTC1 contra la del T30, SY568, T36 y VT, mostró 50.3 %, 51.1 %, 73.1 % y 49.6 % de identidad, y un 62.2 %, 62.5 %, 77.9 % y 60.6 % de similitud, respectivamente. La RdRp del CBG-VTC1 mostró una identidad aminoacídica de 82.9 % con la T30, 72.1 % con SY568, 77.2 % con T36 y 81.1 % con VT, y una similitud de 87.1 %, 75.3 %, 78.2 % y 86.4 %, respectivamente (Cuadro 2). Las secuencias de aminoácidos codificadas por los 10 ORF's restantes del CBG-VTC1 con respecto a T30, SY568, T36 y VT, mostraron 83.2 % de identidad y 89.4 % de similitud, en promedio. La región 3'-UTR está muy conservada, con una identidad nucleotídica del 96 % en todos los casos. La región 5'-UTR fue más variable y la identidad nucleotídica entre CBG-VTC1 y T30, SY568, T36 y VT fue 55 %, 57.9 %, 100 % y 58.9 %, respectivamente (Cuadro 2). El porcentaje de similitud más alto se obtuvo con la cepa T36 catalogada como una raza severa, lo cual concuerda con las características biológicas de nuestro aislamiento CBG-VTC1, que también se sabe es un aislamiento severo por ser capaz de ocasionar severos síntomas en árboles de cítricos, tales

como picado del tallo y declinamiento del árbol. Además, existen estudios en que se utiliza esta región para diferenciar las razas severas de las débiles (Karesev *et al.*, 1994).

El ORF 1a codificó para una poliproteína de 357 kDa, que fue la proteína mas variable entre los aislamientos, constituida por dos proteasas, una metiltransferasa y la helicasa (Vives *et al.*, 1999). Para conocer con cual genoma se encontraba más estrechamente relacionado, la secuencia nucleotídica del aislamiento CBG-VTC1 se comparó contra T36 (severa) en un análisis Dotpath (Figura 3A). Se observó a lo largo del genoma una sucesión de puntos que se alinean para formar una diagonal continua, que permite asumir una relación fuerte entre las dos cepas. Lo contrario ocurrió cuando la secuencia del CBG-VTC1 se comparó con la del T30 (débil) (Figura 3B). Esto refuerza los resultados del análisis del porcentaje de identidad y similitud aminoacídica donde se observó que existen mayores regiones de similitud entre CBG-VTC1 y T36 que entre CBG-VTC1 y T30. Estos resultados podrían sugerir, sin afirmar, que este aislamiento de Tamaulipas potencialmente podría comportarse como una raza Severa. Sin embargo, serían necesarios ensayos biológicos adicionales para poder llegar a una conclusión de este tipo (Mendoza *et al.*, 2003; Rubio *et al.*, 2001).

Cuadro 2. Identidad nucleotídica (nt) y aminoacídica (aa) y similitud (%) entre los diferentes ORFs y regiones UTRs del aislamiento CBG-VTC, comparado con los ORFs de los aislamientos T30, SY568, T36 Y VT CTV.

Cepa	CBG-VTC1		T 30		SY 568		T36		VT				
	Tipo Región	Tamaño (nt)	DÉBIL		INTERMEDIO		SEVERO		DÉBIL				
			I	S	I	S	I	S	I	S			
	kDa	nt I	aa	aa	nt I	aa	aa	nt I	aa	aa	nt I	aa	aa
UTR5'	107	55.0			57.9			100.0			58.9		
P357	9374	357.1	50.3	62.2		51.1	62.5	73.1	77.9		49.6	60.6	
RdRP	1211	46.3	82.9	87.1		72.1	75.3	77.2	78.2		81.1	86.4	
P33	908	33.7	84.2	92.7		84.5	92.7	98.3	99.0		83.2	89.4	
P6	155	5.5	92.2	94.1		92.2	94.1	100.0	100.0		92.2	96.1	
P65	1784	64.7	95.1	97.0		95.3	97.3	99.7	99.8		90.2	93.6	
P61	1604	61.2	92.5	95.0		92.9	95.3	97.0	97.2		87.1	91.4	
P27(CPm)	722	27.3	96.2	97.5		94.6	96.2	100.0	100.0		93.3	95.8	
P25(CP)	671	24.9	93.7	96.0		95.5	97.3	100.0	100.0		94.2	96	
P18	503	18.3	91.0	93.4		90.4	92.8	95.8	97.0		88.6	93.4	
P13	359	13.6	90.8	94.1		87.4	91.6	96.6	96.6		90.8	94.1	
P20	551	20.5	83.6	91.3		85.8	92.3	86.9	91.8		84.2	90.7	
P23	632	23.7	91.0	95.2		89.5	97.1	99.5	99.5		87.1	94.8	
UTR3'	272		96.7			97.8		99.3			97.1		
% de i total			85.0	89.1		85.0	89.0	94.5	95.5		84.1	88.4	

I = Identidad; S = Similitud; aa = Aminoácidos; nt = Nucleótidos.

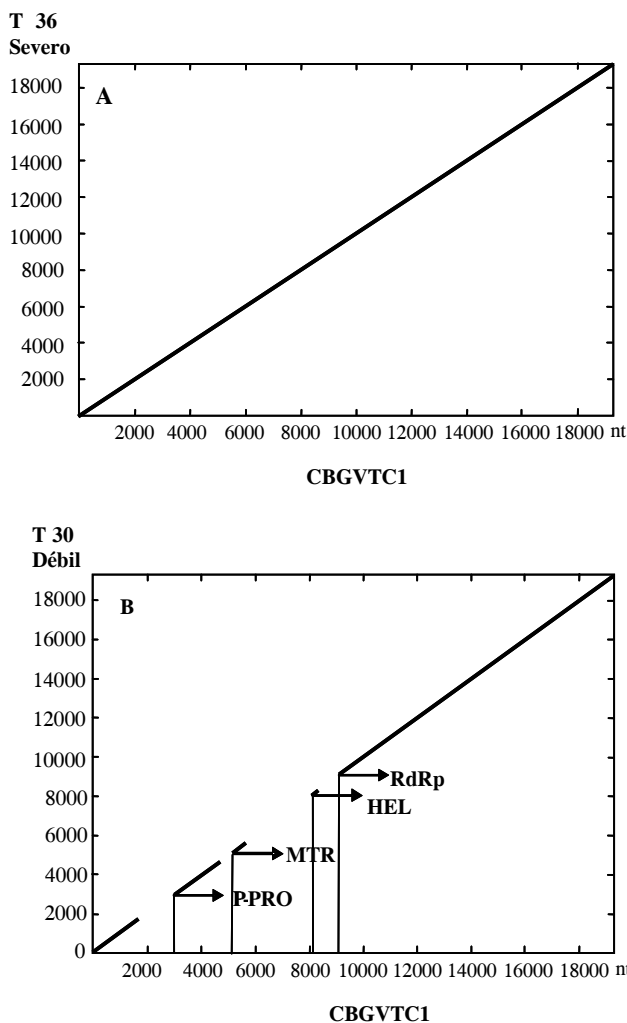


Figura 3. Comparación de la secuencia nucleotídica del aislamiento CBG-VTC1 contra el T36 (A) y de la secuencia nucleotídica del aislamiento CBG-VTC1 contra el T30 (B).

Proteínas asociadas a la replicación

El análisis de la secuencia de la región N-terminal del ORF 1a del VTC que comprende el dominio de las 2 proteasas, las cuales se ha demostrado anteriormente se encuentran asociadas a la replicación, indicó alta similitud con las proteasas de los aislamientos T36, T30 y T385. Se observaron los residuos catalíticos predichos por las secuencias ya reportadas, cisteína e histidina, así como los sitios de separación para ambas proteasas (Figura 4). Los residuos catalíticos predichos son Cys-403 e His-464 para la primera proteasa N-proximal, y Cys-896 e His-956 para la segunda proteasa distal (Gorbalenya *et al.*, 1993; Rozanov *et al.*, 1992; Vives *et al.*, 1999). La separación ocurrió entre Gly-484 y Gly 485 en el dominio “corriente arriba” y entre Gly-976 y Gly-977 en el dominio “corriente abajo” (Figura 4). Normalmente ambos sitios de separación

son precedidos por aminoácidos hidrofóbicos Val y Met, respectivamente. El alineamiento de la región de la poliproteína inmediatamente “corriente abajo” de las proteasas, identificada como dominio de la metiltransferasa (Figura 5), mostró alta similitud con respecto a los dominios de T36, T30 y T385 y se observaron todos los motivos de la metiltransferasa característicos del virus de ARN de cadena positiva (Rozanov *et al.*, 1992; Koonin y Dolja, 1993). La porción carboxilo-terminal del ORF mostró alta similitud con los dominios de la helicasa de los tres aislamientos (Figura 6) y contiene los siete motivos conservados que son típicos de la superfamilia I de las helicasas (Gorbalenya y Koonin, 1993).

Relación con otros genomas del VTC

Se observó un alineamiento múltiple con las secuencias completas de los genomas ya reportadas del VTC, SY568-California, T36-Florida, T36 Qaha-Egipto, Nuaga-Japón, T385-España, VT-Israel y BYV (Beet Yellows). Posteriormente se obtuvo un dendrograma de distancia con el programa Protdist del paquete Phylip, el cual incluyó un análisis de remuestreo “Bootstrap” (Rice *et al.*, 2000). El dendrograma mostró que los aislamientos T385, T30, SY56B, VT y NuagA, que se consideraban como débiles o intermedios, formaron el Grupo A, mientras que aquellos que se consideran severos, T36I, T36F, Qaha y el nuestro CBG-VTC1, formaron un segundo grupo, el Grupo B (Figura 7). En el dendrograma también se muestra claramente que el virus Bty, miembro de la familia closteroviride, sirvió muy bien como raíz de este dendrograma. Además, el análisis de remuestreo para el grupo A, indicó solidez en la formación de este grupo de cepas del VTC, producto de altos porcentajes de identidad entre ellas. Los datos del remuestreo para el grupo B indicaron que al menos en 50 % de casos (578 veces), la cepa CBG-VTC1 siempre se agrupó con las cepas T36 y Qaha, es decir, con las cepas severas.

CONCLUSIONES

El genoma del CBG-VTC1 tiene una longitud de 19 300 pb, codifica para 12 marcos de lectura y su organización mostró sintenia a genomas del VTC reportados. Se observó alta similitud con la secuencia del genoma de una raza catalogada como severa (T36). La región 5'-terminal mostró mayor variabilidad que la región 3'-terminal. El dendrograma de distancia separó en grupos diferentes a las razas severas de las débiles o moderadas.

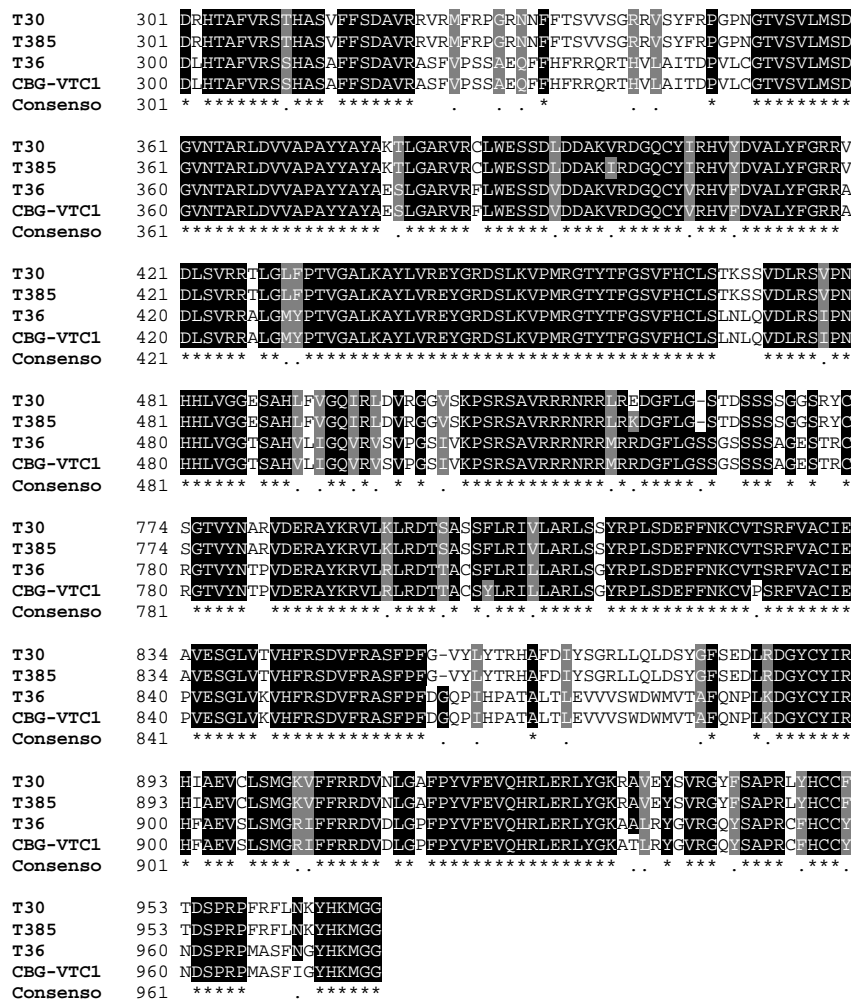


Figura 4. Alineamiento de las secuencias de aminoácidos de las proteasas contiguas del CBG-VTC1 con los dominios de las proteasas de los aislamientos T30, T385 y T36.

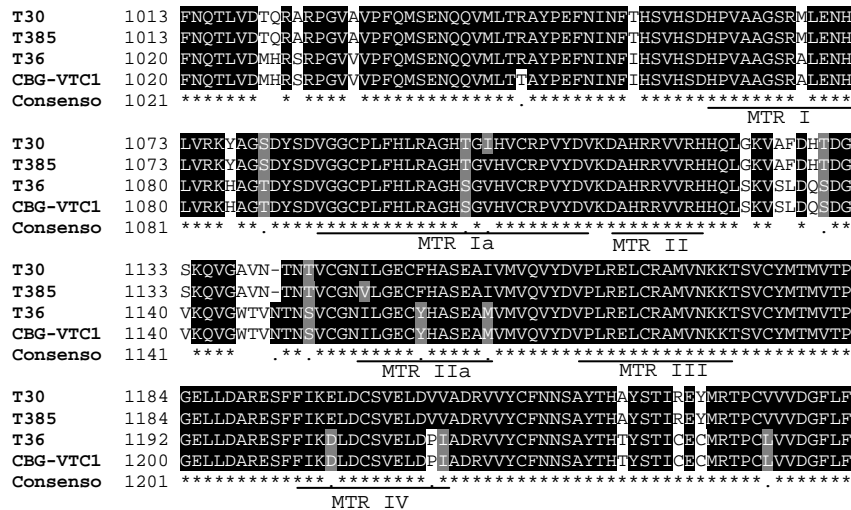


Figura 5. Alineamiento de las secuencias de aminoácidos de la metiltransferasa. Se indican todos los motivos de la metiltransferasa característicos de virus de ARN de cadena positiva.

T30	2675	QFVAANSFLRAVDGKDLTFTNS	EHSLIVYEAPPGGKTHSLVTFYFADYCTKVSCLVVTAN
T385	2675	QFVAANSFLRAVDGKDLTFTNS	EHSLIVYEAPPGGKTHSLVTFYFADYCTKVSCLVVTAN
T36	2684	SFVAANSFLRAVDGKDLTFTNE	EHSLIVYEAPPGGKTHSLVNSYADYCVKVSCLVVTAN
CBG-VTC1	2684	SFVAANSFLRAVDGKDLTFTNE	EHSLIVYEAPPGGKTHSLVNSYADYCVKVSCLVVTAN
Consenso	2701	*****	*****
		HEL I	HEL Ia
T30	2735	KNSQ EISQRISRELM	DRKMLAKHVKTAGRVFTVDSYLMNHLRLTDLDFVDECFMVHA
T385	2735	KNSQ EISQRISRELF	MDRKLAKHVKTAGRVFTVDSYLMNHLRLTDLDFVDECFMVHA
T36	2744	KNSQ EISQRISNELM	GRKLAAYVTDASRVFTVDSYLMNHLRLTQLLFDVDECFMVHA
CBG-VTC1	2744	KNSQ EISQRISNELM	GRKLAAYVTDASRVFTVDSYLMNHLRLTQLLFDVDECFMVHA
Consenso	2761	*****	*****
		HEL II	
T30	2795	GAIGAVVEFTSCKAVVFGDSR	QIHYIHRNDLGVSLHHDIDAFIEPQHRIYGEVSYRCPW
T385	2795	GAIGAVVEFTSCKAVVFGDSR	QIHYIHRNDLGVSLHHDIDAFIEPQHRIYGEVSYRCPW
T36	2804	GAIGAVVEFTSCKAVVFGDSK	QIHYIHRNDLGVSFVADIDAFIQPEHRIYGEVSYRCPW
CBG-VTC1	2804	GAIGAVVEFTSCKAVVFGDSK	QIHYIHRNDLGVSFVADIDAFIQPEHRIYGEVSYRCPW
Consenso	2821	*****	*****
		HEL III	HEL IV
T30	2855	DICEWLSKFYPRHVATANT	GSVKGSSVSEVINGCDDVPYDSSAKYIVYTOAEKNLQKH
T385	2855	DICEWLSKFYPRHVATINT	GSVKGSSVSEVINGCDDVPYDSSAKYIVYTOAEKNLQKH
T36	2864	DICEWLSKFYPRHVATANV	GSVKGSSVSEVINGCDDVPYDKAAKYIVYTOAEKNLQKH
CBG-VTC1	2864	DICEWLSKFYPRHVATANV	GSVKGSSVSEVINGCDDVPYDKAAKYIVYTOAEKNLQKH
Consenso	2881	*****	*****
T30	2915	LGRLTVGRNKAVPVVNTV	HEVQGETYKRVRLVRCCKYQEDTFFCSDNHVVALTRHVDLSIT
T385	2915	LGRLTVGRNKAVPVVNTV	HEVQGETYKRVRLVRCCKYQEDTFFCSDNHVVALTRHVDLSIT
T36	2924	LGRLTVGRNKAVPVVNTV	HEVQGETYKRVRLVRFKYQEDTFFSSKNHIVVALTRHVDLSIV
CBG-VTC1	2924	LGRLTVGRNKAVPVVNTV	HEVQGETYKRVRLVRFKYQEDTFFSSKNHIVVALTRHVDLSIV
Consenso	2941	*****	*****
		HEL V	HEL VI
T30	2975	YSVLNSRRYDKTASSL	DEAREIFDKFRSNHSHGSSLEWYLEKYPTKYGSKASSAPLS
T385	2975	YSVLNSRRYDKTASSL	DEAREIFDKFRSNHSHGSSLEWYLEKYPTKYGSKASSAPLS
T36	2984	YSVLNSRRYDDTAIN	LDRAKEIFDKFRSNHSHGSSLEWYLEKYPTKYGSKASSAPLS
CBG-VTC1	2984	YSVLNSRRYDDTAIN	LDRAKEIFDKFRSNHSHGSSLEWYLEKYPTKYGSKASSAPLS
Consenso	3001	*****	*****

Figura 6. Dominio de las helicasas. Se muestran los siete motivos conservados que son típicos de la superfamilia I de las helicasas.

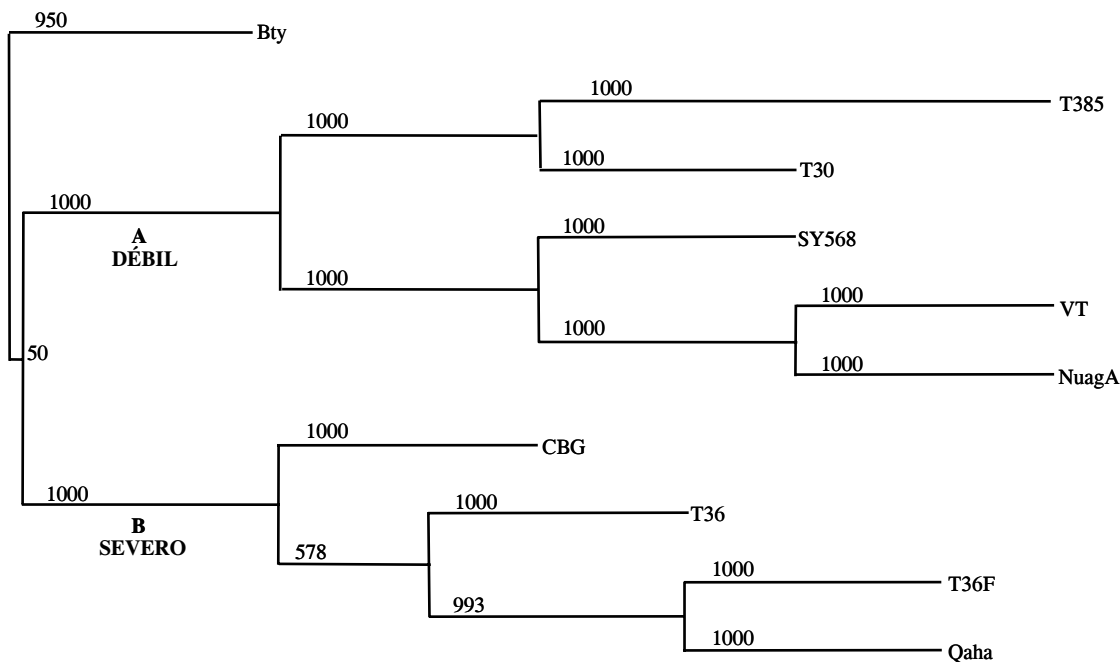


Figura 7. Dendrograma de distancia. Este análisis generó dos grupos; el grupo A conformado por las razas débiles o intermedias (T385, T30, SY568, VT y NuagA) y el grupo B por las razas severas (T36, Qaha y CBG-VTC1).

AGRADECIMIENTOS

A la Coordinación General de Investigación y Posgrado del IPN, Proyecto: 20040239. Por el financiamiento para la investigación y al INIFAP-General Terán, N.L. por facilitar las muestras del tejido infectado por el VTC. Al Dr. John Dagraca (Citrus Research Center, Texas A&M) por su revisión crítica del manuscrito.

BIBLIOGRAFÍA

- Bar-Joseph M, R Marcus, R F Lee (1989) The continuous challenge of citrus tristeza virus control. *Annu. Rev. Phytopathol.* 27:291-316.
- BLAST (2005) www.ncbi.nlm.nih.gov/BLAST/ (Consultada en Marzo, 2006).
- Candresse T, G P Martelli (1995) Beet yellows virus group. *In: Sixth Report of the International Committee on Taxonomy of Viruses.* Arch. Virol. Suppl. 10:461.
- CAP (2005) www.bio.ifom-firc.it/ASSEMBLY/assemble.html (Consultada en Marzo, 2006)
- Claverie J M, C Notredame (2003) *Bioinformatics for Dummies.* Wiley Publishing, Inc.
- Dolja V V, A V Karesev, E V Koonin (1994) Molecular biology and evolution of closteroviruses: sophisticated build-up of large ARN genomes. *Annu. Rev. Phytopathol.* 32: 261-285.
- Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- GenBank (2005) www.ncbi.nlm.nih.gov (Consultada en Noviembre, 2005)
- Goda S K, N P Minton (1995) A simple procedure for gel electrophoresis and Northern blotting of RNA. *Nucleic Acids Res.* 23:3357-3358.
- Gorbalenya A E, E V Koonin (1993) Helicases: Amino acid sequence comparisons and Northern blotting of RNA. *Nucleic Acids Res.* 23:3357-3358.
- Karesev A V, V P Boyko, S Gowda, O V Nikolaeva, M E Hilf, E V Koonin, C L Niblett, K Cline, D J Gumpf, R F Lee, S M Garnsey, D J Lewandowski, W O Dawson (1995) Complete sequence of the *citrus tristeza virus* RNA genome. *Virology* 208:511-520.
- Karesev A V, M E Hilf (1997) Molecular biology of citrus tristeza virus. *In: Filamentous Viruses of Woody Plants.* Research Signpost. P L Monette (ed). Trivandrum. pp:121-131.
- Karesev A V, O V Nikolaeva, E V Koonin, D J Gumpf, S M Garnsey (1994) Screening of the closterovirus genome by degenerate primer-mediated polymerase chain reaction. *J. Gen. Virol.* 75:1415-1422.
- Koonin E V, V V Dolja (1993) Evolution and taxonomy of positive strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* 28:375-430.
- Manjunath K L, H R Pappu, R F Lee, C L Niblett, E L Civerolo (1993) Studies on the coat protein genes of four isolates of citrus tristeza closterovirus from India; cloning, sequencing and expression. *In: Proc. 12th Conf. International Organization of Citrus Virologists, IOCV, Riverside.* pp:16-19.
- Mendoza A, C Salazar, O Alvarado, M A Cruz, H Barrera (2003) Diferenciación molecular de razas severas y débiles de aislamientos del virus de la tristeza de los cítricos en México. *Revista Fitotec. Mex.* 26: 223-230.
- ORF Finder (2005) www.ncbi.nlm.nih.gov/gorf/gorf.html (Consultada en Agosto, 2005).
- Pappu H R, A V Karesev, E J Anderson, S S Pappu, M E Hilf, V Febres, R M G Eckloff, M Mc Caffery, V Boyko, S Gowda, V V Dolja, E V Koonin, D J Gumpf, K Cline, S M Garnsey, W O Dawson, R F Lee, C L Niblett. (1994). Nucleotide sequence and organization of eight 3' open reading frames of the citrus tristeza closterovirus genome. *Virology* 199:35-46.
- Rice, P Longden, I A Bleasby. (2000) "EMBOSS: The European molecular biology open software suite. *Trends in Genetics* 16:276-277
- Rozanov M N, E V Koonin, A E Gorbalenya (1992) Conservation of the putative methyltransferase domain: A hallmark of the "Sindbis-like" supergroup of positive-strand RNA viruses. *J. Gen. Virol.* 73:2129-2134.
- Rubio L, M A Ayllón, P Kong, A Fernández, M Poulek, J Guerri, P Moreno, B W Falk (2001) Genetic variation of *Citrus Tristeza Virus* isolates from California and Spain: Evidence for mixed infections and recombination. *J. Virol.* 75:8054-8062.
- Sambade A, C Lopez, L Rubio, R Flores, J Guerri, P Moreno (2003) Polymorphism of a specific region in gene *p23* of *citrus tristeza virus* allows discrimination between mild and severe isolates. *Arch. Virol.* 148:2325-2340.
- Sambrook J, E F Fritch, T Maniatis (1989) *Molecular Cloning. A Laboratory Manual.* 2nd ed. Cold Spring Harbor Laboratory Press. New York. p ???
- Saitou N, M Nei. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-25.
- Thompson, J D, D G Higgins, T J Gibson. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Vives M C, L Rubio, C López, J Navas-Castillo, M R Albiach-Martí, W O Dawson, J Guerri, R Flores, P Moreno (1999) The complete genome sequence of the major component of a mild citrus tristeza virus isolate. *J. Gen. Virol.* 80:811-816.
- Yang Z-N, D M Mathews, J A Dodds, T E Mirkov (1999) Molecular characterization of an isolate of citrus tristeza virus that causes severe symptoms in sweet orange. *Virus Genes* 19:131-142.