

TESIS DOCTORAL

Marianoel Pereira Gómez

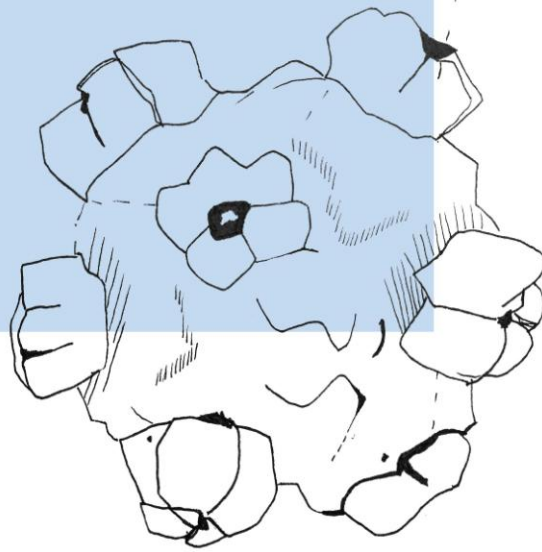
Octubre, 2016

DIRECTORES

Dr. Rafael Sanjuán Verdeguer

Dr. Fernando González Candelas

Genetic and Evolutionary Determinants of Mutation  
Rate in Bacteriophage  $\Phi$  X 174



VNIVERSITAT ID VALÈNCIA



Memoria presentada por Marianoel Pereira Gómez para optar al grado de Doctora en Biotecnología por la Universitat de València.

El presente trabajo se ha llevado a cabo en el grupo de Evolución y Salud de l'Institut Cavanilles de Biodiversitat i Biologia Evolutiva, adscrito a la Universitat de València, bajo la dirección de los doctores Rafael Sanjuán Verdeguer y Fernando González Candelas y ha sido subvencionado por una beca predoctoral de Formación de Profesorado Universitario (FPU) del Ministerio de Educación.




Fdo: Marianoel Pereira Gómez



VNIVERSITAT  
DE VALÈNCIA



Fdo: Rafael Sanjuán Verdeguer



Fdo: Fernando González Candelas



## CONTENTS

<b>AKNOWLEDGMENTS.....</b>	<b>7</b>
<b>RESUMEN.....</b>	<b>9</b>
<b>INTRODUCTION.....</b>	<b>15</b>
1. Evolutionary processes.....	17
2. Experimental model.....	26
<b>HYPOTHESES AND OBJECTIVES.....</b>	<b>31</b>
<b>MATERIAL AND METHODS.....</b>	<b>35</b>
1. Viruses.....	37
2. Cells.....	37
3. Plaque assays.....	38
4. Growth rate estimation: $\phi$ X174 growth assays.....	38
5. Viral mutation rates estimation by the Luria-Delbrück fluctuation test.....	39
6. Life history traits determination.....	42
7. Experimental evolution experiments.....	44
8. Site directed mutagenesis.....	45
9. Methylation analysis.....	46
10. Sequence analysis.....	46
<b>RESULTS AND DISCUSSION.....</b>	<b>47</b>
Chapter I: Lysis and mutation rate.....	49
Chapter II: DNA repair and mutation rate.....	61
Chapter III: GATC avoidance and evolvability.....	69
General discussion.....	81
<b>FINAL CONCLUSIONS.....</b>	<b>91</b>
<b>REFERENCES.....</b>	<b>95</b>
<b>SUPPLEMENTARY DATA.....</b>	<b>107</b>



## ACKNOWLEDGMENTS

Son muchas las personas que de alguna manera han participado en el inicio y desarrollo de las experiencias e ideas que concluyen en la presentación de esta tesis doctoral. A todas ellas quiero expresarles desde ya mi agradecimiento. En primer lugar, me gustaría agradecer a mis profesoras, las doctoras Amparo Cinos y Amparo Latorre, por su respaldo y consejo durante mis épocas de estudiante en el liceo y la universidad. En segundo lugar, agradezco a mis directores de tesis, los doctores Rafael Sanjuán y Fernando González, sin los cuales no hubiera sido posible empezar y realizar este proyecto.

Quiero agradecer además a mis compañeros de laboratorio, José, Ron, Silvia, Marine, Raquel y Jennifer. A los compañeros del grupo Genética Evolutiva, al cual originalmente pertenecía cuando empecé este trabajo, especialmente a los miembros del Chelab. También quiero agradecer a todos mis compañeros de l'Institut Cavanilles, sobre todo al doctor Manuel Serra y a las compañeras del servicio general de l'institut.

También quiero agradecer a los compañeros del laboratorio de Evolución Molecular del Centro de Astrobiología de Madrid y del laboratorio de Virología Molecular del Centro de Investigaciones Nucleares de Montevideo, con los cuales pasé muy lindos momentos dentro y fuera del laboratorio. Especialmente, a las doctoras Ester Lázaro y Pilar Moreno que me han dirigido de forma excelente durante mis estancias en sus laboratorios y por su apoyo posterior.

Además quiero agradecer a mis colegas de Gellentinos que realizaron la corrección gramatical del inglés, y a Vivianna Mazuco y Francisco Supervielle que diseñaron la portada.

Finalmente, pero no menos importante, quiero agradecer especialmente a mi familia y amigos, por acompañarme siempre.





## RESUMEN



La mutación es la fuente última de variabilidad genética y por eso es fundamental en la evolución de los organismos. Aunque la variabilidad genética es necesaria para la adaptación, la mayoría de las mutaciones suelen ser deletéreas y por tanto tienden a reducir la eficacia biológica de las poblaciones debido al lastre genético. La teoría predice que, como consecuencia del equilibrio entre adaptación y lastre genético, debe existir un valor intermedio para la tasa de mutación que permita maximizar el éxito evolutivo, o tasa óptima de mutación.

En general, los virus son uno de los modelos más usados en biología y en particular en biología evolutiva. Esto es debido a una serie de características como son: i) su pequeño tamaño genómico, que hace sencillo comprender las bases moleculares de su evolución y realizar modificaciones genéticas *in vitro*; ii) sus grandes tamaños poblacionales y su rápida tasa de crecimiento, que permiten cultivarlos fácilmente y realizar muchas réplicas independientes; y iii) sus elevadas tasas de mutación, que los hacen evolucionar rápidamente permitiéndonos estudiar la evolución en tiempo real.

Las tasas de mutación virales varían entre  $10^{-8}$  y  $10^{-4}$  sustituciones por nucleótido por ronda de replicación ( $s/n/r$ ) y dependen, en gran parte, de las enzimas implicadas en la replicación. Las ARN polimerasas ARN dependientes (RdRp) y las retrotranscriptasas (RT) presentan las mayores tasas de mutación descritas debido a la ausencia de actividad correctora de errores 3' exonucleasa, lo cual en parte explica que los virus que presentan estadios de ARN en su ciclo de infección tengan tasas de evolución más rápidas que los organismos con genoma de ADN. Sin embargo, hay otros procesos que pueden afectar la tasa de mutación tales como el mecanismo de replicación, la presencia de agentes oxidantes e incluso la introducción de mutaciones mediada por enzimas citidina desaminasas del hospedador de la familia APOBEC3, por citar algunos ejemplos.

Pese a la variedad de factores que pueden modular la tasa de mutación, se ha observado que los organismos que poseen ADN como material genético, incluyendo los virus de ADN, las bacterias y los eucariotas unicelulares, presentan una tasa de mutación por genoma de alrededor de 0,003 sustituciones por genoma y ronda de replicación, lo cual implica que la tasa de mutación está inversamente relacionada con el tamaño del genoma. Esta relación entre tasa de mutación y tamaño genómico se conoce como regla de Drake. Por ejemplo, la mayoría de bacterias tienen tamaños genómicos superiores a 1000 kb y tasas de mutación menores que  $10^{-9} s/n/r$ . Sin embargo, los bacteriófagos de ADN de cadena sencilla, como  $\phi$ X174 y m13, suelen tener unos tamaños genómicos de aproximadamente 6 kb y sus tasas de mutación son del orden de  $10^{-6} s/n/r$ .

El bacteriófago  $\phi$ X174 y otros virus de ADN de cadena sencilla son modelos interesantes para realizar estudios de tasas de mutación y adaptación debido a que sus tasas de mutación son de las más altas entre los virus de ADN y, además, estos virus utilizan la polimerasa del huésped para la replicación viral. Por tanto, el objetivo principal de esta tesis doctoral consistió en investigar los mecanismos de regulación de la tasa de mutación que no dependen de la polimerasa viral, utilizando el bacteriófago  $\phi$ X174 como sistema modelo.

En un trabajo anterior (Domingo-Calap et al., 2012), demostramos que el bacteriófago  $\phi$ X174 desarrolló resistencia al análogo de nucleósido 5-fluorouracilo (5-FU) y que esta resistencia estaba mediada principalmente por sustituciones en el gen que codifica para la proteína de lisis E. Estos resultados nos permitieron plantear la hipótesis de que la acumulación de mutaciones en el genoma viral podría reducirse por cambios en el tiempo de lisis. Para ello, primero diseñamos y construimos una serie de mutantes que portaban remplazamientos del dominio N-terminal de la proteína E mediante mutagénesis dirigida y, a continuación, realizamos varios ensayos para caracterizar diferentes etapas del ciclo viral o rasgos de historia de vida (como, por ejemplo, el tamaño de la progenie, el tiempo de lisis, la tasa de crecimiento, entre otros) y la tasa de mutación de dichos mutantes y el virus silvestre. Los resultados de estos experimentos mostraron que aproximadamente la mitad de los mutantes ensayados producían un aumento en el tiempo de lisis, y dos de ellos (V2A y D8A) también producían resistencia al 5-FU. Los resultados mostraron que un retraso en el tiempo de lisis produjo un incremento del tamaño de la progenie en presencia de 5-FU, explicando el aumento de la eficacia biológica (medida como la tasa de crecimiento) en presencia del mutágeno. Además, estas mutaciones redujeron aproximadamente a la mitad la tasa de mutación comparado al virus silvestre según las estimaciones obtenidas a partir de pruebas de fluctuación en presencia de 5-FU. Finalmente, propusimos un modelo matemático sencillo que se ajusta a los datos observados y explica que un aumento de la progenie, como consecuencia de un mayor tiempo de lisis, permite al virus reducir el número de ciclos de infección necesarios para expandir su tamaño poblacional. En consecuencia, el número de ciclos de replicación en el cual las mutaciones pueden acumularse disminuye contrarrestando el efecto mutagénico del 5-FU y contribuyendo a la resistencia del virus.

La tasa de mutación de  $\phi$ X174 sugiere que el fago evita los mecanismos de reparación post-replicativos de *E. coli*, como el sistema de reparación de apareamientos incorrectos dependiente de la metilación del ADN (MMR por sus siglas en inglés). A diferencia de su hospedador, el fago  $\phi$ X174 no contiene sitios GATC, los cuales son necesarios para la acción del sistema MMR. En un estudio previo (Cuevas et al., 2011) encontramos que la introducción

de siete sitios GATC produjo una reducción de treinta veces en la tasa de mutación. Además, el efecto sobre la tasa de mutación se anuló en células *mutD*, que son deficientes en MMR, indicando que el efecto de los sitios GATC estaba relacionado con la acción del sistema MMR. Para estudiar la relación entre el sistema MMR y la tasa de mutación de  $\phi$ X174 diseñamos y construimos una serie de mutantes en los cuales el variamos el número y localización de sitios GATC en el genoma viral, incluyendo un mutante poseyendo el número de sitios GATC esperados por azar. Esto nos permitió no solo estudiar el efecto la reparación de apareamientos incorrectos en la tasa de mutación del bacteriófago  $\phi$ X174 sino también como afecta a su capacidad adaptativa.

Los resultados de las pruebas de fluctuación mostraron que la restauración del número de secuencias GATC esperadas por azar redujo aproximadamente ocho veces la tasa de mutación sin alterar significativamente la eficacia biológica del virus. Sin embargo, la eficiencia de la maquinaria implicada en la reparación del ADN fue limitada por los bajos niveles de metilación del genoma viral. Además, al construir diferentes virus mutantes, cambiando la localización y el número de sitios GATC introducidos, observamos que los efectos sobre la tasa de mutación no eran aditivos y eran muy variables, con algunas combinaciones que alcanzaban una reducción de 50 veces en la tasa de mutación, hasta algunas que no tenían un efecto significativo. Las mayores reducciones en la tasa de mutación se encontraron en algunos GATC localizados en regiones intergénicas, sugiriendo que los impedimentos estéricos pueden afectar a la acción de las enzimas de reparación. Por otro lado, también observamos que el efecto de la introducción de GATCs en el genoma de  $\phi$ X174 es revertido bajo condiciones de estrés, como son el crecimiento a elevadas temperaturas y la exposición a 5-FU, lo cual sugiere que la adición de secuencias GATC hace al fago sensible al fenómeno de mutagénesis inducida por el estrés. Mientras que a 42°C la tasa de mutación del virus silvestre no fue alterada de forma significativa, la tasa de mutación del virus 20GATC fue significativamente mayor que la de dicho virus a 37°C y también que la del virus silvestre a 42°C. De forma similar, el mutágeno tuvo un mayor efecto sobre la tasa de mutación del virus 20GATC comparado al virus silvestre.

Con el fin de averiguar si la presencia de GATCs afecta a la capacidad evolutiva de  $\phi$ X174, realizamos experimentos de evolución experimental en condiciones constantes de baja y alta intensidad de selección. Para ello, primero realizamos 100 pases seriados en su hospedador natural, *E. coli*. Estas condiciones pretenden simular la intensidad de selección a la que naturalmente está expuesto  $\phi$ X174, al menos en nuestras condiciones de laboratorio. En segundo lugar, realizamos un nuevo experimento de evolución en un nuevo hospedador, *Salmonella typhimurium*, al cual  $\phi$ X174 no está adaptado y, por tanto, en el que las

poblaciones virales estaban sometidas a una alta presión de selección. En ambos casos se evolucionaron cinco líneas independientes por virus y hospedador a partir de poblaciones ancestrales del virus silvestre y del mutante 20GATC. Mediante curvas de crecimiento estándar estimamos la eficacia biológica relativa a su correspondiente población ancestral. Los resultados de la evolución de las poblaciones derivadas del virus silvestre y del mutante 20GATC en *E. coli* mostraron similares trayectorias de la eficacia biológica a lo largo de la evolución. El análisis de las secuencias consenso de las poblaciones evolucionadas en *E. coli* reveló que la mayoría de los sitios GATC se habían mantenido, lo cual indica que, de haber una selección en contra de la presencia de sitios GATC, ésta es baja en nuestras condiciones experimentales. Por otro lado, también mostró que las poblaciones evolucionadas experimentaron un nivel de adaptación tras 100 pases de evolución en *E. coli*, aunque probablemente las diferencias en la eficacia de las mutaciones observadas fueron lo suficientemente pequeñas para no poder detectarlas en nuestras condiciones de crecimiento.

Sin embargo, en la evolución de las poblaciones derivadas del virus silvestre y del mutante 20GATC en *S. typhimurium* observamos que las poblaciones del mutante 20GATC mostraron una menor capacidad de adaptación, ya que el incremento de la eficacia fue menor que el observado en las poblaciones del virus silvestre, indicando que la presencia de GATCs constriñe la adaptabilidad del virus. El análisis de las secuencias consenso de las poblaciones evolucionadas en *S. typhimurium* mostró que el número de mutaciones observadas al final de la evolución fue similar en ambas poblaciones. Sin embargo, algunas mutaciones paralelas en el gen de la proteína H con un potencial efecto beneficioso, puesto que han sido previamente descritas en trabajos de adaptación a *Salmonella*, se encontraban exclusivamente en las poblaciones evolucionadas a partir del virus silvestre, lo cual proporciona una posible base molecular para explicar la menor capacidad adaptativa de las poblaciones derivadas del mutante 20GATC.

En conclusión, los resultados obtenidos a partir de este trabajo contribuyen a dilucidar algunos mecanismos que permiten al bacteriófago  $\phi$ X174 regular su tasa de mutación para poder responder a las diferentes presiones de selección a las que puede estar sometido naturalmente.

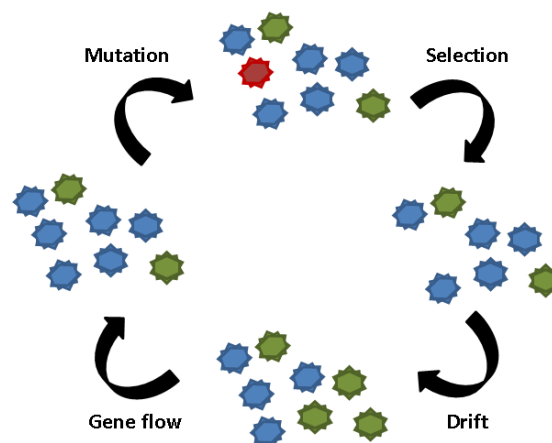
## INTRODUCTION





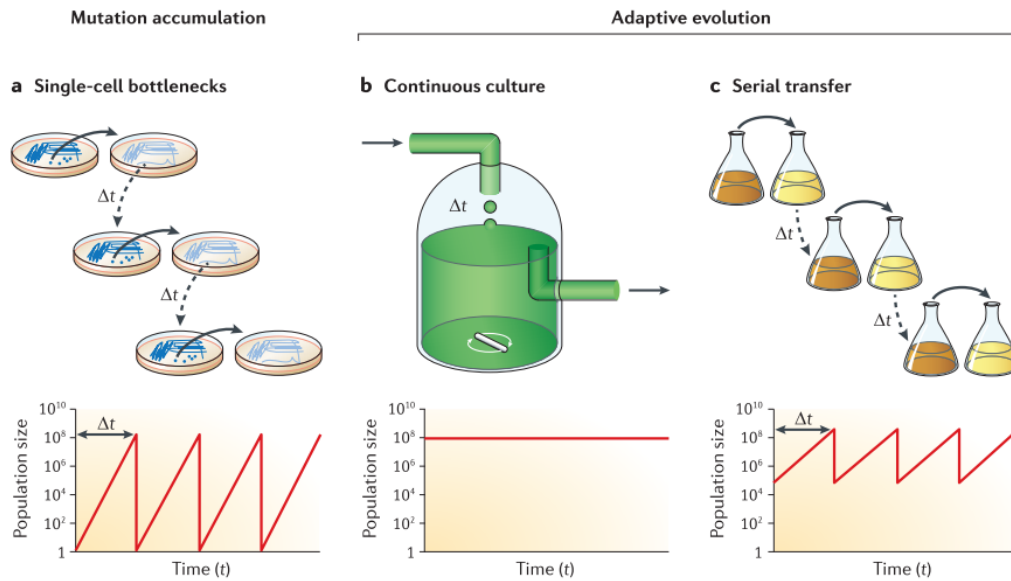
## 1. Evolutionary processes.

Historically, evolutionary biology has focused on natural selection, perhaps because selection was the first evolutionary mechanism described over 150 years ago by Charles Darwin (Darwin, 1859). Later, Sewall Wright suggested that chance could have an important effect, especially in small populations (Wright, 1931). Finally, Motoo Kimura, based on the Wright-Fisher and mutation-drift balance models, developed the neutral theory of molecular evolution (Kimura, 1991, 1983, 1968a), which is based on the principle that many mutations are neutral in terms of fitness. As a result, according to this theory, nucleotide substitutions are a result of random processes driven mainly by drift and mutation. Mutation, natural selection, genetic drift and gene flow are the main processes driving evolutionary change (Figure 1). Gene flow is particularly relevant in viral epidemiology studies but, as this issue is out of the scope of this thesis, it is not discussed here.



**Figure 1. Overview of evolutionary mechanisms.** Mutations arise at different rates and are the ultimate source of genetic variability. Selection takes place when individuals present fitness differences. Drift occurs when the allele frequencies change by chance in finite populations. Gene flow occurs when some individuals enter and/or leave the population.

Experimental evolution is a very useful approach to test evolutionary hypotheses under controlled laboratory conditions. Experimental evolution is typically based on relatively simple experimental designs (Figure 2) and can be performed using a variety of organisms with different complexity such as *Drosophila*, *C. elegans*, and a variety of microorganisms. However, microorganisms, and especially viruses, are more suitable to see the evolutionary process in real time due to their short generation times, large population sizes, and high mutation rates. In addition, viruses have small and relatively simple genomes, allowing both relatively easy genetic manipulation and whole genome sequencing (Buckling et al., 2009; Elena and Lenski, 2003).

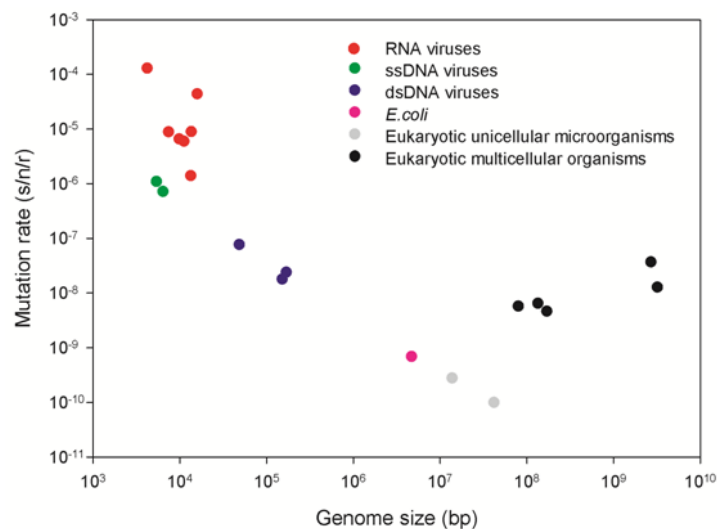


**Figure 2. Types of evolution experiments.** There are three main ways in which populations are propagated during evolution experiments, which lead to different types of genetic dynamics. The mechanics of how populations are maintained in each setup are illustrated for microorganisms (top panels), and representative changes in population sizes over time are also shown for each procedure (bottom panels). Analogous procedures exist for multicellular organisms, although population sizes are generally much smaller. **a)** In mutation accumulation experiments, frequent population bottlenecks lead to the fixation of mutations regardless of their effects on fitness. **b)** In experiments using continuous culture, populations are maintained at a nearly constant size and lead to adaptive evolution in the populations. **c)** In serial transfer experiments, a fraction of the population is periodically transferred to fresh media and allowed to regrow. Such continuous culture growth also leads to adaptive evolution. Figure originally published in Barrick and Lenski (2013) and used here with permission from the Nature Publishing Group.

Viruses are also characterized by being one of the most diverse biological entities observed in nature and they produce infections in very different organisms. Viral infections directly impact on human health and are also relevant in agriculture, cattle, and biotechnological industries. Viruses have very different types of genomes (i.e. RNA vs DNA) and genome polarities (i.e. plus vs minus). DNA viruses can be grouped into three different groups according to the Baltimore classification: i) double-stranded (ds) DNA viruses, such as adenoviruses, herpesviruses, and poxviruses; ii) single stranded (ss) DNA viruses such as microviruses and parvoviruses; and iii) reverse transcribing viruses with an RNA stage, such as hepadnaviruses and caulimoviruses, called pararetroviruses. On the other hand, RNA viruses are classified into four groups: i) positive-strand (+ss) RNA viruses, such as noroviruses, rhinoviruses, flaviviruses and coronaviruses; ii) negative-strand (–ss) RNA viruses, such as rhabdoviruses and orthomyxoviruses; iii) double-strand (ds) RNA viruses such as reoviruses; and iv) reverse transcribing viruses with a DNA stage or retroviruses, such as lentiviruses and deltaretroviruses.

### Viral mutation rates.

Viruses show a wide range of mutation rates. Typically, dsDNA viruses tend to have larger genomes than ssDNA viruses, whereas their mutation rates are lower than those of ssDNA viruses. For example, large dsDNA viruses, such as phage T2 and the eukaryotic herpes simplex virus type-1 (HSV-1), have a mutation rate on the order of  $10^{-7}$ - $10^{-8}$  substitutions per nucleotide per round of copying (s/n/r) (Luria, 1951; Drake, 1991; Drake and Hwang 2005), whereas small ssDNA phages, such as m13 and  $\phi$ X174, show a mutation rate on the order of  $10^{-6}$  s/n/r (Cuevas et al., 2009; Kunkel, 1985; Raney et al., 2004). This inverse relationship between genome size and mutation rate is known as Drake's rule and was observed in DNA based organisms, including bacteria and eukaryotic unicellular microorganisms (Figure 3). Based on this correlation, it has been suggested that the genomic mutation rate is constant across different organisms, averaging 0.003 mutations per round of copy (Drake, 1991). A similar negative relationship was recently suggested in riboviruses (Bradwell et al., 2013), with mutation rates varying between  $10^{-6}$  to  $10^{-4}$  s/n/r (Sanjuán et al., 2010). In contrast, Drake's rule does not seem to apply to eukaryotic multicellular organisms, in which mutation rate increases with genome size (Lynch, 2010).



**Figure 3. Relationship between mutation rate and genome size in different organisms.** Mutation rate is expressed as nucleotide substitutions per round of copying (s/n/r) and the genome size in bases or base pairs (bp). Red, blue and green dots correspond to RNA, dsDNA, and ssDNA viruses, respectively (data taken from Sanjuán et al. 2010 and Bradwell et al. 2013). Estimates of *E. coli* and eukaryotic unicellular microorganisms (*S. cerevisiae* and *N. crassa*) are represented by pink and grey dots, respectively (data taken from Drake 1991). Eukaryotic multicellular organisms are represented by black dots (data taken from Lynch 2010) and correspond to *C. elegans*, *Drosophila*, *A. thaliana*, *M. musculus*, *H. sapiens*.

The evolutionary causes of this negative correlation between mutation rates and genome sizes are not well determined yet. On the one hand, natural selection promotes lower mutation rates due to the deleterious effects of most mutations and, on the other hand, higher mutation rates can contribute to the adaptation process. In addition, the fitness costs imposed by the mechanisms of replication fidelity could promote an increase of the mutation rate (Dawson, 1999; Kimura, 1967; Kondrashov, 2009; Lynch, 2008; Sniegowski et al., 2000). Theory predicts that, as a consequence of a balance between deleterious mutations and adaptability, there should exist an optimum value of the mutation rate which maximizes adaptability and correlates negatively with genome size (Johnson and Barton, 2002; Orr, 2000). In addition, the drift barrier hypothesis proposes that, as drift is inversely related to the effective population size, organisms with large effective population size are expected to show lower mutation rates, providing a unifying explanation for variation of mutation rates in different taxa (Lynch, 2011, 2010; Sung et al., 2012).

Viral mutation rates depend to a large extent on the polymerases involved in viral replication. For instance, RNA-dependent RNA polymerases (RdRp) and reverse transcriptases (RT) show the highest mutation rates described (Flint et al. 2004). This is due to the absence of 3' exonuclease proofreading activity, which produces a replication fidelity increment of ten to a hundred-fold (Kunkel, 2004; Schaaper, 1998) and partly explains why viruses including an RNA stage in their infection cycle tend to evolve faster. Viral polymerase variants with altered base selection have been described in phage T4 (Drake et al., 1969; Speyer, 1965) and several RNA viruses (Arias et al., 2008; Coffey et al., 2011; Graci et al., 2012; Menéndez-Arias, 2009; Pfeiffer and Kirkegaard, 2003).

The viral replication mechanism can also affect mutation accumulation in viral genomes. Viral replication is called “stamping machine” when a single parental DNA molecule is used to produce all progeny and, thus, there is only one round of copying. As a result, mutations accumulate linearly inside cells. Alternatively, viral replication is called geometric or semi-conservative when DNA progeny molecules immediately, become templates such that the number of genome copies doubles after each replication cycle and mutations accumulate geometrically inside cells (Duffy et al., 2008; Sanjuán et al., 2010).

Host factors such as the apolipoprotein-B mRNA-editing catalytic polypeptide-like 3 (APOBEC3) family of cytidine deaminases can also influence viral mutation rates. APOBEC3 constitutes an innate cellular defense against retroviruses, endogenous retro-elements and some DNA viruses, producing mutations in viral DNA by deamination of cytidine residues to uracil (Chiu

and Greene, 2008; Vieira and Soares, 2013). The first studies showing the antiviral function of APOBEC3 were carried out using human immunodeficiency virus type-1 (HIV-1) (Harris et al., 2003; Mangeat et al., 2003; Zhang et al., 2003). Nevertheless, several subsequent studies showed that APOBEC3 is also involved in antiviral responses against hepatitis B virus (Suspène et al., 2005; Turelli et al., 2004), human papillomavirus (Vartanian et al., 2008), human herpesvirus and Epstein-Barr virus (Suspène et al., 2011). Hypermutation resulting from APOBEC3 editing is mainly antiviral because it produces a large number of deleterious missense or nonsense mutations. For example, in HIV-1, it was shown that higher APOBEC3 expression levels are associated with higher CD4 counts and slower progression to disease (Land et al., 2008). However, a small fraction of the edited genomes could be viable and help the virus promote variability, evade immune responses, and become resistant to antiviral treatments (Berkhout and de Ronde, 2004; Kim et al., 2010; Wood et al., 2009).

DNA viruses may also modulate their mutation rates by avoiding host repair mechanisms. For example, the high mutation rate of  $\phi$ X174 is consistent with the avoidance of methyl-directed mismatch repair (MMR), the main post-replication repair mechanism present in *E. coli*, which confers an up to 1000-fold increase in replication fidelity in bacteria and requires the presence of methylated GATC sequences in the genome (Fijalkowska et al., 2012). The bacteriophage  $\phi$ X174 genome sequence has no GATC sequences, hampering viral DNA methylation and therefore avoiding MMR (Cuevas et al., 2011).

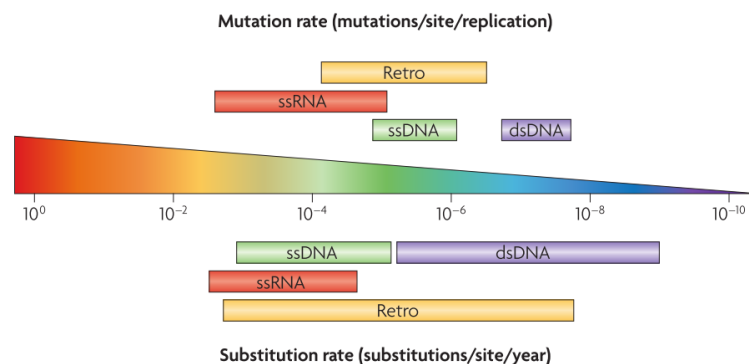
There are additional host factors that might affect viral mutation rates. For instance, unbalanced concentrations of deoxynucleotide triphosphates (dNTPs) were associated with the alteration of different viral polymerases error rates *in vitro* (Bebenek et al., 1992) and *in vivo* (Julias and Pathak, 1998). In addition, in hepatitis C virus (HCV) it was observed that a mutagenic effect was produced by the intracellular presence of ethanol-derived reactive oxygen species (ROS) (Seronello et al., 2011). Finally, mutation rates can show differences among hosts. For example, an approximately four-fold reduction of the mutation rate in insect cells compared to mammalian cells was found for vesicular stomatitis virus (VSV). This difference in mutation rates among host cells seems to be related with the payoffs of the viral transmission mode, explaining the relatively slow evolution rate of VSV (Combe and Sanjuán, 2014).

### **Viral substitution rates.**

DNA viruses have been usually considered as organisms with low mutation and evolution rates compared to those RNA viruses. Nevertheless, recent studies of viral evolution are changing

this established paradigm. For example, it has been observed that some ssDNA virus mutation rate estimates are close to the those of the lowest values shown by RNA viruses (Duffy et al., 2008; Sanjuán et al., 2010).

Similar to mutation rates, substitution rates also vary broadly across viruses (Figure 4). Substitution rates differ from mutation rates and refer to the rate at which genetic changes accumulate in a population, thus depending on both mutation rate and other processes such as the fitness effects of mutations, among others. RNA virus substitution rates range between  $10^{-3}$  and  $10^{-5}$  substitutions/site/year (s/s/y), whereas in DNA viruses substitution rates show higher variability among taxa and range between  $10^{-4}$  and  $10^{-9}$  s/s/y (Duffy et al., 2008). However, there is some potential uncoupling between mutation and evolution rates (Sanjuán et al., 2010). This seems to be the case for several ssDNA viruses, such as canine and human parvovirus (Shackelton et al., 2005; Shackelton and Holmes, 2006), the SEN- virus (Umemura et al., 2002), cotton leaf curl geminivirus (Sanz et al., 1999), and tomato yellow leaf curl virus, (Duffy and Holmes, 2008) as well as some dsDNA viruses, such as variola virus (Firth et al., 2010) and African swine fever virus (Michaud et al., 2013). These high substitution rates could be a consequence of short generation times or strong positive selection rather than high mutation rates (Shackelton et al., 2005). However, mutation rate estimates for these viruses are still lacking.



**Figure 4. Comparison between viral mutation and substitution rates.** The ranges of mutation rates expressed as mutations per site per round of replication are summarized in the upper part of the Figure. Substitution rates are expressed as substitutions per site per year and are summarized in the lower part of the Figure. Figure originally published in Duffy et al. 2008 and reproduced with permission from the Nature Publishing Group.

#### Natural selection versus genetic drift.

Population genetic theory predicts that natural selection drives molecular evolution when the product  $N_e s \gg 1$ , where  $N_e$  and  $s$  are the effective population size and the selection

coefficient, respectively (Ohta, 1992). Thus, if population sizes or selection coefficients are large, evolution is dominated by selection, whereas if population sizes or selection coefficients are small, evolution is driven by drift. Mutations with fitness effects are more often deleterious than beneficial. Studies on the fitness effects of random point mutations performed using several viruses, such as vesicular stomatitis virus (VSV) (Sanjuán et al., 2004), tobacco etch virus (TEV) (Carrasco et al., 2007), and bacteriophages Q $\beta$ ,  $\phi$ X174 (Domingo-Calap et al., 2009) and f1 (Peris et al., 2010), have shown that deleterious mutations with small effects are more frequent than those with large effects. Although the average mutational fitness effect is similar in RNA and small single-stranded DNA viruses, the fraction of lethal mutations seems to be higher, on average, for single stranded RNA viruses, ranging from 29% in Q $\beta$ , up to 40% in VSV and TEV, and only 20% in  $\phi$ X174 and f1 phages. However, more estimates are required to assess the generality of these findings, particularly for larger DNA viruses.

On the other hand, analysis of sequence data from 222 independent viral sequence datasets, which represented a wide variety of DNA and RNA viruses, showed that purifying selection is the main evolutionary force acting on these organisms, although selection is stronger in RNA viruses than in DNA viruses (Hughes and Hughes, 2007). This observation could be a consequence of the fact that DNA viruses are, in general, more robust to mutation than RNA viruses, which show non-redundant and small, compact genomes with strong pleiotropy (Krakauer and Plotkin, 2002). A second explanation could be the existence of higher levels of adaptive evolution in DNA viruses (Shackelton et al., 2005).

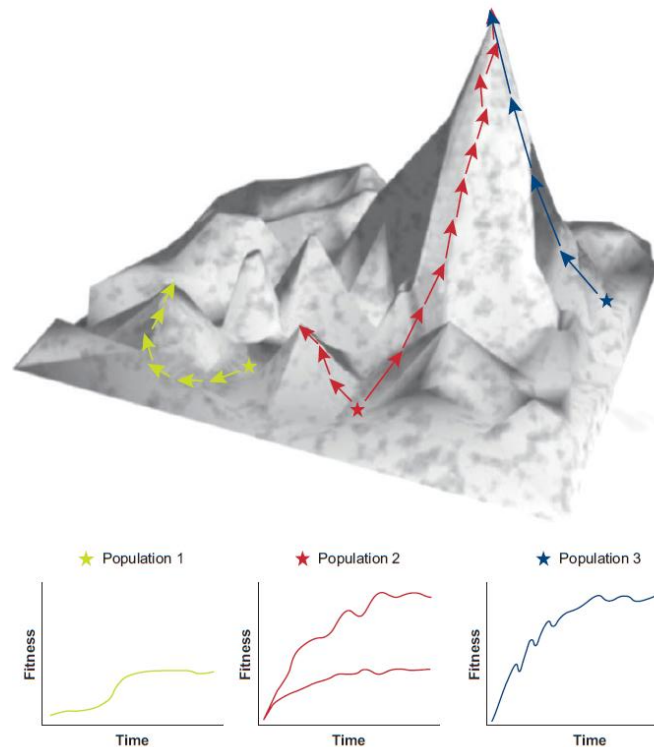
*A priori*, synonymous mutations should evolve neutrally (Kimura, 1977, 1968b) since the protein primary sequence remains unchanged. However, this might not hold in large viral populations where, as mentioned before, evolution is mainly driven by positive selection and synonymous sites can comprise signals for gene expression and encapsidation producing changes on the selective value of the different codons (Mayrose et al., 2013). Furthermore, synonymous sites can be constrained by the presence of secondary structures, as suggested by a comparative study of single stranded viruses that showed that selection at synonymous sites is stronger in RNA viruses than in DNA viruses, likely because selection at the RNA level is stronger in the former (Cuevas et al., 2012). Finally, evolution at synonymous sites can be also constrained by codon usage preferences, which seems to be related to host adaptation in some viruses (Bahir et al., 2009), vector specificity in flaviviruses (Jenkins et al., 2001), and can be used to produce viral attenuation, as shown for poliovirus in a mouse model (Coleman et al., 2008).

Adaptation occurs when beneficial mutations are selected in populations, leading to an increase in average population fitness ( $W$ ). Fitness can be defined as the number of progeny generated per individual per time unit, usually expressed relative to other genotype or organism. The classical Fisher geometric model of adaptation states that adaptive evolution should be a consequence of the fixation of many mutations of small effect because advantageous mutations of small effect should be more frequent than those of large effect (Fisher, 1930). This hypothesis was empirically confirmed using bacteriophage  $\phi 6$  (Burch and Chao, 1999).

In a scenario where drift and selection act jointly, the fate of beneficial mutations is also subjected to accidental loss by drift. In this case, the probability of loss depends on the effective population size and the selection coefficient (Haldane, 1927). Therefore, combining this “threshold” with the classical Fisher model, it is expected that beneficial mutations, which mainly contribute to adaptation, have intermediate effects and that the adaptation rate decreases over time, due to the fact that beneficial mutations become rarer and their effects smaller over time. Another important factor involved in the fixation of beneficial mutation is clonal interference (Gerrish and Lenski, 1998). Clonal interference occurs when different beneficial mutations located in different genomes compete between them, leading to the loss of one competitor and hence to a reduction of adaptation compared to a scenario in which mutations are in the same genome. The effects of clonal interference can thus be offset by recombination. Clonal interference was reported in VSV (Cuevas et al., 2002; Miralles et al., 2000, 1999) and bacteriophage  $\phi X174$  (Pepin and Wichman, 2008).

Experimental evolution has shown that viral adaptation typically follows a pattern whereby fitness initially increases fast until reaching a *plateau*, where there are no significant changes in fitness over time (Bull et al., 1997; Novella et al., 1995) (Figure 5, bottom panel). This suggests that when populations are not well adapted, beneficial substitutions tend to have larger effects, whereas at the end of the adaptation process beneficial nucleotide substitutions tend to have smaller effects or to be rarer (Elena and Lenski, 2003). Independent populations founded from the same genetically homogeneous population under identical experimental conditions can, in principle, evolve to different maximum fitness values (Domingo-Calap and Sanjuán, 2011), or even to different morphological features or fitness in alternative environments, as was observed in *E. coli* (Riley et al., 2001). This is because population fitness trajectories depend on the fitness landscape (Wright, 1932) (Figure 5) more than on the initial fitness of the population.





**Figure 5. Trajectories of adaptation.** The evolutionary trajectories of three independent populations are illustrated in a hypothetical rugged fitness landscape by arrows of different colors. Populations starting from different points can reach the same adaptive peak, in a case of convergent evolution (population 3 and high population 2 dynamics), and in this particular case, the global optimum. Starting from the same point, replicate populations can diverge into two different trajectories and reach completely different adaptive peaks (population 2 dynamics) depending upon the availability of beneficial mutations. A population can be trapped in a local adaptive peak (population 1 and low population 2 dynamics). Figure originally published in Elena and Sanjuán 2007 and reproduced with permission from the Annual Reviews.

As mentioned above, genetic drift should drive evolution when  $N_e s \ll 1$ . Under this scenario, mutations change their frequencies in the populations independently of their adaptive value and deleterious mutations can accumulate in the population producing fitness losses. The best known mutation accumulation mechanism in finite, asexual populations is Muller's ratchet (Muller, 1964). In this case, mutation-free genomes are irreversibly lost by drift, as long as recombination, back and compensatory mutations do not occur, and the population eventually reaches the point where each genome carries at least one deleterious mutation. If this process is continued, it finally can lead to population extinction by a process known as mutational meltdown (Lynch et al., 1993).

Important population size reductions in viruses are produced during host-to-host transmission because populations suffer bottlenecks which can be as extreme as reduction to one infectious particle, such as in HIV (Fischer et al., 2010). However, experiments carried out with equine

influenza virus (Murcia et al., 2010) suggest that bottlenecks can be broader. Bottlenecks can also be produced during vector-borne transmission (Forrester et al., 2012) as well as between organs or tissues within the same host (Ali et al., 2006; Forrester et al., 2012; Li and Roossinck, 2004).

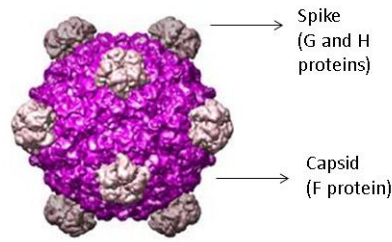
The shifting balance theory is an evolutionary model that explains how populations can evolve as a result of the combined action of genetic drift, mutation and natural selection (Wright, 1982). Under this model, mutation and genetic drift lead to continuous changes in the genetic variability of the populations whereas selection favors beneficial mutations and the elimination of deleterious mutations. As a consequence, episodes of strong genetic drift would play an important role in adaptation because populations would be able to cross through valleys of low adaptive value and then, when the population size increases, populations could reach a higher adaptive peak. In addition, different isolated populations can reach different adaptive peaks, increasing genetic variability. Although evolution is also constrained by the topography of the fitness landscape, it is expected that viral populations may evolve to greater genetic diversity and adaptability compared to the sole action of one of these evolutionary mechanisms.

## **2. Experimental model.**

Bacteriophage  $\phi$ X174 is a model virus belonging to the family *Microviridae*. Microviruses include phages that naturally infect coliform bacteria and have an icosahedral capsid containing a positive polarity, circular, single stranded DNA genome that typically ranges between 5.4 - 6.3 kb in length (Fane et al. 2006; Rokyta et al. 2006).

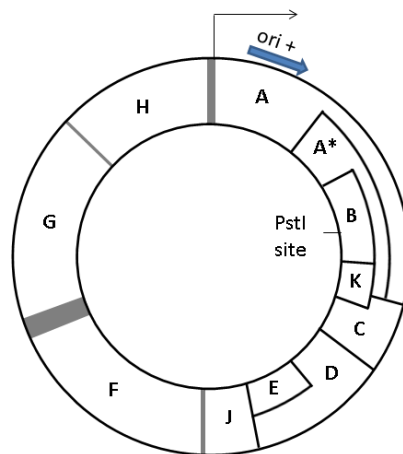
### **Virion structure and genome organization.**

During morphogenesis, F and G proteins (coat and major spike proteins, respectively) are assembled into the procapsid directed by proteins B and D, which are the internal and external scaffolding proteins, respectively. The  $\phi$ X174 virion is formed by 60 copies of the major coat protein F and 12 spikes each composed of five subunits of G protein and one subunit of the H protein (Figure 6). The spikes are involved in host recognition and attachment. The J protein is involved in phage DNA compaction during packaging and subsequently binds to the internal surface of the capsid to produce the displacement of the internal scaffolding protein B. The external scaffolding protein is also displaced when the procapsid collapses (Fane et al. 2006).



**Figure 6.  $\phi$ X174 virion structure.** Capsid atomic structure showing subunits organization on the virion at 3Å resolution. The image was taken from the Virus Particle Explorer Database (VIPERdb) (<http://viperdb.scripps.edu>) (Carrillo-Tripp et al., 2009).

Despite its small genome (5386 nucleotides in length),  $\phi$ X174 encodes eleven genes, of which A\* (Colasanti and Denhardt, 1987) and K proteins (Tessman et al., 1980) have been determined to be non-essential. It presents a high grade of compaction, showing overlapping reading frames in several genes, which can be in different reading frames. Non-coding regions represent only 5% of the total genome size (Figure 7) and contain cis-acting regulatory sequences involved in viral gene expression such as promoters and terminators.

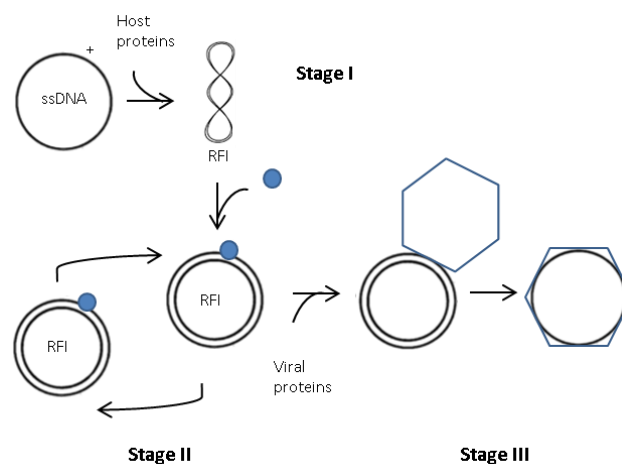


**Figure 7. Schematic representation of the  $\phi$ X174 genome.** Open reading frames are represented by rectangles (B, K, and E genes are in different reading frames), and grey bars indicate intergenic regions. Ori+ indicates the replication origin. The arrow indicates the direction of transcription for all genes. By convention, the first genome position corresponds to the last nucleotide of the unique *PstI* site. The A gene is involved in phage replication (stage II and III). A\* encodes for a N-terminal truncated A protein, a non-essential protein, which may be implicated in the inhibition of host cell DNA replication and superinfection exclusion. The B gene encodes for the internal scaffolding protein, which is also required for procapsid morphogenesis and assembly of early morphogenetic intermediates. The C protein facilitates the switch from stage II to stage III DNA replication and is necessary for stage III DNA synthesis. Gene D encodes for the external scaffolding protein, which is necessary for procapsid morphogenesis. Protein E is involved in cell lysis. F encodes for the major coat protein and G encodes for the major spike protein. Protein H is a DNA pilot protein needed for DNA injection, also called the minor spike protein. Gene J encodes for a DNA binding protein required DNA packaging. K is a non-essential protein, which may be involved in optimizing burst size in various hosts.

### Avoidance of post-replicative repair mechanisms and viral replication.

Unlike large DNA viruses,  $\phi$ X174 does not encode for its own polymerase but instead uses the *E. coli* DNA III holoenzyme, which probably operates with similar replication fidelity on both the host and phage genomes (Fersht, 1979; Fersht and Knill-Jones, 1981). However, the  $\phi$ X174 mutation rate is approximately 1000-fold higher than that of *E. coli*, suggesting that  $\phi$ X174 avoids post-replication repair mechanisms such as methyl-directed mismatch repair (MMR), a machinery that performs specific strand and bidirectional repair through MutHLS system (Jiricny, 2013). Single mismatches or small insertion/deletion loops are recognized by MutS, which interacts with MutL, leading to the activation of the MutH endonuclease, which recognizes GATC motifs that are methylated in the adenosine residue of the parental strand and cleaves the non-methylated daughter strand. This results in degradation of the daughter strand, re-synthesis by DNA polymerase III, and finally sealing of the nicked DNA by a ligase (Fukui, 2010; Li, 2008; Marti et al., 2002; Modrich and Lahue, 1996; Schofield and Hsieh, 2003).

The single stranded DNA of  $\phi$ X174 is replicated by the rolling circle mechanism (Dressler, 1970; Eisenberg et al., 1977), which involves three different stages (Figure 8): during stage I, a double stranded covalently closed circular molecule, called replicative form one (RFI) is synthesized from the single-stranded DNA molecule. Stage I requires 13 host proteins as determined *in vitro* (Shlomai et al., 1981). At stage II, RFI is amplified. This stage requires the viral A protein and host *rep* proteins (Eisenberg et al., 1976). Stage III, which requires viral procapsid and C proteins, involves the simultaneous synthesis and packaging of the single-stranded DNA genome (Mukai et al., 1979).



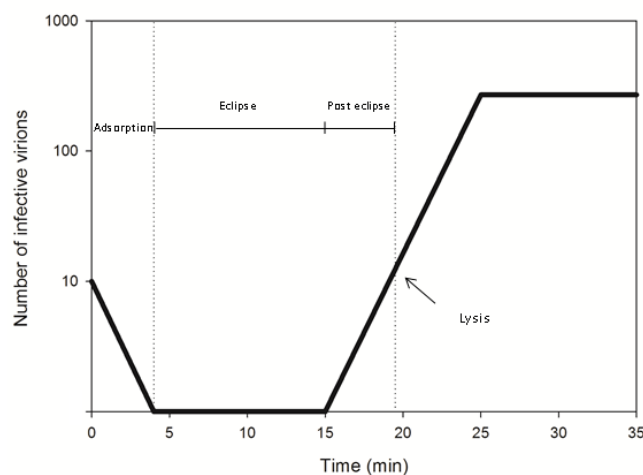
**Figure 8. Bacteriophage  $\phi$ X174 DNA replication by the rolling circle mechanism.** Stage I involves RFI formation. Stage II consists of RF I DNA amplification and requires the viral A protein (blue circles). Stage III involves the simultaneous synthesis and packaging of the single-stranded DNA genome.

### Viral cycle and population growth.

Bacteriophage  $\phi$ X174 is a lytic phage and its viral cycle involves three steps: adsorption, infection phase and lysis (Figure 9). Adsorption comprises phage dispersion until a host cell is found and subsequent host attachment.  $\phi$ X174 attaches to bacteria through a terminal sugar molecule of the bacterial lipopolysaccharides present in the cell wall of *E. coli* C and *Salmonella typhimurium*, which acts as phage receptor (Incardona and Selvidge, 1973; Jazwinski et al., 1975).

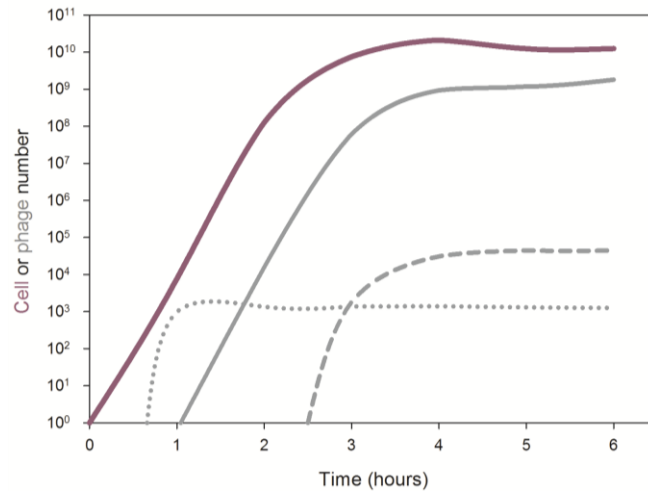
The infection phase was classically divided into eclipse and post eclipse. The eclipse period refers to the loss of viral infectivity, which occurs when the virus enters into the cell and starts producing progeny (i.e. DNA replication and early stages of morphogenesis). Thus, the eclipse time is defined as the time elapsed between viral adsorption and phage maturation. During the post eclipse period phage progeny starts to be assembled inside the host cell and increases in number over time.

The last step in the viral cycle is cell lysis where the new virions are released from the infected cell and dispersed into the environment. The lysis time is a trait that affects in opposite ways two fitness components related to growth rate, the viral generation time and the amount of phage produced per cell (burst size). Therefore, there is a trade-off associated with lysis time: a higher lysis time increases burst size but it also increases generation time and *vice versa* (Bull et al., 2004; Wang et al., 1996).



**Figure 9. General lytic phage cycle.** The black line describes the change in the number of infective virions over time. The first interval, from time zero to the first dashed vertical line, delimits the adsorption phase. The second interval (between the two dashed lines) represents the infection phase (eclipse plus post eclipse periods). The cycle finishes with the host cell lysis and release of the viral progeny.

Like many other biological organisms, viral population typically follows a logistic growth function. In addition to the intrinsic properties of the phage cycle, the  $\phi$ X174 population growth dynamics is dependent on both host cell density and the physiological status of the bacterium. As a consequence of its mechanism of lysis and replication,  $\phi$ X174 can only infect cells under active growth, such as cells in exponential growth phase (Figure 10).



**Figure 10. Dependence of the viral population dynamics on bacterial growth.** The pink solid curve represents the logistic growth of a hypothetical bacterial population. Grey curves represent the viral growth dynamics when the infection starts: at the beginning of the exponential cell growth (dotted lines), at the end of exponential cell growth (dashed line) or at an intermediate time point (solid line).

On a qualitative level, the phage-bacterial growth dynamics showed in Figure 10 can be explained as follows: if bacteria density is relatively low, phages grow slowly due to their low chances of finding susceptible bacteria. Hence, no viral growth is observed or alternatively, a sufficiently large phage inoculum can reduce bacterial density below a threshold necessary for viral growth (dotted line of Figure 10 exemplifies this situation). As bacteria reproduce, viral growth increases due to higher host density and higher phage density contributes further to the phage growth rate (solid grey line of Figure 10 illustrates this case). Finally, when bacterial density is high and the culture reaches the stationary phase, phage infection and growth is impeded (dashed grey line of Figure 10 represents this case).

## **HYPOTHESES AND OBJECTIVES.**





In RNA viruses, serial passaging experiments performed in the presence of nucleoside analogues led to changes in key residues of the viral polymerase, which typically produce an overall increase in replication fidelity or reduced affinity for the drug, thus producing a partial resistance to the mutagenic effects of these treatments (Agudo et al., 2010; Arias et al., 2008; Coffey et al., 2011; Graci et al., 2012; Pfeiffer and Kirkegaard, 2003; Sierra et al., 2007). Bacteriophage  $\phi$ X174 and other ssDNA viruses are interesting models for mutagenesis studies because their mutation rates are the highest among DNA viruses and, in addition, these viruses use the host polymerase for viral replication. Therefore, mechanisms of nucleoside analogue resistance obviously cannot involve changes in the viral polymerase, and hence other mechanisms of resistance may be revealed.

In a previous work, we showed that bacteriophage  $\phi$ X174 can evolve resistance to the nucleoside analogue 5-fluorouracil (5-FU) and that this resistance was mediated by substitutions in the lysis protein E, among other genes (Domingo-Calap et al., 2012). These results led us to hypothesize that mutation accumulation could be reduced by changes in lysis. Therefore, in this PhD thesis we will study the effect of amino acid replacements on the  $\phi$ X174 mutation rate and on life history traits.

As mentioned above, the mutation rate of  $\phi$ X174 suggests that the phage avoids *E. coli* post-replication repair mechanisms, such as methyl-directed mismatch repair (MMR). Unlike its host, bacteriophage  $\phi$ X174 contains no GATC sites, which are required for MMR. In a previous study, we introduced four GATC sites in the  $\phi$ X174 genome and found no effects on viral mutation rate, whereas by increasing the number of GATC sites to seven, we obtained a thirty-fold reduction in viral mutation rate. In addition, this effect on the viral mutation rate was reversed in *mutD* cells, which are MMR deficient, indicating that the effects of GATC motifs were related to MMR (Cuevas et al., 2011). Here, we will explore the interplays between MMR and the mutation rate of bacteriophage  $\phi$ X174 by engineering a phage genome carrying as many GATC sites as expected by chance. This will allow exploring not only the effects of mismatch repair on the mutation rate of the bacteriophage  $\phi$ X174, but also on phage evolution.

The objective of this PhD thesis is to investigate some mechanisms of mutation rate regulation that are not dependent on the viral polymerase, using bacteriophage  $\phi$ X174 as model system. First, we will design and construct a library of sixteen site-directed mutants of the N terminal domain of the E protein to study their effect on both viral 5-FU resistance and mutation rate. To do so, we will perform several assays to determine changes in life-history traits such as the burst size and the lysis time, among others, to compare them to the wild type virus in the absence and the presence of 5-FU as well as the viral mutation rate.

Second, we will study the effect of the mismatch repair mechanism on the  $\phi$ X174 mutation rate. To do so, we will construct several GATC containing mutants including a variant harboring the randomly expected number of GATC sites by chance with minimal effects on protein sequence and determined their mutation rate. Next, we will test the effect of the GATC presence on the viral mutation rate under stress conditions such as high temperature and 5-FU presence.

Third, we will explore the evolutionary consequences of the presence of GATC sites on the  $\phi$ X174 genome. To do so, we will perform experimental evolution experiments in conditions of low and strong selection carrying out five replicate lines. Then, we will quantify the fitness changes during evolution by growth rate assays at different time points. Finally, we will obtain the consensus sequences of the evolved populations to determine the genetic basis of the evolutionary change.

## **MATERIAL AND METHODS**



## **1. Viruses**

### **1.1 $\phi$ X174**

Bacteriophage  $\phi$ X174 was originally obtained from Prof. James J. Bull (University of Texas). Our laboratory strain, here denoted as wild type (WT), has been described previously and was adapted to *Escherichia coli* C IJ1862 strain by long-term serial passaging in our experimental conditions (Domingo-Calap et al., 2009). The full genome sequence of this virus is identical to GenBank accession number GQ153915.

## **2. Cells**

### **2.1 *Escherichia coli* strains.**

*E. coli* C IJ1862 cells were originally obtained from Prof. James J. Bull (University of Texas) and this strain is a permissive host for  $\phi$ X174. *E. coli gro87* strain was acquired from Dr. Bentley A. Fane (University of Arizona) and is a restrictive strain for  $\phi$ X174.

### **2.2 *Salmonella typhimurium* strain.**

*Salmonella typhimurium* LT2 strain IJ750 was obtained from Dr. Sylvain Gandon (Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier), who obtained it from Dr. James J. Bull (University of Texas). This strain is described in previous work (Hone et al., 1987) and is called MS 3849. We used *S. typhimurium* as a new host to which  $\phi$ X174 is not adapted. The culture medium for liquid infections (passaging and growth curves) and plaque assays (titrations) in *S. typhimurium* was complemented with 5mM CaCl<sub>2</sub>.

### **2.3 Cell stock preparation.**

As  $\phi$ X174 is only able to infect and complete its viral cycle in cells in the exponential growth phase, we prepared a stock of exponential phase cells and stored it at  $-70^{\circ}\text{C}$  until the moment of performing the different experiments. The stock preparation protocol is as follows:

A pre-culture (4 mL) seeded from a low number of cells was incubated overnight until the culture reached the stationary phase of bacterial growth. Then, the culture was diluted 100-fold in 350 mL of LB culture medium and incubated at  $37^{\circ}\text{C}$  and 250 rpm agitation until it reached an optical density at 600nm (OD<sub>600</sub>) of 0.7. Subsequently, the flask was rapidly put on ice to stop cell growth, and centrifuged during 15 min at 4000 g and  $4^{\circ}\text{C}$ . After centrifugation, the pellet was resuspended in 8 mL of ice-cold LB with 15% glycerol in a volume 40 times lower

than the original culture volume. Then, the cell suspension was dispensed in 250  $\mu\text{L}$  aliquots and flash-frozen with liquid nitrogen. These aliquots were stored at  $-70^{\circ}\text{C}$ .

### 3. Plaque assays.

When a bacterial culture medium is inoculated with a suspension of bacteriophages containing agar (0.7% v:v), infection develops in a solid medium which limits the spread of the virus and plaques are formed. Plaques are visualized as clear regions in the bacterial lawn. In theory, plaques are produced from a single infectious particle or plaque forming unit (pfu). The plaque assay is widely used for titration. The titer of a virus suspension is defined as the number of infectious particles per volume unit (mL). Since  $\phi\text{X174}$  titers are very high ( $\sim 10^9$  pfu/mL), ten-fold serial dilutions must be made to observe a countable number of plaques. The titer of a phage is given by the following expression:

$$\text{titer}_{(\text{pfu}/\text{mL})} = \frac{\text{number of plaques}}{\text{dilution}} \times 100 \quad (\text{Equation 1})$$

The factor 100 is used to express the titer in pfu/mL units because all plaque assays were carried out with 10  $\mu\text{L}$  of the appropriate dilution. Plaque assays were performed widely in this thesis to determine viral titers at different stages of the experiments and during experimental evolution. Plaque assays were also performed to isolate individual plaques of the different mutants obtained by site-directed mutagenesis.

### 4. Growth rate estimation: $\phi\text{X174}$ growth assays.

Growth assays allow us to obtain an estimate of the viral fitness expressed as the growth rate ( $r$ ). The experiment consists of inoculating bacterial cultures with a known number of viruses and sampling at different time points. In order to determine how the titer changes over time, samples were titrated. All viruses were assayed in the same experimental block and each virus and/or treatment was assayed in triplicate and using the wild type virus as an internal control. To do so, 0.5 mL ( $\sim 5 \times 10^7$  cells) of exponentially growing cells were infected with  $\sim 10^4$  pfu. Two mL microtubes were incubated at  $37^{\circ}\text{C}$  and 650 rpm in a Thermomixer shaker (Eppendorf). Samples were taken at different time points (usually every hour during 4 hours) followed by centrifugation (1 min at 13000g) to remove cells, and the supernatants were aliquoted and stored at  $-70^{\circ}\text{C}$  until titration.

Viral growth was modeled using a logistic function given by:

$$N_t = \frac{K}{(1 + e^{c - rt})} \quad (\text{Equation 2})$$

where  $N_t$  refers to titer at time  $t$ ,  $r$  is the growth rate expressed in hours<sup>-1</sup>,  $K$  denotes the maximal viral titer,  $t$  indicates the post inoculation time expressed in hours, and  $c$  sets the initial conditions.

This kind of function implies that populations initially grow exponentially. Next, as population size increases, populations begin to grow slower due to interactions among members of the population and/or competition for a critical resource (i.e. host cells). Finally, the virus population loses its ability to grow and the titer remains constant. Data obtained from samples at different time points were fitted by nonlinear least-squares regression to the above logistic function using SPSS Software, allowing to infer the values of  $r$ ,  $K$  and  $c$ . Each virus was assayed in the same experimental block and at least three replicates were done.

Viral exponential growth rate ( $r$ ) was sometimes also estimated by adjusting the exponential growth phase to the linear form of the exponential growth function as:

$$r = \frac{\ln(N_t/N_0)}{t} \quad (\text{Equation 3})$$

Where  $N_t$  and  $N_0$  refer to the final and initial number of viruses obtained from fluctuation test experiments, and  $t$  denotes the incubation time in hours.

## 5. Viral mutation rate estimation by the Luria-Delbrück fluctuation test.

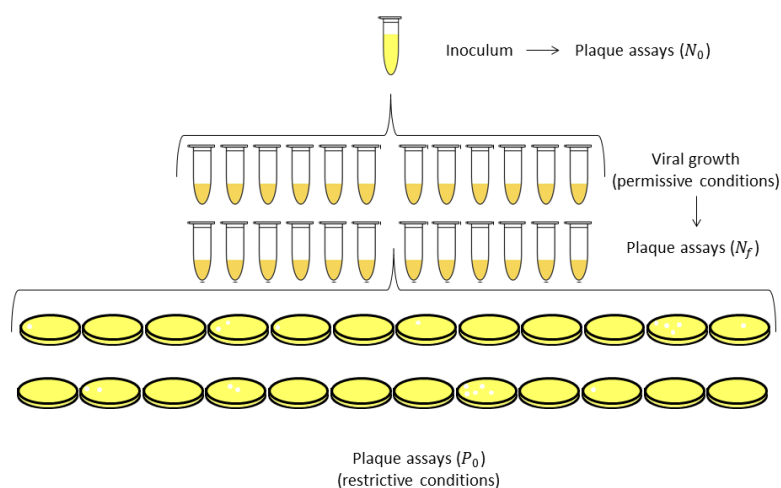
The fluctuation test is a classical experiment originally devised by Luria and Delbrück to show that mutations are produced randomly and not in response to a selection pressure in bacteria growing exponentially (Luria and Delbrück, 1943). Therefore, the fluctuation test assumes that mutations are random in all cultures assayed and allows us to determine the mutant distribution from a series of parallel cultures. Then, mathematical analysis of this distribution allows us to obtain a mutation rate estimate for a given phenotypically selectable trait.

To perform this experiment it is necessary to have a phenotypic marker that is neutral under permissive conditions but acts as a selective agent, allowing quantitation of the number of mutants. In the experimental system used in this thesis, the restrictive strain of *E. coli*, *gro87*, was used. *gro87* cells have a mutation in the *rep* gene, which encodes for a protein with DNA helicase and ssDNA-dependent ATPase activities (Gilchrist and Denhardt, 1987). The *rep* gene is not essential for the bacteria but its expression is necessary for  $\phi$ X174 replication because, when this protein is absent, the  $\phi$ X174 replicative cycle is blocked (Tessman and Peterson, 1976) at stage II of viral DNA synthesis where the replicative form I amplification (RFI) occurs (Eisenberg et al., 1976). Therefore, only mutant phages carrying specific mutations in the N-

terminal domain of the viral A protein will form plaques (Cuevas et al., 2009). Resistant phages to *gro* strains of *E. coli* are called *ogr* mutants for “overgrowth”.

Each fluctuation test consisted at the onset of a block of 24 parallel independent cultures with a relatively small known inoculum that was free of *ogr* mutants (<1000 pfu/culture) and incubated under permissive conditions (37°C and 650 rpm in a Thermomixer shaker (Eppendorf)) until approximately half of the cultures undergo at least one resistance mutation. A pilot growth assay was carried out in the same conditions that the fluctuation test would be performed to determine the viral growth rate. The incubation time was adjusted to achieve a final number of phages ( $N_f$ ) that corresponded to a  $P_0$  of around 0.5 and assuming an *a priori* phenotypic mutation rate (i.e  $10^{-6}$  from Cuevas et al., 2009 for the WT virus). If after perform the fluctuation test was needed to adjust the  $P_0$  value, we modified the  $N_f$  value and recalculate the incubation time using the equation 3.

After incubation, cultures were centrifuged (1 minute at 13000g) to remove cells and plated. The inoculum was plated in triplicate at the appropriate dilution in the permissive host in order to determine the average initial number of viruses per culture. In addition, to test whether the inoculum was free of mutants, the volume equivalent to a full fluctuation test was plated in duplicate in a lawn of the restrictive host. Finally, plaque assays of six cultures were performed in the absence of the phenotypic marker in order to determine the final number of viruses and all cultures were plated in the presence of the restrictive host in order to determine the number of cultures with at least one mutant (Figure 11).



**Figure 11. Luria Delbrück fluctuation test design.** Twenty-four parallel independent cultures are infected with a small phage inoculum ( $N_0$ ) and subsequently independent cultures are incubated under standard or permissive conditions. The supernatant is plated in a restrictive cell host after a final number of phages are produced ( $N_f$ ). The phenotypic mutation rate was estimated using the null-class method ( $P_0$ ).



The null class method was used to calculate the rate of *ogr* mutant occurrence. The null class method is based on the calculation of the proportion of cultures showing no mutants. As mutations are both rare and random events, the number of cultures showing zero or at least one mutant follows a Poisson distribution with parameter  $\lambda = m(N_f - N_0)$  and the expected probability of cultures with no mutants, also called null class ( $P_0$ ) is given by:

$$P_0 = e^{-m(N_f - N_0)} \quad (\text{Equation 4})$$

Where  $m$  is the phenotypic mutation rate and  $N_f$  and  $N_0$  refer to the average number of pfu per culture at the end and the beginning of the experiment respectively.

Therefore, this experiment allows us to empirically determine all the parameters, except the mutation rate to the selectable phenotype, that can be directly inferred by rewriting the previous equation as:

$$m = \frac{-\ln P_0}{N_f - N_0} \quad (\text{Equation 5})$$

To express the mutation rate in substitutions per nucleotide per round of copying  $\mu_{(s/n/r)}$  it is necessary to know the mutational target ( $T_s$ ), which is the total number of different substitutions leading to the *ogr* phenotype, and the mutation rate then is given by:

$$\mu = \frac{3m}{T_s}, \quad (\text{Equation 6})$$

where the number three is a correction factor which refers to the fact that for each position a base can mutate to the other three bases and  $T_s$  is the mutational target for substitutions that can be obtained by sequencing the region involved in the resistance phenotype. In our experimental conditions  $T_s$  is equal to seven (Cuevas et al., 2009).

Alternatively, the phenotypic mutation rate can be inferred using the method of means. The method of means is based on the mean number of mutants of the observed complete distribution. In this case, the distribution of mutants follows a Luria-Delbrück distribution instead of a Poisson distribution. Many mathematical models were developed in order to calculate the Luria-Delbrück distribution depending on different assumptions (Zheng, 1999). The method of means also assumes that mutants are neutral under permissive conditions and binary replication (Drake, 1991). This last assumption is not suitable for calculating the mutation rate of  $\phi$ X174 because the virus replicates its genome by the rolling-circle mechanism (Eisenberg et al., 1977; Gillam et al., 1984). All mutation rate estimates obtained in this study were calculated using the Poisson method because the probability that no mutants are produced depends only on the mutational process, regardless of the replication mode. Data were log transformed for statistical analysis.

Fluctuation tests were performed for different mutants and/or conditions, including stress conditions such as increasing the incubation temperature and the presence of a chemical mutagen (5-fluoruracil, 5-FU). In the estimation of mutation rates using the fluctuation test experiment, it is usually assumed that all mutants are able to form visible plaques with the same efficiency under different treatments. To compare the mutation rate estimates obtained under stress conditions, we applied a correction factor that takes into account the possible bias in the estimation of the number of pfu at the end of the experiment ( $N_f$ ) relative to standard plating conditions ( $z'$ ). The expected probability of the null class ( $P_0$ ) can be rewritten, taking into account this bias in plating efficiency, as:

$$P_0 = e^{-m\left(\frac{N_f}{z'} - N_0\right)} \quad (\text{Equation 7})$$

Where  $z'$  refers to the plating efficiency under stress conditions relative to standard plating conditions and only affects to the  $N_f$  estimate because all titrations were made at permissive conditions. Plating efficiency for each stress condition was calculated by three independent plaque assays carried out under the stress conditions used during fluctuation test experiments and without them.

## **6. Life history traits determination in $\phi$ X174.**

### **6.1 Determination of intracellular virion accumulation, adsorption rate and eclipse time.**

These assays consist of performing curves in an analogous way to the growth curves described previously, allowing us to estimate several life history traits of  $\phi$ X174 such as the rate of virion accumulation within cells, the adsorption rate and the eclipse time. The intracellular virion accumulation rate is the velocity at which the virions are produced inside infected cells and is related to viral eclipse time. The adsorption rate is a measure of the velocity at which free phages are attached to host cells. Eclipse time is the time elapsed between viral adsorption and maturation of the first progeny. Empirically, this eclipse period is observed in the laboratory because no plaques are formed after treating a culture of infected cells with chloroform. Chloroform kills cells by disrupting cell membranes and hence preserves only mature virions.

In this experiment, independent parallel exponentially growing bacterial cultures (0.5 mL,  $\sim 10^8$  cells/mL) were infected at a multiplicity of infection (moi) of  $\sim 1$  virion/cell. After allowing viral adsorption, cultures were diluted  $10^4$ -fold to prevent subsequent infection cycles and incubated at 37°C and 650 rpm in a Thermomixer shaker (Eppendorf). Samples were taken at different time points (every two or three minutes, depending on the growth conditions) until

shortly before lysis time. Samples were treated with chloroform (10% v:v) to lyse cells and titrated in order to determine the number of mature phage particles. All viruses were assayed at the same time in different experimental blocks and three replicates were done for each virus and growth condition. Empirical data were fitted by nonlinear least-squares regression to the following mathematical model:

$$N_t = N_0 e^{-aHt} + \frac{B}{1 + e^{c-Rt}} \quad (\text{Equation 8})$$

where  $N_t$  denotes the viral titer at time  $t$ ,  $N_0$  is the initial experimentally determined titer,  $a$  is the adsorption rate expressed as mL/min,  $H$  indicates the cell density expressed as cells/mL,  $t$  is the time in minutes,  $B$  denotes the burst size or maximal yield per cell,  $R$  is intracellular virion accumulation and  $c$  set the initial conditions.

This mathematical model describes, in the first term of the equation, the beginning of the viral cycle, where titer loss occurs due to viral adsorption. The second term of the equation describes an intracellular logistic growth. This model was fit by nonlinear least-squares regression in SPSS using experimentally determined  $N_0$ ,  $H$ , and  $B$  values to infer  $R$ ,  $c$ , and  $a$ . The inferred values from this model were used to estimate the eclipse period, which is defined as the time period comprising from inoculation to phage maturation. Since at  $t = E$  we have  $N_t = N_0$ , the eclipse time is given by:

$$E = \frac{c - \ln\left(\frac{B}{N_0}\right)}{R} \quad (\text{Equation 9})$$

$N_0$  and  $B$  were experimentally determined and correspond to the initial titer and burst size, respectively. The rest of the parameters were inferred from the model given by equation 8.

## 6.2 Lysis time determination.

To estimate the lysis time, independent parallel exponentially growing bacterial cultures (0.5 ml,  $\sim 10^8$  cells/mL) were infected at an moi of  $\sim 3$  virion/cell in order to favor the rapid infection of the complete culture and incubated at 37°C and 650 rpm in a Thermomixer shaker (Eppendorf). Lysis was calculated by measuring the  $OD_{600}$  using a luminometer (BioPhotometer, Eppendorf) at different time points. All viruses were assayed in the same experimental block and each virus and growth condition was assayed in triplicate.

In order to estimate the average lysis time ( $L$ ), defined as the time required to lyse 50% of the cells, experimental data were fitted by nonlinear least-squares regression to the logistic model given by:

$$OD_{600}(t) = OD_{min} + \frac{(OD_{max} - OD_{min})}{1 + \left(\frac{t}{L}\right)^h} \quad (\text{Equation 10})$$

where  $t$  denotes the post inoculation time in minutes,  $L$  is the lysis time,  $h$  is the maximal slope of the lysis curve (Hill coefficient), and  $OD_{min}$  and  $OD_{max}$  are the initial and final optical densities, respectively. Each virus was assayed at least in triplicate and under different growth conditions.

### 6.3 Burst size determination.

When the viral lytic cycle is fully accomplished, so that additional cycles of infection can no longer occur, it is possible to estimate the burst size or number of new viruses produced by an infected cell. To do so, exponentially growing cells (0.5 ml;  $10^8$  cells/ml) were infected at an moi of 0.5 pfu/cell and, after adsorption, cultures were diluted  $10^4$ -fold to avoid subsequent infection cycles. Samples from time zero were titrated to determine the total number of pfu ( $N_{tot}$ ), i.e. the number of free particles plus the number of pfu that infected cells. In addition, another aliquot of the same time point was treated with 10% v:v chloroform to lyse cells and then was titrated to determine the number of free particles ( $N_{free}$ ) alone, since intracellular viral particles were not mature at this time point and, therefore, did not yield plaques in chloroform. This approach allows for the calculation of the number of infected cells ( $N_c$ ) as:

$$N_c = N_{tot} - N_{free} \quad (\text{Equation 11})$$

Where  $N_{tot}$  refers to the total number of pfu at time zero and  $N_{free}$  is the number free particles at time zero.

Next, infected cultures were incubated at 37°C and 650 rpm in a Thermomixer shaker (Eppendorf) for approximately twice the lysis time of the viruses assayed until completion of the infection cycle and then titrated (without chloroform) to determine the final number of pfu ( $N_1$ ). All viruses were assayed in the same experimental block and each virus and growth condition were assayed in six replicates.

The burst size was calculated as:

$$B = \frac{(N_1 - N_{free})}{N_c} \quad (\text{Equation 12})$$

where  $N_1$  is the final titer,  $N_{free}$  is the number of free pfu and  $N_c$  refers to the number of infected cells.

## 7. Experimental evolution experiments.

Phage populations were evolved by serial passage. To do so, exponentially growing cells (0.5 ml;  $10^8$  cells/mL) were infected with  $\sim 10^6$  pfu and incubated at 37°C and 650 rpm in a Thermomixer shaker (Eppendorf) until the end of the phage population's growth. At this point,

cells were removed by centrifugation (1 min at 13,000g) and the appropriate dilution of the supernatant was used to start the next passage. Bacterial cells that may adapt to phages were thereby discarded and only phage populations should therefore evolve. The last passage of the day was titrated to adjust the viral inoculum of the next passage, if necessary. Each evolution experiment consisted of five independent lines for each virus and host, and 100 passages were performed for all evolution lines. All experimental lineages were evolved in the same block.

Finally, the fitness of evolved viruses relative to the founder ones was expressed as the growth rate increment ( $\Delta r$ ) as:

$$W = r - r_0 \quad (\text{Equation 13})$$

Where  $r$  and  $r_0$  refer to the growth rate of the evolved and founder viruses, respectively.

### **8. Site-directed mutagenesis.**

Site-directed mutagenesis is a technique to introduce specific targeted nucleotide substitutions or even insertions or deletions in a dsDNA molecule. It has many purposes in molecular biology and was used in this thesis to create many different mutants that had a desired property. The basic procedure used here is based on the PCR amplification of the  $\phi$ X174 whole genome using a high-fidelity DNA polymerase and a pair of adjacent, divergent, 5' phosphorylated primers, of which the reverse primer carried the desired specific mutation to be introduced. Finally, amplicons were circularized and transfected by the heat shock method to recover mature mutant virions. Since  $\phi$ X174 is a single stranded DNA virus, it is necessary to isolate its double stranded DNA form (RFI). Double stranded replicative forms are produced inside cells during viral replication and can be isolated using a standard miniprep kit (Qiagen or Macherey-Nagel).

Mutagenic primers designed in this study were on average longer than normal PCR primers (between 25-40 bases), especially the reverse primer, which carried the mutation (for the entire list of the primers used in this study, see Supplementary information 1 and 3). In addition, the reverse primer had the mismatch as close to the center as possible. In order to increase the specificity of the reaction, the GC content and  $T_m$  of the mutagenic primers were higher than 35% and 65°C, respectively (after taking into account the percentage of mismatches). Finally, the high-fidelity DNA polymerase used is highly processive and has an error rate about 50-fold lower than Taq polymerase, as indicated by the manufacturer.

The detailed site-directed mutagenesis protocol is the following:

Primers were phosphorylated using T4 polynucleotide kinase enzyme (PNK; Thermo Scientific) following the manufacturer's instructions. The PNK conditions used were 37°C for 30 min followed by enzyme inactivation during 10 min at 70°C. Next,  $\leq 500$  pg of  $\phi$ X174 dsDNA were used as template together with the primers phosphorylated previously for a PCR-based mutagenesis with Phusion high-fidelity DNA polymerase (Thermo Scientific).

The PCR conditions were: initial denaturation step of 98°C for 2 min, 30 cycles of 98°C for 10 s, 68°C for 30s, 72°C for 3 min, and a final extension of 5 min at 72°C. PCR products were circularized with the Rapid DNA ligation kit (Thermo Scientific) and used for transfecting chemically competent IJ1862 cells by the heat-shock method (42°C, 35 s). Transfected cells were plated and incubated at 37°C until plaques were visible. Finally, individual plaques were isolated and stored at -70°C.

Both the presence of the mutations introduced and the absence of additional changes were confirmed by direct Sanger sequencing of the purified PCR products from the resuspended plaque. The PCR conditions used were 98°C for 2 min, 30 cycles of 98°C for 10 s, 68°C for 30 s, 72°C for 1 min and a final extension of 5 min at 72°C. PCR products were purified with Macherey-Nagel NucleoFast® 96 PCR plates.

## 9. Methylation analysis

The  $\phi$ X174 dsDNA replicative form was quantified using the Quant-iT PicoGreen dsDNA broad range assay kit (Life technologies), and all extracts were brought to the same concentration (30 ng/mL). DNA from each virus was split into three aliquots, which were treated with *XhoI* to linearize the genome, with *XhoI* and *DpnI* to digest methylated GATCs, or with *XhoI* and *MboI* to digest non-methylated GATCs. Double digestions were performed according to the manufacturer instructions (Thermoscientific). A standard plasmid (pIRES, Clontech) was used as a digestion control. A monochrome picture of the gel was transformed to an eight-bit image, and the pixel area and intensity of each band were quantified using ImageJ.

## 10. Sequence analysis.

Sequences from Sanger sequencing were analyzed using the Staden package v1.7.0 (<http://staden.sourceforge.net>). MEGA v5 (<http://www.megasoftware.net>) and GeneDoc v320 (<http://iubio.bio.indiana.edu/soft/molbio/ibmpc/>) were used to perform sequence analysis.

## RESULTS AND DISCUSSION





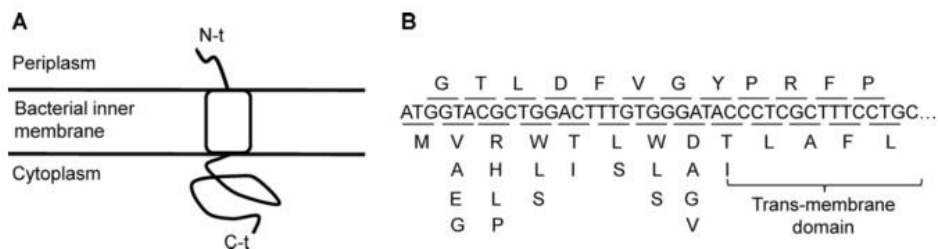
## **Chapter I: Lysis and mutation rate.**



We have previously shown that experimental evolution of  $\phi$ X174 populations in the presence of the nucleoside analogue 5-FU selected for one amino acid replacement in lysis protein E which conferred resistance to the drug (Domingo-Calap et al., 2012). Since viral lysis proteins have not been implicated previously in nucleoside analogue resistance, we investigated the role played by the N terminal domain of the E protein in determining the lytic properties, 5-FU resistance, and the mutation rate of bacteriophage  $\phi$ X174.

**1. V2A and D8A replacements in the N terminal domain of the E protein produce delayed lysis and also confer resistance to 5-FU.**

We designed and constructed a library of sixteen site-directed mutants of the N terminal domain of the E protein, which has been implicated previously in determining lysis time (Tanaka and Clemons, 2012). We changed residues 2 to 9 with no alteration to the amino acid sequence of the overlapping gene D, which is in a different reading frame (Figure 12), in order to determine their effects on lysis time. Of the total mutants, 75% were successfully amplified (12/16) and assayed for lysis time determination. Optical density measurements of infected cultures showed evidence for delayed lysis compared to the WT in 50% cases (6/12), whereas the other six mutations changed lysis time only slightly in either direction (Table 1).



**Figure 12. Single-amino-acid replacements introduced in the N-terminal region of the lysis protein E. (A)** Scheme of the cytoplasmic, transmembrane, and periplasmic domains of the lysis protein E. All substitutions were introduced in the periplasmic domain. **(B)** Nucleotide sequence of the region studied, showing the two overlapping reading frames (top, gene D; bottom, gene E) and the amino acid changes introduced. Figure originally published in Pereira-Gómez and Sanjuán 2014 and reproduced with permission from the American Society for Microbiology.

A preliminary growth assay, carried out to test the ability of these six mutants showing delayed lysis to grow in the presence of 10 ng/ $\mu$ L 5-FU, allowed us to determine the fold increase in titer after 3 h of infection ( $R_{FU}$ ) (Supplementary information 2). This preliminary assay showed that viruses carrying substitutions V2A ( $R_{FU} = 21.0 \pm 7.1$ ) and D8A ( $R_{FU} = 6.3 \pm 1.6$ ) had improved ability to grow in the presence of 5-FU than the non-mutated WT virus ( $R_{FU} = 3.7 \pm 0.9$ ). In order to confirm this, we performed standard growth curves of these viruses in the presence and the

**Table 1. Lysis time of N-terminal mutants of the lysis protein E.**

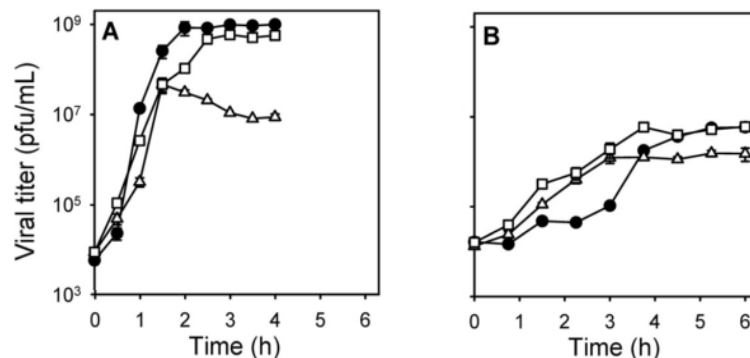
Nucleotide substitution	E protein	Lysis time (min) $\pm$ SEM <sup>1</sup>
None	WT	24.5 $\pm$ 0.2
T572C	V2A	36.2 $\pm$ 0.6***
T572A	V2E	27.8 $\pm$ 0.2
T572G	V2G	NA <sup>2</sup>
G575A	R3H	24.4 $\pm$ 1.0
G575C	R3P	38.0 $\pm$ 0.2**
G575T	R3L	23.6 $\pm$ 0.3
G578C	W4S	NA <sup>3</sup>
G578T	W4L	27.6 $\pm$ 0.4*
C581T	T5I	23.2 $\pm$ 0.1*
T584C	L6S	23 $\pm$ 0.6*
G587T	W7L	35.9 $\pm$ 1.1**
G587C	W7S	36.3 $\pm$ 0.8***
A590T	D8V	NA <sup>3</sup>
A590C	D8A	38.5 $\pm$ 1.0***
A590G	D8G	NA <sup>3</sup>
C593T	T9I	33.8 $\pm$ 1.0**

<sup>1</sup> Asterisks indicate values significantly different from those of the WT (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , as determined by the t test).

<sup>2</sup> Mutagenesis failed.

<sup>3</sup> Low fitness mutants, not achieved high titer, based on titration assays.

absence of 10 ng/ $\mu$ L 5-FU (Figure 13).

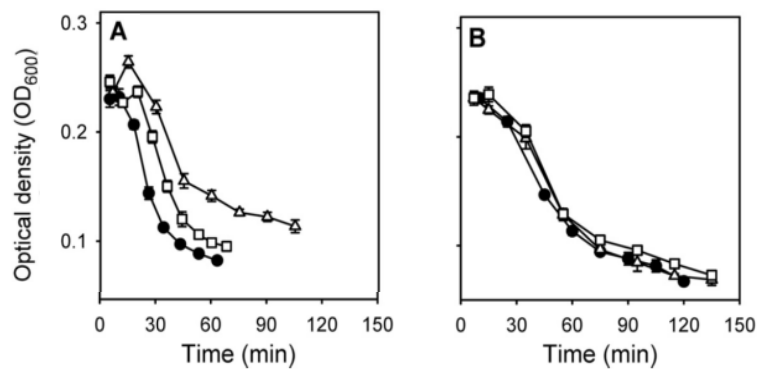


**Figure 13. Standard growth curves of WT (black circles), V2A (white squares), and D8A (white triangles) viruses.** Assays were performed in the absence (A) and the presence (B) of 10ng/ $\mu$ L 5-FU. Black circles, white squares, and white triangles correspond to WT, V2A, and D8A, respectively. Error bars indicate the standard errors of the means. This figure is a variation of that originally published in Pereira-Gómez and Sanjuán 2014 and reproduced with permission from the American Society for Microbiology.

The standard growth curves showed that both substitutions increased the population growth rate ( $r$ ) in the presence of 5-FU but were costly in the absence of the drug (Figure 13 and Table 2). The V2A and WT viruses showed similar maximal viral titers ( $K$ ) in the presence of 5-FU, indicating that growth rate is the main fitness component determining 5-FU resistance, whereas the D8A mutant showed strongly reduced maximal viral titers. Maximal viral titers

depend on the ability of the virus to infect the culture rapidly or to continue replicating as the cell population approaches the stationary phase. Therefore, the interpretation of this result is less clear.

Lysis assays of WT, V2A, and D8A viruses carried out in the presence of 10 ng/ $\mu$ L 5-FU (Figure 14B) showed that the drug produced a strong delay in viral lysis times. V2A and D8A mutants again exhibited significantly delayed lysis compared with the WT, although differences were small (between 3 to 6 min) (Table 2).

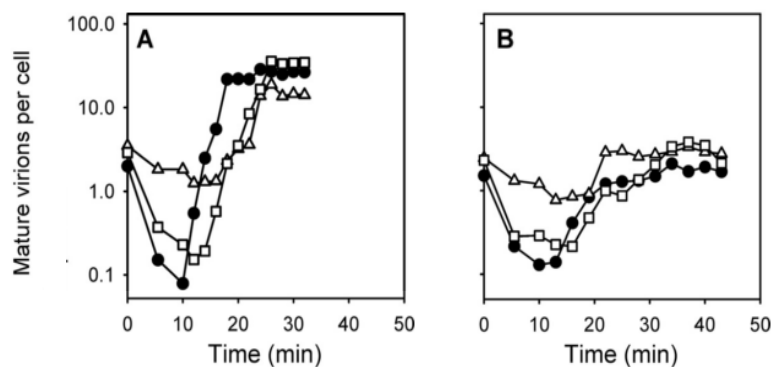


**Figure 14. Lysis assays of WT (black circles), V2A (white squares), and D8A (white triangles) viruses.** Assays were performed in the absence (A) and presence (B) of 10ng/ $\mu$ L 5-FU. Error bars indicate the standard errors of the means. This figure is a variation of the figure originally published in Pereira-Gómez and Sanjuán 2014 and reproduced with permission from the American Society for Microbiology.

## 2. Replacements V2A and D8A produce a delay in the intracellular accumulation of mature virions.

We determined the intracellular accumulation of mature virions during a single infection cycle. The curves showed that during the first ten minutes post inoculation, the number of mature virions declined as a result of viral adsorption. Next, the number of mature virions increased exponentially as newly synthesized virions accumulated inside infected cells, finally plateauing 5 to 10 min before lysis in the absence of 5-FU and 10 to 20 min before lysis in the presence of 5-FU (Figure 15).

We used this experimental data to infer the viral adsorption rate ( $a$ ) and the eclipse time ( $E$ ) by fitting data to a model describing an initial exponential decay term followed by a logistic intracellular growth function (See equation 8 from Material and methods) (Table 2). In the absence of 5-FU, mutants V2A and D8A showed retarded intracellular accumulation of mature virions, mainly due to longer eclipse times than the WT. On the other hand, particle accumulation plateaued later in the V2A and D8A mutants (approximately 26 min post



**Figure 15. Intracellular growth assays of WT (black circles), V2A (white squares), and D8A (white triangles) viruses.** Assays were performed in the absence (A) and presence (B) of 10ng/μL 5-FU. Error bars are not shown for clarity. This figure is a variation of the originally published figure in Pereira-Gómez and Sanjuán 2014 and reproduced with permission from the American Society for Microbiology.

inoculation) than the WT (approximately 18 min post inoculation). Mutants V2A and D8A also seemed to show retarded yet prolonged virion accumulation in the presence of 5-FU, although the inference of eclipse times was complicated by the high error produced by low and fluctuating viral titers. We have been able to infer the eclipse time from only two (WT,  $E = 20.1 \pm 0.2$ ) and one (V2A,  $E = 33.0$ ) of the three replicates due to insufficient viral growth in the other replicates.

### 3. Mutations V2A and D8A increase the viral burst size in the presence of 5-FU.

To assess if delayed lysis time allows viruses to increase the period for accumulation of mature viral particles inside cells and thereby increase the per-cell viral progeny or burst size ( $B$ ), we measured the burst size of each virus in the presence and the absence of 5-FU. For this, we determined the amount of infectious viral particles released from a known number of infected cells. In the absence of 5-FU we observed that the V2A mutant ( $B = 37.2 \pm 7.5$ ) showed greater burst size than the WT ( $B = 25.4 \pm 4.4$ ), whereas the D8A mutant ( $B = 18.2 \pm 1.5$ ) had a lower burst size but in all cases the differences were not statistically significant (Table 2). The lower burst size of the D8A mutant can be explained by less efficient lysis, as is indicated by the corresponding minimal optical density values in the lysis assays. Delayed lysis is expected to provide a payoff between increased burst size and delayed infection of new cells. However, our results suggest that the payoffs of increasing lysis time are absent or, alternatively, we cannot detect them under our laboratory conditions. Assays performed in the presence of 5-FU showed that the drug is extremely harmful and strongly decreased the viral burst size. However, both V2A ( $B = 2.47 \pm 0.19$ ) and D8A ( $B = 3.01 \pm 0.24$ ) mutants showed significantly higher burst size than the WT ( $B = 1.73 \pm 0.19$ ), indicating that mutant viruses are able to produce more progeny per cell than the WT under these conditions.

#### 4. The changes in lysis time and burst size produced by V2A and D8A explain 5-FU resistance.

In theory, since lysis time determines growth rate, viral fitness should be maximized for some intermediate, optimal lysis time (Bull et al., 2004; Chantranupong and Heineman, 2012; Heineman and Bull, 2007). Bull (2006) suggested that this optimum ( $\hat{L}$ ) can be calculated as a function of the eclipse time ( $E$ ) and the maximal population growth rate ( $\hat{r}$ ) as:

$$\hat{L} = E + 1/\hat{r} \quad (\text{Equation 14})$$

This model assumes that mature virions accumulate inside cells linearly at rate  $R$  until they are released by lysis, such that the burst size is:

$$B = R(L - E) \quad (\text{Equation 15})$$

Since our results indicate that intracellular mature virion accumulation was exponential and plateaued before lysis, optimal lysis times cannot be derived from these expressions. Therefore, we used a simple model to predict viral growth rates using the experimentally determined  $B$  and  $L$  values. If cells are not a limiting factor, viral growth follows an exponential growth function that can be described as:

$$N_g = N_0 B^g \quad (\text{Equation 16})$$

Where  $N_0$  is the initial number of pfu and  $g$  the number of viral generations. This equation can be rewritten as:

$$N_g = N_0 e^{rLg} \quad (\text{Equation 17})$$

Where  $r$  corresponds to the population growth rate and  $L$  is the lysis time, which is equivalent to the generation time. Therefore:

$$r = \ln B/L \quad (\text{Equation 18})$$

The predicted population growth rates (Table 2) showed a correlation with the observed rates (Pearson,  $r = 0.986$ ,  $P < 0.001$ ) explaining that the changes in burst size and lysis time produced by V2A and D8A mutations should be deleterious in the absence of 5-FU but beneficial in the presence of the nucleoside analogue (Figure 16).

**Table 2. Characteristics of the WT, V2A, and D8A viruses in the absence/presence of 5-FU.**

Virus and treatment	$r$ ( $\text{min}^{-1}$ )	$K$ (pfu/mL, $\times 10^{-8}$ )	$L$ (min)	$B$ (pfu/cell)	$aH$ (mL/min) <sup>1</sup>	$R$ ( $\text{min}^{-1}$ )	$E$ (min)	$r$ predicted <sup>2</sup>	$r$ predicted <sup>3</sup>
None									
WT	0.130±0.002	9.21±1.12	24.5±0.2	25.4±4.4	0.524±0.062	0.877±0.027	13.6±0.2	0.132	0.123
V2A	0.089±0.002***	5.01±0.50*	36.2±0.6***	37.2±7.5	0.287±0.028*	0.540±0.049***	19.3±0.3***	0.099	0.092
D8A	0.090±0.005**	0.16±0.01***	38.5±1.0***	18.2±1.5	0.111±0.021**	0.400±0.148	20.9±0.9**	0.075	0.064
5-FU									
WT	0.021±0.001	0.059±0.010	43.0±0.8	1.73±0.19	0.379±0.052	0.348±0.057	20.1±0.9 <sup>5</sup>	0.013	0.012
V2A	0.029±0.002*	0.061±0.003	46.3±0.7*	2.47±0.19*	0.279±0.181	0.283±0.113	33.0 <sup>5</sup>	0.020	0.018
D8A	0.030±0.001**	0.015±0.005*	49.0±1.3**	3.01±0.24**	NA	NA	NA	0.022	0.019 <sup>6</sup>

<sup>1</sup>Using a cell density  $H = 10^8$  cells/mL.

<sup>2</sup>Predicted from  $r = \ln B/L$ .

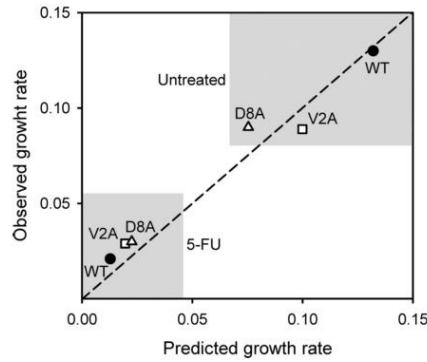
<sup>3</sup> Predicted from  $r = aH(Be^{-Lr} - 1)$

<sup>4</sup> Asterisks indicate values significantly different from those for the WT (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , as determined by the t test)

<sup>5</sup> Calculated from two (WT) and one (V2A) of the three replicates due to insufficient viral growth in the other replicates.

<sup>6</sup> Using  $aH = 0.111$ .





**Figure 16. Expected effects of changes in lysis time and burst size on viral growth rate.** The growth rates of the WT (black circles), V2A (white squares), and D8A (white triangles) viruses were predicted based on their experimentally determined lysis times and burst sizes (x axis) and compared to measured growth rates (y axis) in the presence and in the absence of 10 ng/μL 5-FU. Figure originally published in Pereira-Gómez and Sanjuán 2014 and reproduced with permission from the American Society for Microbiology.

To confirm the FU-dependent fitness effects of V2A and D8A mutations we also determined the intracellular accumulation of virions to infer the adsorption rate and predict the population growth rates (Table 2), using a more complex growth model taken from Heineman and Bull (2007), where the population growth rate is given by:

$$r = aH(Be^{-Lr} - 1) \quad (\text{Equation 19})$$

Where  $a$  is the adsorption rate expressed as mL/min, and  $H$  indicates the cell density expressed as cells/mL. Under this model, the predicted growth rates were similar to those estimated from the simpler model and were also correlated with the observed values (Pearson,  $r = 0.979$ ,  $P < 0.001$ ).

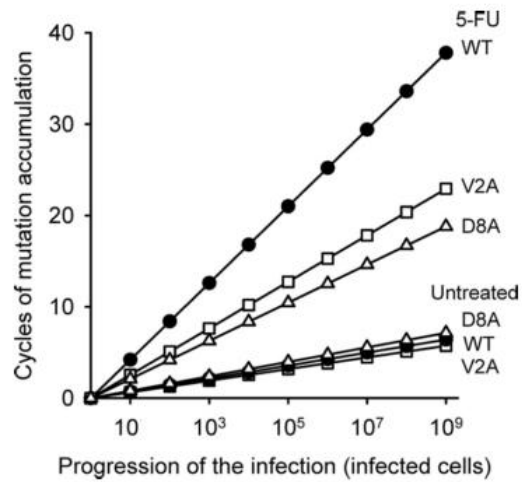
Since optimal lysis times should depend on cell physiology (Bull et al., 2004; Wang et al., 1996), 5-FU is likely to affect them, as it produces multiples changes in the cell metabolism, comprising severe altered replication, transcription, and translation; thymidine deprivation; DNA strand breaks; and anomalous cell wall synthesis, which can lead to lysis (Adelberg and Coughlin, 1956; Drake and Greening, 1970). There are at least two explanations for why the V2A and D8A substitutions increase viral fitness in the presence of 5-FU. The first one is that the drug should slow down phage replication and morphogenesis; therefore, delayed lysis may provide the extra time needed to produce enough mature virions. Second, since both 5-FU and the E protein cause lysis by interfering with cell wall synthesis, 5-FU may lower the threshold concentration of the E protein required for lysis, and the mutant proteins may have higher threshold concentrations than the WT, compensating for the effect of 5-FU.

## 5. V2A and D8A mutations alleviate the mutagenic effect of 5-FU through changes in burst size.

To assess whether V2A and D8A mutations help to counteract the mutagenic effect of the 5-FU, we used the Luria-Delbrück fluctuation test to estimate viral mutation rates in the presence and absence of 5-FU. The fluctuation test results of the WT virus in the absence of 5-FU gave an average spontaneous mutation rate value of  $\mu = 0.86 \pm 0.10 \times 10^{-6}$  s/n/r, which is consistent with previous works (Cuevas et al., 2009; Raney et al., 2004). V2A ( $\mu = 1.26 \pm 0.19 \times 10^{-6}$  s/n/r) and D8A ( $\mu = 0.44 \pm 0.10 \times 10^{-6}$  s/n/r) mutation rate estimates were not significantly different from the WT virus estimates (two-way ANOVA,  $P = 0.399$ ). The 5-FU treatment produced an overall increase in the mutation rate estimates of WT ( $\mu = 27.9 \pm 5.4 \times 10^{-6}$  s/n/r), V2A ( $\mu = 8.73 \pm 0.59 \times 10^{-6}$  s/n/r), and D8A ( $\mu = 16.2 \pm 4.2 \times 10^{-6}$  s/n/r) viruses. In contrast to the results obtained in the absence of 5-FU, the differences among viral mutation rate estimates were significant (two-way ANOVA,  $P = 0.004$ ) (Table 3).

We used the growth model described in equation 16 to test whether the greater burst size conferred by the V2A and D8A substitutions allows the virus to reduce the number of viral generations required for infecting a given number of cells and, therefore, to reduce the rate at which mutations accumulate. We used the average  $N_0$  and  $N_1$  values from fluctuation tests performed with the WT in the absence of 5-FU (Table 3) to calculate the number of viral generations, which was  $g = \frac{\ln \frac{N_1}{N_0}}{\ln B} = 2.39$  generations.

To achieve the same level of population expansion, the V2A and D8A mutants should modify the number of generations by 10% or less ( $g = 2.14$  and  $g = 2.67$  generations, respectively). In contrast, when we did the same calculation taking data for the average  $N_0$  and  $N_1$  values from fluctuation tests performed with the WT in the presence of 5-FU, the calculated number of WT generations was 6.81, whereas for the V2A and D8A mutants the number of generations were 4.13 and 3.39, respectively, which represent 1.7 and 2.0-fold reductions, respectively. These results indicate that, based on burst sizes estimates, the V2A and D8A substitutions should have a small effect on the rate of mutation accumulation in the absence of 5-FU, whereas they should significantly reduce these rates in the presence of the drug (Figure 17).



**Figure 17. Predicted number of viral generations required to infect a given number of host cells for the WT (black circles), V2A (white squares), and D8A (white triangles) viruses in the presence and absence of 5-FU.** For an infection starting with one PFU, generations were calculated as  $g = 1 + \frac{\ln H}{\ln B}$ , where  $H$  is the number of infected cells and  $B$  is the burst size. Figure originally published in Pereira-Gómez and Sanjuán (2014) and reproduced with permission from the American Society for Microbiology.

**Table 3. Mutation rate estimates from fluctuation tests in the absence/presence of 10 ng/μL 5-FU.**

Virus and treatment	Number of tests <sup>1</sup>	$N_0$ (pfu) <sup>2</sup>	$N_1$ (pfu, $\times 10^{-5}$ ) <sup>3</sup>	$P_0$ <sup>4</sup>	$\mu$ (s/n/r, $\times 10^6$ ) <sup>5</sup>
None					
WT	4	110, 432, 178, <u>323</u> <sup>6</sup>	5.2, 3.6, 2.3, <u>13.1</u>	0.33, 0.46, 0.58, <u>0.17</u>	0.864±0.096
V2A	3	192, 382, 468	3.6, 1.8, 1.4	0.38, 0.67, 0.58	1.26±0.19
D8A	5	<u>743</u> , <u>397</u> , <u>198</u> , <u>305</u> , <u>397</u>	<u>9.1</u> , <u>5.0</u> , <u>6.2</u> , <u>3.2</u> , <u>6.6</u>	<u>0.17</u> , <u>0.67</u> , <u>0.58</u> , <u>0.79</u> , <u>0.63</u>	0.436±0.102
5-FU					
WT	6	213, 90, 157, <u>375</u> , <u>343</u> , <u>462</u>	0.042, 0.092, 0.087, <u>0.15</u> , <u>0.23</u> , <u>0.084</u>	0.83, 0.67, 0.71, <u>0.17</u> , <u>0.25</u> , <u>0.54</u>	27.9±5.4
V2A	3	235, 140, 252	0.13, 0.15, 0.13	0.75, 0.75, 0.79	8.73±0.059**
D8A	3	<u>687</u> , <u>407</u> , <u>368</u>	<u>0.025</u> , <u>0.099</u> , <u>0.092</u>	<u>0.96</u> , <u>0.58</u> , <u>0.75</u>	16.2±4.2

<sup>1</sup>Number of independent fluctuation tests performed.

<sup>2</sup>Initial number of pfu per culture.

<sup>3</sup>Final number of pfu per culture.

<sup>4</sup>Fraction of 24 cultures showing no *gro87*-resistant plaques.

<sup>5</sup>For each test, the mutation rate was calculated as  $\mu = [-\ln P_0 / (N_1 - N_0)] \times 3 / T_S$ , where  $T_S = 7$ .

<sup>6</sup>Fluctuation tests were performed in two experimental blocks (differentiated by underlining).

<sup>7</sup>Asterisks indicate significant differences in mutation rates in a t test against the WT (\*\*,  $P < 0.01$ )

## **Chapter II: DNA repair and mutation rate.**



Regardless of the intrinsic replication fidelity of the polymerase, viral mutation rates depend on the access to cellular post-replicative DNA repair mechanisms. In a previous study, we showed that the introduction of four GATC sites in the  $\phi$ X174 genome had no effect on the mutation rate. However, by increasing the number of GATC sites to seven, we observed a 30-fold reduction in the viral mutation rate. In addition, this effect was reverted in MMR-deficient *mutD* cells, indicating that the phage DNA was capable of undergoing MMR (Cuevas et al., 2011). Here, we constructed several  $\phi$ X174 mutants in which the number and location of GATC sites were varied, including a variant carrying the expected number of GATC in the  $\phi$ X174 genome given its size and base composition, with minimal effects on protein sequence in order to study their effect on the mutation rate of  $\phi$ X174.

### 1. Introduction of GATC sites decreases the mutation rate of $\phi$ X174.

The reference genome sequence of  $\phi$ X174 lacks GATC sites. Given its genome size and genomic base composition (5386 nucleotides, 31.3% T, 21.5% C, 24% A and 23.2% G), the expected number of GATC sites by chance should be around 20 ( $0.313 \times 0.215 \times 0.240 \times 0.232 = 3.75 \times 10^{-3} \times 5386 \text{ bp} = 20.18 \text{ GATC sites/genome}$ ). The Poisson probability for finding no GATC sites in the  $\phi$ X174 genome when 20.18 are expected is given by the equation:

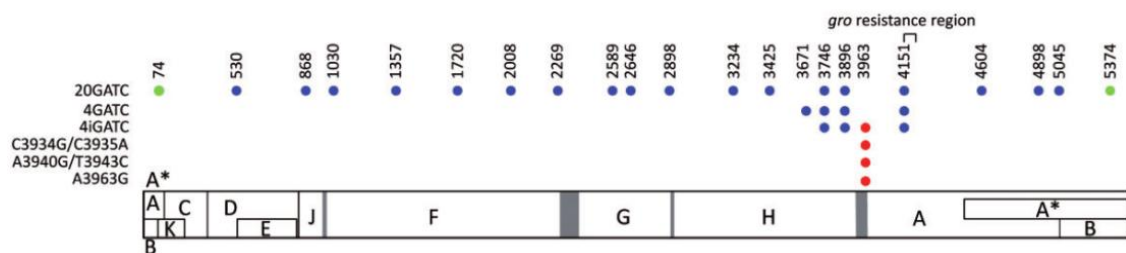
$$P(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (\text{Equation 20})$$

Where  $x$  is the number of GATC sites observed per genome, and  $\lambda$  is the expected number of GATC sites. According to this formula,  $P(0,20.18) = 1.72 \times 10^{-9}$ , an extremely low value which indicates a strong avoidance of GATC sites in the  $\phi$ X174 genome.

In order to determine the effect of the addition of the expected number of GATC sites on the mutation rate of  $\phi$ X174, we constructed a twenty-fold GATC mutant (20GATC) and performed fluctuation tests. The mutant was constructed by sequential addition of GATC sites in the WT virus by mutating a single position using site-directed mutagenesis. Most substitutions were synonymous in order to minimize their effect on protein function and the twenty GATC sites were evenly distributed throughout the phage genome (the largest distance between any two of them was 456 bases) (Figure 18). Since the Dam/MutHLS system can accomplish MMR at a distance of up to 1kb from a GATC (Modrich and Lahue, 1996), the number and distribution of the GATC sites introduced should allow for efficient MMR across the entire viral genome.

The results of the Luria-Delbrück fluctuation test showed an estimated mutation rate for the WT of  $\mu = (1.58 \pm 0.44) \times 10^{-6} \text{ s/n/r}$ , a value consistent with previous studies (Cuevas et al., 2009; Raney et al., 2004). In contrast, the rate of the 20GATC virus ( $\mu = (2.06 \pm 0.15) \times 10^{-7}$

s/n/r) showed a 7.7-fold reduction compared to the WT virus (t-test:  $P < 0.001$ ; Figure 19A; Supplementary information 4). The estimated growth rate using the  $N_1$  and  $N_0$  values obtained from the fluctuation test assays was similar for the WT ( $r = 8.11 \pm 0.29 \text{ h}^{-1}$ ) and the 20GATC viruses ( $r = 8.68 \pm 0.29 \text{ h}^{-1}$ ; t-test:  $P = 0.406$ ; Figure 19B; Supplementary information 4), indicating that these substitutions had no significant impact on the short-term viral fitness. These results confirm our previous findings obtained with a mutant phage carrying seven GATC sites, indicating that phage DNA is susceptible to MMR if the required sequence motifs are present (Cuevas et al., 2011).

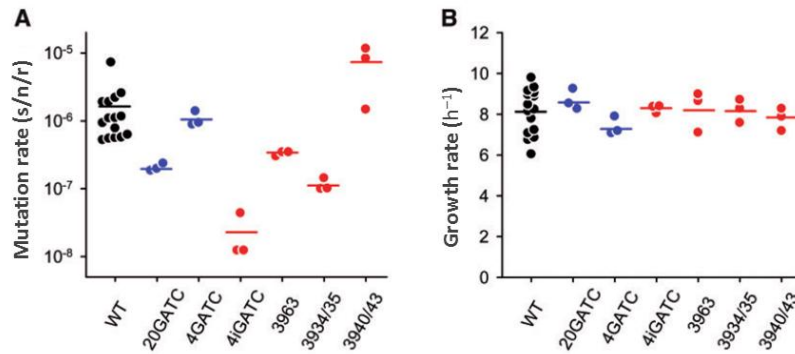


**Figure 18.  $\phi$ X174 genetic map and location of the GATC sites introduced in this study.** Open reading frames are represented by rectangles (B, K, and E genes are in different reading frames), and gray bars indicate intergenic regions. Each GATC is represented by a dot, and its position is indicated on top. GATC motifs that were synonymous in all reading frames are indicated in blue, whereas those producing amino acid replacements in at least one frame are shown in green (A74C produces a K494Q replacement in gene A and is synonymous in gene K; T5374G produces a V465G replacement in gene A and is synonymous in gene B). Mutations falling at intergenic regions are shown in red. The phage has a circular DNA genome but is represented linearly for convenience, where by convention the first position corresponds to the last nucleotide of the unique *Pst*I site. Figure originally published in Pereira-Gómez and Sanjuán 2015 and reproduced under the terms of the Creative Commons Attribution License.

## 2. The effect of the GATC sites is limited by under-methylation of the phage DNA.

Restoration of the expected number of GATC sites produced a relatively low reduction in the mutation rate if MMR were fully efficient (Fijalkowska et al., 2012). Considering that MMR relies on the presence of GATC sites in the vicinity of the mismatch (Modrich and Lahue, 1996), we thought that the addition of GATC sites in the *gro* resistance region, which is the phenotypic marker used to perform Luria-Delbrück fluctuation test, should suffice to yield similarly low mutation rates. Based on this, we constructed a mutant with four GATC sites located within 0.5 kb of known *gro*-resistance mutations (between genome positions 3671 and 4151; Figure 18). However, the mutation rate estimate of this four GATC (4GATC) mutant was significantly higher than the 20GATC virus (t-test:  $P = 0.001$ ) and similar to the WT rate ( $P = 0.832$ ) (Figure 19A; Supplementary information 4).

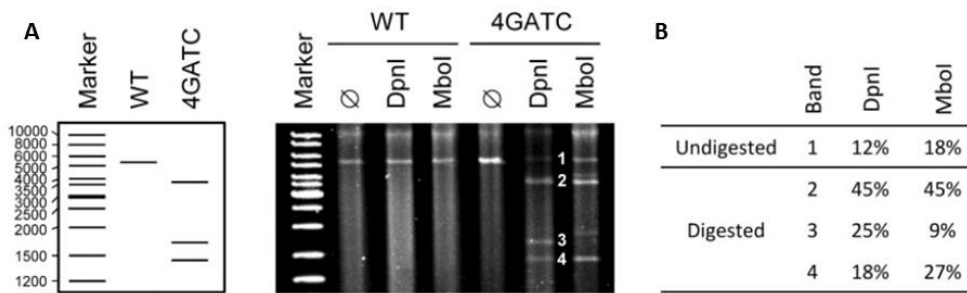




**Figure 19. Effect of different GATC sequence sites on the mutation and growth rate of  $\phi$ X174.** Mutation rates estimated by the Luria–Delbrück fluctuation test (A) and growth rates obtained in these same assays (B) for the WT virus ( $n=15$ ) and GATC mutants ( $n=3$ ) are shown. GATC mutants containing only synonymous substitutions are shown in blue, whereas those mutants containing at least one intergenic substitution are shown in red. Each dot represents an individual estimate and horizontal bars indicate the mean. Figure originally published in Pereira-Gómez and Sanjuán 2015 and reproduced here under the terms of the Creative Commons Attribution License.

Taking this result into account and considering that the adenosines of GATC sites need Dam methylation for MMR action, we sought to determine the methylation status of the replicative form DNA in the 4GATC  $\phi$ X174. To do so, we performed double digestions of the WT and 4GATC viruses using *XhoI* and *DpnI* or *XhoI* and *MboI* restriction endonucleases.  $\phi$ X174 has a unique *XhoI* site at position 162 and, therefore, the action of this endonuclease should linearize the DNA. *DpnI* digests methylated and hemi-methylated GATC sequences whereas *MboI* digests non-methylated GATC sequences. Although *DpnI* produced restriction bands of the expected size, the digestion was only partial, although more detailed analyses would be required to reliably quantify this fraction. In addition, when the same DNA was treated with *MboI*, the amount of digestion products was similar to that obtained with *DpnI*, indicating that a similar fraction of the phage dsDNA was not methylated (Figure 20A).

Image analysis indicated that 82% of the DNA was digested by *MboI* and, therefore, lacked at least one of the four possible methyl groups, whereas 12% was not digested by *DpnI*, and hence lacks all four methyl groups (Figure 20B). Overall, the similar efficiency shown by *DpnI* and *MboI* restriction endonucleases suggests that approximately half of GATC sites were methylated. The partial digestion was not due to a Dam methylation defect in the host cell, since a standard plasmid grown in the same *E. coli* strain was totally digested by *DpnI* and fully resistant to *MboI* (Cuevas et al., 2011). These results suggest that, as opposed to plasmid or chromosomal DNA, GATC-mediated MMR is not fully efficient in  $\phi$ X174 due to under-methylation of the phage DNA. Full viral DNA methylation may be impeded by the fast replication of the phage and the transient nature of the dsDNA replicative form.



**Figure 20. Restriction fragment analysis of the  $\phi$ X174 replicative form.** Phage replicative form DNA linearized with *XhoI* ( $\emptyset$ ), which recognizes a unique site at position 162, with *XhoI* and *DpnI* to cleave methylated or hemi-methylated GATCs, or with *XhoI* and *MboI* to cleave non-methylated GATCs. **A)** Expected (left) and observed (right) restriction fragments for the WT and 4GATC phage dsDNA are shown. Lower size fragments (<300 bp, Fig. 18) were expected but could not be visualized because the amount of input DNA was low. The smear in lanes containing the purified phage DNA probably results from degradation of host DNA. The contrast of the gel image was enhanced to help visualize bands. **B)** Percent abundance of each *DpnI* and *MboI* restriction band (1–4). Band 1 in the *DpnI* lane indicates the non-methylated DNA fraction (i.e., none of the four GATC sites was methylated), whereas in the *MboI* lane, this same band indicates the fully methylated fraction (i.e., the four GATC sites were methylated). Bands were quantified using the raw gel image with no contrast enhancement. Figure originally published in Pereira-Gómez and Sanjuán 2015 and reproduced under the terms of the Creative Commons Attribution License.

### 3. Intergenic GATC sites show highly variable effects on the $\phi$ X174 mutation rate.

Taking into account that 95% of the  $\phi$ X174 genome comprises coding regions, one in twenty 20 GATC sites should occur at non-coding regions by chance. To test the effect of GATC location, we constructed another mutant (4iGATC) in which the synonymous substitution A3671G of the former 4GATC virus was replaced by the substitution A3963G, which is located in the intergenic region between genes H and A (Figure 18). The mutation rate estimate was  $\mu = (2.27 \pm 1.04) \times 10^{-8}$  s/n/r, which is a 50-fold decrease compared to the WT mutation rate (t-test:  $P < 0.001$ ; Figure 19A; Supplementary information 4), indicating that the addition of this single intergenic substitution had a strong effect on the viral mutation rate, compared to the 4GATC virus. This result suggests that intergenic GATCs have stronger effects than those located in gene coding regions, potentially because reduced availability of the MMR machinery to the mismatch could reduce MMR efficiency. Hence, we decided to test solely the effect of the A3963G substitution on the  $\phi$ X174 mutation rate. The mutation rate of the single mutant A3963G was five times lower than that of the WT (t-test:  $P < 0.001$ ; Figure 19A; Supplementary information 4), although still significantly higher than the 4iGATC virus (t-test:  $P = 0.020$ ). This result suggests that the effect of the A3963G substitution was enhanced in the 4iGATC virus by the presence of other neighboring GATC sites.

Next, we constructed another two different single GATC mutants (C3934G/C3935A and A3940G/T3943C), in which mutations were also located in the region between the H and A genes (Figure 18). We found that the mutation rate of the C3934G/C3935A mutant was  $(1.15 \pm 0.14) \times 10^{-7}$  s/n/r, 14 times lower than the WT (t-test:  $P < 0.001$ ; Figure 19A; Supplementary information 4). In contrast, the A3940G/T3943C mutant mutation rate was not different from the WT estimate ( $P = 0.132$ ), despite being located only five bases away from the previous substitutions. Thus, some but not all intergenic GATCs located in the intergenic region between H and A genes are able to promote MMR.

This intergenic region contains the A promoter and the H terminator. The C3934G/C3935A substitution, which had the strongest effect on the mutation rate, was the farthest away from the A promoter and we can speculate that this site, which is located in a more distal position from actively transcribed regions, could potentially be more exposed to methylation and MMR machinery than the other GATC sites. However, the A3940G/T3943C and A3963G substitutions are located approximately at the same distance from the A promoter. A3940G/T3943C is located upstream the promoter, whereas A3963G is downstream the promoter and in a palindromic region with relatively high GC content. This difference in template could potentially produce the observed differences in mutation rates, as it previously was suggested that MMR has increased efficiency in regions with higher GC content (Jones et al., 1987).

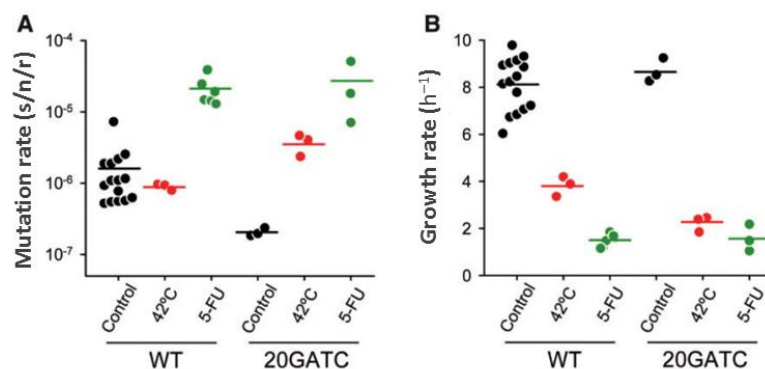
#### **4. The effect of GATC sites on the $\phi$ X174 mutation rate is reverted under stress.**

We finally assessed the effect of two different stresses, heat shock and treatment with 5-FU, on the mutation rate of the WT and 20GATC viruses. The cellular response to heat shock results in the induction of the chaperone-encoding *groE* operon, up-regulating the error-prone DNA polymerase IV, which is involved in the repair of dsDNA breaks, and thereby leading to stress-induced mutagenesis (Layton and Foster, 2005; Malkova and Haber, 2012). On the other hand, besides its mutagenic effect, treatment with 5-FU inhibits thymidylate synthase, causing deoxythymidine monophosphate deprivation, which produce DNA strand breaks, induction of the cellular SOS DNA damage response (DDR), and expression of error-prone DNA repair enzymes (Ahmad et al., 1998; Fonville et al., 2010).

We found that the mutation rate of the WT virus was not significantly affected by a temperature shift from 37°C to 42°C during the incubation time (t-test:  $P = 0.585$ ; Fig. 21A; Supplementary information 5). However, the mutation rate estimate for the 20GATC virus was 20-fold higher at 42°C than at 37°C (t-test:  $P < 0.001$ ) and increased even above the WT level ( $P = 0.003$ ). The heat shock reduced the viral growth rate compared to the standard growth

temperature (37°C) of both viruses. However, the 20GATC virus grew significantly slower than the WT, ( $r = (2.23 \pm 0.19) \text{ h}^{-1}$  and  $r = (3.81 \pm 0.24) \text{ h}^{-1}$  for the 20GATC mutant and WT viruses, respectively; t-test:  $P=0.007$ ; Figure 21B; Supplementary information 5). Finally, the results of the fluctuation tests performed in the presence of 5-FU showed that the drug had an approximately 10-fold higher mutagenic effect on the 20GATC mutant than the WT virus (Figure 21A), abolishing the differences in the mutation rates between these two viruses (t-test:  $P = 0.951$ ).

Overall, these results suggest that GATCs make the phage more sensitive to stress factors, in terms of mutation and growth rates, compared to the WT virus. In bacteria, it has been shown that increases of the mutation rate can promote adaptability (Jolivet-Gougeon et al., 2011; LeClerc et al., 1996; Sniegowski et al., 1997). However, we observed that GATC avoidance does not appear to increase the  $\phi\text{X174}$  mutation rate under all stress conditions tested here, suggesting that GATC avoidance could be selected to favor higher mutation rates in natural populations of  $\phi\text{X174}$ .



**Figure 21. Effect of stress factors on the mutation and growth rate of wild type and 20GATC viruses.** Mutation rates estimated by the Luria–Delbrück fluctuation test (A) and growth rates obtained in these same assays (B) for the WT virus (n=3-15) and the 20GATC mutant (n=3) are shown. Estimates carried out under thermal stress are shown in red, whereas those carried out in the presence of 5-FU are shown in green. Each dot represents an individual estimate and horizontal bars indicate the mean. Figure originally published in Pereira-Gómez and Sanjuán 2015 and reproduced under the terms of the Creative Commons Attribution License.

## **Chapter III: GATC avoidance and evolvability.**



Evolvability is defined as the ability of an organism to incorporate genetic changes that improve its adaptation to the environment (Wagner and Altenberg 1996) and thereby to evolve through natural selection. The results obtained in the previous chapter suggest that the evolutionary optimization of the mutation rate of  $\phi$ X174 may not be the only factor driving GATC counter-selection because, while GATCs are not present in the phage genome in nature, some GATCs had no effect on mutation rate. In addition, although we did not find, in the absence of stress, a deleterious fitness effect in the 20GATC mutant like that found by Deschavanne and Radman (1991), such effects could occur in different environments. Therefore, here we explored how the presence of GATC sites affects the long-term evolvability of  $\phi$ X174.

### **1. Preliminary analyses: Testing for block effect in the fitness assays.**

Fitness assays were performed, for each particular treatment and passage number, in the same block. Possible block or day effects were controlled by including the founder viruses in each block. Because growth rates of WT and 20GATC populations were measured in different experiments, we tested for each host whether growth rates of the founders were affected by a block effect with a one-way ANOVA test. Although there was no difference in WT growth rate between the experiments performed in *E. coli* (one-way ANOVA,  $P = 0.084$ ), there was a significant difference for the 20GATC virus (one-way ANOVA,  $P = 0.004$ ). A Tukey's post hoc test indicated that the means of the growth rates estimated from the assays at passage 25 differed from those assays performed at passage 50 and 100. In contrast, when we analyze the growth rates measured in *S. typhimurium*, we founded differences in WT growth rate between the experiments (one-way ANOVA,  $P < 0.001$ ) whereas we founded no differences for the 20GATC virus (one-way ANOVA,  $P = 0.302$ ). In this case, a Tukey's post hoc test indicated that the mean of the growth rates estimated from the assays at passage 25 differed from those assays performed at passage 50 and 100. In conclusion, we found a block effect, which was not homogeneous for each cell type and virus variant. In order to correct for the block effect, we expressed the fitness changes relative to their founder virus assayed in the same block to perform the statistical analysis of fitness changes during evolution. Some assays were conducted using different cell stocks or the same stock with different storage time. Thus, it is difficult to dissect the effect of time from potential effects of stock.

## **2. GATC sites are not lost in the long term and have no significant impact on $\phi$ X174 evolvability when the selection strength is low.**

To assess the evolutionary factors involved in  $\phi$ X174 GATC avoidance we first performed a “control” evolution experiment using the WT and 20GATC viruses in *E. coli* C. To do so, viruses were serially passaged one hundred times in five independent lines (L1–L5) under conditions that promote the action of natural selection. Viruses, from passages 25, 50, and 100, were assayed for fitness under the same conditions in which the populations were evolved (Supplementary information 6). Since this experiment imposes the selection pressure to which  $\phi$ X174 is normally exposed to in our laboratory and to which, in fact, the virus is already adapted to (Domingo-Calap et al., 2009), we did not expect an improvement in viral fitness over time in these lines. However, although we assumed a relatively low rate of molecular evolution, we also expected the loss of few GATC sites at the end of the evolution experiment. This is because GATC avoidance is selected for in natural populations of  $\phi$ X174 and other DNA enterophages (McClelland, 1985), but the loss rate may be slow because we did not detect previously a deleterious effect of the GATC sites in our experimental conditions.

We found that all WT lines evolved in *E. coli* C had no significantly higher fitness, measured as the growth rate, than their corresponding founder virus during the entire evolution experiment (two-way ANOVA,  $P = 0.994$  at passage 25;  $P = 0.286$  at passage 50, and  $P = 0.064$  at passage 100; Table 6; Figure 22). We detected differences among lines at passage 50 (two-way ANOVA,  $P = 0.049$ ). A Tukey’s post hoc test indicated that none of the WT lines was significant different from another (Tukey’s post hoc test,  $P = 0.731$ ), whereas we found two homogeneous subsets in the GATC lines. The first subset was formed by lines L1 and L5 (Tukey’s post hoc test,  $P = 0.110$ ), and the second by lines L2-L4 (Tukey’s post hoc test,  $P = 0.459$ ). The differences among lines are likely produced by line L5, which is the most different from the other lines, because it is the 20GATC line with the lowest observed fitness change ( $\Delta r = (-0.46 \pm 0.09)$  at passage 50 (Figure 22)). Moreover, we found a positive relationship between the changes in the growth rate and the log of the maximal viral titer (Pearson,  $r = 0.623$ ,  $P < 0.001$ ). Overall, our results showed that GATC sites have minor impact on the viral short-term growth rate.



**Table 6. ANOVA tables.** Fitness changes are the growth rate ( $h^{-1}$ ) for WT and GATC lines relative to that of their corresponding founder population in *E. coli* ( $\Delta r = r - r_0$ ), where  $r$  and  $r_0$  refer to the growth rate of the evolved and founder viruses, respectively. Model: Fitness increment of the evolved lines =  $\mu + \alpha(\text{Virus}) + \beta(\text{Lines}) + \epsilon$ .

**Passage 25**

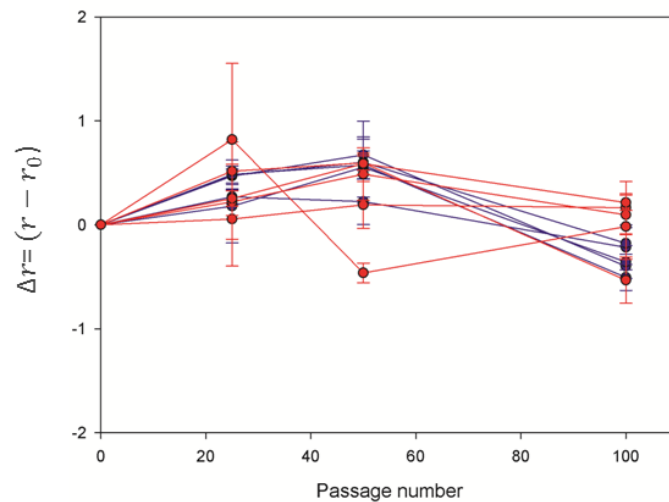
Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	$9.63 \times 10^{-6}$	$9.63 \times 10^{-6}$	$5.9 \times 10^{-6}$	0.994
<b>Lines(virus)</b>	<b>8</b>	1.315	0.164	0.492	0.848

**Passage 50**

Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	0.452	0.452	1.308	0.286
<b>Lines(virus)</b>	<b>8</b>	2.765	0.346	2.463	0.049

**Passage 100**

Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	0.756	0.756	4.633	0.064
<b>Lines(virus)</b>	<b>8</b>	1.306	0.163	1.432	0.244



**Figure 22. Changes in viral fitness of the evolved lines of WT (blue lines) and 20GATC (red lines) viruses during serial passages in *E. coli*.** The difference in growth rate ( $h^{-1}$ ) between each line and the founder is represented at different passage numbers. Each data point corresponds to the average growth rate difference obtained from three independent assays. Error bars are the standard errors of the means. Each evolved line was assayed in triplicate under the same condition in which the population was evolved together with its reference virus.

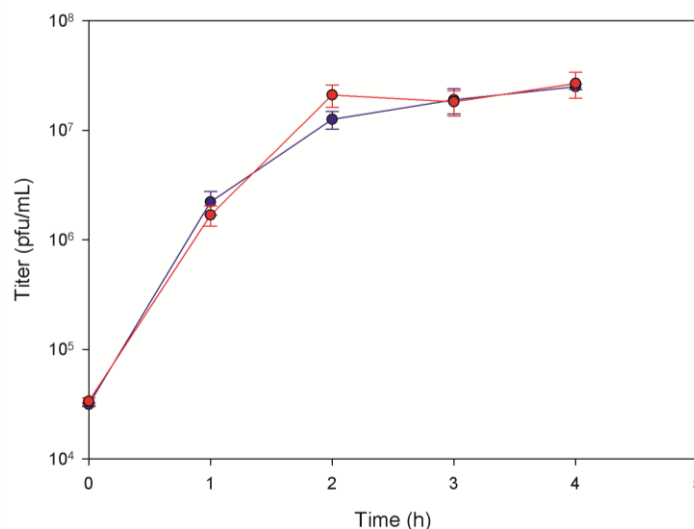
Full genome sequences of the endpoint lines showed parallel substitutions at genome positions 1295, 3042, 3336, and 3896 (Table 6). Parallel evolution is frequent in  $\phi$ X174 and other viruses and suggests a beneficial effect of these mutations (Bull et al., 1997; Cuevas et al., 2002; Rokyta et al., 2005; Wichman and Brown, 2010). The most frequent substitution is the A1295G mutation, which produces the replacement N99D in the F protein, which has been previously found in independent studies of adaptation to *E. coli* carried out in a chemostat (Wichman et al., 2005) and also in serial passage experiments (Domingo-Calap and Sanjuán, 2011; Pepin et al., 2008). The substitutions C3042T and G3336T, which map to the H gene, are also non-synonymous but their effects are less clear. The former was previously found in a line that had been adapted to grow in a chemostat at high temperature in *Salmonella* followed by growth in *E. coli* (Bull et al., 1997), and the latter was reported as a variable site among wild  $\phi$ X174-like phages (Wichman and Brown, 2010). The only synonymous substitution observed (A3896G), which maps to the N-terminus of the A gene, is only present as a polymorphism in lines L1, L4, and L5 of the 20GATC virus. This mutation produces the partial loss of a GATC site, suggesting that the selective disadvantage of GATC sites is low small in under our experimental conditions. Thus, sequence analysis showed that only one of the 20 GATC sites was partially lost and this suggested that  $\phi$ X174 indeed experienced some level of adaptation after 100 passages in *E. coli* C cells, although the fitness changes were probably too small to be observed in our growth assays.

**Table 7. Nucleotide substitutions observed after 100 passages in *E. coli* in each of the five evolved lines.** Genes, genome positions, and amino acid replacements are indicated. In each case we present substitutions compared to the corresponding founder. Nonsynonymous substitutions are shown in red and synonymous substitutions in green. Polymorphisms are indicated with the word “pol”.

		F	H					A
		A1295G (N99D)	C3042T (L38F)	G3336T (A136S)	A3378C (H150N)	C3343G (A1386)	A3896G	
WT	L1							
	L2	pol						
	L3							
	L4							
	L5							
20GATC	L1						pol	
	L2			pol				
	L3							
	L4						pol	
	L5						pol	

### 3. WT and 20GATC viruses show reduced fitness in *Salmonella* compared to *E. coli*.

Since we did not find differences in phage adaptation between the WT and 20GATC viruses in *E. coli*, but GATC sites are apparently intensely counter-selected, we set out to evolve these viruses in a host different from  $\phi$ X174, *S. typhimurium*. *S. typhimurium* is susceptible to  $\phi$ X174 infection (Jazwinski et al., 1975) and has been widely used to study  $\phi$ X174 adaptation (Brown et al., 2013; Crill et al., 2000; Pepin et al., 2008; Wichman et al., 2000). Our WT and 20GATC laboratory strains are not adapted to this host. Therefore, we expect that viral populations show an initial low fitness. In order to measure the effect of a new host on viral fitness, we first performed standard growth curves using the same initial founder populations of the WT and 20GATC viruses in *S. typhimurium*. We found that growth in *S. typhimurium*, measured as the growth rate ( $r$ ) of the WT (t-test,  $P < 0.001$ ) and 20GATC viruses ( $P < 0.001$ ) was strongly reduced compared to their viral growth in *E. coli* (two-way ANOVA,  $P < 0.001$ ). The WT ( $r = 4.12 \pm 0.38 \text{ h}^{-1}$ ) and 20GATC ( $r = 4.05 \pm 0.19 \text{ h}^{-1}$ ) viruses showed a similar growth rate (t-test,  $P = 0.877$ ) and maximal viral titer ( $K = (2.12 \pm 0.43) \times 10^7$  and  $K = (2.40 \pm 0.59) \times 10^7$  for WT and 20GATC, respectively; t-test,  $P = 0.799$ ; Figure 23), indicating that both viral populations will be subjected to a strong and comparable selection pressure during adaptation to *S. typhimurium*.



**Figure 23. Standard growth curves of WT (blue line) and 20GATC (red line) viruses in *S. typhimurium*.** Each virus was assayed in triplicate. Error bars indicate the standard errors of the means.

### 4. GATC sites negatively impact $\phi$ X174 evolvability when the selection is strong.

In order to study viral fitness increases under conditions promoting positive selection, the same founder populations of the WT and 20GATC viruses were serially passaged 100 times at high population sizes, carrying out five independent lines (L1–L5) in *S. typhimurium*. Then, viruses from passages 25, 50, and 100 were assayed for fitness under the same conditions in

which the population was evolved together with their reference virus (Supplementary information 6).

In contrast to the results of the evolution carried out in *E. coli*, the evolved lines of the WT and 20GATC viruses in *S. typhimurium* showed marked differences in growth rate relative to their founder viruses during the entire evolution experiment (two-way ANOVA,  $P < 0.001$  at passage 25;  $P = 0.006$ ; and  $P < 0.001$  at passage 100; Table 8; Figure 24). We also detected differences among lines at passage 50 (two-way ANOVA,  $P = 0.040$ ). A Tukey's *post hoc* test indicated that WT lines formed a single subset (Tukey's *post hoc* test,  $P = 0.202$ ) whereas in the GATC lines we observed two different subsets: the first one included only line L1 ( $P = 1$ ) and the second subset comprised lines L2-L5 ( $P = 0.996$ ). The magnitude of the fitness gain correlated positively with the log of the maximal viral titer (Pearson,  $r = 0.754$ ,  $P < 0.001$ ). Therefore, we conclude that the five GATC lines showed a clear disadvantage compared to WT lines in their ability to adapt to *S. typhimurium* as determined by the increase in their growth rate compared to their relative founder population.

**Table 8. ANOVA tables.** Fitness changes are the growth rate ( $h^{-1}$ ) for WT and GATC lines relative to that of their corresponding founder population in *S. typhimurium* ( $\Delta r = r - r_0$ ), where  $r$  and  $r_0$  refer to the growth rate of the evolved and founder viruses, respectively. Model: Fitness increment of the evolved lines =  $\mu + \alpha(\text{Virus}) + \beta(\text{Lines}) + \epsilon$ .

**Passage 25**

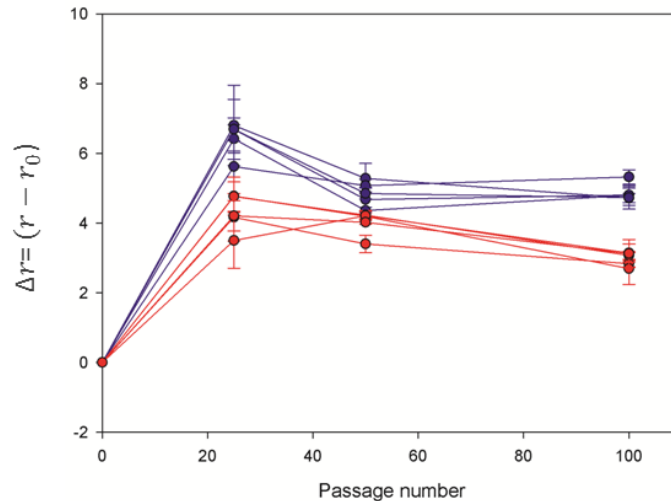
Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	35.046	35.046	46.102	$1.39 \times 10^{-4}$
<b>Lines(virus)</b>	<b>8</b>	6.081	0.760	0.590	0.774

**Passage 50**

Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	5.245	5.245	13.973	0.006
<b>Lines(virus)</b>	<b>8</b>	3.003	0.375	2.596	0.040

**Passage 100**

Effect	DF	SS	MS	F	P
<b>Virus</b>	<b>1</b>	27.037	27.037	173.579	$10^{-6}$
<b>Lines(virus)</b>	<b>8</b>	1.246	0.156	0.691	0.696



**Figure 24. Changes in viral fitness of the evolved lines of WT (blue lines) and 20GATC (red lines) viruses during serial passages in *S. typhimurium*.** The difference in growth rate ( $h^{-1}$ ) between each line and the founder is represented at different passage numbers. Each data point corresponds to the average growth rate difference obtained from three independent assays. Error bars are the standard errors of the mean. Each evolved line was assayed in triplicate under the same condition in which the population was evolved together with their reference virus.

### 5. Molecular basis of the adaptation to *S. typhimurium*.

Full genome sequences of the WT and 20GATC lines after 100 passages showed 18 different substitutions. Of the total mutations observed, 14 were non-synonymous and most substitutions occurred within the genes encoding the phage capsid proteins, such as the F and H genes, although we also observed parallel changes in the J and A genes (Table 7). The distribution of mutations among the genes was significantly different from the expected number of mutations given their lengths in nucleotides ( $\chi^2$  test,  $df = 10$ ,  $P < 0.001$ ). We found parallel changes at positions T869C, T1307A, T1307C, T1308G, A1462C, T2980C, C3183G, A3340G, and T4412C, that account for 50% (9/18) of the total mutations and suggest that these were beneficial mutations. All parallel changes except one synonymous substitution (T4412C) were found in WT lines but, in contrast, GATC lines lacked three of these changes. In particular, WT lines showed a greater variety of substitutions in protein H, which is involved in injection of phage DNA into the host cell (Sun et al., 2014).

The T869C mutation produces the substitution S8P in the J protein, which plays a role in viral attachment to the host cell and DNA packaging (Bernal et al., 2004) and results in the loss of the 868 GATC site in the 20GATC evolved lines. The fact of that the T869C substitution arose in all independent experimental lines suggests that its effect is beneficial, and that the loss of this GATC site was probably not related to a cost related to MMR. This mutation is located in a proline-rich region, which separates two clusters rich in basic residues. A previous genetic

study using different J protein mutants suggested that the function of proline residues might be the optimization of DNA-protein interactions between the viral DNA and the two basic regions (Jennings and Fane, 1997). We can speculate that the addition of a proline in this proline-rich region could improve DNA-protein interactions under the experimental conditions used in the present study.

Mutation T1307G, which produces a replacement in residue 103 of the F protein, is a variable site in natural  $\phi$ X174 populations (Rokyta et al., 2006). This site has been shown to be involved in host attachment and fitness in the same strain of *S. typhimurium* (Crill et al., 2000) and other hosts with variable epistatic effects (Pepin and Wichman, 2007). These previous findings clearly indicate that this mutation should be adaptive in our experimental conditions. We also note that this position evolved more than one amino acid replacement if we consider all evolved lines in *S. typhimurium*, having evolved in 20GATC L4, 20GATC L5 and WT L1, showing that more than one change at this position can be adaptive. The mutation T2980C, was also previously implicated in  $\phi$ X174 growth in *S. typhimurium* (Wichman et al., 2000) and was hypothesized to be involved in ejection (Brown et al., 2013). The last previously reported mutation is A3340G, which was found in all end-point populations evolved in *Salmonella*, and produces the replacement D137G in the H protein. However, its effect may not be host specific, since this mutation was previously reported during long term evolution in both *Salmonella* and *E. coli* hosts (Brown et al., 2013).

Finally, the parallel substitutions A1462C, C3183G and T4412C, which map to F, H and A genes, respectively, had not been described previously and, therefore, we cannot ascertain their likely effects. However, the high rate of missense substitutions, combined with the experimental conditions promoting natural selection and previous observations of high levels of parallel substitutions and amino acid replacements, collectively indicate that most substitutions fixed in the evolved populations were probably beneficial. Nevertheless, further fitness assays of the single mutants, and even double mutants, in the genetic background of the founder clones could determine more accurately the fitness effect of all the substitutions found in the present study.

**Table 9. Nucleotide substitutions observed after 100 passages in *S. typhimurium* in each of the five evolved lines.** Genes, genome positions and amino acid replacements are indicated. In each case we present substitutions compared to their respective founder. Nonsynonymous substitutions are shown in red and synonymous substitutions are shown in green. Polymorphisms are indicated with the word “pol”.

		D/E	J	F						H					A	A/A*				
		T797C (A132/A77P)	T869C (S8P)	A1301G (T101A)	T1307A (Y103N)	T1307C (Y103H)	A1308G (Y103C)	A1462C (Q154H)	T1471C (A157)	A2166C (H389P)	T2980C (V17A)	G3072T (G48S)	T3129C (S67P)	C3183G (Q85E)	A3340G (D137G)	A4168G (Q63R)	T4412C (P144)	T4739C (P253)	T4802C (P274)	
WT	L1																			
	L2					pol		pol												
	L3								pol											
	L4																			
	L5							pol												
20GATC	L1							pol	pol											
	L2																			
	L3							pol										pol		
	L4																			
	L5																			





## **GENERAL DISCUSSION**



Viral mutation rates range from  $10^{-8}$  to  $10^{-4}$  substitutions per nucleotide per round of copying (s/n/r) and determine the evolutionary properties of viruses (Sanjuán et al., 2010). High mutation rates promote faster evolution, higher adaptability, and emergence in new hosts (Holmes, 2009; Pepin et al., 2010). In agreement with this idea, RNA viruses encoding increased fidelity polymerases usually show reduced fitness and pathogenicity (Coffey et al., 2011; Meng and Kwang, 2014; Pfeiffer and Kirkegaard, 2003; Vignuzzi et al., 2006; Zeng et al., 2014). Although, in general, RNA viruses show the highest mutation and evolution rates, some DNA viruses, including bacteriophage  $\phi$ X174 and other ssDNA viruses, are interesting models for both mutation rate and evolutionary studies because their mutation rates are the highest among DNA viruses and close to those of RNA viruses despite using host DNA polymerases (Duffy et al., 2008). Research on viral mutation rates has mainly focused on RNA viruses, rather than DNA viruses, despite the fact that the latter also include important pathogens, such as smallpox viruses, herpesviruses, papillomaviruses, and adenoviruses, among others.

Similarly, a balance between different forces determines the evolution of mutation rates. First, natural selection against deleterious mutations should favor lower mutation rates (Kimura, 1967; Lynch, 2010). However, whereas natural selection can also promote higher mutation rates in specific circumstances, such as high mutation rates with an adaptive value and small mutational load (Sniegowski et al., 2000), studies in bacteria have shown that the maintenance of high mutation rates is only transitory due to mutational load and recombination (LeClerc et al., 1996; Pal et al., 2007; Sniegowski et al., 1997; Wielgoss et al., 2013). Second, mutation rates may be determined by a balance between negative selection and the cost of proof-reading and repair mechanisms (Kimura, 1967). In addition, the drift barrier hypothesis argues that genetic drift prevails over natural selection, limiting mutation rate evolution and therefore preventing further mutation rate reductions in small or low mutation rate populations (Lynch, 2010). Finally, the fitness landscape topology (Clune et al., 2008) or population structure (Jiang et al., 2010) can also determine the evolution of mutation rates.

In spite of these factors, viruses, and in particular RNA viruses, show relatively high mutation rates in natural populations. Based on the concept of error threshold (Eigen, 1971), the quasispecies theory argues that mutation rate should be inversely proportional to genome size, and therefore may contribute to explaining Drake's rule. It has also been argued that high mutation rates in RNA viruses are a consequence of biochemical constraints of RNA replication but the presence of proofreading mechanisms in coronaviruses (Denison et al., 2011; Menéndez-Arias, 2009; Roberts et al., 1988; Steinhauer et al., 1992) does not support this hypothesis. The cost of maintaining mechanisms for replication fidelity was reported in RNA

viruses (Furió et al., 2007, 2005); nevertheless, this cost should be similar in RNA and DNA viruses with similar genome size and, thus, does not explain the differences in mutation rates between RNA and DNA viruses. Overall, the evolutionary mechanisms that determine mutation rate in viruses are multiple and diverse, but it is still not well understood how exactly evolutionary mechanisms drive viral mutation rate evolution in natural populations. In the case of ssDNA viruses, there is an additional factor, the use of host polymerases, which necessarily constrain the ability of these viruses to modify their own mutation rate. The experiments performed in this PhD thesis provided experimental evidence that help us to better understand some of the genetic and evolutionary determinants of the bacteriophage  $\phi$ X174 mutation rate.

In chapter I, we investigated the role of several amino acid replacements in the N-terminal domain of the lysis protein E, since it was previously shown that serial passaging in the presence of the nucleoside analogue 5-FU selected for substitutions in this protein (Domingo-Calap et al., 2012). We have identified two single amino-acid replacements (V2A and D8A) in the lysis protein E of bacteriophage  $\phi$ X174 that modify lysis time and burst size, confer partial resistance to 5-FU, and tend to alleviate the mutagenic effect of the drug. The anti-mutagenic effect was conferred by a delay in lysis time and a concomitant increase in viral burst size, which allowed the virus to produce progeny in fewer infection cycles and thus to effectively reduce the number of replication cycles during which mutations can accumulate.

Changes in burst size and lysis time explain both changes in growth rate and the rate at which mutations accumulate in the viral populations. We propose that the lysis protein may act as a regulator of the mutation rate by inducing changes in lysis time and burst size. Bacteriophage  $\phi$ X174 essentially undergoes a single replication cycle per cell, during which newly synthesized DNA molecules are produced through a rolling circle mechanism from a single template (Dressler, 1970; Eisenberg et al., 1977) regardless of burst size. Hence, greater burst size should allow the virus to reduce the number of viral generations required to infect a population of susceptible cells and, consequently, the overall number of rounds of copying. Since mutation rates were estimated using the null-class method in fluctuation tests, the measured rates should not be sensitive to the number of generations elapsed (Rosche and Foster, 2000; Zheng, 1999). However, we detected differences in the estimates of the mutation rates. This may be because  $P_0$  is sensitive to the actual number of mutants observed in each culture and to the number of generations elapsed if the plaque assay is not fully efficient or, as in this work, only a fraction of the culture volume is used (Bradwell et al., 2013). In this case, if a virus variant has a higher mutation rate, the probability of observing mutants

is higher. The differences in mutation rate between WT and V2A and D8A viruses were likely underestimated because 5-FU reduced plating efficiency during mutant scoring by approximately 2-fold for the WT but had no such effect on the mutant viruses.

The anti-mutagenic effect of the V2A and D8A replacements is comparable to some nucleoside analog resistance mutations previously reported in RNA viruses. For example, serial passaging of poliovirus treated with ribavirin was shown to select for amino acid replacement G64S in the viral polymerase, which increased overall replication fidelity by threefold (Pfeiffer and Kirkegaard, 2003). Similarly, experiments using chikungunya virus treated with the same 5-FU dose used here evolved the C483Y replacement near the conserved active site of the nsP4 replicase, which was shown by molecular sequencing to reduce viral genetic variability in cell culture by slightly less than twofold (Coffey et al., 2011).

The V2A replacement was originally reported in lines serially passaged in the presence of 10 ng/ $\mu$ L 5-FU and, although we cannot discard that the virus was adapted to the physiological changes produced by 5-FU, it is also likely that this mutation was selected by its anti-mutagenic effect. In fact, it was found that the V2A replacement lost its ability to confer 5-FU resistance after cultures were supplemented with deoxythymidine, which should partially reverse the mutagenic effect of 5-FU, among other effects, supporting that its mechanisms of action is produced at the DNA level, including mutagenesis (Domingo-Calap et al., 2012). In agreement with this, we showed here that the increased viral fitness in the presence of 5-FU is explained by changes in lysis time and burst size produced by the mutations V2A and D8A and that these substitutions should also allow the virus to alleviate the mutational load produced by 5-FU over the long term.

In chapter II, we studied the effect of mismatch repair on the mutation rate of bacteriophage  $\phi$ X174, which shows one of the highest mutation rates among DNA-based genomes, with approximately  $10^{-6}$  m/n/r (Cuevas et al., 2009; Raney et al., 2004). This high mutation rate is a consequence, at least in part, of the absence of post replication fidelity mechanisms in ssDNA phages, such as methyl-directed mismatch repair (MMR) performed by the Dam/MutHLS system (Jiricny, 2013), which requires the presence of GATC sites in the genome. We have previously shown that phage  $\phi$ X174 was able to undergo MMR after the addition of seven GATC sites (Cuevas et al., 2011). However, the introduction of the of 20 GATC sites, the expected number by chance given the base frequencies and genome size of  $\phi$ X174, reduced the phage mutation rate by approximately 8-fold. The effect of GATC addition to the  $\phi$ X174 genome conferred to the virus an anti-mutator phenotype that is greater than those previously

reported in experiments performed in presence of nucleoside analogues with RNA viruses (Coffey et al., 2011; Pfeiffer and Kirkegaard, 2003) and  $\phi$ X174 (Domingo-Calap et al., 2012; Pereira-Gómez and Sanjuán, 2014).

Anti-mutator polymerase variants that strongly suppressed the action of nucleoside analogues were also discovered in bacteriophage T4 (Drake et al., 1969; Drake and Greening, 1970). However, these polymerase variants showed a fitness cost due to slower DNA replication than the wild type virus (Mansky and Cunningham, 2000). Finally, anti-mutators were also reported in *E. coli*, with both moderate (Fijalkowska et al., 1993) and strong effects (Quiñones and Piechocki, 1985). Thus, the extent of the reduction of mutation rate afforded by the introduction of GATC motifs is similar or higher than those reported previously without significantly impacting viral fitness in the short-term. However, due to incomplete methylation of the phage genome, this reduction in mutation rate was lower than expected. Both Dam over (McClelland, 1985; Pukkila et al., 1983) and under producer (Bale et al., 1979; Marinus, 2010) mutants of *E. coli* show a mutator phenotype suggesting that cellular Dam methylase levels need to be tightly regulated. Furthermore, the GATC sites showed greatly variable effects on the phage mutation rate, with the highest reduction efficiencies showed by some of the intergenic GATCs, suggesting that steric constraints are important for MMR. Finally, we also found that the reduction in mutation rate provided by the GATC sites was completely reverted in the presence of stress factors, suggesting that addition of GATC sites makes the phage sensitive to stress-induced mutagenesis. Nevertheless, as we have shown that GATC avoidance in  $\phi$ X174 does not appear to increase its mutation rate under stress conditions, the potential evolutionary advantage of such avoidance might not to be the main factor involved in evolution of the mutation rate of  $\phi$ X174.

Previous work has shown that the presence of GATC is deleterious in *dam*<sup>-</sup> *mutH*<sup>+</sup> bacteria, suggesting that when the DNA is unmethylated, *MutH* endonuclease cleaves DNA and interferes with viral replication or encapsidation, therefore promoting GATC avoidance in  $\phi$ X174 (Deschavanne and Radman, 1991). GATC counter selection may also be explained by incomplete DNA methylation produced by the relatively low levels of Dam methylase in the host cells, combined with fast viral replication. Finally, the transitory dsDNA intermediate form of  $\phi$ X174 and similar phages cannot only promote GATC counter selection, but also other restriction sites or avoidance of certain short palindromes, which is not observed in RNA viruses infecting *E. coli*, such as MS2 and Q $\beta$  (McClelland, 1985; Sharp, 1986).

In chapter III, we used the 20 GATC mutant of  $\phi$ X174 to experimentally investigate the evolutionary basis of GATC avoidance. Since we found some GATCs with no effect on the viral mutation rate but which are also absent from the WT phage, our results suggested that evolutionary optimization of the  $\phi$ X174 mutation rate may not be the only factor driving GATC counter selection. We evolved WT and GATC viral populations by serial passages in two different hosts to create conditions which differed in the selection pressure acting on viral populations. We found that viral populations showed no fitness changes in *E. coli*. In contrast, we found that GATC sites delayed adaptation in *S. typhimurium*, indicating that GATC avoidance imposes a strong cost in terms of adaptability to novel hosts in  $\phi$ X174.

Full genome sequence analysis revealed three times more substitutions in lines evolved in *S. typhimurium* than in those evolved in *E. coli*, consistent with the notion that the former constitutes a novel environment offering greater adaptation opportunity. However, we did not find differences between the molecular evolution rates of WT and GATC lines, as these two variants accumulated similar numbers of substitutions after 100 serial passages. Some of the parallel substitutions found in *S. typhimurium* have also been described in previous works (Crill et al., 2000; Rokytka et al., 2006; Wichman et al., 2000), strongly suggesting that changes in this small region are adaptive under our experimental conditions. The H protein has been implicated previously in adaptation to *S. typhimurium* (Wichman et al., 2000). If we discard substitution D137G, which seems to be unrelated to adaptation to *Salmonella*, GATC lines show no changes in protein H, whereas WT lines showed four substitutions in this protein, three of which occurred multiple times. We suggest that the most likely basis of the reduced adaptability of GATC lines lies in the lack of changes in this gene. However, how GATC sites restrict the molecular evolution of this gene remains unclear. In chapter II, we showed that the GATC sites have heterogeneous effects on mutation rate along the viral genome. It is possible that the effect of the engineered GATC motifs on mutation rate were stronger in gene H than in other genome regions.

Finally, we found that almost all GATC sites were maintained after 100 serial passages in both *E. coli* and *S. typhimurium*, indicating weak or no selective disadvantage. This extremely low rate of loss was unexpected. GATC sites are generally rare in  $\phi$ X174 and other enterophages that infect *dam*<sup>+</sup> methylase bacteria (McClelland, 1984). In chapter II, we found that the GATC sites engineered in the  $\phi$ X174 genome show poor Dam methylation, and it has also been shown that GATC sites are deleterious under some conditions (Deschavanne and Radman, 1991). Therefore, GATC sites may be directly detrimental for the virus in some host strains, thereby selecting for GATC avoidance in  $\phi$ X174 and other phages. Overall, these results are

consistent with the idea that GATC avoidance in  $\phi$ X174 may result from the joint action of different selective forces, including mutation rate optimization, deleterious effects in specific environments and promotion of adaptability. Regardless these pressures, our results show that GATC sites impose a cost in terms of adaptability to novel hosts, potentially contributing to the maintenance of high mutation rates in bacteriophage  $\phi$ X174.

The results obtained from this PhD offer several possible lines for future research:

In chapter I we have shown a novel mechanism by which proteins not directly involved in replication can mediate the regulation of mutation rate. Other viruses with a replication mode similar to that of bacteriophage  $\phi$ X174 may also adjust their mutation rates by modifying lysis time and burst size in harsh and mutagenic environments. However, more experiments with other ssDNA phages are necessary to elucidate the generality of these findings.

Despite the fact that in chapter II the relationship between post-replicative repair mechanisms and mutation rate in  $\phi$ X174 was not as expected if MMR was as efficient as it is in the host, we detected an important effect on the phage mutation rate. Future work may elucidate which factors limited MMR in  $\phi$ X174 despite the phage having as many GATC sites as the host. For example, methylation levels of GATC mutants could be enhanced by constructing mutants of this *E. coli* C strain which overexpresses Dam methylase. This will help to further demonstrate the hypothesis of incomplete viral genome methylation and its effect on the 20GATC mutation rate.

MMR is highly conserved throughout evolution. Multiple homologues of *E. coli* MutS and MutL are present in eukaryotes (Marti et al., 2002). In these organisms, MMR is coupled to DNA replication and is involved in DNA repair, recombination and cell cycle arrest and/or apoptosis in response to some types of DNA damage (Li, 2008; Stojic et al., 2004). For instance, small eukaryotic viruses that do not encode their own polymerases, such as parvovirus (Luo and Qiu, 2013), polyomaviruses (Dahl et al., 2005), and papillomaviruses (Gillespie et al., 2012) are known to activate the host DNA damage response for viral replication. In addition, it has been demonstrated that Epstein Barr virus activates the host DNA damage response (Kudoh et al., 2005) and also recruits some MMR proteins during lytic infection (Daikoku et al., 2006). However, the role of the MMR proteins on the viral mutation rate has not been determined, and MMR could also have effects in other genetic processes such as viral recombination. Given the observed effects of repair avoidance in the mutation rate of  $\phi$ X174, we cannot discard that other DNA viruses may avoid DNA repair in similar ways in eukaryotic hosts. Therefore, further work is needed to elucidate the role of DNA repair in eukaryotic DNA viruses. Future studies on



changes in the expression and/or localization of repair-associated proteins will clarify the impact of repair pathways on the mutation rates of eukaryotic DNA viruses.

Whilst the results from chapter III advance farther our knowledge about the long-term evolution and adaptability in bacteriophage  $\phi$ X174, future work using different evolution conditions, such as higher temperatures, the presence of mutagenic compounds, alternating host, or even new hosts, will assess the generality of our findings in this study. This will help to further demonstrate the hypothesis that GATC avoidance applies a strong cost in terms of adaptability in  $\phi$ X174. Moreover, it is necessary to determine more accurately the fitness effects of the parallel substitutions found in the present study by introducing them individually by site-directed mutagenesis in the genetic background of the founder clones. Finally, it would also be interesting to quantify the effect of GATC sites located in the H gene in order to test whether the mutation rate is constant along the genome. This last issue can be addressed using the classical fluctuation test experiment or next-generation sequencing technologies (NGS). The former has the disadvantage of scoring phenotypes associated to only a few nucleotide sites, whereas the latter has been limited by the intrinsic high per-read error rates to study DNA virus mutation rates. However, recently developed, high-fidelity NGS methods, such as Duplex sequencing (Kennedy et al., 2014; Schmitt et al., 2012) and Circular sequencing (Acevedo et al., 2014), solve this problem. Using sequencing, mutation rates can be inferred by starting an infection with a homogeneous viral population and quantifying the genetic diversity produced in one or a few infection cycles. The mutation rate can be estimated by dividing the observed number of mutations by the number of infection cycles produced. The main disadvantage of this approach is that the number of mutations observed is dependent not only on the mutation rate but also on selection. However, selection bias can be corrected if the distribution of mutational fitness effects is determined previously (Sanjuán et al., 2010).



## **FINAL CONCLUSIONS**



- Single amino acid replacements in the N terminus of the lysis protein frequently produced delayed lysis, and two of these substitutions (V2A and D8A) also conferred 5-FU resistance.
- Changes in lysis time and burst size produced by the substitutions V2A and D8A explained 5-FU resistance.
- The substitutions V2A and D8A retarded the intracellular accumulation of mature viral particles.
- In the presence of 5-FU, the substitutions V2A and D8A significantly increased the viral burst size and reduced the estimated phage mutation rate by approximately 2-fold.
- The introduction of the expected number of GATC sites reduced the phage mutation rate by approximately 8-fold. However, their effect was limited by under-methylation of the phage DNA.
- GATC sites located in intergenic regions showed highly variable effects on the mutation rate of the phage.
- The effect of GATC sites on the mutation rate of the phage was reversed at 42°C and in the presence of 5-FU.
- GATC sites had no significant impact on phage adaptability in *E. coli*. Although phage populations were already adapted to this host, we detected parallel sequence changes, suggestive of adaptation.
- *Salmonella* strongly reduced the fitness of WT and 20GATC phages compared to *E. coli*.
- GATC sites retarded phage adaptation to the novel host *S. typhimurium*. However, full-genome sequence analysis did not reveal differences between the rates of molecular evolution of WT and GATC lines.



## REFERENCES





- Acevedo, A., Brodsky, L., Andino, R., 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686–90.
- Adelberg, E.A., Coughlin, C.A., 1956. Bacterial mutation induced by thymine starvation. *Nature* 178, 531–2.
- Agudo, R., Ferrer-Orta, C., Arias, A., de la Higuera, I., Perales, C., Pérez-Luque, R., Verdaguier, N., Domingo, E., 2010. A multi-step process of viral adaptation to a mutagenic nucleoside analogue by modulation of transition types leads to extinction-escape. *PLoS Pathog.* 6, e1001072.
- Ahmad, S.I., Kirk, S.H., Eisenstark, a, 1998. Thymine metabolism and thymineless death in prokaryotes and eukaryotes. *Annu. Rev. Microbiol.* 52, 591–625.
- Ali, A., Li, H., Schneider, W.L., Sherman, D.J., Gray, S., Smith, D., Roossinck, M.J., 2006. Analysis of genetic bottlenecks during horizontal transmission of Cucumber mosaic virus. *J. Virol.* 80, 8345–50.
- Arias, A., Arnold, J.J., Sierra, M., Smidansky, E.D., Domingo, E., Cameron, C.E., 2008. Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J. Virol.* 82, 12346–55.
- Bahir, I., Fromer, M., Prat, Y., Linial, M., 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* 5, 311.
- Bale, A., D'Alarcao, M., Marinus, M.G., 1979. Characterization of DNA adenine methylation mutants of *Escherichia coli* K12. *Mutat. Res.* 59, 157–65.
- Barrick, J.E., Lenski, R.E., 2013. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* 14, 827–39.
- Bebenek, K., Roberts, J.D., Kunkel, T.A., 1992. The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *J. Biol. Chem.* 267, 3589–96.
- Berkhout, B., de Ronde, A., 2004. APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations. *AIDS* 18, 1861–3.
- Bernal, R.A., Hafenstein, S., Esmeralda, R., Fane, B.A., Rossmann, M.G., 2004. The phiX174 protein J mediates DNA packaging and viral attachment to host cells. *J. Mol. Biol.* 337, 1109–22.
- Bradwell, K., Combe, M., Domingo-Calap, P., Sanjuán, R., 2013. Correlation between Mutation Rate and Genome Size in Riboviruses: Mutation Rate of Bacteriophage Q $\beta$ . *Genetics* 195, 243–51.
- Brown, C.J., Millstein, J., Williams, C.J., Wichman, H.A., 2013. Selection affects genes involved in replication during long-term evolution in experimental populations of the bacteriophage  $\phi$ X174. *PLoS One* 8, e60401.
- Buckling, A., Craig Maclean, R., Brockhurst, M. a, Colegrave, N., 2009. The Beagle in a bottle. *Nature* 457, 824–9.
- Bull, J.J., 2006. Optimality models of phage life history and parallels in disease evolution. *J. Theor. Biol.* 241, 928–38.
- Bull, J.J., Badgett, M.R., Wichman, H.A., Huelsenbeck, J.P., Hillis, D.M., Gulati, A., Ho, C., Molineux, I.J., 1997. Exceptional convergent evolution in a virus. *Genetics* 147, 1497–507.
- Bull, J.J., Pfennig, D.W., Wang, I.-N., 2004. Genetic details, optimization and phage life histories. *Trends Ecol. Evol.* 19, 76–82.
- Burch, C.L., Chao, L., 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6.

Genetics 151, 921–7.

- Carrasco, P., de la Iglesia, F., Elena, S.F., 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. *J. Virol.* 81, 12979–84.
- Carrillo-Tripp, M., Shepherd, C.M., Borelli, I.A., Venkataraman, S., Lander, G., Natarajan, P., Johnson, J.E., Brooks, C.L., Reddy, V.S., 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* 37, D436–42.
- Chantranupong, L., Heineman, R.H., 2012. A common, non-optimal phenotypic endpoint in experimental adaptations of bacteriophage lysis time. *BMC Evol. Biol.* 12, 37–50.
- Chiu, Y.-L., Greene, W.C., 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* 26, 317–53.
- Coffey, L.L., Beeharry, Y., Bordería, A. V, Blanc, H., Vignuzzi, M., 2011. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16038–43.
- Colasanti, J., Denhardt, D.T., 1987. Mechanism of replication of bacteriophage phi X174. XXII. Site-specific mutagenesis of the A\* gene reveals that A\* protein is not essential for phi X174 DNA replication. *J. Mol. Biol.* 197, 47–54.
- Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., Mueller, S., 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320, 1784–7.
- Combe, M., Sanjuán, R., 2014. Variation in RNA virus mutation rates across host cells. *PLoS Pathog.* 10, e1003855.
- Crill, W.D., Wichman, H.A., Bull, J.J., 2000. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154, 27–37.
- Cuevas, J.M., Domingo-Calap, P., Sanjuán, R., 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol.* 29, 17–20.
- Cuevas, J.M., Duffy, S., Sanjuán, R., 2009. Point mutation rate of bacteriophage PhiX174. *Genetics* 183, 747–9.
- Cuevas, J.M., Elena, S.F., Moya, A., 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162, 533–42.
- Cuevas, J.M., Pereira-Gómez, M., Sanjuán, R., 2011. Mutation rate of bacteriophage ΦX174 modified through changes in GATC sequence context. *Infect. Genet. Evol.* 11, 1820–2.
- Dahl, J., You, J., Benjamin, T.L., 2005. Induction and utilization of an ATM signaling pathway by polyomavirus. *J. Virol.* 79, 13007–17.
- Daikoku, T., Kudoh, A., Sugaya, Y., Iwahori, S., Shirata, N., Isomura, H., Tsurumi, T., 2006. Postreplicative mismatch repair factors are recruited to Epstein-Barr virus replication compartments. *J. Biol. Chem.* 281, 11422–30.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, Or the preservation of the Favoured Races in the Struggle for Life.* London: John Murray.
- Dawson, K.J., 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor. Popul. Biol.* 55, 1–22.
- Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., Baric, R.S., 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* 8, 270–9.

- Deschavanne, P., Radman, M., 1991. Counterselection of GATC sequences in enterobacteriophages by the components of the methyl-directed mismatch repair system. *J. Mol. Evol.* 33, 125–32.
- Domingo-Calap, P., Cuevas, J.M., Sanjuán, R., 2009. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5, e1000742.
- Domingo-Calap, P., Pereira-Gómez, M., Sanjuán, R., 2012. Nucleoside analogue mutagenesis of a single-stranded DNA virus: evolution and resistance. *J. Virol.* 86, 9640–6.
- Domingo-Calap, P., Sanjuán, R., 2011. Experimental evolution of RNA versus DNA viruses. *Evolution* 65, 2987–94.
- Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7160–4.
- Drake, J.W., Allen, E.F., Forsberg, S.A., Preparata, R.M., Greening, E.O., 1969. Genetic control of mutation rates in bacteriophage T4. *Nature* 221, 1128–32.
- Drake, J.W., Greening, E.O., 1970. Suppression of chemical mutagenesis in bacteriophage T4 by genetically modified DNA polymerases. *Proc. Natl. Acad. Sci. U. S. A.* 66, 823–9.
- Drake, J.W., Hwang, C.B.C., 2005. On the mutation rate of herpes simplex virus type 1. *Genetics* 170, 969–70.
- Dressler, D., 1970. The rolling circle for phiX DNA replication. II. Synthesis of single-stranded circles. *Proc. Natl. Acad. Sci. U. S. A.* 67, 1934–42.
- Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J. Virol.* 82, 957–65.
- Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–76.
- Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523.
- Eisenberg, S., Griffith, J., Kornberg, A., 1977. phiX174 cistron A protein is a multifunctional enzyme in DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3198–202.
- Eisenberg, S., Scott, J.F., Kornberg, A., 1976. An enzyme system for replication of duplex circular DNA: the replicative form of phage phi X174. *Proc. Natl. Acad. Sci. U. S. A.* 73, 1594–7.
- Elena, S.F., Lenski, R.E., 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* 4, 457–69.
- Elena, S.F., Sanjuán, R., 2007. Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* 38, 27–52.
- Fersht, A.R., 1979. Fidelity of replication of phage phi X174 DNA by DNA polymerase III holoenzyme: spontaneous mutation by misincorporation. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4946–50.
- Fersht, A.R., Knill-Jones, J.W., 1981. DNA polymerase accuracy and spontaneous mutation rates: frequencies of purine.purine, purine.pyrimidine, and pyrimidine.pyrimidine mismatches during DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* 78, 4251–5.
- Fijalkowska, I.J., Dunn, R.L., Schaaper, R.M., 1993. Mutants of *Escherichia coli* with increased fidelity of DNA replication. *Genetics* 134, 1023–30.
- Fijalkowska, I.J., Schaaper, R.M., Jonczyk, P., 2012. DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol. Rev.* 36, 1105–21.

- Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., Rambaut, A., 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* 27, 2038–51.
- Fischer, W., Ganusov, V. V., Giorgi, E.E., Hraber, P.T., Keele, B.F., Leitner, T., Han, C.S., Gleasner, C.D., Green, L., Lo, C.-C., Nag, A., Wallstrom, T.C., Wang, S., McMichael, A.J., Haynes, B.F., Hahn, B.H., Perelson, A.S., Borrow, P., Shaw, G.M., Bhattacharya, T., Korber, B.T., 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* 5, e12303.
- Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*. The Clarendon Press.
- Flint, S. J., Enquist, L. W., Racaniello, V.R. & Skalka, A.M., 2004. *Principles of Virology: Molecular Biology, Pathogenesis and Control of Animal Viruses*. ASM PRESS, Washington DC.
- Fonville, N.C., Bates, D., Hastings, P.J., Hanawalt, P.C., Rosenberg, S.M., 2010. Role of RecA and the SOS response in thymineless death in *Escherichia coli*. *PLoS Genet.* 6, e1000865.
- Forrester, N.L., Guerbois, M., Seymour, R.L., Spratt, H., Weaver, S.C., 2012. Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLoS Pathog.* 8, e1002897.
- Fukui, K., 2010. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* 2010 pii: 260512.
- Furió, V., Moya, A., Sanjuán, R., 2007. The cost of replication fidelity in human immunodeficiency virus type 1. *Proc. Biol. Sci.* 274, 225–30.
- Furió, V., Moya, A., Sanjuán, R., 2005. The cost of replication fidelity in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10233–7.
- Gerrish, P.J., Lenski, R.E., 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102–103, 127–44.
- Gilchrist, C.A., Denhardt, D.T., 1987. *Escherichia coli* rep gene: sequence of the gene, the encoded helicase, and its homology with *uvrD*. *Nucleic Acids Res.* 15, 465–75.
- Gillam, S., Astell, C.R., Jahnke, P., Iii, C.A.H., Smith, M., Hutchison, C.A., 1984. Construction and properties of a ribosome-binding site mutation in gene E of phi Construction and Properties of a Ribosome-Binding Site Mutation in Gene E of phiX174 Bacteriophage. *J. Virol.* 52, 892–6.
- Gillespie, K.A., Mehta, K.P., Laimins, L.A., Moody, C.A., 2012. Human papillomaviruses recruit cellular DNA repair and homologous recombination factors to viral replication centers. *J. Virol.* 86, 9520–6.
- Graci, J.D., Gnädig, N.F., Galarraga, J.E., Castro, C., Vignuzzi, M., Cameron, C.E., 2012. Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. *J. Virol.* 86, 2869–73.
- Haldane, J.B.S., 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Proc. Camb. Philos. Soc.* 23, 838–44.
- Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., Malim, M.H., 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* 113, 803–9.
- Heineman, R.H., Bull, J.J., 2007. Testing optimality with experimental evolution: lysis time in a bacteriophage. *Evolution* 61, 1695–709.
- Holmes, E.C., 2009. *The Evolution and Emergence of RNA Viruses.*, 1 edition. ed. Oxford University Press, Oxford, United Kingdom.
- Hone, D., Morona, R., Attridge, S., Hackett, J., 1987. Construction of defined *galE* mutants of *Salmonella* for use as vaccines. *J. Infect. Dis.* 156, 167–74.

- Hughes, A.L., Hughes, M.A.K., 2007. More effective purifying selection on RNA viruses than in DNA viruses. *Gene* 404, 117–25.
- Incardona, N.L., Selvidge, L., 1973. Mechanism of adsorption and eclipse of bacteriophage phi X174. II. Attachment and eclipse with isolated *Escherichia coli* cell wall lipopolysaccharide. *J. Virol.* 11, 775–82.
- Jazwinski, S.M., Lindberg, A.A., Kornberg, A., 1975. The lipopolysaccharide receptor for bacteriophage phiX174 and S13. *Virology* 66, 268–82.
- Jenkins, G.M., Pagel, M., Gould, E.A., de A Zanotto, P.M., Holmes, E.C., 2001. Evolution of base composition and codon usage bias in the genus *Flavivirus*. *J. Mol. Evol.* 52, 383–90.
- Jennings, B., Fane, B.A., 1997. Genetic Analysis of the  $\phi$ X174 DNA Binding Protein. *Virology* 227, 370–7.
- Jiricny, J., 2013. Postreplicative mismatch repair. *Cold Spring Harb. Perspect. Biol.* 5, a012633.
- Johnson, T., Barton, N.H., 2002. The effect of deleterious alleles on adaptation in asexual populations. *Genetics* 162, 395–411.
- Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S., Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., Lobel, B., Guillé, F., Soussy, C.-J., Tenke, P., 2011. Bacterial hypermutation: clinical implications. *J. Med. Microbiol.* 60, 563–73.
- Jones, M., Wagner, R., Radman, M., 1987. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* 115, 605–10.
- Julias, J.G., Pathak, V.K., 1998. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J. Virol.* 72, 7941–9.
- Kennedy, S.R., Schmitt, M.W., Fox, E.J., Kohn, B.F., Salk, J.J., Ahn, E.H., Prindle, M.J., Kuong, K.J., Shen, J.-C., Risques, R.-A., Loeb, L.A., 2014. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* 9, 2586–2606.
- Kim, E.-Y., Bhattacharya, T., Kunstman, K., Swantek, P., Koning, F.A., Malim, M.H., Wolinsky, S.M., 2010. Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J. Virol.* 84, 10402–5.
- Kimura, M., 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. U. S. A.* 88, 5969–73.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Kimura, M., 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267, 275–6.
- Kimura, M., 1968a. Evolutionary rate at the molecular level. *Nature* 217, 624–6.
- Kimura, M., 1968b. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* 11, 247–69.
- Kimura, M., 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* 9, 23–34.
- Kondrashov, A.S., 2009. Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genet. Res.* 66, 53.
- Krakauer, D.C., Plotkin, J.B., 2002. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1405–9.
- Kudoh, A., Fujita, M., Zhang, L., Shirata, N., Daikoku, T., Sugaya, Y., Isomura, H., Nishiyama, Y., Tsurumi,

- T., 2005. Epstein-Barr virus lytic replication elicits ATM checkpoint signal transduction while providing an S-phase-like cellular environment. *J. Biol. Chem.* 280, 8156–63.
- Kunkel, T.A., 2004. DNA replication fidelity. *J. Biol. Chem.* 279, 16895–8.
- Kunkel, T.A., 1985. The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. *J. Biol. Chem.* 260, 5787–96.
- Land, A.M., Ball, T.B., Luo, M., Pilon, R., Sandstrom, P., Embree, J.E., Wachihhi, C., Kimani, J., Plummer, F.A., 2008. Human Immunodeficiency Virus (HIV) Type 1 Proviral Hypermutation Correlates with CD4 Count in HIV-Infected Women from Kenya. *J. Virol.* 82, 8172–82.
- Layton, J.C., Foster, P.L., 2005. Error-prone DNA polymerase IV is regulated by the heat shock chaperone GroE in *Escherichia coli*. *J. Bacteriol.* 187, 449–57.
- LeClerc, J.E., Li, B., Payne, W.L., Cebula, T.A., 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274, 1208–11.
- Li, G.-M., 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85–98.
- Li, H., Roossinck, M.J., 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J. Virol.* 78, 10582–7.
- Luo, Y., Qiu, J., 2013. Parvovirus infection-induced DNA damage response. *Future Virol.* 8, 245–57.
- LURIA, S.E., 1951. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. *Cold Spring Harb. Symp. Quant. Biol.* 16, 463–70.
- Luria, S.E., Delbrück, M., 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 28, 491–511.
- Lynch, M., 2011. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* 3, 1107–18.
- Lynch, M., 2010. Evolution of the mutation rate. *Trends Genet.* 26, 345–52.
- Lynch, M., 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180, 933–43.
- Lynch, M., Bürger, R., Butcher, D., Gabriel, W., 1993. The mutational meltdown in asexual populations. *J. Hered.* 84, 339–44.
- Malkova, A., Haber, J.E., 2012. Mutations arising during repair of chromosome breaks. *Annu. Rev. Genet.* 46, 455–73.
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., Trono, D., 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424, 99–103.
- Mansky, L.M., Cunningham, K.S., 2000. Virus mutators and antimutators: roles in evolution, pathogenesis and emergence. *Trends Genet.* 16, 512–7.
- Marinus, M.G., 2010. DNA methylation and mutator genes in *Escherichia coli* K-12. *Mutat. Res.* 705, 71–6.
- Marti, T.M., Kunz, C., Fleck, O., 2002. DNA mismatch repair and mutation avoidance pathways. *J. Cell. Physiol.* 191, 28–41.
- Mayrose, I., Stern, A., Burdelova, E.O., Sabo, Y., Laham-Karam, N., Zamostiano, R., Bacharach, E., Pupko, T., 2013. Synonymous site conservation in the HIV-1 genome. *BMC Evol. Biol.* 13, 164.
- McClelland, M., 1985. Selection against dam methylation sites in the genomes of DNA of enterobacteriophages. *J. Mol. Evol.* 21, 317–22.

- Menéndez-Arias, L., 2009. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* 1, 1137–65.
- Meng, T., Kwang, J., 2014. Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. *J. Virol.* 88, 5803–15.
- Michaud, V., Randriamparany, T., Albina, E., 2013. Comprehensive phylogenetic reconstructions of African swine fever virus: proposal for a new classification and molecular dating of the virus. *PLoS One* 8, e69662.
- Miralles, R., Gerrish, P.J., Moya, A., Elena, S.F., 1999. Clonal interference and the evolution of RNA viruses. *Science* 285, 1745–7.
- Miralles, R., Moya, A., Elena, S.F., 2000. Diminishing returns of population size in the rate of RNA virus adaptation. *J. Virol.* 74, 3566–71.
- Modrich, P., Lahue, R., 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65, 101–33.
- Mukai, R., Hamatake, R.K., Hayashi, M., 1979. Isolation and identification of bacteriophage phi X174 prohead. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4877–81.
- Muller, H.J., 1964. The relation of recombination to mutational advance. *Mutat. Res.* 106, 2–9.
- Murcia, P.R., Baillie, G.J., Daly, J., Elton, D., Jervis, C., Mumford, J.A., Newton, R., Parrish, C.R., Hoelzer, K., Dougan, G., Parkhill, J., Lennard, N., Ormond, D., Moule, S., Whitwham, A., McCauley, J.W., McKinley, T.J., Holmes, E.C., Grenfell, B.T., Wood, J.L.N., 2010. Intra- and interhost evolutionary dynamics of equine influenza virus. *J. Virol.* 84, 6943–54.
- Novella, I.S., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E., Holland, J.J., 1995. Exponential increases of RNA virus fitness during large population transmissions. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5841–4.
- Ohta, T., 1992. The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* 23, 263–86.
- Orr, H. a, 2000. The rate of adaptation in asexuals. *Genetics* 155, 961–8.
- Pal, C., Maciá, M.D., Oliver, A., Schachar, I., Buckling, A., 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450, 1079–81.
- Pepin, K.M., Domsic, J., McKenna, R., 2008. Genomic evolution in a virus under specific selection for host recognition. *Infect. Genet. Evol.* 8, 825–34.
- Pepin, K.M., Lass, S., Pulliam, J.R.C., Read, A.F., Lloyd-Smith, J.O., 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.* 8, 802–13.
- Pepin, K.M., Wichman, H.A., 2008. Experimental evolution and genome sequencing reveal variation in levels of clonal interference in large populations of bacteriophage phiX174. *BMC Evol. Biol.* 8, 85.
- Pepin, K.M., Wichman, H.A., 2007. Variable epistatic effects between mutations at host recognition sites in phiX174 bacteriophage. *Evolution* 61, 1710–24.
- Pereira-Gómez, M., Sanjuán, R., 2015. Effect of mismatch repair on the mutation rate of bacteriophage phiX174. *Virus Evol.* 1, vev010.
- Pereira-Gómez, M., Sanjuán, R., 2014. Delayed lysis confers resistance to the nucleoside analogue 5-Fluorouracil and alleviates mutation accumulation in the single-stranded DNA bacteriophage x174. *J. Virol.* 88, 5042–9.
- Peris, J.B., Davis, P., Cuevas, J.M., Nebot, M.R., Sanjuán, R., 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185, 603–9.

- Pfeiffer, J.K., Kirkegaard, K., 2003. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7289–94.
- Pukkila, P.J., Peterson, J., Herman, G., Modrich, P., Meselson, M., 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics* 104, 571–82.
- Quiñones, A., Piechocki, R., 1985. Isolation and characterization of *Escherichia coli* antimutators. A new strategy to study the nature and origin of spontaneous mutations. *Mol. Gen. Genet.* 201, 315–22.
- Raney, J.L., DeLongchamp, R.R., Valentine, C.R., 2004. Spontaneous mutant frequency and mutation spectrum for gene A of phiX174 grown in *E. coli*. *Environ. Mol. Mutagen.* 44, 119–27.
- Riley, M.S., Cooper, V.S., Lenski, R.E., Forney, L.J., Marsh, T.L., 2001. Rapid phenotypic change and diversification of a soil bacterium during 1000 generations of experimental evolution. *Microbiology* 147, 995–1006.
- Roberts, J.D., Bebenek, K., Kunkel, T.A., 1988. The accuracy of reverse transcriptase from HIV-1. *Science* 242, 1171–3.
- Rokyta, D.R., Burch, C.L., Caudle, S.B., Wichman, H.A., 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J. Bacteriol.* 188, 1134–42.
- Rokyta, D.R., Joyce, P., Caudle, S.B., Wichman, H.A., 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* 37, 441–4.
- Rosche, W.A., Foster, P.L., 2000. Determining mutation rates in bacterial populations. *Methods* 20, 4–17.
- Sanjuán, R., Moya, A., Elena, S.F., 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8396–401.
- Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. *J. Virol.* 84, 9733–48.
- Sanz, A., Fraile, A., Gallego, J., Malpica, J., Garcia-Arenal, F., 1999. Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. *J. Mol. Evol.* 49, 672–81.
- Schaaper, R.M., 1998. Antimutator mutants in bacteriophage T4 and *Escherichia coli*. *Genetics* 148, 1579–85.
- Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B., Loeb, L.A., 2012. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14508–13.
- Schofield, M.J., Hsieh, P., 2003. DNA mismatch repair: molecular mechanisms and biological function. *Annu. Rev. Microbiol.* 57, 579–608.
- Seronello, S., Montanez, J., Presleigh, K., Barlow, M., Park, S.B., Choi, J., 2011. Ethanol and reactive species increase basal sequence heterogeneity of hepatitis C virus and produce variants with reduced susceptibility to antivirals. *PLoS One* 6, e27436.
- Shackelton, L.A., Holmes, E.C., 2006. Phylogenetic Evidence for the Rapid Evolution of Human B19 Erythrovirus. *J. Virol.* 80, 3666–9.
- Shackelton, L.A., Parrish, C.R., Truyen, U., Holmes, E.C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl. Acad. Sci. U. S. A.* 102, 379–84.
- Sharp, P.M., 1986. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. *Mol. Biol. Evol.* 3, 75–83.



- Shlomai, J., Polder, L., Arai, K., Kornberg, A., 1981. Replication of phi X174 dna with purified enzymes. I. Conversion of viral DNA to a supercoiled, biologically active duplex. *J. Biol. Chem.* 256, 5233–8.
- Sierra, M., Airaksinen, A., González-López, C., Agudo, R., Arias, A., Domingo, E., 2007. Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. *J. Virol.* 81, 2012–24.
- Sniegowski, P.D., Gerrish, P.J., Johnson, T., Shaver, a, 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays* 22, 1057–66.
- Sniegowski, P.D., Gerrish, P.J., Lenski, R.E., 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387, 703–5.
- Speyer, J.F., 1965. Mutagenic DNA polymerase. *Biochem. Biophys. Res. Commun.* 21, 6–8.
- Steinhauer, D.A., Domingo, E., Holland, J.J., 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122, 281–8.
- Stojic, L., Brun, R., Jiricny, J., 2004. Mismatch repair and DNA damage signalling. *DNA Repair (Amst)*. 3, 1091–101.
- Sun, L., Young, L.N., Zhang, X., Boudko, S.P., Fokine, A., Zbornik, E., Roznowski, A.P., Molineux, I.J., Rossmann, M.G., Fane, B.A., 2014. Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. *Nature* 505, 432–5.
- Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G., Lynch, M., 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18488–92.
- Suspène, R., Aynaud, M.-M., Koch, S., Padeloup, D., Labetoulle, M., Gaertner, B., Vartanian, J.-P., Meyerhans, A., Wain-Hobson, S., 2011. Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J. Virol.* 85, 7594–602.
- Suspène, R., Guétard, D., Henry, M., Sommer, P., Wain-Hobson, S., Vartanian, J.-P., 2005. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8321–6.
- Tanaka, S., Clemons, W.M., 2012. Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174. *Mol. Microbiol.* 85, 975–85.
- Tessman, E.S., Peterson, P.K., 1976. Bacterial rep- mutations that block development of small DNA bacteriophages late in infection. *J. Virol.* 20, 400–12.
- Tessman, E.S., Tessman, I., Pollock, T.J., 1980. Gene K of bacteriophage phi X 174 codes for a nonessential protein. *J. Virol.* 33, 557–60.
- Turelli, P., Mangeat, B., Jost, S., Vianin, S., Trono, D., 2004. Inhibition of hepatitis B virus replication by APOBEC3G. *Science* 303, 1829.
- Umemura, T., Tanaka, Y., Kiyosawa, K., Alter, H.J., Shih, J.W.-K., 2002. Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus)(1). *FEBS Lett.* 510, 171–4.
- Vartanian, J.-P., Guetard, D., Henry, M., Wain-Hobson, S., 2008. Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions. *Science* 320, 230–3.
- Vieira, V.C., Soares, M.A., 2013. The role of cytidine deaminases on innate immune responses against human viral infections. *Biomed Res. Int.* 2013,683095.
- Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., Andino, R., 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–8.

- Wagner G.P., Altenberg, L., 1996. Perspective: Complex adaptations and the evolution of evolvability. *Evolution* (N. Y). 50, 967–76.
- Wang, I.-N., Dykhuizen, D.E., Slobodkin, L.B., 1996. The evolution of phage lysis timing. *Evol. Ecol.* 10, 545–58.
- Wichman, H.A., Millstein, J., Bull, J.J., 2005. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170, 19–31.
- Wichman, H.A., Scott, L.A., Yarber, C.D., Bull, J.J., 2000. Experimental evolution recapitulates natural evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 355, 1677–84.
- Wichman, H. a, Brown, C.J., 2010. Experimental evolution of viruses: Microviridae as a model system. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 2495–501.
- Wielgoss, S., Barrick, J.E., Tenaillon, O., Wiser, M.J., Dittmar, W.J., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R.E., Schneider, D., 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. U. S. A.* 110, 222–7.
- Wood, N., Bhattacharya, T., Keele, B.F., Giorgi, E., Liu, M., Gaschen, B., Daniels, M., Ferrari, G., Haynes, B.F., McMichael, A., Shaw, G.M., Hahn, B.H., Korber, B., Seoighe, C., 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog.* 5, e1000414.
- Wright, S., 1982. The shifting balance theory and macroevolution. *Annu. Rev. Genet.* 16, 1–19.
- Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution, in: *Proceedings of the VI International Congress of Genetics* 1: pp. 356–66.
- Wright, S., 1931. Evolution in Mendelian Populations. *Genetics* 16, 97–159.
- Zeng, J., Wang, H., Xie, X., Li, C., Zhou, G., Yang, D., Yu, L., 2014. Ribavirin-resistant variants of foot-and-mouth disease virus: the effect of restricted quasispecies diversity on viral virulence. *J. Virol.* 88, 4008–20.
- Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., Gao, L., 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424, 94–8.
- Zheng, Q., 1999. Progress of a half century in the study of the Luria-Delbrück distribution. *Math. Biosci.* 162, 1–32.

## **SUPPLEMENTARY DATA**



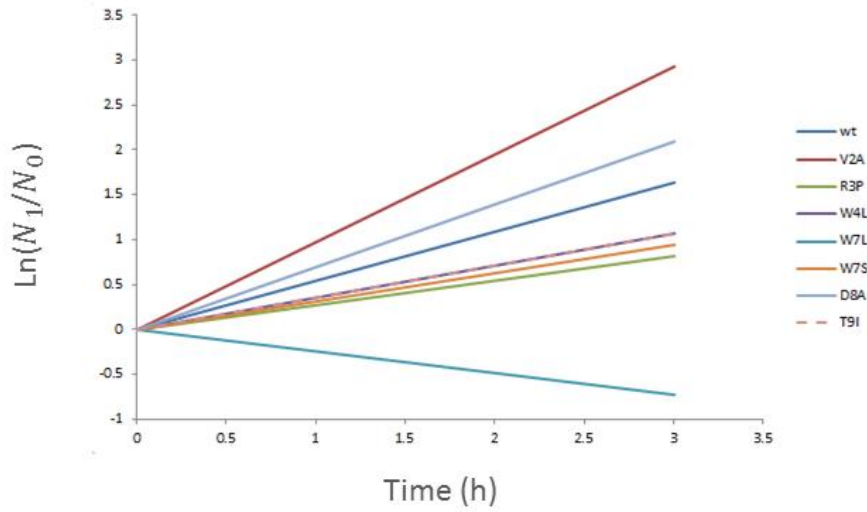
**Supplementary information 1: List of the primers used to create the N-terminal mutants of the lysis protein E by site directed mutagenesis.** Primers were designed in a way that the same forward primer can be used with the corresponding reverse primer to introduce different replacements on the E protein.

Primer name <sup>a</sup>	Sequence (5'→3') <sup>a,b</sup>	Replacement
<b>PX572FU_R</b>	<b>CACAAAGTCCAGCGTgCCATAAACGCAAGCCTCAAC</b>	V2A
<b>PX572_FU_F</b>	<b>GGATACCTCGCTTTCCTGCTCCTG</b>	None
V2E-	CACAAAGTCCAGCGTtCCATAAACGCAAGCC	V2E
V2G-	CACAAAGTCCAGCGTcCCATAAACGCAAGCC	V2G
R3H-	CACAAAGTCCAGtGTACCATAAACGCAAGCC	R3H
R3P-	CACAAAGTCCAGgGTACCATAAACGCAAGCC	R3P
R3L-	CACAAAGTCCAGaGTACCATAAACGCAAGCC	R3L
W4S-	GGTATCCCACAAAGTcGAGCGTACCATAAACG	W4S
W4L-	GGTATCCCACAAAGTcaGAGCGTACCATAAACG	W4L
T5I-	GGTATCCCACAAAaTCCAGCGTACCATAAACG	T5I
L6S-	GGTATCCCACgAAGTCCAGCGTACCATAAACG	L6S
W4_T5_L6+	CTCGCTTTCCTGCTCCTGTTG	None
W7L-	GAAAGCGAGGGTATCCaACAAAGTCCAGCG	W7L
W7S-	GAAAGCGAGGGTATCCgACAAAGTCCAGCG	W7S
D8V-	CAGGAAAGCGAGGGTAaCCCACAAAGTCCAGC	D8V
D8A-	CAGGAAAGCGAGGGTAgCCCACAAAGTCCAGC	D8A
D8G-	CAGGAAAGCGAGGGTAcCCCACAAAGTCCAGC	D8G
T9I-	CAGGAAAGCGAGGaTATCCCACAAAGTCCAGC	T9I
D8_T9+	CTCCTGTTGAGTTTATTGCTGCCG	None

<sup>a</sup> Primers not designed in this thesis, but used to introduce mutations, are in boldface.

<sup>b</sup> sites where the mutation is introduced are in lowercase.

**Supplementary information 2: Preliminary growth assays of the N-terminal region mutants showing delayed lysis in the presence of the 5-FU 10ng/μL.** The natural logarithm of the ratio between the final ( $N_1$ ) and the initial titer ( $N_0$ ) after three hours of incubation is represented.



**Supplementary information 3: List of the mutations introduced by site-directed mutagenesis in the different  $\phi$ X174 GATC-containing mutants.** The first column refers to the  $\phi$ X174 genome position according to the numeration by convention. The second and third columns identify the nucleotide present in the wild type and GATC-containing mutant respectively. The fourth column indicates the mutation type introduced. The fifth and sixth columns identify the amino acid and the gene location of the substitution respectively. The seventh and eighth columns indicate the primer name and their sequence.

Site	WT	GATC	Type of mutation	Amino acid <sup>b</sup>	Gene(s) <sup>a</sup>	Primer name <sup>c</sup>	Sequence (5'->3') <sup>c,d</sup>
74	A	C	Transversion	K494Q (I8)	A,A*,(K)	GATC74F GATC74R	TTGTTTACGAATTAATCGAAGTGGACTGCTG GCAGTAGTAATCCTGCTTgATCAAGATAATTTTTTCGAC
530	C	T	Transition	D47	D	GATC530F GATC530R	CGCTCGCTTGAGGCTTGCGTTTATGG ACGAGCACGAGAGCGaTCAGTAGCAATCCAAAC
868	T	A	Transversion	R7	J	GATC868F GATC868R	GTCTCCGACGCGTTGCGGAGG CAGGGCGAGCGGAGATCGTTTTTACCTTTAG
1030	T	G	Transversion	R10	F	GATC1030F GATC1030R	TCCCATCTTGCTTCTTCTGCTGG AAGGTCATGCGGGATcCGCTCGGC
1357	A	G	Transition	K119	F	<b>GATC1357G+</b> <b>GATC1357G-</b>	<b>GTTTCAGGGTTATTTGAATATCTATAACAAC TATTTTAAAG</b> <b>AAATGCTTAGGGATcTTATTGGTATCAGGGTTAATCTGTGCC</b>
1720	C	A	Transversion	R240	F	GATC1720F GATC1720R	CTGGCTATGATGTTGATGGAAGTACGACC ATGCCAGAGATTAGATcGCATGACAAGTAAAGG
2008	T	A	Transversion	R336	F	GATC2008F GATC2008R	CTAAGAAGTTTAAGATTGCTGAGGGTCCAG ACGAATCACCAGATcCGGAAAACATCCTTCATAG
2269	A	G	Transition	S423	F	GATC2269F GATC2269R	GATTGAGTGTGAGGTTATAACGCCGAAG TTTTATCACGAAGTCATGATCGAATCGCGAGTGG
2589	T	G	Transversion	S65	G	GATC2589F GATC2589R	GGTTGTTTCTGTTGGTCTGATATTGCTTTTATG TGATTAGCGGCTTGTATcGATGATCCATCTGAATGC
2646	C	T	Transition	D84	G	<b>GATC2646T+</b> <b>GATC2646T-</b>	<b>TGTTTGGTTCGCTTTGAGTCTTCTTCG</b> <b>GGCAAAAAATTTAGGATcCGGCATCAAAAGCAATATCAGCACC</b>
2898	T	C	Transition	I168	G	GATC2898F GATC2898R	AAGTGAGGTGATTTATGTTTGGTCTATTGCTG AAGTGGCTGGAGACAAATgATCTCTTAATAACCTGATTC
3234	T	C	Transition	L102	H	GATC3234F GATC3234R	GTCTGCCGCTGATAAAGGAAAGGATACTC TTGCCACCAAGTCCAACCAgATCAAGCAACTATCAG
3425	T	C	Transition	I165	H	GATC3425F GATC3425R	AGAGATTGCTGGCATTcAGTCCGGCG TTTTGAGTCTCATTCTTGCATCTCGGcATCTCTTTCTG
3746	C	A	Transversion	G272	H	<b>GATC3746A+</b> <b>GATC3746A-</b>	<b>TACTGCAAAGGATATTTCTAATGTCTGTCAC</b> <b>GCGCCAATATGAGAAGATCCATACCCTGATTCTG</b>
3896	C	A	Transversion	G322	H	<b>GATC3896A+</b> <b>GATC3896A-</b>	<b>CGTCAGGATTGACACCTCC</b> <b>GTTATTTCTAGACAAATTAGATcCCAATACCATCAGCTTTACCGTC</b>
4151	T	C	Transition	I57	A	<b>GATC4151C+</b> <b>GATC4151C-</b>	<b>CAGGGCGTTGAGTTCGATAATGGTG</b> <b>CATACGAAAAGACAGgATCTCTTCCAAGAGCTTGATGC</b>
4604	C	T	Transition	D208	A,A*	GATC4604F GATC4604R	CCCAATGCTTTGCGTGACTATTTTCG ATTATCATAAAACGCCTCTAATCGaTCGTGAGCCAAACG
4898	C	G	Transversion	P306	A,A*	GATC4898F GATC4898R	CGCAGGACGCTTTTTACGTTCTGG TGTAGCGAACTGCGATcGGCATACTGTAACC
5045	C	T	Transition	D355	A,A*	<b>GATC5045T+</b> <b>GATC5045T-</b>	<b>AGCTAAAGAATGGAACAAC TACTAAAAACC</b> <b>CCTAGACCTTTAGCAGCAAGaTCCATATCTGACTTTTTGTT</b>
5374	T	G	Transversion	V465G (R100)	A,A*,(B)	GATC5374F GATC5374R	GAGTTTTATCGCTTCCATGAC TGCAGGTTGGATcCGCCAATCATTTTTATC
3671	A	G	Transition	Q247	H	GATC3671F GATC3671R	TTAGTCTGAGGTTGACTTAGTTCATCAGC CCTTGCGAGTCATTTCTTTGATcTGGTCATTGGTAAAATACTGAC
3963	A	G	Transition		-	<b>GATC3963G+</b> <b>GATC3963G-</b>	<b>TTCTGTTCTTATTACCCTTCTGAATGTCACGC</b> <b>CCATAAAAAAGCCTCCAAGATcTGGAGGCATGAAAACATACAA</b>
3934/35	C/C	G/A	Transversion/ Transversion		-	GATC3934/35F GATC3934/35R	CCTCCAAATCTTGGAGGC CATGAAAACATACAATTGGGATcGTGTCAATCC
3940/43	A/T	G/C	Transition/ Transition		-	GATC3940/43F GATC3940/43R	CTAATTTGTCTAGGAAATAACCGTCAG GAAAACATACgATcGGGAGGGTGTCAATC

<sup>a</sup> The – symbol indicates an intergenic region.

<sup>b</sup> In overlapping genes also is shown the replacement on the protein where is produced.

<sup>c</sup> Primers not designed in this thesis, but used to introduce some mutations, are in boldface.

<sup>d</sup> Sites where the mutation is introduced are in lowercase.

**Supplementary information 4: Fluctuation test data (mean±SE) of  $\phi$ X174 WT and GATC mutants.**

Virus	Number of tests <sup>1</sup>	$N_0$ (pfu) <sup>2</sup>	$N_1$ (pfu, $\times 10^{-6}$ ) <sup>3</sup>	$P_0$ <sup>4</sup>	$\mu$ (s/n/r, $\times 10^6$ ) <sup>5</sup>	$r$ (h <sup>-1</sup> ) <sup>6</sup>
WT	15	211±25	0.28±0.05	0.52±0.05	1.58±0.44	8.11±0.29
20GATC	3	320±134	3.09±1.03	0.29±0.13	0.21±0.15	8.68±0.29
4GATC	3	117±16	0.11±0.00	0.75±0.05	1.07±0.16	7.38±0.26
4iGATC	3	128±3	4.06±1.10	0.74±0.11	0.03±0.01	8.27±0.11
A3963G	3	197±30	1.15±0.73	0.54±0.23	0.33±0.02	8.24±0.58
C3934G/C3935A	3	326±76	1.17±0.31	0.75±0.04	0.12±0.01	8.19±0.33
A3940G/T3943C	3	395±207	0.21±0.16	0.35±0.20	7.13±2.98	7.78±0.32

<sup>1</sup>Number of independent fluctuation tests performed, each test consisted of 24 independent cultures.

<sup>2</sup>Initial number of pfu per culture.

<sup>3</sup>Final number of pfu per culture.

<sup>4</sup>Fraction of 24 cultures showing no *gro87*-resistant plaques.

<sup>5</sup>For each test, the mutation rate was calculated as  $\mu = [-\ln P_0 / (N_1 - N_0)] \times 3 / T_S$ , where  $T_S = 7$ .

<sup>6</sup>Growth rate is given by  $r = \ln(N_1/N_0)/t$ , where  $t$  is the incubation time in hours.



**Supplementary information 5: Fluctuation test data (mean±SE) of  $\phi$ X174 WT and 20GATC mutant under stress conditions.**

Virus	Stress inductor	Number of tests <sup>1</sup>	$N_0$ (pfu) <sup>2</sup>	$N_1$ (pfu, $\times 10^{-6}$ ) <sup>3</sup>	$P_0$ <sup>4</sup>	$\mu$ (s/n/r, $\times 10^6$ )	$r$ (h <sup>-1</sup> ) <sup>6</sup>
WT <sup>7</sup>	None	15	211±25	0.28±0.05	0.52±0.05	1.58±0.44	8.11±0.29
WT	42°C	3	187±22	5.18±1.58	0.38±0.11	0.90±0.06	3.81±0.24
WT <sup>8</sup>	5-FU	6	273±58	0.01±0.00	0.53±0.11	20.9±4.12	1.50±0.11
20GATC <sup>7</sup>	None	3	320±134	3.09±1.03	0.29±0.13	0.21±0.15	8.68±0.29
20GATC <sup>9</sup>	42°C	3	133±19	0.38±0.15	0.71±0.13	3.72±0.69	2.23±0.19
20GATC <sup>9</sup>	5-FU	3	109±36	0.03±0.03	0.60±0.28	25.5±13.2	1.57±0.33

<sup>1</sup> Number of independent fluctuation tests performed, each test consisted of 24 independent cultures.

<sup>2</sup> Initial number of pfu per culture.

<sup>3</sup> Final number of pfu per culture.

<sup>4</sup> Fraction of 24 cultures showing no *gro87*-resistant plaques.

<sup>5</sup> For each test, the mutation rate was calculated as  $\mu = [-\ln P_0 / (N_1 - N_0)] \times 3 / T_S$ , where  $T_S = 7$ .

<sup>6</sup> Growth rate is given by  $r = \ln(N_1/N_0)/t$ , where  $t$  is the incubation time in hours.

<sup>7</sup> From Supplementary information 4.

<sup>8</sup> From Table 3.

<sup>9</sup> Corrected for differences in plating efficiency at 42°C and 5-FU

**Supplementary information 6: Growth assays of the WT (blue circles) and GATC evolved lines (red squares) in both hosts. Top) Growth assays of the evolved lines in *E. coli* at passages 25, 50 and 100. Bottom) Growth assays of the evolved lines in *S. typhimurium*. Founders are represented by filled symbols, and the evolved lines are represented by open symbols. Each evolved line was assayed in triplicate under the same condition in which the population was evolved together with their reference virus (i.e., founder). Bars represent the standard errors of the means.**

