



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

**Epistasia multidimensional, rugosidad de
los paisajes adaptativos y factores del
huésped que determinan la eficacia viral**

Tesis presentada por

Héctor Cervera Benet

Programa de Doctorado en Biotecnología

Tesis dirigida por

Prof. Santiago F. Elena Fito





Santiago F. Elena Fito, Doctor en Ciencias Biológicas y Profesor de Investigación del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Biología Molecular y Celular de Plantas (IBMCP), centro mixto del CSIC y de la Universitat Politècnica de València,

CERTIFICA:

Que Don Héctor Cervera Benet, Licenciado en Biología por la Universitat de València, ha realizado bajo mi supervisión la Tesis Doctoral titulada “*Epistasia multidimensional, rugosidad de los paisajes adaptativos y factores del huésped que determinan la eficacia viral*”.

Y para que así conste, en cumplimiento de la legislación vigente, firma el presente certificado en Valencia a 13 de Septiembre de 2017.

Fdo. Santiago F. Elena Fito

“A esta conservación de las variaciones y diferencias individualmente favorables y la destrucción de las que son perjudiciales, la he llamado selección natural o supervivencia de los más aptos”.

- **Charles Darwin**

“La biología evolutiva versa sobre los cambios que importan, los cambios que contribuyen al ajuste extraordinario de los organismos a su ambiente. El objetivo último del biólogo evolutivo es explicar la diversidad inmensa y fantástica que vemos en la naturaleza. Lo que desvelaba a Darwin por las noches era el problema de cómo montar un ojo por selección natural.”

- **Andrew Berry**

“Nada tiene sentido en biología si no es a la luz de la evolución.”

- **Theodosius Dobzhansky**

“La vida se abre camino.”

- **Jeff Goldblum interpretando al Dr. Ian Malcolm**

“Diseño debe ser simple. Elegante. Implementación, no tanto. Secuencias deben cambiar. Intención estática, producto fluido. Siempre posible mejorar. Factores estresantes revelan fallos. Fallos revelan potencial. Siempre mejorando. Bueno.”

- **Abathur, maestro de evolución**

Agradecimientos

Durante estos cuatro años he aprendido que la realización de una tesis doctoral no es atribuible a una sola persona, ya que todos aquellos que la rodean allanan el camino que conduce al acto de defensa de la Tesis Doctoral. En ese sentido, cabe agradecer el haber llegado hasta aquí a compañeros de trabajo, amigos y demás personas.

En primer lugar, dar las gracias al equipo que me ha acompañado durante estos años. A Julia y Anouk por todas las veces que me socorrieron durante mis primeros pasos en la biología evolutiva de sistemas y, muy especialmente, por su amistad. A Jose, con quien he compartido estos años de principio a fin y con quien he explorado el mapamundi del comedor al completo mientras comíamos. A Susana, por su inestimable ayuda y conocimiento del procedimiento de cuantificación. A Paqui y Paula, que con mucha dedicación y esfuerzo han tenido lista cualquier cosa que pudiera necesitar en el laboratorio y, muy especialmente, se han encargado de sembrar, trasplantar y cuidar mis plantitas. También a Silvia, Fernando, Guillermo B. y Guillermo R. por resolver mis dudas y los conocimientos que me han brindado.

En segundo lugar, dar las gracias a Santiago, a quien más he de agradecer. Primero, por aceptarme en su laboratorio. Segundo, por contribuir de forma positiva y constructiva a mi formación. Y tercero, por su infinita paciencia, pues reconozco ser una persona “especial”.

Por supuesto, también debo agradecerle a Carmelo, mi tutor de la UPV, su enorme implicación en cuanto a la gestión de la Tesis Doctoral.

Continuando con los agradecimientos, no puedo olvidarme de mis amigos, especialmente Víctor y Hugo, ni de los monitores del gimnasio ni, de nuevo, de mis compañeros de laboratorio, que además de acompañarme en esta aventura han estado conmigo y me han ayudado a levantarme en el momento en que más lo necesitaba.

No quiero terminar sin agradecer el esfuerzo de Claudia, persona que me animó a perseguir mis sueños y con la cual recorrí todos los pueblos de su Asturias natal, un sábado por la mañana, en busca de un ayuntamiento cuyas puertas estuvieran abiertas para poder entregar la solicitud que me ha llevado a defender hoy la Tesis Doctoral. A ella, que me empujó a empezar. A ella, para quien las expresiones matemáticas e^∞ y $\infty \times \infty$ tenían significado. A ella, que aunque ya no esté a mi lado, fue la mujer más importante de mi vida.

Índice

| | |
|---|-----------|
| Agradecimientos | 8 |
| Prólogo | 24 |
| Resumen / Resum / Abstract | 28 |
| 1. Introducción | 46 |
| 1. Introducción al mundo de los virus. | 48 |
| 2. Clasificación de los virus. | 49 |
| 3. Virus de ARN. | 51 |
| 3.1 Arquitectura del genoma. | 51 |
| 3.2 Ciclo de vida. | 54 |
| 3.3 La familia <i>Potyviridae</i> . | 58 |
| 4. Mutaciones como fuente de adaptación y evolución. | 64 |
| 4.1 Estimaciones de la tasa de mutación de los virus. | 65 |
| 4.2 Interacciones entre mutaciones. | 67 |
| 5. Los paisajes adaptativos de eficacia. | 69 |

| | |
|---|-----------|
| 2. Objetivos | 82 |
| 3. Material y métodos | 88 |
| 1. Generación de clones infecciosos de TEV. | 90 |
| 1.1 Origen. | 90 |
| 1.2 Vector plasmídico. | 91 |
| 1.3 Generación de los clones por mutagénesis dirigida. | 92 |
| 1.3.1 Mutagénesis dirigida. | 93 |
| 1.3.2 Genotipos de virus. | 95 |
| 1.4 Generación de los clones por evolución mediante pases seriados en <i>Arabidopsis thaliana</i> . | 97 |
| 1.4.1 Evolución mediante pases seriados. | 97 |
| 1.5 Generación de los clones por preparación de una población clonal inicial de TEV. | 99 |
| 2. Hospedadores. Condiciones de crecimiento e inoculación. | 99 |
| 2.1 <i>Nicotiana tabacum</i> var. Xanthi. | 100 |
| 2.2 <i>Arabidopsis thaliana</i> ecotipo Ler-0. | 103 |
| 2.3 <i>Chenopodium quinoa</i> . | 104 |

| | |
|--|------------|
| 3. Purificación, cuantificación de la carga viral y secuenciación del ARN genómico del virus. | 107 |
| 4. Evaluación de la eficacia y análisis estadísticos. | 110 |
| 4.1 Eficacia relativa. | 111 |
| 4.2 Análisis estadísticos. | 112 |
| 5. Paisajes de eficacia. | 115 |
| 5.1 Representación de los paisajes de eficacia. | 115 |
| 5.2 Evaluación de la robustez del paisaje de eficacia. | 116 |
| 5.3 Cálculo de la epistasia. | 118 |
| 6. Análisis transcriptómico. | 120 |
| 6.1 Hibridación de micromatrices. | 120 |
| 6.2 Análisis de genes expresados diferencialmente. | 121 |
| 6.3 Análisis funcional a partir de los listados de genes. | 122 |
| 6.4 Validación de la expresión génica mediante RT-qPCR. | 124 |

4. Resultados y discusión _____ 128

Capítulo 1. Las poblaciones de virus de ARN pueden escapar eficientemente de un óptimo local de eficacia. _____ 130

- 1. Paralelismo y divergencia en la eficacia. _____ 134
- 2. Contribuciones de la selección, la casualidad y la historia al resultado de la evolución. _____ 135
- 3. Diversificación genotípica. _____ 139
- 4. Discusión. _____ 142

Capítulo 2. Efecto de la especie huésped en la topografía del paisaje de eficacia para un virus vegetal de ARN. _____ 146

- 1. Correlaciones físicas y pleiotropía entre los huéspedes. _____ 151
- 2. Diferencias en la epistasia del paisaje entre los huéspedes. _____ 154
- 3. Relación entre la pleiotropía antagonista y las interacciones epistáticas. _____ 160
- 4. Discusión. _____ 161

Capítulo 3. Cambios en la eficacia debidos a la variación genética dentro de una población clonal de un virus de ARN. _____ 168

1. La eficacia promedio de los subclones no predice la eficacia de toda la población clonal de forma temprana. _____ 172
2. Las diferencias en la eficacia entre los subclones y la población clonal desaparecen después de largos períodos de replicación. _____ 174
3. Discusión. _____ 175

Capítulo 4. Factores del huésped *N. tabacum* que determinan la eficacia viral de TEV. _____ 182

1. Las diferencias en la eficacia están asociadas a diferencias en los perfiles transcriptómicos del huésped. _____ 187
2. El número y categoría funcional de los genes dependen de la eficacia viral. _____ 194
3. Correlación entre la expresión génica y la eficacia viral. _____ 203
4. Validación de los valores de expresión normalizados para ocho genes mediante RT-qPCR. _____ 207
5. Discusión. _____ 216

5. Conclusiones _____ 230

6. Publicaciones _____ 236

7. Referencias _____ 240

Prólogo

Los virus son parásitos intracelulares obligados cuyo pequeño genoma codifica la información necesaria para poder replicarse, pero no pueden hacerlo solos, sino que deben aprovecharse de la maquinaria metabólica de su huésped para tal fin (Koonin, 2012). Constituyen el grupo de organismos más abundante y extendido del planeta (Edwards y Rohwer, 2005). También hay que destacar que su diversidad biológica es más grande que la de cualquier otro tipo de patógeno conocido. De ahí que se puedan encontrar infectando todas las formas de vida, ya sean bacterias, plantas o animales.

Desde una perspectiva evolutiva, una de las principales características de cualquier virus es su capacidad para mutar (Sanjuán et al., 2010). En particular, los virus de ARN son los que presentan una mayor diversidad genética a nivel poblacional. Una consecuencia directa de su enorme capacidad evolutiva es su habilidad para escapar del sistema inmune y evadir estrategias antivirales (Elena y Sanjuán, 2008). Incluso, pueden enfrentarse al desafío ambiental que supone la adaptación a un nuevo huésped (Holmes, 2009).

Debido a su ciclo de vida, los virus han desarrollado numerosas formas de interactuar con sus huéspedes. En ocasiones, se produce la interacción de un virus y huésped particulares que conducen a una situación de emergencia viral, es decir, la aparición de nuevos aislados con propiedades radicalmente distintas. Por este motivo, la mayoría de las enfermedades de etiología viral son causadas por virus de ARN (Domingo y Holland, 1997) y son una amenaza para la salud humana la agricultura y la ganadería. Entre las enfermedades infecciosas emergentes que afectan a las plantas, los virus causan casi la mitad de ellas (Anderson et al., 2004).

Todas estas características, junto con tiempos de generación muy cortos y grandes tamaños poblacionales, ha hecho de estos patógenos un excelente modelo para estudiar la evolución. Y es que, comprender los mecanismos evolutivos por los cuales un virus puede convertirse en emergente es el primer paso para diseñar estrategias de control y terapias antivirales (McDonald y Linde, 2002).

Resumen

Somos conscientes de que existen claras diferencias en la organización del material hereditario en todos los organismos y que puede alterarse en escalas de tiempo relativamente cortas. En otras palabras, nos referimos a la evolución de la arquitectura del genoma y predecirla es particularmente difícil, debido a la falta de conocimiento de algunos de sus principios fundamentales. Sí poseemos información sobre las tasas de mutación y recombinación, el papel de la deriva genética o la distribución de los efectos mutacionales. Sin embargo, la información sobre la topografía de los paisajes adaptativos sigue siendo escasa y no sabemos cómo puede condicionar la evolución del virus. Los paisajes adaptativos son un concepto fundamental en biología evolutiva que trata de discernir el número de vías mutacionales accesibles para alcanzar picos de eficacia (es decir, soluciones adaptativas). Durante mucho tiempo, los paisajes fueron sólo construcciones teóricas, pero actualmente pueden caracterizarse las propiedades de los paisajes empíricos gracias a la aportación de información precisa sobre los distintos genotipos. En relación a dicha aportación, uno de los parámetros más importantes es la eficacia viral, que se ha convertido en un cálculo estándar en virología y es esencial para evaluar el potencial evolutivo de un genotipo. Básicamente, define la ventaja (o desventaja) de un genotipo dado con respecto a uno de referencia. Por lo general, las poblaciones de virus se componen de una mezcla de genotipos cuya abundancia en la población depende de las diferencias de eficacia, las interacciones entre genotipos y la deriva genética. Pero, ¿cuántos picos de eficacia puede poseer un paisaje adaptativo? ¿Serían los mismos en otras condiciones ambientales? ¿Es equitativa la contribución de la adaptación, la historia y el azar en el proceso evolutivo? ¿Hasta qué punto son relevantes las interacciones epistáticas para determinar la rugosidad del paisaje adaptativo? ¿Y la regulación de los genes del huésped durante la infección? ¿Podemos identificar los genes del mismo cuya expresión depende de la eficacia viral? ¿Pertenece a la misma clase funcional? Para contestar a éstas y muchas otras preguntas, hemos dividido los resultados de esta tesis en cuatro capítulos.

La primera parte del trabajo tratará de mostrar cómo se mueven las poblaciones virales en un paisaje de eficacia rugoso y si lo hacen eficientemente. También se evaluará si pueden escapar de la cuenca de atracción de un pico de eficacia óptima local. Para ello, partiendo de un conjunto de genotipos del virus, que se sitúan a distintas distancias del óptimo local en un paisaje muy rugoso, observaremos si la evolución recompensa mantenerse en el óptimo local o, por el contrario, premia la exploración de regiones distantes del paisaje. En cualquier caso, determinaremos la contribución de la adaptación, la contingencia histórica y los eventos fortuitos tales como las mutaciones en el resultado final.

La segunda parte volverá a centrarse en las propiedades del paisaje de eficacia empírico. En este caso, compararemos las topografías de los paisajes de dos huéspedes, original y alternativo, en busca de similitudes y diferencias. Así, analizaremos si se producen ganancias o pérdidas de eficacia en distintas condiciones y observaremos si destaca algún tipo de interacción epistática. Además, sin olvidarnos de todas las demás características que definen un paisaje de eficacia, trataremos de realizar un ejercicio de valoración global para esclarecer si son más importantes los detalles macroscópicos del paisaje o los microscópicos.

El propósito del tercer capítulo se aleja momentáneamente del tema principal de esta tesis, los paisajes de eficacia empíricos, para tratar de evaluar *in vivo* las diferencias de eficacia entre componentes subclonales de una población. Para hacerlo, trataremos de aislar centenares de subclones biológicos individuales de una población viral mediante ensayos de infectividad en un huésped alternativo, en el que el virus objeto de estudio produce lesiones locales. Hecho esto, infectaremos el huésped original con cada uno de dichos componentes subclonales para evaluar su eficacia según el número de días transcurridos tras la inoculación.

Por último, habiendo analizado en profundidad los paisajes de eficacia, ahondaremos en el estudio de las diferencias entre genotipos virales a nivel de la respuesta molecular del huésped a la infección. Para tal objetivo, compararemos los transcriptomas de plantas infectadas con siete genotipos virales, que diferirán en su eficacia. Nuestra expectativa inicial es obtener un gran volumen de datos, dadas las características de la tecnología de micromatrices de oligos, y emplearlos para averiguar si los perfiles transcriptómicos son más parecidos o disímiles según la eficacia de los genotipos. Por otra parte, de existir diferencias, pudiéramos intentar dar una explicación lógica si logramos identificar un conjunto de genes correlacionados con la eficacia y las categorías funcionales a las que pertenecen.

Resum

Som conscients que hi ha clares diferències en l'organització del material hereditari en tots els organismes i que pot alterar-se en escales de temps relativament curtes. En altres paraules, ens referim a l'evolució de l'arquitectura del genoma i predir-la és particularment difícil, a causa de la falta de coneixement d'alguns dels seus principis fonamentals. Sí que posseïm informació sobre les taxes de mutació i recombinació, el paper de la deriva genètica o la distribució dels efectes de les mutacions. No obstant això, la informació sobre la topografia dels paisatges adaptatius continua sent escassa i no coneguem com pot condicionar l'evolució del virus. Els paisatges adaptatius són un concepte fonamental en biologia evolutiva que tracta de aclarir el nombre de vies mutacionals accessibles per a arribar als pics d'eficàcia (és a dir, solucions adaptatives). Durant molt de temps, els paisatges van ser només construccions teòriques, però ara poden caracteritzar-se les propietats dels paisatges empírics gràcies a l'aportació d'informació precisa dels distints genotips. En relació a la dita aportació, un dels paràmetres més importants és l'eficàcia viral, que s'ha convertit en un càlcul estàndard en virologia i és essencial per a avaluar el potencial evolutiu d'un genotip. Bàsicament, definix l'avantatge (o desavantatge) d'un genotip respecte a un de referència. Generalment, les poblacions de virus es componen d'una mescla de genotips l'abundància dels quals en la població depén de les diferències d'eficàcia, les interaccions entre genotips i la deriva genètica. Però, quants pics d'eficàcia pot posseir un paisatge adaptatiu? Serien els mateixos en altres condicions ambientals? És equitativa la contribució de l'adaptació, la història i l'atzar en el procés evolutiu? Fins a quin punt són rellevants les interaccions epistàtiques per a determinar la rugositat del paisatge adaptatiu? I la regulació dels gens de l'hoste durant la infecció? Podem identificar els gens de l'hoste l'expressió dels quals depén de l'eficàcia viral? Pertanyen a la mateixa classe funcional? Per a contestar a aquestes i moltes altres preguntes, hem dividit els resultats d'esta tesi en quatre capítols.

La primera part del treball tractarà de mostrar com es mouen les poblacions virals en un paisatge d'eficàcia rugós i si ho fan eficientment. També s'avaluarà si poden escapar de la conca d'atracció d'un pic d'eficàcia òptima local. Per a aixó, partint d'un conjunt de genotips del virus que se situen a distintes distàncies de l'òptim local en un paisatge molt rugós, observarem si l'evolució recompensa mantindre's en l'òptim local o, al contrari, premia l'exploració de regions distants del paisatge. En tot cas, determinarem la contribució de l'adaptació, la contingència històrica i els esdeveniments fortuits com ara les mutacions en el resultat final.

La segona part tornarà a centrar-se en les propietats del paisatge d'eficàcia empíric. En este cas, compararem les topografies dels paisatges de dos hostes, original i alternatiu, a la cerca de similituds i diferències. Així, analitzarem si es produeixen guanys o pèrdues d'eficàcia en distintes condicions i observarem si destaca algun tipus d'interacció epistática. A més, sense oblidar-nos de totes les altres característiques que definixen un paisatge d'eficàcia, tractarem de realitzar un exercici de valoració global per a aclarir si són més importants els detalls macroscòpics del paisatge o els microscòpics.

El propòsit del tercer capítol s'allunya momentàniament del tema principal d'esta tesi, els paisatges d'eficàcia empírics, per a tractar d'avaluar *in vivo* les diferències d'eficàcia entre components subclonals d'una població. Per a fer-ho, tractarem d'aïllar centenars de subclons biològics individuals d'una població viral per mitjà d'assajos d'infectivitat en un hoste alternatiu, en el que el virus objecte d'estudi produeix lesions locals. Fet açò, infectarem l'hoste original amb cada un d'estos components subclonals per a avaluar la seua eficàcia segons el nombre de dies transcorreguts després de la inoculació.

Finalment, després d'analitzar en profunditat els paisatges d'eficàcia, aprofundirem en l'estudi de les diferències entre genotips virals a nivell de la resposta molecular de l'hoste a la infecció. Per a tal objectiu, compararem els transcriptomas de plantes infectades amb set genotips virals, que es diferencien en la seua eficàcia. La nostra expectativa inicial és obtindre un gran volum de dades, donades les característiques de la tecnologia de micromatrius d'oligos, i emprar-los per a esbrinar si els perfils transcriptòmics són més pareguts o dissímils segons l'eficàcia dels genotips. D'altra banda, d'existir diferències, poguérem intentar donar una explicació lògica si aconseguim identificar un conjunt de gens correlacionats amb l'eficàcia i les categories funcionals a què pertanyen.

Abstract

We are aware that there are clear differences in the organisation of hereditary material in all organisms and that it can be altered at relatively short timescales. In other words, predicting the evolution of genome architecture is particularly difficult, due to a lack of knowledge of some of its fundamental principles. We have information about mutation and recombination rates, the role of genetic drift, or the distribution of mutational effects. However, information about the topography of adaptive landscapes remains scarce and we do not know how it can condition the evolution of the virus. Adaptive landscapes are a fundamental concept in evolutionary biology that seeks to discern the number of mutational pathways accessible to achieve fitness peaks (*i.e.*, adaptive solutions). For a long time, landscapes were only theoretical constructions, but today the properties of empirical landscapes can be characterized thanks to the contribution of precise information on the different genotypes. One of the most important parameters in relation to this contribution is viral fitness, which has become a standard parameter in virology and is essential for assessing the evolutionary potential of a genotype. Basically, it defines the advantage (or disadvantage) of a given genotype over a reference genotype. In general, virus populations are composed of a mixture of genotypes whose abundance in the population depends on differences in fitness, interactions between genotypes and genetic drift. But how many fitness peaks can have an adaptive landscape? Would they be the same in other environmental conditions? Is equal the contribution of adaptation, history and chance in the evolutionary process? To what extent are epistatic interactions relevant to determining the roughness of the adaptive landscape? What about regulation of host genes during infection? Can we identify the host genes whose expression depends on viral fitness? Do they belong to the same functional class? To answer these and many other questions, we have divided the results of this thesis into four chapters.

The first part of the work will try to show how viral populations move in a rugged fitness landscape and if they do so efficiently. It will also be evaluated whether they can escape from the catchment area of a local peak of maximum fitness. To this end, starting from a set of genotypes of the virus, which are located at different distances from the local optimum in a very rugged landscape, we will observe whether evolution rewards staying in the local optimum or, on the other way, rewards the exploration of another regions of the landscape. In either case, we will determine the contribution of adaptation, historical contingency and fortuitous events such as mutations in the final result.

The second part will focus again on the properties of the empirical fitness landscapes. In this case, we will compare the topographies of the landscapes of two hosts, the original and an alternative one, in search of similarities and differences. Thus, we will analyze if there are gains or losses in fitness under different conditions and we will look if any epistatic interaction is highlighted. Furthermore, without forgetting all the other characteristics that define a fitness landscape, we will try to carry out a global valuation exercise to clarify whether macroscopic or microscopic details of the landscape are more important.

The purpose of the third chapter momentarily moves away from the main theme of this thesis, empirical fitness landscapes, to try to assess *in vivo* the differences in fitness between subclonal components of a population. To do this, we will attempt to isolate hundreds of individual biological subclones of a viral population through infectivity assays in an alternative host, where the virus produces local lesions. Once this is done, we will infect the original host with each of these subclonal components to assess their fitness based on the number of days after inoculation.

Finally, having thoroughly analyzed the empirical fitness landscapes, we will study the differences between viral genotypes at the level of molecular response in infected hosts. For this purpose, we will study the profiles of plants infected with seven viral genotypes, which will differ in their fitness. Our initial expectation is to obtain a large volume of data, given the characteristics of microarray technology, and use them to find out if the transcriptomic profiles are more similar or dissimilar according to the genotypes fitness. On the other hand, if there are differences, we could try to give a logical explanation if we can identify a set of genes correlated with fitness and the functional categories to which they belong.

1. Introducción

1. Introducción al mundo de los virus.

Las formas de vida en la Tierra se cuentan por millones y resultan de lo más diversas. Cuando hablamos de ellas habitualmente pensamos en el mundo macroscópico que somos capaces de percibir directamente con nuestros sentidos, es decir, pensamos en la gran diversidad de animales y plantas que encontramos a lo largo y ancho del globo. Todos ellos han debido lidiar en el transcurso de millones de años contra condiciones climáticas cambiantes, competidores en su mismo nicho ecológico o depredadores. Sin embargo, existe una lucha constante de todo organismo vivo de la que no se tuvo constancia hasta hace poco más de un siglo. Dicho conflicto surge del descubrimiento en el ecosistema de unos organismos conocidos como virus.

Los virus son parásitos intracelulares obligados que no son capaces de replicarse por sí solos, pues carecen de maquinaria propia para ejercer la biosíntesis de proteínas. Para tal fin deben aprovecharse de la maquinaria metabólica de aquel organismo que le sirve como huésped. Así pues, desarrollan la mayor parte de su ciclo vital en el interior de las células del mismo. Por lo general, los virus se componen de una cadena de ácidos nucleicos protegidos por una cubierta conformada por proteínas y, en algunos casos, de lípidos procedentes de la membrana celular. A dicha cubierta se le denomina típicamente cápsida, y además de proteger el material genético del virus también puede funcionar como vehículo de transmisión. Al conjunto de ambos componentes, material genético y cápsida, se le denomina virión y, típicamente, se encuentra en el medio extracelular y es capaz de iniciar una infección. La diversidad de los virus es enorme, tanto en lo que respecta a su material genético (que al no poseer mecanismos de corrección y reparación es propenso a errores durante la replicación) como a su cubierta (cuya composición puede variar desde unidades multiméricas de un solo polipéptido hasta complejos heteroméricos). Por ejemplo, los miembros de la familia *Parvoviridae* se encuentran entre los virus

Introducción

más pequeños, con un genoma lineal de una sola cadena de aproximadamente 5 kilobases (Berns, 1990). Del otro extremo, los *Pandoravirus* se sitúan como algunos de los virus más grandes del mundo, tanto en tamaño físico como en su genoma, que puede llegar a llegar a alcanzar 2,5 megabases (Philippe et al., 2013). Estos patógenos se encuentran distribuidos alrededor de todo el planeta infectando cualquier todo tipo de bacterias, arqueas, hongos, plantas y animales, e incluso a otros virus (La Scola et al., 2008; Claverie y Abergel, 2009). Esta última afirmación no debe llevarnos a error; es habitual que un mismo virus logre infectar de forma sistémica más de un organismo distinto, incluso decenas, comprendidos siempre dentro de la misma familia o género, pero no es habitual que logre infectar organismos muy distantes filogenéticamente. Como todo en Biología, siempre existen excepciones a la norma, como los *Tospovirus*, que además de replicarse en la planta huésped también lo hacen en los insectos que actúan como vectores de transmisión (Li et al., 2014; Flenniken, 2014).

2. Clasificación de los virus.

Una de las características más relevantes de los virus es que la composición y estructura de sus genomas presentan una enorme variabilidad (Levine, 1991; Levy et al., 1994; Fields et al., 1996; Cann, 2005). En primer lugar, puede estar compuesto tanto por ADN como por ARN, el cual pudiera ser monocatenario o bicatenario. En caso de hallarnos en presencia de un virus monocatenario también hay que valorar una característica adicional, y es que la cadena puede poseer polaridad positiva, al igual que el ARN mensajero (ARNm), o negativa.

La clasificación mostrada en la **Figura 1** corresponde al sistema de Baltimore, la cual se basa en la estructura y composición del genoma viral (Baltimore, 1971). Este esquema da cuenta de que cada genoma debe sintetizar un ARNm para que pueda ser traducido por los ribosomas del huésped.

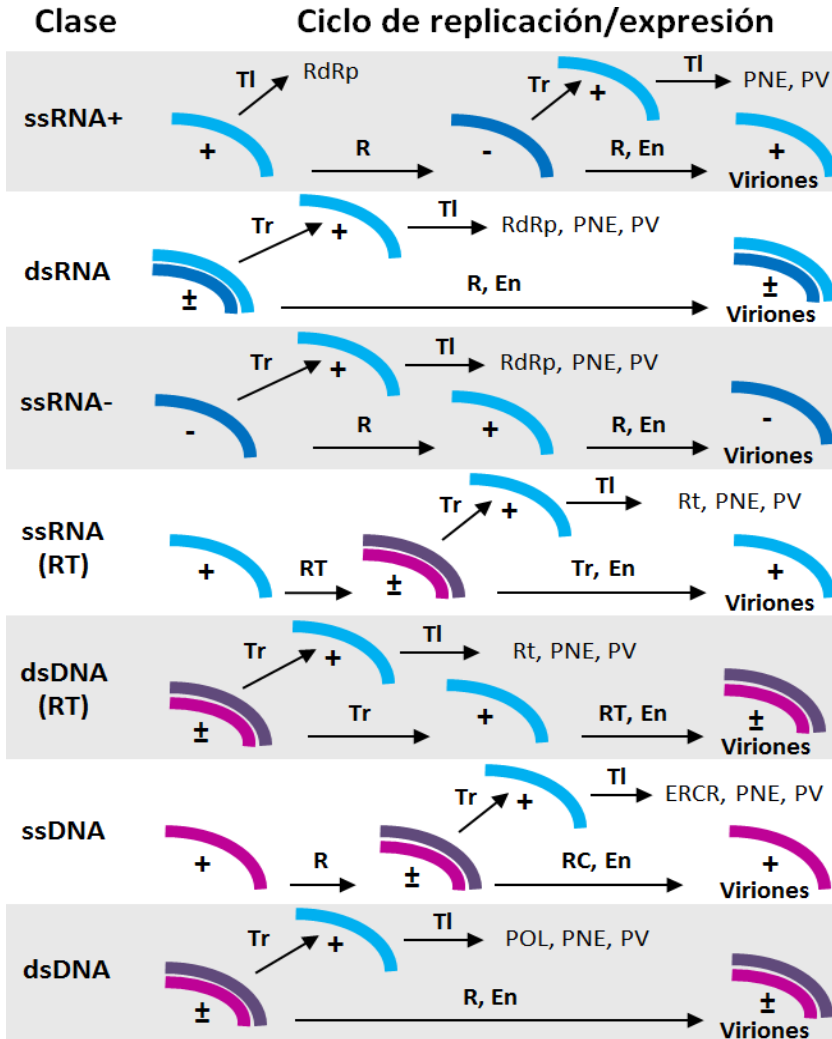


Figura 1. Representación de la diversidad de los virus. Se muestran las siete clases junto con sus ciclos de replicación y expresión. **Tl**, traducción; **Tr**, transcripción; **R**, replicación; **RC**, replicación en círculo rodante; **RT**, transcripción inversa; **Rt**, retrotranscriptasa; **POL**, polimerasa; **En**, encapsidación; **NSP**, proteínas no estructurales; **RdRp**, ARN-polimerasa dependiente de ARN; **VP**, proteínas del virión; **PNE**, proteína no estructural; **RCRE**, endonucleasa de replicación en círculo rodante; (+), cadena positiva; (-), cadena negativa; (±), bicatenario. (Figura extraída de [Koonin y Dolja, 2013](#)).

Introducción

Hoy día sabemos que la mayoría de los virus almacenan su información genética en forma de ARN. Además, como ya avanzamos, gozan de una enorme dispersión y se hallan infectando todo tipo de plantas y animales. Centrando nuestra atención en los virus de ARN monocatenario de polaridad positiva, todos codifican en su genoma una ARN-polimerasa dependiente de ARN (RdRp) implicada en la replicación, ya que sintetiza la cadena de polaridad negativa complementaria que sirve de molde para dar lugar a nuevas cadenas de polaridad positiva (Brinton et al., 1994). Respecto a éstos, se clasifican actualmente en tres órdenes y 33 familias, además de constituir el tipo de virus más común que infecta plantas (Modrow et al., 2013; Saxena and Lomonosoff, 2014).

Para el cumplimiento de los objetivos de esta tesis, hemos elegido como modelo experimental un virus de ARN de polaridad positiva, el virus del grabado del tabaco (*Tobacco etch virus* o TEV), miembro del género *Potyvirus* de la familia *Potyviridae*. La familia *Potyviridae* pertenece, a su vez, a la superfamilia tipo *Picornia*, a la que pertenecen otros virus tan importantes y conocidos como el de la hepatitis C, la fiebre aftosa o la polio.

3. Virus de ARN.

Tras introducir los diversos tipos de virus que podemos encontrar y habiendo destacado los virus de ARN de cadena simple positiva como nuestro modelo experimental, procederemos a describir de forma breve algunas de sus características.

3.1 Arquitectura del genoma.

La característica más notoria de los virus de ARN quizá sea el tamaño de su genoma, que suele ser pequeño, típicamente alrededor de 10 kilobases (Holmes, 2009; Elena et al., 2011; Campillo-Balderas

et al., 2015). Si nos detenemos a observar aquellos que son capaces de provocar una infección en plantas, que en su mayoría son virus de ARN, el tamaño medio oscila alrededor de las 10 kilobases (Figura 2). Hoy día no se sabe cuál es la razón para tal limitación, pero no faltan teorías al respecto. Por ejemplo, pudiera deberse simplemente a la fragilidad química del ARN (Cann, 2005), a la necesidad de desplazarse a través de los plasmodesmos (Roossinck, 2008), una menor complejidad enzimática que conlleva mayor variabilidad (Szathmáry y Maynard Smith, 1995), al aumento de la velocidad de replicación (Holmes, 2003) o a la reducción del número de mutaciones deletéreas limitando el tamaño del genoma (Holmes, 2003). Además, cabe destacar que los virus de ARN codifican un número limitado de genes. Para el caso de virus de ARN de polaridad positiva, que son los que nos interesan, desde 3 hasta 12 genes (Flint et al., 2015).

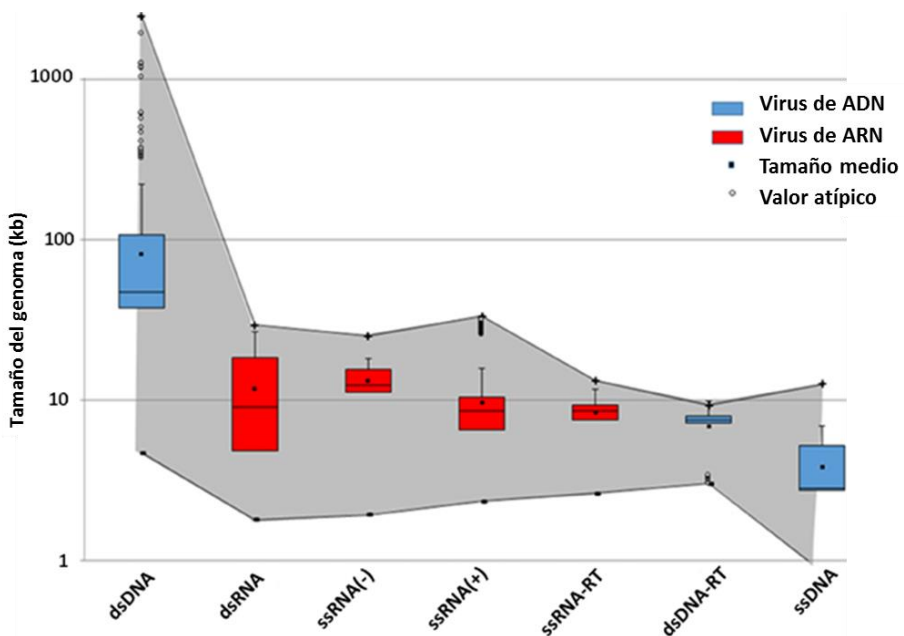


Figura 2. Representación con diagramas de cajas del tamaño máximo (límite superior) y mínimo (límite inferior) del genoma de las siete clases de virus. (Figura adaptada de Campillo-Balderas et al., 2015).

Introducción

Otra de las características de las que debemos hablar al referirnos a la arquitectura del genoma es la superposición de genes (Lynch, 2006). Como puede observarse en la **Figura 3**, dicho fenómeno permite comprimir información de modo que aumente el número de proteínas que pueden traducirse sin aumentar el tamaño del genoma (Belshaw et al., 2007; Brandes y Linial, 2016).

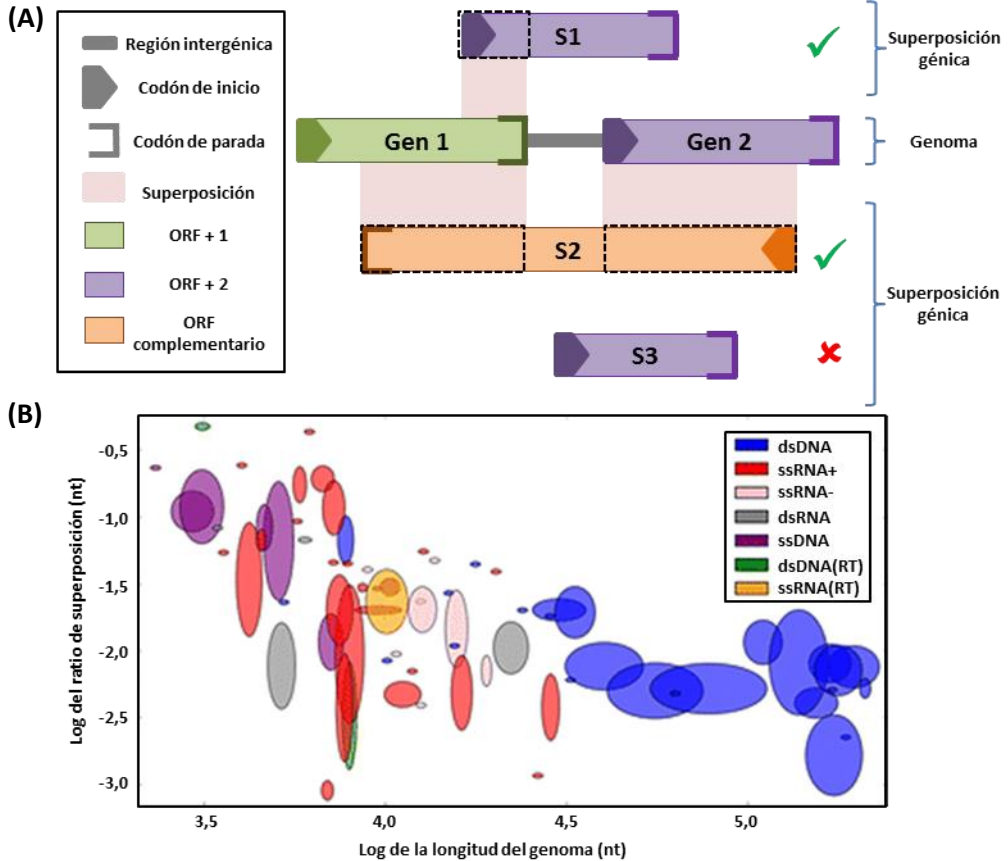


Figura 3. Representación de la superposición génica en la que los marcos de lectura solapantes de dos genes conducen a la expresión de una proteína nueva en otro marco de lectura (A). Diagrama de dispersión que muestra la correlación negativa entre las longitudes del genoma y la tasa de superposición en las siete clases de virus (B). (Figura adaptada de Brandes y Linial, 2016).

3.2 Ciclo de vida.

El ciclo de vida completo de un virus de ARN puede describirse en tres fases, que son la llegada al organismo huésped, el inicio y progreso de la replicación viral y el movimiento entre células para establecer una infección sistémica (Hull, 2002). A continuación, detallaremos brevemente estas fases tomando como ejemplo un virus de ARN de polaridad positiva (TEV) y su huésped vegetal natural, la planta de tabaco (*Nicotiana tabacum* L.).

Respecto a la llegada y entrada en el organismo huésped, la situación más común es que se produzca la transmisión activa del virus mediada por un vector animal. En nuestro ejemplo, los pulgones (también llamados áfidos) actúan como vectores para TEV (Figura 4). Aunque muchas especies pueden transmitir de manera no persistente el virus, la especie *Myzus persicae* (pulgón verde del melocotón) cobra especial relevancia en dicha tarea, pues posee tamaños poblacionales muy grandes y comparte en el medio natural con su huésped. El mecanismo por el cual interaccionan el estilete del pulgón y la proteína multifuncional **HC-Pro** del virus ha sido ampliamente estudiado (Brault et al., 2010; Blanc et al., 2014).

Una vez en el medio intracelular, el virus debe iniciar el proceso de replicación, el cual se halla totalmente detallado en la Figura 5. En resumidas cuentas, primero tiene lugar la desencapsidación del material genético en el citoplasma de las células del huésped. Este tipo de virus no encapsidan la RdRp. Así, en un primer momento, la cadena de polaridad positiva sirve como molde análogo al ARNm que sí puede ser traducido por los ribosomas de la célula huésped, es decir, el material genético del virus se traduce por primera vez y se obtiene una poliproteína. De ésta, surgen el resto de proteínas virales mediante un procesamiento proteolítico, pues el genoma de TEV codifica para tres proteinasas, **P1**, **HC-Pro** y **Nla-Pro**, que escinden la poliproteína en sitios específicos (Carrington et al., 1989).

Introducción

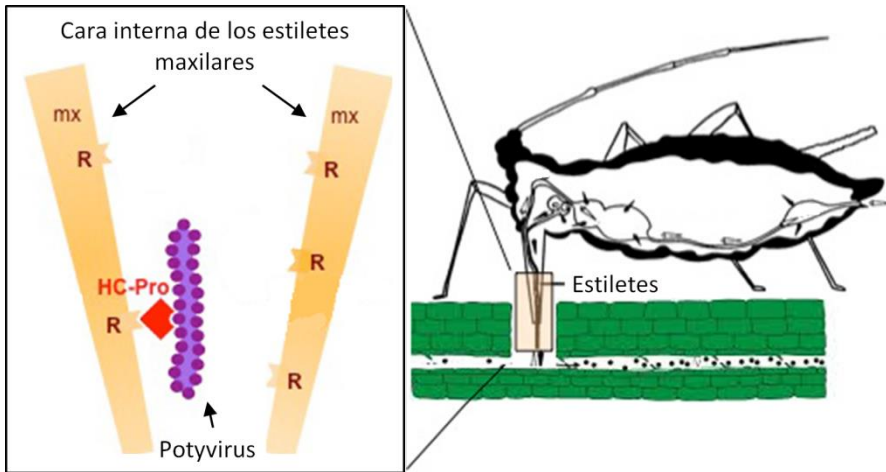


Figura 4. Representación del acoplamiento del virión de TEV en el estilete del pulgón. El virión (extracelular) se une indirectamente a las moléculas diana (**R**) mediante una proteína previamente codificada por el virus (intracelular) en un huésped anterior. Al clavar el estilete para alimentarse del floema de la planta huésped, la proteína multifuncional **HC-Pro** ejerce de enlace molecular entre la molécula diana y el virión, lo que permite que, al desplazarse a otra planta para seguir alimentándose, el pulgón traslade el virión. (Figura adaptada de umr-bgpi.cirad.fr/equipes/equipe2-detail-projet.uzest-uk.htm).

Una vez se hallan presentes todas las proteínas virales en el medio intracelular, la RdRp (**N1b**) emplea el molde de polaridad positiva inicial para ejercer la transcripción (en lugar de la traducción). Con todo ello, se sintetiza la cadena de polaridad negativa complementaria que, a su vez, servirá para dar lugar a nuevas cadenas de polaridad positiva (Brinton et al., 1994). El resultado es una producción asimétrica de ARN en que el número de cadenas de polaridad positiva termina excediendo el número de cadenas negativas, es decir, se producen unas pocas cadenas intermediarias negativas que sirvan como plantillas para la

producción de progenie de polaridad positiva (Elena et al., 2008; Martínez et al., 2011). Esto es característico de los virus de ARN de polaridad positiva, ya que a continuación toda la progenie podrá encapsidarse para formar viriones completos que podrán establecer una infección sistémica.

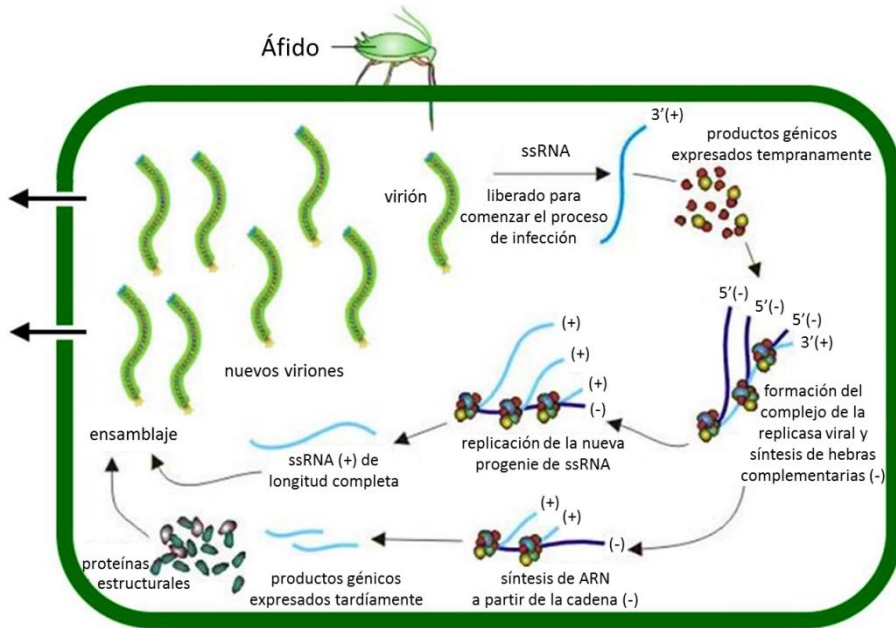


Figura 5. Representación esquemática del ciclo completo de un virus de ARN monocatenario de polaridad positiva en una célula vegetal. El ciclo parte de la introducción del virión en la célula por parte del vector animal, el pulgón (situado en el margen superior). Continúa en el sentido de las agujas del reloj con la traducción inicial del material genético, con la posterior transcripción que da lugar a la replicación y culmina con la encapsidación de la progenie. (Figura extraída de Lalić, 2012).

Introducción

En la última fase, el virus se desplaza desde la primera célula infectada hasta producir la infección sistémica de la planta. En la dinámica de la infección sistémica, el foco primario de la infección en la hoja, que se corresponde con la zona en que se alimentó el pulgón, crece radialmente al desplazarse los viriones a las células vecinas a través de los plasmodesmos. Llega un momento en que, al alcanzar el sistema vascular del huésped, pueden extenderse a otros tejidos e iniciar de nuevo el proceso de replicación en ellos. Se considera en este momento que la infección es sistémica. Sin embargo, la rapidez con que se produce la infección sistémica depende de la naturaleza de la interacción entre el virus y el organismo huésped (Rodrigo et al., 2014). Hoy día no se sabe con certeza cómo funciona el mecanismo por el cual los viriones consiguen desplazarse de una célula a otra. Sin embargo, recientemente se han publicado diversos estudios que apuntarían a que el complejo formado por el virión y la helicasa viral **CI** se uniría a **P3N-PIPO**, a través de un dominio de **CI**. Después, **P3N-PIPO** transportaría el complejo a la membrana plasmática uniéndose a **PCaP1** (Ide et al., 2007). El complejo unido a **PCaP1** se anclaría a la membrana plasmática mediante una reacción de miristoilación en el residuo de glicina que ocupa la posición 2 en el extremo N-terminal (Nagasaki et al., 2008) y se acumularía en los plasmodesmos debido a la afinidad de la proteína **CI** por éstos (Vijayapalani et al., 2012). Así, **PCaP1** permitiría el movimiento célula a célula de los viriones (Figura 6). El problema que suscita esta hipótesis es que se ha probado con un potyvirus distinto en un modelo experimental distinto, concretamente, el virus del mosaico del nabo (*Turnip mosaic virus* o TuMV) y su huésped natural, *Arabidopsis thaliana*. En resumen, sabemos que la coordinación intracelular de los elementos virales (Mäkinen y Hafrén, 2014) y la variabilidad intra-específica de la proteína **P3N-PIPO** son de gran relevancia biológica durante el establecimiento de la infección (Hillung et al., 2013), pero no se ha reproducido el experimento para el potyvirus TEV y el huésped *N. tabacum* (principalmente debido a que no se posee la anotación completa del genoma de *N. tabacum*) ni se ha profundizado en cómo el virus interactúa o regula los genes de su huésped para tal fin.

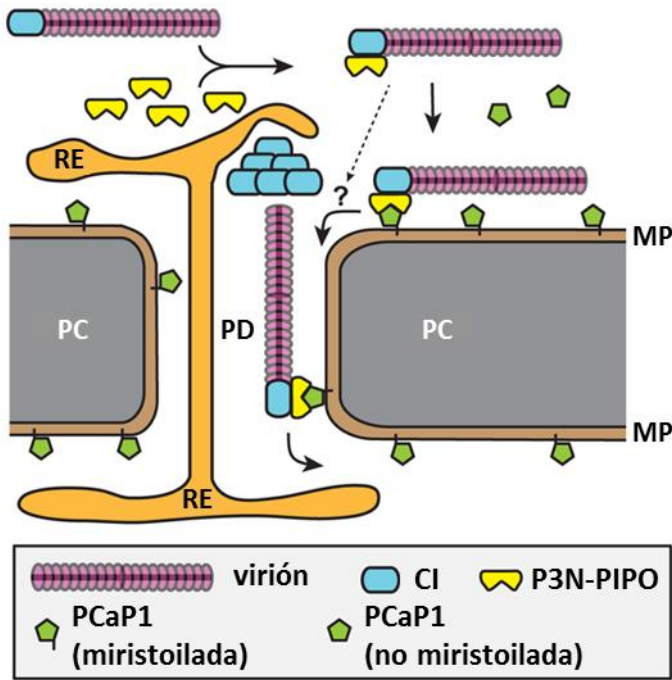


Figura 6. Esquema que muestra la hipótesis de la interacción entre PCaP1 y P3N-PIPO para llevar a cabo el movimiento del TuMV a través de los plasmodesmos. **RE**, retículo endoplasmático; **PD**, plasmodesmo; **MP**, membrana plasmática; **PC**, pared celular. (Figura extraída de [Vijayapalani et al., 2012](#)).

3.3 La familia *Potyviridae*.

La familia *Potyviridae* se compone en su totalidad de virus de ARN de cadena simple y polaridad positiva que infectan plantas. A día de hoy, engloba 195 especies divididas en 8 géneros (**Figura 7**): *Brambyvirus* (que incluye una única especie), *Bymovirus* (seis), *Ipomovirus* (seis), *Macluravirus* (ocho), *Poacevirus* (tres), *Potyvirus* (160), *Rymovirus* (tres) y *Tritimovirus* (seis). Todas ellas suman 193 especies, las dos especies restantes no han sido asignadas todavía a ningún género.

Introducción

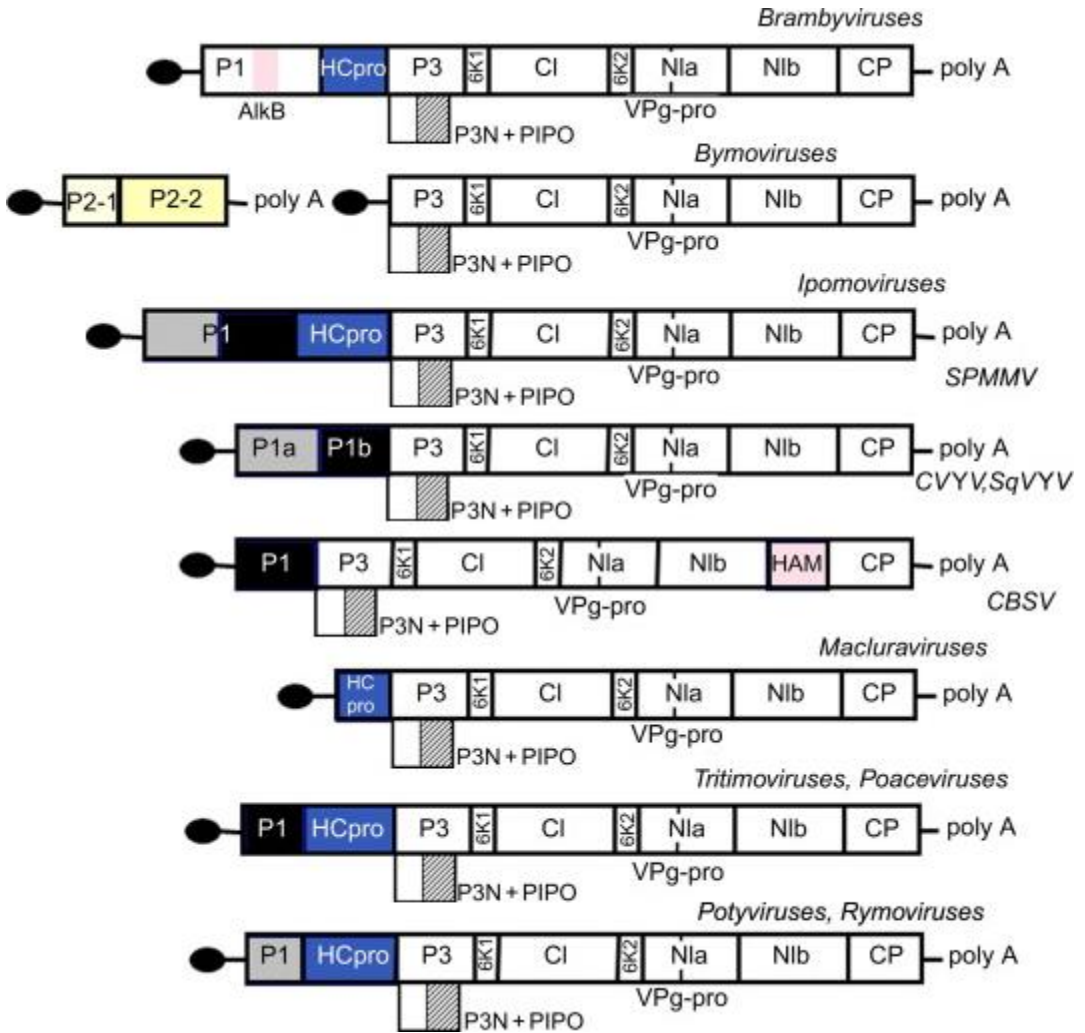


Figura 7. Esquema de la organización genómica de los ocho géneros de la familia *Potyviridae*. Todos los virus comprendidos en estos géneros tienen un genoma monopartito, a excepción del género *Bymovirus*, que posee un genoma bipartito. El género *Ipomovirus*, aunque sólo incluye seis especies, es el más diverso, por ello se representan tres especies tipo de dicho género. Los virus de esta familia codifican entre 8 y 10 genes. El orden de los genes de los diferentes géneros dentro de la familia *Potyviridae* está muy bien conservado. (Figura extraída de [Revers y García, 2005](#)).

Como ya avanzamos, para el cumplimiento de los objetivos de esta tesis, elegimos como modelo experimental TEV. Éste pertenece al género *Potyvirus*, el más grande de todos los incluidos dentro de la familia *Potyviridae*. Además, se halla distribuido por todo el globo y tiene un amplio rango de huéspedes, principalmente plantas en la familia *Solanaceae* (como los géneros *Nicotiana* y *Capsicum*) (Shukla et al., 1994). La infección por parte de TEV induce habitualmente un grabado necrótico en la hoja (**apartado 2 de Material y Métodos**).

Los viriones de TEV están formados por unas 2000 unidades proteicas y tienen estructura filamentosa de simetría helicoidal (**Figura 8**). Miden aproximadamente 680-900 nanómetros de largo y 11-15 de diámetro (Dougherty et al., 1988; King et al., 2011).

Respecto a la arquitectura del genoma de TEV, éste posee 9.539 nucleótidos y contiene una sola pauta lectura que codifica para una poliproteína de aproximadamente 350 kDa que, como ya describimos, es procesada por tres proteinasas para dar lugar a las proteínas maduras **P1**, **HC-Pro**, **P3**, **6K1**, **CI**, **6K2**, **VPg**, **NIa-Pro**, **NIb** y **CP**. Además, hay una pauta de lectura alternativa que da lugar a la proteína **P3N-PIPO** (solapada dentro de la región que codifica **P3**). Los extremos del genoma están flanqueados por regiones no codificantes, estando asociado el extremo 5' con la proteína viral **VPg** y el extremo 3' con una cola de poliadenina (**Figura 9**).

No podemos terminar este apartado sin destacar el amplio abanico de funciones de las proteínas virales de TEV (**Tabla 1**). Por ejemplo, **HC-Pro** también está implicada en el anclaje al estilete de los pulgones para facilitar su transmisión, la supresión del silenciamiento génico o el desarrollo de síntomas (Llave et al., 2000; Redondo et al., 2001). Otro ejemplo, la proteína **VPg**, junto con la formación de una estructura secundaria en la región no codificante 5', sirve como anclaje para los ribosomas del huésped con tal de iniciar la traducción del ARN viral. Por último, la proteína de la cápsida (**CP**) interviene en la transmisión por pulgones, la protección y el movimiento del genoma viral (Callaway et al., 2001).

Introducción

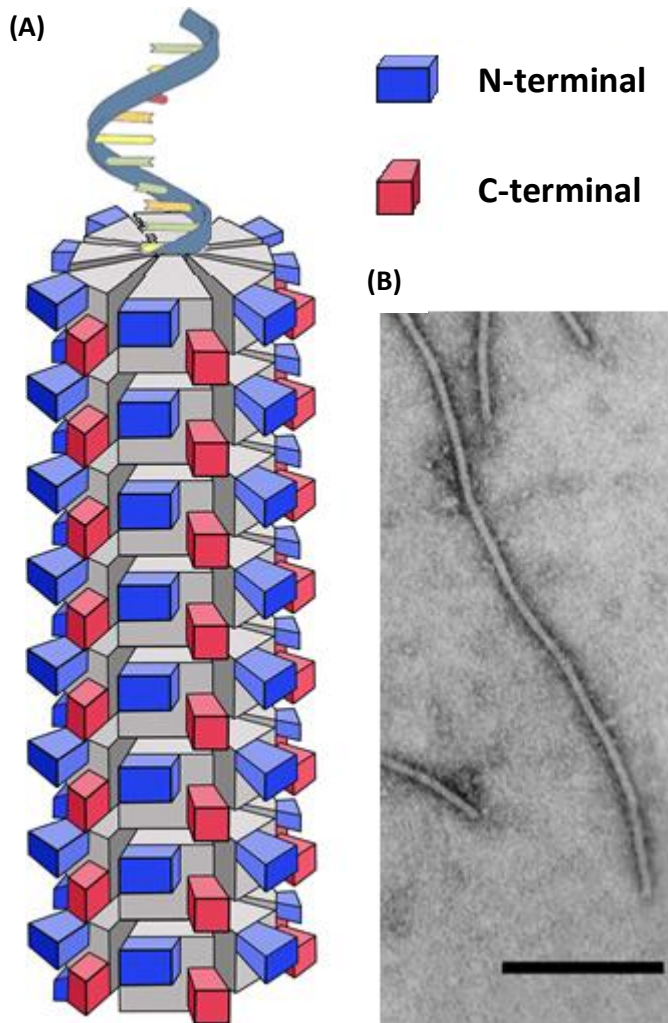


Figura 8. Esquema de un virión típico del género *Potyvirus* en que se presenta la disposición de los extremos terminales de la proteína de la cápsida. El extremo N-terminal consta de aproximadamente 30 aminoácidos, mientras que el extremo C-terminal consta de 19 aminoácidos (A). Fotografía obtenida mediante microscopía electrónica de contraste negativo de un virión del género *Potyvirus* (B). (Figura adaptada de Shukla y Ward, 1989; Dekker et al., 1993; King et al., 2012; Lalić, 2012).

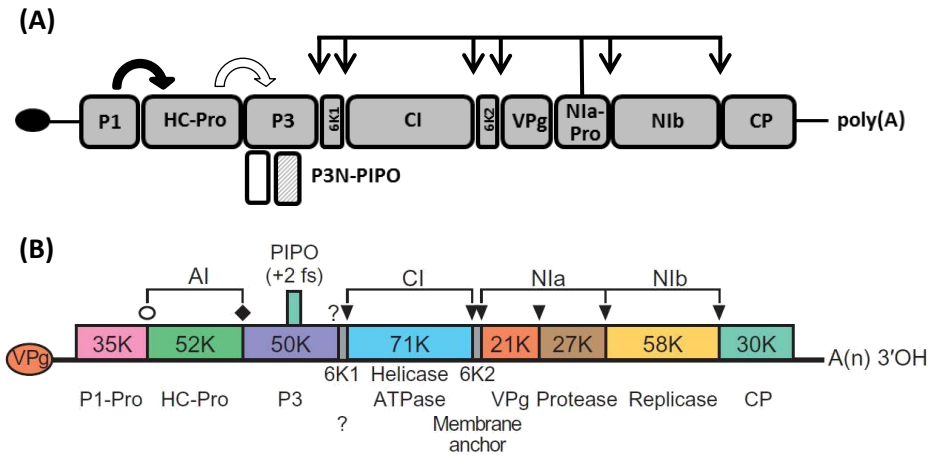


Figura 9. Esquema de la arquitectura del genoma de TEV, en que cada cuadro representa una proteína madura individual que se obtiene tras el procesamiento de la poliproteína por parte de las tres proteinasas. La flecha curva negra corresponde a la escisión llevada a cabo por la serina proteinasa **P1**, la flecha curva blanca a la escisión llevada a cabo por la cisteína proteinasa **HC-Pro** y las flechas verticales negras a las escisiones llevadas a cabo por la cisteína proteinasa **Nia-Pro**. Las cajas inferiores muestran el marco de lectura alternativo que da lugar a la proteína **P3N-PIPO** (A). Mapa genómico de TEV en que se muestra la función más relevante asociada con la proteína madura. En este caso, se especifica el tipo de actividad proteolítica llevada a cabo por las proteinasas: **P1**, responsable de la escisión en **Tyr/Phe-Ser** (○); **HC-Pro**, escisión en **Gly-Gly** (◆); **Nia-Pro**, escisión en **Gln/Glu-(Ser/Gly/Ala)** (▼). La proteína implicada y el tipo de cuerpo de inclusión se muestran encima del mapa genético; **AI**, inclusión amorfa; **CI**, cuerpo de inclusión de forma cilíndrica localizado en el citoplasma; **Nia** y **Nib**, proteínas que se agregan en el núcleo para formar un cuerpo de inclusión nuclear. También se especifica el peso de cada proteína madura (en kDa) y vuelve a mostrarse el marco de lectura alternativo que da lugar a **P3N-PIPO** (B). En ambos se muestran las regiones flanqueantes no codificantes. (Figura adaptada de [Revers y García, 2005](#); [King et al., 2012](#); [Lalić, 2012](#); [Hillung, 2015](#); [Willemsen, 2016](#)).

Introducción

Tabla 1. Resumen de las funciones de las proteínas virales de TEV. (Tabla extraída de [Hillung, 2015](#)).

| Proteína | Funciones |
|-----------------|---|
| P1 | <ul style="list-style-type: none">▪ Serina proteinasa▪ Factor accesorio para la amplificación del virus▪ Adaptación al huésped |
| HC-Pro | <ul style="list-style-type: none">▪ Cisteína proteinasa▪ Factor ayudante para la transmisión por pulgones▪ Supresión del silenciamiento del ARN▪ Aumento de cantidad de partículas virales |
| P3 | <ul style="list-style-type: none">▪ Amplificación del virus▪ Adaptación al huésped |
| P3N-PIPO | <ul style="list-style-type: none">▪ Movimiento célula a célula del virus |
| 6K1 | <ul style="list-style-type: none">▪ Posible modulación de la actividad de P3 |
| CI | <ul style="list-style-type: none">▪ Helicasa del ARN▪ Replicación del ARN▪ Formación de estructuras de molinillo con aspas▪ Movimiento del virus célula a célula |
| 6K2 | <ul style="list-style-type: none">▪ Proliferación de vesículas de membrana▪ Transporte hacia la membrana |
| VPg | <ul style="list-style-type: none">▪ Proteína ligada al extremo 5' del genoma viral▪ Cebador de replicación viral▪ Traducción del ARN▪ Movimiento del virus |
| NlaPro | <ul style="list-style-type: none">▪ Cisteína proteinasa▪ ADNasa |
| Nlb | <ul style="list-style-type: none">▪ ARN-polimerasa dependiente de molde de ARN (replicasa viral) |
| CP | <ul style="list-style-type: none">▪ Protección del ARN genómico▪ Movimiento del virus▪ Transmisión por pulgones |

4. Mutaciones como fuente de adaptación y evolución.

La razón del éxito evolutivo de los virus de ARN se basa en su capacidad para generar enjambres mutantes altamente polimórficos que respondan eficientemente ante los cambios ambientales (como pudieran ser, un cambio en las defensas del huésped o, incluso, un cambio de huésped). Esta capacidad no puede explicarse como consecuencia de un determinado factor o elemento, sino que se trata de la combinación de tres factores distintos.

El primer factor sería la elevada tasa de mutación ([Sanjuán et al., 2010](#)), que permite la aparición de nuevos genotipos virales sobre los cuales podrá operar la selección natural y la deriva genética. Tratándose de virus de ARN, el evento de mutación más frecuente es la sustitución nucleotídica, generalmente debido a que las RdRp son propensas a cometer errores durante la replicación ([Flint et al., 2015](#)). Las incorporaciones incorrectas durante la amplificación del genoma viral por parte de las RdRp resultan en un aumento de la tasa de mutación de hasta dos órdenes de magnitud, en comparación con las ADN-polimerasas ([García-Díaz y Bebeneck, 2007](#)). Es lógico pensar que cuanto mayor sea la tasa de error mayores oportunidades surgirán para lograr la adaptación. Sin embargo, la gran mayoría de estas mutaciones son deletéreas. Aquí entra en juego el siguiente factor, el tamaño de la población ([Wasik y Turner, 2013](#)), pues una carga mutacional excesiva, a pesar de permitir la exploración rápida del espacio genotípico, puede generar grandes cantidades de mutaciones deletéreas y genotipos inviables que pueden poner en peligro la viabilidad de poblaciones pequeñas ([Gabriel et al., 1993](#)). Así, en una población de gran tamaño esperamos que las mutaciones deletéreas tengan una probabilidad insignificante de fijarse, porque la selección natural tiende a eliminarlas. Sólo aquellas que resulten beneficiosas y escapen a la pérdida estocástica de la deriva genética contribuirán a la adaptación ([Elena y Sanjuán, 2005](#)). El último factor a considerar sería el tiempo de generación, el cual es muy corto para los virus de ARN ([Wasik y Turner, 2013](#)).

4.1 Estimaciones de la tasa de mutación de los virus.

La tasa de mutación espontánea es un parámetro clave para entender la estructura genética de las poblaciones virales en el tiempo. Contemplando todo el espectro de virus conocidos, aquellos cuyo genoma está conformado por ARN son los que poseen las tasas de mutación más elevadas (**Figura 10**). Por ejemplo, se estima que la tasa de mutación para los virus de ADN oscila entre 10^{-8} y 10^{-6} sustituciones por nucleótido y ronda de replicación (Drake, 1991; Sanjuán et al., 2010). Centrándonos en el grupo que nos interesa, la tasa de mutación de los virus de ARN es mayor y se estima entre 10^{-6} y 10^{-4} (Drake et al., 1998; Sanjuán et al., 2010). Ello significa que hay aproximadamente una mutación por genoma y ciclo de replicación. Como ya adelantamos, la variabilidad entre los genomas individuales es una consecuencia inevitable de la replicación de los virus de ARN. Así pues, el elemento que más influye en la tasa de mutación es la RdRp que, a diferencia de las ADN polimerasas dependientes de ADN, no posee ningún mecanismo de corrección del material genético y es más propensa a cometer errores (Drake et al., 1998).

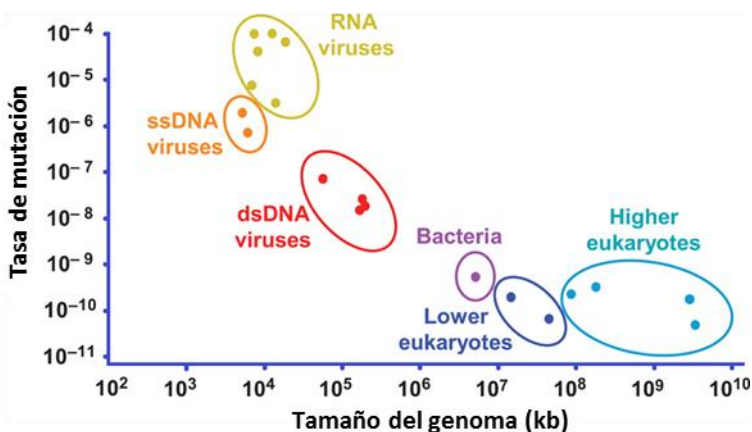


Figura 10. Gráfico representativo de la relación entre la tasa de mutación y el tamaño del genoma de diferentes organismos. (Figura adaptada de Gago et al., 2009).

Además de la RdRp, hay otros dos elementos que condicionan una mayor o menor tasa de mutación, aunque influyen en menor medida. El primero es la presencia de estructuras secundarias de ARN, para las cuales se ha mostrado que portan más mutaciones, o son susceptibles de adquirir más, que el resto del genoma viral (Pathak y Temin, 1992). Incluso una mutación sinónima en una estructura secundaria pudiera conducir a un genotipo letal, lo que indica que la presencia de éstas en el genoma del virus puede afectar a la tasa de mutación (Simmonds et al., 2004). El segundo es el organismo huésped en el que tiene lugar la replicación (Pita et al., 2007). Respecto a este último factor, cabe destacar que la elevada capacidad de los virus de ARN para generar enjambres mutantes es especialmente importante para la supervivencia de aquellos cuyo huésped es una planta anual, ya que les permite adaptarse y alojarse en diferentes huéspedes en las épocas en que el huésped principal no es accesible (Schneider y Roossinck, 2001).

Como hemos visto, las tasas de mutación de los virus de ARN son altas, pero para nuestro modelo experimental, TEV, la tasa de mutación se estima en $3 \cdot 10^{-5}$ sustituciones por nucleótido y ronda de replicación (Sanjuán et al., 2009; Tromas et al., 2014b), un valor algo más bajo de lo esperado que se encuentra en el rango inferior de la distribución de los virus de ARN y muy cercano al rango superior de los virus de DNA monocatenarios (Figura 10). Quitando de la importancia de la RdRp, en que una mutación pudiera conducir a un cambio drástico de la eficacia o a un genotipo letal, se postula que la tasa de mutación inusualmente baja para el caso concreto de TEV depende más de la regulación ejercida por la planta de tabaco huésped que de la presencia de estructuras secundarias en el genoma viral. Esta idea se apoya en la evidencia de que las mutaciones en la región no codificante 5', donde hallamos las estructuras secundarias de TEV (Figura 11), provocan pérdidas de eficacia menores que las mutaciones a lo largo del resto del genoma (Bernet y Elena, 2015). De este modo, la interacción entre la región 5' y los ribosomas del huésped sería robusta a las mutaciones.

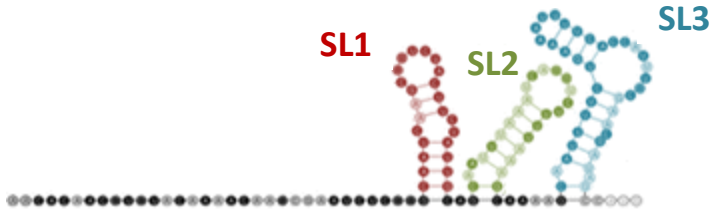


Figura 11. Esquema de las estructuras secundarias de ARN en la región no codificante 5'. Cada una se representa con un color distinto (SL1, SL2 y SL3). Estas tres estructuras definen el sitio de anclaje del ribosoma del huésped y son esenciales para la expresión de todos los genes virales. **SL**, estructura tallo-lazo. (Figura adaptada de [Bernet y Elena, 2015](#)).

4.2 Interacciones entre mutaciones.

Las mutaciones se pueden calificar como neutrales, beneficiosas o deletéreas. Las mutaciones neutrales se acumulan por deriva genética a velocidad constante y únicamente dependen del tamaño de la población y de la tasa de mutación ([Kimura, 1983](#)). Las mutaciones beneficiosas o deletéreas son consecuencia de la acción de la selección positiva o negativa, respectivamente, sobre ellas. La selección positiva favorece la fijación de mutaciones adaptativas de carácter beneficioso, mientras que la selección negativa elimina o tiende a eliminar las deletéreas. En el caso de los virus, pudiéramos distinguir tres componentes que participarían en dicha selección, siendo el más importante el organismo huésped, que provee el ambiente en que se ejercerá la selección. Los otros dos serían el tamaño de la población y su heterogeneidad ([Ohta y Ina, 1995](#)).

Asumiendo que el medio ambiente se reduce en su mayor parte al huésped, no todos los huéspedes potenciales de un virus (sean diferentes especies o diferentes genotipos dentro de la misma especie) son igualmente susceptibles a la infección ([Agrawal y Lively, 2003](#)). En este contexto, se denomina especialista a un virus que

posee uno o muy pocos huéspedes potenciales, ya que ha evolucionado para adaptarse a una o dos especies huéspedes o genotipos (Turner and Elena, 2000), mientras que si ha evolucionado para adaptarse a un gran abanico de especies huéspedes se le denomina generalista (Remold et al., 2008). La transferencia del virus entre huéspedes no conlleva directamente una infección sistémica, sino que resulta en una infección puntual y limitada. Si este nuevo ambiente permite la replicación del virus de forma mínimamente eficiente, la adaptación del virus al nuevo huésped va a estar determinada en gran medida por interacciones entre alelos de distintos locus, es decir, procesos epistáticos (Figura 12).

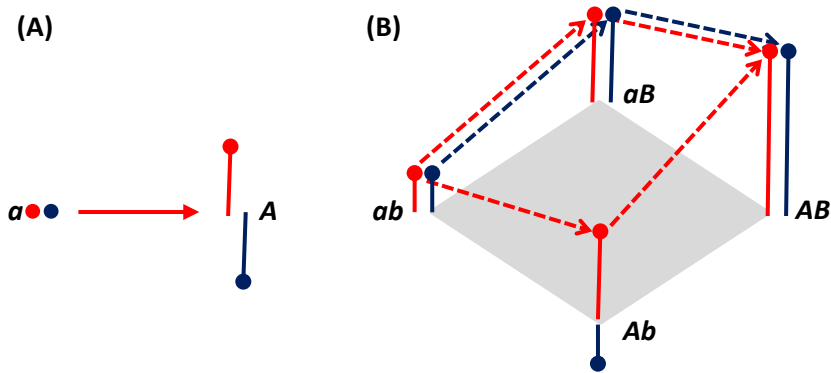


Figura 12. Interacción entre alelos de uno o varios locus en dos ambientes diferentes. El alelo silvestre se denota con letra **minúscula** y el mutante con letra **MAYÚSCULA**. Los ambientes son denotados como huésped azul y huésped rojo. Si sólo se halla implicado un locus, **A** es beneficioso en el huésped rojo. Al haber implicado un único locus con un único alelo no hay ningún tipo de interacción epistática (A). Si en la expresión de un carácter fenotípico se hallan implicados dos locus, puede darse la interacción epistática. En el ejemplo mostrado, el mutante doble **AB** es el genotipo más eficaz en cualquiera de los huéspedes considerados. Por el contrario, el mutante simple **aB** conserva su eficacia en ambos huéspedes pero el mutante simple **Ab** es deletéreo en el huésped azul (B).

Introducción

Por el momento, no entraremos a analizar en detalle ni los tipos de epistasia ni el efecto de las mutaciones sobre la eficacia, puesto que más adelante nos extenderemos en la explicación y los cálculos matemáticos relacionados con ambos (**apartado 5 de Material y Métodos**).

Volviendo al tema que nos ocupa, si el virus no es capaz *a priori* de replicarse en el huésped nuevo, su proliferación va a depender de la preexistencia de mutantes que sean capaces de iniciar la replicación y recuperar la eficacia de la población viral (**Figura 13**). Respecto a este conjunto de mutantes, que forman un grupo de virus con genomas estrechamente relacionados, la unidad de selección no es el virus individual, sino el conjunto de genotipos interconectados en el paisaje definido por el huésped ([Miralles et al., 1997](#)). Este enjambre mutante constituye una reserva de posibilidades evolutivas, de las cuales se espera que al menos una sea eficaz en las condiciones ambientales impuestas por el nuevo huésped. La propiedad de los virus de mantener diferentes genotipos mutantes relacionados en la población sin que afecten a su fenotipo se denomina robustez mutacional, considerada una característica esencial en la capacidad adaptativa de los virus, pues les permite hacer frente a las variaciones acaecidas en su entorno ([Elena et al., 2006](#)).

5. Los paisajes adaptativos de eficacia.

Como hemos visto, la epistasia y la pleiotropía interactúan mutuamente de una manera muy intrincada ([Remold y Lenski, 2004](#); [Lalić y Elena, 2013](#)). En el último apartado de la introducción, que figura a continuación, trataremos de ofrecer una visión global de los paisajes de eficacia y cómo las interacciones epistáticas entre dos loci determinan la rugosidad del mismo.

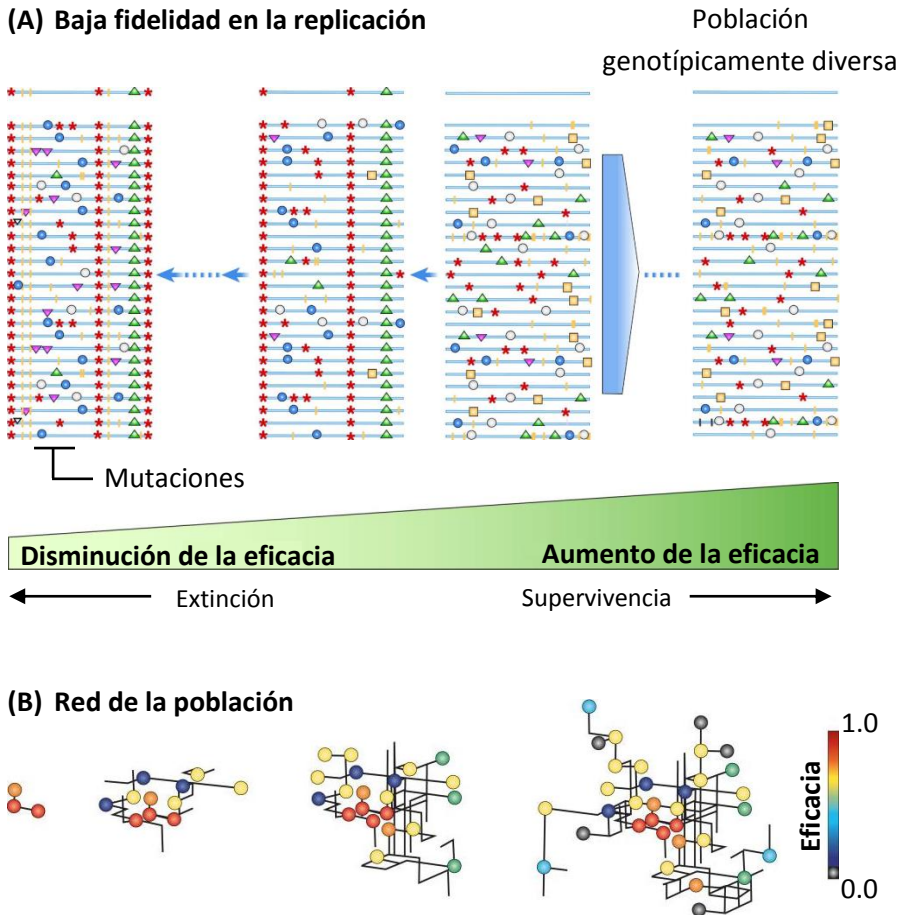


Figura 13. Esquema de la eficacia de un enjambre mutante viral. Los genomas mutantes se muestran como líneas horizontales y los símbolos ilustran las diversas mutaciones debidas a la poca fidelidad de la RdRp. La línea superior es la secuencia consenso. La flecha grande muestra el pase de toda la población al nuevo huésped, que resulta en un amplio abanico de posibilidades que conlleva la ganancia de eficacia. Los cuellos de botella (flechas pequeñas) resultan en una disminución de la eficacia **(A)**. Red en las que los genotipos mutantes (círculos) de diversa eficacia están conectados por cambios mutacionales (líneas negras) **(B)**. (Figura adaptada de Domingo y Wain-Hobson, 2009; Domingo et al., 2012; Lauring et al., 2013).

Introducción

La eficacia es un parámetro complejo empleado inicialmente para describir cuantitativamente la capacidad reproductiva y el potencial evolutivo de un organismo en un entorno dado (Orr, 2009; Linnen y Hoekstra, 2010). Aplicado a organismos tales como los virus de plantas, definiría la capacidad de éstos para producir una progenie infecciosa en el entorno proporcionado por un huésped (Domingo y Holland, 1997). Sin embargo, la medición de este parámetro es muy compleja y no tiene en cuenta sólo uno o dos componentes, sino que describe los efectos combinados de todas las propiedades fenotípicas para la supervivencia y reproducción de un determinado genotipo en un determinado huésped (Lenski, 1991). Este conjunto de propiedades se define como eficacia replicativa (Domingo, 2010) y tiene en cuenta la entrada al medio intracelular, la traducción, la replicación, la encapsidación, el movimiento entre células o tejidos distantes, el silenciamiento génico y la transmisión, entre otras. Todas estas propiedades tienen en común la interacción con la maquinaria metabólica del huésped (Ahlquist, 2003; Wang y Li, 2012), salvo la entrada y la transmisión, que en el caso de virus de plantas, dependen del vector.

En el campo de la Virología, la eficacia replicativa se ha convertido en un cálculo estándar para comparar diversos genotipos virales durante el proceso de infección (Holland et al., 1991). La comparación de dichos genotipos de acuerdo a su rendimiento en un huésped particular se ejerce en base a un genotipo de referencia (Wargo y Kurath, 2012). Dicho cálculo ha sido fundamental durante el transcurso de las últimas tres décadas para la comprensión de muchos procesos evolutivos, añadiendo un componente cuantitativo al efecto de los cuellos de botella (Chao, 1990), la acumulación de mutaciones deletéreas (Duarte et al., 1992), la dinámica de adaptación a nuevos huéspedes (Novella et al., 1995a), el coste de ampliar la gama de huéspedes (Agudelo-Romero et al., 2008b; Novella et al., 1995b; Turner y Elena, 2000), la complejidad del genoma (Pesko et al., 2015; Zwart y Elena, 2015; Willemsen et al., 2016b), las duplicaciones (Willemsen et al., 2016a), el escape hacia

otros picos de eficacia (Novella et al., 2004), la topografía de los paisajes y la robustez mutacional (Sanjuán et al., 2007).

Una lista que incluya los valores de la eficacia replicativa de un conjunto de genotipos es difícil de interpretar por sí misma. De ahí que se idearan los paisajes de eficacia, que son unos gráficos tridimensionales, visualmente muy atractivos, que describen el proceso de la evolución indicando el movimiento de las poblaciones en un espacio compuesto por una superficie plana, valles y picos (Wright, 1932).

En la práctica, el paisaje describe la eficacia de todos los posibles genotipos que componen el llamado espacio genotípico. En una proyección bidimensional, el eje vertical, que se corresponde con la altura de los picos, indica la eficacia de cada genotipo, que se definen por el eje horizontal (Figura 14).

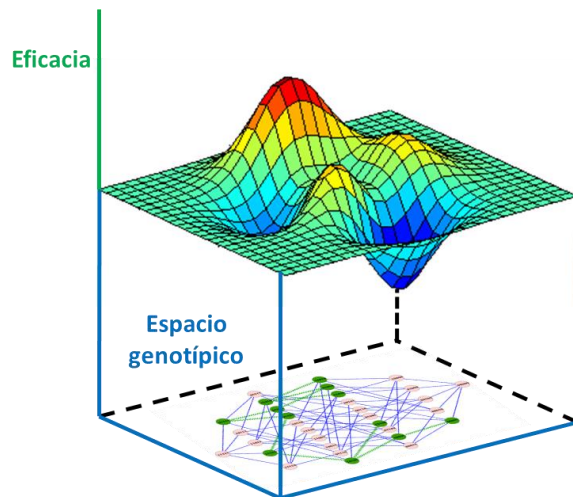


Figura 14. Esquema de un paisaje de eficacia. En el eje vertical indica la eficacia relativa al genotipo de referencia. El eje horizontal que da profundidad al gráfico representa el espacio genotípico, el cual se ha extendido hacia abajo para ilustrar la vinculación de la variación al nivel de nucleótidos con sus consecuencias en la eficacia de las poblaciones. (Figura adaptada de vincentcastric.weebly.com).

Introducción

El paisaje de eficacia es una aproximación visual de las características más destacables de nuestro enjambre mutante en su nuevo ambiente. Sin embargo, otro apartado fundamental es inferir un paisaje empíricamente a partir de la construcción de las relaciones entre todos los genotipos posibles. Nos referimos a la representación de la topografía empírica del paisaje de eficacia (**Figura 15**). La adaptación es siempre un proceso de escalada donde una población es empujada hacia la cima del pico de eficacia más cercano. En este marco, la evolución se produce a través de una red continua de genotipos intermedios que representan cada uno de los pasos mutacionales que conducen al aumento de eficacia. Sin embargo, los genotipos intermedios situados en los valles, que se caracterizan por su baja eficacia o letalidad, dificultan este proceso. Este otro tipo de gráfico detalla las trayectorias mutacionales por las que podría haber ocurrido la adaptación ([Weinreich et al., 2006](#)).

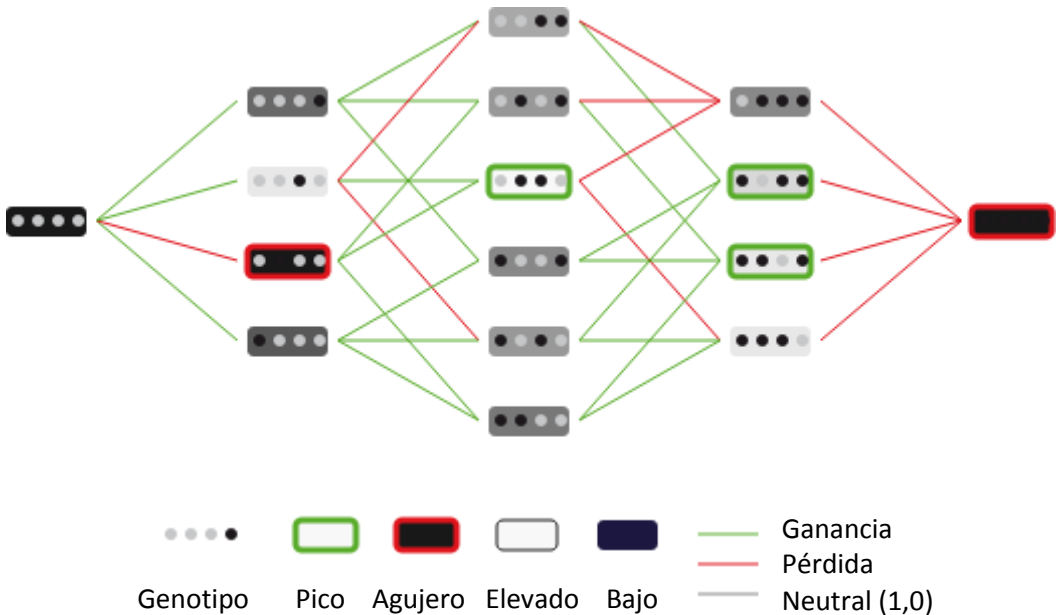


Figura 15. Ejemplo de la topografía empírica de un paisaje de eficacia que hemos computado utilizando el servidor web MAGELLAN. Para este ejemplo se ha introducido un número de cuatro loci.

Así, sabiendo que el genoma de TEV se compone de 9539 nucleótidos (Lalić y Elena, 2015), el número de genotipos que pudiéramos incluir en un paisaje adaptativo es de 4^{9539} . Puesto que es muy aparatoso manejar semejante volumen de genotipos, habitualmente se construyen los paisajes para una pequeña colección de mutantes simples y los mutantes que corresponden a la combinación de estas mutaciones simples (Flynn et al., 2013).

Establezcamos una serie de ejemplos para introducir la visualización e interpretación de estos paisajes. En la **Figura 16** tenemos la representación tridimensional de un paisaje situada a la izquierda, y a la derecha tenemos su correspondiente topografía empírica. La representación tridimensional cuenta con el eje vertical (medida de eficacia) y el eje horizontal (espacio genotípico). Respecto a la topografía empírica, cada conjunto de letras es un genotipo viral, con **minúsculas** representando los alelos silvestres y **MAYÚSCULAS** los alelos mutantes. Los genotipos se ordenan comenzando por el genotipo silvestre y terminando con el genotipo que acumula todas las mutaciones posibles. Cada fila representa genotipos con igual número de mutaciones. Los genotipos marcados en rojo son aquellos con una eficacia inferior a la del silvestre (incluyendo letales), en negro los que tienen una eficacia igual al silvestre, y en verde los que son significativamente más eficaces que el silvestre. Las líneas verdes sólidas representan caminos adaptativos (que siempre conectan genotipos beneficiosos), las líneas negras discontinuas conectan genotipos no significativamente distintos en cuanto a eficacia (rutas neutras) y las líneas rojas representan caminos deletéreos.

Para proceder con el ejemplo, supongamos que un virus posee un genoma conformado por sólo tres genes, que denominaremos **a**, **b** y **c**. Dicho genoma, que sería el silvestre, lo seleccionaríamos como genotipo de referencia a la hora de establecer comparaciones con otros genomas. En un determinado momento se presenta un cambio ambiental representado por un nuevo huésped. En este punto, pueden sucederse varias situaciones distintas.

Introducción

En primer lugar, pudiera darse el caso de que cualquier mutación en cualquier gen del genoma viral tuviese una eficacia independiente de qué alelos haya en el resto de genes. Este escenario da lugar a un paisaje conformado por un amasijo de picos locales de eficacia, aunque no necesariamente de la misma altura (todos los genotipos son viables pero pudieran presentar diferencias en su eficacia replicativa). Este paisaje define el caso extremo del modelo de “castillo de naipes” (Aita et al., 2001; Kingman, 1987), en el que se consideraría que un virus podría adaptarse al nuevo huésped a partir de cualquier cambio en su genoma (Figura 16.A).

En segundo lugar, pudiera darse el caso menos radical del modelo puramente multiplicativo. Esto significa que sólo un genotipo será viable (en nuestro ejemplo es el doble mutante **AbC**). Este escenario da lugar a un paisaje conformado por un sólo pico global de eficacia, que necesariamente implica que el resto de genotipos serán siempre deletéreos. Este paisaje define el caso extremo del modelo del “Monte Fuji” (Aita et al., 2000; Aita et al., 2001), en el que un virus sólo dispondría de un único camino evolutivo (Figura 16.B).

Ninguno de los extremos anteriores sucede en la naturaleza, sino que las topografías de los paisajes empíricos reflejan escenarios aleatorios no correlacionados, situados entre ambos casos extremos. Así, se conforman paisajes con diferentes picos que representan soluciones alternativas al problema de la adaptación (Gavrilets, 2010). Los genotipos que se sitúan cercanos en el espacio genotípico se diferencian por un cambio mutacional, de modo que la eficacia asociada a esos genotipos da lugar a una superficie dispuesta entre ellos más pronunciada o más plana. Los genotipos que tienen altos valores de eficacia ocupan los picos del paisaje, mientras que los de muy baja eficacia o letales ocupan los valles (Figura 16.C).

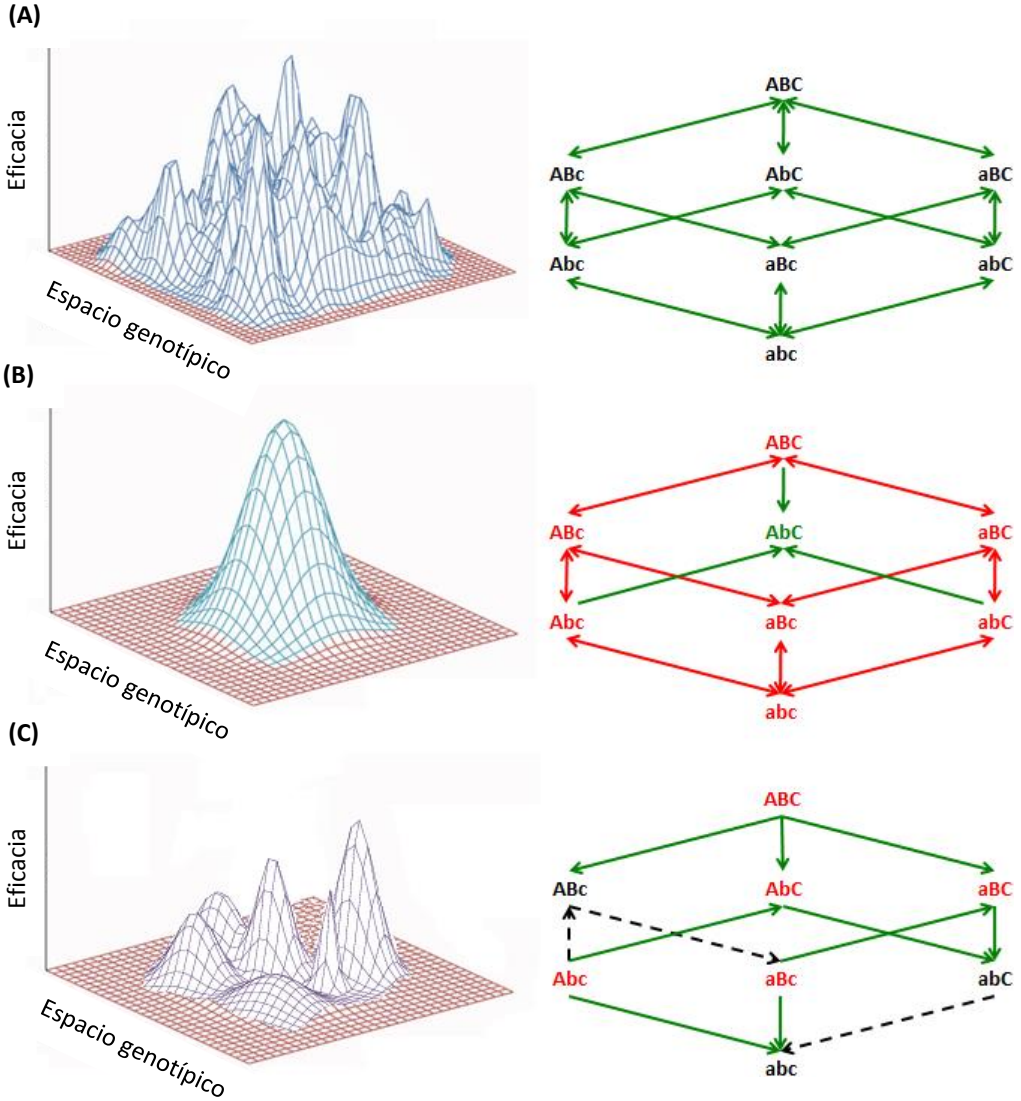


Figura 16. Representación tridimensional de la eficacia de un enjambre mutante en el entorno proporcionado por un nuevo huésped (**gráficos situados a la izquierda**). Representación de la topografía empírica del paisaje de eficacia local que describe la adaptación de un virus a su nuevo huésped (**gráficos situados a la derecha**).

Introducción

Respecto a estos escenarios, y para cualquier paisaje de eficacia que pretendamos analizar, la probabilidad de que una mutación aleatoria sea beneficiosa es muy pequeña. En principio, la evolución se produciría dada la acumulación de un gran número de mutaciones con pequeños efectos beneficiosos (Fisher, 1930). Dicho esto, cuanto menor sea el efecto fenotípico mutacional, mayor será la probabilidad de que la mutación sea beneficiosa. Sin embargo, además de beneficiosas, las mutaciones deben escapar a la pérdida estocástica. Al igual que ocurría anteriormente, cuanto menor sea el efecto de la mutación, mayor será la probabilidad de escapar a esa pérdida (Kimura, 1983). Con todo ello, las mutaciones de efecto pequeño, o incluso intermedio, son las que contribuirían en mayor medida a la adaptación (Orr, 1998).

Hasta ahora, hemos visto cómo pueden afectar las mutaciones a la eficacia replicativa de una población viral, pero el paisaje de eficacia subyacente puede verse influenciado por muchos otros factores como, por ejemplo, las interacciones epistáticas. Así, el paisaje que define el modelo del “Monte Fuji” (Aita et al., 2000; Aita et al., 2001) se compondría de una superficie lisa con un único pico de eficacia, que indicaría la ausencia de interacciones epistáticas o únicamente la presencia de interacciones de magnitud (De Visser y Krug, 2014) (Figura 17). Sin embargo, en presencia de interacciones epistáticas de signo (en las que la eficacia asociada a cada mutación depende del fondo genético) o de signo recíproco (en las que ambas mutaciones consideradas son individualmente deletéreas pero mutuamente beneficiosas) se crean curvaturas en el paisaje y se promueve el establecimiento de múltiples picos de elevada eficacia separados por valles de baja eficacia (De Visser et al., 2011; Weinreich et al., 2005) (Figura 17). En este paisaje rugoso las diferentes poblaciones pueden alcanzar diferentes picos, dependiendo de la accesibilidad a los mismos en relación con su posición inicial. En otras palabras, la evolución puede ser predecible dado que el número de trayectorias accesibles desde un punto concreto del paisaje es limitado (De Visser y Krug, 2014). De hecho, el

número de picos de eficacia en un paisaje depende, en última instancia, de si las mutaciones interactúan multiplicativa o epistáticamente (De Visser y Krug, 2014). Cabe destacar que cuanto mayor sea el número de interacciones epistáticas de signo (Weinreich et al., 2005) o signo recíproco (Poelwijk et al., 2011), más rugosos serán los paisajes y el número de trayectorias accesibles será más limitado (Kauffman y Levin, 1987; Szendro et al., 2013; Weinreich et al., 2013). Por ejemplo, al cambiar el patrón de interacciones epistáticas es posible acceder a trayectorias mutacionales que pueden ser beneficiosas en un huésped pero no en otro (Mustonen y Lässig, 2009; Matuszewski et al., 2014; Laughlin y Messier, 2015; Steinberg y Ostermeier, 2016).

De todos los conceptos introducidos en este apartado se desprende la idea de que la rugosidad de los paisajes adaptativos es crítica para predecir si las poblaciones alcanzarán el óptimo de eficacia global o, por el contrario, a través de trayectorias evolutivas alternativas, quedarán atrapadas en picos de eficacia locales o subóptimos (Poelwijk et al., 2007; Kvitek y Sherlock, 2011). Aunque hemos destacado las mutaciones y las interacciones epistáticas como los factores más importantes que influyen en la determinación de los paisajes de eficacia, no son los únicos. Por ejemplo, para explicar el movimiento de las poblaciones virales entre picos de eficacia debemos introducir el factor de la deriva genética. En poblaciones pequeñas, la deriva genética puede promover la fijación de una mutación deletérea si es seguida por la fijación de una mutación compensatoria que restaura la eficacia (Burch y Chao, 1999). De esta manera, pudieran modificarse el tipo de interacciones epistáticas en poblaciones pequeñas aisladas en un pico de eficacia local, lo que les permitiría atravesar valles de baja eficacia (dada la fijación de la mutación deletérea) y establecerse en otro pico (dada la fijación posterior e inmediata de la mutación beneficiosa) (Wright, 1932).

Introducción

Como es de esperar, la reproducibilidad de la evolución en paisajes tan complejos disminuye a medida que aumenta el número de posibles picos, es decir, conforme aumenta la rugosidad del paisaje. Dicho con otras palabras, la probabilidad de que la adaptación siga el mismo camino en un paisaje muy accidentado es menor, y debido a ello, disminuye nuestra capacidad de predecir el resultado evolutivo.

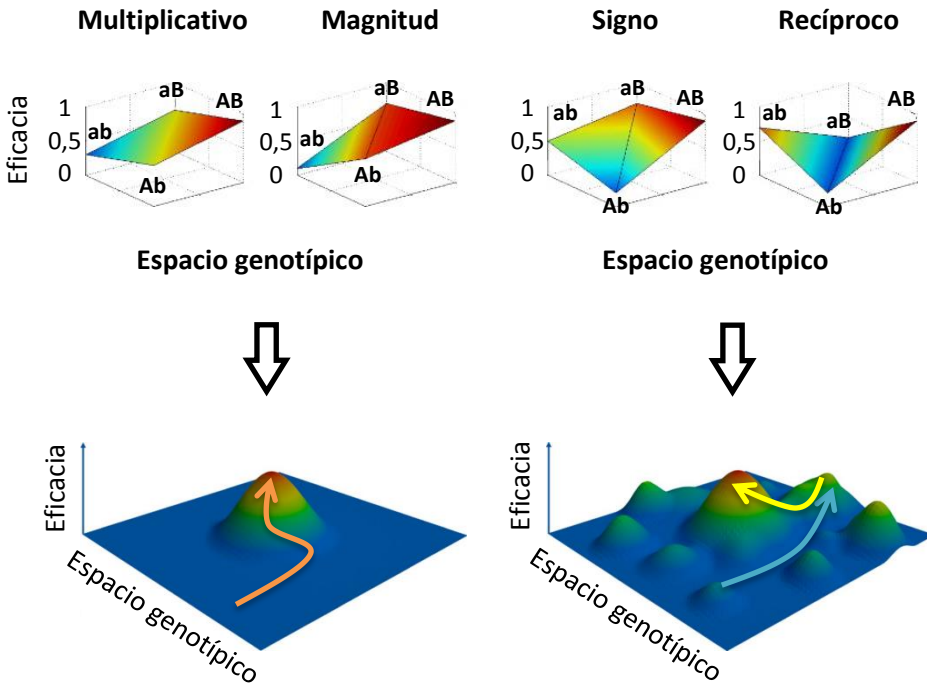


Figura 17. Ejemplo de los caminos evolutivos a seguir por una población viral para adaptarse al ambiente definido por un huésped nuevo. Cada conjunto de letras es un genotipo viral, con **minúsculas** representando los alelos silvestres y **MAYÚSCULAS** los alelos mutantes. El genotipo silvestre es **ab** y el genotipo mutante adaptado es el doble mutante **AB**. En ausencia de epistasia o en el caso de epistasia de magnitud, los genotipos intermedios (los mutantes simples **Ab** y **aB**) tienen valores de eficacia entre los genotipos silvestre (**ab**) y doble mutante (**AB**). En este caso, la selección natural

favorece y conduce la adaptación hacia el único pico óptimo global, ya que ambos genotipos intermedios tienen mayor eficacia que el silvestre. En los casos en que observamos epistasia de signo o signo recíproco uno o ambos genotipos intermedios tienen menor eficacia en comparación con el silvestre. En nuestro ejemplo, en el escenario de epistasia de signo el genotipo intermedio de baja eficacia (el mutante simple Ab) será seleccionado negativamente. Sin embargo, en el escenario de epistasia de signo recíproco ambos mutantes simples serán seleccionados en contra. Estos escenarios darán lugar a un paisaje de eficacia más rugoso. (Figura adaptada de [Gutiérrez y Maere, 2014](#); [Pesce et al., 2016](#)).

2. Objetivos

El objetivo general de esta tesis es el estudio de los factores que determinan la topografía de los paisajes de eficacia y ahondar en el análisis de los genes que pudieran resultar determinantes de la eficacia viral. Para ello, emplearemos una colección de genotipos mutantes del virus del grabado del tabaco (TEV) que serán evaluados en huéspedes diferentes.

Los objetivos concretos que se corresponden con cada uno de los capítulos de la presente tesis se enumeran a continuación:

Capítulo 1.

- Evaluación de la eficacia replicativa de 25 linajes evolutivos independientes de TEV en el huésped *A. thaliana*.
- Análisis de la contribución de la selección, el azar y la contingencia histórica al resultado de la evolución.
- Exploración del efecto de la rugosidad sobre la topografía empírica del paisaje de eficacia.

Capítulo 2.

- Evaluación de la eficacia replicativa de 32 genotipos mutantes en el ambiente proporcionado por dos huéspedes distintos.
- Caracterización, comparación y relación de la situación espacial de los genotipos en ambos paisajes de eficacia.
- Exploración de los efectos de la pleiotropía sobre la topografía empírica del paisaje de eficacia de ambos huéspedes.

Objetivos

Capítulo 3.

- Evaluación de la eficacia replicativa de 164 subclones biológicos individuales aislados de la misma población clonal inicial.
- Exploración de la dinámica de adaptación dentro del huésped y análisis de la eficacia de los componentes subclonales en el tiempo.
- Explorar el efecto de la baja densidad de la población viral.

Capítulo 4.

- Análisis transcriptómico de plantas de tabaco infectadas con siete genotipos de TEV que difieren en su eficacia y severidad de síntomas inducidos.
- Relación entre eficacia y expresión diferencial de genes.
- Identificación y caracterización de genes candidatos cuya expresión correlaciona positiva o negativamente con la eficacia viral.

3. Material y Métodos

1. Generación de clones infecciosos de TEV.

En el transcurso de esta tesis se han empleado diversos conjuntos de clones infecciosos de TEV cuyo origen ha partido siempre del vector plasmídico pMTEV. Sin embargo, en función de las necesidades experimentales, se han iniciado los experimentos partiendo de una población viral ancestral, de una colección de clones generados por mutagénesis dirigida o de una colección de clones generados por evolución mediante pases seriados.

En esta sección detallaremos todos los procedimientos relacionados con el trabajo con el vector plasmídico pMTEV (**apartado 1.2**), así como los procedimientos por los cuales obtenemos los distintos grupos de clones infecciosos de TEV para la iniciación de los experimentos (**apartados 1.3, 1.4 y 1.5**).

1.1. Origen.

De los 37 genotipos de TEV utilizados en el transcurso de esta tesis doctoral, 31 genotipos provienen de las observaciones previas de las mutaciones fijadas en el primer experimento de evolución de TEV en *A. thaliana*, realizado por este mismo grupo de investigación y descrito en [Agudelo-Romero et al. \(2008a\)](#). En resumen, tras 17 pases seriados, el genoma viral consenso difería sólo en cinco posiciones del genoma ancestral (**Tabla 2**), tres de los cuales correspondían a mutaciones no sinónimas. Sobre dichas observaciones, se construyeron todos los genotipos iniciales por mutagénesis dirigida que describiremos en el **apartado 1.3.2** ([Lalić y Elena, 2015](#)) con los que iniciar los experimentos de evolución que se presentan en esta tesis.

Tabla 2. Mutaciones puntuales utilizadas para generar los genotipos incluidos en esta tesis doctoral. Cada genotipo se indica mediante un código binario en el que ○ indica el alelo de tipo silvestre y ● el alelo mutante. Detalles adicionales sobre estas mutaciones se pueden encontrar en [Agudelo-Romero et al., 2008a](#) y [Lalić y Elena, 2015](#).

| Etiqueta | Mutación | Gen | Cambio aminoácido* |
|----------|----------|----------------|--------------------|
| ●○○○○ | U 357 C | <i>P1</i> | sinónimo |
| ○●○○○ | C 3140 U | <i>P3</i> | A 999 V |
| ○○●○○ | C 3629 U | <i>6K1</i> | T 1162 M |
| ○○○●○ | C 6037 U | <i>VPg</i> | L 1965 F |
| ○○○○● | C 6906 U | <i>Nla-Pro</i> | sinónimo |

*La numeración se corresponde con el residuo del aminoácido en la poliproteína.

1.2. Vector plasmídico.

El genoma de TEV utilizado para generar todos los clones de virus, se aisló originalmente de plantas de *N. tabacum* ([Carrington et al., 1993](#)). En la presente tesis se ha utilizado el clon infeccioso pMTEV (número de acceso en GenBank DQ986288; [Bedoya y Daròs, 2010](#)), que contiene un cADN de longitud completa de TEV y una cola de poli-T de 44 nucleótidos de longitud seguida por una diana de restricción *Bgl*III ([Figura 18](#)). Este plásmido representa el esqueleto sobre el que se introdujeron diversas mutaciones puntuales en función de las necesidades experimentales anteriores llevadas a cabo por este mismo grupo de investigación y constituye el pilar básico para la obtención *in vitro* de los virus mutantes con los cuales se desarrollará el trabajo de esta tesis.

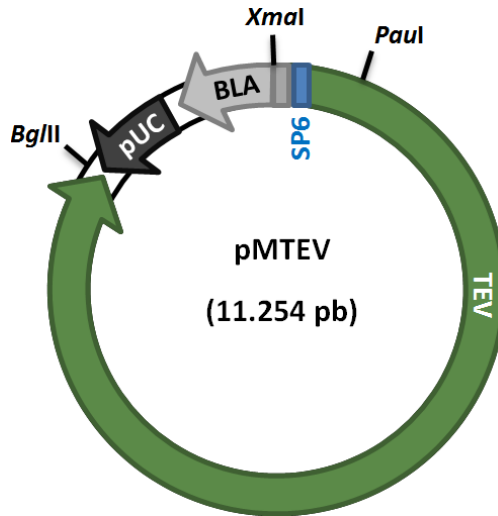


Figura 18. Vector de transcripción pMTEV con un promotor SP6 (caja azul), cADN de TEV (flecha verde), origen de replicación del plásmido pUC19 (flecha negra), marcador de selección de la β -lactamasa (BLA, flecha gris clara) con su promotor (caja gris oscura) y una diana de restricción *Bgl*III. En el esquema también se indican las posiciones de algunas otras dianas de restricción, particularmente interesantes. (Figura reproducida de [Bedoya y Daròs, 2010](#)).

1.3. Generación de los clones por mutagénesis dirigida.

La mutagénesis dirigida mediante PCR es una técnica de biología molecular utilizada para crear mutaciones puntuales en una cadena de ADN modificando los cebadores empleados. Nuestro objetivo será modificar o eliminar la funcionalidad de la proteína o factor codificados en la secuencia de ADN, que a la hora de realizar los análisis correspondientes, nos permitirá evaluar los efectos de dicha modificación o supresión en su correspondiente marco de estudio (por ejemplo, evaluar el efecto de la pérdida de función de una proteína de un virus vegetal en su huésped natural).

1.3.1. Mutagénesis dirigida.

Los 37 genotipos de TEV utilizados en el transcurso de esta tesis doctoral se construyeron por rondas sucesivas de mutagénesis dirigida mediante PCR a partir del plásmido pMTEV. Para construir los clones se utilizaron cebadores mutagénicos (Lalić y Elena, 2015) con cambios específicos de un solo nucleótido y la ADN polimerasa de alta fidelidad de copia Phusion (Thermo Scientific), siguiendo las instrucciones proporcionadas por el fabricante.

El perfil de PCR para realizar la mutagénesis consistió en una desnaturalización de 30 s a 98 °C, seguida de 30 ciclos de 10 s a 98 °C, 30 s a 60 °C y 3 min a 72 °C, terminando con una elongación de 10 min a 72 °C. A continuación, los productos de dicha PCR se incubaron con *DpnI* (Fermentas) durante 2 h a 37 °C para digerir el plásmido molde parental metilado. Se transformaron células electrocompetentes de *Escherichia coli* DH5 α con 1-2 μ l de estos productos de reacción y se plaquearon sobre agar Luria-Bertani (medio LB sólido) suplementado con 100 mg/ml de ampicilina. Las colonias bacterianas se inocularon en 8 ml de medio líquido LB-ampicilina y se hicieron crecer durante 16 h en un agitador orbital (37 °C, 225 rpm).

Las purificaciones de los plásmidos se llevaron a cabo utilizando *Wizard Plus SV Minipreps ADN Purification System* (Promega) y *NucleoBond Xtra Midi EF* (Macherey Nagel), siguiendo en ambos casos las instrucciones del fabricante. La elección de un sistema u otro depende únicamente del volumen de la muestra. Cuando éste era pequeño, consistente en 1-10 ml de cultivo de clones transformados, se empleaba el primer sistema. Consecuentemente, el volumen final eluido y su concentración serán pequeños. Cuando el volumen de la muestra era más grande, de un máximo de 200 ml de cultivo de clones transformados, se optaba por el segundo de los sistemas. En este caso el volumen final eluido y su concentración serán grandes.

Posteriormente, los transcritos se precipitaron (1,5 volúmenes de agua tratada con pirocarbonato de dietilo (DEPC), 1,5 volúmenes de LiCl 7,5 M, EDTA 50 mM) y se resuspendieron en agua tratada con DEPC (Carrasco et al., 2007a). La integridad y la cantidad de ARN se evaluaron mediante electroforesis en gel de agarosa. Se confirmó la incorporación de la mutación (o mutaciones si se repetía el proceso para la generación de mutantes múltiples) mediante la secuenciación de un fragmento de aproximadamente 800 pb que contenía el nucleótido mutagenizado.

Por último, el plásmido se linearizó con *Bgl*II y se transcribió *in vitro* usando el sistema *mMESSAGE mMACHINE SP6* (Ambion) con el fin de obtener ARN infeccioso de cada genotipo de virus (Carrasco et al., 2007b). Además, la secuencia consenso de cada transcripción se confirmó por secuenciación. Para ello, primero se realizó una retrotranscripción (RT) utilizando la transcriptasa inversa de M-MuLV (Thermo Scientific) y un cebador inverso fuera de la región de interés para la obtención de un fragmento de 800 pb que contuviera el sitio de mutación. La PCR se realizó con la ADN polimerasa de alta fidelidad de copia Phusion (Thermo Scientific) y parejas de cebadores apropiadas para cada producto transcrito.

Para los experimentos relativos a los **Capítulos 1 y 2**, en que empleamos un número grande de genotipos, la secuenciación se realizó en GenoScreen (Lille, Francia; www.genoscreen.com) con un equipo ABI 3730XL (Thermo Scientific). En el caso del experimento descrito en el **Capítulo 4**, en que el conjunto muestral se compone únicamente de seis genotipos mutantes, la secuenciación se realizó en el Servicio de secuenciación del IBMCP. Los moldes se marcaron con Big Dyes v3.1 y se resolvieron en un equipo ABI 3130XL (Thermo Scientific). La visualización del cromatograma y análisis posteriores se realizaron con GENEIOUS v. R9.1 (www.geneious.com).

1.3.2. Genotipos de virus.

Los cinco genotipos que figuran a continuación (**Tabla 3**) conforman el inicio de la evolución experimental y el punto de partida del primer experimento de esta tesis (ver **Capítulo 1**).

Tabla 3. Características de los cinco genotipos mutantes que se emplearon en la evolución experimental en el huésped alternativo *A. thaliana*. La notación binaria indica si en una posición determinada hallamos el alelo silvestre (○) o el alelo mutante (●).

| Etiqueta | Mutación | Gen | Cambio aminoácido* |
|----------|----------|----------------|--------------------|
| ○●○○○ | C 3140 U | <i>P3</i> | A 999 V |
| ○○○●○ | C 6037 U | <i>VPg</i> | L 1965 F |
| ○○○○● | C 6906 U | <i>Nla-Pro</i> | sinónimo |
| ●○○○○ | U 357 C | <i>P1</i> | sinónimo |
| | C 3629 U | <i>6K1</i> | T 1162 M |
| | C 6037 U | <i>VPg</i> | L 1965 F |
| ○●●○○ | C 3140 U | <i>P3</i> | A 999 V |
| | C 3629 U | <i>6K1</i> | T 1162 M |
| | C 6906 U | <i>Nla-Pro</i> | sinónimo |

*La numeración se corresponde con el residuo del aminoácido en la poliproteína.

Asimismo, de la combinación de las mutaciones simples presentes en la **Tabla 2**, se construyeron otros 26 genotipos mutantes de TEV por rondas sucesivas de mutagénesis dirigida a partir del plásmido pMTEV (**Tabla 4**). Dichos mutantes comprenden todas las combinaciones posibles de mutaciones dobles, triples, cuádruples y el mutante quíntuple. En total, este material infeccioso se compone de 32 genotipos (**Lalić y Elena, 2015**) que serán utilizados para definir los paisajes adaptativos en diferentes huéspedes en los **Capítulos 1 y 2**, constituyendo además el bloque central de la discusión del **Capítulo 2**.

Tabla 4. Listado completo de los genotipos generados por combinación de las cinco mutaciones fijadas en TEV adaptado a *A. thaliana*, con los cuales se llevaron a cabo inoculaciones en el huésped original *N. tabacum* y en el huésped alternativo *A. thaliana*. Las posibilidades en cada posición son o (alelo silvestre) y ● (alelo mutante). En el lado izquierdo, el genotipo silvestre de TEV queda representado como ooooo, mientras que en el lado derecho hallamos el aislado adaptado de *A. thaliana*, mutante quintuple con las cinco mutaciones, representado como ●●●●●.

| WT | Mutantes simples | Mutantes dobles | Mutantes triples | Mutantes cuádruples | Mutante quintuple |
|-------|------------------|-----------------|------------------|---------------------|-------------------|
| ooooo | ●oooo | ●●ooo | ●●●oo | ●●●●o | ●●●●● |
| | o●ooo | ●o●oo | ●●o●o | ●●●o● | |
| | oo●oo | ●oo●o | ●●oo● | ●●o●● | |
| | ooo●o | ●ooo● | ●o●●o | ●o●●● | |
| | oooo● | o●●oo | ●o●o● | o●●●● | |
| | | o●o●o | ●oo●● | | |
| | | o●oo● | o●●●o | | |
| | | oo●●o | o●●o● | | |
| | | oo●o● | o●o●● | | |
| | | oo●●● | oo●●● | | |

Los seis genotipos mutantes que conforman el punto de partida del experimento de análisis transcriptómico (ver **Capítulo 4**) también se construyeron por mutagénesis dirigida a partir de pMTEV, tal como se describe en [Torres Barceló et al., 2008](#) (para los genotipos **HC-Pro:AS13**, **HC-Pro:CLA2** y **HC-Pro:CLA11**) y [Carrasco et al., 2007a](#) (para los genotipos **CI:PC48**, **CI:PC55** y **Nib:PC95**) (**Tabla 5**). Entrando en detalle, **HC-Pro:AS13** es el único genotipo que posee dos mutaciones. Además, los genotipos **HC-Pro:AS13** y **HC-Pro:CLA2** se han descrito como asociados con un fenotipo hiposupresor del silenciamiento génico, mientras que **HC-Pro:CLA11** se ha descrito como hipersupresor. Los genotipos **Nib:PC95** y **CI:PC55** se han descrito como fenotipos de baja eficacia y **CI:PC48** como fenotipo de gran eficacia.

Tabla 5. Localización de las mutaciones y características de los mutantes que se emplearon en el experimento de análisis transcriptómico.

| Mutante | Mutación | Gen | Cambio aminoácido* |
|---------------------|----------------------|---------------|--------------------|
| HC-Pro:AS13 | A 1951 C A 1954 C | HC-Pro | E 603 A D 604 A |
| HC-Pro:CLA2 | U 1629 C | HC-Pro | V 496 A |
| HC-Pro:CLA11 | G 2382 A | HC-Pro | E 747 K |
| CI:PC48 | U 4101 A | CI | H 1319 Q |
| CI:PC55 | A 5115 U | CI | K 1657 N |
| Nib:PC95 | G 8398 U | Nib | A 2752 S |

*La numeración se corresponde con el residuo del aminoácido en la poliproteína.

1.4. Generación de los clones por evolución mediante pases seriados en *A. thaliana*.

Los experimentos de pases seriados que simulan la transmisión horizontal se caracterizan por producir aumentos en la virulencia del parásito debido a una mayor competencia intra-huésped entre las cepas patógenas, la disociación entre la tasa de crecimiento intra-huésped, la tasa de transmisión y la falta de innovación evolutiva en el huésped (Ebert, 1998).

1.4.1. Evolución mediante pases seriados.

Se utilizaron plantas de tabaco (*N. tabacum* L. var. Xanthi NN) para la producción de un *stock* de partículas de virus de cada uno de los cinco genotipos considerados en el primer experimento de esta tesis. Se inocularon lotes de plantas de tabaco de ocho semanas de edad con 5 mg de ARN de cada genotipo viral por abrasión de la

tercera hoja verdadera con Carborundum. Diez días después de la inoculación (dpi), todas las plantas infectadas se recogieron y se agruparon para cada genotipo del virus. El tejido vegetal se congeló con nitrógeno líquido, se homogeneizó usando mortero y mazo y se dividió en alícuotas en tubos de 1.5 ml. Para preparar el suero de inoculación, se añadió 1 ml de tampón fosfato de potasio 50 mM (pH 8.0) por gramo de tejido vegetal homogeneizado, se centrifugaron a 4 °C y 10.000 g durante 10 minutos, se recogió la fase líquida superior y se mezcló con Carborundum al 10%.

Para iniciar el experimento de evolución, fueron inoculadas mecánicamente entre 15 y 25 plantas de *A. thaliana* L. ecotipo *Ler-0*, en la etapa de crecimiento 3,5 según la escala de Boyes (Boyes et al., 2001), con las preparaciones de virus anteriores. Las plantas se mantuvieron en un invernadero de nivel de bioseguridad 2 a 25°C y un período de luz de 16 h. El estado de infección se determinó por RT-PCR en un solo paso a partir de una muestra tomada a los 14 dpi, como se describe en Lalić et al. (2010). En este punto, se seleccionaron cinco plantas infectadas con cada genotipo para iniciar cinco linajes de evolución independientes. Ello suma un total de 25 linajes. Las plantas enteras infectadas se recogieron a 21 dpi, que corresponde al tiempo de transmisión utilizado en experimentos anteriores de este mismo grupo de investigación (Agudelo-Romero et al., 2008a). En total, se homogeneizaron 200 mg de tejido en nitrógeno líquido usando un mortero y un mazo, se mezclaron con 1 ml de tampón de inoculación (fosfato de potasio 50 mM, pH 8.0 y Carborundum al 10%) y se usaron para inocular el siguiente conjunto de plantas *A. thaliana* *Ler-0*. Este procedimiento se repitió cuatro veces más, sumando un total de cinco pases en serie.

1.5. Generación de los clones por preparación de una población clonal inicial de TEV.

Se generó una población clonal de TEV inoculando un conjunto de plantas de tabaco con ARN infeccioso del aislado silvestre generado por transcripción *in vitro*. A los 8 dpi, después de que aparecieran los primeros síntomas, se recogieron hojas sintomáticas y ápices de veinticinco plantas en bolsas de plástico (después de retirar la hoja inoculada). El tejido entero se mezcló, se congeló en nitrógeno líquido, se molió con mortero y se almacenó a -80°C. Este material vegetal aglomerado se compone del genotipo silvestre de nuestro virus y un gran número de mutantes biológicos individuales denominados subclones. Es importante mencionar que algunos subclones pueden corresponderse realmente al mismo genotipo, ya que la probabilidad de volver a muestrear el mismo genotipo depende en última instancia de su frecuencia poblacional. El objetivo último de poseer dicha población clonal es poder aislar posteriormente componentes subclonales mediante un ensayo de formación de lesiones en hojas de *Chenopodium quinoa Willd* en un sistema *in vivo*, que es equivalente al ensayo de cultivos celulares para virus animales. Este material infeccioso fue utilizado como punto de partida en el tercer experimento de esta tesis (ver **Capítulo 3**).

2. Hospedadores. Condiciones de crecimiento e inoculación.

Durante el transcurso de esta tesis se emplearon como huéspedes de TEV tres especies de plantas distintas en función de los supuestos iniciales y los objetivos de cada uno de los experimentos. Dichos experimentos se realizaron con el huésped natural de TEV, *N. tabacum*, y dos huéspedes alternativos, *C. quinoa* y *A. thaliana*.

Las plantas de tabaco y quenopodio se mantuvieron en un invernadero de nivel de bioseguridad 2, mientras que las plantas de *Arabidopsis* se mantuvieron en un fitotrón a 25° C y con un período de luz de 16 h.

2.1. *Nicotiana tabacum* L. var. Xanthi.

El protocolo de inoculación de TEV en plantas de tabaco puede partir de muestras de tejido vegetal infectado o transcritos de ARN.

Se utilizaron plantas de *N. tabacum* para la producción de grandes *stocks* de viriones para cada uno de los genotipos considerados en esta tesis. Los transcritos de ARN se mezclaron con un volumen 1:10 de tampón de inoculación (compuesto por fosfato de potasio 50 mM y Carborundum en una proporción de 100 mg/ml). Se inocularon lotes de plantas de *N. tabacum* de 8 semanas de edad (**Figura 19**) con aproximadamente 5 µg de ARN de cada genotipo viral por abrasión de la tercera hoja verdadera (**Figura 20**).



Figura 19. Detalle de plantas de *N. tabacum* de 8 semanas de edad.

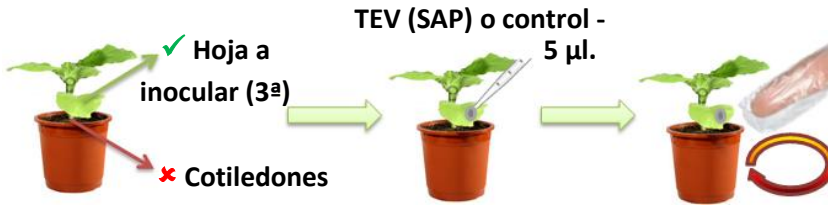


Figura 20. Inoculación del virus del grabado del tabaco (TEV) en plantas de tabaco.

La inoculación de todos los genotipos relativos al mismo experimento se realizó en un único bloque experimental, a excepción de los experimentos de aislamiento de componentes subclonales (ver **Capítulo 3**), en que las inoculaciones se dividieron en cuatro bloques (como se detallará más adelante en el **apartado 2.3**) y de análisis transcriptómico (ver **Capítulo 4**), en que las inoculaciones se realizaron en dos bloques experimentales, el primero de los cuales incluía los genotipos **HC-Pro:AS13**, **HC-Pro:CLA2**, **HC-Pro:CLA11**, **Nib:PC95** y sus controles y el segundo incluía **Cl:PC48**, **Cl:PC55** y sus controles. En cualquier caso, todas las plantas se encontraban en etapas de crecimiento similares. Posteriormente, las plantas se mantuvieron en un invernadero de bioseguridad 2. Todas las plantas de tabaco infectadas mostraron síntomas (**Figura 21**) al cabo 5-8 dpi (salvo genotipos letales o asintomáticos como **HC-Pro:AS13**). A los 8 dpi, las hojas y los ápices de cada planta se recogieron individualmente en bolsas de plástico (después de retirar la hoja inoculada), con la excepción de las plantas infectadas con **HC-Pro:AS13** que se recogieron a los 15 dpi. A continuación, el tejido vegetal se congeló con nitrógeno líquido, se homogenizó usando un molino mezclador MM 400 (Retsch) y se dividió en alícuotas en tubos de 1.5 ml (conteniendo 100 mg de tejido cada uno). Estas alícuotas de tejido infectado con TEV se almacenaron a -80°C .

Los sueros para inoculación se prepararon añadiendo a las preparaciones anteriores 1 ml de tampón de fosfato de potasio 50 mM (pH 8,0) por g de tejido vegetal homogeneizado. Luego se

centrifugaron a 4°C y 10.000 g durante 10 minutos. Se pipeteó la fase líquida superior y se mezcló con Carborundum al 10%. Estos sueros de inoculación contienen las poblaciones virales que se utilizaron para inocular de forma mecánica otras plantas de *N. tabacum*, entre 3 y 31 plantas (12 de media, según la eficacia del genotipo) de *A. thaliana* o plantas de *C. quinoa*.



Figura 21. Detalle de una planta de tabaco sana (izquierda) y una planta de tabaco infectada con el virus del grabado del tabaco (TEV).

Los genotipos letales y los experimentos fallidos producen el mismo resultado, la falta de infección. Para hacer frente a esta posible fuente de error, procedimos como se describe en [Lalić et al. \(2011\)](#) y [Carrasco et al. \(2007a\)](#). En resumen, se estimó la probabilidad de fallar un experimento de inoculación utilizando transcritos de ARN de pMTEV de tipo silvestre y un gran número de plantas. Luego, utilizando esta probabilidad, se aplicó la distribución de probabilidad de Bernoulli para evaluar la probabilidad de fallar todos los intentos de inoculación después de un número determinado de ensayos. En todos los casos, esta probabilidad fue menor del 0,01.

2.2. *Arabidopsis thaliana* L. ecotipo Ler-0.

El protocolo de inoculación de TEV en plantas de *A. thaliana* ha partido siempre de muestras de tejido vegetal infectado. Así, los sueros de inoculación se prepararon a partir de los tejidos infectados de tabaco, si se trató de iniciar un experimento de evolución, o bien a partir de los tejidos de plantas de *A. thaliana* previamente infectadas, si se trató de continuar un experimento de evolución mediante pases seriados. En cualquier caso, la preparación de los sueros, la metodología del experimento de evolución, el número y condiciones de las plantas se detallan en el **apartado 1.4.1**.

Únicamente cabe añadir cómo se seleccionaron las hojas que se van a inocular y como se practicó la inoculación. Para *N. tabacum* seleccionábamos la tercera hoja verdadera (sin contar los cotiledones) mientras que para *A. thaliana* seleccionamos las tres hojas más grandes de la roseta (**Figura 22**). Sobre la superficie de las hojas seleccionadas, depositamos 5 µl de la muestra e hicimos un poco de fricción con una varilla de cristal (tratando únicamente de realizar pequeñas heridas por abrasión) (**Figura 23**). El genotipo silvestre de TEV suele ser asintomático en *A. thaliana*, es decir, la planta se infecta de forma sistémica pero no suele observarse sintomatología clara (aunque en función del virus o la presencia de mutaciones, pudieran observarse síntomas a partir de los 7 dpi). El estado de infección se determinó siempre por RT-PCR en un solo paso a partir de una muestra tomada a los 14 dpi, consistente en dos o tres hojas de la parte aérea de la planta, y se recogió el tejido entero de las plantas infectadas a los 21 dpi. Este tejido se procesó para inocular el siguiente lote de plantas (ya fueran otras plantas de *A. thaliana* o inocular el huésped original *N. tabacum*) o bien se extrajo para procedimientos analíticos posteriores.



Figura 22. Detalle de la roseta de *A. thaliana* en la etapa de crecimiento 3,5 según la escala de Boyes.

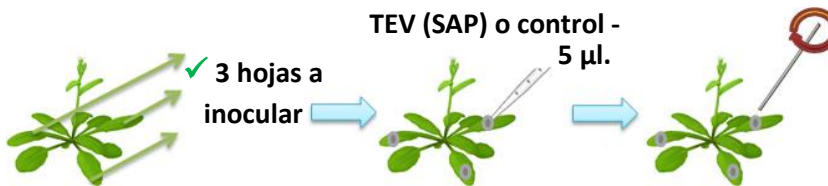


Figura 23. Inoculación del virus del grabado del tabaco (TEV) en plantas de *A. thaliana*.

2.3. *Chenopodium quinoa* Willd.

En el protocolo de inoculación de TEV en plantas de quenopodio también hemos partido de tejido vegetal previamente infectado, sólo que en este caso dicho tejido siempre provino de plantas de tabaco. Este material infeccioso constituye una población clonal con la que realizaremos un ensayo de formación de lesiones en hojas de quenopodios. Nuestro objetivo es aislar los subclones biológicos que forman parte de la población clonal, infectar plantas de tabaco con ellos y realizar los análisis pertinentes. La **Figura 24** muestra una representación esquemática del protocolo experimental seguido en este experimento.

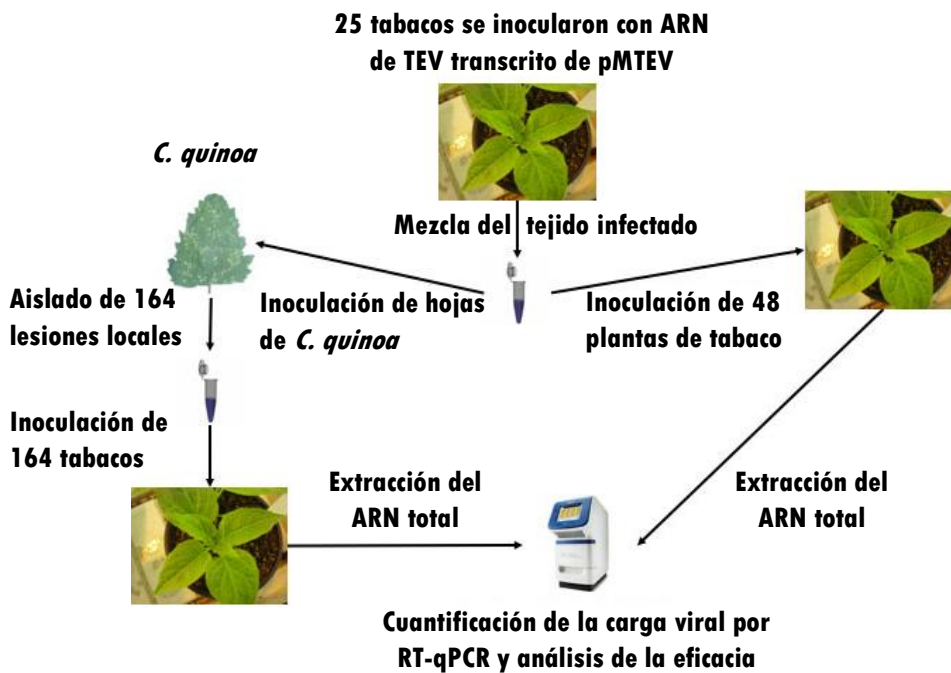


Figura 24. Esquema del protocolo experimental seguido para aislar componentes subclonales y muestras aleatorias de la población clonal.

En resumen, se procedió a aislar los componentes subclonales de TEV mediante el método de ensayo de dilución-inoculación en el huésped *C. quinoa* (Kleczkowski, 1950; De la Iglesia y Elena, 2007). Para empezar, se trituraron 100 mg de tejido de tabaco infectado con TEV en un mortero añadiendo 1 ml de tampón fosfato de potasio. A continuación, se inocularon por abrasión nueve hojas completamente desarrolladas de cada una de las plantas de *C. quinoa* (Figura 25). Cabe destacar que para este experimento dispusimos de 4 plantas de *C. quinoa* de 4 semanas de edad por cada bloque experimental. La inoculación se realizó añadiendo Carborundum, a la concentración descrita previamente, a una preparación de tejido

infectado sin diluir, una preparación diluida 1/10 y otra diluida 1/100 (Figura 26). Se inocularon tres hojas de cada dilución para minimizar los efectos en las plantas (Kleczkowski, 1950). Además, se inocularon dos hojas adicionales únicamente con tampón de inoculación que servirían como controles negativos. A los 9 dpi, se recogieron individualmente las lesiones locales que se observaron claramente aisladas, las cuales se congelaron inmediatamente con nitrógeno líquido, se molieron en tubos de 1.5 ml con un pequeño mazo de plástico y se mantuvieron congeladas hasta el momento de inocular las plantas de tabaco (Figura 27). Dichas plantas se inocularon inmediatamente tras aislar el número convenido de lesiones. Para ello, se añadieron 20 μ l de tampón de inoculación a cada tubo y se inocularon en tabacos, como se ha explicado anteriormente. Paralelamente, se procesaron de la misma manera varias alícuotas de la población clonal inicial de tejido infectado de tabaco y se usaron para inocular otras plantas de tabaco. El experimento de inoculación completo se dividió en cuatro bloques, cada bloque conteniendo una serie de inoculaciones con subclones y un número de inoculaciones con muestras de la población clonal. Todas las plantas de tabaco infectadas mostraron síntomas claros a los 4-6 dpi. La carga viral de las plantas infectadas se evaluó después con 5 (cincuenta y dos plantas infectadas con subclones y diecisiete infectadas con la población clonal inicial), 7 (cinco y tres plantas), 9 (siete y tres plantas) y 12 (cien y veinticinco plantas) dpi.



Figura 25. Inoculación del virus del grabado del tabaco (TEV) en hojas de *C. quinoa*.



Figura 26. De izquierda a derecha, hojas de quenopodio inoculadas con una preparación de tejido infectado sin diluir, una preparación diluida 1/10 y otra diluida 1/100.



Figura 27. Esquema del aislamiento de lesiones y preparación de las mismas para su inoculación inmediata en plantas de tabaco.

3. Purificación, cuantificación de la carga viral y secuenciación del ARN genómico del virus.

Para los experimentos detallados en los **Capítulos 1 y 3**, partimos de 100 mg de tejido de cada planta procesada y la extracción de ARN se realizó utilizando el sistema *InviTrap Spin Plant*

ARN Mini Kit (Stratec Molecular) siguiendo las instrucciones del fabricante. La razón de emplear este sistema es que provee dos soluciones de lisis diferentes que, en función de la especie vegetal y el tejido del que se pretende extraer ARN total, pueden incrementar el rendimiento de la extracción. Es importante considerar este punto ya que en ambos experimentos utilizamos diversos huéspedes. La concentración de ARN total se ajustó a 50 ng/μl para cada muestra y la cuantificación de la carga viral se midió mediante RT-qPCR absoluta usando curvas estándar (Lalić et al., 2010).

Las curvas estándar se construyeron usando el ARN de TEV ancestral transcrito del plásmido pMTEV (como se describió en el **apartado 1.3.1**). Para ello, dicho ARN de TEV fue diluido en preparaciones de ARN total de plantas sanas, previamente ajustadas también a 50 ng/μl. Estas plantas recibieron el mismo tratamiento que las demás plantas en el experimento (se inocularon mecánicamente con tampón de inoculación). En total, se realizaron diez diluciones en serie para la construcción de las curvas estándar. Las reacciones de RT-qPCR se realizaron en un volumen final de 20 μl utilizando el sistema *One Step SYBR Prime Script RT-PCR II Perfect Real Time* (Takara) y siguiendo las instrucciones del fabricante. Los cebadores directo (q-TEV-F; 5'-TTGGTCTTGATGGCAACGTG-3') e inverso (q-TEV-R; 5'-TGTGCCGTTCAAGTGTCTTCCT-3') fueron diseñados para amplificar un fragmento de 71 nucleótidos en el extremo 3' del genoma de TEV con objeto de cuantificar únicamente genomas completos, evitando cuantificar genomas incompletos (Lalić et al., 2011). Las amplificaciones se realizaron en placas de noventa y seis pocillos, conteniendo cada una las muestras objeto de estudio, sus respectivos controles y su curva estándar. Se hicieron dos réplicas técnicas de cada planta para el experimento relativo al **Capítulo 1** y tres réplicas técnicas para el experimento detallado en el **Capítulo 3**. Las amplificaciones se realizaron usando un equipo ABI StepOne Plus (Applied Biosystems), de acuerdo con el siguiente perfil: la fase RT consistió en 5 min a 42 °C y 10 s a 95 °C, seguido de la fase de PCR

Material y Métodos

consistente en 40 ciclos de 5 s a 95 °C y 34 s a 60 °C, terminando con la fase final con 15 s a 95 °C, 1 min a 60 °C y 15 s a 95 °C.

Para los experimentos detallados en los **Capítulos 2 y 4**, la metodología es la misma salvo que empleamos otros productos comerciales. Partimos también de 100 mg de tejido pero la extracción de ARN se llevó a cabo con el sistema *Agilent Plant ARN Isolation Mini Kit* (Agilent Technologies) siguiendo las instrucciones del fabricante. La razón de emplear este sistema radica en que el ARN purificado con el mismo es ideal para análisis de expresión génica mediante *microarrays*, que se llevará a cabo posteriormente con las muestras del experimento de análisis transcriptómico (**Capítulo 4**). Dicho sistema permite obtener el ARN vegetal total con un alto valor de RIN (*ARN Integrity Number*), que determina una gran calidad del ARN purificado. Además, para cualquier muestra purificada de ambos experimentos, en que únicamente se emplearon plantas de tabaco, este sistema permite deshacerse de la práctica totalidad del ADN genómico.

Prosiguiendo con la metodología, el ajuste de la concentración de las muestras y la construcción de los estándares se realizó como se detalla en el párrafo anterior. Para estos dos experimentos, la cuantificación de la carga viral se realizó en un volumen final de 20 µl utilizando el sistema *GoTaq 1-Step RT-qPCR System* (Promega) siguiendo las instrucciones del fabricante. Los cebadores empleados fueron los mismos que en el caso anterior. Se hicieron tres réplicas técnicas de cada planta para ambos experimentos. Las amplificaciones se realizaron usando de nuevo el equipo ABI StepOne Plus (Applied Biosystems), de acuerdo con este otro perfil: la fase RT consistió en 15 min a 37°C y 10 min a 95 °C, seguido de la fase de PCR consistente en 40 ciclos de 10 s a 95 °C, 34 s a 60 °C y 30 s a 72 °C, terminando con la fase final con 15 s a 95 °C, 1 min a 60 °C y 15 s a 95 °C.

Los resultados de todas las cuantificaciones se examinaron utilizando el software StepOne v. 2.2.2 (Applied Biosystems). Además, cada muestra de ARN empleada en la cuantificación se secuenció de nuevo en esta etapa de la misma manera que se detalla en el **apartado 1.3.1.** (a excepción de las muestras provenientes de las lesiones locales del experimento de componentes subclonales).

4. Evaluación de la eficacia y análisis estadísticos.

Para todos los genotipos y subclones biológicos considerados en esta tesis, se extrajo el ARN total y se cuantificó la acumulación de genomas virales por RT-qPCR como describimos en el apartado anterior. En todos los casos, se cuantificó la acumulación de virus como pg de ARN de TEV por cada 100 ng de ARN vegetal total a unos determinados dpi. Es muy importante la determinación del parámetro t , días de crecimiento de la planta tras el tratamiento (dpi), ya que hemos de asegurarnos de que las poblaciones virales estudiadas se hallen en la fase de crecimiento exponencial, que depende del huésped considerado (ya que TEV alcanza una meseta cuasi-estacionaria más rápidamente en *N. tabacum* que en *A. thaliana*) y los genotipos mutantes utilizados. Estos valores de cuantificación se utilizaron entonces para calcular la eficacia de los genotipos objeto de estudio en relación con la de un genotipo de referencia. Cabe destacar que los parámetros empleados en el cálculo de la eficacia relativa y la modelización estadística difieren en función de las necesidades analíticas de cada uno de los experimentos de esta tesis.

4.1. Eficacia relativa.

Para el caso de los experimentos de los **Capítulos 2 y 4**, que involucran cuantificaciones de ARN de TEV en plantas de tabaco, los valores de cuantificación se utilizaron para calcular la eficacia de los genotipos mutantes de TEV en relación a la del genotipo salvaje. Para ello, se utilizó la expresión $W = \sqrt[t]{R_t / R_0}$, donde R_0 y R_t representan la acumulación estimada para el mutante o virus de referencia en el momento de la inoculación (R_0) y después de t días de crecimiento tras la inoculación (R_t) (Carrasco et al., 2007a). Para el caso de los experimentos del **Capítulo 1** y, de nuevo, el **Capítulo 2**, que involucran cuantificaciones de ARN de TEV en plantas de *A. thaliana*, la eficacia de los genotipos mutantes se calculó en relación a la del genotipo ○●○○●. Se tomó dicho genotipo como referencia porque, en la construcción del paisaje de eficacia, constituye el pico de eficacia local óptima. Recordemos que hay que prestar atención a la determinación del parámetro t . Así, en el **Capítulo 1** corresponde a $t = 15$ para los cinco genotipos mutantes evolucionados en *A. thaliana*. En el **Capítulo 2**, corresponde a $t = 21$ dpi para los genotipos inoculados en *A. thaliana* y $t = 5$ dpi para aquellos inoculados en *N. tabacum* (tiempos en que ambas poblaciones virales se hallaban en la fase de crecimiento exponencial en su respectivo huésped). Por último, en el **Capítulo 4**, en el que únicamente inoculamos plantas de tabaco, corresponde a $t = 8$ dpi para todos los genotipos con la excepción de **HC-Pro:AS13**, que se cuantificó a $t = 15$ dpi (para asegurar que este genotipo también se hallaba en la fase de crecimiento exponencial).

Por el contrario, el cálculo de la eficacia difiere del resto en el **Capítulo 3**, en el que nuestro objetivo era aislar los subclones biológicos. En este caso, la cuantificación de las cargas virales a determinados dpi (V_t), obtenidas por RT-qPCR, se transformaron en parámetros de crecimiento malthusianos usando la expresión $m = \frac{1}{t} \ln V_t$. Puesto que estamos interesados en evaluar el rendimiento de los componentes subclonales en relación con

muestras aleatorias de la población clonal, la eficacia relativa de un subclon i en la muestra tomada en t dpi se calculó como $W_{it} = e^{m_{it} - m_t^{pop}}$, donde m_t^{pop} constituye el parámetro de crecimiento malthusiano promedio estimado para las muestras de la población tomadas en t dpi.

4.2. Análisis estadísticos.

Como avanzamos anteriormente, cada experimento parte de unos supuestos iniciales y objetivos únicos. Así, la modelización estadística es diferente en cada uno de los experimentos reflejados en esta tesis y procederemos a detallarlos por separado.

Para el experimento del **Capítulo 1**, la magnitud del efecto en los modelos ANOVA se evaluó utilizando el estadístico η_p^2 , que representa la proporción de la variabilidad total atribuible a un determinado factor mientras controlamos el resto de factores. Cuando dichos valores corresponden a $\eta_p^2 < 0,05$ se considera un efecto pequeño, $0,05 \leq \eta_p^2 < 0,15$ efecto medio y $\eta_p^2 \geq 0,15$ gran efecto. Cuando realizamos múltiples ensayos de la misma hipótesis nula, como cuando comparamos la eficacia de los linajes evolucionados independientemente con la eficacia del genotipo inicial o la eficacia del pico adaptativo más cercano, ajustamos los niveles de significancia usando la corrección Holm-Bonferroni. En estos casos, como veremos más adelante en el apartado de resultados, sólo se proporcionarán los valores de significatividad estadística, P , más significativos (los valores por encima de este valor de corte corregido no se consideraron significativos). Asimismo, el modelo lineal generalizado (GLM) incorpora dos factores aleatorios. El genotipo del virus (G) y el linaje (L , que representa la réplica biológica, es decir, diferentes plantas infectadas con el mismo genotipo viral G). El factor L se anidó dentro del factor G como sigue:

$$W_{ijkl} = \mu + G_i + L(G)_{ij} + \xi_{ijkl}$$

Para el **Capítulo 2**, también se realizó un GLM. Dicho modelo incorpora tres factores aleatorios. El primero, la especie huésped (H , que pudiera corresponderse con *A. thaliana* o con *N. tabacum*). El segundo, el genotipo del virus (G , que para este experimento son 32 en total). Y el último factor, la planta (P , que representa la réplica biológica, es decir, diferentes plantas de la especie huésped H fueron infectadas con el genotipo viral G). H y G se consideraron factores ortogonales, mientras que P se anidó dentro de la interacción de H y G . La ecuación del modelo figura a continuación:

$$W_{ijkl} = \mu + G_i + H_j + (G \times H)_{ij} + P(G \times H)_{ijk} + \xi_{ijkl}$$

Para el **Capítulo 3** es importante destacar que las diferencias entre los subclones son biológicamente significativas, ya que cada uno se corresponde potencialmente con una secuencia genómica diferente (aunque existe la probabilidad de que algunos subclones pueden compartir la misma secuencia en la población clonal), mientras que las diferencias entre grandes muestras tomadas de toda la población clonal son en su mayoría estadísticamente significativas. Antes de progresar con los análisis, debimos profundizar los datos relativos a la eficacia para averiguar si cumplían con los supuestos de normalidad y homocedasticidad de las varianzas. Los datos no poseían una distribución Normal (prueba de Kolmogorov-Smirnov de una muestra; $D = 0,201$ y $P < 0,001$) ni las varianzas eran homogéneas entre los grupos (prueba de Levene, $F_{139,496} = 21,520$ y $P = 0,001$). Dadas estas circunstancias, se optó por un GLM que incorporaba tres factores aleatorios. El primer factor, el origen del inóculo para infectar la planta de tabaco (S , es decir, un subclón o una muestra aleatoria de la población). El segundo factor, la réplica biológica (R , es decir, el subclón individual o la muestra de la población) que se anidó dentro de S . Y el tercer

factor, dpi (T , que tratamos como una covariable y en la interacción de $S \times T$). La ecuación de dicho modelo sería la siguiente:

$$W_{ijkl} = \mu + S_i + R(S)_{ij} + T_k + (S \times T)_{ij} + \xi_{ijkl}$$

En último lugar, para el **Capítulo 4**, el modelo lineal generalizado (GLM) continúa siendo similar a los anteriores e incorpora dos factores aleatorios. El primero, el genotipo del virus (G) y el segundo, la planta (P , que representa la réplica biológica, es decir, diferentes plantas infectadas con el mismo genotipo viral G). El factor P se anidó dentro del factor G . La ecuación del modelo se expresaría como sigue:

$$W_{ijkl} = \mu + G_i + P(G)_{ij} + \xi_{ijkl}$$

Para las ecuaciones de los **Capítulos 2, 3 y 4**, μ es el valor de la media y ξ_{ijkl} es el error asociado a las mediciones individuales l (estimado a partir de las repeticiones técnicas de la RT-qPCR). Para el modelo del **Capítulo 2** se asumió una distribución Normal y una función de enlace de identidad basándose en el criterio de información Bayesiano. En todos ellos, la significancia estadística de cada factor fue evaluada mediante una prueba de razón de verosimilitudes que se distribuye según una función de probabilidad χ^2 . Por último, la magnitud de los diferentes factores se evaluó utilizando el estadístico η_p^2 . La varianza total entre los diferentes factores se determinó por máxima verosimilitud.

Todos los análisis estadísticos se realizaron con el software IBM SPSS, versión 23.

5. Paisajes de eficacia.

En biología evolutiva se emplean los paisajes de eficacia para visualizar la relación entre distintos genotipos y su éxito reproductivo. En este caso particular, en que un potyvirus constituye nuestro objeto de estudio, cada genotipo tiene una tasa de replicación bien definida (lo que llamamos eficacia). Esta eficacia queda representada en el paisaje con una determinada altura. Asimismo, los genotipos similares los encontraremos cercanos en dicho paisaje, mientras que los que son muy diferentes se encontrarán lejos uno del otro. Al conjunto de todos los genotipos, su grado de similitud y sus valores de eficacia se le llama paisaje de eficacia.

5.1. Representación de los paisajes de eficacia.

La manera en que hemos elegido representar los paisajes de eficacia es confeccionando un gráfico en que cada nodo corresponde a un genotipo específico. Para simplificar el gráfico, continuaremos utilizando la notación empleada en la **Tabla 2**, en la que en lugar de representar los distintos genotipos referenciando los nucleótidos o aminoácidos, únicamente hemos indicado si en una posición determinada hallamos el alelo silvestre o el alelo mutante, es decir, las posibilidades en cada posición son o (alelo silvestre) y ● (alelo mutante). Con esta notación binaria, el genotipo silvestre de TEV queda representado como ooooo, mientras que el aislado adaptado de *A. thaliana*, mutante quintuple con las cinco mutaciones de la **Tabla 2**, queda representado como ●●●●●. Las líneas rectas de color en el gráfico binario representan pasos mutacionales de tamaño uno, es decir, dichas líneas conectan genotipos que difieren en un solo alelo.

5.2. Evaluación de la robustez del paisaje de eficacia.

La robustez de los paisajes de los dos huéspedes estudiados se evaluó utilizando tres enfoques diferentes. El primero es la media entre pendiente-rugosidad, θ (Aita et al., 2001). El parámetro θ mide el grado en que la pendiente de un determinado pico sobresale de la superficie promedio en la que se encuentra. Un valor de $\theta \gg 1$ significa que dicho pico emerge de una superficie plana y un valor de $\theta \leq 1$ indica que la pendiente del pico no difiere sustancialmente de la superficie de fondo, es decir, se halla rodeada de muchos picos pequeños con pendientes similares (Aita et al., 2001). El segundo es la correlación entre los niveles de eficacia de los genotipos, ρ (Ferretti et al., 2016). El parámetro ρ mide la similitud entre los niveles de eficacia de los genotipos que ocupan posiciones cercanas en el paisaje. Si esta correlación es perfecta ($\rho \approx 1$), entonces el paisaje es suave. Como los efectos de epistasia son cada vez más frecuentes a lo largo de la evolución experimental, la correlación medida por ρ se reduce y su valor tiende a $\rho < 0$ en el caso de epistasia de signo o reciprocidad (Figura 28) (Ferretti et al., 2016). El tercero únicamente tiene en cuenta las frecuencias de los diferentes tipos de interacciones epistáticas.

Valorando los efectos de las mutaciones sobre la topografía del paisaje, podemos definir cuatro tipos diferentes de interacciones epistáticas (Figura 28). En el caso de que no hubiera epistasia, la eficacia de los mutantes múltiples corresponde a la multiplicación de la eficacia de cada mutación. En el ejemplo, la eficacia del doble mutante (●●) es la multiplicación de los efectos de la eficacia de ambas mutaciones en el fondo genético de tipo silvestre, es decir, la multiplicación de los niveles de eficacia para los genotipos ●○ y ○● (Figura 28.A). Si el efecto de las mutaciones consideradas es diferente de la expectativa multiplicativa, hablamos de una interacción epistática de magnitud positiva (efecto mayor al esperado en el mutante múltiple; epistasia positiva) o de magnitud negativa (efecto menor; epistasia negativa). Continuando con el ejemplo, la eficacia

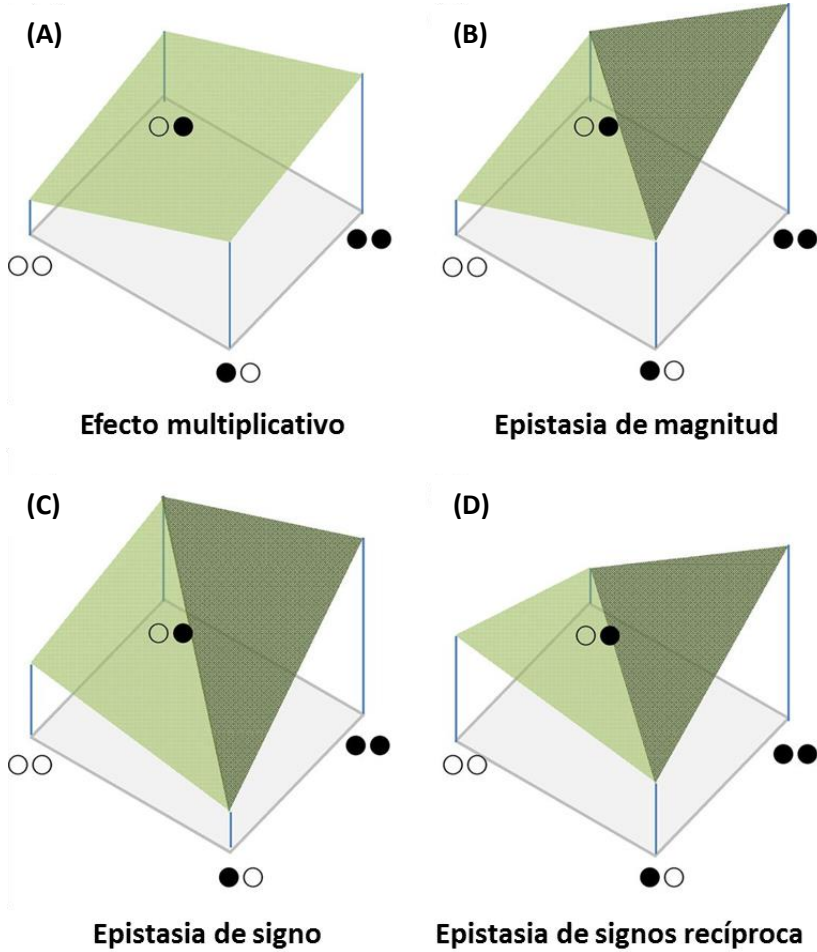


Figura 28. Diferentes tipos de epistasia entre dos genes que definen la eficacia de un genotipo. El alelo silvestre se representa con o y el alelo mutante con •.

del doble mutante (••) es mayor de lo esperado (epistasia positiva) (Figura 28.B). En los casos en que no hay epistasia u observamos epistasia de magnitud, los efectos de las mutaciones simples son individual e incondicionalmente beneficiosos. Sin embargo, el efecto de una de las mutaciones puede ser beneficioso en un fondo genético pero deletéreo en otro, lo que conocemos como interacción

epistática de signo (**Figura 28.C**). El último caso a valorar es si ambas mutaciones son deletéreas por sí mismas pero beneficiosas cuando se combinan, en cuyo caso hablamos de epistasia de signos recíproca (**Figura 28.D**). Como comentábamos anteriormente, la frecuencia de los efectos de epistasia aumenta a lo largo de la evolución experimental, lo que conlleva la visualización de un paisaje adaptativo más accidentado.

La representación gráfica de los paisajes adaptativos y la estimación de los parámetros definidos en este apartado se obtuvieron utilizando el servidor web MAGELLAN ([Brouillet et al., 2016](#)).

5.3. Cálculo de la epistasia.

La magnitud de la epistasia entre las mutaciones i y j se calculó de la siguiente manera:

$$\varepsilon_{ij} = W_{00}W_{ij} - W_{i0}W_{0j}$$

W_{i0} y W_{0j} son los niveles de eficacia relativa de los genotipos mutantes simples de cada mutación, W_{ij} es la eficacia relativa del doble mutante y W_{00} es la eficacia del genotipo de referencia; el genotipo silvestre en el caso de TEV inoculado en *N. tabacum* y el genotipo $\circ\bullet\circ\bullet$ en *A. thaliana* ([Kouyos et al., 2007](#); [Gao et al., 2010](#)). El segundo término de la ecuación ($W_{i0}W_{0j}$) corresponde a la eficacia esperada que, bajo la hipótesis de efectos independientes multiplicativos (no hay epistasia), es igual a la eficacia observada, dando como resultado un valor $\varepsilon_{ij} = 0$. Las desviaciones de la hipótesis nula indican mutaciones antagonistas (epistasia de magnitud negativa, $\varepsilon_{ij} > 0$) o sinérgicas (epistasia de magnitud positiva, $\varepsilon_{ij} < 0$).

Material y Métodos

Para aquellos genotipos que contienen más de dos mutaciones, se puede usar una ecuación muy similar:

$$\varepsilon_{i(k)} = W_{00}W_{i(k)} - W_iW_{(k)}$$

En este caso, $W_{i(k)}$ corresponde a la eficacia del genotipo que contiene k mutaciones más la mutación i . Así, $\varepsilon_{i(k)}$ es el valor de epistasia entre la mutación i y el fondo genético que contiene las k mutaciones. Por ejemplo, el genotipo ●●●● podría obtenerse de tres maneras: mediante la inserción de la mutación ○○○● en el fondo genético ●○○○, la mutación ○○●○○ en el fondo genético ●○○●, o la mutación ●○○○○ en el fondo genético ○○●●, lo que significa que podemos evaluar tres casos de epistasia para este genotipo triple mutante. Teniendo en cuenta todos los genotipos mutantes múltiples, la descomposición de todas las interacciones posibles generó 75 posibles combinaciones genéticas para las que se calculó el valor de epistasia. Con todo ello, siguiendo las condiciones matemáticas detalladas en [Poelwijk et al. \(2011\)](#), se evaluó si los casos para los que se estimó un coeficiente de epistasia significativo también correspondieron a interacciones epistáticas de signo o recíprocas.

La epistasia de orden superior también se evaluó utilizando el servidor Web MAGELLAN ([Brouillet et al., 2016](#)) y calculando los coeficientes de Walsh ([Weinreich et al., 2013](#)). Los coeficientes de Walsh de orden cero representan la eficacia media en todos los genotipos; los de primer orden son equivalentes a los coeficientes de selección, que representan los efectos de eficacia de mutaciones individuales; los coeficientes de segundo orden son equivalentes a coeficientes epistáticos entre pares de mutaciones, es decir, ε_{ij} ; por último, los coeficientes de orden superior equivalen a las interacciones epistáticas de orden superior entre las mutaciones presentes en un genotipo. En el caso de paisajes de eficacia multiplicativa (en que no hay epistasia), todos los coeficientes de Walsh de orden 2 son iguales a cero, y por lo tanto el paisaje es

suave. En contraste, la robustez es máxima en un paisaje completamente aleatorio, con presencia de muchos picos locales que representan todo tipo de interacciones epistáticas. Los paisajes de eficacia reales se encuentran entre estos dos extremos, no siendo ni lisos ni exageradamente accidentados (Weinreich et al., 2013).

6. Análisis transcriptómico.

Los componentes virales, tanto proteínas como ácidos nucleicos, deben establecer múltiples interacciones entre ellos y con un gran número de factores del huésped (Whitham y Wang, 2004). En respuesta, las células tratan de compensar dichos efectos regulando positiva o negativamente la expresión de sus genes. Las micromatrices de ARN constituyen una moderna tecnología que permite estudiar las complejas interacciones planta-patógeno, de cuyo análisis podemos identificar cambios en los patrones de expresión de genes del huésped como consecuencia de la infección viral (Rowland y Jones, 2001; Wise et al., 2007).

6.1. Hibridación de micromatrices.

La extracción del ARN total se realizó como se describió anteriormente y se evaluó la integridad del ARN, antes y después de la hibridación, usando un bioanalizador Agilent 2100 (Agilent). Las muestras de ARN se hibridaron en 7 micromatrices *Gene Expression 4x44K* (AMADID: 021113), genotípicamente diseñadas para *N. tabacum*, que contenían 43.803 sondas (cada una compuesta por 60 nucleótidos). Dichas micromatrices se utilizaron en un diseño experimental de un solo color de acuerdo con las directrices MIAME (Brazma et al., 2001). Se analizaron tres réplicas biológicas por cada genotipo viral mutante. A su vez, se analizaron cuatro réplicas

biológicas para el aislado silvestre (controles positivos) y cuatro réplicas biológicas de plantas tratadas con tampón de inoculación (controles negativos). Se amplificó y se marcó el ARN de la muestra (200 ng) con el sistema *Agilent Low Input Quick Amp Labeling Kit* (Agilent 5190-2306). Paralelamente, se usó el sistema *Agilent One color Spike-In Kit* (Agilent 5188-5282) para evaluar la eficiencia del etiquetado y la hibridación. La hibridación y el lavado de las micromatrices se realizaron con los sistemas *Gene Expression Hybridization Kit* (Agilent 5188-5242) y *Gene Expression Wash Buffers* (Agilent 5188-5326) como se detalla en el protocolo *G4140-90040 GeneExpression one Color v6.6 proporcionado por Agilent*. Después de lavar y secar, las micromatrices fueron escaneadas en un equipo GenePix 4000B (Axon), estableciendo 5 μm de resolución. Las imágenes se extrajeron con el software *Feature Extraction 9.5.1* (Axon). Las hibridaciones de microarrays se llevaron a cabo en el Servicio de Genómica del IBMCP.

6.2. Análisis de genes expresados diferencialmente.

Las comparaciones entre micromatrices se realizaron con el software Babelomics 5. Para ello, los archivos que contenían los datos de las señales de hibridación de micromatrices se cargaron juntos para estandarizar los valores de expresión utilizando la normalización por cuantiles. Este proceso generó una matriz de expresión normalizada con los genes dispuestos en filas y las réplicas experimentales para cada genotipo en columnas. Antes de comparar los niveles de expresión de dos genotipos, se restó el nivel de expresión correspondiente al control negativo (plantas inoculadas únicamente con tampón de inoculación).

Después, se calculó la expresión diferencial de plantas infectadas con un genotipo mutante, incluidas las réplicas biológicas, comparándolo con el control negativo o plantas infectadas con el aislado silvestre de TEV, también con sus réplicas. Este cálculo se realizó con el paquete estadístico Limma (Smyth, 2004) y aplicando la

corrección Benjamini-Hochberg (Benjamini y Hochberg, 1995) (valor global $P < 0,05$). Además, con objeto de descartar aquellos genes cuya expresión entre ambos genotipos es muy pequeña aunque significativa, introducimos un criterio adicional de la expresión en las condiciones de interés respecto al correspondiente control ($FC = \log_2 \text{muestra/control} = 1$). Dicho umbral excluye del resultado todos aquellos genes de las plantas infectadas con un genotipo mutante cuya expresión no refleje un aumento o disminución mínimos del 50% del valor de expresión de la planta sana o plantas infectadas con el aislado silvestre de TEV (en función de con cual se compare). Finalmente, de la comparación de cada par de muestras obtenemos unos listados de genes expresados diferencialmente (DEGs), divididos en listados que únicamente incluyen genes sobreexpresados o reprimidos.

Por último, se repitió el análisis comparativo de las plantas infectadas con genotipos mutantes e infectadas con el aislado silvestre. En esta ocasión, se añadió un parámetro adicional para identificar los genes que muestran una correlación significativa entre eficacia y expresión ($P < 0,05$). De este modo, se generaron dos nuevas listas. La primera muestra los genes cuya expresión correlaciona positivamente con la eficacia viral, esto es, cuanto mayor sea la eficacia viral mayor será la expresión del gen en la planta. La segunda muestra los genes cuya expresión correlaciona negativamente con la eficacia viral, esto es, cuanto mayor sea la eficacia viral menor será la expresión del gen en la planta.

6.3. Análisis funcional a partir de los listados de genes.

Cuando se inició este estudio, teníamos disponible la vieja anotación del mapa genético del tabaco y también la anotación de Agilent. Tras revisar el proceso de anotación del chip de Agilent y advertir que la anotación del genoma estaba incompleta, decidimos construir nuestra propia anotación para el genoma del tabaco.

Material y Métodos

En primer lugar, la asociación de sondas individuales con los correspondientes genes del tabaco nos viene proporcionada por la información que Agilent da sobre su chip (ficheros 021113_D_AA_20130122.txt y 021113_D_GeneList_20130122.txt). A partir del archivo de las secuencias de las sondas de Agilent (021113_D_Fasta_20130122.txt), realizamos un BLAST contra el cADN de tabaco (Ntab-BX_AWOK-SS_Basma.mrna.annot.fna) para asociar cada sonda con un ARNm de tabaco de los definidos en la base de datos de solanáceas Solgenomics (ya que el objetivo era realizar una re-asociación del chip Agilent utilizando la información más reciente sobre secuencias de tabaco). El objetivo de ese nuevo archivo de anotación era tener la secuencia diana del tabaco para cada una de las diferentes sondas en el chip, de manera que tendríamos una secuencia en formato fasta para cada sonda (40.430 en total si eliminamos las secuencias no encontradas). Sin embargo, nos encontramos con diferentes sondas que remitían al mismo gen en el tabaco (por ejemplo, las sondas A_95_P217927 y A_95_P046006 tienen como diana el mismo gen: EB438730, por lo que el gen diana aparecerá dos veces). En determinado número de casos, cada sonda se asociaba con un ARNm diferente (continuando con el ejemplo, la sonda A_95_P217927 remitía al ARNm 6000 y la sonda A_95_P046006 al ARNm 7000). De hecho, el total de 40.430 sondas diferentes corresponden a 37.757 dianas diferentes en el tabaco.

A continuación, puesto que la anotación de tabaco es incompleta, procedimos a construir la anotación de genes ortólogos de tabaco en *Arabidopsis*. Realizamos otro BLAST contra la base de datos específica de cADN de *A. thaliana* TAIR10 para asociar el ARNm de tabaco con el correspondiente ortólogo de *A. thaliana* (siempre y cuando tuviera un e -valor $< 10^{-4}$). El proceso de anotación fue asistido por el Servicio de Bioinformática del IBMCP.

Para determinar las categorías de ontología génica (GOs) que se hallaban sobre-representadas dentro de las listas de DEGs

empleamos el servidor AgriGO (Tian et al., 2017). Una vez cargadas las listas, utilizamos la prueba exacta de Fisher aplicando la corrección Benjamini-Yekutieli (Benjamini y Yekutieli, 2001) ($P < 0,05$). Para la representación gráfica, construimos un esquema en un plano que incluye las categorías más relevantes, representadas como círculos cuyo tamaño es proporcional al número total de genes pertenecientes a esa categoría (en escala logarítmica).

Para terminar, con las dos listas de genes cuya expresión se correlaciona con la eficacia viral, positiva o negativamente, realizamos una serie de gráficos circulares en que se representa la proporción de genes con una determinada función biológica o función molecular. Estas figuras se realizaron cargando las listas en el servidor PANTHER (Mi et al., 2016).

6.4. Validación de la expresión génica mediante RT-qPCR.

Para dicho procedimiento, realizamos una nueva extracción de ARN de cada planta a partir de 100 mg de tejido fresco como se describió anteriormente. La concentración de la muestra purificada de ARN total se ajustó de nuevo a 50 ng/ μ l para cada muestra.

Una vez que seleccionamos los genes adecuados para la validación experimental decidimos cuantificar el ARN maduro post-transcripcional (teniendo en cuenta que las secuencias con las que hemos anotado el tabaco fueron las secuencias de ARNm, pudiéramos tener problemas con las cuantificaciones si los cebadores anidasen exactamente en un sitio de procesamiento de intrones). Para conocer la secuencia de ARN maduro de los genes seleccionados, cargamos las secuencias de los ARNm de nuestro interés y realizamos BLASTn y BLASTx contra la base de datos específica de *N. tabacum* (taxid: 4097). Los cebadores se diseñaron manualmente para cada gen sobre la base de estas secuencias de ARN maduros usando el programa OLIGO versión 7.

El nivel de expresión se midió mediante RT-qPCR utilizando como calibradores dos genes de expresión constitutiva (Schmidt y Delaney, 2010). El primero de ellos codifica la proteína ribosomal L25 (número de acceso en GenBank L18908), cuyos cebadores directo (NtL25-F; 5'-CCCCTCACCACAGAGTCTGC-3') y reverso (NtL25-R; 5'-AAGGGTGTGTTGTCCTCAATCTT-3') se seleccionaron para amplificar un fragmento de 51 nucleótidos. El segundo de ellos codifica el factor de elongación 1 α (número de acceso en GenBank AF120093), para el cual los cebadores directo (NtEF1 α -F; 5'-TGAGATGCACCACGAAGCTC-3') y reverso (NtEF1 α -R; 5'-CCAACATTGTCACCAGGAAGTG-3') amplifican también un fragmento de 51 nucleótidos.

Las cuantificaciones se realizaron en un volumen final de 20 μ l utilizando el sistema *GoTaq 1-Step RT-qPCR System* (Promega) y siguiendo las instrucciones del fabricante. Los cebadores directo e inverso para cada gen seleccionado se diseñaron para amplificar un fragmento de 68-137 nucleótidos en la secuencia de ARN diana. Las amplificaciones se realizaron usando de nuevo el equipo ABI StepOne Plus (Applied Biosystems), de acuerdo con el siguiente perfil: la fase RT consistió en 15 min a 37 °C y 10 min a 95 °C, seguido de la fase de PCR consistente en 40 ciclos de 10 s a 95 °C, 34 s a 60 °C y 30 s a 72 °C, terminando con la fase final con 15 s a 95 °C, 1 min a 60 °C y 15 s a 95 °C. Las amplificaciones se realizaron individualmente para cada gen objeto de estudio (con el conjunto de cebadores correspondiente) en una placa de 96 pocillos que contenía tres réplicas biológicas y dos réplicas técnicas por planta analizada. Además, se incluyeron en cada placa los dos genes de expresión constitutiva, necesarios para evaluar la calidad del procedimiento. Una vez completada la cuantificación procedimos a realizar la cuantificación al mismo nivel para todo el conjunto muestral, pero puesto que cada placa contiene la cuantificación de un solo ARN maduro junto con los genes constitutivos, hemos de establecer como valor de referencia del umbral de medición 0,105586. Los resultados de estas cuantificaciones también se examinaron utilizando el software StepOne v. 2.2.2 (Applied Biosystems).

4. Resultados y Discusión

Capítulo 1.

Las poblaciones de virus de ARN pueden escapar eficientemente de un óptimo local de eficacia.

Para evaluar la contribución de los eventos históricos, la presión de selección y el azar sobre el resultado de la evolución experimental de las poblaciones virales, seleccionamos cinco genotipos de TEV situados en diferentes posiciones en el paisaje de eficacia empírica mostrado en la **Figura 29**.

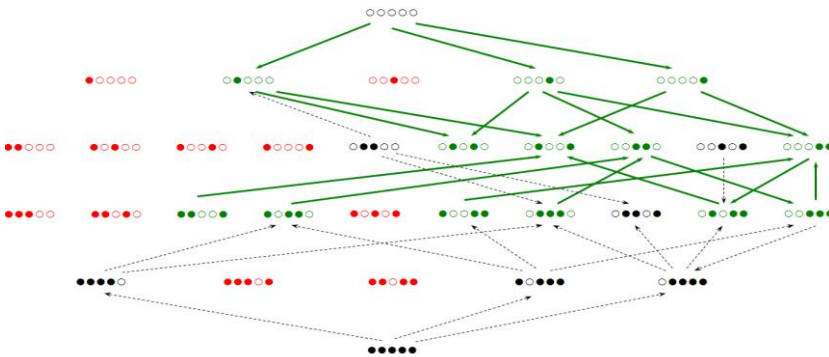


Figura 29. Topografía empírica del paisaje de eficacia local que describe la adaptación de TEV a su nuevo huésped *A. thaliana*. Cada cadena es un genotipo viral, con o representando los alelos silvestres y ● los alelos mutantes (**Tabla 2**). Los genotipos se ordenan comenzando por el virus ancestral proveniente del tabaco (o o o o o) y terminando con el genotipo quintuple mutante (● ● ● ● ●). Cada fila representa genotipos con igual número de mutaciones. Los genotipos marcados en rojo son aquellos con una eficacia inferior a la del genotipo ancestral (incluyendo letales), los marcados en negro tienen una eficacia igual al genotipo ancestral, y los marcados en verde son significativamente más eficaces que el ancestral. Las líneas verdes sólidas representan caminos adaptativos (que siempre conectan genotipos beneficiosos) y las líneas negras discontinuas conectan genotipos no significativamente mejores que el ancestral (rutas neutras). El paisaje construido es altamente epistático y contiene dos picos adaptativos independientes y no conectados correspondientes a los genotipos o ● o o ● y o o ● o o. El pico o ● o o ● se corresponde con la eficacia máxima de este paisaje local. (Figura adaptada de [Lalić y Elena, 2015](#)).

Tres de los genotipos seleccionados (●●○○○, ○○○○● y ○●●●●) se localizaban a un solo paso mutacional del óptimo de eficacia local (genotipo ○●○○●), el genotipo ○○○○● se hallaba a tres pasos mutacionales del óptimo de eficacia local, pero sólo a un salto de distancia del segundo pico de eficacia (○●●●○), y el genotipo ●●●●○ se encontraba a cinco pasos mutacionales del óptimo local y, de nuevo, a sólo un salto del segundo pico de eficacia. A partir de cada uno de estos cinco genotipos, se iniciaron cinco linajes independientes de evolución mediante cinco pases seriados en el mismo huésped, *A. thaliana*. Después de esta fase de evolución experimental, se evaluó la eficacia de los 25 linajes evolucionados tomando como referencia la eficacia del genotipo situado en el óptimo de eficacia local (○●○○●). Los resultados de todos estos ensayos de eficacia se recopilan en la **Figura 30**.

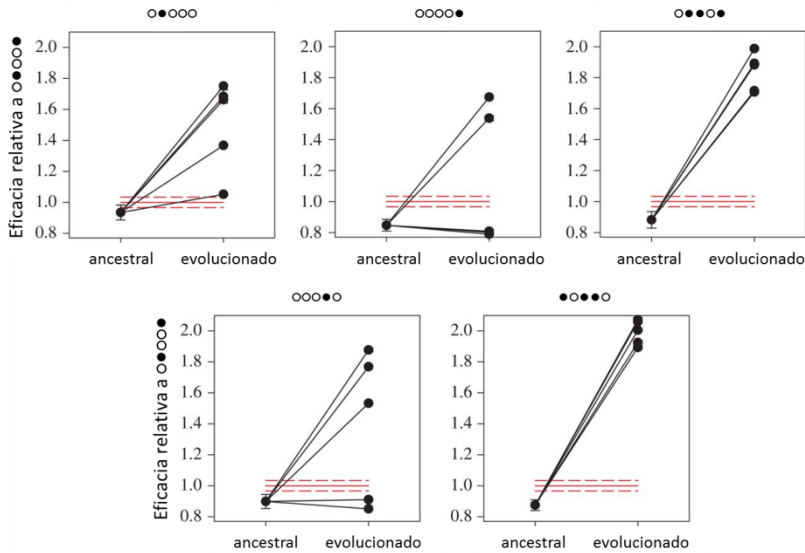


Figura 30. Evolución de la eficacia de los genotipos de TEV a diferentes distancias del óptimo de eficacia local (genotipo ○●○○●). La línea roja continua representa la eficacia del óptimo local y las líneas rojas discontinuas representan el error estándar de la media.

Diecinueve de los 25 linajes evolucionados alcanzaron valores de eficacia significativamente mayores que el del genotipo de referencia $\bullet\bullet\bullet\bullet\bullet$ situado en el máximo local de eficacia, cuyo valor queda representado con líneas rojas en la **Figura 30** (pruebas t de dos muestras, todas corregidas para múltiples pruebas $P \leq 0,002$).

Asimismo, se observaron dos linajes independientes, uno a partir de $\bullet\bullet\bullet\bullet\bullet$ y otro de $\bullet\bullet\bullet\bullet\bullet$, que alcanzaron valores de eficacia indistinguibles del máximo local (pruebas t de dos muestras, ambas corregidas $P \geq 0,05$). Otros cuatro linajes independientes, tres que evolucionaron a partir del genotipo $\bullet\bullet\bullet\bullet\bullet$ y uno a partir del genotipo $\bullet\bullet\bullet\bullet\bullet$, conservaron valores de eficacia tras la evolución que todavía eran significativamente inferiores al máximo local (pruebas t de dos muestras, todos corregidos $P \leq 0,001$).

1. Paralelismo y divergencia en la eficacia.

La **Figura 30** sugiere la existencia de un paralelismo fenotípico evolutivo que depende de la posición inicial en el paisaje. Para evaluar la contribución de la divergencia genética en contraposición al paralelismo fenotípico, empleamos el índice $I_W = \sigma_W / \langle \Delta W \rangle$, donde σ_W es la desviación genética estándar (o divergencia genética) para la eficacia entre los linajes evolucionados y $\langle \Delta W \rangle$ es una medida de paralelismo fenotípico que contempla el cambio promedio de la eficacia entre los linajes evolucionados y los genotipos iniciales (Vasi et al., 1994). Los linajes que evolucionaron a partir del genotipo $\bullet\bullet\bullet\bullet\bullet$ poseen un valor de $I_W > 1$, lo que indica poco paralelismo entre la eficacia y la divergencia genética observada entre los linajes (**Tabla 6** y **Figura 30**). En todos los demás casos obtenemos un valor de $I_W < 1$, lo que indica que el paralelismo fenotípico fue mayor de lo esperado debido a la divergencia genética entre los linajes (**Tabla 6** y **Figura 30**). Los linajes evolucionados del genotipo $\bullet\bullet\bullet\bullet\bullet$ fueron aquellos en que se observó mayor paralelismo fenotípico (los cinco linajes independientes prácticamente alcanzaron el mismo valor de

eficacia), mientras que los linajes del genotipo ○○○●○ registraron el menor valor de paralelismo fenotípico. Esto sugiere que, hasta cierto punto, el resultado de la evolución depende en realidad del punto de partida en el paisaje adaptativo.

Tabla 6. Evaluación de la divergencia genética y el paralelismo fenotípico entre los cinco linajes evolucionados a partir del mismo genotipo de partida. La eficacia se expresa en relación con el genotipo óptimo local ○●○○●. Los errores representan $\pm 1 d. e.$

| Genotipo de partida | Eficacia inicial (W_0) | Cambio promedio en eficacia (ΔW) | Varianza genética (σ_G^2) ^a | Índice de paralelismo (I_W) |
|---------------------|----------------------------|--|---|---------------------------------|
| ○●○○○ | 0,934 \pm 0,153 | 0,570 \pm 0,429 | 0,068 \pm 0,043 | 0,458 \pm 0,491 |
| ○○○●○ | 0,899 \pm 0,171 | 0,489 \pm 0,623 | 0,184 \pm 0,117 | 0,878 \pm 1,397 |
| ○○○○● | 0,847 \pm 0,124 | 0,275 \pm 0,543 | 0,158 \pm 0,100 | 1,446 \pm 3,314 |
| ●○○●○ | 0,875 \pm 0,178 | 1,116 \pm 0,254 | 0,005 \pm 0,003 | 0,063 \pm 0,035 |
| ○●●○○ | 0,882 \pm 0,168 | 0,955 \pm 0,283 | 0,012 \pm 0,008 | 0,114 \pm 0,070 |

^a Estimadores de máxima verosimilitud.

2. Contribuciones de la selección, el azar y la historia al resultado de la evolución.

La diversificación fenotípica entre los linajes evolucionados puede resultar de la contribución de tres factores evolutivos: la adaptación por selección natural, los eventos fortuitos como las mutaciones y la deriva, y la historia evolutiva previa experimentada por una población. La contribución de dichos factores al cambio evolutivo es una cuestión que aún se debate a día de hoy (Travisano et al., 1995) y es particularmente difícil aplicarlos cuando el objeto de estudio son virus (Meyer et al., 2012; Morley et al., 2015). Para averiguar el porcentaje de contribución de estos tres factores a los patrones de cambio en eficacia observados en este experimento para

TEV, aplicamos los métodos estadísticos desarrollados por [Travisano et al. \(1995\)](#). La [Figura 31.A](#) muestra la eficacia de cada linaje independiente antes y después de los cinco pases seriadados de la evolución experimental. El patrón que observamos es consistente con un fuerte efecto de la contingencia histórica (diferencias en la eficacia de los genotipos iniciales), cuya importancia prácticamente desaparece después de la evolución. Para modelizar las contribuciones de la adaptación, el azar y la historia estimamos el cambio en la gran media de la eficacia $0,681 \pm 0,430$ (1 d. e.), que reflejaría la adaptación. Con un modelo ANOVA anidado (**apartado 4.2 de Material y Métodos**) estimamos los componentes de la varianza correspondientes al azar (diferencias entre linajes dentro de los genotipos) e historia (diferencias entre genotipos) ([Tabla 7](#)).

Tabla 7. Modelo ANOVA anidado de la varianza para estimar la contribución del azar y la historia al patrón observado de diversificación en la eficacia viral.

| Fuente de variación | SS ^a | g.l. | F | p-valor |
|--------------------------|-----------------|------|---------|---------|
| Interceptación (μ) | 122,963 | 1 | 574,703 | < 0,001 |
| Genotipo | 4,865 | 4 | 5,684 | 0,003 |
| Linaje | 4,279 | 20 | 771,342 | < 0,001 |
| Error | 0,007 | 25 | | |

| Fuente de variación | η_p^2 ^b | $1 - \beta$ ^c | Componente de la varianza ^d | d. e. de la varianza ^d |
|--------------------------|-------------------------|--------------------------|--|-----------------------------------|
| Interceptación (μ) | 0,966 | 1 | | |
| Genotipo | 0,532 | 0,942 | 0,0759 | 0,0619 |
| Linaje | 0,998 | 1 | 0,1068 | 0,0338 |
| Error | | | $2,770 \times 10^{-4}$ | $7,845 \times 10^{-5}$ |

^a Suma de cuadrados tipo 3.

^b η_p^2 mide la magnitud del efecto.

^c Potencia estadística de la prueba realizada.

^d Estimadores de máxima verosimilitud para componente y su error.

La **Figura 31.B** muestra las contribuciones relativas de los tres factores. Después de cinco pases seriadados de evolución experimental, observamos que la gran media de la eficacia de los 25 linajes ha aumentado significativamente (prueba z , $P < 0,001$), lo que sugiere que el factor de mayor contribución a la diversificación fenotípica es la adaptación. El efecto de la historia aún es significativo, aunque menor que el efecto del azar (**Tabla 7**, comparar los valores de η_P^2). Ello se debe principalmente a la presencia de los cuatro linajes (tres linajes que evolucionaron a partir del genotipo $\circ\circ\circ\bullet$ y uno a partir del genotipo $\circ\circ\bullet\circ$) cuya eficacia continuaba siendo inferior en relación al genotipo de referencia $\circ\bullet\circ\bullet$ (**Figura 30** y puntos cercanos a la línea roja en la **Figura 31.A**). El efecto del azar fue muy significativo (**Tabla 7**) y tuvo un gran impacto en la variabilidad de la eficacia observada. Estos resultados son congruentes con la débil influencia de las historias evolutivas experimentales en la adaptación de TEV a un nuevo huésped ([Bedhomme et al., 2013](#)).

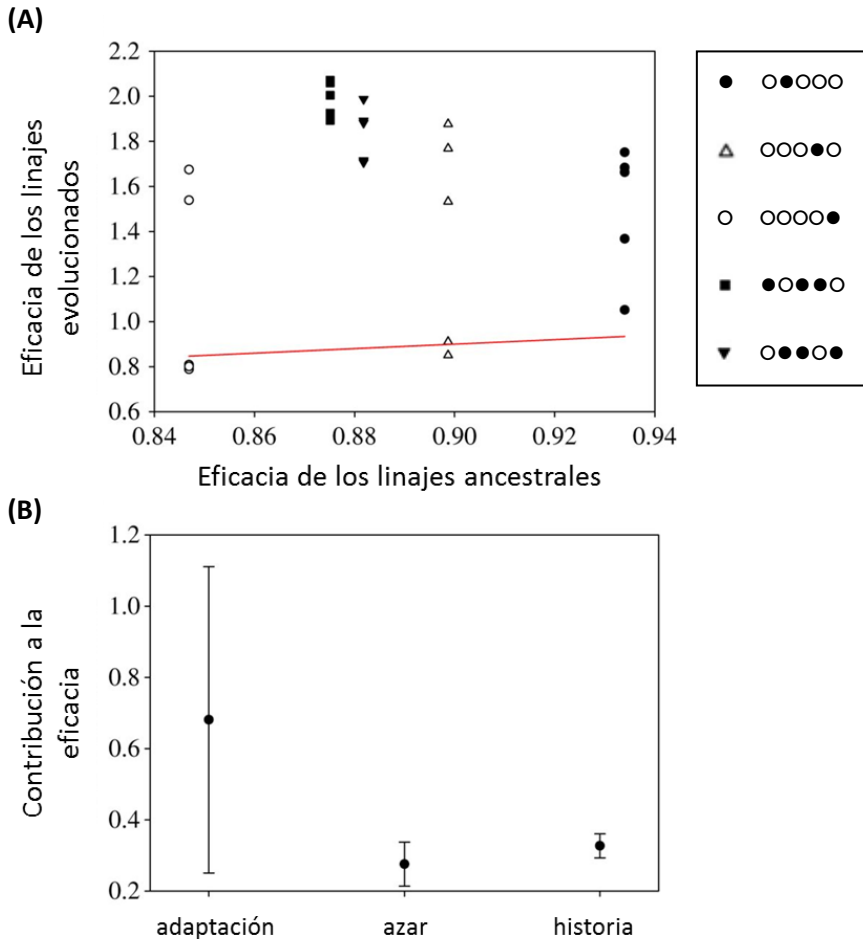


Figura 31. Representación de la magnitud de la contribución de la adaptación, el azar y la historia evolutiva previa en la evolución y diversificación de poblaciones experimentales de TEV evolucionadas en *A. thaliana*. (A) Valores de eficacia en los 25 linajes evolucionados después de cinco pases seriados en plantas de *A. thaliana* (respecto al genotipo óptimo local ○●○○●). La línea roja corresponde a la hipótesis de diferencias de eficacia que sólo pudieran explicarse por efectos históricos sin ninguna contribución del azar y la adaptación. Cada genotipo está representado por diferentes símbolos (leyenda). (B) Contribuciones relativas de la adaptación, el azar y la historia en la evolución de la eficacia. Las barras de error representan $\pm 1 d. e.$

3. Diversificación genotípica.

Buscando apoyo para la hipótesis previa de que la proximidad al máximo de eficacia local condicionaría el conjunto de mutaciones fijadas tras el proceso evolutivo, nos preguntamos si la contingencia es todavía detectable a nivel genotípico. Para tratar esta cuestión, obtuvimos la secuencia consenso poblacional de los 25 linajes evolucionados. Una secuencia de consenso poblacional representa un "genotipo promedio" en el que sólo se detectan mutaciones por encima de una frecuencia dada en la población; es por tanto una construcción estadística que podría no tener existencia real en la población. La **Figura 32** muestra una representación esquemática del genoma consenso de TEV y de las mutaciones fijadas dentro de cada linaje.

Ninguno de los linajes evolucionados a partir del genotipo ○○○○ fijó la mutación ○○○○●, que los hubiera llevado al óptimo local en un único paso mutacional; sino que cuatro linajes adquirieron cada uno una serie de mutaciones que elevaron significativamente su eficacia en mayor o menor grado. Hay que destacar el linaje 4, que constituye la excepción, en el que otra serie mutaciones diferentes permitió que adquiriera un valor de eficacia indistinguible del óptimo local (**Figura 30**). Curiosamente, los linajes 4 y 5 habían fijado el mismo par de mutaciones: **P1/G418A** que dio lugar al cambio **P1/G92S** (una mutación conservadora, ya que ambos aminoácidos son diferentes, pero poseen propiedades bioquímicas similares; un radical corto y apolar) y **N1b/G7668A** que dio lugar a una mutación sinónima.

La situación es más compleja si centramos nuestra atención en los linajes evolucionados a partir del genotipo ○○○○●. Los linajes 1 y 2 fijaron la mutación ○●○○○, aunque el linaje 1 había revertido la mutación ○○○○● al alelo ancestral. Del mismo modo, en el linaje 3 también se observó la reversión de la mutación ○○○○● al alelo ancestral. Así, sólo el linaje 2 ha evolucionado hasta convertirse en ○●○○● y representa el más claro ejemplo observado en este trabajo

de determinismo evolutivo a nivel genómico. Sin embargo, la eficacia del linaje 2 resultó ser significativamente mayor que el máximo de eficacia local, probablemente debido a una única mutación adicional fijada en la proteína de cubierta (CP) (**Figura 32**). Curiosamente, los linajes 3 y 5 fijaron la mutación $\circ\circ\circ\bullet\circ$. Por último, los linajes 1 y 4 habían fijado la mutación **3'UTR/G9444A**.

Respecto a los linajes evolucionados a partir del genotipo $\circ\bullet\bullet\circ\bullet$, ninguno de los cinco había revertido la mutación $\circ\circ\bullet\circ\circ$ al alelo ancestral ni se fijaron mutaciones convergentes entre ellos. Únicamente revertir la mutación $\circ\circ\bullet\circ\circ$ los hubiera llevado al óptimo local; a pesar de ello, todos los linajes aumentaron significativamente su eficacia fijando únicamente una (linajes 3 y 5) o dos mutaciones (linajes 1,2 y 4) nuevas.

Ninguno de los linajes a partir del genotipo $\circ\circ\circ\bullet\circ$ había revertido dicha mutación ni adquirieron las mutaciones $\circ\bullet\circ\circ\circ$ ni $\circ\circ\circ\circ\bullet$. En cambio, pudimos observar en estos linajes tres ejemplos de nuevas mutaciones convergentes. En primer lugar, los linajes 2 y 4 fijaron la mutación **P3/U2537C** que dio lugar al cambio **P3/I798T** (un radical no polar por uno polar). En segundo lugar, los linajes 1 y 2 fijaron la mutación **CP/G8556C** que reemplaza **CP/K2804N** (radical básico por uno polar sin carga neta). En último lugar, más interesante, los linajes 1, 2, 4 y 5 fijaron la mutación **CP/G9103U** que da como resultado el cambio de aminoácido **CP/A2987S** (radical no polar por uno polar).

Para terminar, todos los linajes evolucionados a partir del genotipo $\bullet\circ\bullet\circ\circ$ habían fijado la mutación sinónima **CI/A3975U**. Además, los linajes 1, 2, 4 y 5 tienen una mutación fija **VPg/A6175C** que dio lugar al cambio de aminoácido **VPg/N2011H** (radical polar por básico).

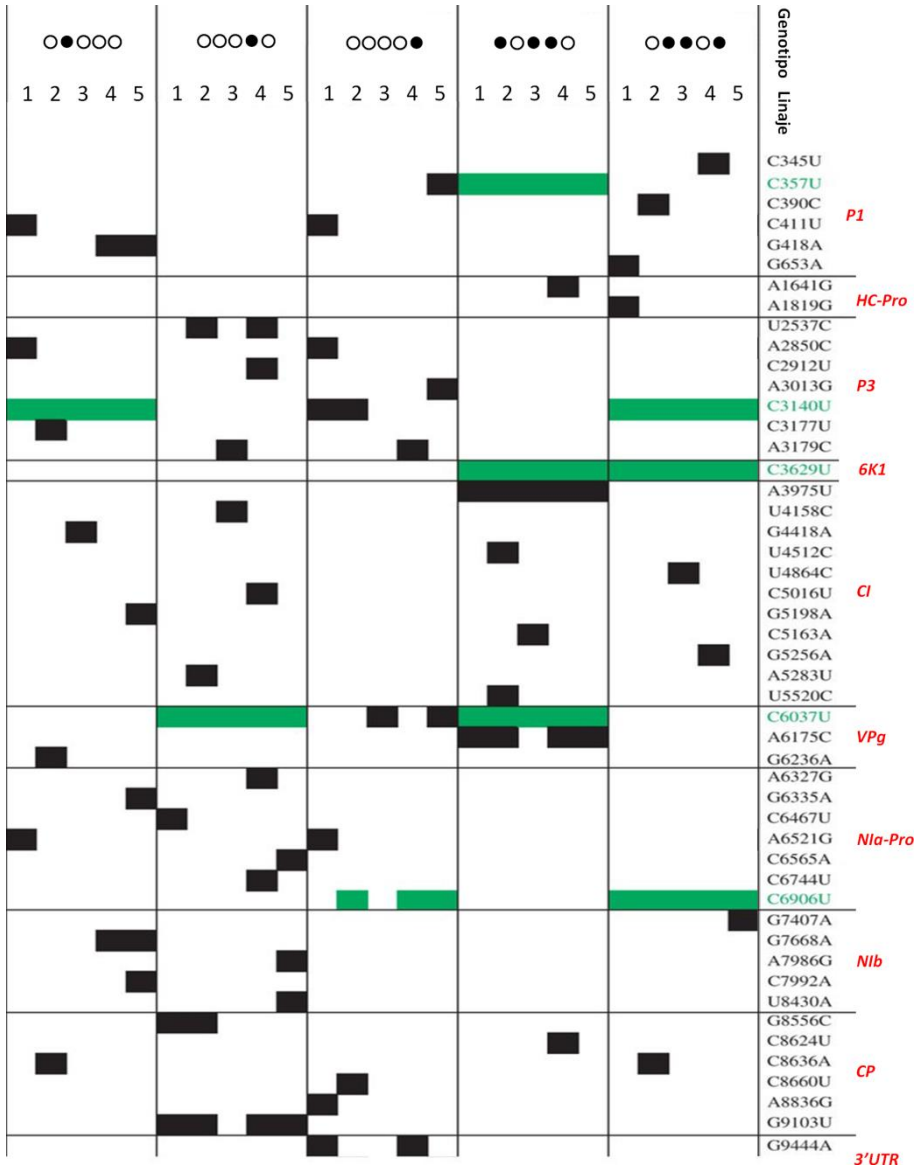


Figura 32. Mutaciones fijadas en los linajes después de cinco pases adicionales de evolución experimental en *A. thaliana*. Los diferentes cistrones en el genoma de TEV se indican en rojo en la parte superior. Las mutaciones originalmente presentes en cada genotipo están indicadas por celdillas verdes. Las celdillas negras indican nuevas mutaciones fijadas durante el experimento de evolución.

Con todo lo descrito anteriormente, se concluye que la presencia de una determinada mutación, o conjunto de mutaciones, en el camino hacia un pico local adaptativo, no determina la fijación de las mutaciones subsecuentes. Por el contrario, en 24 de 25 casos analizados, los linajes evolucionados fijaron nuevos conjuntos de mutaciones (con la única excepción del linaje 3 del genotipo ○○○○●), explorando partes distantes del paisaje. Sin embargo, también detectamos cierto nivel de paralelismo a nivel molecular, con algunas mutaciones fijadas en más de un linaje.

4. Discusión.

En el contexto del trabajo realizado, en que pretendemos arrojar luz sobre los mecanismos evolutivos que impulsan la adaptación de los virus a nuevos huéspedes, hemos descrito los resultados de experimentos diseñados específicamente para evaluar el efecto de la rugosidad del paisaje adaptativo, siendo ésta la primera vez que se realiza tal estudio para un virus vegetal de ARN. Nuestros resultados sugieren que, independientemente de la proximidad de las poblaciones virales a los picos adaptativos locales, existen posibilidades de que puedan alejarse de la cuenca de atracción del pico cercano y explorar nuevas regiones del paisaje. Hemos observado que estos movimientos son en su mayoría debidos a la adaptación, mientras que los efectos del azar y la historia evolutiva previa (contingencia genética) tuvieron una contribución menor, aunque significativa. Particularmente, la historia evolutiva previa resultó ser un factor de enorme importancia en algunos de los linajes estudiados, cuya eficacia siempre fue inferior a la del genotipo de referencia. Asimismo, hemos observado paralelismo a nivel genotípico y fenotípico. Por último, la contribución a la variación genética dependía del genotipo de partida, pero esta dependencia no

se correlacionó con la proximidad del genotipo de partida al pico de eficacia local (se exploraron partes distantes del paisaje de eficacia).

Una pregunta que debemos hacernos es por qué la contingencia histórica no tuvo un mayor impacto en la evolución, especialmente para aquellas poblaciones que estaban a sólo un paso mutacional del pico adaptativo local. No podemos ofrecer una respuesta concreta pero podemos apuntar a dos posibilidades, no mutuamente excluyentes. En primer lugar, el paisaje empírico restringido a combinaciones de sólo cinco mutaciones ([Lalić y Elena, 2015](#)) no es la única solución posible para que TEV pueda adaptarse a *A. thaliana*, sino que existirían otras opciones, tal vez incluso mejores, que no fueron exploradas en el experimento de evolución por pases seriados originalmente realizado por [Agudelo-Romero et al. \(2008a\)](#). Cabe señalar que el experimento original dio como resultado sólo un linaje evolutivo exitoso, es decir, no hemos reproducido el experimento original y no tenemos evidencia de que pudiéramos reproducir este paseo adaptativo particular. En segundo lugar, es posible que cerca de un pico óptimo local la dinámica adaptativa se ralentice, dando lugar a una probabilidad de fuga estocástica comparable a la de un proceso aleatorio. Entonces, la población deambula en el espacio genotípico comenzando un nuevo camino adaptativo tras cada fuga exitosa ([Weinreich y Chao, 2005](#); [Jain y Krug, 2007](#)). Sea cual sea el caso, nuestros resultados confirman una vez más el tremendo potencial adaptativo y la capacidad evolutiva de los virus de ARN.

Capítulo 2.

Efecto de la especie huésped en la topografía del paisaje de eficacia para un virus vegetal de ARN.

En este capítulo, evaluaremos y compararemos la topografía de los paisajes adaptativos de 32 genotipos de TEV en dos huéspedes distintos; el huésped original, *N. tabacum*, y el nuevo huésped, *A. thaliana*. De hecho, el paisaje de eficacia que utilizaremos aquí para *A. thaliana* es el mismo que describimos previamente en el **Capítulo 1**, con idénticos genotipos y huésped. La diferencia radica en la infección del huésped original con estos 32 genotipos de TEV. De este modo, con los cálculos de cuantificación y eficacia, construimos el paisaje de eficacia para *N. tabacum* y pudimos proceder a comparar ambos paisajes estadísticamente. La **Figura 33** muestra ambos paisajes adaptativos. Como viene siendo habitual, cada cadena de caracteres representa un genotipo; o representa el alelo silvestre en el locus correspondiente mientras que ● corresponde al alelo mutante en ese mismo locus. Los genotipos en una caja verde corresponden a los picos de eficacia locales. Las líneas verdes corresponden a mutaciones beneficiosas, las líneas rojas a mutaciones deletéreas y las líneas naranjas a cambios neutros (en la dirección del genotipo ooooo al genotipo ●●●●●).

Asimismo, la **Tabla 8** contiene los datos estadísticos que describen la topografía de dichos paisajes. Como en casos anteriores, un pico de eficacia se define como aquel genotipo cuyos vecinos tienen una menor eficacia. En el caso del paisaje adaptativo de *A. thaliana*, observamos dos picos de eficacia de diferente altura correspondientes a los genotipos o●ooo y oo●oo (Lalić y Elena, 2015). En el paisaje adaptativo del huésped original, *N. tabacum*, observamos cuatro picos de eficacia correspondientes a los genotipos ●oooo, o●ooo, ooo●● y ●●●●o. Podemos evaluar la robustez del paisaje mediante diferentes medidas (**Tabla 8**). Por ejemplo, la relación entre la pendiente media y la rugosidad, θ , (Aita et al., 2001) mostró valores similares para ambos huéspedes ($\theta > 1$), lo que indica que los paisajes tienen una topografía rugosa. Recientemente se propuso una medida de epistasia calculada como la correlación entre los efectos de eficacia de un determinado genotipo y todos sus vecinos interconectados a una sola mutación de distancia. Dicha

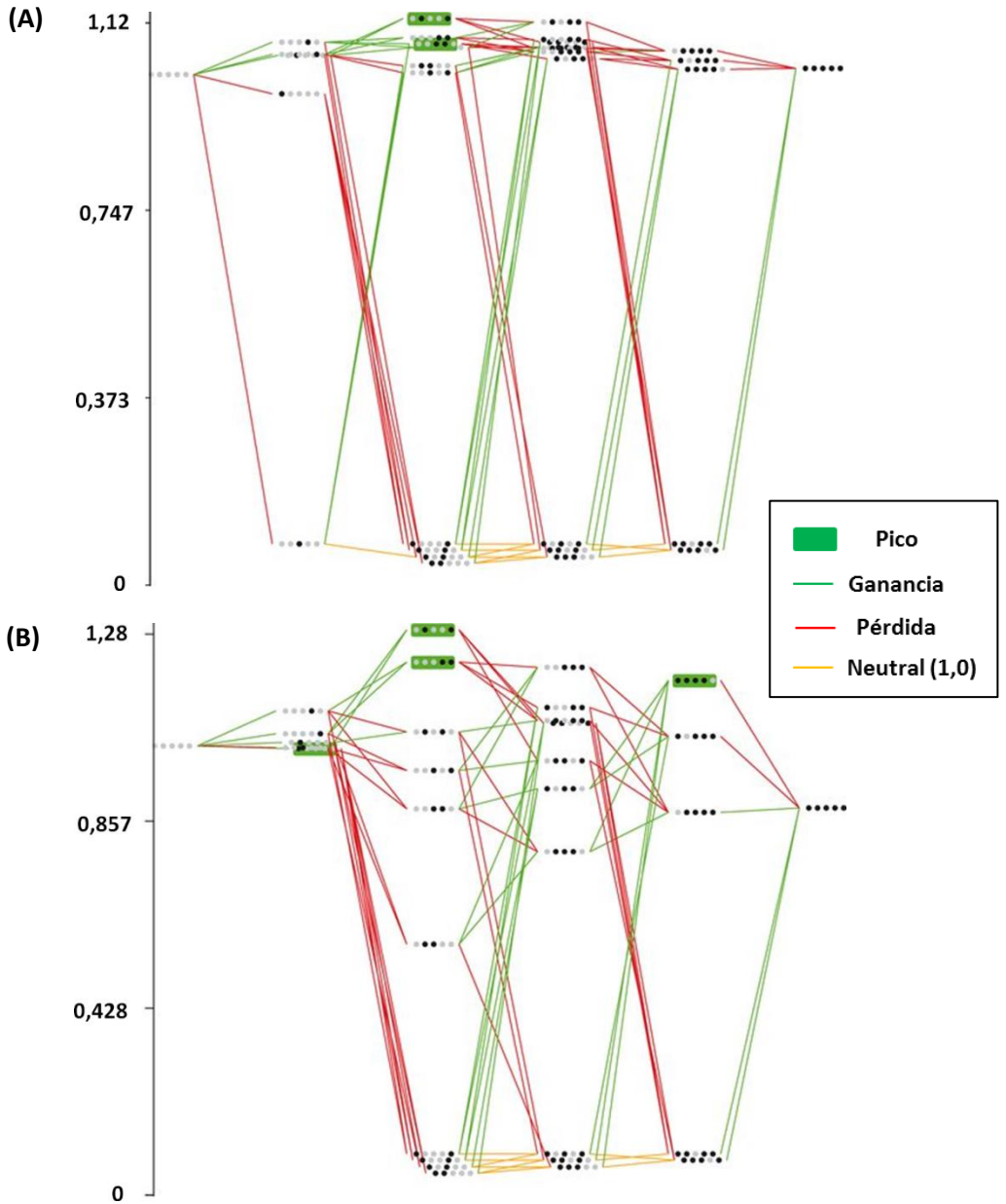


Figura 33. Paisajes de eficacia empíricos para las cinco mutaciones fijadas en TEV durante su evolución experimental en *A. thaliana*. La eficacia de los 32 genotipos se evaluó en el nuevo huésped (A) y en el original, *N. tabacum* (B).

medida fue definida con el parámetro ρ (Ferretti et al., 2016). En nuestro caso, ambos valores para dicho parámetro fueron positivos y cercanos a cero, lo que sugiere la existencia de muchos casos de epistasia de magnitud. Otra medida intuitiva de la robustez de un paisaje es calcular la frecuencia de cada tipo de interacción epistática entre todos los pares de mutaciones posibles. Así, para un paisaje liso todas las interacciones serán no epistáticas y, conforme aumente la rugosidad del paisaje, las interacciones de signo y de signo recíproco serán más comunes (Weinreich et al., 2005; Poelwijk et al., 2011). La **Tabla 8** indica que la mayoría de las mutaciones interactúan epistáticamente en ambos huéspedes, siendo la epistasia de magnitud el tipo más común de interacción en ambos paisajes. La epistasia de signo fue el segundo tipo de interacción más frecuente en *A. thaliana*, mientras que la epistasia de signo recíproco fue el segundo tipo de interacción más frecuente en *N. tabacum*. Estos resultados sugieren que el paisaje definido por las cinco mutaciones puntuales que conforman los 32 genotipos es más rugoso en el huésped original que en el nuevo huésped. En los siguientes apartados detallaremos estos resultados de forma más amplia.

Tabla 8. Resumen de los parámetros estadísticos que describen la topografía de ambos paisajes.

| Parámetro estadístico ^a | <i>A. thaliana</i> | <i>N. tabacum</i> |
|--|--------------------|-------------------|
| Estadísticos generales | | |
| Nº de picos | 2 | 4 |
| Estadísticos epistáticos | | |
| Relación entre pendiente y rugosidad (θ) | 1,902 | 1,697 |
| Correlación entre los efectos de eficacia (ρ) | 0,119 | 0,111 |
| Frecuencia de interacciones multiplicativas | 0,013 | 0,013 |
| Frecuencia de epistasia de magnitud | 0,662 | 0,575 |
| Frecuencia de epistasia de signo | 0,212 | 0,188 |
| Frecuencia de epistasia de signo recíproco | 0,113 | 0,225 |

^a Parámetros calculados utilizando el servidor web MAGELLAN.

1. Correlaciones y pleiotropía entre los huéspedes.

Para explorar aún más la relación entre las topografías de los dos paisajes que se muestran en la **Figura 33**, evaluamos la similitud de los efectos de eficacia estimados para cada genotipo en cada especie huésped (**Figura 34**). Los valores de eficacia están significativamente correlacionados entre los dos paisajes (correlación de Pearson: $r = 0,891$; 30 gl ; $P < 0,001$). Sin embargo, esta correlación es debida a la existencia de un grupo de genotipos que son letales en ambos huéspedes (genotipos en el margen inferior de los paisajes de eficacia de la **Figura 33**). Si estos genotipos se eliminan del análisis, la correlación ya no es significativa ($r = 0,338$; 20 gl ; $P = 0,124$). Las líneas discontinuas en la **Figura 34** representan la eficacia relativa del aislado silvestre de TEV en ambos huéspedes (genotipo que ocupa exactamente el centro del diagrama). Dichas líneas dividen el plano en cuatro regiones, cada una de las cuales engloba los genotipos con los valores de eficacia más grandes o más pequeños que el del genotipo silvestre (oooo) para cada huésped. Para representarlo más claramente las cuatro regiones se dividen como sigue: la región superior izquierda engloba los genotipos con mayor eficacia que el genotipo silvestre en *N. tabacum* y menor que éste en *A. thaliana*; la región superior derecha engloba genotipos cuya eficacia es mayor en ambos huéspedes; la región inferior izquierda engloba los genotipos cuya eficacia es menor en ambos huéspedes; la región inferior derecha engloba los genotipos con menor eficacia que el genotipo silvestre en *N. tabacum* y mayor que éste en *A. thaliana*. Doce genotipos tenían valores de eficacia relativa más grandes que el del tipo silvestre en ambos huéspedes y por lo tanto eran incondicionalmente beneficiosos (región superior derecha). En contraste, diez genotipos fueron incondicionalmente deletéreos, con valores de eficacia más pequeños que el del tipo silvestre en ambos huéspedes (región inferior izquierda). Respecto a estos diez genotipos cabe añadir que nueve de ellos fueron letales en ambos huéspedes y el genotipo oo●oo fue letal en *A. thaliana* pero ligeramente deletéreo ($-0,5 \%$ de efecto) en *N. tabacum*. Estos 22

genotipos, que ocupan los cuadrantes superior derecho e inferior izquierdo de la **Figura 34**, explican la correlación descrita anteriormente. Los casos en los otros dos cuadrantes son más interesantes, ya que representan ejemplos de pleiotropía antagonista, es decir, genotipos beneficiosos en un huésped pero que son deletéreos en el otro. De forma breve, el genotipo ●oooo fue beneficioso en *N. tabacum* pero perjudicial en *A. thaliana*. Además, ocho genotipos tenían valores de eficacia mayores al del genotipo silvestre en *A. thaliana* pero eran deletéreos en el huésped original, lo cual sugiere que estos genotipos probablemente fueron generados y seleccionados positivamente durante el proceso de adaptación al nuevo huésped.

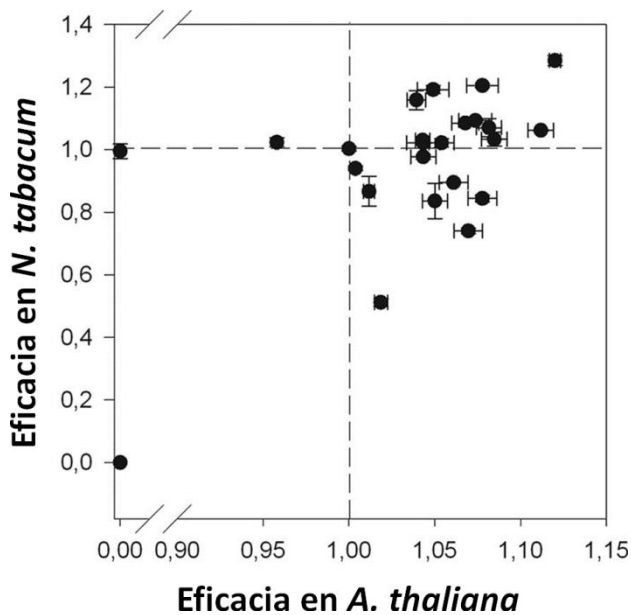


Figura 34. Valores de eficacia medidos para los 32 genotipos y en ambos huéspedes. La eficacia se expresa en relación con la del genotipo silvestre (oooo). Las líneas discontinuas corresponden a la eficacia del genotipo silvestre en cada huésped. El punto 0,0 incluye los nueve casos de genotipos incondicionalmente letales. Las barras de error representan $\pm 1 d. e.$

Antes de cualquier análisis estadístico adicional, los datos de eficacia se verificaron para comprobar que no existía violación de los supuestos de normalidad y homoscedasticidad de las varianzas. Se observó que los datos no estaban distribuidos normalmente (prueba de Kolmogorov-Smirnov de una muestra: $D = 0,248$; $P < 0,001$), ni las varianzas eran homogéneas entre los grupos (prueba de Levene: $F_{229,399} = 13,980$; $P < 0,001$). Un modelo general linearizado, con distribución gamma y función de enlace de logaritmo (ambas opciones seleccionadas porque tenían el valor más pequeño para el criterio de información bayesiano entre el conjunto de modelos alternativos comprobados), se ajustó a los datos de eficacia para evaluar las contribuciones relativas de los genotipos y las especies huéspedes a la variabilidad observada en la eficacia. La **Tabla 9** muestra los resultados de este análisis. En general, existen diferencias altamente significativas entre los 32 genotipos ($P < 0,001$), que constituyen el factor que más contribuye a las diferencias observadas en la eficacia ($\eta_p^2 = 0,860$). El 77,8 % de la varianza total observada es explicada únicamente por las diferencias genéticas entre los genotipos virales (factor G). La contribución neta de las especies huéspedes a la eficacia también es significativa ($P = 0,037$), aunque la magnitud del efecto es muy pequeña ($\eta_p^2 = 0,006$). Tan sólo el 0,2 % de la varianza total observada es explicada únicamente por las especies huéspedes (factor H) y, en consecuencia, el poder estadístico asociado con esta prueba es demasiado bajo para que el resultado se considere realmente fiable. Sin embargo, existe un efecto altamente significativo ($P < 0,001$) de las especies huéspedes que depende de cada genotipo, y la magnitud de este efecto de interacción también es grande ($\eta_p^2 = 0,500$). El 16,7 % de la varianza total observada puede explicarse por la interacción de las especies huéspedes y los genotipos virales (factor $G \times H$). El hecho de que dicha interacción contribuya a la eficacia en un grado mucho mayor que el de las especies huéspedes consideradas individualmente (0,2 %) tiene una consecuencia importante: los dos paisajes difieren en detalles pequeños y

relevantes más que en detalles generales o amplios. Por último, las diferencias entre plantas de cada especie huésped inoculadas con el mismo genotipo viral son también significativas ($P < 0,001$) y tienen una magnitud comparable a la del factor G ($\eta_p^2 = 0,848$). A pesar de ello, sólo explica el 4,2 % de la varianza total observada (factor $P(G \times H)$).

Tabla 9. Resumen del modelo general linearizado ajustado a los datos.

| Factor | LRT ^a | g.l. | p-valor | η_p^2 ^b | $1 - \beta$ ^c |
|--------------------------|------------------|------|---------|-------------------------|--------------------------|
| Interceptación (μ) | 3.979,285 | 1 | < 0,001 | 0,982 | 1 |
| G | 2.397,695 | 31 | < 0,001 | 0,860 | 1 |
| H | 4.341 | 1 | 0,037 | 0,006 | 0,183 |
| $G \times H$ | 1.344,481 | 20 | < 0,001 | 0,500 | 1 |
| $P(G \times H)$ | 1.168,930 | 177 | < 0,001 | 0,848 | 1 |

^a Estimadores de verosimilitud.

^b η_p^2 mide la magnitud del efecto asociado a cada factor.

^c Potencia estadística de la prueba realizada.

2. Diferencias en la epistasia del paisaje entre los huéspedes.

A continuación, exploraremos la relación entre los valores de epistasia y sus diferentes tipos entre las especies huéspedes para todos los genotipos que portan dos o más mutaciones. El cálculo de la epistasia entre pares de mutaciones es sencillo, pero para los genotipos con más de dos mutaciones el cálculo se vuelve un poco más complicado (**apartado 5.3 de Material y Métodos**). Tomando como ejemplo un mutante triple, recordemos que había tres casos diferentes en los que cada mutación individual podía introducirse en los correspondientes genotipos mutantes dobles. Si contamos todos los casos posibles para todos los genotipos, debemos analizar un total de 75 posibilidades diferentes (Lalić y Elena, 2015). La **Figura 35** muestra estos datos e ilustra la existencia de una correlación

significativa entre los coeficientes de epistasia medidos para ambos huéspedes ($r = 0.718$; 73 gl ; $P < 0,001$). Sin embargo, algunos de los casos analizados no se ajustan a la diagonal esperada bajo la hipótesis de que el huésped no produce ningún efecto en la epistasia. La mayoría de estos casos (10) poseían epistasia de magnitud negativa en *N. tabacum* que se transformó en efecto multiplicativo o incluso epistasia de magnitud positiva en *A. thaliana*.

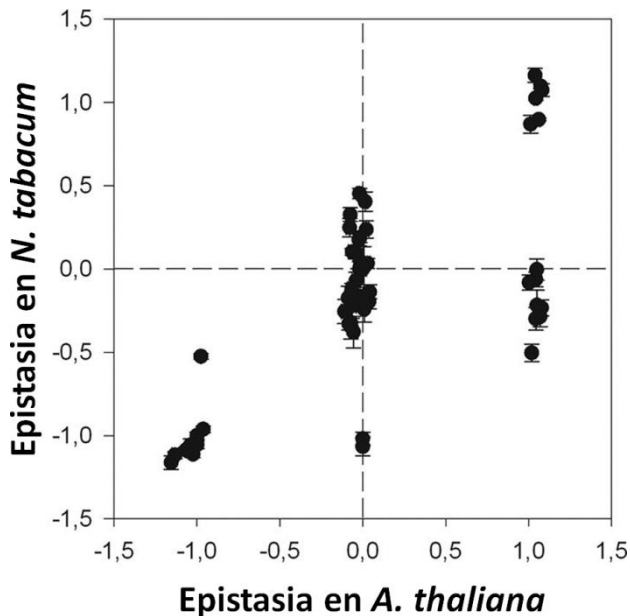


Figura 35. Distribución de las interacciones epistáticas para ambos huéspedes. Las líneas discontinuas corresponden al caso de los efectos multiplicativos de la eficacia (no epistasia). Las barras de error representan $\pm 1 \text{ d.e.}$

Las interacciones genéticas en *A. thaliana* se pueden clasificar de la siguiente manera: 40 casos de interacciones multiplicativas, 26 de epistasia de magnitud, 4 de epistasia de signo y 5 de epistasia de signo recíproco (Lalić y Elena, 2015) (Tabla 10). Para *N. tabacum*, los recuentos son los siguientes: 46 casos de interacciones

multiplicativas, 11 de epistasia de magnitud, 7 de epistasia de signo y 11 de epistasia de signo recíproco. Si analizamos estos números estadísticamente, las distribuciones de recuentos por categoría son significativamente diferentes entre huéspedes ($\chi^2 = 10,05$, 3 gl , $P = 0,008$). La **Tabla 10** muestra el tipo de transición experimentado por cada tipo de interacción epistática entre ambos huéspedes, consecuencia de la adaptación experimental de TEV a *A. thaliana*. La mayoría de las interacciones fueron del mismo tipo en ambos huéspedes (65,3 %; prueba binomial: $P = 0,011$), debido principalmente a la congruencia en el número de casos de interacciones multiplicativas (no epistasia). Curiosamente, si analizamos los casos en que cambiaron los tipos de interacciones genéticas, el 57,5 % redujeron la rugosidad del paisaje en *A. thaliana* (por ejemplo, produciéndose un cambio de epistasia de signo o signo recíproco por epistasia de magnitud).

Tabla 10. Matriz de transición entre casos de interacciones epistáticas en cada huésped. La suma de los valores de las filas corresponde al número total del tipo de interacciones epistáticas descrito para *N. tabacum*. La suma de los valores de las columnas corresponde al número total del tipo de interacciones descritas en *A. thaliana*. El desglose de los números de una fila o columna indica que número de interacciones son compartidos y que número de interacciones cambian entre ambos huéspedes.

| Tipo de epistasia para <i>N. tabacum</i> | Nº de casos para <i>A. thaliana</i> | | | |
|--|-------------------------------------|----------|-------|-----------------|
| | Multiplicativa | Magnitud | Signo | Signo recíproco |
| Multiplicativa | 37 | 8 | 1 | 0 |
| Magnitud | 2 | 9 | 0 | 0 |
| Signo | 1 | 4 | 0 | 2 |
| Signo recíproco | 0 | 5 | 3 | 3 |

Para continuar analizando estos cambios en los patrones de epistasia, nos centraremos en las interacciones por parejas debido a su simplicidad. Cinco combinaciones de dos mutaciones resultaron en una reducción de la rugosidad del paisaje para *A. thaliana* en comparación con la de *N. tabacum*; en tres casos se produjo el cambio de epistasia de signo recíproco por epistasia de magnitud y en un caso de epistasia de signo a multiplicativa. Cabe destacar que cuatro de los cinco casos implicaban la mutación sinónima **P1/U357C** (●○○○○). No podemos ofrecer ninguna explicación sobre por qué el efecto de una mutación sinónima depende tanto de la presencia de mutaciones en otros genes. Otro caso interesante es la mutación no sinónima **6K1/T1126M** (○○●○○). El péptido 6K1 es requerido para la replicación viral y se localiza junto con los elementos de replicación viral unidos al cloroplasto 6K2 y **Nib** en las primeras etapas de la infección (Cui y Wang, 2016). Los efectos en la eficacia viral resultantes de la interacción entre esta mutación particular de 6K1 y las otras cuatro mutaciones fueron siempre dependientes del huésped. Si analizamos esta mutación en combinación con la mutación sinónima **P1/U357C** o la mutación no sinónima **P3/A999V** (○●○○○), observamos que las interacciones genéticas cambiaron de epistasia de signo en *N. tabacum* a interacciones multiplicativas o de magnitud en *A. thaliana*. Sin embargo, cuando hallamos esta mutación junto con la mutación no sinónima **VPg/L1965F** (○○○●○) o junto con la mutación sinónima **NlaPro/C6906U** (○○○○●), aumenta la rugosidad del paisaje en el nuevo huésped dándose el cambio de interacciones multiplicativas por epistasia de signo o de magnitud, respectivamente. Una vez más, no podemos explicar por qué el efecto de esta mutación depende de mutaciones sinónimas en otros genes. El efecto de la mutación **6K1/T1126M** sobre la eficacia viral es ligeramente diferente en presencia de la mutación **P3/A999V** en el nuevo huésped, sugiriendo alguna forma de interacción entre estos dos genes (directa o indirecta) que aún no se ha probado experimentalmente (Elena y Rodrigo, 2012). El aumento de eficacia promovido por la mutación **6K1/T1126M** es mayor en presencia de la mutación **VPg/L1965F** en *A. thaliana*, sugiriendo también que

podieran ejercer funciones relacionadas con la determinación de la eficacia de TEV en el nuevo huésped. La interacción entre estas dos proteínas tampoco se ha confirmado experimentalmente (Elena y Rodrigo, 2012). Sin embargo, en ambos casos, pudiera existir una interacción indirecta mediada por la proteína CI (Elena y Rodrigo, 2012).

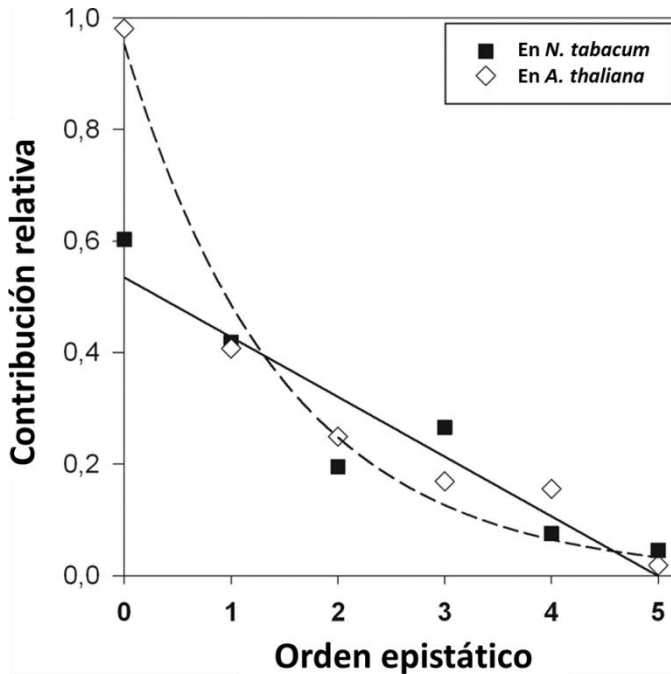


Figura 36. Contribución relativa de epistasis de orden creciente a la variabilidad observada en eficacia en ambos huéspedes. Los valores de la eficacia fueron normalizados para hacer que la eficacia media sea igual a 1 (coeficiente de Walsh de orden cero) en el nuevo huésped. Los coeficientes de primer orden son análogos a los coeficientes de selección para mutaciones únicas y los de segundo orden son análogos a los coeficientes de epistasia para pares de mutaciones. Los términos de orden superior son equivalentes a la epistasia entre un número creciente de mutaciones. Los coeficientes de Walsh se calcularon con el servidor web MAGELLAN.

Hasta ahora, nos hemos centrado en las interacciones entre parejas de mutaciones individuales o entre una mutación y un grupo de mutaciones. [Weinreich et al. \(2013\)](#) advirtieron que este enfoque pudiera ser engañoso, ya que los productos de muchos genes interactúan de muy diversas formas para determinar la eficacia de los individuos. Así, la epistasia de orden superior debe ser un componente fundamental de la arquitectura genética de la eficacia. Utilizando el método de los coeficientes de Walsh ([Weinreich et al., 2013](#)), se evaluó la contribución de la epistasia de orden superior en los dos paisajes. La **Figura 36** compara el peso de cada coeficiente de Walsh con la variabilidad en la eficacia observada en ambos paisajes. Los coeficientes de orden cero representan la eficacia media en todos los genotipos. En este caso, la eficacia media es mayor en el nuevo huésped que en el ancestral. Esto es lógico, ya que estos genotipos fueron seleccionados positivamente en *A. thaliana* (al menos aquellos que no son letales y pueden tener una existencia real en la población). Los coeficientes de primer orden corresponden a coeficientes de selección para mutaciones únicas. En ambos paisajes, las interacciones de hasta cuatro órdenes contribuyen de manera notable al patrón observado de eficacia, lo que ilustra la complejidad de las interacciones entre los genes para determinar la eficacia de TEV en ambos huéspedes. Curiosamente, las interacciones de segundo orden, que corresponden a coeficientes de epistasia para pares de mutaciones, parecen ser cualitativamente más importantes en *A. thaliana*, mientras que las interacciones de tercer orden, que representan el efecto de una mutación dada sobre la superficie definida por otras dos mutaciones, parecen ser más importantes en *N. tabacum*. La diferencia en la importancia de las interacciones de segundo y tercer orden en ambos huéspedes apoya aún más la idea de que el paisaje es menos rugoso en el nuevo huésped que en el ancestral. Los coeficientes de cuatro órdenes reflejan el efecto que una superficie definida por un par de mutaciones ejerce sobre la superficie definida por otro par de mutaciones. Las interacciones de cuatro órdenes también parecen ser cualitativamente más importantes en *A. thaliana* que en *N. tabacum*. Por último, los

coeficientes de quinto orden reflejan el efecto de un par de mutaciones sobre otra superficie definida por tres mutaciones. En este caso, al igual que los coeficientes de primer orden, no parecen contribuir en gran medida al patrón de eficacia observado en ambos paisajes.

3. Relación entre la pleiotropía antagonista y las interacciones epistáticas.

La pleiotropía y la epistasia tienen fuertes paralelismos porque el efecto pleiotrópico de un alelo depende más de la especie huésped y el efecto epistático de ese alelo depende más del fondo genético viral. De hecho, se ha postulado que la pleiotropía es un requisito previo para la epistasia (De Visser et al., 2011; Martin et al., 2007). Esta dependencia es obvia para el caso de pleiotropía antagonista, donde mutaciones con efecto positivo en el nuevo huésped tienen un efecto negativo en el ancestral (Remold, 2012). Además, en el contexto de la evolución compensatoria, la pleiotropía antagonista es una condición previa para la epistasia de signos, ya que permite compensar los efectos pleiotrópicos negativos de las mutaciones previamente seleccionadas con otros efectos adicionales (De Visser et al., 2011).

Con todo lo visto, es interesante analizar si los 8 genotipos que muestran evidencia de pleiotropía antagonista (Figura 34) también cambiaron el tipo de interacciones epistáticas entre ambos huéspedes. Estos 8 genotipos eran portadores de la mutación no sinónima **6K1/T1126M**. De hecho, todos ellos mostraron dicho cambio: el genotipo $\circ\bullet\bullet\circ\circ$ cambió de epistasia de magnitud negativa a positiva, y los otros 7 genotipos de magnitud positiva a negativa.

Por otro lado, entre los 18 genotipos que no mostraban evidencia de pleiotropía antagonista, 14 no cambiaron el tipo de interacciones epistáticas entre las especies huéspedes, y 4 si lo hicieron (3 de magnitud negativa a positiva y sólo 1 de magnitud positiva a negativa). La prueba exacta de Fisher confirma que los cambios de epistasia se enriquecen significativamente entre los genotipos que muestran pleiotropía antagonista en comparación con los genotipos que no lo muestran ($P < 0,001$).

4. Discusión.

Los resultados descritos anteriormente ilustran claramente que los cambios en las especies huéspedes dan lugar a perturbaciones en la topografía del paisaje de eficacia de un virus vegetal de ARN. Así, el paisaje definido por las cinco mutaciones puntuales que conforman los 32 genotipos fue distinto en ambos huéspedes, siendo más robusto en el original que en el nuevo. Las diferencias entre los dos paisajes, sin embargo, eran locales más que globales, con genotipos concretos cambiando su altura relativa en el paisaje y el tipo de interacciones epistáticas con sus vecinos. Con todo ello concluimos que la topografía del paisaje de eficacia depende del huésped, lo cual apoya la idea de que los paisajes entre huéspedes son dinámicos (Laughlin y Messier, 2015) y con fluctuaciones de los picos de eficacia (Mustonen y Lässig, 2009) en lugar de estáticos. Ello no significa que ambos paisajes no compartan características comunes; por ejemplo, ambos comparten la existencia de agujeros de eficacia debidos a genotipos letales. Las topografías de ambos paisajes empíricos coinciden bastante bien con las expectativas de un paisaje aleatorio no correlacionado, situado entre el caso extremo del modelo de “castillo de naipes” (Aita et al., 2001; Kingman, 1987), en el que la eficacia de cada genotipo es absolutamente independiente de la eficacia de los demás genotipos, y el caso menos radical del modelo puramente multiplicativo o del “Monte Fuji” (Aita et al., 2000; Aita et al., 2001). Los paisajes

empíricos que hemos construido combinan las propiedades tanto del modelo de “castillo de naipes” como del modelo del “Monte Fuji”.

Los efectos pleiotrópicos antagonistas se aceptan como la causa principal de las compensaciones de eficacia entre el huésped nuevo y el original, que conduce generalmente a la especialización del virus en el nuevo huésped (Bedhomme et al., 2015). En nuestro estudio, aproximadamente el 26 % de los genotipos tenían un efecto pleiotrópico (reflejado en su eficacia) y, a excepción de uno (el genotipo ●○○○○), todos ellos eran genotipos beneficiosos en el huésped nuevo pero deletéreo en el original. Estos resultados resaltan la importancia de la pleiotropía antagonista en el proceso de adaptación a un nuevo huésped a costa de reducir la eficacia en el original. Otros autores consideran que la compensación de la eficacia ha sido sobrevalorada como el mecanismo que explica la especialización de un virus a un nuevo huésped (Fry, 1996; Smith-Tsurkan et al., 2010). De hecho, se ha propuesto que los paisajes de eficacia incongruentes pueden ser una mejor explicación para la evolución de virus especialistas que infectan huéspedes alternativos (Smith-Tsurkan et al., 2010). Nuestros resultados muestran que estas dos hipótesis pueden reconciliarse, ya que algunos genotipos representan ejemplos claros de pleiotropía antagonista, a la vez que ambos paisajes son incongruentes en algunos detalles particulares. De hecho, ambas hipótesis no son mutuamente excluyentes, ya que la pleiotropía antagonista contribuye en gran medida a la incongruencia entre los paisajes.

Pudiéramos afirmar también que la pleiotropía antagonista en el huésped y la epistasia a nivel genómico en el virus están íntimamente relacionadas, definiendo así una situación conocida como pleiotropía epistática (Remold, 2012). De hecho, en este mismo grupo de investigación, se obtuvieron anteriormente resultados similares al analizar la eficacia y las interacciones epistáticas de un gran conjunto de mutaciones aleatorias en el genoma de TEV (Elena y Lalić, 2013). La pleiotropía epistática tiene dos implicaciones

importantes. En primer lugar, permite la evolución tanto de virus generalistas como especialistas, dependiendo del huésped de la población del virus. En segundo lugar, y muy importante para limitar la aparición de nuevos virus, cuando la epistasia es de signo recíproco, la rugosidad de los paisajes adaptativos disminuye la capacidad de las poblaciones virales para escapar de la especialización hacia una situación de generalismo. La larga historia evolutiva de un virus en el huésped original puede resultar en una adaptación hacia un pico específico de eficacia en dicho huésped que involucra la mayoría de los genes virales, si no todos. Tal población debiera encontrarse muy lejos de alcanzar un pico generalista, a muchos pasos mutacionales de distancia.

En los últimos años, los biólogos evolucionistas hemos comenzado a abordar la topografía de los paisajes de eficacia desde una perspectiva empírica (De Visser y Krug, 2014). Desafortunadamente, la cantidad de información que hemos generado sobre los paisajes de eficacia es todavía muy limitada. Los paisajes de eficacia empíricos se han explorado exhaustivamente y han sido ampliamente estudiados sólo para el VIH-1 (virus de la inmunodeficiencia humana de tipo 1), en que se analizaron las mutaciones que permiten el acoplamiento a receptores alternativos de la superficie celular (Da Silva, 2010; Da Silva et al., 2010; Da Silva y Wyatt, 2014) y la adaptación frente a diferentes fármacos antivirales (Kouyos et al., 2012). Relacionado con este último punto, al igual que ocurre con los resultados que hemos expuesto para TEV, los paisajes para VIH-1 también difirieron en detalles pequeños pero relevantes más que en detalles generales o amplios (Kouyos et al., 2012).

Cabe preguntarse cómo una población viral alcanza el máximo de eficacia global en un paisaje accidentado sin quedarse atrapada en algún pico de eficacia secundario o subóptimo. Nuestros resultados muestran que un cambio en la especie huésped aplanan el paisaje de eficacia, facilitando el acceso a ciertos picos de eficacia que de otro modo podrían haber permanecido inaccesibles en el huésped

original. Existen otros mecanismos posibles para mejorar eficientemente la eficacia en tales paisajes que no necesariamente requieren moverse mediante pasos mutacionales individuales. Estos saltos de largo alcance se conocen como túneles estocásticos en grandes poblaciones (Proulx, 2011). La recombinación es el mecanismo más obvio para que se produzca tal efecto túnel, ya que puede combinar mutaciones beneficiosas en un solo genotipo. Generalmente, los virus de ARN de una única cadena de sentido positivo poseen tasas de recombinación típicamente altas (Simon-Loriere y Holmes, 2011). La tasa de mutación también es alta en este tipo de virus, generalmente una por genoma y ronda de replicación (Sanjuán et al., 2010). Para TEV, la tasa de recombinación está al mismo nivel que la tasa de mutación (Tromas et al., 2014b). Así, la tasa de mutación combinada con la tasa de replicación, ambas muy altas, en tamaños poblacionales grandes, hacen probable que se pueda crear un mutante doble que contenga dos mutaciones beneficiosas, lo que permitiría el efecto túnel. En el caso de TEV, la tasa de mutación genómica (U) es 0,601 (Tromas y Elena, 2010). Sabiendo que el tamaño de TEV es 9.539 nucleótidos y suponiendo que sólo una proporción muy pequeña de todas las posibles mutaciones individuales son beneficiosas, digamos sólo una por genoma, la probabilidad de encontrar un genoma que contenga dos mutaciones beneficiosas sería $U_b^2 = (0,601/9.539)^2 = 3,97 \times 10^{-9}$. Desde la perspectiva evolutiva, el número que importa es el producto NU_b^2 , donde N es el tamaño del censo de la población. Este producto da el número de individuos en la población que son mutantes dobles. Para TEV, N varía enormemente entre los huéspedes, pero en el caso de los ecotipos susceptibles de *A. thaliana*, siempre es mayor que 10^8 y puede ser tan grande como 10^{10} genomas por planta (Hillung et al., 2012), lo que hace que NU_b^2 sea muy probablemente mayor que 1 durante el transcurso de la mayoría de las infecciones (resultará en al menos un genoma con dos mutaciones beneficiosas para una población con $N = 2,52 \times 10^8$ y 40 genomas con dos mutaciones beneficiosas para una población con $N = 10^{10}$).

Antes de terminar cabe considerar dos aspectos adicionales que pudieran plantear un problema en este estudio en particular. El primero de ellos es que *A. thaliana* no es un huésped natural de TEV. No consideramos que sea un problema real, ya que este estudio y todos los anteriores realizados con el mismo sistema experimental planta-virus (Agudelo-Romero et al., 2008a; Lalić y Elena, 2015; Lalić et al., 2010; Hillung et al., 2012; Hillung et al., 2014; Hillung et al., 2015; Hillung et al., 2016) se ocupan de los determinantes evolutivos y las consecuencias de la emergencia y adaptación viral a un huésped completamente nuevo, es decir, buscamos un huésped no natural a propósito para iniciar el experimento. El segundo puede ser un tema a considerar, y es que todos nuestros experimentos se realizaron bajo condiciones de invernadero controladas que son óptimas para la replicación y acumulación del virus. Es bien sabido que *A. thaliana* en su ambiente natural, y otros huéspedes silvestres en general, soportan menos carga viral (menor replicación) que los cultivos o huéspedes cultivados bajo condiciones de invernadero (Pagán et al., 2008). En este sentido, nuestros argumentos anteriores para una exploración eficiente del paisaje basada en el túnel estocástico pueden no funcionar bien en el ambiente natural si se reducen los niveles de replicación. Por lo tanto, generalizar nuestros hallazgos y conclusiones a un contexto ecológico natural no es inmediato.

A pesar de la importancia que tiene la dinámica evolutiva del virus, se sabe muy poco sobre la topografía de los paisajes adaptativos en relación a su eficacia, incluso sobre los efectos que pueden tener diferentes especies huéspedes. En resumidas cuentas, con todo este trabajo podemos constatar que ambos paisajes comparten algunas propiedades macroscópicas, como la existencia de genotipos letales y una topografía accidentada, pero difieren en detalles microscópicos principalmente debido a la existencia de muchos cambios de epistasia de magnitud o signo.

Como consideración final, la recopilación de información sobre la estructura y la topología de los paisajes adaptativos de virus de ARN, su dependencia de factores externos y sobre cómo modulan la evolución del virus puede ser fundamental para desarrollar nuevas estrategias antivirales y tratamientos clínicos personalizados y predecir y contener enfermedades emergentes.

Capítulo 3.

Cambios en la eficacia debidos a la variación genética dentro de una población clonal de un virus de ARN.

Los resultados que se detallan en este capítulo se obtuvieron a partir de una población clonal de TEV, generada inoculando una única planta de tabaco, y a partir de 164 subclones biológicos individuales de esta población clonal aislados mediante un ensayo de formación de lesiones en hojas de *C. quinoa*. Para evaluar la distribución de los valores de eficacia relativa para subclones individuales y muestras aleatorias de una población clonal original a distintos dpi (Figura 37), ajustamos un modelo estadístico a los datos de eficacia y procedimos al análisis de los componentes de varianza correspondientes. La Tabla 11 resume los resultados del ajuste del modelo y las pruebas de significancia para todos los factores involucrados en el modelo.

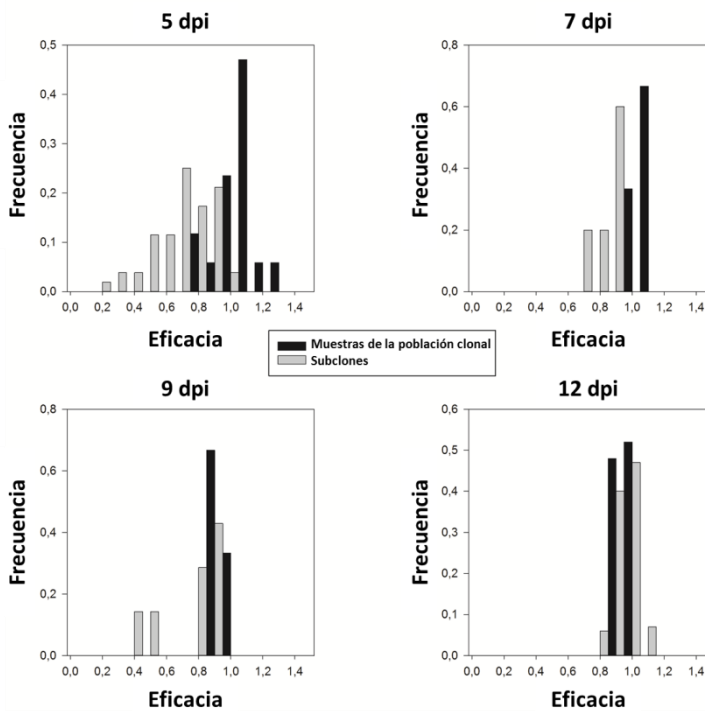


Figura 37. Distribución de valores de eficacia relativa en diferentes tiempos de muestreo (dpi) para subclones individuales (barras grises) y muestras aleatorias de la población clonal original (barras negras).

Tabla 11. Modelo general linearizado de los datos de eficacia relativa.

| Factor | LRT ^a | g.l. | p-valor | η_p^2 ^b | $1 - \beta$ ^c |
|-----------------------------|------------------|------|---------|-------------------------|--------------------------|
| Interceptación (μ) | 1.218,534 | 1 | < 0,001 | 0,872 | 0,163 |
| Tipo de muestra (S) | 386,746 | 1 | < 0,001 | 0,340 | 1 |
| dpi (T) | 275,726 | 1 | < 0,001 | 0,352 | 1 |
| Pendientes ($S \times T$) | 285,508 | 1 | < 0,001 | 0,362 | 1 |
| Réplicas ($R(S)$) | 626,917 | 138 | < 0,001 | 0,627 | 1 |

^a Estimadores de verosimilitud.

^b η_p^2 mide la magnitud del efecto asociado a cada factor.

^c Potencia estadística de la prueba realizada.

En primer lugar, el análisis muestra que existen diferencias significativas entre la eficacia de los componentes subclonales aislados individualmente y la eficacia de la población clonal en su conjunto (factor **Tipo de muestra (S)** en la **Tabla 11**). De hecho, la eficacia relativa media para los subclones es $0,870 \pm 0,003$, mientras que la eficacia relativa media estimada para la población clonal es $1,001 \pm 0,006$. Este resultado sugiere que la eficacia de toda la población clonal no puede predecirse simplemente haciendo un promedio de la eficacia de los subclones individuales que la componen.

En segundo lugar, existe un efecto temporal significativo (**dpi (T)** en la **Tabla 11**) y, lo que es más interesante, el tiempo afecta a la magnitud de las diferencias entre subclones y la población clonal de la siguiente forma (**Pendientes ($S \times T$)** en la **Tabla 11**): las estimaciones de eficacia para la población clonal no cambian con el tiempo, mientras que la eficacia de los subclones aumenta con el número de dpi. Este punto se expondrá con más detalle en las secciones siguientes.

En tercer lugar, existen diferencias significativas entre la eficacia relativa de las réplicas biológicas dentro de cada grupo

(Réplicas ($R(S)$) en la **Tabla 11**), esto es, entre subclones y/o entre muestras aleatorias tomadas de toda la población clonal. Llegados a este punto, cabe preguntarse qué causa estas diferencias. Para tratar dicha cuestión, calculamos la varianza para el componente genético y la eficacia relativa entre los subclones y la población clonal a diferentes dpi. En el caso de los subclones, la contribución genética a las diferencias observadas en la eficacia oscila entre $(3,240 \pm 0,051) \times 10^{-2}$ a 5 dpi y $(3,086 \pm 0,031) \times 10^{-3}$ a 12 dpi. Dicho resultado sugiere que las diferencias genéticas entre los subclones son enormes poco después de la inoculación, pero dichas diferencias se difuminan a medida que se acumulan más variantes genéticas que tienden a ser homogéneas en términos de eficacia (un orden de magnitud menos diverso). En otras palabras, los subclones tienden a recuperar la eficacia a medida que se replican dado que algunas de las variantes genéticas desplazan a las demás. En contraste, la contribución genética a las diferencias de eficacia observadas entre muestras independientes de toda la población clonal se sitúa entre 0 (a 5 y 7 dpi) y $(4,090 \pm 0,178) \times 10^{-4}$ a 12 dpi. Como se esperaba, las muestras aleatorias de una misma población clonal eran realmente homogéneas tras la inoculación, aunque comienzan a divergir lentamente, ya que acumulan algunas variantes genéticas de manera independiente durante el progreso de la infección (debido únicamente a errores propios de la replicación).

1. La eficacia promedio de los subclones no predice la eficacia de toda la población clonal de forma temprana.

La **Figura 38** ilustra el efecto de la duración de la infección sobre las diferencias en la eficacia relativa entre el subclón medio y toda la población clonal. A 5 dpi, la distribución de los valores de eficacia relativa estimados a partir de la población clonal completa es simétrica ($g_1 = -0,337 \pm 0,550$; $t_{16} = 0,613$; $P = 0,548$) y

mesocúrtica ($g_2 = 1,360 \pm 1,063$; $t_{16} = 1,279$; $P = 0,220$), tal y como se esperaba para una distribución Normal. La distribución se centra alrededor de un valor de eficacia relativa media de $1,036 \pm 0,030$. Por el contrario, la distribución de los valores de eficacia entre los subclones es moderadamente sesgada ($g_1 = -0,778 \pm 0,330$; $t_{51} = 2,358$; $P = 0,022$), es decir, la eficacia de la mayoría de los subclones está por debajo del valor medio. Sin embargo, la distribución sigue siendo mesocúrtica ($g_2 = 0,252 \pm 0,650$; $t_{51} = 0,388$; $P = 0,700$), lo que indica que la mayoría de los valores se encuentran cerca del centro de la distribución en lugar de en las colas. La eficacia relativa media de un subclón elegido al azar es de $0,689 \pm 0,025$, por lo tanto, la eficacia promedio de los subclones a 5 dpi es un 33,49 % menor que la eficacia esperada en la población clonal después del mismo número de dpi. Todo ello representa una diferencia altamente significativa en la centralidad (prueba de Mann-Whitney: $U = 38$, $P < 0,001$) y la forma (prueba de Kolmogorov-Smirnov para dos muestras: $D = 0,825$, $P < 0,001$) entre ambas distribuciones. La diferencia observada en la eficacia media de la población clonal ($W = 1,036$) y los subclones ($W = 0,689$) indica que las mutaciones presentes en los subclones individuales tienden a ser deletéreas, de ahí que observemos un descenso global en la eficacia.

A continuación, repetimos el mismo análisis en intervalos de tiempo más amplios (7 y 9 dpi), aunque contamos con un conjunto muestral más pequeño. Aplicando el mismo modelo contrastamos que las diferencias entre los subclones y la población clonal aún eran significativas a 7 y 9 dpi. En el primer caso, 7 dpi, la distribución de la eficacia entre los subclones tiene una media de $0,886 \pm 0,034$, que es significativamente menor que la eficacia promedio estimada para muestras independientes de la población clonal, $1,000 \pm 0,002$. Las diferencias también se observan en los parámetros de centralidad (prueba de Mann-Whitney: $U = 0$, una cola $P = 0,018$) y forma (prueba de Kolmogorov-Smirnov: $D = 1$, $P = 0,047$). A los 9 dpi, ambas poblaciones siguen siendo significativamente diferentes, con

valores medios de eficacia de $0,794 \pm 0,077$ para los subclones y $1,000 \pm 0,006$ para la población clonal, incluyendo diferencias en centralidad (prueba de Mann-Whitney: $U = 0$, una cola $P = 0,009$) y forma (prueba Kolmogorov-Smirnov: $D = 1$, $P = 0,030$).

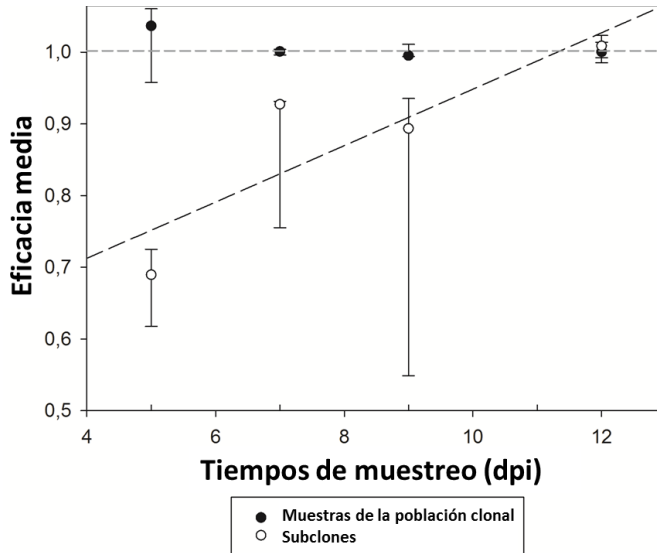


Figura 38. Evolución de la media de la eficacia para subclones y muestras tomadas de toda la población clonal. Las barras de error representan intervalos de confianza del 95% de la media y se obtuvieron usando el método de remuestreo “bootstrap” (1.000 pseudoréplicas). Las líneas discontinuas representan la pendiente en ambos grupos.

2. Las diferencias en la eficacia entre los subclones y la población clonal desaparecen después de largos períodos de replicación.

Como se muestra en la **Tabla 11** y comentamos anteriormente, la duración de la infección tuvo un efecto significativo en la eficacia relativa para la población clonal y los subclones individuales. La

Figura 38 muestra claramente que las diferencias observadas después de la inoculación desaparecen a los 12 dpi, cuando se ha alcanzado la fase estacionaria de replicación viral. En este punto, la distribución de los valores de eficacia relativa para muestras independientes tomadas de la población clonal sigue siendo simétrica ($g_1 = 0,079 \pm 0,464$; $t_{24} = 0,170$; $P = 0,866$) y mesocúrtica ($g_2 = -0,862 \pm 0,902$; $t_{24} = 0,956$; $P = 0,349$), es decir, continúa siendo una distribución Normal. Dicha distribución se centra alrededor de un valor medio de $1,000 \pm 0,005$. Respecto a la población de subclones, la distribución de los valores de eficacia relativa también fue simétrica ($g_1 = -0,126 \pm 0,241$; $t_{99} = 0,523$; $P = 0,602$) y mesocúrtica ($g_2 = -0,260 \pm 0,478$; $t_{99} = 0,544$; $P = 0,588$), con un valor medio de $1,009 \pm 0,006$. No se encontraron diferencias significativas entre ambas distribuciones ni en la centralidad (prueba de Mann Whitney: $U = 1,089$, $P = 0,320$) ni en la forma (prueba de Kolmogorov-Smirnov para dos muestras: $D = 0,300$, $P = 0,055$). De todo ello, podemos afirmar que los subclones, inicialmente diversos, han compensado las mutaciones deletéreas que pudiesen portar para converger hacia un valor de eficacia relativa similar al de la media de la población clonal de la que fueron aislados.

3. Discusión.

Hace dos décadas, [Duarte et al. \(1994\)](#) describieron por primera vez la gran heterogeneidad fenotípica entre los componentes subclonales de una población del virus de la estomatitis vesicular (VSV). La variabilidad encontrada entre los genomas individuales, los cuales se esperaba que fueran clones, fue una consecuencia inevitable de la replicación de los virus de ARN, propensos a errores con tasas de mutación que normalmente están dentro del rango de 0,1-1 mutaciones por genoma replicado ([Sanjuán](#)

et al., 2010). Además, la gran mayoría de las mutaciones aleatorias tuvieron un efecto negativo en la eficacia viral y, debido a ello, concluyeron que la eficacia media de los componentes subclonales era significativamente menor que la eficacia de la población contemplada en su conjunto.

Anteriormente, se aislaban componentes subclonales en cultivos celulares artificiales (Duarte et al., 1994). En nuestro estudio, hemos aplicado dicha premisa y dichas observaciones del aislamiento de componentes subclonales a un virus de ARN vegetal y, por primera vez, en un sistema in vivo, es decir, reproduciendo una situación biológica realista (la infección de plantas de *C. quinoa* para el aislamiento de los componentes subclonales seguida de la infección de plantas de tabaco, el huésped natural de TEV). Poco después de la infección, nuestros resultados reprodujeron completamente los de Duarte et al. (1994), pudiendo observarse diferencias muy significativas entre los subclones y la población clonal de la que fueron aislados, siendo los subclones menos aptos que la población clonal. Además, también observamos una gran variabilidad genética relacionada con la eficacia, indicativo de diferentes mutaciones que se fijan en cada subclon. También observamos que después de largos períodos de infección las diferencias genéticas entre los subclones desaparecen. De nuevo, la replicación propensa a errores de estos subclones dio lugar a enjambres de mutantes sobre los cuales operó la presión de selección positiva para llevar la eficacia media de estos nuevos enjambres mutantes recién creados de nuevo al mismo valor de eficacia que la población original en el huésped natural. Como discutimos en los capítulos anteriores, la historia evolutiva de TEV en su huésped natural resulta en una adaptación hacia un pico específico de eficacia en dicho huésped conformando una población que, aunque diversa, debiera encontrarse muy lejos de alcanzar un pico generalista, pues la rugosidad del paisaje adaptativo disminuye la capacidad de los subclones biológicos para escapar de la especialización hacia una situación de generalismo.

Para proseguir con las conclusiones de este estudio, primero debemos preguntarnos por el posible efecto de aislar componentes subclonales en un huésped, *C. quinoa*, diferente del huésped natural en el que evaluamos la eficacia, *N. tabacum*. Tanto en el **Capítulo 1** de la presente tesis como en otros trabajos de este mismo grupo de investigación, se ha mostrado que la inoculación de TEV en un nuevo huésped conduce a cambios en su eficacia y, consecuentemente, a la adaptación (Agudelo-Romero et al., 2008b; Bedhomme et al., 2012; Hillung et al., 2014). Sin embargo, para que ello suceda se requieren tamaños de población efectiva muy grandes, de modo que puedan generarse un número determinado de mutaciones beneficiosas para mejorar la eficacia en el nuevo huésped, que éstas sobrevivan a la deriva genética y que aumenten su frecuencia en la población hasta lograr fijarse. Durante los ensayos de lesiones locales en las hojas de *C. quinoa* hemos impuesto cuellos de botella que hacen que la adaptación de TEV a este huésped sea altamente improbable.

Continuando con el apartado anterior, los cuellos de botella son comunes durante la infección de plantas individuales mediada por insectos que actúan como vectores (Moury et al., 2007; Betancourt et al., 2008; Sacristán et al., 2011), durante la propagación célula a célula en la hoja inoculada (Miyashita y Kishino, 2010; Tromas et al., 2014a), durante el movimiento sistémico a través del floema y posterior colonización de los tejidos distales (Hall et al., 2001a; Hall et al., 2001b; Sacristán et al., 2003; French y Stenger, 2005; González-Jara et al., 2009; Ali y Roossinck, 2010; Gutiérrez et al., 2010; Gutiérrez et al., 2012; Gutiérrez et al., 2015; Tromas et al., 2014a) e incluso en semillas (Fabrè et al., 2014). En todos estos casos, los cuellos de botella ejercen un gran efecto y el número de genomas transmitidos varía dentro del rango de unidades o decenas (Zwart y Elena, 2015). Estos fuertes cuellos de botella minimizan la eficacia de la selección para eliminar los alelos deletéreos, que se producen constantemente durante la replicación del ARN genómico, recordemos que es propenso a errores, y dan

como resultado la puesta en marcha del fenómeno conocido como trinquete de Muller. Éste consiste en la pérdida de la clase genotípica menos mutada en cada generación por acción combinada de la mutación y la deriva (ausencia de selección purificadora). El trinquete de Muller ha sido ampliamente descrito operando en poblaciones de virus de ARN bajo las condiciones demográficas adecuadas (Chao, 1990; Duarte et al., 1992; Clarke et al., 1993; Yuste et al., 1999), incluyendo los virus de plantas (De la Iglesia y Elena, 2007). La velocidad a la que opera el trinquete de Muller depende de un proceso de retroalimentación conocido como colapso mutacional (Lynch y Gabriel, 1990; Gabriel et al., 1993; Lynch et al., 1993): cuanto mayor es el número de mutaciones presentes en la población, menor es el número de individuos viables en esa población y, en consecuencia, menor es el tamaño de la población efectiva y más fuerte el cuello de botella. A pesar de la presencia de cuellos de botella durante las infecciones de las plantas, los virus vegetales de ARN no se extinguen. Ello es debido a dos aspectos que ya describimos en el apartado de resultados. En primer lugar, hemos observado que la evolución compensatoria tiene lugar tan pronto como los subclones se replican generando un nuevo enjambre mutante. En segundo lugar, dichos enjambres evolucionan y alcanzan de nuevo, en el huésped original, niveles de eficacia idénticos a la población original. La conclusión es que cuanto más larga sea la duración de la infección, más posibilidades tendrá el enjambre mutante de generar la combinación adecuada de mutaciones compensatorias o revertir las mutaciones deletéreas fijadas durante la transmisión, de modo que se seleccionen los individuos viables rápida y eficientemente, dando como resultado la recuperación de la eficacia dentro del huésped. De hecho, se ha demostrado para el alphavirus chikungunya que la recuperación del efecto deletéreo de las mutaciones fijadas es dependiente de la fidelidad de la replicasa viral RdRp (Coffey et al., 2011). En resumen, los virus de tipo silvestre creaban enjambres mutantes con gran capacidad adaptativa y altamente virulentos, mientras que virus que poseían una RdRp de alta fidelidad no fueron capaces de generar

enjambres mutantes ni de recuperar su eficacia recuperar (produjeron enjambres mutantes más homogéneos) (Coffey et al., 2011).

Debemos hacer dos consideraciones sobre cómo opera el trinquete de Muller en poblaciones virales finitas. En primer lugar, la distribución de las diferencias de eficacia entre los componentes subclonales representa una muestra sesgada de la distribución real de los efectos mutacionales asociados a mutaciones únicas para TEV en su huésped natural (Carrasco et al., 2007b). Aunque las distribuciones reales incorporan una proporción sustancial de mutaciones letales para TEV (Carrasco et al., 2007b) y VSV (Sanjuán et al., 2004a), la muestra generada en este estudio sólo contiene genotipos viables, es decir, aquellos capaces de generar una lesión local visible en hojas de *C. quinoa*. Dicho problema también fue evaluado en Duarte et al. (1994) con VSV. La existencia de alelos letales dentro de un enjambre mutante reduce aún más su tamaño efectivo de población, ya que los genotipos letales no contribuyen a la próxima generación, acelerando así el trinquete de Muller. En segundo lugar, los componentes subclonales pueden contener más de una mutación, y es probable que los enjambres de mutantes recién creados contuvieran genotipos con más de una mutación. Si las mutaciones interactúan epistáticamente, especialmente si lo hacen de manera sinérgica, y los efectos mutacionales son todos idénticos, el trinquete de Muller dejará de operar sobre la población (Kondrashov, 1994). Sin embargo, si los efectos mutacionales siguen alguna distribución continua, el trinquete de Muller funcionará independientemente de la forma en que interactúen dichas mutaciones (Butcher, 1995). Con todo lo explicado anteriormente deducimos que los efectos mutacionales son variables para TEV y VSV, pero para ambos virus las interacciones epistáticas entre mutaciones son en su mayoría de tipo antagónico (epistasia de magnitud negativa) (Sanjuán et al., 2004b; Lalić y Elena, 2012), aspecto que analizamos a fondo en el **Capítulo 2**, manteniendo el funcionamiento del trinquete de Muller.

Debemos hacer una consideración final para evitar malentendidos. Como hemos comentado, se espera que el trinquete de Muller opere sólo en poblaciones virales pequeñas en las que la selección es débil y la deriva desempeña un papel importante. Esto no debe confundirse con la mutagénesis letal, un fenómeno completamente diferente. La mutagénesis letal en las poblaciones virales es un proceso determinista que es independiente del tamaño de la población y sólo depende de una tasa de mutación excesivamente alta y de un número muy pequeño de réplicas viables por virión parental (Bull et al., 2007).

Capítulo 4.

Factores del huésped natural *N. tabacum* que determinan la eficacia de TEV.

Antes de comenzar a describir en detalle los experimentos incluidos en este capítulo, y puesto que vamos a utilizar una colección de mutantes distinta a las de los capítulos anteriores, vamos a detenernos un momento en comentar en más detalle las diferencias en la eficacia y la sintomatología de los seis genotipos mutantes aquí empleados. La **Tabla 3 (apartado 1.3.2 de Material y Métodos)** muestra la información más relevante de dichos genotipos y la **Figura 39** ilustra su situación a lo largo del genoma de TEV. Todos ellos difieren del genotipo silvestre en una mutación no sinónima, a excepción del mutante **HC-Pro:AS13**, que posee dos mutaciones no sinónimas que resultan en cambios del correspondiente amino ácido a alanina.

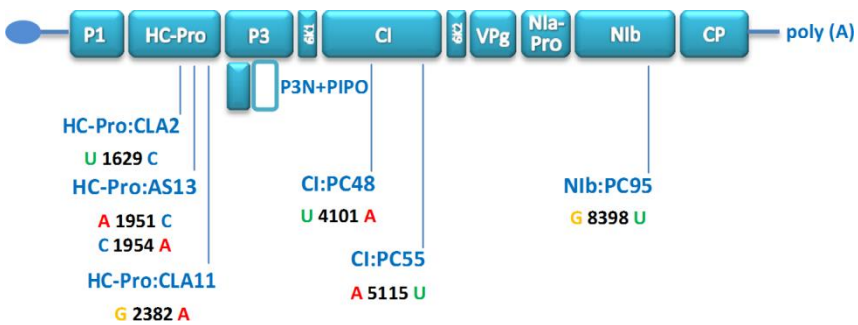


Figura 39. Representación del genoma TEV en la que se indica la localización de las mutaciones no sinónimas de los mutantes empleados.

Cabe destacar los tres genotipos con los valores de eficacia más bajos (**HC-Pro:AS13**, **HC-Pro:CLA2** y **HC-Pro:CLA11**), puesto que contienen mutaciones en la proteína multifuncional **HC-Pro**, cuyo papel más relevante es la supresión del silenciamiento génico de la planta (Revers y García, 2005). Los dos genotipos con mutaciones en la proteína **CI** (**CI:PC55** y **CI:PC48**), helicasa implicada en el movimiento del virus de célula a célula (Revers y García, 2005), poseen un valor de eficacia similar al del genotipo silvestre, por lo

que los efectos deletéreos de estas mutaciones pueden considerarse leves. El mutante **Nib:PC95**, con una mutación en la replicasa **Nib** (Revers y García, 2005), posee un valor de eficacia intermedio (Figura 40.A).

A pesar de que la mayoría de los casos implican una única mutación, los valores de eficacia y la gravedad de los síntomas inducidos son muy variados (Figura 40.A).

Encontramos diferencias significativas entre los valores de eficacia de los siete genotipos virales ($\chi^2 = 373,006$, 6 gl , $P < 0,001$) y entre las plantas inoculadas con un mismo genotipo viral ($\chi^2 = 2927,885$, 14 gl , $P < 0,001$). Asimismo, el porcentaje de la varianza total en la eficacia explicado por diferencias genéticas entre los genotipos es del 70,56 %, mientras que las diferencias entre plantas inoculadas con el mismo genotipo viral explican el 29,16 % de la varianza. Sólo el 0,28 % restante de la variabilidad observada es debida a mediciones erróneas.

En relación a la descripción de la sintomatología (Figura 40.B), todos los genotipos indujeron un patrón de grabado severo muy similar al inducido por el genotipo silvestre, a excepción de **HC-Pro:AS13**. En lo que respecta a este mutante, las tres plantas que se analizaron en este estudio mostraron un número reducido de pequeñas manchas cloróticas locales, aunque en otros experimentos llevados a cabo por este mismo grupo de investigación se ha descrito como fenotipo asintomático (Torres-Barceló et al., 2010).

Si pretendemos valorar conjuntamente el valor de eficacia y la sintomatología inducida, no podremos definir una correlación entre la eficacia y los síntomas del virus, una situación ya descrita previamente para este sistema experimental (Carrasco et al., 2007b).

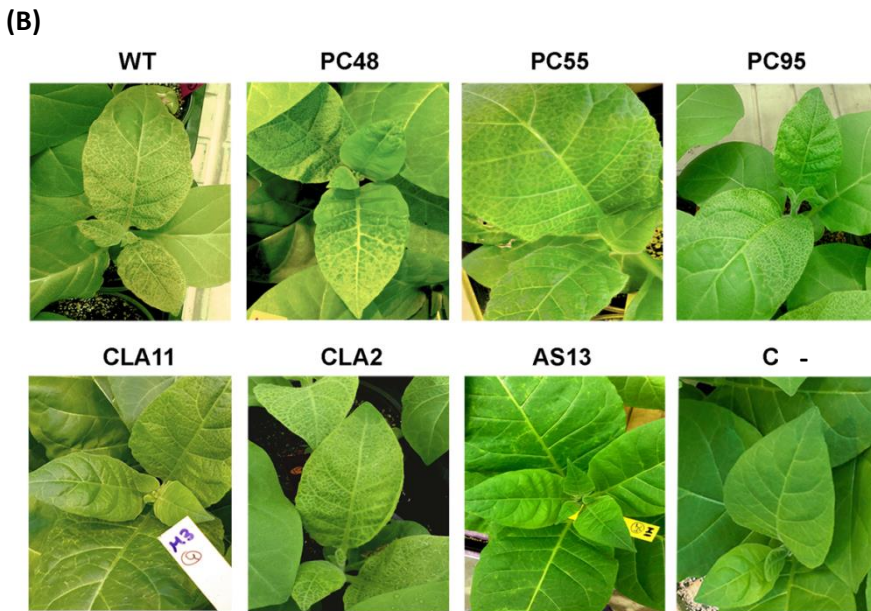
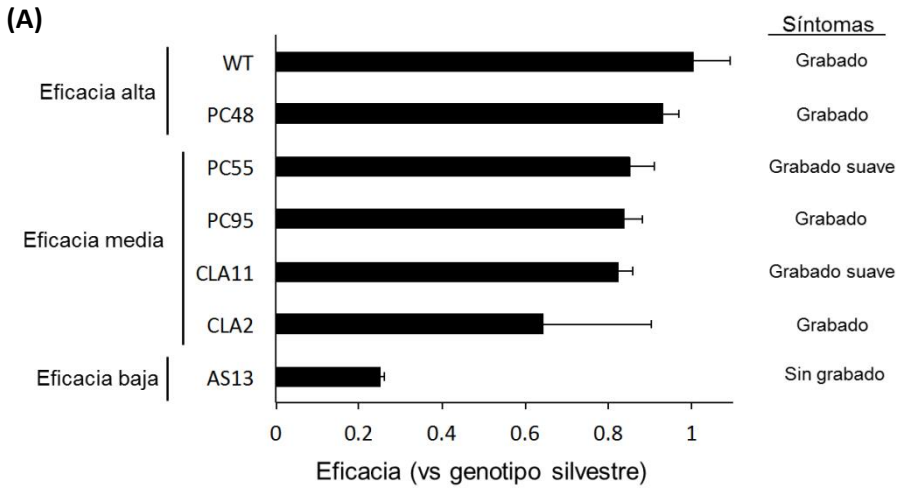


Figura 40. (A) Valores de eficacia estimados para cada uno de los genotipos de TEV en el huésped natural *N. tabacum*. Las barras de error representan $\pm 1 d.e.$ (B) Detalle de la sintomatología inducida por los siete genotipos de TEV. El panel etiquetado como C- corresponde con una planta sana.

1. Las diferencias en la eficacia están asociadas a diferencias en los perfiles transcriptómicos del huésped.

En primer lugar, hemos tratado de averiguar si las diferencias en la eficacia viral de los distintos mutantes se refleja en los perfiles de expresión de los genes de la planta huésped. Al infectar una planta de tabaco con el genotipo silvestre de TEV, cuya eficacia es $W = 1,004 \pm 0,003$, esperamos que la interacción entre los componentes del virus y los factores del huésped sea óptima (ejerciendo la regulación siempre a favor del genotipo silvestre de TEV, que se haya completamente adaptado a *N. tabacum*). Al infectar otras plantas de tabaco con mutantes de TEV, cuya eficacia se va reduciendo paulatinamente, esperamos que dicha interacción sea menos óptima y ello se refleje en perfiles de expresión génica diferentes. Incluso, esperaríamos que la eficacia decreciente de los mutantes se correlacionara con mayores diferencias entre los perfiles de expresión génica en el huésped.

De forma breve, tal y como detallamos en el **apartado 2.1** y siguientes de **Material y Métodos**, se infectaron plantas de tabaco con cada uno de los siete genotipos. A los 8 dpi, las hojas y los ápices de cada planta se recogieron individualmente, a excepción de las plantas infectadas con **HC-Pro:AS13**, que se recogieron a los 15 dpi. Tras ello, se extrajo el ARN total de las muestras, se evaluó su calidad, se ajustó su concentración y, finalmente, estos ARNs se hibridaron a micromatrices de Agilent. Una vez extraídas las imágenes, los valores de intensidad se normalizaron utilizando las herramientas implementadas en Babelomics 5.0 (Alonso et al., 2015).

La **Figura 41** muestra el agrupamiento de los datos de expresión para los genes que cambiaron significativamente respecto a la planta control sana no infectada (método de la distancia promedio no ponderada, UPGMA; $\pm 50\%$ expresión). Se compararon las plantas infectadas con los siete genotipos virales con las plantas sanas (prueba ANOVA; $P < 0,05$). Sólo se incluyen los genes que muestran diferencias significativas entre todas las plantas analizadas.

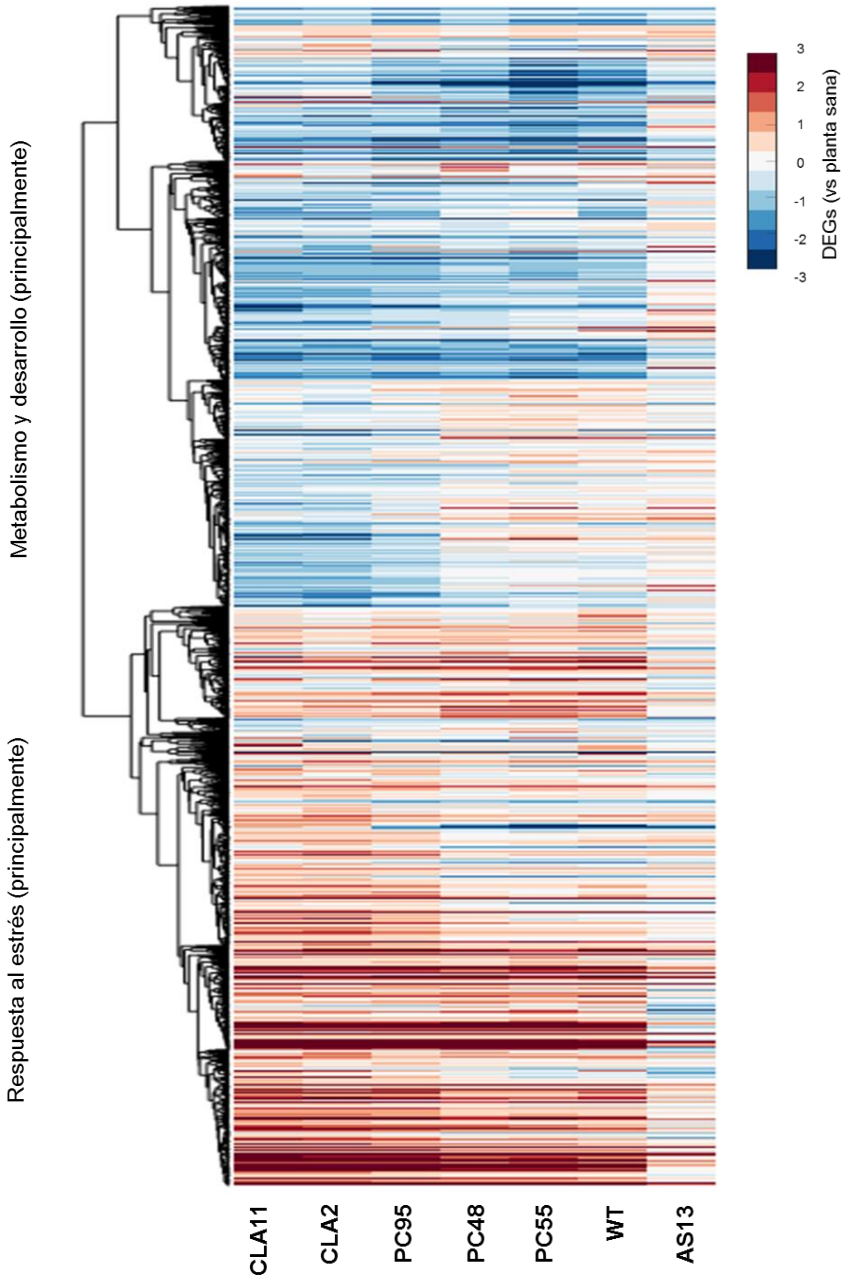


Figura 41. Representación del agrupamiento de los valores de expresión de los genes (de las plantas infectadas comparadas con las plantas sanas) para todos los genotipos virales.

Resultados y Discusión

Podemos distinguir dos grandes grupos de genes. En azul se muestran los genes cuya expresión ha disminuido $\geq 50\%$, es decir, genes reprimidos, principalmente relacionados con procesos metabólicos y de desarrollo. En rojo se muestran los genes cuya expresión ha aumentado $\geq 50\%$, es decir, genes sobreexpresados, principalmente genes relacionados con respuestas de estrés.

El siguiente paso fue tratar de determinar el grado de similitud de los perfiles transcriptómicos de las plantas. Para ello, calculamos todos los coeficientes de correlación de Pearson para cada par de muestras empleando el valor medio de expresión de todos los genes de la micromatriz. Dichas correlaciones se utilizaron como medida de similitud para construir un dendrograma (Figura 42).

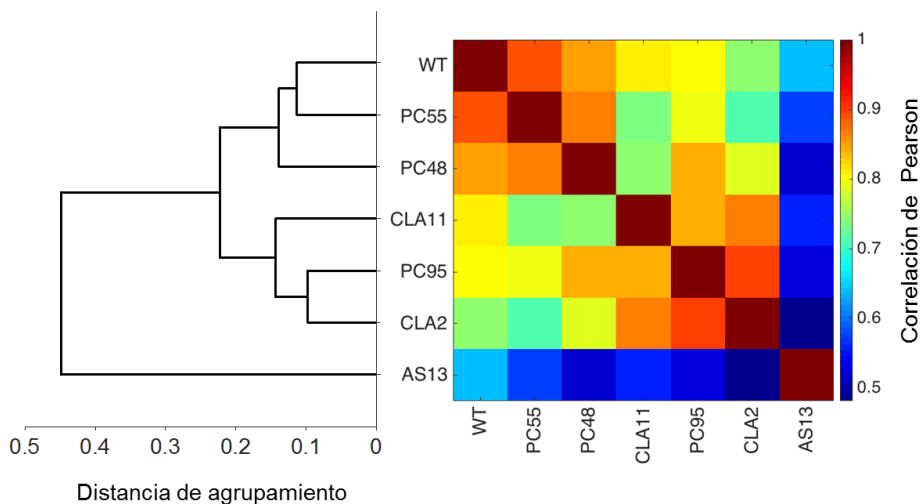


Figura 42. Dendrograma que agrupa los genotipos de TEV en función del grado de similitud de los perfiles de expresión de las plantas infectadas con cada uno de ellos.

El color en la **Figura 42** representa el valor del coeficiente de correlación de Pearson entre el par de perfiles comparados. Obviamente, cuando comparamos una muestra consigo misma el valor de correlación es 1, que determina una correlación perfecta (diagonal representada en color rojo). Para el resto de casos, los perfiles de expresión más correlacionados se destacan en color amarillo o naranja. Ello indica que los dos perfiles transcricómicos son muy similares, es decir, el efecto inducido por los dos genotipos virales afecta a la expresión de un grupo concreto de DEGs respecto a las plantas sanas. En caso contrario, el coeficiente de correlación será muy pequeño y se indicará en color azul. Ello indica que los genes con expresión diferencial no coinciden en las dos muestras que se comparan.

Atendiendo al grado de similitud entre pares de muestras podemos distinguir tres grupos claramente diferenciados en el dendrograma de la **Figura 42**. El primer grupo está constituido por los tres genotipos virales con los valores de eficacia más altos (**WT**, **CI:PC55** y **CI:PC48**). Los genotipos de **HC-Pro:CLA11**, **HC-Pro:CLA2** y **Nib:PC95**, que pudiéramos calificar como genotipos de eficacia intermedia, conforman un segundo grupo. Finalmente, las plantas infectadas con el genotipo **HC-Pro:AS13** constituyen el tercer grupo, ya que muestran un perfil de expresión génica muy diferente al del resto de las muestras. Dentro de los grupos, el valor del coeficiente de correlación fue de $r > 0,85$, mientras que entre los grupos los valores de la correlación fueron algo menores, $0,65 < r < 0,75$. El valor del coeficiente de correlación entre plantas infectadas con el genotipo **HC-Pro:AS13** con el resto de genotipos siempre fue $r < 0,65$.

En el dendrograma se posicionan los genotipos virales ordenados de acuerdo con estos valores de correlación, lo cual coincide con el orden decreciente de los valores de eficacia, a excepción de los genotipos **HC-Pro:CLA11** y **Nib:PC95** (que intercambian posiciones).

Podemos realizar otros tipos de análisis para continuar estudiando las similitudes entre los perfiles de expresión, por ejemplo, un análisis de componentes principales (PCA) a partir de todos los datos de expresión génica.

El PCA es un método estadístico utilizado para reducir las dimensiones de un conjunto de datos. Técnicamente, el PCA busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. En otras palabras, convierte un conjunto de observaciones de variables correlacionadas en un conjunto de valores de variables sin correlación lineal (ortogonales) llamadas componentes principales.

Una vez introducidos todos los datos de expresión génica y realizado el análisis, el porcentaje de la varianza total observada explicada por el primer componente principal ($PC1$) es del 81 %. En el espacio tridimensional definido por los tres primeros componentes principales ($PC1$, $PC2$ y $PC3$) el porcentaje aumenta hasta el 93 %. La **Figura 43** muestra la distribución de los perfiles de expresión génica en el espacio definido por los tres primeros componentes principales. Los resultados del PCA son equivalentes a los obtenidos con los dos métodos de agrupamiento anteriores, clasificándose los genotipos en tres grupos claramente diferenciados. En primer lugar, los genotipos **WT**, **CI:PC48** y **CI:PC55** se sitúan muy cercanos entre ellos y se caracterizan por poseer valores positivos para $PC1$ y valores negativos para $PC2$ y $PC3$. Estos tres genotipos, que poseen los valores de eficacia más altos, conforman un primer grupo. Seguidamente, hallamos los genotipos **Nib:PC95**, **HC-Pro:CLA11** y **HC-Pro:CLA2**, en el espacio definido por valores positivos para $PC1$ y $PC2$ y valores negativos para $PC3$. Estos genotipos, cuya eficacia hemos calificado de intermedia, forman el segundo grupo. Por último, tal y como ocurría en las clasificaciones anteriores, los datos de expresión génica del genotipo **HC-Pro:AS13** son muy diferentes respecto al resto de genotipos. Ello conlleva una vez más la clasificación de **HC-Pro:AS13** como único miembro de un grupo

aislado (caracterizado por un valor de eficacia muy bajo). Así, en el espacio tridimensional, **HC-Pro:AS13** posee valores negativos para *PC1* y *PC2* y valores positivos para *PC3*. Si observamos la **Figura 43** con detenimiento, nos percatamos de que los genotipos se sitúan en el espacio definido por los tres PCs siguiendo una trayectoria de aumento de los valores de eficacia. A lo largo de dicha trayectoria, los valores para *PC1* cambian de negativos a positivos, mientras que los valores para *PC3* experimentan el cambio en sentido contrario. Esta trayectoria sugiere que la sobreexpresión o represión de un conjunto de genes dados se asocia con grandes diferencias en la eficacia viral.

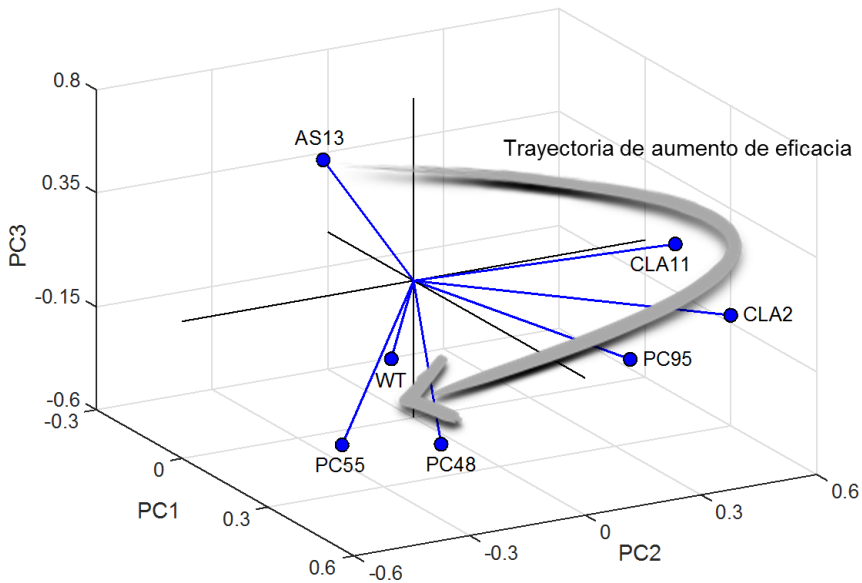


Figura 43. Representación de los tres componentes principales resultado de un PCA usando los datos mostrados en el dendrograma de similitud de los perfiles de expresión de las plantas infectadas. Los tres primeros componentes principales destacados en el análisis explican hasta el 93% de la varianza total observada. Las líneas azules enlazan cada genotipo con el centro del espacio tridimensional. La flecha gris representa una trayectoria putativa de aumento de la eficacia viral en el espacio de los PCs.

Siguiendo con nuestra hipótesis de trabajo, si el genotipo silvestre de TEV ha evolucionado para optimizar su interacción con los genes del huésped, es lógico pensar que pequeñas diferencias en la eficacia viral pudieran corresponderse con pequeñas diferencias entre los transcriptomas. De este modo, esperaríamos que la eficacia y los transcriptomas de plantas infectadas con genotipos cuya eficacia sea elevada sean muy similares. Asimismo, grandes diferencias en la eficacia conllevarían enormes diferencias a nivel transcriptómico. Para finalizar con este apartado de resultados, hemos procedido a calcular la correlación entre el grado de similitud de los perfiles y la eficacia (**Figura 44.A**: $r = 0,826$, 5 gl , $P = 0,005$). También calculamos la correlación entre la distancia de agrupación de los genotipos y la eficacia (**Figura 44.B**: $r = -0,857$, 5 gl , $P = 0,003$). Ambas correlaciones resultaron significativas. Los resultados de ambos análisis se muestran en la **Figura 44**.

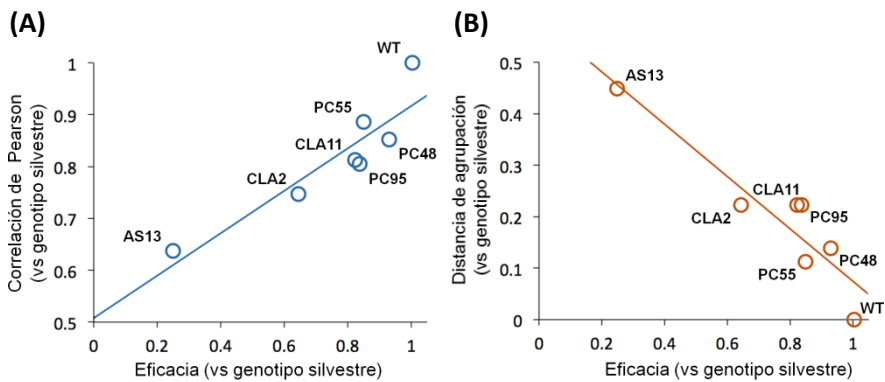


Figura 44. (A) Correlación entre el grado de similitud de los perfiles de expresión génica de todos los genotipos y la eficacia relativa respecto al genotipo silvestre. (B) Correlación entre la distancia de agrupación de todos los genotipos y la eficacia relativa respecto al genotipo silvestre.

2. El número y categoría funcional de los genes dependen de la eficacia viral.

La **Figura 45** muestra el número de genes expresados diferencialmente (DEGs), tanto reprimidos como sobreexpresados, en relación con el transcriptoma de plantas sanas. El número de DEGs reprimidos oscila entre 531 (para **HC-Pro:AS13**) y 2.809 (para **HC-Pro:CLA11**), mientras que el número de DEGs sobreexpresados oscila entre 781 (para **HC-Pro:AS13**) y 2.696 (para **HC-Pro:CLA11**).

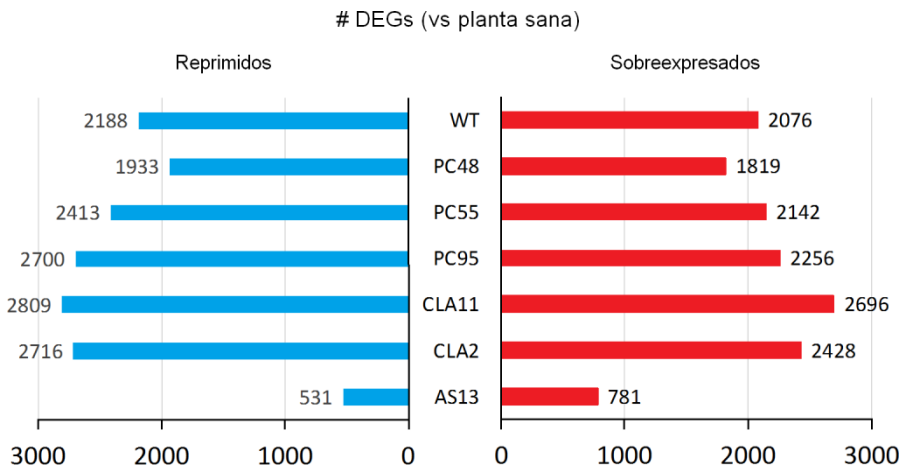


Figura 45. Resultado del análisis de expresión diferencial para cada genotipo de TEV en comparación con el perfil de expresión génica de plantas sanas. El resultado se divide en dos cifras, que representan el número de DEGs reprimidos (azul) y sobreexpresados (rojo).

La pregunta que debemos hacernos ahora es si los DEGs son comunes o diferentes entre genotipos dependiendo de la eficacia del genotipo viral correspondiente. Como en el apartado anterior, esperaríamos que dos genotipos que muestran pequeñas diferencias en la eficacia viral se correspondieran con los mismos DEGs en sus perfiles de expresión, es decir, para genotipos con una eficacia similar las listas de DEGs deberían ser iguales o muy parecidas. Del mismo

modo, esperaríamos que dos genotipos de distinta eficacia den lugar a listas de DEGs cuyos genes sobreexpresados y reprimidos sean muy distintas. Para responder a esta pregunta, hemos construido la **Figura 46**, que muestra el número de DEGs en común entre todos los pares de transcriptomas de plantas infectadas. La figura muestra un patrón de modularidad, con tres grupos bien definidos. Como de costumbre, el primer grupo contiene los tres genotipos con mayor eficacia (**WT**, **CI:PC55** y **CI:PC48**). El segundo grupo contiene los tres genotipos con eficacia intermedia (**HC-Pro:CLA11**, **HC-Pro:CLA2** y **Nib:PC95**). El tercer grupo de nuevo está formado sólo por el genotipo **HC-Pro:AS13**. El número de DEGs compartidos dentro de cada uno de los grupos fue $> 75\%$. El número de DEGs compartidos entre grupos fue $< 60\%$. Como era de esperar, el número de DEGs compartidos entre **HC-Pro:AS13** y los otros dos grupos es especialmente bajo, siendo $\leq 20\%$.

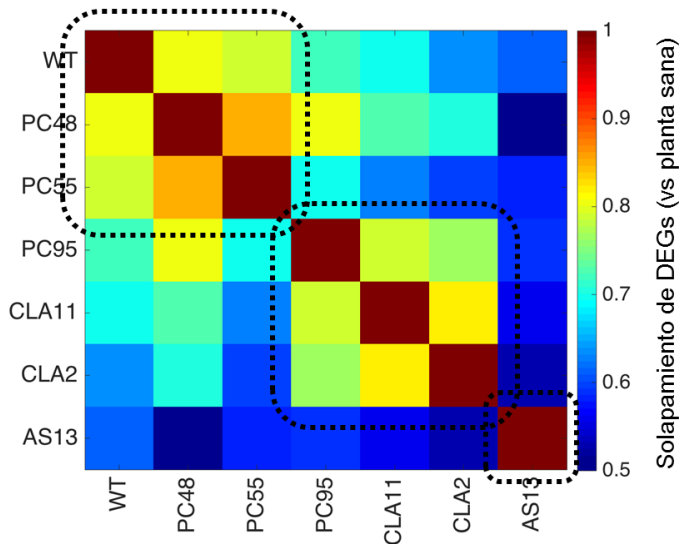


Figura 46. Grado de solapamiento entre pares de listas de DEGs obtenidas para plantas infectadas con cada genotipo viral. Las líneas delimitan engloban los tres módulos en que se pueden dividir los datos.

El color en la **Figura 46** representa el grado de solapamiento entre el par de listas que se comparan. Obviamente, cuando comparamos una lista de DEGs consigo misma la coincidencia es perfecta (diagonal representada en color rojo). Para el resto de casos, las parejas de listas que más DEGs poseen en común se destacan en color amarillo o naranja. Ello indica que ambas listas son muy similares. En caso contrario, el número de DEGs en común será muy pequeño y se indicará en color azul.

En la figura se posicionan los genotipos virales ordenados de acuerdo con el grado de solapamiento, lo cual coincide con el orden decreciente de los valores de eficacia.

A continuación, queremos determinar si el número de DEGs también es dependiente de las diferencias en la eficacia entre el genotipo silvestre y los genotipos mutantes. Siguiendo el mismo razonamiento que en la sección anterior, vamos a comparar el número de DEGs en relación con el transcriptoma de plantas infectadas con el genotipo silvestre. La **Figura 47** muestra el resultado de este análisis. El número de DEGs reprimidos oscila entre 0 y 1.120 (para **HC-Pro:AS13**), mientras que el número de DEGs sobreexpresados oscila entre 0 y 881 (para **HC-Pro:AS13**). Como era de esperar, **CI:PC48**, **CI:PC55** y **Nib:PC95** alteran los mismos genes, aunque **Nib:PC95** lo hace con una magnitud diferente (como se muestra en la sección anterior). El único gen que se ha encontrado el análisis de expresión diferencial en común entre **CI:PC48** y el genotipo silvestre es **EB438953** (cuyo gen ortólogo en *A. thaliana* es *At1g69760*, una proteína supresora llamada **SRP40**). El número de genes no comunes aumenta conforme aumenta la diferencia entre los valores de eficacia de los genotipos que se comparan. Así, encontramos una correlación negativa y significativa entre el número de diferencias en las listas de DEGs y las diferencias en la eficacia viral ($r = -0,934, 5 gl, P < 0,001$) (**Figura 48**).

Resultados y Discusión

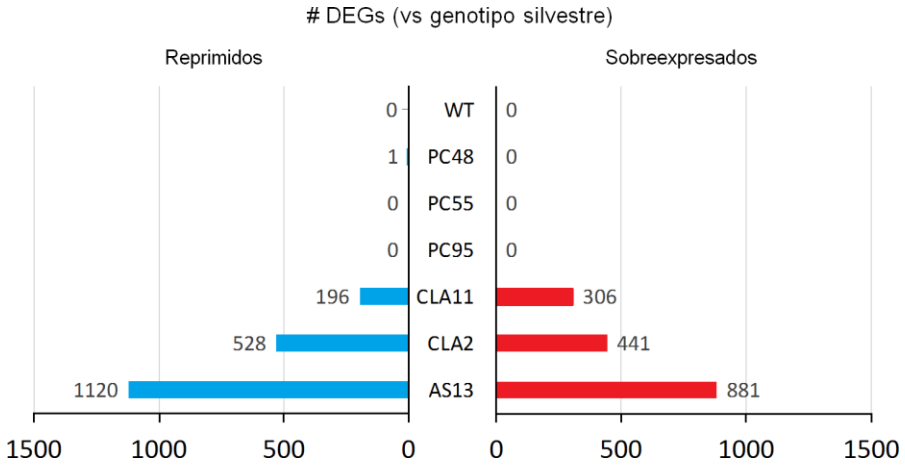


Figura 47. Resultado del análisis de expresión diferencial para cada genotipo mutante de TEV en comparación con el perfil de expresión génica de plantas infectadas con el genotipo silvestre de TEV, distinguiendo entre DEGs reprimidos (azul) y sobreexpresados (rojo).

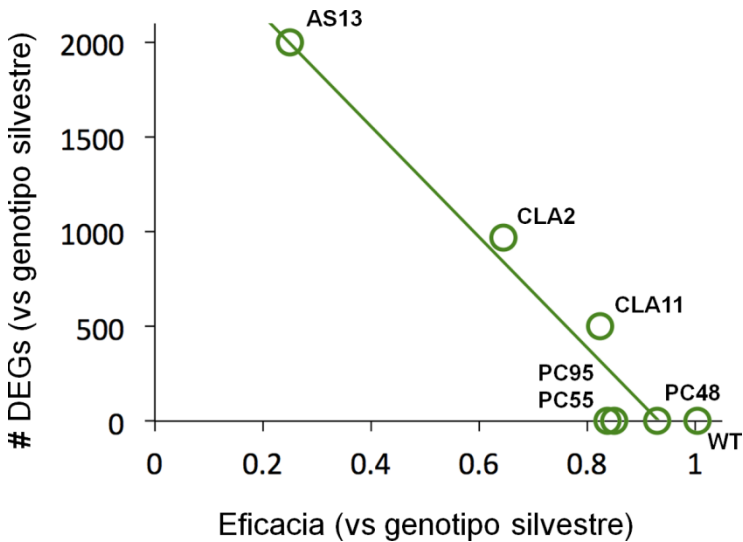
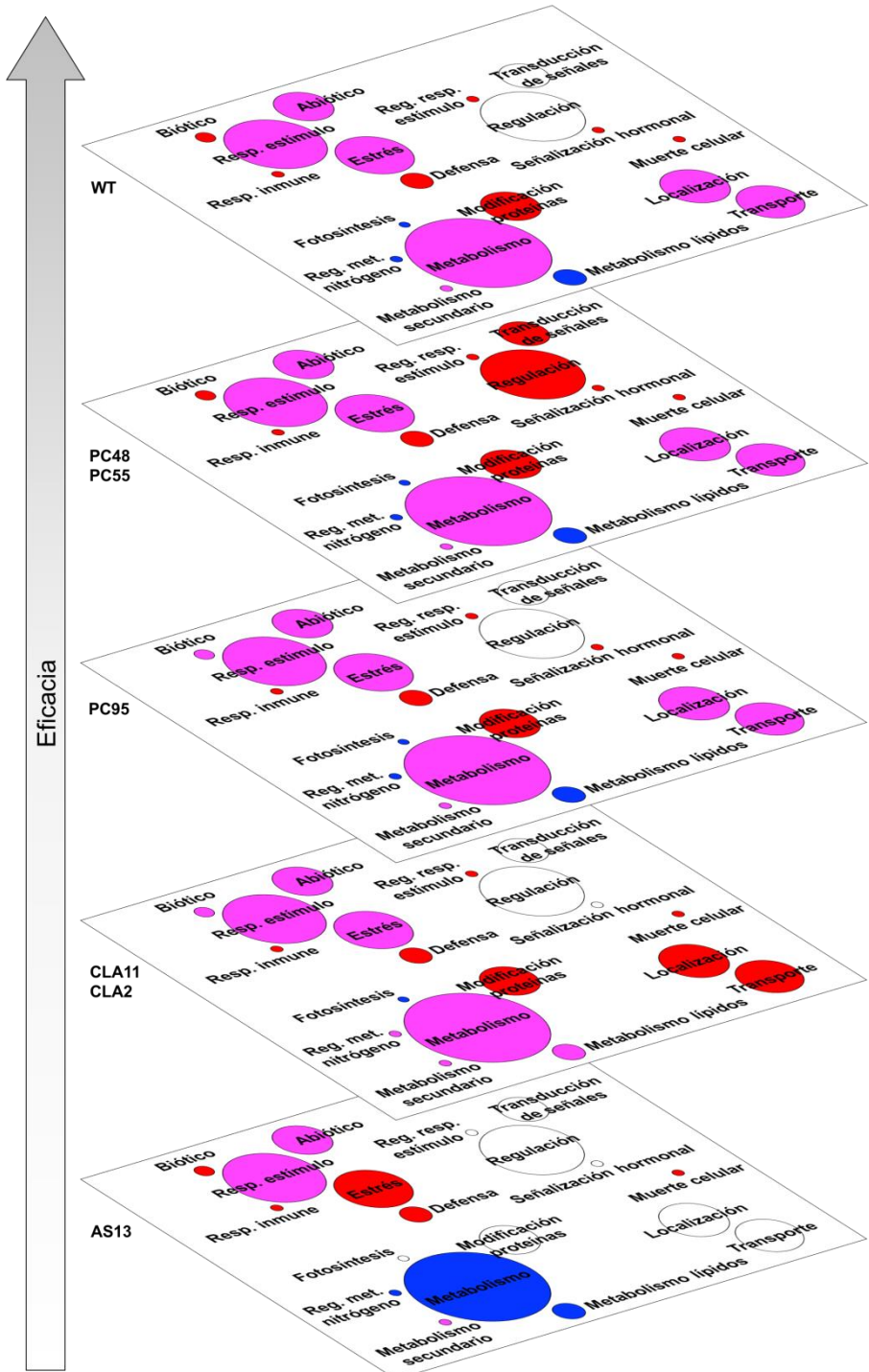


Figura 48. Gráfica que ilustra la correlación entre el número total de DEGs en relación con el transcriptoma de plantas infectadas con el genotipo silvestre y la eficacia relativa respecto al genotipo silvestre.

Para concluir con este apartado, exploraremos el tipo de genes incluidos en cada una de las listas. Con ello pretendemos averiguar si la categoría funcional de los genes también se ve condicionada de alguna manera por la eficacia viral.

Para arrojar luz sobre esta cuestión, realizamos un análisis de las categorías funcionales para los genes de las plantas infectadas con cada genotipo viral. La **Figura 49** ilustra las diferencias entre las categorías funcionales más representativas (para todas las muestras comparadas con las plantas sanas). Los diferentes planos se ordenan de arriba hacia abajo desde el genotipo con la eficacia más alta hasta el genotipo con la eficacia más baja. En cada plano, se muestran las categorías funcionales alteradas en las plantas infectadas con ese genotipo. La superficie de cada círculo es proporcional al número de DEGs incluidos en cada categoría. Para todo el conjunto, el mayor número de DEGs corresponde a la categoría de procesos metabólicos (**GO:0008152**) y el menor corresponde a la fotosíntesis (**GO:0015979**). Asimismo, el color rojo denota categorías que se han enriquecido en su mayoría con DEGs sobreexpresados, el color azul denota categorías que se han enriquecido en su mayoría con DEGs reprimidos y el color morado denota categorías enriquecidas con DEGs de ambos tipos. Si analizamos el plano de las plantas infectadas con el genotipo silvestre, podemos distinguir siete categorías funcionales indicadas en color rojo: regulación de estrés biótico (**GO:0002831**), defensa (**GO:006952**), sistema inmune (**GO:0002376**), modificación de proteínas (**GO:0036211**), señalización mediada por hormonas (**GO:0009755**), muerte celular (**GO:008219**) y regulación de la respuesta a estímulo (**GO: 0048583**). Únicamente hallamos tres categorías marcadas en color azul: fotosíntesis (**GO:0015979**), metabolismo de lípidos (**GO:0006629**) y la regulación metabólica de nitrógeno (**GO:0051171**). Por último, hay siete categorías denotadas en color morado: procesos metabólicos (**GO:0008152**), estrés inespecífico (**GO:0006950**), respuesta a estímulo (**GO:0050896**), estrés abiótico (**GO:0009628**), localización (**GO:0051641**), transporte (**GO:0008150**) y metabolismo secundario (**GO:0019748**).



← **Figura 49.** Representación de las categorías funcionales de los genes destacados en el análisis de expresión diferencial, siempre en relación al transcriptoma de plantas sanas. Cada plano comprende un genotipo (o a dos, si ambos transcriptomas son similares). Los planos se organizan en función de la eficacia viral. La flecha indica el sentido del aumento de la eficacia. Los campos coloreados indican un número de DEGs significativos: rojo para una mayoría de genes sobreexpresados, azul si la mayoría de genes están reprimidos y morado si ambos tipos se encuentran igualmente representados.

Por lo tanto, en el plano correspondiente al genotipo silvestre de TEV, encontramos sobreexpresados genes implicados en diferentes aspectos de las vías de defensa de la planta, y reprimidos unos pocos genes implicados en algún aspecto metabólico y en la fotosíntesis.

En el segundo plano se emplazan los genotipos **CI:PC48** y **CI:PC55**. La diferencia más notable con el resto de los genotipos analizados es el enriquecimiento de genes sobreexpresados en las categorías de transducción de señales (**GO:0007165**) y regulación de la expresión génica (**GO:0010468**).

El tercer plano pertenece al genotipo **Nib:PC95** y la distribución de los colores en las categorías funcionales es muy similar a la del primer plano; únicamente cambia de rojo a morado la categoría de regulación de estrés biótico (**GO:0002831**), lo que significa que a esta categoría ahora están contribuyendo más genes reprimidos que en el caso del genotipo silvestre.

Recordemos que el genotipo silvestre y los genotipos mutantes **CI:PC48**, **CI:PC55** y **Nib:PC95** alteran los mismos genes (**Nib:PC95** lo hace con una magnitud diferente). Así, si anteriormente encontramos que pequeñas diferencias en la eficacia viral se correspondieran con

los mismos DEGs en el análisis de expresión diferencial (independientemente de si se realizaba la comparación con la planta sana o con la infectada con el genotipo silvestre), esperaríamos observar una distribución similar de categorías funcionales en los tres primeros planos. En efecto, la distribución de los DEGs en las 19 categorías funcionales es casi idéntica (únicamente observamos cambios en las tres categorías citadas en el párrafo anterior).

Continuando con el análisis de los planos, el cuarto engloba los genotipos **HC-Pro:CLA2** y **HC-Pro:CLA11**. Dicho plano difiere de los tres primeros planos en cuatro categorías funcionales: no observamos ningún enriquecimiento de DEGs en la señalización mediada por hormonas (**GO:0009755**), la regulación metabólica de nitrógeno (**GO:0051171**) cambia de azul a morado, y destaca el enriquecimiento de genes sobreexpresados en las categorías de localización (**GO:0051641**) y transporte (**GO:0008150**). Además, al igual que ocurría con el genotipo mutante **Nib:PC95**, la categoría de regulación de estrés biótico (**GO:0002831**) también cambia de rojo a morado (respecto a los dos primeros planos).

Finalmente, el quinto y último plano corresponde al genotipo **HC-Pro:AS13**, que recordemos posee la eficacia viral más baja e induce síntomas de forma tardía y leve. Sabemos que **HC-Pro:AS13** constituye un caso especial debido a que las listas de genes sobreexpresados y reprimidos son muy distintas respecto a las listas de los demás genotipos. Lo que esperaríamos entonces es que el plano de distribución de categorías funcionales de **HC-Pro:AS13** fuera muy diferente de los demás. En efecto, en comparación con el resto de planos, observamos una disminución del enriquecimiento en las categorías de localización (**GO:0051641**), transporte (**GO:0008150**), modificación de proteínas (**GO:0036211**), respuesta a estímulo (**GO:0050896**) y fotosíntesis (**GO:0015979**). Además, observamos un enriquecimiento de genes reprimidos en procesos metabólicos (**GO:0008152**). Por último, hay un enriquecimiento de genes sobreexpresados en estrés de tipo inespecífico (**GO:0006950**). En

comparación con alguno de los planos en particular, también podemos observar muchos más cambios, por ejemplo, la categoría de regulación de estrés biótico (**GO:0002831**) cambia de morado a rojo (respecto al tercer y cuarto planos). En resumen, todos los genes relacionados con estrés inespecífico, estrés biótico, defensa, respuesta inmune y muerte celular están sobreexpresados, en detrimento de aquellos relacionados con cualquier proceso metabólico, que se hallan reprimidos. El plano de la distribución de categorías funcionales de los genes de la planta a la que infecta da una visión general que explicaría cómo la planta conseguiría defenderse frente al patógeno, o por el contrario, por qué el patógeno no podría superar las defensas de su huésped.

3. Correlación entre la expresión génica y la eficacia viral.

Si valoramos todos los resultados de las secciones anteriores de este capítulo, es evidente que existe una respuesta transcriptómica por parte de la planta frente a la infección por parte de TEV. Sin embargo, el perfil transcriptómico es diferente en función del genotipo con que se inoculó la planta. En unas ocasiones, un grupo de genes puede cambiar su expresión ante un genotipo concreto (por ejemplo, transducción de señales y regulación de la expresión génica frente a los genotipos **CI:PC48** y **CI:PC55**), y en otras ocasiones parece que la expresión varía de forma continua en el mismo sentido en que se produce la variación continua de la eficacia (por ejemplo, los genes de estrés de tipo inespecífico). En otras palabras, debiera ser posible correlacionar la expresión de los genes con la eficacia viral, de modo que aquellos genes que correlacionen significativamente debieran reflejar aumentos o disminuciones continuas de su expresión en relación a la disminución continua del valor de eficacia.

Un análisis de correlación no permite concluir que realmente exista una dependencia funcional entre la eficacia viral y la expresión de un gen individual del huésped. Sin embargo, sí puede proporcionarnos una lista de genes candidatos que pudiéramos considerar como “**genes determinantes de la eficacia viral**”. Bajo el supuesto planteado en el párrafo anterior, los genes cuya expresión correlacionara significativa y positivamente con la eficacia viral pudiéramos definirlos como posibles “**determinantes positivos**”, es decir, cuanto mayor sea la eficacia del genotipo de TEV mayor será la expresión de este gen en la planta huésped. Por el contrario, los genes cuya expresión se correlacionará significativa y negativamente con la eficacia viral pudiéramos clasificarlos como posibles “**determinantes negativos**”, es decir, cuanto mayor sea la eficacia del genotipo de TEV menor será la expresión de este gen en la planta huésped.

Con todo lo visto, calculamos el coeficiente de correlación entre la eficacia viral y el valor de la expresión génica normalizada de un único gen (puntuación z). El cálculo de la correlación se repitió para todos los genes de *N. tabacum* (**Figura 45**). Así, identificamos 326 genes con correlación positiva y 154 genes con correlación negativa (prueba ANOVA; $P < 0,05$) (**Figura 50**).

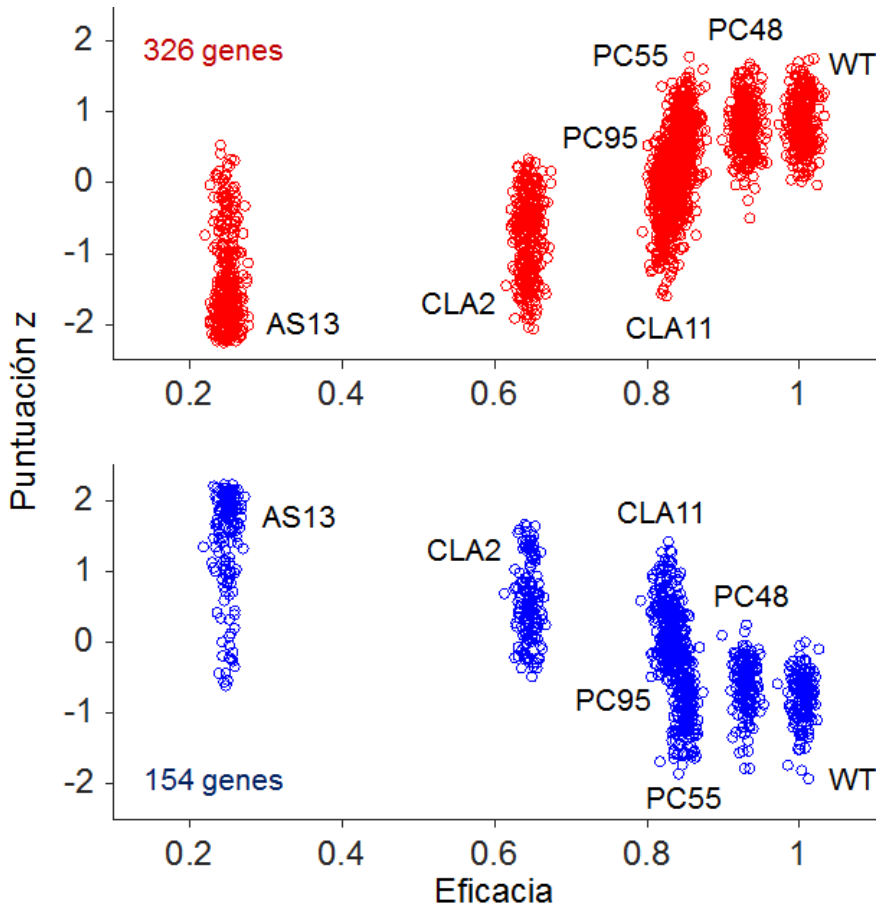


Figura 50. Correlación entre la expresión y la eficacia viral. El panel superior muestra genes correlacionados positivamente. El panel inferior; genes correlacionados negativamente.

A continuación, hemos vuelto a realizar un análisis de las categorías funcionales con estos dos nuevos conjuntos de genes. La particularidad de este análisis radica en los datos empleados: en este caso no se trata de muchas listas de genes individuales para cada genotipo viral, sino que se trata de una única lista que incluye genes que correlacionan con la eficacia de cada uno de los siete genotipos virales. Esta lista puede dividirse en dos de forma sencilla; si

ordenamos los valores del coeficiente de correlación no paramétrico de Spearman, aquellos genes cuyo valor sea $r_s \geq 0,786$ serán considerados como posibles “**determinantes positivos**” y aquellos que posean $r_s \leq -0,786$ serán considerados como posibles “**determinantes negativos**”. Con estas dos listas procedimos a realizar el análisis de enriquecimiento funcional.

Una vez obtenidas las dos listas de categorías funcionales, decidimos dividir las en dos subconjuntos para poder analizar más detalladamente y por separado las categorías relacionadas con funciones biológicas y funciones moleculares. De este modo, pudimos observar diferencias significativas en la distribución de genes correlacionados positiva y negativamente en categorías relacionadas con funciones biológicas (diagramas de sectores de la columna izquierda de la **Figura 51**; $\chi^2 = 29,225$, 6 gl, $P < 0,001$), siendo la diferencia entre ambos gráficos de magnitud moderada (V de Cramér = 0,304). Si comparamos ambos gráficos, la regulación de procesos biológicos (**GO:0065008**) y del desarrollo (**GO:0032502**) se hallan fuertemente representados entre los “**determinantes positivos**” en comparación con los “**determinantes negativos**”. Por el contrario, éstos últimos se enriquecen más en las categorías de respuesta a estímulo (**GO:0050896**) y localización (**GO:0051641**). Centrándonos en las funciones moleculares, también observamos diferencias significativas en la distribución de genes correlacionados (diagramas de sectores de la columna derecha de la **Figura 51**; $\chi^2 = 36,720$, 6 gl, $P < 0,001$). En este caso, la diferencia entre ambos gráficos es de mayor magnitud (V de Cramér = 0,341). Si comparamos ambos gráficos, la categoría de unión a ácidos nucleicos (**GO:003676**) se halla más enriquecida entre los “**determinantes positivos**” y las categorías de actividad catalítica (**GO:003824**) y transporte (**GO:0005215**) se hallan más enriquecidas entre los “**determinantes negativos**”.

Valorando los resultados de este apartado, parece que los “**determinantes positivos**” estarían más implicados en la regulación transcripcional de las defensas del huésped, mientras que los “**determinantes negativos**” lo estarían en actividades catalíticas y de transporte. Todo ello sugiere una reorientación de los recursos por parte del huésped para hacer frente a la infección que depende de la eficacia del virus.

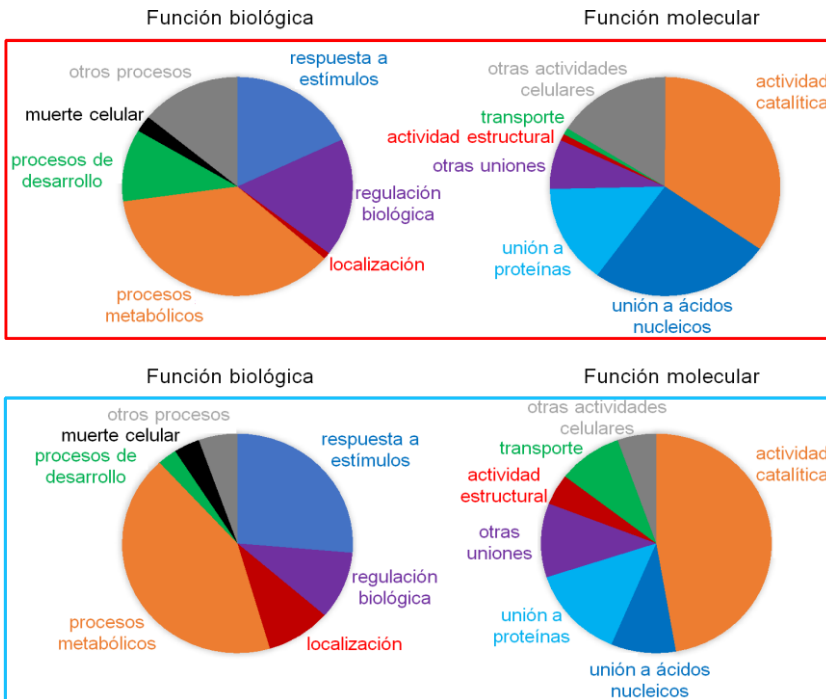


Figura 51. Diagramas de sectores mostrando la distribución de categorías relacionadas con funciones biológicas y moleculares. Remarcadas en rojo, figuran las gráficas correspondientes a genes correlacionados positivamente. En azul, las correspondientes a genes correlacionados negativamente.

4. Validación de los valores de expresión normalizados para ocho genes mediante RT-qPCR.

Todos los apartados anteriores se sustentan sobre los datos de expresión normalizados que obtuvimos de la hibridación de siete micromatrices diseñadas para contener el transcriptoma de *N. tabacum*. Para finalizar este capítulo, quisimos validar estos resultados con un método alternativo. Así, seleccionamos ocho genes correlacionados positiva y negativamente (cuatro de cada clase; **Tabla 12**) que cubrieran un rango de coeficientes de correlación significativos amplio, es decir, desde genes cuya correlación es significativa y muy alta hasta genes cuya correlación es significativa, aunque baja (**Tabla 13**). Los ocho genes están involucrados en diferentes categorías funcionales o se expresan en diferentes etapas del desarrollo, por lo que no esperamos un sesgo en la selección de dicha muestra.

Tabla 12. Genes seleccionados para validación por RT-qPCR. En rojo, genes que correlacionan positivamente con la eficacia viral, y en azul, los que correlacionan negativamente.

| Sonda | Gen | Nombre |
|--------------|-------|---------------------------------|
| A_95_P156197 | VQ29 | VQ MOTIF-CONTAINING PROTEIN 29 |
| A_95_P217927 | DCL2 | DICER-LIKE 2 |
| A_95_P156817 | GASA1 | GAST1 protein homolog 1 |
| A_95_P164432 | PLAT1 | PLAT domain protein 1 |
| A_95_P146147 | ADK2 | ADENOSINE KINASE 2 |
| A_95_P245082 | GBSS1 | GRANULE BOUND STARCH SYNTHASE 1 |
| A_95_P191007 | AGL20 | AGAMOUS-LIKE 20 |
| A_95_P005611 | FDM1 | FACTOR OF ADN METHYLATION 1 |

Tabla 13. Resumen de los parámetros estadísticos que describen los genes que se seleccionaron para validación mediante RT-qPCR.

| Gen | Umbral | Pendiente | r_s |
|--------------|----------------|------------------|------------------|
| VQ29 | 4,81701 | 5,964776 | 0,892857 |
| DCL2 | 2,6719 | 2,982678 | 0,892857 |
| GASA1 | 2,93044 | 3,282595 | 0,857143 |
| PLAT1 | 5,44308 | 6,689926 | 0,785714 |
| ADK2 | 0,8396 | -0,859033 | -0,857143 |
| GBSS1 | 4,33993 | -4,148685 | -0,821429 |
| AGL20 | 3,181 | -3,744371 | -0,785714 |
| FDM1 | 1,06181 | -1,41815 | -0,785714 |

Los genes correlacionados positivamente (o “**determinantes positivos**”) que seleccionamos son: **DCL2**, **VQ29**, **GASA1** y **PLAT1**. Muy brevemente, **DCL2** está implicado en la respuesta defensiva frente a los virus (mediada por ARNs interferentes) y el mantenimiento de la metilación del ADN (Parent et al., 2015). **VQ29** también está implicado en la defensa, mediada esta vez por ácido jasmónico, e interacciona con gran número de factores de transcripción (Li et al., 2014). **GASA1** codifica una proteína regulada por hormonas (brasinosteroides y giberelinas) probablemente implicada en la elongación celular (Bouquin et al., 2001) y en la respuesta al estrés abiótico mediada por la señalización de especies reactivas de oxígeno (ROS) (O’Brien et al., 2012). **PLAT1** codifica una lipasa/lipoxigenasa que promueve la tolerancia al estrés abiótico (Hyun et al., 2015).

Los genes correlacionados negativamente (o “**determinantes negativos**”) que seleccionamos: **ADK2**, **AGL20**, **FDM1** y **GBSS1**. También de forma muy breve, **ADK2** codifica para una adenosina quinasa implicada en el metabolismo de la adenosina (Schoor et al., 2011) e interviene en el silenciamiento génico mediado por metilación (Eckardt, 2003). **AGL20** es un activador de la transcripción implicado en el desarrollo de la diferenciación del meristemo de la inflorescencia y en la tolerancia al frío (Lee et al., 2000). **FDM1** es una proteína capaz de metilar el ADN y también participa en la defensa de la planta mediante silenciamiento génico (Xie et al., 2012). **GBSS1** es un gen implicado en la biosíntesis de glucano y responsable de la síntesis de amilasa y otros procesos del desarrollo vegetal (Denyer et al., 1996).

Los detalles de los genes de expresión constitutiva, el tamaño de los amplicones y demás aspectos de la metodología de la validación experimental pueden revisarse en el **apartado 6.4 de Material y Métodos**. Los detalles específicos referentes a los cebadores empleados para amplificar los ocho genes que hemos seleccionado se hallan en la **Tabla 14**.

Para calcular la expresión relativa de cada gen se empleó el método $\Delta\Delta C_T$. Recordemos de forma muy breve dicha metodología. Las amplificaciones se realizaron individualmente para cada gen (con su propio conjunto de cebadores), incluyendo en cada placa los dos genes de expresión constitutiva. Una vez concluidas todas las cuantificaciones, recalculamos los valores de cuantificación empleando para ello el valor promedio de la línea base (0,1056).

Finalmente, para poder comparar fácilmente los niveles de expresión obtenidos por ambos métodos (micromatrices y RT-qPCR), todos los datos fueron transformados en puntuaciones *z*. Los resultados se muestran en la **Figura 52** (genes que correlacionan positivamente) y en la **Figura 53** (aquellos que correlacionan negativamente).

Tabla 14. Características de los cebadores empleados para amplificar los ocho genes seleccionados para validación por RT-qPCR.

| Gen | Cebador | Secuencias 5'-3' | Tamaño amplicón |
|--------------|---------|----------------------------|-----------------|
| VQ29 | 1F | CCTTCATACCATTTCCTCACTTCA | 104 |
| | 1R | CGGTAGATTTTTGGAGGTGTTGG | |
| DCL2 | 1F | AGCTCTGCCCTGACTCTCTGCC | 68 |
| | 1R | GCTTTAATGCCGTCTCCAATGC | |
| GASA1 | 1F | CAATGGCGATTTCTAAAACCTCTCT | 74 |
| | 1R | AATGGATTGGACTTCATCGAGG | |
| PLAT1 | 1F | CATGGGAGTAGCACAAGTTAACC | 85 |
| | 1R | TTCAGTGCCATAAATAGAAGAAACGG | |
| ADK2 | 3F | ATTTTCCTCTTGTTGGGCTTTCC | 130 |
| | 3R | CAATATGGGGGATGGGAAGTGG | |
| GBSS1 | 2F | CGCAATATCACACATGGCAAGC | 123 |
| | 2R | TGGTTCCTCAGCCCTATCTGTG | |
| AGL20 | 1F | CTGAGGGAGAAGTTTGGTGGCC | 72 |
| | 1R | TGTACAAAGAGCTTCCCCTTCC | |
| FDM1 | 2F | GGAGGAATGGGGGGATGAAATC | 108 |
| | 2R | CAGCGTGGCCTTCCTTCCTTCC | |

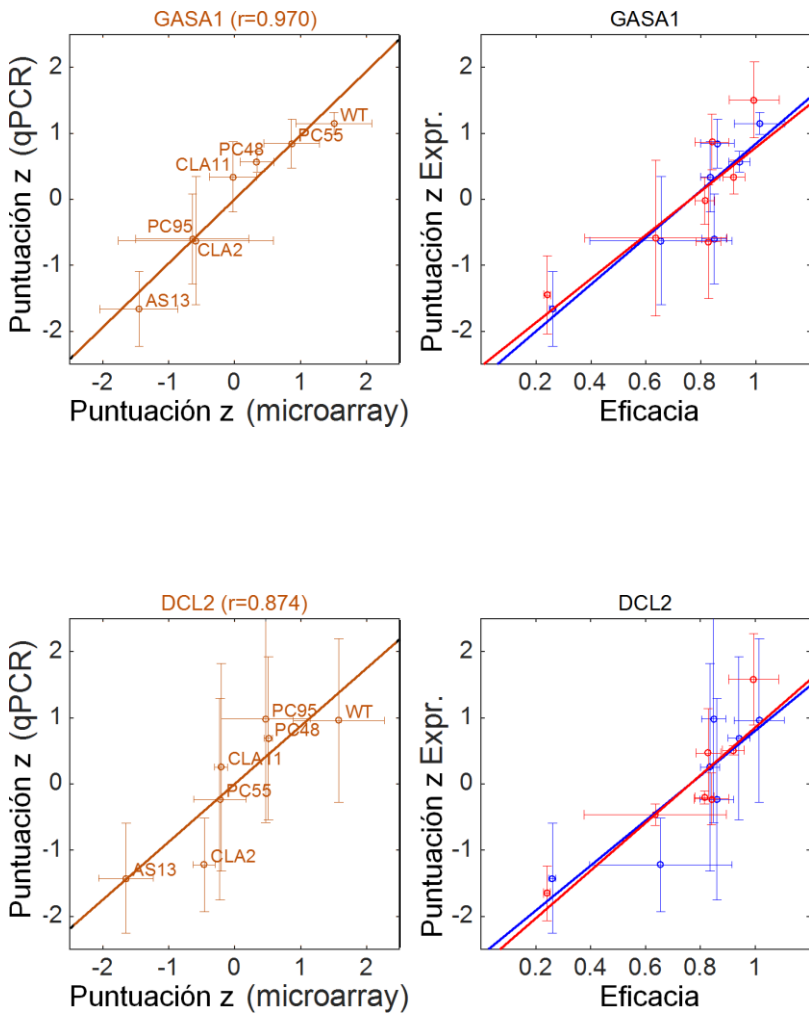
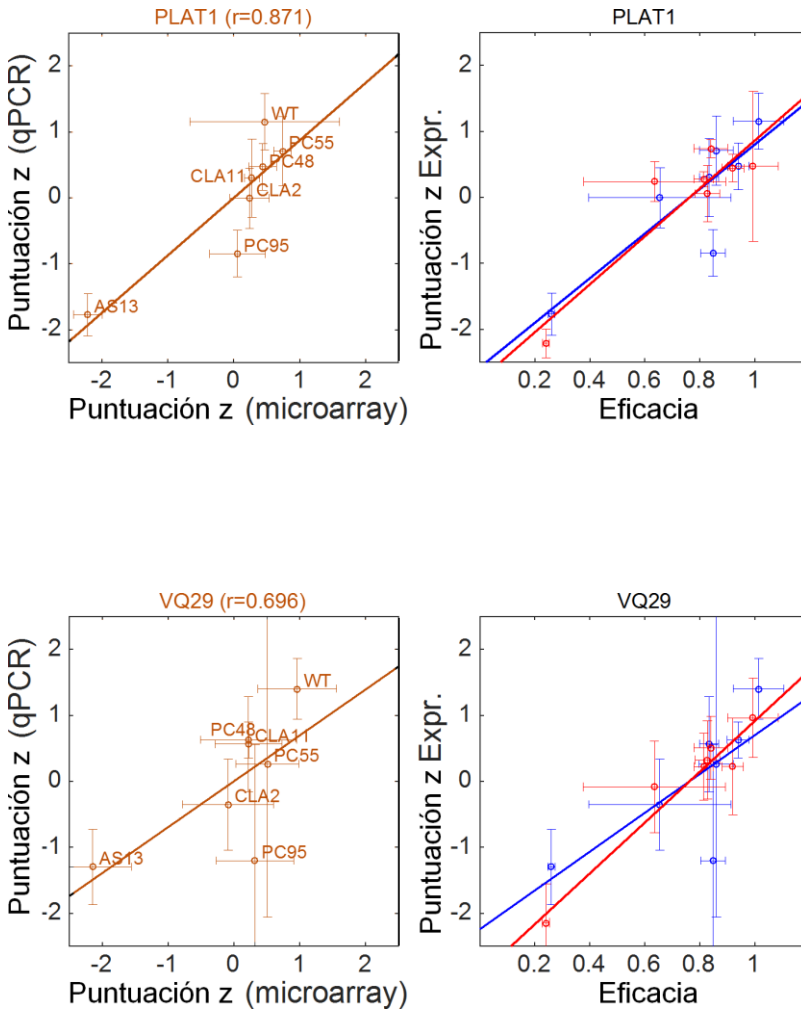


Figura 52. Validación de los datos transcriptómicos para cuatro “determinantes positivos”. El gráfico izquierdo muestra la correlación (r) entre los niveles de expresión normalizados medidos con las micromatrices de ARN (eje x) y con RT-qPCR (eje y). La línea continua representa la hipótesis nula de igualdad de los valores de expresión (la coincidencia perfecta entre las dos cuantificaciones).



El gráfico derecho muestra los gráficos de correlación superpuestos de la expresión génica (en rojo la medición con micromatrices y en azul con RT-qPCR) con la eficacia viral. Cuanto más cercanas se sitúen las pendientes de las dos líneas, la similitud entre los datos de expresión será mayor. Las barras de error representan $\pm 1 d. e.$

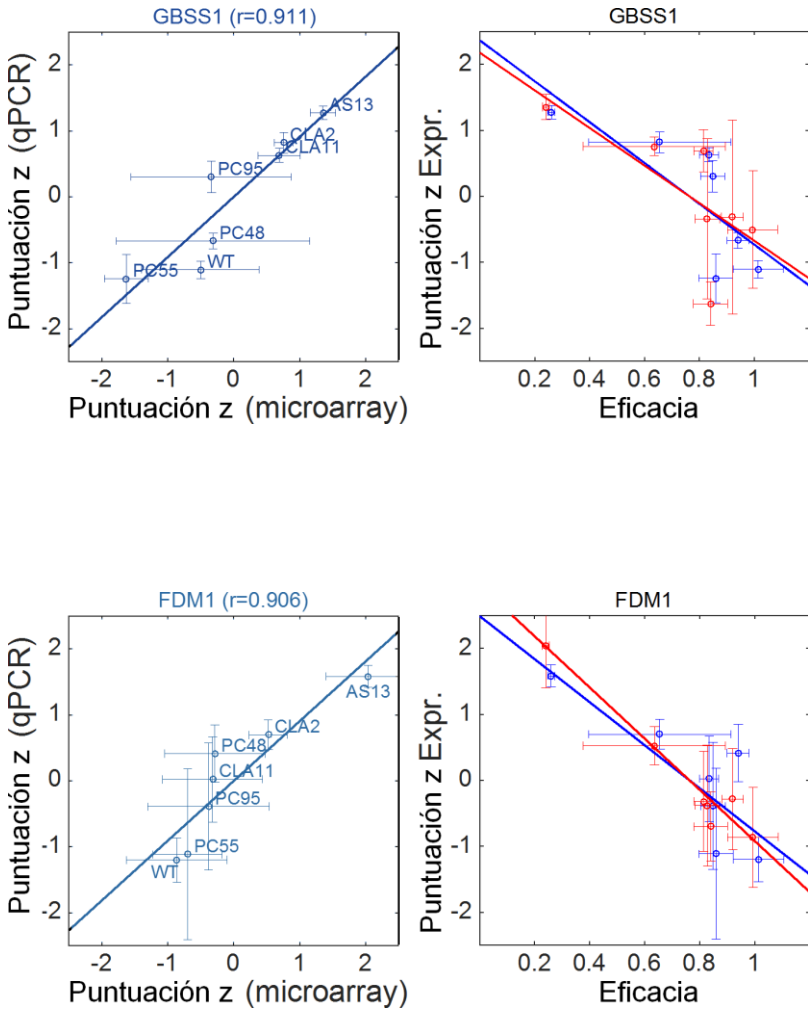
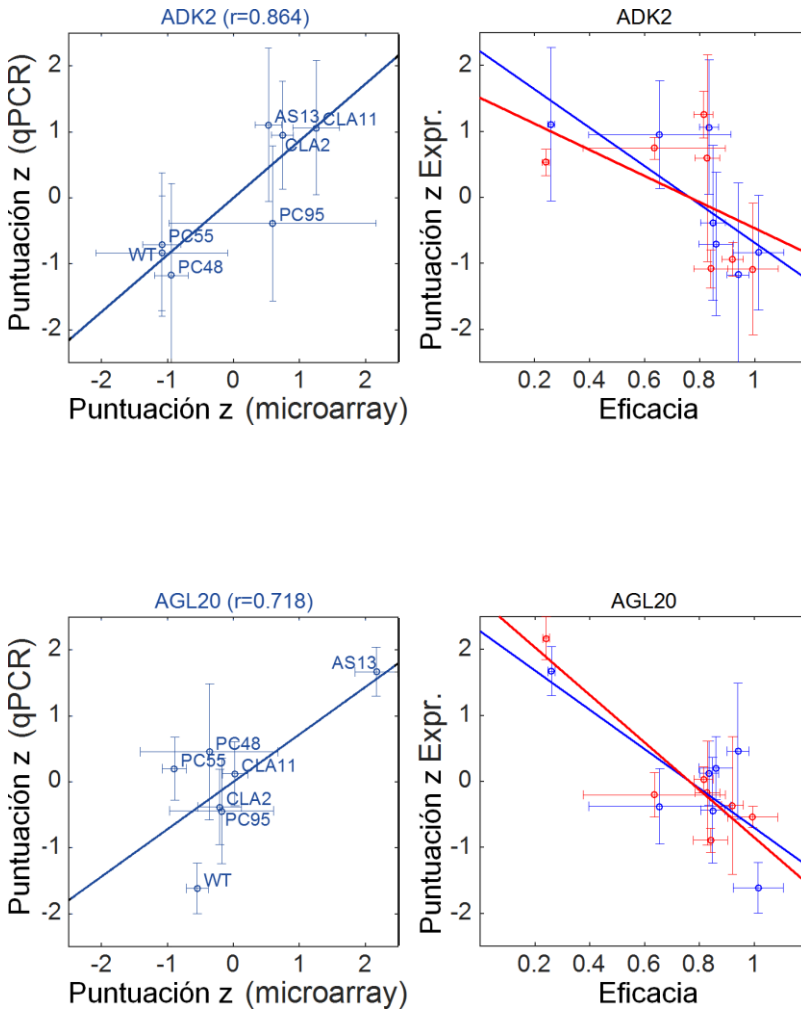


Figura 53. Validación de los datos transcriptómicos para cuatro “determinantes negativos”. El gráfico izquierdo muestra la correlación (r) entre los niveles de expresión normalizados medidos con las micromatrices de ARN (eje x) y con RT-qPCR (eje y). La línea continua representa la hipótesis nula de igualdad de los valores de expresión (la coincidencia perfecta entre las dos cuantificaciones).



El gráfico derecho muestra los gráficos de correlación superpuestos de la expresión génica (en rojo la medición con micromatrices y en azul con RT-qPCR) con la eficacia viral. Cuanto más cercanas se sitúen las pendientes de las dos líneas, la similitud entre los datos de expresión será mayor. Las barras de error representan $\pm 1 d.e.$

Ambas figuras anteriores ilustran la comparación de las dos medidas de expresión para cuatro genes de cada grupo. Se han construido dos gráficos distintos para cada gen validado. Las gráficas pertenecientes al mismo gen se sitúan en la misma fila. Así, la gráfica izquierda ilustra la correlación (r) entre los niveles de expresión normalizados medidos con las micromatrices (eje x) y con RT-qPCR (eje y), mientras que la gráfica derecha muestra los gráficos de correlación superpuestos de la expresión génica (en rojo la medición con micromatrices y en azul con RT-qPCR) con la eficacia viral.

Respecto a la gráfica izquierda, la línea continua representa la hipótesis nula de igualdad de los valores de expresión, es decir, la coincidencia perfecta entre las dos cuantificaciones, que se correspondería con una línea de regresión con valor $r = 1$. En los ocho casos analizados, ambas puntuaciones z están correlacionadas de forma significativa, aunque el rango de valores de significación es amplio. Para el caso de los genes que correlacionan positivamente con la eficacia viral, el menor valor de correlación entre las puntuaciones z pertenece al gen **VQ29** ($r = 0,696$) y el mayor valor al gen **GASA1** ($r = 0,970$). Para aquellos genes que correlacionan negativamente con la eficacia viral, el valor más pequeño lo registra el gen **AGL20** ($r = 0,718$) y el valor más grande el gen **GBSS1** ($r = 0,911$). Todos los valores de correlación de Pearson se calcularon bajo los mismos supuestos (5 gl , $P \leq 0,041$). Por lo general, la mayoría de las correlaciones posee un valor $r > 0,85$ y, aunque todos los resultados son significativos, si somos algo más restrictivos con la corrección Bonferroni-Holm del nivel de significación global, el gen **VQ29** pudiera no seguir siendo significativo. Dicha diferencia en el valor de r pudiera deberse únicamente a fluctuaciones experimentales o a pequeñas diferencias de eficiencia de amplificación entre los cebadores específicos (con los que se validó cada muestra por separado).

Respecto a la gráfica derecha, las líneas continuas representan el mejor ajuste lineal entre los valores de expresión normalizados y la eficacia viral. Así, cuanto mayor sea la superposición entre ambas líneas, mayor será la similitud entre los datos de expresión calculados mediante ambos métodos cuantitativos. Para todos los genes objeto de estudio, ambos gráficos de correlación se superponen sin dar lugar a diferencias significativas (prueba no paramétrica de Wilcoxon, $P \geq 0,499$; prueba t de Student, $P \geq 0,285$). La superposición de las líneas de correlación de los genes **VQ29** y **ADK2** es inferior a la observada para otros genes, pero la diferencia no es lo suficientemente grande como para que el análisis estadístico la resalte como significativa.

En conclusión, la correlación observada entre la expresión génica del huésped y la eficacia viral es consistente entre ambos métodos experimentales utilizados para evaluar los niveles de expresión génica.

5. Discusión.

Muy a menudo, los síntomas derivados de una enfermedad no suelen ser resultado directo de una mutación en un único gen, sino que suelen ser reflejo de múltiples perturbaciones causadas por esa mutación en una extensa y compleja red de interacciones (Barabási et al., 2011), en la que intervienen elementos tan diversos como genes, proteínas, metabolitos o ARNs (Finzer, 2017). Las inducidas por patógenos vegetales no son una excepción (Bailer y Haas, 2009; Friedel y Haas, 2011), y no son pocas las ocasiones en que se ha estudiado el alcance de dichas perturbaciones empleando un modelo planta-patógeno (Tan et al., 2007; Viswanathan y Früh, 2007; Agudelo-Romero et al., 2008a; Elena y Rodrigo, 2012; Bernet y Elena, 2015; Martínez et al., 2016; Willemsen et al., 2016a; Hillung et al., 2016).

El proceso y el progreso de la infección por parte de un virus dependen directamente de su eficacia (Orr, 2009), característica de mayor relevancia a la hora de engendrar enjambres mutantes con los que esperar conseguir un genotipo que supere las defensas del huésped o adaptarse a uno nuevo (Linnen y Hoekstra, 2010). La eficacia viral de distintos genotipos es una medida muy empleada en estudios relacionados con biología evolutiva de sistemas. A grandes rasgos, dicha medida describe la acumulación absoluta de viriones en un tiempo dado (Holland et al., 1991; Wargo y Kurath, 2012) y puede incluso correlacionarse con el desarrollo de los síntomas en el huésped (Carrasco et al., 2007b).

Los resultados expuestos a lo largo de este capítulo nos permiten explicar parte de la diferencia de eficacia que distingue a cada uno de los mutantes. Sin embargo, dichas diferencias pueden deberse a muchos factores, no todos ellos contemplados en este estudio. En primer lugar, es posible que la diferencia de un mutante particular pudiera explicarse únicamente por el efecto directo de una mutación en alguna proteína viral. Este caso no es el más habitual (Barabási et al., 2011) y depende de que la mutación sea no sinónima (que en el caso más extremo, podría llevar a cambios en la conformación tridimensional de la proteína y dificultar o suprimir su función). Todos los mutantes empleados en este estudio poseen una o dos mutaciones no sinónimas pero no disponemos de información relativa a posibles cambios en la conformación de las proteínas con los que podamos mostrar una consecuencia directa en el valor de la eficacia.

Una segunda posibilidad consistiría en que las diferencias de eficacia pudieran explicarse por mutaciones que modificaran la configuración estructural del virión. En este caso, la interacción de una proteína alterada pudiera ver reducida su capacidad para unirse o interactuar con el resto de proteínas del virus o factores de la maquinaria celular del huésped. El problema continúa siendo el mismo que el que expusimos anteriormente; *a priori*, es difícil ofrecer

una explicación concluyente sobre esta posibilidad sin ningún tipo de información adicional más que la secuencia de los mutantes.

La tercera posibilidad es que una o varias mutaciones en el genoma viral condujeran a una interacción sub-óptima con unos pocos elementos que, como si de un efecto mariposa se tratase, alteraran de forma radical la compleja red de interacciones entre el patógeno y su huésped. Como parásitos intracelulares que son, los virus necesitan secuestrar todo tipo de factores celulares y recursos, reprogramar los patrones de expresión de los genes del huésped en su propio beneficio e interferir con las defensas celulares. Todos estos procesos se integran en la red de interacciones y el efecto sobre uno de ellos puede significar la diferencia entre una replicación ineficiente o una infección exitosa. Estudios anteriores se centraban únicamente en la comparación de las redes de un mutante dado y el genotipo silvestre (Geng et al., 2016). He aquí el interés de nuestro trabajo, ya que analizamos la relación entre la eficacia viral de siete genotipos y el perfil transcriptómico del huésped.

En términos generales, hemos observado que la expresión de los genes de la planta se divide en dos bloques claramente diferenciados: el primero de ellos constituido por todos aquellos genes relacionados con la respuesta al estrés y el segundo, por los relacionados con metabolismo y desarrollo. Es interesante detenerse aquí y ahondar con un poco más en dos de los genotipos mutantes empleados en este estudio, pues consideramos que son de gran interés. El primero que queremos destacar es el mutante **Nib:PC95**. Recordemos que poseía una mutación no sinónima en la replicasa **Nib** (Revers y García, 2005; Carrasco et al., 2007a) y un valor de eficacia intermedio ($W = 0,838 \pm 0,004$). La comparación del transcriptoma de una planta inoculada con este virus y una planta sana daba lugar a 4.956 DEGs, mientras que la comparación de la planta inoculada con el genotipo silvestre y la planta sana destacaba 692 DEGs menos. En principio, esos 692 genes debieron ser la clave para que **Nib:PC95** se agrupara con los mutantes simples para

HC-Pro. El genotipo silvestre y los mutantes **CI** prácticamente tenían el mismo número de DEGs y, además, la amplísima mayoría se encontraba en las tres listas. Sin embargo, hablando de números absolutos, la mayor diferencia registrada entre los tres correspondía a 803 genes diferentes. Volviendo al punto anterior, no es posible que siendo menor la diferencia en cuanto a números (692 genes frente a 803) **Nib:PC95** se clasifique junto a otros genotipos. En realidad, la explicación es sencilla. Si en vez de analizar los números absolutos determinamos el número de genes compartidos entre **Nib:PC95** y los mutantes **CI**, la diferencia es de más 2.974 genes (el número de DEGs compartidos entre grupos era menor al 60 %). Ello explica que observemos **Nib:PC95** junto a **HC-Pro:CLA11** y **HC-Pro:CLA2**. Es más, los genotipos virales se posicionaron ordenados de acuerdo con el orden decreciente de los valores de eficacia, situándose **Nib:PC95** en la posición central (de acuerdo a su eficacia). Siguiendo el mismo razonamiento, si comparamos el número de DEGs en relación con el transcriptoma de plantas infectadas con el genotipo silvestre, **CI:PC48**, **CI:PC55** y **Nib:PC95** alteran los mismos genes, es decir, no hay ninguna diferencia entre ellos. Técnicamente, en esta nueva comparación, **Nib:PC95** debiera agruparse con los genotipos **CI** y silvestre, pero al llevar a cabo el cálculo de los coeficientes de correlación de Pearson de todos los genes **Nib:PC95** sigue agrupándose con **HC-Pro:CLA11** y **HC-Pro:CLA2**. La interpretación puede resultar difícil. La diferencia se haya realmente no en el número de DEGs sino en la magnitud con que se expresan los genes. **Nib:PC95** constituye el punto de inflexión entre un genotipo calificado como muy eficaz y otro calificado como poco eficaz, en el que un determinado número de genes se hallan alterados en el margen de detección de la expresión diferencial ($0,58 > \text{cambio relativo} < 1$). Es muy probable que ello explique únicamente parte de la pérdida de eficacia, mientras que el porcentaje restante de dicha pérdida pudiera explicarse por la presencia de la mutación no sinónima en la replicasa (que de conllevar un cambio estructural en la proteína conduciría a un número menor de ciclos de replicación por unidad de tiempo).

Lamentablemente, no podemos ampliar la información sobre el efecto de mutaciones simples en la replicasa viral, pues no poseemos datos de ningún otro mutante para **Nib**. Desde un primer momento, se inoculó un gran volumen de plantas con otros dos genotipos mutantes simples para **Nib** (**Nib:PC81** y **Nib:PC96**; Carrasco et al., 2007a). Sin embargo, nunca obtuvimos ninguna infección ni ningún resultado positivo de detección de partículas virales por RT-PCR, lo que sugiere que ambos genotipos, que procedían de una colección de 66 mutantes generados simultáneamente, eran letales.

El segundo es el mutante **HC-Pro:AS13**, único mutante doble incluido en este estudio (Revers y García, 2005; Torres Barceló et al., 2008) y cuya eficacia es la más baja de todo el conjunto estudiado ($W = 0,250 \pm 0,004$). En este caso, la comparación del transcriptoma de una planta inoculada con este virus y una planta sana daba lugar a sólo 1.312 DEGs, mientras que la comparación de cualquier otro genotipo con la planta sana encontraba aproximadamente 4.000 DEGs. Tal diferencia en números absolutos pudiera explicar por sí sola que **HC-Pro:AS13** forme un grupo propio. Pero aún hay más, si exploramos las listas de DEGs de todos los genotipos no sólo llama la atención el número absoluto de DEGs sino, lo que es más llamativo, el hecho de que sólo unos 260 DEGs son compartidos con los otros dos grupos ($\leq 20\%$). A pesar de que el número de DEGs es relevante, debemos pensar que el factor de mayor contribución para clasificarlo en un grupo aparte es el pequeño número de genes que comparte con el resto de genotipos. Sin embargo, si comparamos el número de DEGs en relación con el transcriptoma de plantas infectadas con el genotipo silvestre, obtenemos 2.001 DEGs. Éste es el mayor número de DEGs de todo el conjunto, que supera en más del doble el número de DEGs destacados del análisis de los otros mutantes para **HC-Pro**. De ellos, la mitad son genes sobreexpresados relacionados con estrés inespecífico, estrés biótico, defensa, respuesta inmune y muerte celular, y la otra mitad genes reprimidos relacionados con procesos metabólicos. Consideramos que el elevado número de DEGs, así

como sus funciones, puede explicar por sí sólo gran parte de la diferencia observada en la eficacia viral de **HC-Pro:AS13** y porqué es tan distinto del resto de genotipos.

Profundizando en los resultados del análisis de las categorías funcionales, cabe llamar la atención al lector sobre un término que pudiera llevar a confusión o generar controversia. Nos referimos a los **“genes determinantes de la eficacia viral”**. Recordemos que obtuvimos una lista de genes cuya expresión correlacionaba significativamente con la eficacia viral, que dividimos en **“determinantes positivos”** y **“determinantes negativos”**. Hay que puntualizar dos aspectos. El primero de ellos es remarcar que los genes “determinantes” no tienen por qué ser también genes expresados diferencialmente, al menos no en todos los mutantes. De hecho, la situación más común que nos encontramos al analizar los genes es que los mutantes situados en el extremo de menor eficacia sí poseen expresión diferencial significativa (mutantes para **HC-Pro**), mientras que los de mayor eficacia no llegan al límite necesario para considerar las diferencias como significativas. Incluso, hay casos en que ningún mutante posee expresión diferencial pero la pérdida o ganancia de eficacia se produce de forma equitativa y constante, dando lugar a la correlación. Al considerar un gen concreto que correlaciona, si calculamos la media de valores de FC para todos los mutantes, obtenemos valores en el rango 0,1489 – 5,44308. Respecto a la segunda puntualización, cuando se analizan los genes que figuran en ambas listas, se pone de manifiesto que la regulación de la mitad de ellos es realmente importante para el proceso de infección de TEV, lo que llamaríamos **“determinantes”**, mientras que la expresión de la mitad restante únicamente es una consecuencia de la infección que en ningún momento contribuye al establecimiento de la infección, lo que serían genes **“causales”**. Establezcamos un ejemplo de ambas clases.

El gen **WIN2** (HOPW1-1-INTERACTING 2) codifica una proteína con actividad fosfatasa y modula las defensas de la planta contra patógenos bacterianos. En el genoma de *Arabidopsis* se han identificado tres proteínas **WIN**, denominadas **WIN1**, **WIN2** y **WIN3**. Se ha mostrado que sólo **WIN2** y **WIN3**, cuya diana es la proteína efectora HopW1-1 de *Pseudomonas syringae* Pto DC3000, funcionan regulando la vía de señalización contra patógenos biotróficos mediante la inducción del gen **HWI1**, que activa la vía de señalización del ácido salicílico (Zhou et al., 1998). Además, **WIN2** tiene mayor especificidad por HopW1-1 que **WIN3** (Lee et al., 2008; Denancé et al., 2013). Respecto a tabaco, sólo hallamos un gen ortólogo de **WIN2** correlacionado negativamente con la eficacia viral (TA18770_4097). Esto podría sugerir que no hay otros genes similares en el tabaco, pero lo cierto es que sí los hay, aunque no están correlacionados con la eficacia viral. Si exploramos todo el genoma del tabaco encontramos otros tres genes ortólogos de **WIN1** (BP535526, TA16431_4097 y TA16432_4097) y uno más de **WIN2** (FG157464). Según esto, el genoma del tabaco tiene al menos cinco genes pertenecientes a la familia HOPW1-1-INTERACTING, pero sólo el gen TA18770_4097 se correlaciona negativamente con el aumento de la eficacia. Podemos sugerir que TEV reprime **WIN2** para evitar la activación de **HWI1** y suprimir la señalización del ácido salicílico. Con toda esta información, podemos considerar que el gen TA18770_4097 es un auténtico “**determinante negativo de la eficacia viral**”, ya que el virus trata de reprimir su expresión en beneficio propio.

Como ejemplo de gen “**causal**”, tomemos al gen **DCL2** (DICER-LIKE 2), que codifica una proteína que interviene en el mecanismo de silenciamiento de ARN, dando lugar a ta-siARNs (*trans-acting small interfering RNAs*). Sin entrar en más detalles, sólo hallamos un gen ortólogo de **DCL2** correlacionado positivamente con la eficacia viral (EB438730). Según esto, cuanto mayor es el número de viriones mayor es la expresión de este gen. Esta situación es contraproducente para el virus. Lo que ocurre en este caso es que se

produce la activación de los genes que funcionan en las defensas innatas del huésped. Así, podemos considerar que la expresión del gen **EB438730** es “**causa de la infección viral**”, ya que se trata de un gen implicado en una ruta de actuación inespecífica antiviral y el virus ni trata ni puede alterar su expresión en beneficio propio, en todo caso, intentar paliar sus efectos mediante el secuestro de siARNs que hace **HC-Pro**.

Volviendo a las listas de “**genes determinantes de la eficacia viral**”, debemos incluir aquellos genes que reflejan un beneficio para el virus (redirección de los recursos defensivos, regulación de la homeostasis celular, alteración de la expresión génica de determinados factores de transcripción, el desvío de recursos desde rutas metabólicas concretas...) bajo el término “**determinantes**”, mientras que aquellos que intervienen en procesos activados como consecuencia de una infección debida a un patógeno inespecífico, o incluso en respuestas a estreses abióticos (activación de las defensas basales, rutas complejas reguladas por hormonas, silenciamiento génico temprano, metilación...) se agrupan como “**causales**”. Por ello, no se puede asumir que exista realmente una dependencia funcional entre la eficacia viral y la expresión de un gen individual del huésped, de ahí que los consideremos como “**posibles**”.

Cabe añadir que no sólo se validaron los ocho genes aquí presentados, sino que en realidad fueron dieciséis. Sin embargo, dado el volumen de la información y las figuras asociadas a éstos sólo se muestran datos de ocho de ellos. Por ejemplo, de entre los genes cuyos datos no se muestran en este trabajo, también validamos **AGO1** (ARGONAUTE 1), que tiene un espectro muy amplio de funciones, tales como silenciamiento génico (Mallory et al., 2015; Qu et al., 2008), la formación y diferenciación de los meristemos (Kidner y Martienssen, 2005) o el desarrollo foliar (Bohmert et al., 1998). Éste poseía el valor de expresión media más bajo de entre todos los “**determinantes positivos**” validados ($FC = 1,04742$; $pendiente = 0,799311$; $r = 0,785714$). Del lado de los “**determinantes**

negativos” también se validó **RBCS3B** (RUBISCO SMALL SUBUNIT 3B) ($FC = 1,4362$; $pendiente = -1,776481$; $r = -0,857143$), el cual está implicado en la fijación de carbono durante la fotosíntesis y el mantenimiento del nivel de Rubisco (Zhan et al., 2014). No sólo se validaron genes posibles **“determinantes”** sino que también se validaron genes que no guardaban ningún tipo de correlación, es decir, genes **“no determinantes”** para los cuales no existían diferencias significativas en los datos de expresión. Entre éstos, que también resultan de utilidad para probar la validación experimental, se encuentran **RAV2** (RELATED TO ABI3/VP1 1), que codifica un factor de transcripción que es sobreexpresado en respuesta a temperaturas bajas e interviene en la floración (Matías-Hernández et al., 2014; Mittal et al., 2014), y **LHCB1.3** (LIGHT-HARVESTING CHLOROPHYLL A/B-PROTEIN 1.3), que codifica para una proteína involucrada en la fotosíntesis cuya diana es la clorofila (Rugnone et al., 2013; Mitra et al., 2009).

El gran volumen de datos generado en el transcurso de este trabajo da pie a un elevado número de aspectos que discutir, ya sea desde una perspectiva global (en que valoramos todos los perfiles transcriptómicos) o muy específica (en que valoramos un único gen). Tal es volumen que resulta imposible resumir toda esta información. Por ello, nos limitaremos a desarrollar brevemente algunos ejemplos que hemos considerado de gran interés.

Desde una perspectiva global, nos interesamos desde el primer momento en el gen que codifica la proteína **PCaP1**. Hace poco tiempo se describió un nuevo tipo de proteína en *A. thaliana* que era rica en residuos de glutamato y lisina (Ide et al., 2007). Dicha proteína, a la que se llamó **PCaP1**, era hidrófila y podía unir específicamente cationes de Ca^{2+} , incluso en presencia de altas concentraciones de otros cationes como Mg^{2+} . El análisis histoquímico reveló que se expresaba en la mayoría de los órganos y se localizaba en la membrana plasmática, aunque no poseía dominio transmembrana (Ide et al., 2007). Pudo comprobarse in vitro que

PCaP1 se asociaba con la membrana plasmática en condiciones naturales y se liberaba de ella en presencia de altas concentraciones de Ca^{2+} (Ide et al., 2007). La interpretación de dicho mecanismo de unión era más complicada de lo que parecía en un primer momento, ya que al profundizar en el mecanismo de unión se descubrió que **PCaP1** se une a la membrana plasmática mediante una reacción de miristoilación en el residuo de glicina que ocupa la posición 2 en el extremo N-terminal (Nagasaki et al., 2008). Al buscar esta secuencia en las bases de datos existentes, se constató la presencia de proteínas homólogas en varias especies, incluyendo *N. tabacum*. He aquí el interés que suscitó el estudio de la expresión del gen que codifica esta proteína, pues sabemos que la proteína viral **P3N-PIPO** se une a la proteína viral **CI**, que a su vez, se halla unida a la proteína de cubierta de un virión completo. Sin embargo, ninguno de ellos puede unirse a la membrana plasmática, sino que dicha función recae sobre la proteína **PCaP1**. Así, **PCaP1** forma un complejo con los elementos virales, se produce la miristoilación, se une de forma estable a la membrana plasmática que circunda los plasmodesmos y permite el movimiento célula a célula de los viriones (Vijayapalani et al., 2012). Desde el punto de vista del virus, sabemos que la coordinación intracelular de sus elementos (Mäkinen y Hafrén, 2014) y la variabilidad intra-específica de la proteína **P3N-PIPO** son de gran relevancia biológica durante el establecimiento de la infección (Hillung et al., 2013). Lo que no sabemos es qué ocurre desde el punto de vista de la planta. Si exploramos el genoma anotado de *N. tabacum* enseguida advertimos la presencia de tres genes ortólogos de **PCaP1** (**TA13430_4097**, **CV016893** y **EB451650**). En realidad, los dos primeros genes se corresponden con el mismo ARNm de los definidos en la base de datos de solanáceas Solgenomics (ARNm 6000). El tercero se correspondía con el ARNm 7000. De estos tres, **TA13430_4097** apareció en la lista de genes correlacionados positivamente con la eficacia viral, lo que sugiere que el virus regulaba la expresión de **PCaP1** para facilitar el movimiento célula a célula, es decir, sería un auténtico “**determinante positivo**”. Aquí se nos plantea un problema, pues

TA13430_4097 y **CV016893** remiten ambos al ARNm 6000, esto es, se trataba de dos sondas diferentes que hibridaban en dos lugares distintos del mismo ARNm. Dado el interés que nos suscitó **PCaP1** tratamos de validarlo, pero tras diseñar cuatro parejas de cebadores distintos, no obtuvimos valores de eficiencia adecuados y el ruido de fondo experimental sobrepasaba el margen de tolerancia impuesto. Sin embargo, un análisis estadístico de los datos de expresión de ambos genes concluyó que no existían diferencias significativas. Es más, al explorar los datos para **TA13430_4097** advertimos que no existía expresión diferencial (media del \log_2 -cambio en expresión = 1,0267) y la razón por la cual correlacionaba era porque la expresión subía muy poco a poco, aunque de forma constante ($r = 0,854$). Así pues, pensamos en la posibilidad de que **PCaP1** fuera un elemento imprescindible para el movimiento de célula a célula pero no el más importante. Dicho esto, cabía la posibilidad de que otros genes involucrados, como la **N-miristoiltransferasa**, pudieran expresarse diferencialmente y destacar como el elemento más importante. Desafortunadamente, aún no hemos explorado esta posibilidad a día de hoy.

Para terminar, ejemplificaremos cómo podemos obtener información del análisis individual de los mutantes. Por ejemplo, el mutante **HC-Pro:AS13** posee una lista de DEGs que apenas guarda un 20 % de similitud con las listas del resto de mutantes. La situación con el 80 % de genes restantes es la siguiente: ninguno muestra expresión diferencial en los genotipos analizados, salvo **HC-Pro:AS13**, en el que la expresión experimenta un aumento o disminución realmente drásticos. Por ejemplo, **ADF4** (ACTIN DEPOLYMERIZING FACTOR 4) interviene en la señalización defensiva innata contra bacterias (Tian et al., 2009; Day et al., 2011), nematodos (Clément et al., 2009) u hongos (Miklis et al., 2007) y **MAC3A** (MOS4-ASSOCIATED COMPLEX 3A) también es parte de la señalización de la respuesta inmune innata (Monaghan et al., 2009; Monaghan et al., 2010; Kourmpetis et al., 2011; Xu et al., 2012; Duplan y Rivas, 2014; Jones y Dangl, 2006; Abramovitch et al., 2003). Ambos genes son la diana de

efectores bacterianos como AvrPphB ([Porter et al., 2012](#); [Porter y Day, 2013](#)) y pudiéramos pensar que son regulados negativamente por el virus para afectar la cascada de señalización, que culmina en el núcleo para activar la resistencia. El hecho es que esta información es correcta únicamente para **HC-Pro:AS13**. Para el resto de genotipos, ambos genes ni se correlacionan con el aumento de la eficacia ni muestran expresión diferencial.

5. Conclusiones

En esta tesis se han estudiado los factores que configuran los paisajes de eficacia viral y los genes determinantes de dicha eficacia. El gran volumen de información que hemos obtenido en este trabajo se ha cimentado sobre el empleo de una colección de genotipos y subclones biológicos individuales, lo que ha requerido de un gran componente estadístico y bioinformático para comprenderla.

Los resultados obtenidos pueden resumirse en las siguientes conclusiones:

1. En un paisaje rugoso, las poblaciones virales pueden alejarse de los picos de eficacia para explorar regiones distantes del mismo, incluso aunque se sitúen inicialmente muy próximas a un máximo local de eficacia.
2. El factor de mayor contribución al movimiento en los paisajes de eficacia es la adaptación. La contribución de los factores historia y azar, aunque mucho menor, también resulta significativa.
3. El cambio de la especie huésped da lugar a perturbaciones en la topografía del paisaje de eficacia.
4. El paisaje definido por el huésped original *N. tabacum* es más rugoso que el definido por el huésped nuevo *A. thaliana*, principalmente debido a genotipos que cambian su altura relativa en el paisaje y al tipo de interacciones epistáticas con sus vecinos.
5. Los paisajes comparten propiedades macroscópicas, como la existencia de genotipos letales y una topografía rugosa, pero difieren en detalles microscópicos principalmente debido a la existencia de muchos cambios de epistasia de magnitud o signo.

Conclusiones

6. Las variaciones de eficacia de los componentes subclonales son consecuencia de una gran variabilidad genética, siendo diferentes las mutaciones que se fijan en cada uno.
7. La mayoría de las mutaciones aleatorias generadas durante la replicación viral tienen un efecto negativo en la eficacia viral de los componentes subclonales.
8. La replicación propensa a errores da lugar a enjambres de mutantes, cuya eficacia global disminuye, sobre los que opera la selección positiva para restaurar el valor de eficacia media, desapareciendo las diferencias genéticas entre subclones tras largos periodos de infección.
9. Una o varias mutaciones en el genoma viral pueden conducir a una interacción sub-óptima con uno o unos pocos elementos que pueden alterar de forma radical la red de interacciones entre el patógeno y su huésped.
10. La expresión de los genes de la planta infectada se divide entre los relacionados con la respuesta al estrés ("**causales**") y los relacionados con metabolismo y desarrollo ("**determinantes**").
11. Los genotipos virales se clasifican en grupos claramente diferenciados y ordenados de acuerdo al grado de solapamiento del efecto que generan sobre la expresión génica de la planta y sus valores de correlación asociados, coincidiendo con el orden de los valores de eficacia.

6. Publicaciones

Cervera H. and Elena S.F. (2016). Genetic variation in fitness within a clonal population of a plant RNA virus. *Virus Evol.* 2: vew006. (doi.org/10.1093/ve/vew006)

Cervera H., Lalić J. and Elena S.F. (2016). Efficient escape from local optima in a highly rugged fitness landscape by evolving RNA virus populations. *Proc. R. Soc. B.* 283: 20160984. (doi.org/10.1098/rspb.2016.0984)

Cervera H., Lalić J. and Elena S.F. (2016). Effect of host species on the topography of fitness landscape for a plant RNA virus. *J. Virol.* 90: 10160-10169. (doi.org/10.1128/JVI.01243-16)

Cervera H., Ambrós S., Bernet G.P., Rodrigo G. and Elena S.F. (2017). Viral fitness contains high information about the virus-host interaction. En preparación.

7. Referencias

- Abramovitch R., Kim Y., Chen S., Dickman M. and Martin G. (2003). Pseudomonas type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. The EMBO journal, Vol. 22 No. 1 pp. 60-69. (doi.org/10.1093/emboj/cdg006)
- Agrawal A.F. and Lively C.M. (2003). Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proc Biol Sci 270: 323–334. (doi.org/10.1098/rspb.2002.2193)
- Agudelo-Romero P., Carbonell P., Pérez-Amador M.A. and Elena, S.F. (2008a). Virus adaptation by manipulation of host's gene expression. PLoS ONE 3, e2397.
- Agudelo-Romero P., De la Iglesia F. and Elena S.F. (2008b). The pleiotropic cost of host specialization in tobacco etch potyvirus. Infection Genetics and Evolution, 8: 806–14.
- Ahlquist P., Noueiry A.O., Lee W.M., Kushner D.B. and Dye B.T. (2003). Host factors in positive-strand RNA virus genome replication. J. Virol. 77:8181-8186.
- Aita T., Iwakura M. and Husimi Y. (2001). A cross-section of the fitness landscape of dihydrofolate reductase. Protein Eng 14:633–638. (doi.org/10.1093/protein/14.9.633)
- Aita T., Uchiyama H., Inaoka T., Nakajima M., Kokubo T. and Husimi Y. (2000). Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysis. Biopolymers 54:64–79.
- Ali A. and Roossinck M.J. (2010). Genetic bottlenecks during systemic movement of cucumber mosaic virus vary in different host plants. Virology, 404: 279–83.

Referencias

- Alonso R., Salavert F., Garcia-Garcia F., Carbonell-Caballero J., Bleda M., Garcia-Alonso L., Sanchis-Juan A., Perez-Gil D., Marín-Garcia P., Sanchez R., Cubuk C., Hidalgo M.R., Amadoz A., Hernansaiz-Ballesteros R.D., Alemán A., Tarraga J., Montaner D., Medina I. and Dopazo, J. (2015). BABELOMICS 5.0 functional interpretation for new generations of genomic data. *Nucl Acids Res* 43, W117–W121. (doi.org/10.1093/nar/gkv384)
- Anderson P.K., Cunningham A.A., Patel N.G., Morales F.J., Epstein P.R. and Daszak P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19: 535–544.
- Bailer S.M. and Haas J. (2009). Connecting viral with cellular interactomes. *Curr Opin Microbiol* 12, 453-459.
- Baltimore D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35: 235–241.
- Barabási A.L., Gulbahce N. and Loscalzo J. (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12, 56-68. (doi.org/10.1038/nrg2918)
- Bedhomme S., Hillung J. and Elena S.F. (2015). Emerging viruses: why they are not jacks of all trades? *Curr Opin Virol* 10:1–6. (doi.org/10.1016/j.coviro.2014.10.006)
- Bedhomme S., Lafforgue G. and Elena S.F. (2012). Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Molecular Biology and Evolution*, 29: 1481–92.
- Bedhomme S., Lafforgue G. and Elena S.F. (2013). Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. *BMC Evol. Biol.* 13, 46. (doi.org/10.1186/1471-2148-13-46)

- Bedoya L. and Daròs J.A. (2010). Stability of tobacco etch virus infectious clones in plasmid vectors. *Virus Research* 149, 234–240.
- Belshaw R., Pybus O.G. and Rambaut A. (2007). The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17: 1496–1504. (doi.org/10.1101/gr.6305707)
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57, 289-300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 29, 1165-1188.
- Bernet G.P. and Elena S.F. (2015). Distribution of mutational fitness effects and of epistasis in the 5' untranslated region of a plant RNA virus. *BMC Evol. Biol.* 15: 274. (doi.org/10.1186/s12862-015-0555-2)
- Berns K.I. (1990). Parvovirus replication. *Microbiological Reviews*, p. 316-329.
- Betancourt M., Fereres A., Fraile A. and García-Arenal F. (2008). Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *Journal of Virology*, 82: 12416–21.
- Blanc S., Drucker M. and Uzeit M. (2014). Localizing viruses in their insect vectors. *Annu. Rev. Phytopathol.* 52:403–25. (doi.org/10.1146/annurev-phyto-102313-045920)
- Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M. and Benning C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* 17:170-180.

Referencias

- Bouquin T., Meier C., Foster R., Nielsen M.E. and Mundy J. (2001). Control of specific gene expression by gibberellin and brassinosteroid. *Plant Physiol* 127, 450-458.
- Boyes D.C., Zayed A.M., Ascenzi R., McCaskill M.J., Hoffman N.E., Davis K.R. and Görlach J. (2001). Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510. (doi.org/10.1105/tpc.13.7.1499)
- Brandes N. and Linial M. (2016). Gene overlapping and size constraints in the viral world. *Biology Direct.* 11:26. (doi.org/10.1186/s13062-016-0128-3)
- Brault V., Uzest M., Monsion B., Jacquot E. and Blanc S. (2010). Aphids as transport devices for plant viruses. *C R Biol* 333, 524-38. (doi.org/10.1016/j.crv.2010.04.001)
- Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J. and Vingron M. (2001). Minimum information about a microarray experiment (MIAME) - Toward standards for microarray data. *Nat. Genet.* 29: 365–371.
- Brinton M.A., Calisher C.H. and Rueckert R. (1994). *Positive-Strand RNA Viruses*. Springer-Verlag.
- Brouillet S., Annoni H., Ferreti L. and Achaz G. (2016). MAGELLAN: a tool to explore small fitness landscapes. *bioRxiv* 2016:031583.
- Bull J.J., Sanjuán R. and Wilke C.O. (2007). Theory of lethal mutagenesis for viruses. *Journal of Virology*, 81: 2930–39.

- Burch C.L. and Chao L. (1999). Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics* 151: 921–927.
- Butcher D. (1995). Muller’s ratchet, epistasis and mutation effects. *Genetics* 141: 431–7.
- Callaway A., Giesman-Cookmeyer D., Gillock E.T., Sit T.L. and Lommel S.A. (2001). The multifunctional capsid proteins of plant RNA viruses. *Annu. Rev. Phytopathol.* 39: 419–60. (doi.org/10.1146/annurev.phyto.39.1.419)
- Campillo-Balderas J.A., Lazcano A. and Becerra A. (2015). Viral Genome Size Distribution Does not Correlate with the Antiquity of the Host Lineages. *Front. Ecol. Evol.* 3:143. (doi.org/10.3389/fevo.2015.00143)
- Cann A.J. (2005). *Principles of Molecular Virology*. Academic Press, San Diego, California, USA.
- Carrasco P., Daròs J.A., Agudelo-Romero P. and Elena S.F. (2007a). A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J. Virol. Methods* 139:181–188.
- Carrasco P., De la Iglesia F. and Elena S.F. (2007b). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in tobacco etch virus. *Journal of Virology*, 81: 12979–84.
- Carrington J.C., Freed D.D. and Sanders T.C. (1989). Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and in vitro. *Journal of Virology*. 63(10):4459-4463.

Referencias

- Carrington J.C., Haldeman R., Dolja V.V. and Restrepo-Hartwig M.A. (1993). Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of Tobacco etch potyvirus in vivo. *J. Virol.* 67: 6995–7000.
- Chao L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348: 454–5.
- Clarke D.K., Duarte E.A., Moya A., Elena S.F., Domingo E. and Holland J. (1993). Genetic bottleneck and population passages cause profound fitness differences in RNA viruses. *Journal of Virology*, 67: 222–8.
- Claverie J.M. and Abergel C. (2009). Mimivirus and its Virophage. *Annual Review of Genetics*, 43: 49–66. (doi.org/10.1146/annurev-genet-102108-134255)
- Clément M., Ketelaar T., Rodiuc N., Banora M.Y., Smertenko A., Engler G., Abad P., Hussey P.J. and de Almeida Engler J. (2009). Actin-Depolymerizing Factor2-mediated actin dynamics are essential for rootknot nematode infection of Arabidopsis. *Plant Cell* 21: 2963–2979, doi: 10.1105/tpc.109.069104
- Coffey L.L., Beeharry Y., Bordería A.V., Blanc H. and Vignuzzi M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proceedings of the National Academy of the USA*, 108: 16038–43.
- Cui H. and Wang A. (2016). Plum pox virus 6K1 protein is required for viral replication and targets the viral replication complex at the early stage of infection. *J Virol* 90:5119–5131. (doi.org/10.1128/JVI.00024-16)
- Da Silva J. (2010). An adaptive walk by human immunodeficiency virus type 1 through a fluctuation fitness landscape. *Evolution* 64:1160–1165. (doi.org/10.1111/j.1558-5646.2009.00885.x)

- Da Silva J. and Wyatt S.K. (2014). Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. *J Evol Biol* 27:604–615. (doi.org/10.1111/jeb.12329)
- Da Silva J., Coetzer M., Nedellec R., Pastore C. and Mosier D.E. (2010). Fitness epistasis and constraints on adaptation in a Human immunodeficiency virus type 1 protein region. *Genetics* 185:293–303. (doi.org/10.1534/genetics.109.112458)
- Day B., Henty J.L., Porter K.J. and Staiger C.J. (2011). The pathogen-actin connection: A platform for defense signaling in plants. *Annu Rev Phytopathol* 49: 483–506, doi: 10.1146/annurev-phyto-072910-095426
- De la Iglesia F. and Elena S. F. (2007). Fitness declines in tobacco etch virus upon serial bottleneck transfers. *Journal of Virology*, 81: 4941–7.
- De Visser J.A.G.M. and Krug J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* 15:480–490. (doi.org/10.1038/nrg3744)
- De Visser J.A.G.M., Cooper T.F. and Elena S.F. (2011). The causes of epistasis. *Proc Biol Sci* 278:3617–3624. (doi.org/10.1098/rspb.2011.1537)
- Dekker E.L., Derks A.F., Asjes C.J., Lemmers M.E., Bol J.F. and Langeveld S.A. (1993). Characterization of potyviruses from tulip and lily which cause flower-breaking. *Journal of General Virology*. 74(5), 881–887. (doi.org/10.1099/0022-1317-74-5-881)
- Denancé N., Sánchez-Vallet A., Goffner D. and Molina A. (2013). Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. *Front Plant Sci.*, May 24;4:155. (doi.org/10.3389/fpls.2013.00155)

Referencias

- Denyer K., Clarke B., Hylton C., Tatge H. and Smith A.M. (1996). The elongation of amylose and amylopectin chains in isolated starch granules. *Plant J* 10, 1135-1143. (doi.org/10.1046/j.1365-313X.1996.10061135.x)
- Domingo E. (2010). Mechanisms of viral emergence. *Vet. Res.* 41:38. (doi.org/10.1051/vetres/2010010)
- Domingo E. and Holland J.J. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51: 151–178.
- Domingo E. and Wain-Hobson S. (2009). The 30th anniversary of quasispecies. Meeting on “Quasispecies: past, present and future”. *EMBO reports* 10, 444-448. (doi.org/10.1038/embor.2009.61)
- Domingo E., Sheldon J. and Perales C. (2012). Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* 76: 159–216. (doi.org/10.1128/MMBR.05023-11)
- Dougherty W.G., Carrington J.C., Cary S.M. and Parks T.D. (1988). Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* 7:1281-1287.
- Drake J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbe. *Proc. Natl. Acad. Sci. USA* 88: 7160–7164.
- Drake J.W., Charlesworth B., Charlesworth D. and Crow J.F. (1998). Rates of spontaneous mutation. *Genetics* 148: 1667–1686.
- Duarte E.A., Clarke D.K., Moya A., Domingo E. and Holland J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller’s ratchet. *Proceedings of the National Academy of Sciences of the USA*, 89: 6015–9.

- Duarte E.A., Novella I.S., Ledesma S., Clarke D.K., Moya A., Elena S.F., Domingo E. and Holland J.J. (1994). Subclonal components of consensus fitness in an RNA virus clone. *Journal of Virology*, 68: 4295–301.
- Duplan V. and Rivas S. (2014). E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. *Frontiers in plant sciences*, Vol. 5, Article 42. (doi.org/10.3389/fpls.2014.00042)
- Ebert D. (1998). Experimental evolution of parasites. *Science* 282: 1432–1435.
- Eckardt N.A. (2003). Viral Defense and Counterdefense: A Role for Adenosine Kinase in Innate Defense and RNA Silencing. *Plant Cell*, 15(12): 2758–2761. (doi.org/10.1105/tpc.151211)
- Edwards R.A. and Rohwer F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3: 504–510.
- Elena S. F., Carrasco P., Daròs J.A. and Sanjuán R. (2006). Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* 7:168-173.
- Elena S.F. and Lalić J. (2013). Plant RNA virus fitness predictability: contribution of genetic and environmental factors. *Plant Pathol* 62:S10 –S18.
- Elena S.F. and Rodrigo G. (2012). Towards an integrated molecular model of plant-virus interactions. *Curr Opin Virol* 2:719–724. (doi.org/10.1016/j.coviro.2012.09.004)
- Elena S.F. and Sanjuán R. (2005). Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of Virology*, 79: 11555–8. (doi.org/10.1128/JVI.79.18.11555-11558.2005)

Referencias

- Elena S.F. and Sanjuán R. (2008). The effect of genetic robustness on evolvability in digital organisms. *BMC Evol. Biol.* 8: 284.
- Elena S.F., Agudelo-Romero P., Carrasco P., Codoñer F.M., Martín S., Torres-Barceló C. and Sanjuán R. (2008). Experimental evolution of plant RNA viruses. *Heredity* 100: 478–483. (doi.org/10.1038/sj.hdy.6801088)
- Elena S.F., Bedhomme S., Carrasco P., Cuevas J.M., de la Iglesia F., Lafforgue G., Lalić J., Pròsper A., Tromas N. and Zwart M.P. (2011). The evolutionary genetics of emerging plant RNA viruses. *Mol Plant Microbe Interact.* 24(3):287-93. (doi.org/10.1094/MPMI-09-10-0214)
- Fabré F., Moury B., Johansen E.I., Simon V., Jacquemond M. and Senoussi R. (2014). Narrow bottlenecks affect pea seedborne mosaic virus populations during vertical seed transmission but not during leaf colonization. *PLoS Pathogens*, 10: e4003833.
- Ferretti L., Schmiegelt B., Weinreich D.M., Yamauchi A., Kobayashi Y., Tajima F. and Achaz G. (2016). Measuring epistasis in fitness landscapes: the correlation of fitness effects of mutations. *J Theor Biol* 396:132–143. (doi.org/10.1016/j.jtbi.2016.01.037)
- Fields B.N., Knipe D.M. and Howley P.M. (1996). *Fundamental Virology*. Lippincott-Raven.
- Finzer P. (2017). How we become ill. *EMBO Rep* 18, 515-518.
- Fisher R.A. (1930). *The Genetical Theory of Natural Selection*. Oxford, Clarendon Press.
- Flenniken M.L. (2014). Honey bee-infecting plant virus with implications on honey bee colony health. *MBio*. 5:e00877-14. (doi.org/10.1128/mBio.00898-13)

- Flint J., Racaniello V.R., Rall G.F. and Skalka A.M. (2015). Principles of Virology. Fourth Edition. ASM Press, Washington, DC.
- Flynn K.M., Cooper T.F., Moore F.B.G. and Cooper V.S. (2013). The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet* 9:e1003426. (doi.org/10.1371/journal.pgen.1003426)
- French R. and Stenger D.C. (2005). Population structure within lineages of wheat streak mosaic virus derived from a common founding event exhibit stochastic variation inconsistent with the deterministic quasi-species model. *Virology*, 343: 179–89.
- Friedel C.C. and Haas J. (2011). Virus-host interactomes and global models of virus-infected cells. *Trends Microbiol* 19, 501-508.
- Fry J.D. (1996). The evolution of host specialization: are trade-offs overrated? *Am Nat* 148:S84–S107. (doi.org/10.1086/285904)
- Gabriel W., Lynch M. and Bürger R. (1993). Muller’s ratchet and mutational meltdowns. *Evolution* 47: 1744–57.
- Gago S., Elena S.F., Flores R. and Sanjuán R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* 323: 1308.
- Gao H., Granka J.M. and Feldman M.W. (2010). On the classification of epistatic interactions. *Genetics* 184:827–837. (doi.org/10.1534/genetics.109.111120)
- García-Díaz M. and Bebenek K. (2007). Multiple functions of ADN polymerases. *CRC Crit. Rev. Plant Sci.* 26, 105–122. (doi.org/10.1080/07352680701252817)
- Gavrilets S. (2010). High-dimensional fitness landscapes and the origins of biodiversity. *Toward an Extended Evolutionary Synthesis*. MIT Press, Cambridge, MA.

Referencias

- Geng C., Wang H.Y., Liu J., Yan Z.Y., Tian Y.P., Yuan X.F., Gao R. and Li X.D. (2016). Transcriptomic changes in *Nicotiana benthamiana* plants inoculated with the wildtype or an attenuated mutant of Tobacco vein banding mosaic virus. *Molecular Plant Pathology*. (doi.org/10.1111/mpp.12471)
- González-Jara P., Fraile A., Canto T. and García-Arenal F. (2009). The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *Journal of Virology*, 83: 7487–94.
- Gutiérrez J. and Maere S. (2014). Modeling the evolution of molecular systems from a mechanistic perspective. *Trends in Plant Science*. Vol.19, No.5. (doi.org/10.1016/j.tplants.2014.03.004)
- Gutiérrez S., Piroilles E., Yvon M., Baecker V., Michalakis Y. and Blanc S. (2015). The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *Journal of Virology*, 89: 9665–75.
- Gutiérrez S., Yvon M., Piroilles E., Garzo E., Fereres A., Michalakis Y. and Blanc S. (2012). Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. *PLoS Pathogens*, 8: e1003009.
- Gutiérrez S., Yvon M., Thébaud G., Monsion B., Michalakis Y. and Blanc S. (2010). Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathogens*, 6: e1001113.
- Hall J.S., French R., Hein G.L., Morris T.J. and Stenger D.C. (2001a). Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology*, 282: 230–6.

- Hall J.S., French R., Morris T.J. and Stenger D.C. (2001b). Structure and temporal dynamics of populations within wheat streak mosaic virus isolates. *Journal of Virology*, 75: 10231–243.
- Hillung J. (2015). Viral emergence as a consequence of the interaction between the genetic variabilities of the virus and host. (roderic.uv.es/handle/10550/47826)
- Hillung J., Cuevas J.M. and Elena S.F. (2012). Transcript profiling of different *Arabidopsis thaliana* ecotypes in response to tobacco etch potyvirus infection. *Front Microbiol* 3:229. (doi.org/10.3389/fmicb.2012.00229)
- Hillung J., Cuevas J.M. and Elena S.F. (2015). Evaluating the within-host fitness effects of mutations fixed during virus adaptation to different ecotypes of a new host. *Philos Trans R Soc Lond B Biol Sci* 370:20140292. (doi.org/10.1098/rstb.2014.0292)
- Hillung J., Cuevas J.M., Valverde S. and Elena S.F. (2014). Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution* 68:2467–2480. (doi.org/10.1111/evo.12458)
- Hillung J., Elena S.F. and Cuevas J.M. (2013). Intra-specific variability and biological relevance of P3N-PIPO protein length in potyviruses. *BMC Evol Biol.*; 13: 249. (doi.org/10.1186/1471-2148-13-249)
- Hillung J., García-García F., Dopazo J., Cuevas J.M. and Elena S.F. (2016). The transcriptomics of an experimentally evolved plant-virus interaction. *Sci Rep* 6:24901. (doi.org/10.1038/srep24901)
- Holland J.J., de la Torre J.C., Clarke D.K. and Duarte E. (1991). Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J Virol* 65, 2960-2967.

Referencias

- Holmes E.C. (2009). The evolutionary genetics of emerging viruses. *Annu. Rev. Ecol. Evol. Syst.* 40: 353–72. (doi.org/10.1146/annurev.ecolsys.110308.120248)
- Holmes, E.C. (2003). Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* 11: 543–546.
- Hull, R. (2002) *Plant Virology* (4th edition). Academic Press, London, UK.
- Hyun T. K., Albacete A., van der Graaff E., Eom S.H., Großkinsky D.K., Böhm H., Janschek U., Rim .Y, Ali W.W., Kim S.Y. and Roitsch T. (2015). The Arabidopsis PLAT domain protein 1 promotes abiotic stress tolerance and growth in tobacco. *Transgenic Res* 24, 651-663. (doi.org/10.1007/s11248-015-9868-6)
- Ide Y., Nagasaki N., Tomioka R., Suito M., Kamiya T. and Maeshima M. (2007). Molecular properties of a novel, hydrophilic cation-binding protein associated with the plasma membrane. *J Exp Bot* 58 (5): 1173-1183. (doi.org/10.1093/jxb/erl284)
- Jain K. and Krug J. (2007). Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics* 175, 1275–1288. (doi.org/10.1534/genetics.106.067165)
- Jones J.D. and Dangl J.L. (2006). The plant immune system. *Nature* 444, 323–329. (doi.org/10.1038/nature05286)
- Kauffman S.A. and Levin S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128, 11–45. (doi.org/10.1016/S0022-5193(87)80029-2)
- Kidner C.A. and Martienssen R.A. (2005). The role of ARGONAUTE1 (AGO1) in meristem formation and identity. *Dev. Biol.* 280:504-517. (doi.org/10.1016/j.ydbio.2005.01.031)

- Kimura M. (1983). Diffusion model of intergroup selection, with special reference to evolution of an altruistic character. *Proc. Natl. Acad. Sci. U.S.A.* 80:6317-6321.
- King A.M.Q., Adams M.J., Carstens E.B. and Lefkowitz E.J. (2012). Virus taxonomy: classification and nomenclature of viruses. In: Ninth Report of the International Committee on Taxonomy of Viruses, Elsevier, San Diego, USA.
- King A.M.Q., Lefkowitz E., Adams M.J. and Carstens E.B. (2011). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
- Kingman J. (1987). A simple model for the balance between selection and mutation. *J Appl Probab* 15:1–12.
- Kleczkowski A. (1950). Interpreting relationships between the concentrations of plant viruses and numbers of local lesions. *Journal of General Microbiology*, 4: 53–69.
- Kondrashov A.S. (1994). Muller’s ratchet under epistatic selection. *Genetics*, 136: 1469–73.
- Koonin E.V. (2012). The logic of chance: the nature and origin of biological evolution. FT Press Science, NJ, USA.
- Koonin E.V. and Dolja V.V. (2013). A virocentric perspective on the evolution of life. *Curr Opin Virol.* 3(5): 546–557. (doi.org/10.1016/j.coviro.2013.06.008)
- Kourmpetis Y., Dijk A., Ham R. and Braak C. (2011). Genome-Wide Computational Function Prediction of Arabidopsis Proteins by Integration of Multiple Data Sources. *Plant Physiology* vol. 155 no. 1 271-281. (doi.org/10.1104/pp.110.162164)

Referencias

- Kouyos R.D., Leventhal G.E., Hinkley T., Haddad M., Whitcomb J.M., Petropoulos C.J. and Bonhoeffer S. (2012). Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genet* 8:e1002551. (doi.org/10.1371/journal.pgen.1002551)
- Kouyos R.D., Silander O.K. and Bonhoeffer S. (2007). Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol Evol* 22:308–315. (doi.org/10.1016/j.tree.2007.02.014)
- Kvitek D.J. and Sherlock G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7: e1002056. (doi.org/10.1371/journal.pgen.1002056)
- La Scola B., Desnues C., Pagnier I., Robert C., Barrassi L., Fournous G., Merchat M., Suzan-Monti M., Forterre P., Koonin E. and Raoult D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100-104. (doi.org/10.1038/nature07218)
- Lalić J. (2012). Nonlinearities in plant RNA virus fitness. (roderic.uv.es/handle/10550/25037)
- Lalić J. and Elena S.F. (2012). Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity*, 109: 71–77.
- Lalić J. and Elena S.F. (2013). Epistasis between mutations is host-dependent for an RNA virus. *Biol Lett* 9:20120396. (doi.org/10.1098/rsbl.2012.0396)
- Lalić J. and Elena S.F. (2015). The impact of high-order epistasis in the within-host fitness of a positive sense plant RNA virus. *J. Evol. Biol.* 28, 2236–2247. (doi.org/10.1111/jeb.12748)

- Lalić J., Agudelo-Romero P., Carrasco P. and Elena S.F. (2010). Adaptation of tobacco etch potyvirus to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. *Phil. Trans. R. Soc. B* 365, 1997–2008. (doi.org/10.1098/rstb.2010.0044)
- Lalić J., Cuevas J.M. and Elena S.F. (2011). Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet* 7:e1002378. (doi.org/10.1371/journal.pgen.1002378)
- Laughlin D.C. and Messier J. (2015). Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends Ecol Evol* 30:487–496. (doi.org/10.1016/j.tree.2015.06.003)
- Lauring A.S., Frydman J. and Andino R. (2013). The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology* 11, 327–336 (doi.org/10.1038/nrmicro3003)
- Lee H., Suh S.S., Park E., Cho E., Ahn J.H., Kim S.G., Lee J.S., Kwon Y.M. and Lee I. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev* 14, 2366–2376.
- Lee M.W., Jelenska J. and Greenberg J.T. (2008). *Arabidopsis* proteins important for modulating defense responses to *Pseudomonas syringae* that secrete HopW1-1. *The Plant Journal*, 54: 452–465. (doi.org/10.1111/j.1365-313X.2008.03439.x)
- Lenski R. E. (1991). Quantifying fitness and gene stability in microorganisms. *Biotechnology* 15:173-192.
- Levine, A.J. (1991). *Viruses*. Scientific American Press Library.
- Levy J.A., Fraenkel-Conrat H.F. and Owens R.A. (1994). *Virology*, 3d ed. Prentice Hall.

Referencias

- Li J.L., Cornman R.S., Evans J.D., J. S. Pettis J.S., Zhao Y., Murphy C., Peng W.J., Wu J., Hamilton M., Boncristiani H.F., Zhou L., Hammond J. and Chen Y.P. (2014). Systemic spread and propagation of a plant-pathogenic virus in European honeybees, *Apis mellifera*. MBio. 5:e00898-13. (doi.org/10.1126/science.1239181)
- Li Y., Jing Y., Li J., Xu G. and Lin R. (2014). Arabidopsis VQ MOTIF-CONTAINING PROTEIN 29 represses seedling deetiolation by interacting with PHYTOCHROME-INTERACTING FACTOR 1. Plant Physiol 164, 2068-2080. (doi.org/10.1104/pp.113.234492)
- Linnen, C.R. and Hoekstra H.E. (2010). Measuring natural selection on genotypes and phenotypes in the wild. Cold Spring Harb Symp Quant Biol 74, 155-168.
- Llave C., Kasschau K.D. and Carrington J.C. (2000). Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. Proc. Natl. Acad. Sci. USA 97: 13401–13406. (doi.org/10.1073/pnas.230334397)
- Lynch M. (2006). Streamlining and simplification of microbial genome architecture. Annu. Rev. Microbiol. 60: 327–349. (doi.org/10.1146/annurev.micro.60.080805.142300)
- Lynch M. and Gabriel W. (1990). Mutational load and the survival of small populations. Evolution, 44: 1725–37.
- Lynch M., Bürger R., Butcher D. and Gabriel W. (1993). The mutational meltdown in asexual populations. Journal of Heredity, 84: 339–44.
- Mäkinen K. and Hafrén A. (2014). Intracellular coordination of potyviral RNA functions in infection. Front Plant Sci. 2014; 5: 110. (doi.org/10.3389/fpls.2014.00110)

- Mallory A.C., Hinze A., Tucker M.R., Bouché N., Gascioli V., Elmayan T., Laressergues D., Jauvion V., Vaucheret H. and Laux T. (2009). Redundant and Specific Roles of the ARGONAUTE Proteins AGO1 and ZLL in Development and Small RNA-Directed Gene Silencing. *PLoS Genet* 5(9): e1000646. (doi.org/10.1371/journal.pgen.1000646)
- Martin G., Elena S.F. and Lenormand T. (2007). Distribution of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet* 39: 555–560. (doi.org/10.1038/ng1998)
- Martínez F., Rodrigo G., Aragonés V., Ruiz M., Lodewijk I., Fernández U., Elena S.F. and Daròs J.A. (2016). Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. *BMC Genomics* 17: 87.
- Martínez F., Sardanyés J., Elena S.F. and Daròs J.A. (2011). Dynamics of a plant RNA virus intracellular accumulation: stamping machine vs. geometric replication. *Genetics* 188:637– 646. (doi.org/10.1534/genetics.111.129114)
- Matías-Hernández L., Aguilar-Jaramillo A.E., Marín-González E., Suárez-López P. and Pelaz S. (2014). RAV genes: regulation of floral induction and beyond. *Ann Bot.* 114(7):1459-70. (doi.org/10.1093/aob/mcu069)
- Matuszewski S., Hermisson J. and Kopp M. (2014). Fisher’s geometric model with a moving optimum. *Evolution* 68:2571–2588. (doi.org/10.1111/evo.12465)
- McDonald B.A. and Linde C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40: 349–379.

Referencias

- Meyer J.R., Dobias D.T., Weitz J.S., Barrick J.E., Quick R.T. and Lenski R.E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335, 428–432. (doi.org/10.1126/science.1214449)
- Mi H., Huang X., Muruganujan A., Tang H., Mills C., Kang D. and Thomas P.D. (2016). PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucl Acids Res* 45, D183-D189.
- Miklis M., Consonni C., Bhat R.A., Lipka V., Schulze-Lefert P., et al. (2007). Barley MLO modulates actin-dependent and actin-independent antifungal defense pathways at the cell periphery. *Plant Physiol* 144: 1132–1143. (doi.org/10.1104/pp.107.098897)
- Miralles R., Moya A. and Elena S.F. (1997). Is group selection a factor modulating the virulence of RNA viruses? *Genet. Res.* 69:165-172.
- Mitra A., Han J. and Zhang Z.J. (2009). The intergenic region of *Arabidopsis thaliana* cab1 and cab2 divergent genes functions as a bidirectional promoter. *Planta*. 229(5):1015-22. (doi.org/10.1007/s00425-008-0859-1)
- Mittal A., Gampala S.S., Ritchie G.L., Payton P., Burke J.J., Payton P. and Rock C.D. (2014). Related to ABA-Insensitive3(ABI3)/Viviparous1 and AtABI5 transcription factor coexpression in cotton enhances drought stress adaptation. *Plant Biotechnol J.* 12(5):578-89. (doi.org/10.1111/pbi.12162)

- Miyashita S. and Kishino H. (2010). Estimation of the size of genetic bottlenecks in cell-to cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *Journal of Virology*, 84: 1828–37
- Modrow S., Falke D., Truyen U. and Schätzl H. (2013). *Molecular virology. Chapter 14 - Viruses with Single-Stranded, Positive-Sense RNA Genomes.* pp 185-349. (doi.org/10.1007/978-3-642-20718-1_14)
- Monaghan J., Germain H., Weihmann T. and Li X. (2010). Dissecting plant defence signal transduction: modifiers of *snc1* in *Arabidopsis*. *Canadian Journal of Plant Pathology*, 32:1, 35-42. (doi.org/10.1080/07060661003621001)
- Monaghan J., Xu F., Gao M., Zhao Q., Palma K., Long C., Chen S., Zhang Y. and Li X. (2009). Two Prp19-Like U-Box Proteins in the MOS4-Associated Complex Play Redundant Roles in Plant Innate Immunity. *PLoS Pathog* 5(7): e1000526. (doi.org/10.1371/journal.ppat.1000526)
- Morley V.J., Mendiola S.Y. and Turner P.E. (2015). Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proc. R. Soc. B* 282, 20150801. (doi.org/10.1098/rspb.2015.0801)
- Moury B., Fabré F. and Senoussi R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proceedings of the National Academy of the USA*, 104: 17891–6.
- Mustonen V. and Lässig M. (2009). From fitness landscapes to seascapes: nonequilibrium dynamics of selection and adaptation. *Trends Genet* 25:111–119. (doi.org/10.1016/j.tig.2009.01.002)

Referencias

- Nagasaki N., Tomioka R. and Maeshima M. (2008). A hydrophilic cation-binding protein of *Arabidopsis thaliana*, AtPCaP1, is localized to plasma membrane via N-myristoylation and interacts with calmodulin and the phosphatidylinositol phosphates PtdIns(3,4,5)P(3) and PtdIns(3,5)P(2). *FEBS Journal*. May; 275 (9): 2267 - 82. (doi.org/10.1111/j.1742-4658.2008.06379.x)
- Novella I.S., Clarke D.K., Quer J., Duarte E.A., Lee C.H., Weaver S.C., Elena S.F., Moya A., Domingo E. and Holland J.J. (1995b). Extreme fitness differences in mammalian and insect hosts after continuous replication of Vesicular stomatitis virus in sandfly cells. *J Virol* 69, 6805-6809.
- Novella I.S., Duarte E.A., Elena S.F., Moya A., Domingo E. and Holland J.J. (1995a). Exponential increases of RNA virus fitness during large populations transmissions. *Proc Natl Acad Sci USA* 92, 5841-5844.
- Novella I.S., Gilbertson D.L., Borrego B., Domingo E. and Holland J.J. (2004). Adaptability costs in immune escape variants of Vesicular stomatitis virus. *Virus Res* 107, 27-34. (doi.org/10.1016/j.virusres.2004.06.007)
- O'Brien J. A., Daudi A., Finch P., Butt V.S., Whitelegge J.P., Souda P., Ausubel F.M. and Bolwell G.P. (2012). A peroxidase-dependent apoplastic oxidative burst in cultured *Arabidopsis* cells functions in MAMP-elicited defense. *Plant Physiol* 158, 2013-2027. (doi.org/10.1104/pp.111.190140)
- Ohta T. and Ina. Y. (1995). Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. *J. Mol. Evol.* 41:717-720.

- Orr H.A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935–949.
- Orr H.A. (2009). Fitness and its role in evolutionary genetics. *Nat Rev Genet* 10, 531-539.
- Pagán I., Alonso-Blanco C. and García-Arenal F. (2008). Host responses in life-history traits and tolerance to virus infection in *Arabidopsis thaliana*. *PLoS Pathog* 4:e1000124. (doi.org/10.1371/journal.ppat.1000124)
- Parent J.S., Bouteiller N., Elmayan T. and Vaucheret H. (2015). Respective contributions of *Arabidopsis* DCL2 and DCL4 to RNA silencing. *Plant J*, 81: 223–232. (doi.org/10.1111/tpj.12720)
- Pathak V.K. and Temin H.M. (1992). 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate. *J. Virol.* 66: 3093–3100.
- Pesce D., Lehman N, de Visser J.A.G.M. (2016). Sex in a test tube: testing the benefits of in vitro recombination. *Phil. Trans. R. Soc. B* 371: 20150529. (doi.org/10.1098/rstb.2015.0529)
- Pesko K., Voigt E.A., Swick A., Morley V.J., Timm C., Yin J. and Turner P.E. (2015). Genome rearrangements affects RNA virus adaptability on prostate cancer cells. *Front Genet* 6, 121. (doi.org/10.3389/fgene.2015.00121)
- Philippe N., Legendre M., Doutre G., Couté Y., Poirot O., Lescot M., Arslan D., Seltzer V., Bertaux L., Bruley C., Garin J., Claverie J.M. and Abergel C. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* Vol. 341, Issue 6143, pp. 281-286. (doi.org/10.1126/science.1239181)

Referencias

- Pita J.S., de Miranda J.R., Schneider W.L. and Roossinck M.J. (2007). Environment determines fidelity for an RNA virus replicase. *J. Virol.* 81: 9072–9077. (doi.org/10.1128/JVI.00587-07)
- Poelwijk F.J., Kiviet D.J., Weinreich D.M. and Tans S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445: 383-386. (doi.org/10.1038/nature05451)
- Poelwijk F.J., Tanase-Nicola S., Kiviet D.J. and Tans S.J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J Theor Biol* 272:141–144. (doi.org/10.1016/j.jtbi.2010.12.015)
- Porter K. and Day B. (2013). Actin branches out to link pathogen perception and host gene regulation. *Plant Signaling & Behavior*, 8:3, e23468. (doi.org/10.4161/psb.23468)
- Porter K., Shimono M., Tian M. and Day B. (2012). Arabidopsis Actin-Depolymerizing Factor-4 Links Pathogen Perception, Defense Activation and Transcription to Cytoskeletal Dynamics. *PLoS Pathog* 8(11): e1003006, doi: 10.1371/journal.ppat.1003006
- Proulx S.R. (2011). The rate of multi-step evolution in Moran and Wright- Fisher populations. *Theor Pop Biol* 80:197–207. (doi.org/10.1016/j.tpb.2011.07.003)
- Qu F., Ye X. and Morris T.J. (2008). Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *PNAS*. Vol. 105 no. 38, 14732–14737. (doi.org/10.1073/pnas.0805760105)

- Redondo E., Krause-Sakate R., Yang S.J., Lot H., Le Gall O. and Candresse T. (2001). Lettuce mosaic virus pathogenicity determinants in susceptible and tolerant lettuce cultivars map to different regions of the viral genome. *Mol. Plant Microbe Interact.* 14:804-810. (doi.org/10.1094/MPMI.2001.14.6.804)
- Remold S.K. (2012). Understanding specialism when the jack of all trades can be the master of all. *Proc Biol Sci* 279:4861–4869. (doi.org/10.1098/rspb.2012.1990)
- Remold S.K. and Lenski R.E. (2004). Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat Genet* 36:423–426. (doi.org/10.1038/ng1324)
- Remold S.K., Rambaut A. and Turner P.E. (2008). Evolutionary genomics of host adaptation in vesicular stomatitis virus. *Mol Biol Evol* 25:1138–1147. (doi.org/10.1093/molbev/msn059)
- Revers F. and García J. A. (2005). Molecular biology of potyviruses. *Adv Virus Res* 92, 101-199.
- Rodrigo G., Zwart M.P. and Elena S.F. (2014). Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J R Soc Interface.* 11(98):20140555. (doi.org/10.1098/rsif.2014.0555)
- Roossinck, M.J. (2008). *Plant Virus Evolution*. Springer, Berlin, Germany.
- Rowland O. and Jones J.D. (2001). Unraveling regulatory networks in plant defense using microarrays. *Genome Biol.* 2:REVIEWS1001.

Referencias

- Rugnone M.L., Faigón-Soverna A., Sanchez S.E., Schlaen R.G., Hernando C.E., Seymour D.K., Mancini E., Chernomoretz A., Weigel D., Más P. and Yanovsky M.J. (2013). LNK genes integrate light and clock signaling networks at the core of the *Arabidopsis* oscillator. *Proc Natl Acad Sci U S A*. 16;110(29):12120-5. (doi.org/10.1073/pnas.1302170110)
- Sacristán S., Díaz M., Fraile A. and García-Arenal F. (2011). Contact transmission of tobacco mosaic virus: a quantitative analysis of parameters relevant for virus evolution. *Journal of Virology*, 85: 4974–1.
- Sacristán S., Malpica J.M., Fraile A. and García-Arenal F. (2003). Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. *Journal of Virology*, 77: 9906–11.
- Sanjuán R., Agudelo-Romero P. and Elena S.F. (2009). Upper-limit mutation rate estimation for a plant RNA virus. *Biol. Lett.* 5:394-396. (doi.org/10.1098/rsbl.2008.0762)
- Sanjuán R., Cuevas J.M., Furió V., Holmes E.C. and Moya A. (2007). Selection for robustness in mutagenized RNA viruses. *PLoS Genet* 3, 939-946. (doi.org/10.1371/journal.pgen.0030093)
- Sanjuán R., Moya A. and Elena S.F. (2004a). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the USA*, 101: 8396–401.
- Sanjuán R., Moya A. and Elena S.F. (2004b). The contribution of epistasis to the architecture of fitness in an RNA virus. *Proceedings of the National Academy of Sciences of the USA*, 101: 15376–9.

- Sanjuán R., Nebot M.R., Chirico N., Mansky L.M. and Belshaw R. (2010). Viral mutation rates. *J Virol* 84:9733–9748. (doi.org/10.1128/JVI.00694-10)
- Saxena P. and Lomonossoff G.P. (2014). Virus infection cycle events coupled to RNA replication. *Annual Review of Phytopathology* Vol. 52:197-212. (doi.org/10.1146/annurev-phyto-102313-050205)
- Schmidt G.W. and Delaney S.K. (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol Genet Genomics*, 283:233–241.
- Schneider W.L. and Roossinck M.J. (2001). Genetic diversity in RNA viral quasispecies is controlled by host-virus interactions. *J. Virol.* 75: 6566–6571. (doi.org/10.1128/JVI.75.14.6566-6571.2001)
- Schoor S., Farrow S., Blaschke H., Lee S., Perry G., von Schwartzberg K., Emery N. and Moffatt B. (2011). Adenosine kinase contributes to cytokinin interconversion in *Arabidopsis*. *Plant Physiol* 157, 659-672. (doi.org/10.1104/pp.111.181560)
- Shukla D.D. and Ward C.W. (1989). : Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Adv. Virus Research* (1989) 36, S. 273–314.
- Shukla D.D., Ward C.W. and Brunt A.A. (1994). *The Potyviridae*. Wallingford: CAB International.
- Simmonds P., Tuplin A. and Evans D.J. (2004). Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA* 10: 1337–1351. (doi.org/10.1261/rna.7640104)

Referencias

- Simon-Loriere E. and Holmes E.C. (2011). Why do RNA viruses recombine? *Nat Rev Microbiol* 9:617–626. (doi.org/10.1038/nrmicro2614)
- Smith-Tsurkan S.D., Wilke C.O. and Novella I.S. (2010). Incongruent fitness landscapes, not tradeoffs, dominate the adaptation of Vesicular stomatitis virus to novel host types. *J Gen Virol* 91:1484-1493. (doi.org/10.1099/vir.0.017855-0)
- Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl. Genet Mol Biol* 3, 3.
- Steinberg B. and Ostermeier M. (2016). Environmental changes bridge evolutionary valleys. *Sci Adv* 2:e1500921. (doi.org/10.1126/sciadv.1500921)
- Szathmáry E. and Maynard Smith J. (1995). The major evolutionary transitions. *Nature* 374, 227–232. (doi.org/10.1038/374227a0)
- Szendro I.G., Schenk M.F., Franke J., Krug J. and de Visser J.A.G.M. (2013). Quantitative analysis of empirical fitness landscapes. *J. Stat. Mech.* P01005. (doi.org/10.1088/1742-5468/2013/01/P01005)
- Tan S.L., Ganji G., Paeper B., Proll S. and Katze M.G. (2007). Systems biology and the host response to viral infection. *Nat Biotech* 25, 1383-1389. (doi.org/10.1038/nbt1207-1383)
- Tian M., Chaudhry F., Ruzicka D.R., Meagher R.B., Staiger C.J., et al. (2009). Arabidopsis actin-depolymerizing factor AtADF4 mediates defense signal transduction triggered by the *Pseudomonas syringae* effector AvrPphB. *Plant Physiol* 150: 815–824 (doi.org/10.1104/pp.109.137604)

- Tian T., Liu Y., Yan H., You Q., Yi X., Du Z., Xu W. and Su Z. (2017). AgriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucl Acids Res.* (doi.org/10.1093/nar/gkx382).
- Torres-Barceló C., Daròs J.A. and Elena S.F. (2010). Compensatory molecular evolution of HC-Pro, an RNA-silencing suppressor from a plant RNA virus. *Mol. Biol. Evol.* 27(3): 543–551. (doi.org/10.1093/molbev/msp272)
- Torres-Barceló C., Martín S., Daròs J.A. and Elena S.F. (2008). From hypo- to hypersuppression: effect of amino acid substitutions on the RNA-silencing suppressor activity of the tobacco etch potyvirus HC-Pro. *Genetics.* 180(2): 1039–1049. (doi.org/10.1534/genetics.108.091363)
- Travisano M., Mongold J.A., Bennet A.F. and Lenski R.E. (1995). Experimental tests of the roles of adaptation, chance and history in evolution. *Science* 267, 87–90. (doi.org/10.1126/science.7809610)
- Tromas N. and Elena S.F. (2010). The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* 185:983–989. (doi.org/10.1534/genetics.110.115915)
- Tromas N., Zwart M.P., Lafforgue G. and Elena S.F. (2014a). Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genetics*, 10: e1004186.
- Tromas N., Zwart M.P., Poulain M. and Elena S.F. (2014b). Estimation of the in vivo recombination rate for a plant RNA virus. *J Gen Virol* 95:724–732. (doi.org/10.1099/vir.0.060822-0)
- Turner P.E. and Elena S.F. (2000). Cost of host radiation in an RNA virus. *Genetics* 156:1465–1470.

Referencias

- Vasi F., Travisano M. and Lenski R.E. (1994). Long-term experimental evolution in *Escherichia coli*. Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat.* 144, 432–456. (doi.org/10.1086/285685)
- Vijayapalani P., Maeshima M., Nagasaki-Takekuchi N. and Miller W.A. (2012). Interaction of the Trans-Frame Potyvirus Protein P3N-PIPO with Host Protein PCaP1 Facilitates Potyvirus Movement. *PLoS Pathog.* Apr; 8(4): e1002639. (doi.org/10.1371/journal.ppat.1002639)
- Viswanathan K. and Früh K. (2007). Viral proteomics: global evaluation of viruses and their interaction with the host. *Exp Rev Proteomics* 4, 815-829.
- Wang R.Y. and Li K. (2012). Host factors in the replication of positive-strand RNA viruses. *Chang Gung Med J.* 35(2):111-24.
- Wargo A.R. and Kurath G. (2012). Viral fitness: definitions, measurements, and current insights. *Curr Opin Virol* 2, 538-545.
- Wasik B.R. and Turner P.E. (2013). On the biological success of ciruses. *Annual Review of Microbiology*, 67: 519–41. (doi.org/10.1146/annurev-micro-090110-102833)
- Weinreich D.M. and Chao L. (2005). Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* 59, 1175–1182. (doi.org/10.1111/j.0014-3820.2005.tb01769.x)
- Weinreich D.M., Delaney N.F., Depristo M.A. and Hartl D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312: 111–114. (doi.org/10.1126/science.1123539)

- Weinreich D.M., Lan Y., Wylie C.S. and Heckendorn R.B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev* 23:700–707. (doi.org/10.1016/j.gde.2013.10.007)
- Weinreich D.M., Watson R.A. and Chao L. (2005). Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174.
- Whitham S.A. and Wang Y. (2004). Roles for host factors in plant viral pathogenicity. *Curr. Opin. Plant Biol.* 7:365-371.
- Willemsen A. (2016). Experimental evolution of genome architecture and complexity in an RNA virus. (roderic.uv.es/handle/10550/53654)
- Willemsen A., Zwart M.P., Higuera P., Sardanyés J. and Elena S.F. (2016a). Predicting the stability of homologous gene duplications in a plant RNA virus. *Genome Biol Evol* 8, 3065-3082. (doi.org/10.1093/gbe/evw219)
- Willemsen A., Zwart M.P., Tromas N., Majer E., Daròs J.A. and Elena S.F. (2016b). Multiple barriers to the evolution of alternative gene orders in a positive-strand RNA virus. *Genetics* 202: 1503-1521.
- Wise R.P., Moscou M.J., Bogdanove A.J. and Whitham S.A. (2007). Transcript profiling in host-pathogen interactions. *Annu Rev Phytopathol.* 45:329-369.
- Wright S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Intl. Congress Genetics* 1: 356–366.

Referencias

- Xie M., Ren G., Zhang C. and Yu B. (2012). The ADN- and RNA-binding protein FACTOR of ADN METHYLATION 1 requires XH domain-mediated complex formation for its function in RNA-directed methylation. *Plant J* 72, 491-500. (doi.org/10.1111/j.1365-313X.2012.05092.x)
- Xu F., Xu S., Wiermer M., Zhang Y. and Li X. (2012). The cyclin L homolog MOS12 and the MOS4-associated complex are required for proper splicing of plant resistance genes. *Plant J*. 70, 916–928. (doi.org/10.1111/j.1365-313X.2012.04906.x)
- Yuste E., Sánchez-Palomino S., Casado C., Domingo E. and López-Galíndez C. (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *Journal of Virology*, 73: 2745–51.
- Zhan G. M., Li R.J., Hu Z.Y., Liu J., Deng L.B., Lu S.Y. and Hua W. (2014). Cosuppression of RBCS3B in Arabidopsis leads to severe photoinhibition caused by ROS accumulation. *Plant Cell Rep* 33, 1091-1108. (doi.org/10.1007/s00299-014-1597-4)
- Zhou N., Tootle T.L., Tsui F., Klessig D.F. and Glazebrook J. (1998). PAD4 functions upstream from salicylic acid to control defense responses in Arabidopsis. *Plant Cell*, 10, 1021–1030. (doi.org/10.1105/tpc.10.6.1021)
- Zwart M.P. and Elena S.F. (2015). Matter of size: genetic bottlenecks in virus infection and their potential impact on evolution. *Annual Review of Virology*, 2: 161–79.

