DETERMINACIÓN DE LA ESTRUCTURA PRIMARIA DEL *VIRUS LATENTE DEL CLAVEL* E IDENTIFICACIÓN DE REGIONES REGULADORAS ESENCIALES EN EL CICLO DE INFECCIÓN

Trabajo de Fin de Máster

Máster en Biotecnología Molecular y Celular de Plantas

Sergio Galindo Trigo

Directores: Jesús Ángel Sánchez Navarro

Vicente Pallás Benet

Tutor: José Miguel Mulet Salort

Septiembre 2014

Índice

1. Introducción.....4

1.1. CONCEPTO DE VIRUS.....4

1.2. REPLICACIÓN DE LOS VIRUS DE RNA DE PLANTAS.....4

1.2.1. Primera fase: Infección.....5

1.2.2. Segunda fase: Desencapsidación.....6

1.2.3. Tercera fase: Traducción de las proteínas de replicación.....6

1.2.4. Cuarta fase: Replicación del RNA viral.....7

1.2.5. Quinta fase: Producción de proteínas virales estructurales y de movimiento.....9

1.2.6. Sexta fase: Ensamblaje de los viriones.....10

1.3. FAMILIA *BETAFLEXIVIRIDAE*.....10

1.3.1. Propiedades generales.....10

1.3.2. Organización genómica y replicación.....11

1.3.3. Propiedades biológicas.....12

1.4. GÉNERO *CARLAVIRUS*.....13

1.4.1. Propiedades generales.....13

1.4.2. Expresión génica y ciclo replicativo.....13

1.5. EL *VIRUS LATENTE DEL CLAVEL* (CLV).....14

1.6. VIGS Y VECTORES VIRALES.....16

1.6.1. Desarrollo de vectores de VIGS.....17

1.7. PROTECCIÓN CRUZADA CON VIRUS.....18

1.7.1. Silenciamiento génico y protección cruzada.....19

2. MOTIVACIÓN Y OBJETIVOS.....20

3. MATERIALES Y MÉTODOS.....23

3.1. MATERIAL BIOLÓGICO.....23

3.1.1. Procedencia y manejo de los virus.....23

3.1.2. Cepas bacterianas utilizadas.....23

3.1.3. Oligonucléotidos.....23

3.1.4. Plásmidos.....25

3.2. TESTS DE INFECCIÓN.....27

3.2.1. Ensayos de hibridación Dot-Blot.....27

3.2.2. Amplificiación por RT-PCR.....27

3.3. PURIFICACIÓN DE VIRIONES.....28

3.4. PURIFICACIÓN DE RNA DE DOBLE CADENA.....28

3.5. DISEÑO DE OLIGONUCLEÓTIDOS DEGENERADOS.....29

3.6. ESTRATEGIA GENERAL DE CLONACIÓN.....30

3.6.1. Amplificación de fragmentos de CLV.....30

3.6.2. Purificación de productos de PCR.....31

3.6.3. Ligación.....31

3.6.4. Transformación y selección de colonias positivas.....31

3.6.5. Producción de DNA y purificación del plásmido.....32

3.6.6. Comprobación de los plásmidos purificados.....33

3.6.7. Estrategia empleada para la clonación del cDNA del genoma completo de CLV.....33

3.7. AMPLIFICACIÓN DE EXTREMOS DEL CDNA.....35

3.8. NORTHERN BLOT.....37

3.9. ESTUDIO BIOINFORMÁTICO.....37

4. RESULTADOS Y DISCUSIÓN.....38

4.1. CARACTERIZACIÓN DE LA ESTRUCTURA PRIMARIA DEL *VIRUS LATENTE DEL CLAVEL*.....38

4.2. ANÁLISIS FILOGENÉTICO DEL *VIRUS LATENTE DEL CLAVEL*.....42

4.3. ESTUDIO DE REGIONES REGULADORAS ESENCIALES PARA EL CICLO DE INFECCIÓN DEL *VIRUS LATENTE DEL CLAVEL*.....46

4.3.1. Determinación del inicio de transcripción de los RNA subgenómicos del *Virus latente del clavel*.....46

4.3.2. Estimación de la longitud de la cola poli(A) del *Virus latente del clavel*.....52

5. PERSPECTIVAS.....55

6. CONCLUSIONES.....56

7. REFERENCIAS.....57

ANEXO I: Genoma completo del *Virus latente del clavel*.....60

ANEXO II: Secuencias de las proteínas codificadas por el *Virus latente del clavel*....64

1. Introducción

1.1. CONCEPTO DE VIRUS

Los virus son agentes infecciosos de pequeño tamaño que solamente son capaces de replicarse dentro de las células vivas de otros organismos y por tanto son parásitos de la maquinaria celular de sus huéspedes. Los virus pueden infectar todos los tipos de formas de vida, desde el reino animal, plantas, bacterias y arqueas.

Desde que en 1892, Dmitri Iwanowsky describiera cómo un patógeno de naturaleza desconocida pero no bacteriano infectaba las plantas de tabaco (virus que hoy se conoce como *Virus del mosaico del tabaco* o TMV), cerca de 5000 virus se han descrito ya en gran detalle. Estos aparecen en cualquier ecosistema y constituyen la entidad biológica más abundante del planeta Tierra.

Las partículas víricas, o viriones, constan de hasta tres componentes básicos: i) el material genético, que puede ser DNA o RNA, de simple o doble cadena, de polaridad positiva o negativa, monopartito o multipartito, lineal o circular, ii) proteínas de cubierta que recubren y protegen el material genético y, en algunos casos, iii) una envoltura de naturaleza lipídica y glucídica que cubre la cubierta proteica y que obtienen del huésped.

1.2. REPLICACIÓN DE LOS VIRUS DE RNA DE PLANTAS

Teniendo en cuenta que la mayoría (cerca de un 80%) de los virus poseen un genoma de RNA de simple cadena y de polaridad positiva (single strand RNA (+); ssRNA(+)), es comprensible que la mayoría de los estudios de la replicación de los virus se hayan llevado a cabo en este tipo de virus. Los virus con genoma ssRNA(+) constituyen casi un tercio del total de géneros de virus, entre los que se encuentran tanto virus animales como aquellos que infectan plantas o microorganismos. A pesar

de sus amplias diferencias en lo referente a la morfología de las partículas virales, gama de huéspedes y síntomas producidos, el proceso de la replicación es esencialmente el mismo para todos ellos (Buck, 1996; Haseloff et al., 1984), por lo que en la actualidad se suelen considerar como un único bloque.

La compactación genética en este tipo de virus es tal que una misma proteína puede llevar a cabo múltiples funciones durante el ciclo de replicación. El RNA genómico presenta una función dual, por una parte sirve como molde para la traducción de las proteínas virales, y por otra como molde para la transcripción de la hebra de RNA de polaridad negativa a partir de la cual se da la replicación del RNA. Y, sin embargo, no se producen interferencias entre ambos procesos.

A continuación resumo brevemente los pasos clave en el proceso de replicación viral (Figura 1), que ha sido exhaustivamente revisado previamente en numerosas ocasiones (Mackenzie, 2005; Mandahar, 2006). Como parásitos intracelulares, los virus deben introducirse en la célula huésped y explotar la maquinaria molecular endógena así como las fuentes de energía disponibles. La replicación involucra proteínas del virus y de la célula huésped, e implica las consecuentes múltiples interacciones virus-huésped durante cada una de las fases del ciclo.

1.2.1. Primera fase: Infección

El ciclo infectivo de un virus de plantas se inicia con la entrada de partículas virales en las células vegetales. La entrada del virus en la célula se realiza a través de lesiones, que dañen la cutícula y la pared de pectocelulosa de la célula vegetal, bien mecánicamente o con la ayuda de vectores biológicos (insectos, nemátodos, hongos). No existen evidencias de la presencia de receptores específicos en la membrana plasmática que desencadenen el proceso de infección. Durante el proceso de infección se requiere de la liberación del virión en algún compartimento celular

concreto o la interacción de la partícula vírica con algún componente celular que la dirija al sitio en el que se dará la replicación del mismo. El extremo 5' del virión se suele encontrar parcialmente liberado de proteínas de cubierta en el momento en que entra a la célula.

1.2.2. Segunda fase: Desencapsidación

Con la desencapsidación se consigue desenmascarar el RNA viral (viral RNA; vRNA) al separarse de las proteínas de cubierta. Este proceso se da por mecanismos de desensamblaje tanto co-traduccional, como co-transcripcionales. El vRNA liberado es recogido por los ribosomas, que se encargan de traducirlo y producir los componentes de la replicasa viral. Son, por tanto, fases parcialmente solapantes dentro del ciclo del virus. De esta forma, a medida que se va desestructurando la cubierta proteica, la maquinaria de traducción celular se va uniendo al RNA evitando así su desprotección completa y la posible degradación por nucleasas celulares. Todavía se desconocen los factores que inician el desensamblaje de la cubierta; sin embargo se ha propuesto que este proceso podría desencadenarse a través de un mecanismo de desestabilización como consecuencia de cambios sutiles en el ambiente, tales como cambios en el pH o en la concentración de calcio en el interior celular (Inoue et al., 2011; Mundry et al., 1991; Shaw, 1999).

1.2.3. Tercera fase: Traducción de las proteínas de replicación

En esta fase se producen las proteínas víricas de replicación del RNA, incluyéndose aquí a las RNA polimerasas dependientes de RNA (RdRp). Esto ocurre al comienzo de la infección ya que dentro de la célula no existen previamente maquinarias capaces de realizar la tarea de replicación de genomas de RNA virales. En caso de los virus de ssRNA(+), la RdRp sintetiza el RNA genómico de polaridad negativa complementario al vRNA original y que a su vez servirá de molde para la síntesis de nuevos ssRNA(+). Esta enzima también es la responsable de sintetizar los

RNAs subgenómicos (sgRNAs), los cuales funcionan como mRNA para la traducción de proteínas virales cuya secuencia está localizada bien en regiones internas o bien en el extremo 3' del genoma (Sztuba-Solinska et al., 2011). Durante la actividad de la RdRp se produce un secuestro de la molécula de RNA genómico del proceso de traducción para poder transcribir en sentido 3'-5' y generar la molécula de polaridad negativa. Esto sucede en complejos de replicación, que se localizan normalmente asociados al sistema de endomembranas tales como el retículo endoplasmático (endoplasmic reticulum; ER), Golgi, vacuola, cloroplasto, peroxisoma o endosoma (Hwang et al., 2008; Netherton et al., 2007). El complejo de la replicasa se une a la región 3' no traducible (3'-UTR) y comienza la síntesis del RNA viral de polaridad negativa.

1.2.4. Cuarta fase: Replicación del RNA viral

Para que esta fase comience se requieren algunos pasos previos: se debe dar un reconocimiento específico de elementos *cis* en el RNA viral positivo (los promotores de replicación) por parte del complejo de replicación formado por la RdRp y/o los factores celulares asociados. En ese momento se cambia de un estado traduccional del RNA en el que los ribosomas están generando copias de la RdRp a uno replicativo. La replicación genera moléculas complementarias de polaridad negativa, produciendo así moléculas de RNA de doble cadena (dsRNA).

Estas son separadas por la acción helicasa, haciendo posible la síntesis de nuevas cadenas de polaridad positiva tomando como molde la de polaridad negativa recién formada. Así, de manera reiterativa se pasa por un estado conocido como intermedio replicativo (RI) de RNA, la estructura de dsRNA.

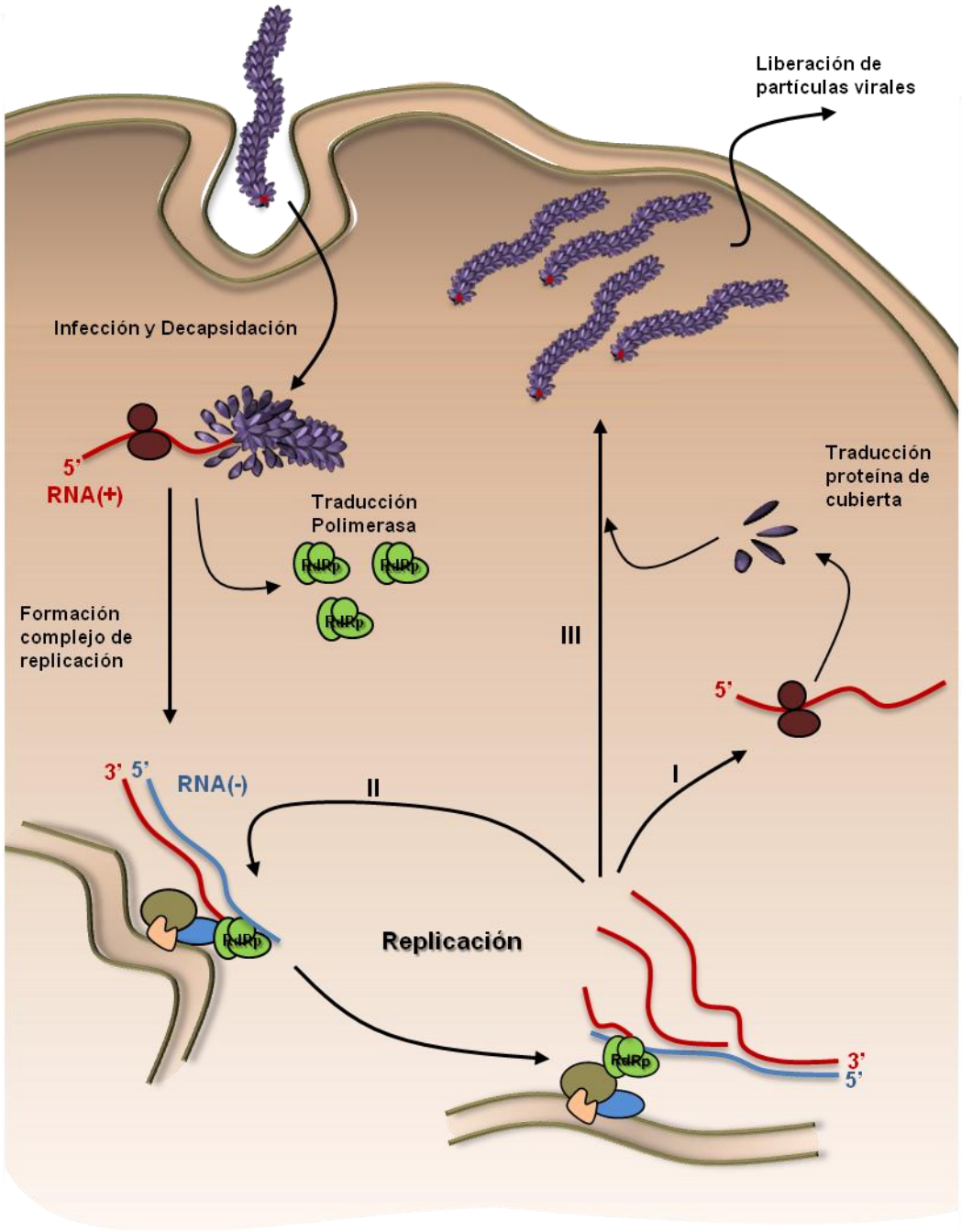


FIGURA 1. Replicación de los virus de ssRNA(+). Representadas todas las fases del proceso de replicación, desde la entrada a la célula, la desencapsidación simultánea a la traducción de la RNA polimerasa viral, la posterior formación del complejo de transcripción asociado a membranas celulares, la síntesis posterior de la progenie de moléculas de RNA genómico viral de polaridad positiva con sus tres posibles destinos: I) actuar como mRNA para la síntesis de proteínas virales, II) alimentar el mecanismo de síntesis de hebras de polaridad negativa virales, o III), la encapsidación y formación de viriones maduros que posteriormente podrán ser liberados al exterior.

El complejo de la RdRp necesita de ciertas señales para comenzar la replicación, estas se suelen encontrar cerca del extremo 3' del RNA genómico en la polaridad positiva, como pueden ser estructuras de tipo tRNA, colas poli(A), pequeñas horquillas o pequeñas secuencias primarias sin estructuras complejas. Cuando la síntesis de RNA se inicia desde promotores internos en el RNA de polaridad negativa se generan los sgRNAs.

La progenie de RNA de polaridad positiva se genera a través de un mecanismo asimétrico que permite la síntesis simultánea de varias cadenas, que en ciertos estadios parecen "colgar" de la negativa a partir de la que se transcriben. Esto resulta en una descompensación en la formación de cadenas positivas y negativas (llegando a ser proporciones de hasta 1000:1 (Buck, 1996)). Los nuevos genomas de RNA positivo se separan del molde por la actividad helicasa, y son capeados por la actividad metiltransferasa. Esta progenie de moléculas puede entrar a 3 rutas diferentes: i) actuar como mRNA y servir como molde para la síntesis de proteínas virales (quinta fase), ii) ser convertida en dsRNA y alimentar la maquinaria de replicación, o iii) encapsidarse y producir los viriones (sexta fase).

1.2.5. Quinta fase: Producción de proteínas virales estructurales y de movimiento

Junto con las polimerasas virales, los virus expresan un reducido número de proteínas requeridas para los diferentes procesos virales, tales como transporte, defensa, encapsidación, etc. Dado el reducido tamaño de los genomas virales, los virus han desarrollado toda una serie de mecanismos que les capacita tanto para competir con los genes celulares como para lograr una eficiente expresión y replicación del genoma viral (Bustamante and Hull, 1998; Drugeon et al., 1999; Firth and Brierley, 2012). Entre ellos cabe destacar, (a) la organización multipartita del genoma con RNAs policistrónicos o *ambisense*, (b) la síntesis de sgRNAs, (c) el

procesamiento proteolítico de una poliproteína, (d) el solapamiento de pautas abiertas de lectura (open reading frame; ORF), (e) la síntesis alternativa de proteínas de mayor tamaño obviando codones de parada ("read-through"), (f) el desplazamiento de lectura (frameshift) o (g) la iniciación interna de la traducción (internal ribosome entry site; IRES). El hecho de que las distintas fases de la traducción estén bajo el control de un programa espacio-temporal hace posible que los virus hagan un uso más eficiente de su capacidad codificante.

1.2.6. Sexta fase: Ensamblaje de los viriones

La unión de las proteínas de la cápsida a las moléculas de ssRNA positivas requiere del reconocimiento específico del ssRNA por parte de las subunidades de la cápsida. En ocasiones se pueden llegar a formar partículas vacías. En el caso de otros tipos de virus existe una fase posterior de maduración en la que se añade la cobertura membranosa al gemar en alguna de las membranas celulares.

1.3. FAMILIA *BETAFLEXIVIRIDAE*

La familia *Betaflexiviridae*, junto con las familias *Alphaflexiviridae*, *Gammaflexiviridae* y *Tymoviridae*, se localiza dentro del orden *Tymovirales*, compuesto por virus que en la mayoría de los casos infectan plantas y cuyos genomas están compuestos por una única molécula ssRNA(+). En los más recientes estudios sobre la taxonomía de virus se incluyen dentro de la familia *Betaflexiviridae* a los géneros *Capillovirus*, *Carlavirus*, *Citrivirus*, *Foveavirus*, *Trichovirus* y *Vitivirus* (King et al., 2012).

1.3.1. Propiedades generales

Las propiedades distintivas de los viriones de esta familia son: i) en cuanto a su morfología, se pueden describir como filamentos flexibles, de unos 12-13 nm de diámetro y de entre 600 y 1000 nm de longitud, dependiendo del género viral. En lo

respectivo a ii) su composición en ácidos nucleicos, contienen una única molécula ssRNA(+) de entre 5,9 y 9,0 kb, que representan aproximadamente un 5% de la masa total del virión. La molécula de RNA contiene un grupo 7-metilguanosina trifosfato, conocido como estructura CAP, en su extremo 5', mientras que en el extremo 3' presenta una cola de poliadeninas. Además, algunos representantes del género *Carlavirus* tienen dos RNAs subgenómicos (sgRNAs) de aproximadamente 3,0 y 1,5 kb, que posiblemente se encapsiden en partículas de menor tamaño. Por último, en los referente a iii) las proteínas, todos los miembros presentan una única proteína de cubierta (CP) de un tamaño que varía entre 18 y 44 kDa.

1.3.2. Organización genómica y replicación

Dependiendo del género del que se trate, el número de genes puede variar entre 3 y 6, pero en todos ellos el expresado por la 5' ORF proximal, comparte un gran porcentaje de identidad con otras polimerasas virales. Estas proteínas, de entre 190 y 250 kDa, contienen dominios conservados para las actividades metiltransferasa (Mtr), helicasa (Hel) y RNA-polimerasa RNA-dependiente (RdRp). En la mayoría de los miembros aparecen dominios AlkB y Papain-like proteasa entre los dominios Mtr y Hel (Figura 2).

Los procesos de transporte viral los llevan a cabo proteínas que expresadas por los ORFs de menor tamaño que aparecen en las proximidades del extremo 3' del gen de la polimerasa. Pueden concentrarse en una única proteína de movimiento (MP) de la superfamilia "30K" (como en el caso de *Capillovirus*, *Citrivirus*, *Trichovirus* y *Vitivirus*), o en tres proteínas pertenecientes al "triple gene block" (TGB). Por último, el gen de la CP siempre aparece aguas abajo de la MP, y en ciertos géneros, como *Carlavirus* y *Vitivirus*, presentan un ORF 3' proximal que expresa una proteína con un motivo "zinc-finger" de unión a ácidos nucleicos, y que en *Vitivirus* se ha comprobado que tiene capacidad de suprimir el silenciamiento de RNA (Zhou et al., 2006).

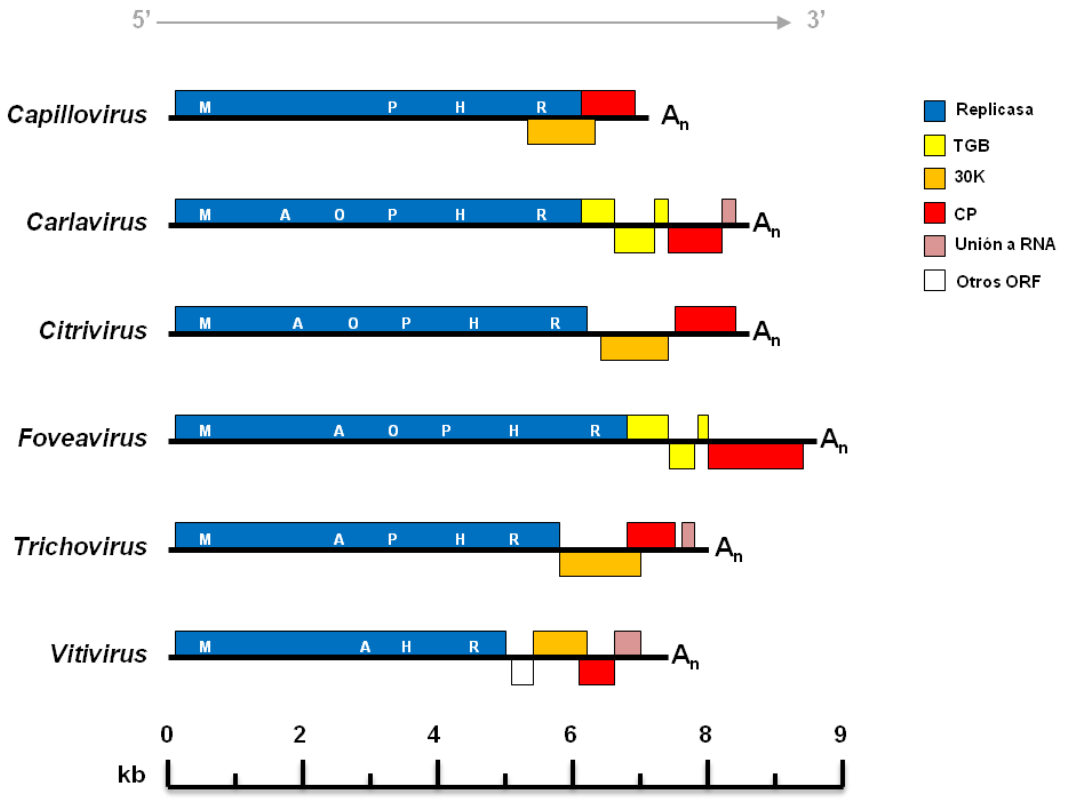


FIGURA 2. Representación esquemática de la organización genómica de los géneros pertenecientes a la familia *Betaflexiviridae*. Se representan con los bloques los ORFs. Anotados los dominios del ORF1 con posibles funciones metiltransferasa (M), AlkB (A), OTu-like peptidasa (O), papain-like proteasa (P), RNA-helicasa (H) y RNA polimerasa RNA-dependiente. Figura adaptada de (Martelli et al., 2007).

Todos los ORFs aguas abajo de la polimerasa se traducen desde sgRNAs que se localizan frecuentemente en los tejidos infectados, incluso junto a sus formas asociadas de doble cadena de RNA (dsRNA). Para todos los miembros de la familia *Betaflexiviridae* se postula una replicación citoplasmática, siendo el producto del ORF1 la única proteína para la que se ha demostrado su implicación directa en dicho proceso.

1.3.3. Propiedades biológicas

La gama de hospedadores para los miembros individuales de la familia *Betaflexiviridae* es relativamente limitado, siendo mayoritariamente hospedadores leñosos, a excepción de los pertenecientes al género *Carlavirus*. Muchos de ellos

causan síntomas moderados. La inoculación mecánica ha resultado exitosa en todas las especies en las que se probado con diferentes grados de dificultad. Pueden utilizar como vectores de transmisión diferentes especies de hongos y áfidos.

1.4. GÉNERO *CARLAVIRUS*

Se trata de uno de los géneros que componen la familia *Betaflexiviridae*. Nombrado a partir de uno de sus integrantes, el *Virus latente del clavel* (*Carnation latent virus*), a día de hoy se han encontrado más de 50 especies virales dentro de este género. En general se caracterizan por la presencia de 6 ORFs, tres de los cuales se corresponden con el TGB, además de una CP, una proteína de unión a RNA y una replicasa de gran tamaño (Figura 2). Todas las ORFs se expresan a partir de sgRNAs, a excepción de la replicasa.

1.4.1. Propiedades generales

La morfología típica de los viriones de este género es la de filamentos flexibles, de unos 12 a 13 nm de diámetro por 470 a 1000 nm de longitud, compuesto por la asociación de múltiples unidades de la CP al genoma linear y de ssRNA(+) de entre 5,8 y 9,0 kb. El extremo presenta una estructura CAP mientras que en el extremo 3' presenta una cola de poliadeninas.

1.4.2. Expresión génica y ciclo replicativo

El RNA del virión es infectivo y hace las veces de genoma y RNA mensajero viral, a partir del cual se transcribe directamente la polimerasa; mientras que el resto de ORFs se supone que se transcriben como sgRNAs monocistrónicos (Figura 3).

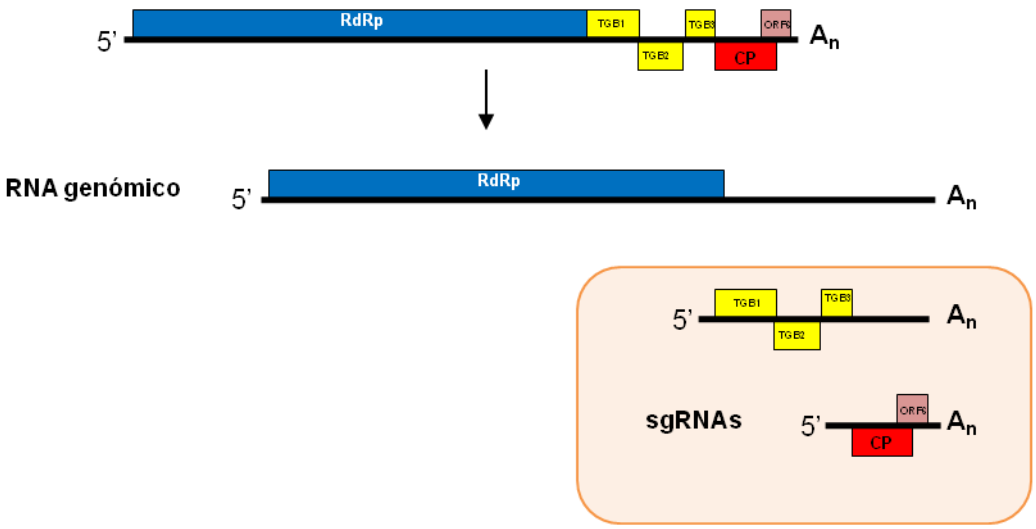


FIGURA 3. Esquema de la organización genómica de los carlavirus y las formas intermediarias a partir de las cuáles se lleva a cabo la traducción de las proteínas (el RNA genómico para la replicasa; los subgenómicos para las proteínas de movimiento, de cubierta y de unión a RNA).

1.5. EL *VIRUS LATENTE DEL CLAVEL* (CLV)

El *Virus latente del clavel* (CLV, del inglés, *Carnation latent virus*) es el virus representativo del género *Carlavirus*. Aislado a partir de variedades comerciales de *Dianthus caryophyllus* L. en 1954 al estudiar otros virus que afectaban al clavel, el CLV se encontró en stocks de claveles asintomáticos y supuestamente sanos que en estudios serológicos parecían estar dando resultados confusos (Kassanis, 1954). Los primeros estudios realizados acerca de este virus determinaron algunas propiedades físicas como la forma de "bastón" de sus viriones, la ausencia de síntomas en el huésped experimental empleado, clavel del poeta (*Dianthus barbatus* L.), su relación serológica con otros miembros del género *Carlavirus*, como el *Virus S de la patata*, y la posibilidad de transmisión mediante inoculación mecánica o a través del áfido *Myzus persicae* de manera no persistente (Kassanis, 1955). Más tarde, en estudios más exhaustivos se determinó que las partículas víricas tienen un tamaño de

aproximadamente 650 nm de longitud y 12-13 nm de diámetro (Figura 4, A) (Brandes et al., 1959).

A pesar de que el CLV provocaba infecciones asintomáticas, se pudo identificar a *Chenopodium quinoa* como una planta indicadora adecuada para el estudio de este virus. Las plantas de *C. quinoa* infectadas con CLV muestran síntomas muy leves en hojas inoculadas (puntos amarillentos de unos 3 mm de diámetro tras una semana desde la inoculación mecánica) así como en el resto de tejidos distales (moteado amarillo, rugosidad y leve curvatura de hojas, morfologías irregulares en las hojas, y un crecimiento ralentizado general (Figura 4, B). No se han llevado a cabo estudios exhaustivos sobre la gama de posibles huéspedes para el CLV, reduciéndose a especies de los géneros *Chenopodium* y *Dianthus* (Hollings and Stone, 1965).

En lo referente al genoma del CLV, se supone una organización similar a la que presentan otros carlavirus, un genoma de ssRNA(+) de aproximadamente 8.0-8.5 kb, con la distribución típica de ORFs para este género. Solamente se dispone de la secuencia genómica parcial de unas 2.1 kb en total (Figura 4, C), de fragmentos correspondientes a i) un fragmento de la replicasa localizado cerca de su extremo 3' de 536 nucleótidos (Pham et al., 2008), ii) el ORF correspondiente al TGB3 de 263 nucleótidos (Foster et al., 1990), y iii) el extremo 3' final del virus incluyendo los ORFs correspondientes a las proteínas de la cubierta y de unión a RNA (Meehan and Mills, 1991).

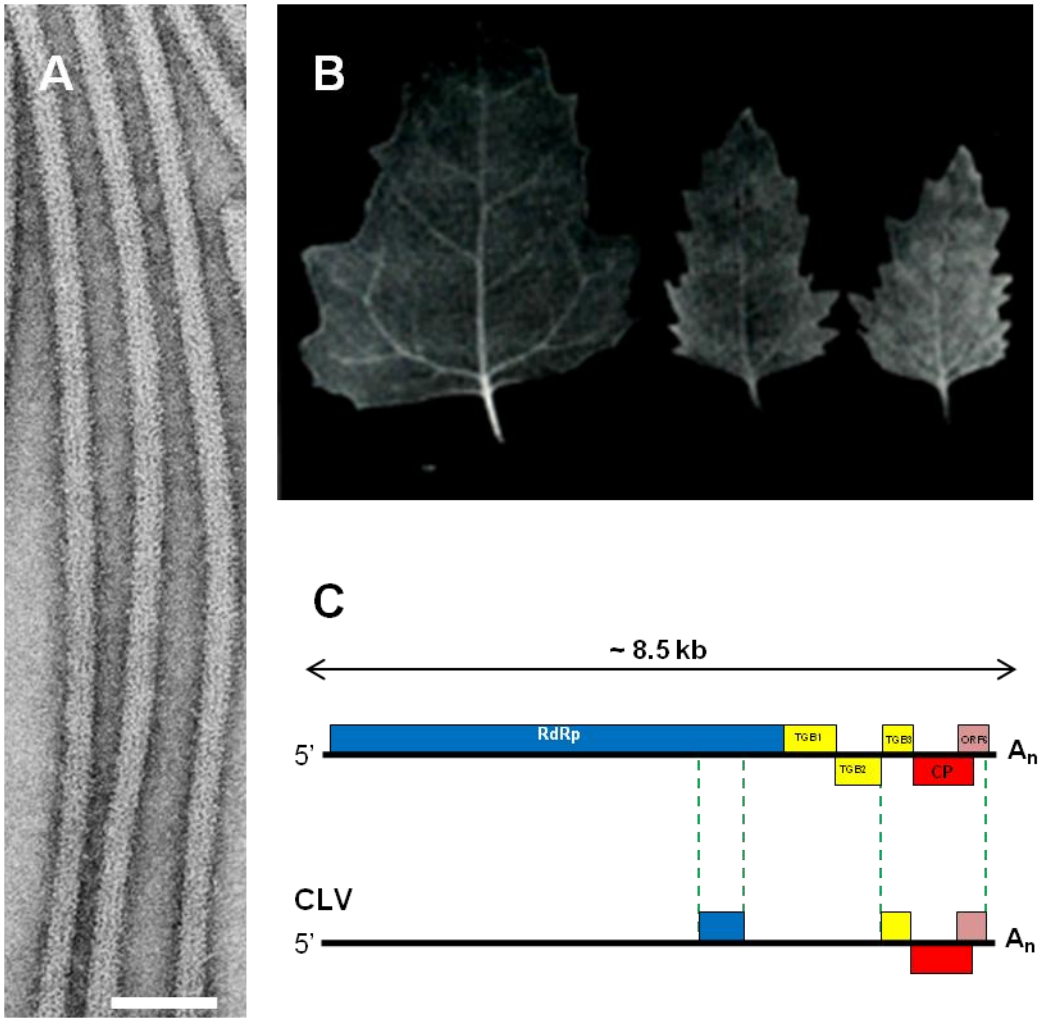


FIGURA 4. Características del *Virus latente del clavel CLV*: A, fotografía de microscopía electrónica de partículas víricas de CLV (barra de 50 nm) (Martelli et al., 2007); B, *Chenopodium quinoa* como huésped experimental para CLV. A la izquierda, hoja sana, a la derecha hojas sistémicas de plantas infectadas en las que se observan un detrimento en el desarrollo (Hollings and Stone, 1965); C, esquema representativo del conocimiento de la secuencia genómica del CLV hasta la fecha. Arriba el esquema de la estructura que se presume tendrá el virus, debajo los fragmentos depositados en las bases de datos públicas.

1.6. VIGS Y VECTORES VIRALES

El silenciamiento génico inducido por virus o VIGS (del inglés, *Virus Induced Gene Silencing*) explota el sistema de defensa innato del que disponen las plantas de silenciamiento génico postranscripcional o PTGS (del inglés, *Posttranscriptional Gene Silencing*) en la lucha contra la proliferación intracelular de los virus y su movimiento

extracelular. El VIGS se basa en el fenómeno del RNA de interferencia (RNAi), nombre que hace referencia a la interferencia en la expresión génica mediada por pequeños RNA que se da de manera específica y secuencial. Una de las funciones principales de esta ruta de PTGS de las plantas es la defensa contra los virus, en la que este sistema es dirigido contra los genes virales con el objetivo de frenar la infección (Ding and Voinnet, 2007; Pallas, 2013). En el VIGS se aprovecha el mecanismo de defensa desatado contra el RNA viral como una herramienta en genética reversa, permitiendo analizar la función de determinados genes en la planta. En comparación con otras herramientas, el VIGS puede generar *knock-downs* funcionales en cuestión de semanas sin la necesidad de transformar plantas genéticamente (Baulcombe, 1999).

1.6.1. Desarrollo de vectores de VIGS

El proceso a seguir suele ser el siguiente: se clona una parte del gen a analizar en un derivado infeccioso del virus, ya sea en vectores virales basados en DNA (en el caso de los virus de DNA), o basados en cDNA (en el caso de virus de RNA); estos se introducen en las plantas por inoculación mecánica de transcritos in vitro del vector, agroinfiltración o métodos biolísticos. Durante el proceso de infección que se desencadena el RNA viral pasa a fase de replicación por acción de la RdRp, de manera que se obtienen las formas de dsRNA que son entonces reconocidas por el sistema de PTGS (Waterhouse and Fusaro, 2006). Una ribonucleasa celular (DICER) se encarga del reconocimiento del dsRNA y las procesa en pequeños RNA de interferencia de 20-25 nucleótidos (siRNA). Estos pequeños RNAs son reconocidos posteriormente por el complejo multiproteico RISC (RNA-induced silencing complex) que las separa en siRNAs de cadena simple. Luego, la asociación complejo RISC-siRNA de cadena simple rastrea las poblaciones de RNA en busca de una secuencia complementaria y la degrada. A su vez, estas secuencias siRNA se amplifican e incluso transportan por los tejidos, desencadenando un PTGS a nivel sistémico (Kalantidis et al., 2008).

Además de la homología entre la secuencia que se incluya en el vector viral y aquella contra la que queremos dirigir el silenciamiento, existen otros requisitos para que el VIGS se produzca eficientemente: i) la longitud de esta secuencia debe ser de al menos 23 nucleótidos con una identidad del 100% (Thomas et al., 2001), y ii) la orientación del inserto (Lacomme et al., 2003).

1.7. PROTECCIÓN CRUZADA CON VIRUS

La protección cruzada es el fenómeno mediante el cual se suprimen los síntomas de enfermedad causados por una cepa de un virus al infectar la planta con otra cepa de ese virus, o un virus muy relacionado que sea asintomático o muy poco agresivo (Dodds et al., 1985). Este proceso se conoce desde hace más de un siglo, sin embargo, el mecanismo por el que se rige aún no se ha esclarecido por completo. Desde las primeras hipótesis que proponían que la presencia previa de un virus podría suponer una competición por determinados recursos que los virus emplean específicamente (Bawden, 1934), pasando por aquellas que sugerían que podrían competir por sitios específicos de la célula para su replicación (Kassanis and White, 1974), o incluso otras en las que se proponía que la expresión de la CP del virus atenuado podría ser crucial para que la protección cruzada ocurriera (Sherwood, 1987), hasta la propuesta de que era la homología a nivel de secuencia de nucleótidos entre las cepas de virus la que podría estar dirigiendo este proceso (Palukaitis and Zaitlin, 1984); se ha seguido sin conseguir dar una respuesta concreta al mecanismo que dirige la protección cruzada.

1.7.1. Silenciamiento génico y protección cruzada

En 2006, Gal-On y Shiboleth propusieron y revisaron un modelo que explica la protección cruzada y que tiene en cuenta el silenciamiento génico por RNA. En él sugieren que es probable que durante el ciclo de replicación del virus, en el momento en que se sintetizan las formas de dsRNA virales, estas puedan ser reconocidas por el sistema celular de tipo DICER. Como muchos virus suelen disponer de herramientas para detener o hacer menos efectivo el sistema de silenciamiento génico celular para así favorecer su avance, estableciéndose una "carrera" entre la dispersión de la señal de silenciamiento por los tejidos y el avance del virus,Gal-On y Shiboleth proponen que, en el caso de la protección cruzada, la cepa protectora estaría escapando del mecanismo de silenciamiento génico de la planta mientras coloniza sus tejidos, y a su vez, dejando los tejidos por los que ha pasado "preparados" (con la maquinaria de silenciamiento activada) para la llegada de la cepa atacante. Es un modelo que explica con claridad por qué la protección cruzada solamente funciona entre cepas de virus estrechamente relacionadas y la necesidad de un intervalo de tiempo entre ambas inoculaciones (Gal-On and Shiboleth, 2006).

2. MOTIVACIÓN Y OBJETIVOS

El clavel (*Dianthus caryophyllus* L.) es uno de los principales cultivos en floricultura a escala mundial. Es una planta de la familia *Caryophyllaceae* y pertenece al género *Dianthus*, en el cual se han clasificado más de 300 especies. Muchas especies se distribuyen por toda Europa y Asia, incluso a ciertas partes de América y África. Además, algunas especies de este género, incluyendo a *D. caryophyllus, D. barbatus, D. chinensis, D. plumarius, D. superbus* y sus híbridos son ampliamente empleados como cultivares en horticultura. Debido a su interés económico, numerosas nuevas variedades de clavel se han ido seleccionando en función de características como el color, tamaño, fragancia, o longevidad de las flores. A principios de 2014 se publicó el primer borrador del genoma del cultivar "Francesco" de *D. caryophyllus* (Yagi et al., 2014).

Este cultivo se ve afectado por la infección de múltiples virus, como el *Virus del moteado del clavel* o CarMV (del inglés, *Carnation mottle virus*), o el *Virus de las manchas anilladas del clavel* o CRSV (del inglés, *Carnation ringspot virus*), que en ciertos casos conllevan considerables pérdidas económicas en la fase de cultivo en viveros del clavel. En la lucha contra las virosis en los cultivos con más impacto comercial se suele recurrir a la búsqueda de variedades o especies cercanas resistentes a una enfermedad concreta, para luego intentar introgresar el material genético responsable de la misma. Estos programas de mejora pueden ser costosos y de larga duración, y definitivamente ineficaces cuando se trata de cultivos para los que no se pueden encontrar esos genes de resistencia por motivos de incompatibilidad. Aunque en estos casos la transgénesis puede ser la solución más sencilla, no es en muchos casos una metodología popular. Por esto, aproximaciones como la protección cruzada aparecen en estas situaciones como una solución plausible a los problemas planteados.

Como se ha comentado en la sección 1.5. del presente trabajo, el *Virus latente del clavel* (CLV), por su capacidad de provocar infecciones asintomáticas en los cultivos de clavel comercial podría considerarse como un buen candidato para desarrollar a partir de él un sistema viral de protección cruzada en el que una versión modificada del CLV se utilizaría como cepa protectora frente a infecciones por otros virus que sí provocan pérdidas en el cultivo de clavel.

Para ello es necesario conseguir un sistema eficiente de inoculación e infección de CLV que, mientras conserve sus propiedades asintomáticas, confiera resistencia cruzada frente a virus que no estén íntimamente relacionados a nivel de secuencia con él, como CarMV o CRSV. Esa ausencia de homología de secuencia podemos superarla si integramos las metodologías de desarrollo de vectores virales para VIGS con las bases de la protección cruzada. La idea consiste en incluir en el vector desarrollado a partir de CLV fragmentos de la secuencia del genoma del virus contra el que se pretende conferir resistencia, de manera que si CLV despierta el mecanismo de silenciamiento génico en el clavel, este quede preparado para detectar y degradar el RNA del virus atacante en una infección posterior, dando lugar así a la protección cruzada. Por otro lado, el desarrollo de un vector viral basado en CLV, abre también la posibilidad de expresar o silenciar proteínas de interés en clavel, aspectos muy a tener en cuenta tanto en la selección de nuevas variedades como en estrategias de control frene a otras enfermedades.

A pesar de tratarse del miembro tipo del género *Carlavirus* y de un virus sobre el que se han realizado varios estudios acerca de la gama de huéspedes, estructura de las partículas víricas y propiedades serológicas, entre otras, paradójicamente no se dispone de ningún aislado clonado en vectores plasmídicos ni de la secuencia completa del genoma del CLV, requerimientos básicos para desarrollar una estrategia como la que se describe más arriba. Por todo esto, y con el fin de dar los primeros pasos dentro de la fase de desarrollo de la herramienta molecular de protección

cruzada, en este trabajo de fin de máster se abordan los siguientes objetivos concretos:

2.1. Caracterización de la estructura primaria del *Virus latente del clavel*.

2.2. Análisis filogenético del CLV, basado tanto en la secuencia de nucleótidos como en la de aminoácidos de las diferentes ORFs.

2.3. Caracterización molecular de regiones reguladoras del CLV esenciales para su ciclo de replicación celular: determinación del inicio de los sgRNAs así como del tamaño de la cola poli(A).

3. MATERIALES Y MÉTODOS

3.1. MATERIAL BIOLÓGICO

3.1.1. Procedencia y manejo de los virus

Como fuente de inóculo se utilizó tejido infectado (aislado PV-0365) procedente de la Colección Alemana de Microorganismos y Cultivos Celulares del Leibniz Institute (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures). Las muestras liofilizadas recibidas se trataron siguiendo las instrucciones del proveedor: para la inoculación, el material vegetal liofilizado se reconstituyó con unas gotas del buffer de inoculación (fosfato sódico 0,05 M, pH 7,0, EDTA 1 mM, DIECA 5 mM y ácido tioglicólico 5 mM) en un mortero a baja temperatura. Se maceró hasta obtener una pasta uniforme y se añadieron unos 3-4 ml de buffer de inoculación y se procedió a la inoculación mecánica sobre hojas sanas de *Chenopodium quinoa* y/o *Dianthus barbatus*, previamente tratadas con el abrasivo carborundum. Posteriormente se rociaron las plantas con agua para eliminar el exceso de inóculo y evitar una evaporación excesiva. Las plantas inoculadas se incubaron en condiciones de 16 horas de luz a 25 ºC/8 de oscuridad a 20ºC.

3.1.2. Cepas bacterianas utilizadas

En todos los procesos de selección y purificación del DNA plasmídico se utilizó la cepa DH5α de *E. coli* (Invitrogen, genotipo cromosómico: *fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*).

3.1.3. Oligonucléotidos

En la siguiente tabla se recoge el conjunto de oligonucleótidos que se ha empleado para los experimentos descritos en el presente trabajo.

Referencia	Oligonucleótido (5' - 3')	Posición en el genoma de CLV	Propósito (C/S/E/O)	Comentarios
2567	AAGAGACTCGAGGAAGATGTGG	7212	C	
2568	AAAATACGGAAAAATATTTAAAACGC	8459 (comp)	C	
2569	CCTGACACAGCCGATTGTTCC	7283 (comp)	C	
2570	TGTACTTGTGACAGATGCAAGCG	6950	C	
2571	ACTGCAAACCGTGAACCTAAGG	7156 (comp)	C	
2572	CAGATTATGAAGCGTTTGATGCAAG	5310	C	
2574	ATGCACT(A/C)ACATAC(A/C)G(A/C/G)AG(C/T)CC	~140	C	Oligonucleótido degenerado
2575	CA(C/T)GACGAG(T/C/A)T(G/A)CACTACTGG	~ 700	C	Oligonucleótido degenerado
2576	GG(C/T)GA(T/C)CC(A/T)TG(C/T)CAGAGT(G/A)ATTA(C/T)G	~4300	C	Oligonucleótido degenerado
2605	ATGTGTGTGCCTCACAGGG	4461	S	
2606	TCAGAGGCTTTACAGACC	6411 (comp)	E	Caracterización del inicio del sgRNA1
2607	CCTGACGCAACCGATTGCTCC	7283 (comp)	E	Caracterización del inicio del sgRNA2
2608	AACCTATAGCTCAGCGTGGC	3971 (comp)	C	
2609	GGAACGTGCTGCCTCTCC	4933 (comp)	S	
2611	CGCTCAATGTCACTTGGG	3925 (comp)	C	
2661	GGGTCACCGCTCTAAGTCGC	4272 (comp)	C	
2662	TCCTTTTCACAAACAAGCGC	2393 (comp)	C	
2663	CATCTCAAATTTTACAGCTCC	2420 (comp)	C	
2664	ATCCAGGATGTAATTATCGC	3035	S	
2696	ACAAACCTGCTTCAATAAGC	232 (comp)	E	Caracterización del extremo 5'
2697	GAGAGAGAGCATAATTGAAG	196 (comp)	E	Caracterización del extremo 5'
2698	TTCCTTTTCCAGTGGTCAGC	1032	C	
2703	GTAAAATACGGAAAAATATTTAAA	8454 (comp)	O	Amplificación y clonación del genoma completo del

	ACGC			CLV
2704	AAGTCAATAGCCCGATGTCC	8276	E	Caracterización del extremo 3'
2706	TTTGGTAAGCCGTGGGCTCAGGC	1126 (comp)	C	
2707	AATAGTCAAGATCGCCTCG	3136 (comp)	S	
2708	TAGCATTGCACTTAGATTGC	4404 (comp)	S	
2721	GTCGTATCCAGTGCAGGGTCCGAGGGTATTCGCACTGGATACGACTTTTTTTTTTTTTTT	-	O	Oligonucleótido *Stem-loop*
2723	TCTACAACCGCTAACGACCC	6173 (comp)	S	
2733	CGATAAACAAAACAAAACAAAACAACGC	1	O	Amplificación y clonación del genoma completo del CLV
2740	GATGCGAAACGCTTGAACAGG	8158 (comp)	E	Caracterización del inicio del "sgRNA3"
461	GGACGAAGCTTGTATATGGAATTGGTT	-	O	Comprobación de insertos en el plásmido pUC18 modificado
URP	GTGCAGGGTCCGAGGT	-	O	*Universal Reverse Primer*, utilizado para hacer nested-PCR sobre reacciones con el oligonucleótido *Stem-loop*

TABLA 1. Conjunto de oligonucleótidos utilizado para este trabajo. Anotado el propósito de cada uno de ellos: C, utilizados para la amplificación de fragmentos del genoma viral y clonación para su posterior secuenciación; S, diseñados estrictamente con el propósito de secuenciar; E, empleados en la determinación de extremos del RNA (comienzo de sgRNAs, extremos 5' y 3' genómicos, cola poli(A)); y O, otros.

3.1.4. Plásmidos

Para la realización del presente trabajo se utilizaron los plásmidos pTZ57R/T y pUC18, ambos preparados para su replicación y mantenimiento en *Escherichia coli*, presentan las pautas de lectura abierta (ORFs) *bla* para conferir resistencia al antibiótico ampicilina y *lacZ* para facilitar la selección de las colonias positivas por color azul/blanco (Figura 5, A). El plásmido pTZ57R/T (Thermo Scientific InsTAclone PCR

Cloning Kit) permite la clonación directa de productos de PCR que contienen una adenina protuberante en el extremo 3', generados por determinadas Taq DNA polimerasas. Este plásmido se utilizó en las tareas de obtención de la secuencia genómica del virus, determinación de la secuencia inicial de los sgRNA, estimación de la longitud de la cola poli(A) y clonación del cDNA del genoma viral completo. En el caso del plásmido pUC18, se utilizó una versión modificada del mismo con el propósito de clonar el cDNA del genoma viral completo. La modificación consiste en la sustitución de la MCS del plásmido por un inserto de 1,2 kb (por lo que la discriminación por color se pierde) que contiene un casete de expresión transitoria compuesto por el promotor 35S del *Virus del mosaico de la coliflor* (*Cauliflower mosaic virus*, CaMV) y la secuencia del terminador del gen inhibidor de proteinasa II de *S. tuberosum* (Popit) (Figura 5,B).

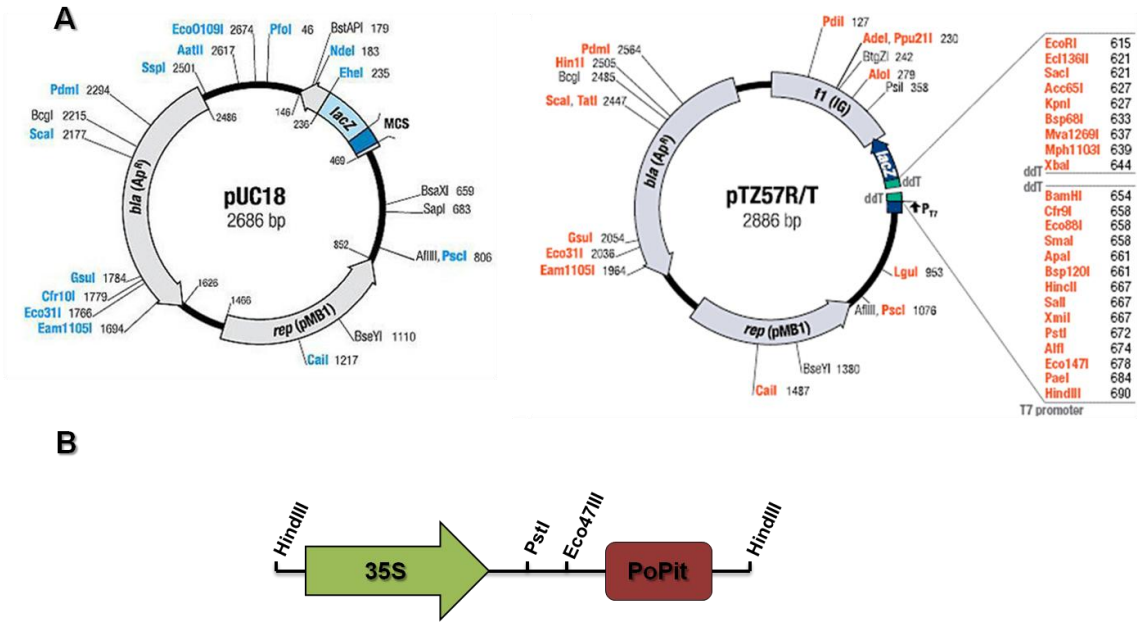


FIGURA 5. Representación esquemática de los plásmidos utilizados en el presente trabajo. A, estructura de los plásmidos pUC18 y pTZ57R/T (Imágenes de la página web de Thermo Scientific) en donde las dianas de restricción y su ubicación en la secuencia plasmídica están indicadas. Las diferentes ORFs están indicadas con flechas; B, inserto presente en la versión modificada de pUC18, en donde se muestran el promotor 35, el terminador PoPit y las dianas de restricción utilizadas.

3.2. TESTS DE INFECCIÓN

3.2.1. Ensayos de hibridación Dot-Blot

Las plantas inoculadas con tejido infectado con CLV fueron analizadas mediante ensayos de hibridación tipo Dot Blot tal y como se ha descrito previamente (Sanchez-Navarrro et al., 1996). Brevemente, las hojas se homogeneizaron con tampón de extracción (citrato sódico 50 mM pH 8,5; EDTA 1mM), utilizando una relación de 1 ml por cada 100 mg de tejido. 1 μl del extracto se aplicó directamente sobre la membrana de nylon cargada positivamente (ROCHE), procediendo a fijar covalentemente con luz UV (700 x 100 $\mu J/cm^2$) tras dejar secar la membrana durante unos minutos. Los pasos de prehibridación e hibridación se llevaron a cabo tal y cómo se ha descrito previamente (Pallas et al., 1998), utilizando una sonda de RNA marcada con digoxigenina diseñada para hibridar con los 393 nucleótidos finales del extremo 3' del virus (Sanchez-Navarro et al., 1999). Para la detección quimioluminiscente se utilizó el substrato CSPD (ROCHE), siguiendo las indicaciones de fabricante.

3.2.2. Amplificación por RT-PCR

Las muestras de tejido a analizar se procesaron siguiendo el protocolo de extracción de RNA con sílica (Rott and Jelkmann, 2001). Posteriormente, los extractos se analizaron mediante RT-PCR utilizando el kit comercial de RT-PCR SuperScript III OneStep RT-PCR System-Platinum Taq DNA polymerase (Invitrogen Life Technologies), siguiendo las indicaciones del fabricante en un termociclador Mastercycler Pro (Eppendorf) y empleando los oligonucleótidos específicos 2567 y 2568 (Tabla 1) para amplificar el extremo 3' de 1,3 kb de CLV. Por último se comprobó el resultado de la amplificación por electroforesis en gel de agarosa-TAE al 1%.

3.3. PURIFICACIÓN DE VIRIONES

La purificación de partículas víricas se llevó a cabo según el protocolo previamente descrito (Diez et al., 1998). Brevemente, 0,5 g de tejido infectado de *Chenopodium quinoa recogido* a los 15 días post inoculación (dpi) se trituró a temperatura ambiente con 1 ml de buffer de extracción (fosfato potásico 100 mM pH 7,1; ácido ascórbico 100 mM y EDTA 20 mM) preparado en el momento. Tras la homogenización, se añadieron 0,5 ml de cloroformo:butanol (1:1), se mezcló mediante vórtex durante 1 minuto y se procedió a centrifugar durante 10 minutos a 14000 rpm. El sobrenadante obtenido se mezcló con 1/5 del volumen total de polietilenglicol (PEG)-20000 al 30% (p/v) y se incubó durante 15 minutos en hielo. Tras la incubación, las muestras se centrifugaron a 6500 rpm durante 10 minutos, tras lo cual, se eliminó el sobrenadante y se procedió a resuspender el precipitado obtenido con 200 μl de tampón de elución (fosfato sódico 10 mM pH 7,0; EDTA 1 mM). Tras incubar las muestras durante 30 minutos y centrifugar a 6500 rpm durante 5 minutos, se separó el sobrenadante enriquecido en partículas virales y se almacenó a -80ºC.

3.4. PURIFICACIÓN DE RNA DE DOBLE CADENA

Para purificar la fracción de RNA de doble cadena se partió de 3,5 g de *C. guinoa* infectado a 15 dpi. El tejido se trituró con nitrógeno líquido y se mezcló con 10 ml de buffer STE 1X, 1 ml de SDS 10 %, 500 μl de bentonita 2 % y 9 ml de fenol-STE 1X, mediante agitación durante 30 minutos. Tras centrifugar 15 minutos a 10000 rpm, se procedió a separar la fase acuosa a la que se le añadió un 1/5 de volumen de etanol (96 %) y 2,5 % (p/v) de celulosa CF-11. Tras agitar durante una hora, se procedió a centrifugar a 3000 rpm durante 5 minutos, descartando el sobrenadante. El precipitado obtenido se lavó con 20 ml de STE 1X al 16% de etanol, procediendo a centrifugar a 3000 rpm durante 5 minutos. El pellet obtenido se resuspendió en 1 ml de

agua estéril, se incubó a 70 ºC durante 4 minutos y se centrifugó 3 minutos a 13000 rpm. El sobrenandante enriquecido con dsRNAs se guardó a -20ºC o -80ºC.

3.5. DISEÑO DE OLIGONUCLEÓTIDOS DEGENERADOS

Para amplificar mediante RT-PCR segmentos del CLV de los que se desconocía las secuencias flanqueantes, se recurrió a la utilización de oligonucleótidos degenerados diseñados en base a un alineamiento de secuencias múltiple de los genomas completos de los carlavirus con más similitud al CLV. Para ello se utilizaron las secuencias disponibles en la base de datos del GenBank del CLV, correspondientes a: i) un fragmento de la replicasa (FJ555525.1, (Pham et al., 2008)), ii) el gen 7K, y iii) el extremo 3' del virus, que contiene los ORFs de la proteína de cubierta y de unión a RNA (X55331.1, X52627.1, (Foster et al., 1990; Meehan and Mills, 1991), y mediante búsquedas por similitud de secuencia de nucleótidos para todas ellas con el algoritmo Blastn (Altschul et al., 1997) en el NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi), se recopilaron las secuencias con mayor similitud, correspondientes a los virus: *Carlavirus Garlic latent virus* (NC_003557), *Coleus vein necrosis virus* (NC_009764), *Butterbur mosaic virus* (NC_01352) y *Helleborus net necrosis virus* (NC_012038). El alineamiento de las secuencias recopiladas utilizando el programa ClustalX 2.1 (Larkin MA. *et al*., 2007) permitió seleccionar los bloques mejor conservados conteniendo secuencias de alrededor de 20 nucleótidos en base a los cuales se diseñaron los oligonucleótidos degenerados (2574, 2575 y 2576, Tabla 1) (Figura 6). El conjunto de oligonucleótidos degenerados diseñados se recoge en la Tabla 1.

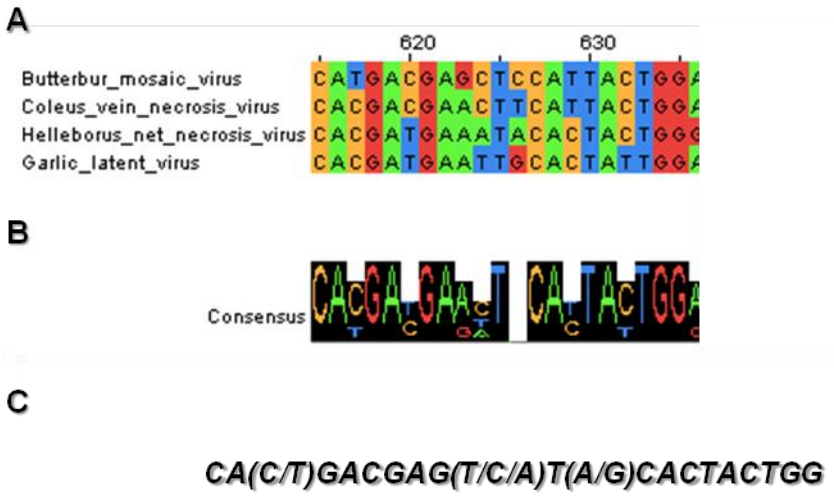


FIGURA 6. Diagrama del método de diseño de oligonucleótidos degenerados. A, alineamiento de las secuencias genómicas de *Garlic latent virus* (NC_003557), *Coleus vein necrosis virus* (NC_009764), *Butterbur mosaic virus* (NC_01352) y *Helleborus net necrosis virus* (NC_012038). B, Logo de la secuencia consenso para el alineamiento. C, Oligonucleótido diseñado en base al mismo.

3.6. ESTRATEGIA GENERAL DE CLONACIÓN

3.6.1. Amplificación de fragmentos de CLV

El paso de obtención del cDNA correspondiente a secuencias parciales del genoma de CLV se llevó a cabo mediante RT-PCR a partir de RNA total o dsRNA extraído de material vegetal infectado con el CLV (Secciones 3.2.2., 3.4., respectivamente), utilizando las combinaciones de oligonucleótidos específicos y degenerados (Tabla 1) que permitieran la amplificación de diferentes fragmentos solapantes del CLV y que cubrieran todo el genoma viral. Cuando se amplificaban fragmentos de más de 2 kb se recurrió al kit comercial de RT-PCR SuperScript III HiFi OneStep RT-PCR System-Platinum Taq DNA polymerase (Invitrogen Life Technologies). Los productos amplificados se analizaron en geles de agarosa al 1% en tampón TAE 1X (Tris-acetato 40 mM, EDTA 1 mM) con 0,05 µl de bromuro de etidio

y se visualizaron en un transiluminador “Vilber Lourmat, ECX-20-M”. El marcador de peso molecular utilizado fue el de 1Kb plus DNA Ladder (Thermo Scientific).

En los casos en los que la concentración obtenida de producto de PCR no fuera suficiente, se recurrió a un segundo paso de amplificación, empleando la misma combinación de oligonucleótidos utilizada en el paso de RT-PCR, utilizando la polimerasa de alta fidelidad PrimeSTAR® HS (Takara Bio Inc.) siguiendo las recomendaciones del fabricante en un termociclador “Mastercycler Pro” (Eppendorf).

3.6.2. Purificación de productos de PCR

Los productos de PCR obtenidos en los diferentes pasos de amplificación se purificaron mediante extracción con fenol-cloroformo y precipitación con etanol y acetato sódico. Cuando en el análisis de los amplicones se observó más de una banda, se procedió a purificar el producto de PCR del tamaño esperado a partir de geles de agarosa-TAE al 1%, utilizando el kit comercial “GeneJET™ Gel Extraction Kit” (Thermo Scientific) y siguiendo el protocolo descrito por el fabricante.

3.6.3. Ligación

Los productos de PCR purificados se ligaron directamente mediante TA-*cloning* en el vector pTZ57R/T (Thermo Scientific) utilizando la enzima DNA ligasa del fago T4 (Promega). Las reacciones se llevaron a cabo en volúmenes de 6 µl, con 0,5 µl del plásmido (100 ng), 2 µl del inserto, 3 µl de tampón 2X y 0,5 µl de la ligasa (1-3 U/µl; Promega). La ligación se llevó a cabo durante 16 h a 16ºC.

3.6.4. Transformación y selección de colonias positivas

Los productos de ligación se transformaron mediante electroporación en *E. coli* (cepa DH5α) que presenta, entre otras, las mutaciones *endA1*, que inactiva una endonucleasa intracelular que degrada el DNA plasmídico; *hsdR17*, que elimina la endonucleasa de restricción del sistema de modificación-restricción *EcoKI*, y por tanto

los fragmentos de DNA que poseen metilaciones en los sitios *EcoKI* no son degradados; y además *recA*, que elimina la posibilidad de recombinación homóloga. 3 µl de la reacción de ligación se mezclaron con 50 µl de células competentes de *E. coli* en hielo. Posteriormente se electroporaron en un electroporador "Gene Pulser Xcell Electroporation System" (BIO-RAD), en cubetas de 0,1 mm y con pulsos de 1350 V. Inmediatamente se resuspendieron en 300 µl de medio Luria Bertani (LB) compuesto por 10 g/l de triptona, 5 g/l de extracto de levadura y 10 g/l de NaCl, procediendo a incubarlas a 37ºC durante media hora. Por último se crecieron a 37 ºC durante 16 horas en medio sólido de LB agar suplementado con 50 mg/ml de ampicilina para la selección de las colonias transformadas con el plásmido empleado en la ligación.

Para seleccionar las colonias transformadas que contenían el inserto esperado, se realizó un análisis por PCR de colonias sobre aquellas de color blanco (y que, por tanto, incluyen un inserto que interrumpe el ORF *lacZ*). Seleccionamos varias colonias blancas al azar y utilizando el kit comercial "GoTaq DNA polymerase - GoTaq Green MasterMix" (Promega), en reacciones de 5 µl, con la combinación de oligonucleótidos: directo utilizado para la amplificación por RT-PCR - reverso M13/pUC (complementario a la región flanqueante aguas abajo del MCS del plásmido); o directo M13/pUC (complementario a la región flanqueante aguas arriba del MCS del plásmido) - reverso empleado en la amplificación por RT-PCR. Tras el análisis en geles de agarosa de los productos de la PCR de colonias, aquellas con una banda específica para el tamaño de inserto esperado fueron seleccionadas como colonias positivas.

3.6.5. Producción de DNA y purificación del plásmido

Colonias positivas para el inserto de interés se seleccionaron y se crecieron en 5 ml de medio LB suplementado con ampicilina en agitación, a 37 ºC y durante 16 horas. Los cultivos se centrifugaron a 6000 rpm durante 30 segundos a temperatura

ambiente y en tubos de 2 ml. Descartado el sobrenadante se procedió a la preparación de MiniPreps a partir de los pellets de bacteria utilizando el kit comercial "GeneJET™ Plasmid Miniprep Kit" (Thermo Scientific), siguiendo las indicaciones del fabricante para el aislamiento y purificación del plásmido. El plásmido obtenido se resuspendió en 50 µl.

3.6.6. Comprobación de los plásmidos purificados

En un último paso de comprobación, 5 µl de MiniPrep se sometieron a una reacción enzimática de restricción, utilizando las enzimas *Bam*HI y *Xba*I (Thermo Scientific), cuyos sitios de corte están presentes en la MCS del plásmido pTZ57R/T flanqueando el sitio de TA-*cloning*. Todas las reacciones se llevaron a cabo en volúmenes finales de 10 µl, con tampón TANGO 1X (Thermo Scientific) y 0,5 µl de cada una de las enzimas, incubando a 37 ºC durante 2 horas. En paralelo se llevaron a cabo reacciones de restricción sobre la misma MiniPrep con una sola enzima (*Bam*HI o *Xba*I) para linearizar el plásmido. Los productos de la restricción se analizaron en geles de agarosa-TAE al 1%, como se ha explicado en secciones anteriores, seleccionando como positivas aquellas que desprendieran un fragmento del tamaño esperado en la doble restricción, o bien que desprendieran varios fragmentos de menor tamaño y para los que la diferencia de tamaño entre la banda del vector linearizado y el tamaño del plásmido pTZ57R/T fuera el correspondiente al tamaño del inserto esperado.

3.6.7. Estrategia empleada para la clonación del cDNA del genoma completo de CLV

El procedimiento descrito en los apartados anteriores de la sección 3.6. hace referencia a la estrategia de amplificación y clonación seguida en este trabajo en todos los casos, excepto para la clonación del cDNA correspondiente al genoma completo del CLV. Aunque la intención con la que aquí aparece, es la de conseguir la secuencia

de nucleótidos genómica de un único clon del CLV, para poder aprovechar el esfuerzo puesto en la clonación para objetivos posteriores del proyecto, se decidió utilizar un procedimiento alternativo.

En lugar de utilizar el plásmido pTZ57R/T, se empleó una versión modificada del plásmido pUC18, en el que se ha introducido un inserto en la MCS que contiene un casete de expresión compuesto por el promotor 35S de CaMV y el terminador PoPiT (ver sección 3.1.4.). Este plásmido se sometió a una reacción de restricción con la enzima Eco47III, que genera extremos romos. El producto de la digestión se extrajo de un gel de agarosa y se purificó siguiendo las recomendaciones del kit comercial "GeneJETTM Gel Extraction Kit" (Thermo Scientific). A continuación y con el objetivo de reducir la tasa de religación del plásmido, el plásmido digerido y purificado se desfosforiló en un volumen final de 10 µl conteniendo 1 µl del tampón 10X y 1 unidad de la enzima fosfatasa alcalina FastAP (Promega Corp). La reacción de desfosforilación se mantuvo durante 20 min a 37ºC, tras lo cual se procedió a inactivar la fosfatasa incubando la reacción a 75ºC durante 10 min.

El cDNA correspondiente al genoma completo del CLV se sintetizó mediante una reacción de RT-PCR utilizando la muestra enriquecida en dsRNA y el kit comercial de RT-PCR SuperScript III HiFi OneStep RT-PCR System-Platinum Taq DNA polymerase (Invitrogen Life Technologies), por tratarse de un fragmento a amplificar de 8,5 kb. El fragmento de cDNA obtenido se procesó de acuerdo a lo descrito en la sección 3.6.2. y se sometió a una reacción de fosforilación de los extremos 5' con la enzima T4 Polynucleotide Kinase (Promega Corp.), de acuerdo al protocolo proporcionado por el fabricante. A continuación, el producto resultante se fenolizó, se precipitó en etanol y se resuspendió en un volumen reducido de agua estéril para incrementar así la eficiencia de ligación. El resto del proceso se llevó a cabo como se describe en las secciones 3.6.3. y 3.6.4., excepto por los oligonucleótidos empleados en comprobación del inserto (2704 - 461, Tabla 1), y la selección visual de las

colonias, ya que en este caso el vector no permite la discriminación por color. Para la comprobación de los plásmidos purificados se llevó a cabo una reacción de restricción con la enzima *Apa*I (Thermo Scientific).

3.7. AMPLIFICACIÓN DE EXTREMOS DEL CDNA

La amplificación de extremos de cDNA se utilizó para determinar la secuencia de nucleótidos de los extremos 5' y 3' del CLV, el inicio de los sgRNAs y estimación de la longitud de la cola poli(A) presente en el extremo 3' del CLV. Las regiones correspondientes a los extremos 5' y 3' del CLV se amplificaron utilizando el kit comercial 5'/3' RACE System (Roche), siguiendo las indicaciones proporcionadas por el fabricante. Para la amplificación del extremo 3' se aprovechó la presencia de una cola poli(A) en la estructura primaria del CLV. La utilización de un oligo dT junto con el oligo *sense* (2704, ver Tabla 1) en la reacción de RT-PCR, permitió obtener un amplicón de 300 nucleótidos. Para el extremo 5', se requirió un paso previo de poli-adenilación del extremo 3' de la primera cadena de cDNA obtenida con el oligo 2696. Tras la poliadenilación, se realizó una reacción de PCR, utilizando los cebadores oligo-dT y 2696, la primera cadena poliadenilada y la polimerasa de alta fidelidad PrimeSTAR® HS (Takara Bio Inc.), siguiendo las recomendaciones del fabricante.

La utilización del kit comercial 5'/3' RACE System (Roche), implica la utilización de un cebador oligo-dT, degenerado para la última base del extremo 3'. Este cebador está diseñado para unirse justo al inicio de la cola poliadenilada, en donde el último nucleótido que precede a la cola, se determina por la selección de cebadores dentro del *pool* que forma el oligo-dT. Para poder determinar el tamaño aproximado de la cola poli(A) del CLV así como de los extremos 5'y 3', sin utilizar un cebador degenerado, decidimos adaptar el protocolo descrito para la detección y cuantificación de microRNAs (miRNAs) mediante *stem-loop* RT-PCR (Varkonyi-Gasic et al., 2007). Este protocolo se basa en el uso de un oligonucleótido cuyo extremo 3' es complementario

al extremo 3' de los miRNAs, y su extremo 5' presenta una zona de autocomplementariedad de 15 nucleótidos, de forma que se genera una estructura secundaria en el oligonucleótido que favorece en gran medida su unión específica al extremo 3' final del miRNA. En el nuevo protocolo, que denominaremos *stem-loop* oligo-dT, sustituimos la zona 3'' del oligonucleótido por una secuencia de 15 timinas, de manera que se uniera al extremo de la cola poli(A) o a los extremos de los tractos poliadenilados (Figura 7). Asimismo, de manera alternativa al proceso de síntesis de cDNA mediante poliadenilación de la primera cadena de cDNA, recurrimos a la poliadenilación de los extremos 3' de las muestras de dsRNA con la enzima polyA Polymerase, Yeast (Affymetrix), sobre los que pudimos llevar a cabo la amplificación de los extremos 5' directamente mediante una única reacción de RT-PCR.

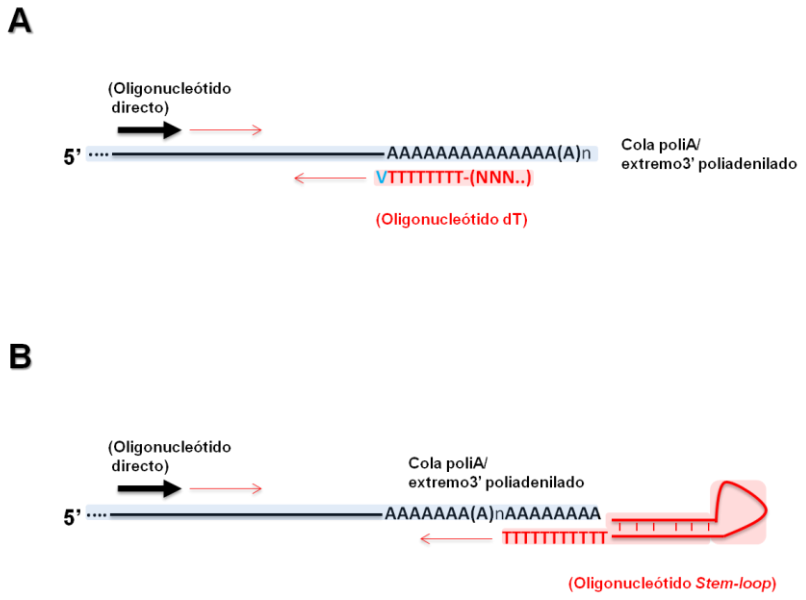


FIGURA 7. Comparación de las dos estrategias utilizadas en este trabajo para la amplificación de los extremos 5' y 3'. A, estrategia de amplificación con el oligonucleótido dT, diseñado para unirse específicamente al extremo interno de la cola poli(A) o tracto poliadenilado. Denotada con V la posición 3' del oligonucleótido dT por ser variable (adenina, guanina o citosina); con N el extremo 5' no complementario a la zona de adeninas que se incluye para facilitar una posterior reamplificación de los productos de PCR obtenidos. B, estrategia de amplificación adaptada a partir de (Varkonyi-Gasic et al., 2007) en la que se utiliza el oligonucleótido *stem-loop* con el extremo 3' con una secuencia de timinas que asegura la unión al extremo 3' de la molécula de RNA.

3.8. NORTHERN BLOT

Las muestras de RNA total de plantas infectadas junto con las fracciones purificadas de virión y dsRNA se separaron en geles desnaturalizante de formaldehído al 1.5 % de agarosa y se transfirieron a membranas de nylon cargadas positivamente (ROCHE) (Sambrook et al., 1989). Las membranas, conteniendo los ácidos nucleicos, se dejaron secar unos minutos y se fijaron covalentemente por exposición a radiación UV (700 x 100 $\mu J/cm^2$), procediendo a revelarlas tal y como se ha descrito en la sección 3.2.1.

3.9. ESTUDIO BIOINFORMÁTICO

Los estudios de filogenia se realizaron a partir de secuencias de nucleótidos y aminoácidos, utilizando en ambos casos las secuencias obtenidas en el presente trabajo para el CLV (Anexos I y II) y recopilando de las bases de datos del GenBank las correspondientes al resto de carlavirus. Primero se realizó un alineamiento de secuencias múltiple con el programa ClustalX2 (Larkin et al., 2007), con los parámetros por defecto del programa. Del alineamiento se infirió la historia filogenética y se representó en forma de árboles filogenéticos en el programa MEGA5.2 (Tamura et al., 2011) mediante el método de *neighbor-joining* (Saitou and Nei, 1987), utilizando un test de *bootstrap* (Felsenstein, 1985) con 10000 réplicas, calculando las distancias evolutivas con el método de corrección de Poisson (Zuckerkandl and Pauling, 1965). Se incluyeron las secuencias correspondientes al *Apricot latent virus* como *outgroup*.

4. RESULTADOS Y DISCUSIÓN

4.1. CARACTERIZACIÓN DE LA ESTRUCTURA PRIMARIA DEL *VIRUS LATENTE DEL CLAVEL*

Para determinar la secuencia de nucleótidos completa del genoma del CLV se procedió a diseñar oligonucleótidos específicos en base a la secuencia parcial del CLV conocida, así como oligonucleótidos degenerados, en base a regiones conservadas en los 4 miembros del género *Carlavirus* más cercanos filogenéticamente al CLV (sección 3.5.). En el diseño de los oligonucleótidos se tuvo en cuenta que la amplificación por RT-PCR de los fragmentos del virus generase amplicones parcialmente solapantes, a partir de las cuales se pudiera reconstruir el genoma completo del CLV.

El siguiente paso se procedió a la amplificación mediante RT-PCR, de los diferentes fragmentos de cDNAs, utilizando como muestras extractos de RNA totales o fracciones purificadas de RNAs de doble cadena, obtenidas de plantas de *C. quinoa* infectadas con CLV. Los cinco amplicones obtenidos (Figura 8) se clonaron en el plásmido pTZ57R/T, procediendo posteriormente a determinar la secuencia de ácidos nucleicos en el Servicio de secuenciación del DNA y análisis de la expresión génica del Instituto de Biología Molecular y Celular de Plantas (IBMCP). Para aquellos clones en donde no se pudo determinar la secuencia completa del cDNA mediante los cebadores que flanquean la región de clonaje múltiple, se procedió a diseñar progresivamente nuevos oligonucleótidos específicos en base a la nueva secuencia que se iba obteniendo, pudiendo cubrir de esta forma el genoma completo del CLV, a excepción de los extremos 5' y 3'. La Figura 8 muestra un esquema del genoma viral junto la estrategia de clonación de los distintos fragmentos del virus.

Mediante el método RACE (*Rapid Amplification of cDNA Ends*) se determinaron las secuencias de los extremos 5' y 3', utilizando para ello las parejas de cebadores

oligo-dT/2696 y 2705/oligo-dT, respectivamente (Figura 8, B). Posteriormente, se volvieron a secuenciar ambos extremos mediante un método alternativo adaptado de un protocolo para la caracterización de miRNAs (Varkonyi-Gasic et al., 2007), utilizando un oligonucleótido *stem-loop* (ver sección 3.7.) en lugar del cebador oligo–dT (protocolo *stem-loop* oligo-dT; Sección 3.7.). La secuencia de nucleótidos obtenida con ambas técnicas fue idéntica, a excepción del nucleótido inicial y final de los extremos 5' y 3', respectivamente. Dado que el oligo-dT del método RACE presenta una posición variable en su extremo 3', y que el método adaptado de (Varkonyi-Gasic et al., 2007) no tiene posibilidad de interferir en esa posición, tomamos las secuencias obtenidas por este último método como las válidas. La variación de secuencias observada entre ambos métodos pone también de manifiesto que el método RACE, ampliamente utilizado para caracterizar los extremos de diferentes moléculas de RNA, está sujeto a errores en los nucleótidos iniciales, algo especialmente relevante en secuencias de RNA virales, en donde se localizan los promotores de las polimerasas virales (Bol, 2005; Flick et al., 1996; Surana et al., 2014).

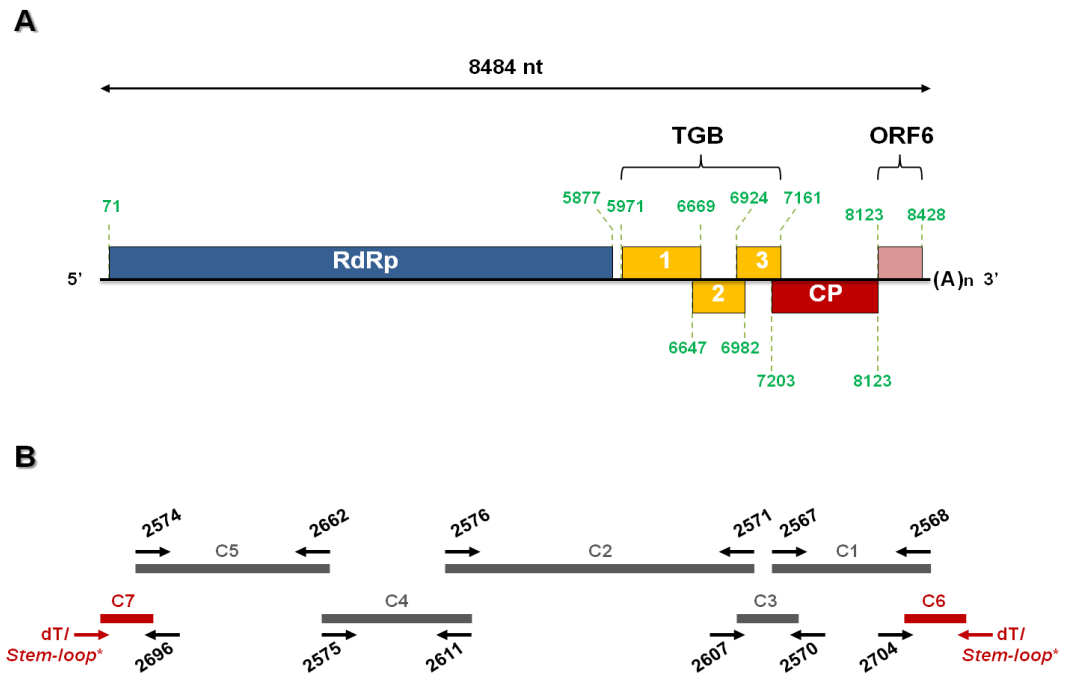


FIGURA 8. Organización genómica del CLV. A, diagrama esquemático del genoma del CLV: la línea continua representa el genoma de RNA, las cajas representan los ORFs,

cuyos probables productos se indican como RdRp (RNA polimerasa RNA dependiente), TGB 1, 2 y 3 (*Triple Gene Block*), CP (proteína de cubierta) y ORF6. En verde se indican la posición de los nucleótidos iniciales y finales para cada uno de los ORFs. B, estrategia de clonación genómica y localización de los clones de cDNA utilizados para secuenciar el CLV: con C1-C5 se representan las secuencias generadas por RT-PCR a partir de las combinaciones de oligonucleótidos específicos y degenerados; C6 y C7 representan los clones terminales 3' y 5' generados por RACE o mediante el método *stem-loop* oligo-dT. Los números denotan la referencia del oligonucleótido empleado en cada caso (Tabla 1).

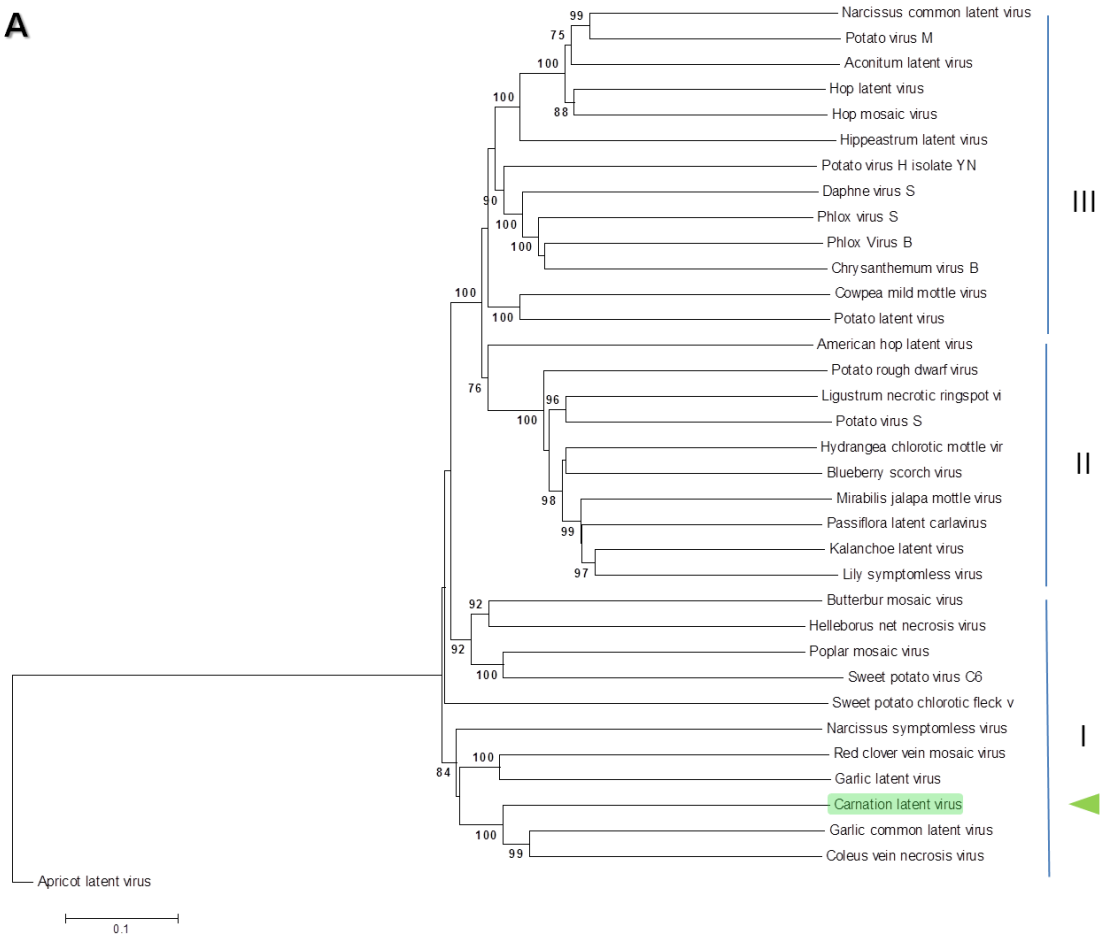
Debido a la dilación y dificultad de obtener las amplificaciones esperadas mediante RT-PCR de los fragmentos del genoma del CLV para su clonación y caracterización, se utilizaron diversas fuentes de RNA viral, desde extractos de RNA total realizados sobre el tejido liofilizado recibido de la colección alemana DSMZ en primera instancia, pasando por extractos de RNA total de tejido fresco de *C. quinoa* infectado, hasta la purificación de dsRNA. La diversa procedencia de los fragmentos amplificados y secuenciados implicaba que la secuencia completa de CLV obtenida inicialmente pudiera presentar variaciones de residuos no presentes en el mismo aislado viral. Por ello y una vez conocidos los extremos 5' y 3' del CLV, decidimos realizar la amplificación completa del virus utilizando oligonucleótidos directo y reverso diseñados en base a ambos extremos. Mediante RT-PCR se amplificaron las 8484 pares de bases del cDNA en una única reacción. Con el objetivo de aprovechar la clonación en posteriores experimentos de infectividad, se clonó en un casete de expresión con promotor 35S del *Virus del mosaico de la coliflor* y terminador PoPiT del gen inhibidor de proteinasa II de *S. tuberosum* en un vector pUC18 modificado previamente (ver secciones 3.1.4. y 3.6.7.). Una vez clonado se obtuvo la secuencia del genoma completo utilizando los oligonucleótidos específicos previamente diseñados. De la comparación con la secuencia previamente obtenida se encontraron varios cambios puntuales a lo largo del genoma que, sin embargo, no afectaban a la estructura interna del mismo.

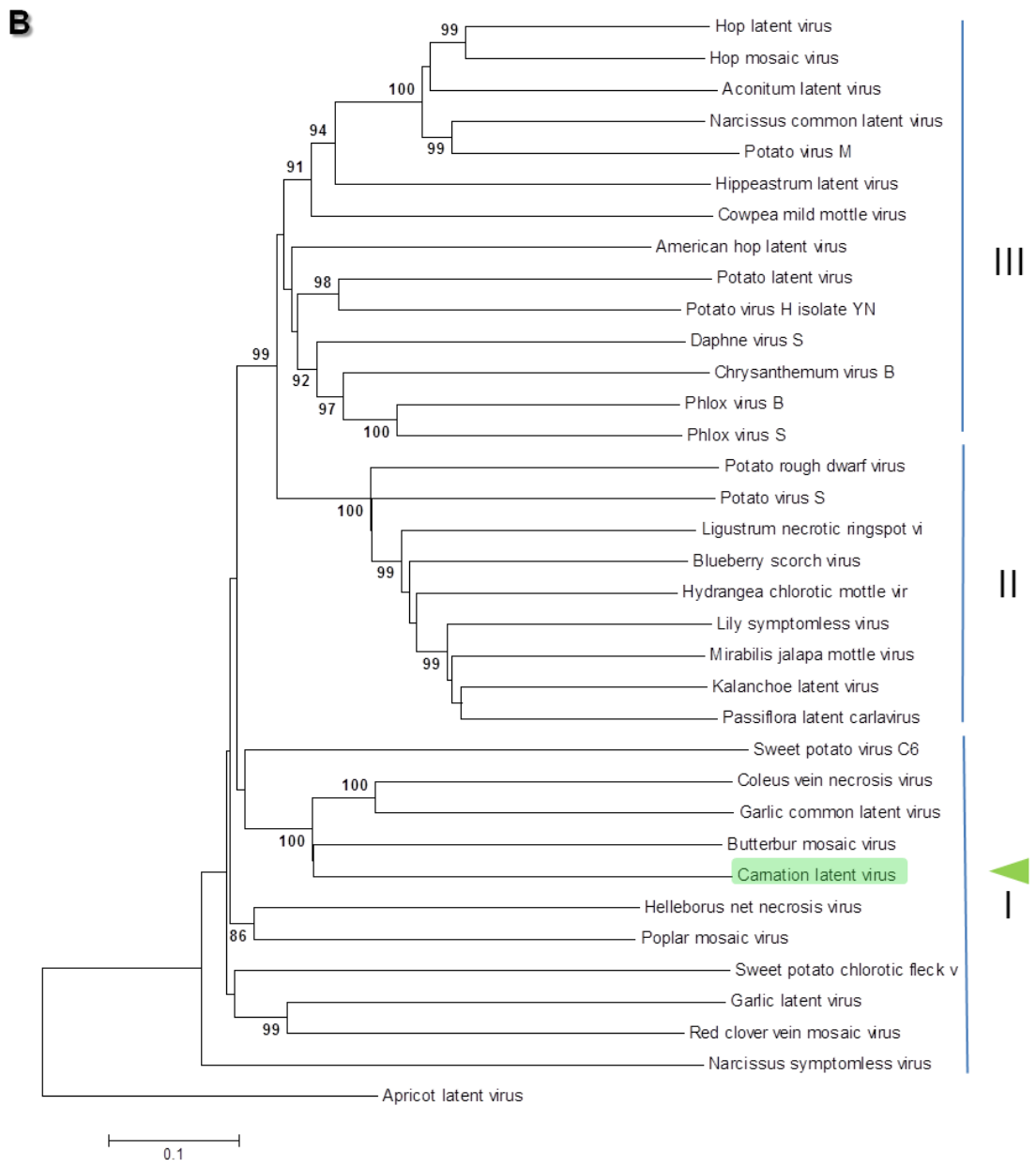
Del análisis de la secuencia obtenida se pudo demostrar que comparte la misma organización genómica que el resto de miembros del género *Carlavirus*. La longitud total del genoma del CLV, excluyendo la cola poli(A), es de 8484 nucleótidos, y contiene un total de 6 ORFs adyacentes o solapantes. El ORF situado en el extremo 5' comienza a 71 nt y se extiende hasta el nt 5877. Utilizando el análisis de dominios conservados del Conserved Domain Database del NCBI (http://www.ncbi.nlm.nih.gov/cdd/) se determinó que este ORF expresa una proteína de replicación de 1959 aminoácidos y 222,6 kDa, que contiene los dominios metiltransferasa, 2OG-Fe(II) oxigenasa, cisteína proteasa tipo-OTU, endopeptidasa de carlavirus, helicasa y RNA-polimerasa RNA-dependiente. Aguas abajo del primer ORF aparecen 3 ORFs solapantes, desde las posiciones 5971 a 6669, 6647 a 6982, y 6924 hasta 7161, correspondientes a tres proteínas de 233, 112, y 66 aminoácidos de 26,0, 12,0, y 6,9 kDa, respectivamente. Estos forman el *triple gene block*, con probable implicación en el movimiento célula a célula y a larga distancia del CLV (King et al., 2012), como se ha descrito en el resto de carlavirus. Solapando con el TGB3 aparece el quinto ORF, desde el nt 7203 hasta el nt 8123, que expresa la proteína de cubierta de 307 aminoácidos y 39,0 kDa de masa molecular. Por último, en el extremo 3' del genoma se encuentra el denominado ORF6, desde 8123 nt a 8428 nt, que expresa una proteína de 102 aminoácidos y 16,0 kDa para la que se predice un dominio similar a zinc-finger por contener un motivo de cuatro cisteínas C-X_2-C-X_9-C-X_4-C, que podría servir para la unión de RNA y que, además, en otros carlavirus se ha demostrado que actúa como supresor del silenciamiento génico mediado por RNA en la planta hospedadora (Zhou et al., 2006). Las regiones UTR 5' y 3' constan de 70 y 56 nt, respectivamente. Todas estas características son similares al resto de especies dentro del género *Carlavirus* (King et al., 2012).

4.2. ANÁLISIS FILOGENÉTICO DEL *VIRUS LATENTE DEL CLAVEL*

Con la intención de comprender la relación entre el CLV y el resto de carlavirus estudiados hasta la fecha se llevaron a cabo estudios de filogenia utilizando tanto la secuencia de nucleótidos del genoma completo, como las de aminoácidos de las proteínas de replicación y cubierta.

Se construyeron árboles filogenéticos utilizando el alineamiento de secuencias múltiples con el programa CLUSTALX2, y posteriormente se estudiaron las distancias filogenéticas y visualizaron los árboles con el programa MEGA (versión 5.2). En la Figura 9 aparecen los árboles obtenidos. En el árbol generado a partir del alineamiento de las secuencias de nucleótidos del genoma completo de los *Carlavirus* se pone de manifiesto la presencia de tres grupos bien diferenciados (I, II y III), con valores de *bootstrap* superiores al 87 %, en donde CLV quedaría englobado dentro del grupo I en íntima relación el *Virus de la necrosis de la nervadura del cóleo* y el *Virus común latente del ajo* (Figura 9, A). Este conjunto se agrupa próximo a otros virus como el *Virus latente del ajo*, el *Virus asintomático del narciso* y el *Virus del mosaico de la nervadura del trébol rojo*. Resultados similares se obtuvieron cuando se utilizó la secuencias de aminoácidos de la replicasa viral (Figura 9, B) o de cubierta (Figura 9, C) indicando la solidez de los tres grupos filogenéticos observados. Similar agrupamiento ha sido descrito recientemente utilizando la secuencia de aminos ácidos de la replicasa viral y de la proteína de cubierta (Li et al., 2013).


A
Narcissus common latent virus
Potato virus M
Aconitum latent virus
Hop latent virus
Hop mosaic virus
Hippeastrum latent virus
Potato virus H isolate YN
Daphne virus S
Phlox virus S
Phlox Virus B
Chrysanthemum virus B
Cowpea mild mottle virus
Potato latent virus
American hop latent virus
Potato rough dwarf virus
Ligustrum necrotic ringspot vi
Potato virus S
Hydrangea chlorotic mottle vir
Blueberry scorch virus
Mirabilis jalapa mottle virus
Passiflora latent carlavirus
Kalanchoe latent virus
Lily symptomless virus
Butterbur mosaic virus
Helleborus net necrosis virus
Poplar mosaic virus
Sweet potato virus C6
Sweet potato chlorotic fleck v
Narcissus symptomless virus
Red clover vein mosaic virus
Garlic latent virus
Carnation latent virus
Garlic common latent virus
Coleus vein necrosis virus
Apricot latent virus
III
II
I
0.1


B
Hop latent virus
Hop mosaic virus
Aconitum latent virus
Narcissus common latent virus
Potato virus M
Hippeastrum latent virus
Cowpea mild mottle virus
American hop latent virus
Potato latent virus
Potato virus H isolate YN
Daphne virus S
Chrysanthemum virus B
Phlox virus B
Phlox virus S
Potato rough dwarf virus
Potato virus S
Ligustrum necrotic ringspot vi
Blueberry scorch virus
Hydrangea chlorotic mottle vir
Lily symptomless virus
Mirabilis jalapa mottle virus
Kalanchoe latent virus
Passiflora latent carlavirus
Sweet potato virus C6
Coleus vein necrosis virus
Garlic common latent virus
Butterbur mosaic virus
Carnation latent virus
Helleborus net necrosis virus
Poplar mosaic virus
Sweet potato chlorotic fleck v
Garlic latent virus
Red clover vein mosaic virus
Narcissus symptomless virus
Apricot latent virus
III
II
I
0.1

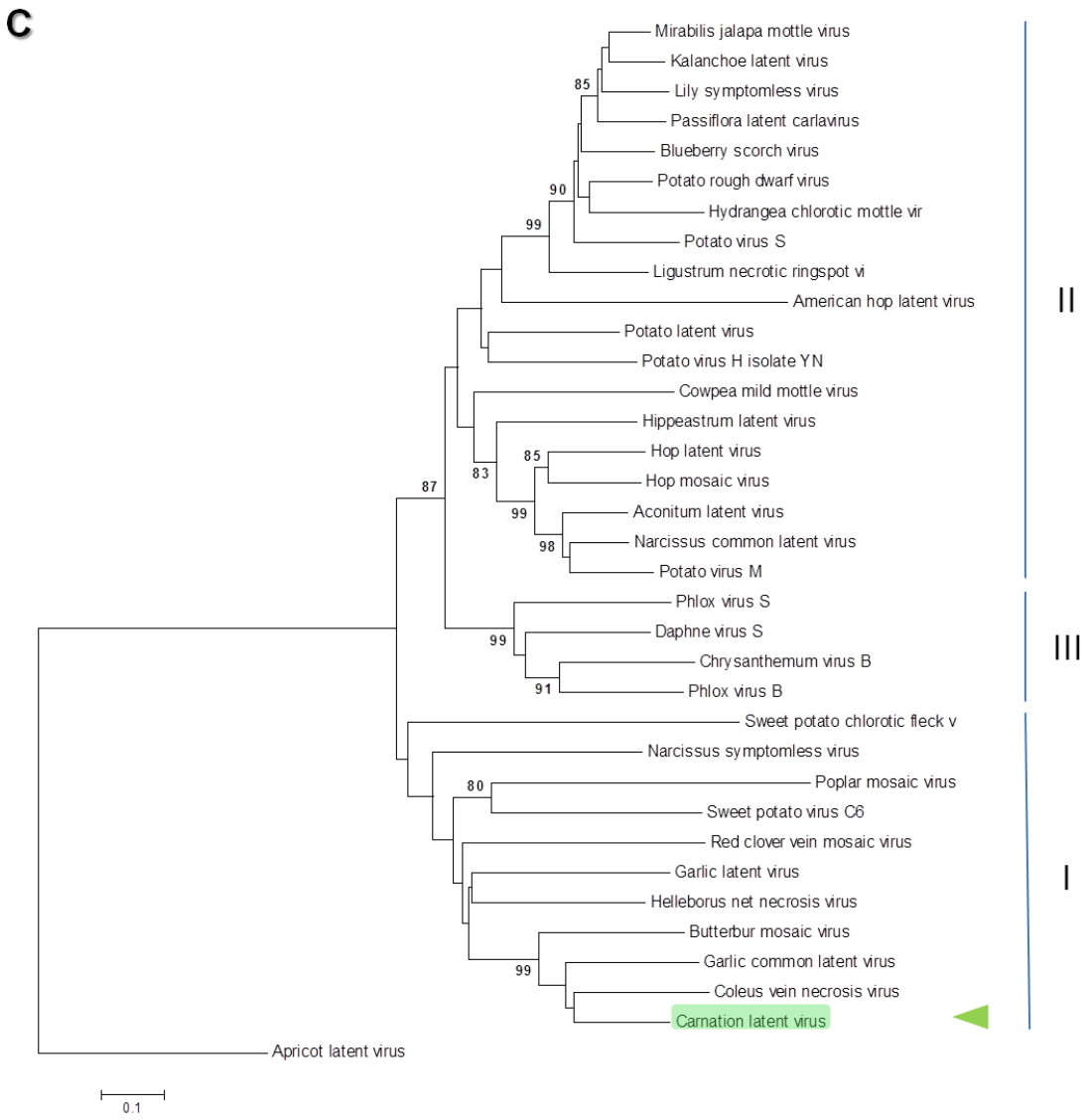


FIGURA 9. Análisis filogenético del CLV. Se muestran la localización del CLV y sus relaciones con el resto de virus del género *Carlavirus* en base a A, las secuencias de nucleótidos del genoma completo, B, las secuencias de aminoácidos de la proteína de replicación viral, y C, las secuencias de aminoácidos de la proteína de cubierta. Las especies de los virus aquí estudiados y sus números de accesión del GenBank son estos: *Aconitum latent virus* (NC_002795), *American hop latent virus* (KF964324), *Blueberry scorch virus* (NC_003499), *Chrysanthemum virus B* (NC_009087), *Coleus vein necrosis virus* (NC_009764), *Cowpea mild mottle virus* (NC_014730), *Daphne virus S* (NC_008020), *Garlic latent virus* (NC_003557), *Helleborus net necrosis virus* (NC_012038), *Hippeastrum latent virus* (NC_011540), *Hop latent virus* (NC_002552), *Hop mosaic virus* (NC_010538), *Kalanchoe latent virus* (NC_013006), *Ligustrum necrotic ringspot virus* (NC_010305), *Lily symptomless virus* (NC_005138), *Narcissus common latent virus* (NC_008266), *Narcissus symptomless virus* (NC_008552), *Passiflora latent carlavirus* (NC_008292), *Phlox virus B* (NC_009991), *Phlox virus S* (NC_009383), *Poplar mosaic virus* (NC_005343), *Potato latent virus* (NC_011525), *Potato rough dwarf virus* (NC_009759), *Potato virus M* (NC_001361), *Potato virus S*

(NC_007289), *Red clover vein mosaic virus* (NC_012210*), Sweet potato chlorotic fleck virus* (NC_006550), *Sweet potato virus C6* (NC_018448), *Mirabilis jalapa mottle virus* (NC_016080), *Garlic common latent virus* (NC_016440), *Butterbur mosaic virus* (NC_01352), *Hydrangea chlorotic mottle virus* (NC_012869), *Potato virus H isolate YN* (NC_018175), *Apricot latent virus* (NC_014821). Se realizaron los alineamientos de secuencias utilizando el programa CLUSTALX2 con los parámetros por defecto, los árboles filogenéticos se construyeron con el método de *neighbor-joining* (NJ) y se evaluó mediante *bootstrapping* con 10000 réplicas utilizando para todo ello el programa MEGA (versión 5.2). Se utilizó las secuencias del *Apricot latent virus* como *outgroup* para la generación de los árboles. En las figuras aparecen los porcentajes del *bootstrap* superiores al 75% en las ramas correspondientes. Sombreado en verde y señalado con la cuña la posición del CLV en los árboles filogenéticos.

4.3. ESTUDIO DE REGIONES REGULADORAS ESENCIALES PARA EL CICLO DE INFECCIÓN DEL *VIRUS LATENTE DEL CLAVEL*

4.3.1. Determinación del inicio de transcripción de los RNA subgenómicos del *Virus latente del clavel*

Para el CLV, al igual que para el resto de las especies del género *Carlavirus*, se predice la síntesis de sgRNAs como parte de su ciclo de replicación (ver secciones 1.4.2. y 1.5.) y su producción se entiende como un mecanismo utilizado por los virus para acercar el extremo 5' del RNA al comienzo de ORFs, facilitando de esta manera *la traducción de los mismos. En publicaciones recientes se han descrito elementos que formarían parte del hipotético promotor para la síntesis de los sgRNAs. Se trata de elementos en *cis* constituidos por un hexanucleótido (5'-UUAGGU-3') que aparece aguas arriba del inicio de ORFs que se traducen a partir de sgRNAs en otros virus. El motivo está altamente conservado entre los miembros de las familias *Alphaflexiviridae* y *Betaflexiviridae*, lo cual pone de manifiesto su posible relevancia a la hora de determinar los inicios de transcripción de sgRNAs (Komatsu et al., 2012).

Tras caracterizar la estructura primaria del CLV (ver sección 4.1.), procedimos a determinar la presencia y localización del hexanucleótido 5'-UUAGGU-3' en el genoma viral. El motivo aparece 4 veces, comenzando en las posiciones 830 nt, 5529

nt, 5944 nt y 7158 nt (Figura 10, A). Las dos posiciones situadas más cerca del extremo 5' del virus (830 nt y 5529 nt) se localizan a lo largo del ORF correspondiente a la replicasa viral; mientras que las otras dos (5944 nt y 7158 nt), solapan con el final de los ORF de la replicasa y el tercer gen del TGB, respectivamente, y a pocos nucleótidos de distancia del siguiente ORF (21 nt del primer gen del TGB, y 39 nt del gen de la CP, respectivamente). También se analizó la presencia del hexanucleótido en los genomas del resto de carlavirus. Para ello recurrimos a un alineamiento múltiple de secuencias genómicas de los carlavirus (ver secciones 3.9. y 4.2.) observando que el motivo 5'-UUAGGU-3' aparece altamente conservado en dos posiciones a lo largo del alineamiento. Estas posiciones se localizan pocos nucleótidos aguas arriba del comienzo de los ORFs del primer gen del TGB y de la CP.

Con la intención de obtener evidencias de la producción de sgRNAs por parte del CLV durante su ciclo de replicación, realizamos un ensayo Northern blot sobre extracciones de RNA total de tejido de *C. quinoa* y clavel (*Dianthus barbatus* L.) infectados con CLV, fracciones enriquecidas con dsRNA y muestras de RNA total extraído a partir de viriones. Las membranas se hibridaron con una sonda RNA, marcada con digoxinenina y complementaria a los 393 nucleótidos finales del extremo 3' del CLV (Sanchez-Navarro et al., 1999). Dado que todos los sgRNAs se transcriben hasta el extremo final 3' del virus, la sonda permite la detección tanto del RNA genómico como de los diferentes RNAs subgenómicos. Como se muestra en la Figura 10 (B), las preparaciones de RNA total de *C. quinoa* (menor señal) y clavel infectados, así como la fracción enriquecida de dsRNA, presentaron una señal clara de hibridación para un fragmento de RNA de aproximadamente 8 - 9 kb, que se correspondería con el RNA genómico completo del CLV. Además se detectaron 3 bandas de menor tamaño: una banda de muy reducida intensidad de 2,5 - 2,7 kb; otra banda muy definida de 1,2 - 1,5 kb; y una última banda de 600 - 700 nt. Los tamaños de los sgRNAs que se espera que el CLV genere son ligeramente superiores a 2514 nt y

1282 nt, distancias desde los comienzos de los ORFs hasta el final 3' del genoma viral. Por todo esto parece plausible que las bandas de 2,5 - 2,7 kb y 1,2 - 1,5 kb detectadas se correspondan con los sgRNAs de CLV, existiendo claras diferencias en los niveles de acumulación para ambas especies. No se observaron bandas inespecíficas en el control negativo correspondiente al extracto de RNA total procedentes de de *C. quinoa* no inoculado. Sin embargo, tampoco se observaron bandas en la fracción de RNAs extraídos de viriones, a pesar de confirmar que el stock utilizado es infectivo mediante inoculación mecánica y de detectar la presencia del virus mediante RT-PCR, reflejando posiblemente una baja concentración de los RNA virales. (Figura 10, C).

Teniendo en consideración que i) para los miembros del género *Carlavirus* se propone la producción de dos sgRNAs (para la traducción del TGB, y para la traducción de la CP y ORF6) (King et al., 2012), ii) que el hexanucleótido arriba comentado aparece a pocos nucléotidos de distancia de los comienzos de los siguientes ORFs (Komatsu et al., 2012), iii) que solamente los motivos en el genoma del CLV en posiciones 5944 nt y 7158 nt están conservados en el género *Carlavirus*, y iv) que mediante Northern blot conseguimos detectar la presencia de RNA de tamaños cercanos a los que presentarían los sgRNAs producidos a partir de esas posiciones, desde este trabajo nos planteamos determinar los comienzos de transcripción de los sgRNA del CLV.

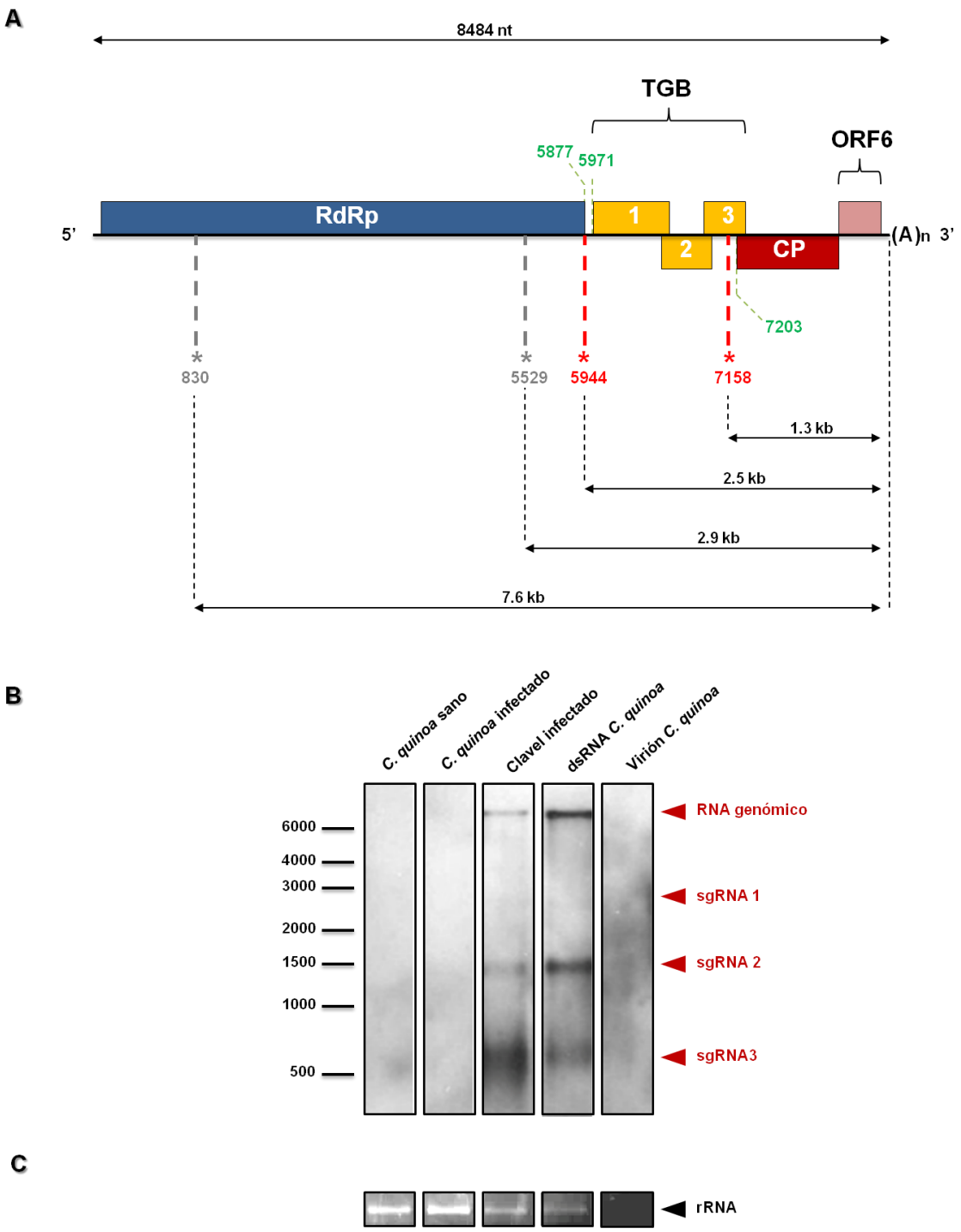


FIGURA 10. Estudio de los RNA subgenómicos del CLV. A, diagrama esquemático en el que se muestra la posición en el genoma de CLV de los motivos 5'-UUAGGA-3' (asteriscos). Anotados en color rojo aquellas posiciones conservadas en los carlavirus. B, hibridación de la sonda diseñada en base a los últimos 393 nt del extremo 3' del CLV con muestras correspondientes a extractos de RNA total de plantas de *C. quinoa* no inoculadas, infectadas, plantas de clavel (*Dianthus barbatus*, L.) infectadas, y muestras de las fracciones purificadas de dsRNA y viriones extraídas a partir de tejido infectado de *C. quinoa*. Señaladas con las cuñas las bandas detectadas en la hibridación correspondientes al RNA genómico, y subgenómicos. C, control de carga del RNA en el ensayo de Northern blot.

La aproximación seguida se basó en el método *stem-loop* oligo-dT para la amplificación de extremos del cDNA que adaptamos de (Varkonyi-Gasic et al., 2007) y que utilizamos para la determinación de la secuencia del extremo 5' del genoma del CLV (ver secciones 3.7. y 4.1.). Sobre la muestra de la fracción purificada de dsRNA se llevó a cabo una reacción de poliadenilación de los extremos 3' del RNA, seguida por una amplificación mediante RT-PCR utilizando el oligonucleótido *stem-loop* y oligonucleótidos específicos reversos diseñados a 530 nt del supuesto comienzo del sgRNA1, y a 130 nt en el caso del sgRNA2 (Figura 11, A). Los productos de PCR generados del tamaño correspondiente a los esperados se clonaron y secuenciaron. Las secuencias obtenidas se correspondían con los inicios de los ORFs esperados, pudiéndose determinar que los sgRNAs de CLV se transcriben a partir de los nucleótidos 5965 (sgRNA1) y 7199 (sgRNA2). Estas posiciones se encuentran entre los motivos 5'-UUAGGU-3' y los inicios de los siguientes ORFs, tal y como se ha descrito para los sgRNAs de virus pertenecientes al género *Carlavirus*. De forma adicional, y motivado por la detección de una banda de menor tamaño (600 nt) en el ensayo Northern blot en aquellas muestras en las que se detectaban las formas genómicas y subgenómicas del CLV, se decidió estudiar la existencia de un posible tercer sgRNA de menor tamaño que contendría el ORF6 del CLV. Se siguió un procedimiento análogo al descrito aquí para los sgRNA1 y 2, utilizando un oligonucleótido reverso situado a unos 300 nt del posible origen del sgRNA (región 8160-8180 nt del genoma viral). En este caso no obtuvimos el producto de PCR esperado (en torno a 300 pb) sino una banda de aproximadamente 1,2 kb (Figura 11, B), que con gran probabilidad se corresponde con el comienzo del sgRNA2. Aunque no se puede descartar por completo la existencia de un sgRNA de menor tamaño, nuestros resultados hasta el momento indican lo contrario.

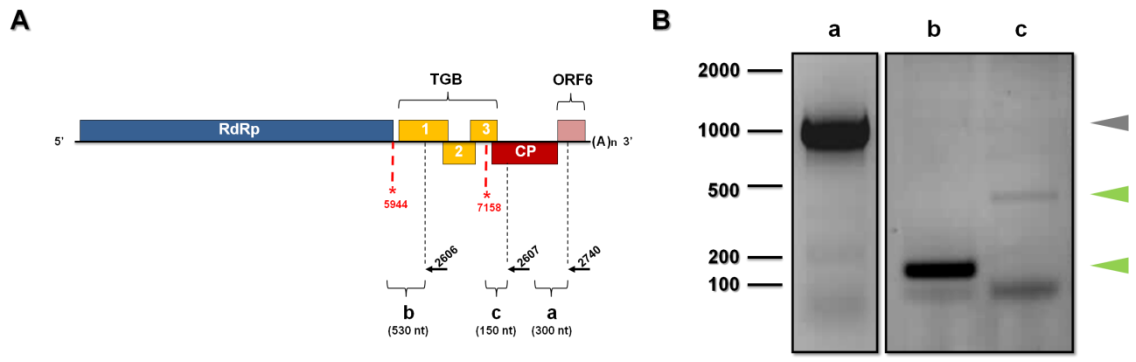


FIGURA 11. Determinación de los inicios de transcripción de los sgRNAs de CLV. A, esquema de la disposición de los posibles inicios de transcripción y los oligonucleótidos reversos (Ver Tabla 1) utilizados para su amplificación. B, fotografía de la electroforesis en gel de agarosa correspondiente al análisis de los amplicones obtenidos para los extremos 5' de los sgRNA. Marcados con cuñas verde aquellas bandas que se corresponden con los productos de amplificación esperados y que se clonaron posteriormente. Con una cuña gris se marca el producto mayoritario del intento de amplificación del extremo 5' del hipotético sgRNA3.

Todos estos resultados tomados en conjunto nos indican que el CLV sintetiza en su ciclo de replicación, al menos dos sgRNAs de 2520 nt y 1286 nt, respectivamente, que comienzan a transcribirse desde las posiciones 5965 nt para el sgRNA1, y 7199 nt para el sgRNA2, y cuya transcripción podría estar dirigida por los hexanucleótidos 5'-UUAGGU-3' situados aguas arriba de los mismos en las posiciones 5944 nt y 7158 nt. El resto de ocasiones en que aparece este motivo disperso en el genoma de los carlavirus, si bien no se ajusta al modelo propuesto como motivo que precede al inicio del sgRNA, no podemos descartar que estén vinculados de alguna manera en la síntesis de los dos sgRNAs caracterizados (ej. estructura secundaria del RNA, reclutamiento complejo replicasa viral, etc), un aspecto que se analizará una vez obtenido el clon infectivo.

4.3.2. Estimación de la longitud de la cola poli(A) del *Virus latente del clavel*

Las secuencias de las regiones 3' no traducibles (3'-UTR) de virus de RNA de polaridad positiva desempeñan un papel de gran importancia en la amplificación del

RNA durante el ciclo replicativo del virus, ya sea por terminar en una estructura tipo tRNA o en una cola poli(A) (Eggen et al., 1989; Sriskanda et al., 1996).

Como ocurre con el mRNA eucariótico, se espera que la cola poli(A) de los carlavirus desempeñe un papel en la estabilización del RNA, en la iniciación de la traducción, e incluso en el reconocimiento de ciertos elementos del complejo de la replicasa. De hecho hay experimentos en otros virus de RNA de polaridad positiva en los que se demuestra cómo versiones de los mismos con la cola poli(A) truncada pierden gran parte o toda la infectividad (Eggen et al., 1989; Guilford et al., 1991). A día de hoy no hay datos disponibles acerca del tamaño de la cola poli(A) en miembros del género *Carlavirus* y prácticamente nula en el resto de virus de plantas. Ya que se trata de un atributo de gran relevancia para el ciclo de replicación del virus desde este trabajo nos planteamos estimar la longitud de la cola poli(A) del CLV.

La estrategia con la que se llevó a cabo la estimación se basa en el método de caracterización de miRNAs (Varkonyi-Gasic et al., 2007) que adaptamos en este trabajo (ver sección 3.7.), con el que, mediante una reacción de RT-PCR podemos amplificar los extremos 3' poliadenilados de RNA. Por ser un elemento implicado en la replicación viral del cual se desconoce el mecanismo por el que es añadido al extremo 3' del virus o por el cual se replica en conjunto con el resto del genoma viral, utilizamos RNA viral procedente de tres muestras distintas en cuanto la naturaleza y estadio del ciclo replicativo del virus: la fracción purificada de dsRNA en la que predomina la formas intermediarias de la replicación de dobles cadenas, la purificación de viriones en la que se encuentra aquellas moléculas de RNA viral que son encapsidadas, y el extracto de RNA total de plantas de *C. quinoa* infectadas.

Para la amplificación del extremo 3' en conjunto con la cola poli(A) se utilizó la combinación de oligonucleótidos 2704 (el cual anilla a 200 nt del extremo 3' del genoma viral) y *stem-loop* oligo-dT, que por su diseño con estructura secundaria se

une al extremo final de la cola poli(A) (Figura 12, A). Como se observa en la Figura 12 (B), de las reacciones de RT-PCR utilizando las tres muestras diferentes antes comentadas, se obtuvo un resultado similar. El producto de amplificación mayoritario fue una banda de un tamaño ligeramente superior a los 300 pb, en el que se incluyen los 189 pb del extremo 3' del genoma viral, además de los oligonucleótidos empleados para su amplificación. Teniendo todo lo anterior en cuenta, el tamaño que se estima mediante esta aproximación para la cola poli(A) del CLV es de 75 - 100 nt.

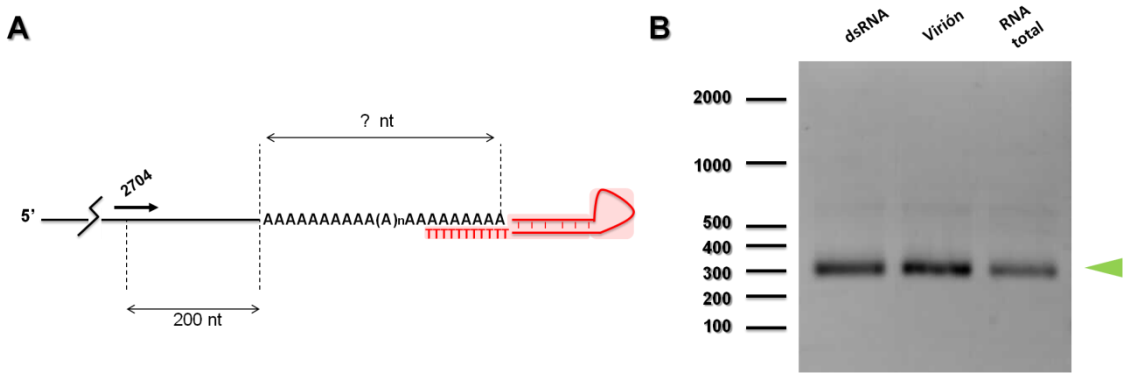


FIGURA 12. Estimación de la longitud de la cola poli(A) del CLV. A, diagrama de la estrategia seguida. Representado con una flecha el oligonucleótido directo utilizado, y anotado su número de referencia (Tabla 1). Representado en rojo el oligonucleótido *stem-loop* que se utilizó como oligonucleótido reverso para la reacción de RT-PCR. B, electroforesis en gel de agarosa de los productos de la RT-PCR sobre las muestras de dsRNA, virión y RNA total. Marcado con una cuña el producto de amplificación obtenido mayoritario.

Debido a la baja resolución de los geles de agarosa utilizados en este ensayo y a que la secuenciación de regiones con alto contenido de nucleótidos repetidos afecta a las reacciones de secuenciación, no pudimos determinar con mayor exactitud el tamaño. Sería interesante repetir el ensayo, analizando los productos de RT-PCR en geles de acrilamida con un porcentaje de acrilamida óptimo para la resolución de fragmentos de tamaños cercanos a las 250 - 350 pb. De esta manera podríamos determinar si efectivamente, como indican nuestros resultados, no se observan

diferencias significativas en los tamaños de cola poli(A) en los estados del ciclo de replicación que representan las distintas muestras utilizadas.

5. PERSPECTIVAS

En el presente trabajo se describen los primeros pasos necesarios para la caracterización molecular de un virus de plantas. Como se puso de manifiesto en la sección 2., los objetivos a largo plazo del proyecto, dentro del cual se engloba este Trabajo de Fin de Máster, incluyen el desarrollo de un vector viral diseñado a partir del *Virus latente del clavel*, con posibles aplicaciones dentro de la protección cruzada o expresión de proteínas recombinantes en las plantas hospedadoras. En este trabajo se ha determinado la secuencia completa del genoma del virus y, además, se ha clonado el cDNA genómico completo en un plásmido preparado para experimentos de expresión transitoria. Los siguientes pasos a seguir dentro del proyecto de generación del vector viral incluyen, i) clonar el cDNA genómico completo en plásmidos preparados para experimentos de transcripción *in vitro*, ii) llevar a cabo ensayos de infectividad por agroinfiltración e inoculación mecánica de los transcritos del virus, iii) diseñar las estrategias de modificación del genoma del virus para las aplicaciones antes comentadas (determinar las posiciones en dónde poder incluir secuencias de al menos 21 nucleótidos de virus que provocan pérdidas en el cultivo de clavel, como CarMV, por ejemplo), iv) llevar a cabo las modificaciones y probar que no suponen un perjuicio en la capacidad de infección y sintomatología del virus, y v) ensayar la eficacia del vector viral en la aplicación concreta de interés.

6. CONCLUSIONES

1. A partir de la secuencia completa de nucleótidos del genoma del *Virus latente del clavel* se pudo determinar que este virus presenta una organización genómica típica de los miembros del género *Carlavirus*, con un genoma de 8484 residuos de nucleótido de tamaño que consta de seis marcos de lectura abiertos correspondientes a la proteína de replicación, las proteínas implicadas en el movimiento viral, la proteína de cubierta y la proteína de unión a RNA.
2. Los resultados del análisis filogenético dentro del género *Carlavirus* de las secuencias genómicas, y de las proteínas de replicación y cubierta pusieron de manifiesto que los virus pertenecientes a este género viral se distribuyen en tres grupos bien definidos, en donde el CLV aparece íntimamente relacionado con el *Virus de la necrosis de la nervadura del cóleo* y el *Virus latente del ajo*.
3. La determinación de la secuencia de los extremos 5' y 3' mediante el protocolo RACE y el protocolo *stem-loop* oligo-dT, puesto a punto en el presente trabajo, permitió determinar la secuencia completa de los extremos del RNA viral así como constatar la presencia de errores en la determinación de la secuencias mediante el protocolo RACE.
4. Con el estudio acerca de la conservación posicional de motivos promotores de la transcripción de RNAs subgenómicos dentro del género *Carlavirus*, el ensayo de Northern blot y la posterior caracterización molecular por amplificación de los extremos 5', se determinó que el CLV produce dos sgRNAs durante su ciclo de infección, de 2,5 y 1,2 kb, cuyos inicios de transcripción se localizan en las posiciones 5965 nt y 7199 nt de su genoma, respectivamente.
5. Mediante amplificación por RT-PCR del extremo 3' del RNA del *Virus latente del clavel* mediante el protocolo *stem-loop* oligo-dT, se pudo estimar la longitud de la cola poli(A) entre 75 y 100 residuos de nucleótido.

7. REFERENCIAS

Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res (1997) 25:3389-3402.

Baulcombe DC. Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol (1999) 2:109-113.

Bawden FC. Studies on a Virus Causing Foliar Necrosis of the Potato. roceedings of the Royal Society of London. Series B, Biological Sciences (1934) 116:20.

Bol JF. Replication of alfamo- and ilarviruses: role of the coat protein. Annu Rev Phytopathol (2005) 43:39-62.

Brandes J, Wetter C, Bagnall RH, Larson RH. Size and shape of the particles of Potato virus S, Potato virus M, and Carnation latent virus. Phytopathology (1959) 49:4.

Buck KW. Comparison of the replication of positive-stranded RNA viruses of plants and animals. Adv Virus Res (1996) 47:159-251.

Bustamante PI, Hull R. Plant virus gene expression strategies. Electron. J. Biotech. (1998) 1:17.

Diez J, Marcos JF, Pallas V. Carmovirus isolation and RNA extraction. Methods Mol Biol (1998) 81:211-217.

Ding SW, Voinnet O. Antiviral immunity directed by small RNAs. Cell (2007) 130:413-426.

Dodds JA, Lee SQ, Tiffany M. Cross protection between strains of cucumber mosaic virus: effect of host and type of inoculum on accumulation of virions and double-stranded RNA of the challenge strain. Virology (1985) 144:301-309.

Drugeon G, et al. The strategies of plant virus gene expression: models of economy. Rev. Plant Sci. (1999) 148:11.

Eggen R, Verver J, Wellink J, Pleij K, van Kammen A, Goldbach R. Analysis of sequences involved in cowpea mosaic virus RNA replication using site-specific mutants. Virology (1989) 173:456-464.

Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution (1985) 39:18.

Firth AE, Brierley I. Non-canonical translation in RNA viruses. J Gen Virol (2012) 93:1385-1409.

Flick R, Neumann G, Hoffmann E, Neumeier E, Hobom G. Promoter elements in the influenza vRNA terminal structure. RNA (1996) 2:1046-1057.

Foster GD, Meehan BM, Mills PR. Nucleotide sequence of the 7K gene of carnation latent virus. Plant Mol Biol (1990) 15:937-939.

Gal-On A, Shiboleth YM. Cross-protection. In: Natural Resistance Mechanisms of Plant to Viruses (2006).

Guilford PJ, Beck DL, Forster RL. Influence of the poly(A) tail and putative polyadenylation signal on the infectivity of white clover mosaic potexvirus. Virology (1991) 182:61-67.

Haseloff J, Goelet P, Zimmern D, Ahlquist P, Dasgupta R, Kaesberg P. Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. Proceedings of the National Academy of Sciences (1984) 81:4358-4362.

Hollings M, Stone OM. Chenopodium quinoa Willd. as an Indicator Plant for Carnation latent virus. Plant Pathology (1965) 14:66-68.

Hwang YT, McCartney AW, Gidda SK, Mullen RT. Localization of the Carnation Italian ringspot virus replication protein p36 to the mitochondrial outer membrane is mediated by an internal targeting signal and the TOM complex. BMC Cell Biol (2008) 9:54.

Inoue T, Moore P, Tsai B. How viruses and toxins disassemble to enter host cells. Annu Rev Microbiol (2011) 65:287-305.

Kalantidis K, Schumacher HT, Alexiadis T, Helm JM. RNA silencing movement in plants. Biol Cell (2008) 100:13-26.

Kassanis B. A Virus Latent in Carnation and Potato Plants. Nature (1954) 173:1097-1098.

Kassanis B. Some Properties of Four Viruses Isolated from Carnation Plants. Annals of Applied Biology (1955) 43:103-113.

Kassanis B, White RF. Inhibition of acquired resistance to tobacco mosaic virus by actinomycin D. J Gen Virol (1974) 25:323-324.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses (2012) San Diego: Elsevier Academic Press.

Komatsu K, et al. Infection of capilloviruses requires subgenomic RNAs whose transcription is controlled by promoter-like sequences conserved among flexiviruses. Virus Res (2012) 167:8-15.

Lacomme C, Hrubikova K, Hein I. Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. Plant J (2003) 34:543-553.

Larkin MA, et al. Clustal W and Clustal X version 2.0. Bioinformatics (2007) 23:2947-2948.

Li YY, et al. Discovery and Characterization of a Novel Carlavirus Infecting Potatoes in China. PLoS One (2013) 8:e69255.

Mackenzie J. Wrapping Things up about Virus RNA Replication. Traffic (2005) 6:967-977.

Mandahar CL. Multiplication of RNA plant viruses. (2006) Dordrecht, The Netherlands: Springer.

Martelli GP, Adams MJ, Kreuze JF, Dolja VV. Family Flexiviridae: a case study in virion and genome plasticity. Annu Rev Phytopathol (2007) 45:73-100.

Meehan BM, Mills PR. Nucleotide sequence of the 3'-terminal region of carnation latent virus. Intervirology (1991) 32:262-267.

Mundry KW, Watkins PA, Ashfield T, Plaskitt KA, Eisele-Walter S, Wilson TM. Complete uncoating of the 5' leader sequence of tobacco mosaic virus RNA occurs rapidly and is required to initiate cotranslational virus disassembly in vitro. J Gen Virol (1991) 72 (Pt 4):769-777.

Netherton C, Moffat K, Brooks E, Wileman T. A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. Adv Virus Res (2007) 70:101-182.

Pallas V. Una breve historia del fenómeno del silenciamiento del RNA. Contribuciones de la Virología a su descubrimiento. Virología (2013) 16:7.

Pallas V, Mas P, Sanchez-Navarro JA. Detection of plant RNA viruses by nonisotopic dot-blot hybridization. Methods Mol Biol (1998) 81:461-468.

Palukaitis P, Zaitlin M. A model to explain 'cross-protection' phenomenon shown by plant viruses and viroids. In: Plant-Microbe Interactions (1984) New York: Macmillan.

Pham KTK, Verhoeven JTJ, de Kock M. Partial nucleotide sequence of Carnation latent virus from Dianthus. Submitted to the EMBL/GenBank/DDBJ databases (2008).

Rott ME, Jelkmann W. Characterization and detection of several filamentous viruses of Cherry: Adaptation of an alternative cloning method (DOP-PCR), and modification of an RNA extraction protocol. European Journal of Plant Pathology (2001) 107:9.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol (1987) 4:406-425.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. 2nd ed. (1989) NY: Cold Spring Harbor Laboratory Press.

Sanchez-Navarro JA, Canizares MC, Cano EA, Pallas V. Simultaneous detection of five carnation viruses by non-isotopic molecular hybridization. J Virol Methods (1999) 82:167-175.

Sanchez-Navarrro JA, Cano E, Pallas V. Non-radioactive molecular hybridization detection of carnation mottle virus in infected carnations and its comparison to serological and biological techniques. Plant Pathology (1996) 45:7.

Shaw JG. Tobacco mosaic virus and the study of early events in virus infections. Philos Trans R Soc Lond B Biol Sci (1999) 354:603-611.

Sherwood JL. Demonstration of the Specific Involvement of Coat Protein in Tobacco Mosaic Virus (TMV) Cross Protection using a TMV Coat Protein Mutant. Journal of Phytopathology (1987) 118:358-362.

Sriskanda VS, Pruss G, Ge X, Vance VB. An eight-nucleotide sequence in the potato virus X 3' untranslated region is required for both host protein binding and viral multiplication. J Virol (1996) 70:5266-5271.

Surana P, Satchidanandam V, Nair DT. RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanistically important pre-initiation state. Nucleic Acids Res (2014) 42:2758-2773.

Sztuba-Solinska J, Stollar V, Bujarski JJ. Subgenomic messenger RNAs: mastering regulation of (+)-strand RNA virus life cycle. Virology (2011) 412:245-255.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol (2011) 28:2731-2739.

Thomas CL, Jones L, Baulcombe DC, Maule AJ. Size constraints for targeting post-transcriptional gene silencing and for RNA-directed methylation in Nicotiana benthamiana using a potato virus X vector. Plant J (2001) 25:417-425.

Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods (2007) 3:12.

Waterhouse PM, Fusaro AF. Plant science. Viruses face a double defense by plant small RNAs. Science (2006) 313:54-55.

Yagi M, et al. Sequence Analysis of the Genome of Carnation (Dianthus caryophyllus L.). DNA Research (2014) 21:231-241.

Zhou ZS, Dell'Orco M, Saldarelli P, Turturo C, Minafra A, Martelli GP. Identification of an RNA-silencing suppressor in the genome of Grapevine virus A. Journal of General Virology (2006) 87:2387-2395.

Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. J Theor Biol (1965) 8:357-366.

ANEXO I: GENOMA COMPLETO DEL *VIRUS LATENTE DEL CLAVEL*

```
   1 CGATAAACAA AACAAAACAA CGCAAACACA TCTGAAACAC TTTAAGCAGA
  51 GCTAAACTAT TTCAGATCTG ATGGCGTTAA CTTACAGAAG TCCGATCGAA
 101 GGTGCACTAG ACTCATTCAC AACCTCAGAA CAGAGCTTAA TCGCTTCAAC
 151 AGCTGTAGGT GGCTACAAGC GGCTTGAGGA GAATAACTTC CAGTTCTTCA
 201 ATTATGCTCT CTCTCCGCAT CAAAAGCAAA AGCTTATTGA AGCAGGTTTG
 251 TATTTGAGTC CCTTTTCAGG AGTACCTCAT TCTCACCCAT GTTGTAAAAC
 301 TTTAGAAAAC CATATATTGT ACACTGTATT GCCTTCATAT ATTGATAGCA
 351 CTTTTTATTT TGTAGGTATT AAAGAACGTA AATTACTTTT TCTTAAAGAT
 401 AGGAATAGTG GATTGTCTGC ATGCGGCCTA ATAAATAGGT ACGTTACAAG
 451 TAAAGATTAC TCTAGGTACC AAAGCACATT TCATGCAAGC GCTAGCGAGG
 501 AATTGAGTGG GAATTTAAGA GCGCTCTCAG AGAGCTCGGA AACATTAAGA
 551 AGCTTGATCC CACCTTGTAT ATCGAAAAAT GCACGGAACC TCTTCCTTCA
 601 TGATGAAGTT CATTACTGGT CCACGAAGGA CTTGAAGAGG TTTTTGGGCG
 651 CAGTCAGGCC CTCACGTTTA TTGTGCACTG TGGTGGTTCC TCCAGAGATA
 701 TTGGGGGGCT GCCGACAGAG TTTAAATCCT TGGTGTTATG AGTTCAAGAT
 751 TCAGGGAAAG GATTTACACT TCTACCCAGA TGGGGTTCGC TCTGAGGGTT
 801 ATGTGCAACC CATGACGTGT TGCGAGCTTT TAGGTTCCAA TAGAGTGGTG
 851 CTTGAGGACG GTACGGTTTA TGCAATAGAT ATGCTGCATA GTGTGTTTAG
 901 TCATCACCTC ATATCGGTGA CTAGAACAAA TGCCTTGTGC GAGAAGGAGA
 951 GGAGTTTCTC TTGCTTTGAG GCTACGACGA GCACTGGGCT CGTTAACTTT
1001 GGGAAGAAGA AGGTTGGGCC TTGCTTAATG ATTCCTTTTC CAGTGGTCAG
1051 CAGGATTTAC CGATATTTAA GGACTCTAAT GAAGCCAGAC ATGGAGTCTG
1101 CAATGGCGAA GTTGAGTCAA ATTTTGCCTG AGCCCACGGC TTACCAAATT
1151 AAATTTGTGC AGGAATTCAG TGAACTAGTG ATAAAGACAC CCGTGCACAG
1201 GTCAATGTTG AGCACTAGTT TCATGGTTGA ACTTGGTGCA TTTGTGACCT
1251 CTGTCTTTCC TAGCTGCTTT GCACACAAAT TTCAAAGCTT GAAAATCAAG
1301 AGTTTGGATG CATTCGTCCG AACTATGAAG CCGTACTGTG TCGACGTGAA
1351 GCTGAGTGAT TATGGGAGCT TTATTGAGAG AATTGTTTGG TTGGATAGTA
1401 CATTAATCGG GATTGAGCCA GAGTTGGACG TTGTCAGCAA AATGGAGCGG
1451 TTGTACACTG GGTTCGACGA AGGTGAGCAG ATTGTTAGGC AAAGTCGCTA
1501 TACTCTCGAA CCTTATGCTG ATGAGGTAGA GTTTAACTTT AGGAGGTTGC
1551 ATAACATTTG CTTGAGGTTC TTGAGGGGGA TGATTTATCA AGGCATGGAG
1601 TTGCCAAGTT ATTCAGCTTG TATTGAGGTT CTTAAAACAG CGGTCCGGAG
1651 GAGTAGGTTG CTGGAACTAT GTCTTGAAAG GTTCCTACAG GAGCCAAGAG
1701 AATTGATTAA GGTGTACCGT GATTTCTGCA GGGTCGCTCG TCGTTATGTA
1751 GGTCCTTGCC CCAGCAGCTT CGCCTTCTCT CCTATGTTGT GGTTTAGCTT
1801 GCGCAGCAGC CGCAGAGTTA ACATTAATTA TCTAGATAGT GTTGCTCACG
1851 GGCATTCAAC CGTGTACTCC CTTAAGGCCC CATGGGCCGA CGTTGTGCGA
1901 AGTCTGGGCG CGTGTGGTGC AAATTTTATC GAATTGGAAG AATGCAAGGA
1951 TCTTCAGGGT GAAGTCACCT ATAGGGGCGA GGTAGACGAA GGGCTTAATC
2001 AGTGGGAGTT AAGTAGACGA GAGCTGGATC ACCTGCATGT TGGGGCTTCG
2051 GTTAAGTTGA CCTTTTGGAG AGCTTTTGGC ACGGAATACC TCAACCTTGA
2101 GGCCTGTGAC CAGATGAAAG GTAGGGAGGT CAGCTGGTTT ACCAGACGTG
2151 GTAATAAGGA CTACTCTTAC AATGGAGGCA ATCATGTAGA TCGCGGGTGG
2201 CCCGAGGTTT TGGACCACTT CATGGAACTG AATGAGATAC CCCTCGATTA
2251 TTACGACTGC TGCCTTTTCC AGGAATACAA GCATGGAGGC GCTATAGGTT
2301 TCCACAAGGA TGATGAGGAT ATTTTTCTTA ATGGTGGTTC AATTTACACG
```

```
2351 GTCTCTTATG ATGGGGATTG CGTTTTCGGG TTCAAGCACG AGGCGCTTGT
2401 TTGTGAAAAG GATTTGAGTG GAGCTGTAAA ATTTGAGATG CCTGAGGGGT
2451 TCCAAAAAAC TCATTACCAT CGGGTGCGCA ATTGTTCAGA GGGCAGGAAG
2501 AGCCTAACTT TCCGCAGACT CAAACGAAGT ATGGGCCCTG ACGTTGCGCC
2551 TTCTGTGGTT CAGCAAAAGG AGGACGCTAG TGTTGGAACA CAGACCGTTT
2601 CAGTTGTCGA AGAGCTGTAT GGTTCAAGCT TGAGATACAC GCCCTGGGAT
2651 GGAAAGGGTG TTGTGTCTAT TCATGACGTG CCTGGGGATG GAAACTGCTT
2701 CTTCCACTGC TTAGCTCTGT TCTTGCGTGG GACGCATAAA GAGCTGCGTG
2751 AGGCAGTGCG TCGTGAAGTG GCCCGCCACA AGCTGGATGA TCAGGCCTTG
2801 CACAAGCAAT TGGAGAATGG AGTCTACTCT GAGACTCAGA TTATCGCTTT
2851 TGCCTGCAAG GTTTTTAAAC TGGAGGTTAA AATTTTTGAT TTTAACGCAA
2901 AGGCTGTCGT CGATGTGAGC GGAAGTGATG CTGAGTATTT CATCGAACTA
2951 CTACTGCTTG ACGAGCACTT CACTTTGTGC TGCCTTGAGG AGAATTGCAT
3001 TTTCAAATCT ATTGCAGACT GCTTGGGCCG CAGTATCCAG GATGTAATTA
3051 TCGCAATGGG CAAGGGTGCG AACCAGGAGC TGTACAAGAG TGTCGTTTGT
3101 AGATCTGGTA GATCTGGGGT GGATGTGACA ATGCTCGAGG CGATCTTGAC
3151 TATTTTTGGA ATTGAAGGGC ATATGGCAGT GGGTGGTGAT TTAGTCGTTT
3201 TGAACGAGGG GGGCCGAATA CCGGCTAGTT TCGTCCTCAG TGAGAAGCAT
3251 GTAGCCTTCA AGAGCCTCGG TGGATCTGAT AAACCACTTA AGGTGCACCG
3301 ATCAGGTCTG AGCGGGGAGC TGATCGATGT GAACGTTATG CCTGTGCGAA
3351 GAAAAAGGAG CAACGAGCTC GTTTTGAAGG TTGCTGGGTC AAGCGTGGAC
3401 TTCATAGCCT CAAAAGGGAA AGCTTCACTA CTGAGCAGGA GTCTGCATCA
3451 AGGTTCAACG GGTTTGATTT GCTCAGAGCT CTTCTCCAAC CACAAAGAGA
3501 TAGATGTTGA GAGTCTGGGG AGAAACGATG GGTTCAACCT TAAGGTTAAC
3551 CTTATCCTTG GTACTTTTGG GAGTGGAAAA AGCACGGTTT ACAAGAAGTT
3601 CCTGTGTAAA TCGATAGGCA AGCACGTAGT TTTCGTCTCA CCGCGTAGAG
3651 TCCTAGCTGA CATGTTCAGG GAAATTGTCA ATGATGTGTG TTGTGACTCG
3701 AAAGAACGCG AGAAGTGGAA AATTATGACT TTTGAGGTCT TCGCTCTCAA
3751 AATGAGTTCA ATCACGGGCG CGACAGTTAT TATTGATGAA ATTCAGCTTT
3801 TTCCGCCAGG TTATCTAGAT CTTATCTGCT TCTGCTTACG CGATTGCAAC
3851 TCGCTTTACC TCGCTGGGGA TCCTTGCCAG AGTGATTACG ACAATGAGCG
3901 GGATAGGGGT ATTTTTGGGG ATGCCCCAAG TGACATTGAG CGGGTTTTGA
3951 CAATGCAGTG CTATCGTTTT GCCACGCTGA GCTATAGGTT CTTGAACGCT
4001 GACTTTTTGG GTAGATTACC TTGTGACATG CTTGAAAGCG GTTTCACTAT
4051 CAATGAGCCC TGCTTTGACT TTGCAGAATC CTATGAAGAC CAAGCAATTT
4101 ACGAGGGTTA TGGTGCTGTC TTAGTGTCAA GCTTTAATGA AAAAACTGCG
4151 TTCAAGTGCT TGGTCCAGCC TGCAGTAAAA GTACTTACTT TTGGGGAAAG
4201 CACTGGCTTA ACCTATGATA AGGTTGTTAT CGTAATATCT GAACCATCCA
4251 TGGCCGTGTC TGAGAATCGT TGGGTCACCG CTCTAAGTCG CGCTAGGAAG
4301 TTTGTCAGAT TCGTGTTGAC GATAGATTGC CATGCGAGCA GGCTGGCTGC
4351 AATTTTTGAA GGGAGGCACC TGGGGCGGTT TTTAACTAGA GAGGCTAGCG
4401 CCCGCAATCT AAGTGCAATG CTACCCGGGA ACGCCAGGTT TGTCAAGGAG
4451 CTAATGGGTA ATGTGGGGGC CTCCCAGGGT GTCTTTGAAG AAAAACTGCG
4501 TGGGGATCCC TGGTTGAAGT CGCAACTCTT TCTGGGGCAA GATGAGGATG
4551 TGCAGGAAGC TGACGTGCAG GAGTATGAGG GAGTGCTACA GATTTTTAAG
4601 ACCCACATTC CACGTTGTGA GATGGAAGGA GTGCGCGCTG AGTGGGTGCA
4651 CAAATTCATG TCTAAGGATG TGAGAGAGAA AAGGATTGGT TACCTCGTCT
4701 CTGAACAGTT TGCAGACCAG CACAGCAGGA ATCGCGGTAA GCACCTGACC
4751 AACGCAGCTG AGCGGTTTGA GGCGATTTAC CCAAAACATA AGAACTCGGA
4801 TACAGTGACC TTCATTATGG GCGCGCGAAA ACGCTTAAGA TTTTCCAAAC
4851 CTTCAGTTGA AATGAGGAAG TACGTCGAGG CCAGAGTTTA TGGCGAAATG
```

4901 CTGTTGAAAG AGTTCTTAAA ACGAGTTCCC CTGAAGAGGA AGCACGATCC
4951 TAACATGATG GCAGAGGCTC TACGGGATTT CGAATTGAAA AAAGTGAGCA
5001 AGAGTGCAGC CGTGATTGAG AATCATGCAG GGCGCTCTTG CAGGGACTGG
5051 CTCATCGACA CAGGTCTGGT TTTCATGAAG TCCCAGCACT GCACGAAGTA
5101 CGAGAAGAGG TTCACAGAAG CGAAAGCTGC GCAGAGCATT GTGTGCTTTC
5151 AACATTCGGT TTTGATCAGA TTTGCACCAT ACATGCGCTA TATTGAGAAG
5201 AAGCTCAAGG AAGCCATACC AGGTCACTAC TACATACACA GCGGCAAGGG
5251 GCTGGATGAG TTAAACAAGT GGGTTAAGGA GTATGGCTTT AATGGTGTGT
5301 GTACTGAATC AGATTATGAA GCATTTGATG CAAGTCAAGA TGAGTATATC
5351 CTAGCTTTTG AGGTGAAGAC CATGGAATAC TTGGGTATAC CGAAAGATTT
5401 GATCGCTGAT TATATCTTTA TCAAGACACA TCTCGGGTCA AAATTGGGTA
5451 ACTTTGCTAT CATGCGGTTT TCAGGAGAGG CAAGCACCTT CCTATTTAAT
5501 ACGATGGCTA ACATGTTATT CACCTTTATT AGGTACGACT TAAACGGTAA
5551 GGAAGCCATA TGCTTTGCTG GTGATGACAT GTGCGCTTCT AAACGCCTGA
5601 AGAAGAGTGA GAGGATGAGT TTTTTCCTCG ATAAGCTAAA ACTGAAAGCC
5651 AAGGTGTGCT TTACTGAAAA ACCGACCTTC TGCGGGTGGA ATCTCTGTAG
5701 TGATGGCATA TACAAGAAGC CGCAACTGGT TTATGAGAGG TTATGCATTG
5751 CTAAGGAAAC AAATAATCTG CCAGCTTGCA TTGATAATTA TGCTATCGAG
5801 GTTTCTTATG CTTACAAATT AGGGGAATTG GCCATCAACA GGATGGATGA
5851 CGAGGAGATC AAGAGCTACT ACAATTGTGT GAAAGTGATA ATTAAAAATA
5901 AGCACTTAAT GAAATCAAGT GTTGCAGATA TTTTCAAATC TGCTTAGGTT
5951 TACGCACTGT AATTGATTAG ATGGAAGTAG TCTGTAAAGT ATTAGCTGAT
6001 AGAGGTTTTC AACGTGTGCG TAGTAAGTTA AGTGTGCCAA TAATTGTGAA
6051 TTGCGTGCCT GGTGCGGGTA AGACGAGCGC AATCTTTGAT ATTTTAGCCA
6101 GTGATAGCCG GTTTAAAGCT GTCACTTTTG GGGCGAAGCA AGAACGGAAG
6151 ATTGAGGGCC TAACCATTGA CAGGGTCGTT AGCGGTTGTA GACCACCTGC
6201 AGAGTTCTTA ATTGTCGACG AGTACACTGA AGGAGATTAT GGGGAGCTCG
6251 ATCCTTTCGT GATTTTTGGT GACCCGAATC AGTCAAGAAC GGACTCCTGT
6301 TTGAGGCCCC ACTGGTTGTC ATTCAAGAAC TTCCGTTTCG GCAGATCCAC
6351 TTGCACTTAT TTAAACAAGT TGGGCTTTAA GATTGAGAGT GAAAAAGAGG
6401 ATTCCTTTAA GGTCTGTAAA GCCTCTGAGT GTGATATTGA GGGCCAGCTG
6451 ATTGTGTACG GCCAGCAGGC TTTAACCCTC GCTAGTTGGT ACGGTTTGGA
6501 GTTCCTGTGC CTCGAAGCCG CACGCGGTAG GACTTTTGAG GTTTGCACCG
6551 TCCTAACGGA TTACGAGAGT GTCACAGAAG AAATCAGGGG TGATCTTTTT
6601 GTGTTGCTCA CAAGACATAG GAGTAAGTGT GTCGTACTGA ACGGAGATGC
6651 CACTCTTGCC GCCACCTGAT CACACCAAAG CAGTTCTGGC GATCTCAATT
6701 GGTGTTGTGG TTGCAATTGT GATTTTCGCC CTACGTTCAA ATAATCTGCC
6751 TGCGGTTGGT GATAACGTGC ATAGCTTACC CCACGGGGGT TTCTACAAGG
6801 ACGGTACAAA AACTGTGTTC TACGGCTCTA AAGTAAACGC GGGGTCAGTC
6851 TCAGAACACA CAGCTCGAAT TCTATCAAAA CCGTATGCTG CCTTTGCTCT
6901 CATCTGTGGA GTACTTTTCC TGAGCTTTTT CAACAAGAGG AAGTGCAATT
6951 GTACCTGTGA CAGATGCAAG CGTCTGGCCT AATACTGGTG GCGCTCTTTA
7001 GTGCGGCAGT CTCTTATTTG GCACTTTTGC ATTTAAGTAG CAGTAGTAGT
7051 TCGTGTGTAG TTGTAGTGAC AGGCGAGTCC TTTAGAATAA GCGGTTGTGA
7101 TTTCACAGAG GAATTTATTG GATTTGCAAA AACCTTGAGG GTTGCGAATT
7151 CACAACCTTA GGTTCACGGT TTGCAGTTTG AAATAATATA CATACATACA
7201 TAATGCCTGA AAAGAGACCC GAGGAAAATG TGGGATCGCA AGGGCAAACC
7251 GGCCCCAACA ATCCACCACA ACAACATCAG GGGGAGCAAT CGGTTGCGTC
7301 AGGGATGGTT TCAGTCAATG AGGTTGATCT TAGGAATCTT CAAAACCGCG
7351 TGGAGGAGGC TGAGAAATTC ATGCAGCGTT TTCAAAAGTT AAAGGAATTC
7401 AACTCCCAAA ACCTCACAGC TGGCGAATTG AAAAATGGTG GTTTTGAAAG

```
7451 CGGTAGACCT CCAGCTAAAA TTTCTGAGCA TTTACGTGGC TCAACTGCGA
7501 ATGTCTTTAC AAGGCCTTCT CTAGATGCAC TACAGATGAT GGACTTCAAA
7551 CCGGAGTCAA ACATGATTGT AACAGCTGAA GAACTTGCAG CCATAACAGC
7601 CAAGTTGGAA GGTCTGGGTG TACCCACTGA GCGTAGTGCA CCAATCTGTT
7651 GGGCTGTCGC ACGTTACTGT GCAAACACAA GTACCTCACC CTACACCGAC
7701 CCAAAAGGGG TGTTTGAATT TCCTGGGGGT GCTATAACTA GAGACGCAGT
7751 CTTTGCTGTC ATAAGAGAAG TCACAACTTT ACGGGCCTTC TGCAGGGCTT
7801 TCGCGCCCAT CACTTGGAAC CAGATGCTGT TCGCCAAATC CCCACCTGAG
7851 AATTGGCAAG CAAAGGGATA CACCTATGAA ACAAGGTATG CCGCTTTTGA
7901 TGTGTTTGAC TTCGTGCAAA ACCCTGCGGC TATACAACCT CTGGAGGGTC
7951 TGCTGAGGAT TCCAACAGCC GAGGAAAAAA TAGCACACGC TACAAACAAA
8001 AGGCTGGCCC TTGATAGGAA CCGTAGAAAT GCAAGATTCT CTAGCACTGA
8051 TAGCCTCGTG ACAGGCGGGA TGTATGGTAA GGATATTAAG ACGAACTTTA
8101 ACGGTTCGAA CAATTCAGAC TAATGCGTGA AAGAAAGCTA AGAAAGACGC
8151 TTGAGGACCT GTTCAAGCGT TTCGCATCAG TGCAACATGG GCACTCTGAT
8201 TGTGTTAATA TTATAATAGC AAAAATAAAG AGTGGTCAAC CTGGGGAGAG
8251 TAAATATGCA CGGAGACGTA GAGCTAAGTC AATAGCCCGA TGTCCAAGGT
8301 GCGCACGCGT TTCTCCTGGG TTTTACTTCA CCACTCGTTG CGATGGTAAG
8351 ACATGCAGAC CTGGTTTATC AGCCCGACCA GATTTGTTAG AATTTATTGG
8401 GATTGACTTG TGTGTAAGAT CTAAGTAATA TATAAGCACT TAAGTATAAA
8451 AAGCAAATGC GTTTTAAATA TTTTTCCGTA TTTT
```

ANEXO II: SECUENCIAS DE LAS PROTEÍNAS EXPRESADAS POR LAS SEIS PAUTAS DE LECTURA ABIERTAS DEL *VIRUS LATENTE DEL CLAVEL*

1. RNA POLIMERASA RNA DEPENDIENTE

```
   1 MALTYRSPIE GALDSFTTSE QSLIASTAVG GYKRLEENNF QFFNYALSPH
  51 QKQKLIEAGL YLSPFSGVPH SHPCCKTLEN HILYTVLPSY IDSTFYFVGI
 101 KERKLLFLKD RNSGLSACGL INRYVTSKDY SRYQSTFHAS ASEELSGNLR
 151 ALSESSETLR SLIPPCISKN ARNLFLHDEV HYWSTKDLKR FLGAVRPSRL
 201 LCTVVVPPEI LGGCRQSLNP WCYEFKIQGK DLHFYPDGVR SEGYVQPMTC
 251 CELLGSNRVV LEDGTVYAID MLHSVFSHHL ISVTRTNALC EKERSFSCFE
 301 ATTSTGLVNF GKKKVGPCLM IPFPVVSRIY RYLRTLMKPD MESAMAKLSQ
 351 ILPEPTAYQI KFVQEFSELV IKTPVHRSML STSFMVELGA FVTSVFPSCF
 401 AHKFQSLKIK SLDAFVRTMK PYCVDVKLSD YGSFIERIVW LDSTLIGIEP
 451 ELDVVSKMER LYTGFDEGEQ IVRQSRYTLE PYADEVEFNF RRLHNICLRF
 501 LRGMIYQGME LPSYSACIEV LKTAVRRSRL LELCLERFLQ EPRELIKVYR
 551 DFCRVARRYV GPCPSSFAFS PMLWFSLRSS RRVNINYLDS VAHGHSTVYS
 601 LKAPWADVVR SLGACGANFI ELEECKDLQG EVTYRGEVDE GLNQWELSRR
 651 ELDHLHVGAS VKLTFWRAFG TEYLNLEACD QMKGREVSWF TRRGNKDYSY
 701 NGGNHVDRGW PEVLDHFMEL NEIPLDYYDC CLFQEYKHGG AIGFHKDDED
 751 IFLNGGSIYT VSYDGDCVFG FKHEALVCEK DLSGAVKFEM PEGFQKTHYH
 801 RVRNCSEGRK SLTFRRLKRS MGPDVAPSVV QQKEDASVGT QTVSVVEELY
 851 GSSLRYTPWD GKGVVSIHDV PGDGNCFFHC LALFLRGTHK ELREAVRREV
 901 ARHKLDDQAL HKQLENGVYS ETQIIAFACK VFKLEVKIFD FNAKAVVDVS
 951 GSDAEYFIEL LLLDEHFTLC CLEENCIFKS IADCLGRSIQ DVIIAMGKGA
1001 NQELYKSVVC RSGRSGVDVT MLEAILTIFG IEGHMAVGGD LVVLNEGGRI
1051 PASFVLSEKH VAFKSLGGSD KPLKVHRSGL SGELIDVNVM PVRRKRSNEL
1101 VLKVAGSSVD FIASKGKASL LSRSLHQGST GLICSELFSN HKEIDVESLG
1151 RNDGFNLKVN LILGTFGSGK STVYKKFLCK SIGKHVVFVS PRRVLADMFR
1201 EIVNDVCCDS KEREKWKIMT FEVFALKMSS ITGATVIIDE IQLFPPGYLD
1251 LICFCLRDCN SLYLAGDPCQ SDYDNERDRG IFGDAPSDIE RVLTMQCYRF
1301 ATLSYRFLNA DFLGRLPCDM LESGFTINEP CFDFAESYED QAIYEGYGAV
1351 LVSSFNEKTA FKCLVQPAVK VLTFGESTGL TYDKVVIVIS EPSMAVSENR
1401 WVTALSRARK FVRFVLTIDC HASRLAAIFE GRHLGRFLTR EASARNLSAM
1451 LPGNARFVKE LMGNVGASQG VFEEKLRGDP WLKSQLFLGQ DEDVQEADVQ
1501 EYEGVLQIFK THIPRCEMEG VRAEWVHKFM SKDVREKRIG YLVSEQFADQ
1551 HSRNRGKHLT NAAERFEAIY PKHKNSDTVT FIMGARKRLR FSKPSVEMRK
1601 YVEARVYGEM LLKEFLKRVP LKRKHDPNMM AEALRDFELK KVSKSAAVIE
1651 NHAGRSCRDW LIDTGLVFMK SQHCTKYEKR FTEAKAAQSI VCFQHSVLIR
1701 FAPYMRYIEK KLKEAIPGHY YIHSGKGLDE LNKWVKEYGF NGVCTESDYE
1751 AFDASQDEYI LAFEVKTMEY LGIPKDLIAD YIFIKTHLGS KLGNFAIMRF
1801 SGEASTFLFN TMANMLFTFI RYDLNGKEAI CFAGDDMCAS KRLKKSERMS
1851 FFLDKLKLKA KVCFTEKPTF CGWNLCSDGI YKKPQLVYER LCIAKETNNL
1901 PACIDNYAIE VSYAYKLGEL AINRMDDEEI KSYYNCVKVI IKNKHLMKSS
1951 VADIFKSA
```

2. TRIPLE GENE BLOCK 1

```
  1 MEVVCKVLAD RGFQRVRSKL SVPIIVNCVP GAGKTSAIFD ILASDSRFKA
 51 VTFGAKQERK IEGLTIDRVV SGCRPPAEFL IVDEYTEGDY GELDPFVIFG
101 DPNQSRTDSC LRPHWLSFKN FRFGRSTCTY LNKLGFKIES EKEDSFKVCK
151 ASECDIEGQL IVYGQQALTL ASWYGLEFLC LEAARGRTFE VCTVLTDYES
201 VTEEIRGDLF VLLTRHRSKC VVLNGDATLA AT
```

3. TRIPLE GENE BLOCK 2

```
  1 MPLLPPPDHT KAVLAISIGV VVAIVIFALR SNNLPAVGDN VHSLPHGGFY
 51 KDGTKTVFYG SKVNAGSVSE HTARILSKPY AAFALICGVL FLSFFNKRKC
101 NCTCDRCKRL A
```

4. TRIPLE GENE BLOCK 3

```
  1 MQASGLILVA LFSAAVSYLA LLHLSSSSSS CVVVVTGESF RISGCDFTEE
 51 FIGFAKTLRV ANSQP
```

5. PROTEÍNA DE CUBIERTA

```
  1 MPEKRPEENV GSQGQTGPNN PPQQHQGEQS VASGMVSVNE VDLRNLQNRV
 51 EEAEKFMQRF QKLKEFNSQN LTAGELKNGG FESGRPPAKI SEHLRGSTAN
101 VFTRPSLDAL QMMDFKPESN MIVTAEELAA ITAKLEGLGV PTERSAPICW
151 AVARYCANTS TSPYTDPKGV FEFPGGAITR DAVFAVIREV TTLRAFCRAF
201 APITWNQMLF AKSPPENWQA KGYTYETRYA AFDVFDFVQN PAAIQPLEGL
251 LRIPTAEEKI AHATNKRLAL DRNRRNARFS STDSLVTGGM YGKDIKTNFN
301 GSNNSD
```

6. ORF6

```
  1 MRERKLRKTL EDLFKRFASV QHGHSDCVNI IIAKIKSGQP GESKYARRRR
 51 AKSIARCPRC ARVSPGFYFT TRCDGKTCRP GLSARPDLLE FIGIDLCVRS
101 K
```